

PROTEIN INTERACTIONS, MOLECULAR NETWORKS AND SYSTEMS BIOLOGY

Chairs: Benno Schwikowski and Denis Thieffry

Protein Interactions, Molecular Networks and Systems Biology	1
G-1. Phenotypic effects of network rewiring in regulatory hierarchies	4
G-2. Bringing order to disorder: comparative genomics and genetic interactions uncover three biologically distinct forms of protein disorder.....	5
G-3. Prediction of genetic interactions in yeast using machine learning	6
G-4. Prioritizing candidate genes by network analysis of differential expression using machine learning approaches	7
G-5. The human E3 ubiquitin ligase enzyme protein interaction network	8
G-6. Fast motif enumeration and clustering in integrated networks	9
G-7. Dynamic Deterministic Effects Propagation Networks: learning signalling pathways from longitudinal protein array data	10
G-8. Systematic mapping of multiple specificity in peptide recognition modules reveals new binding modes of protein domains.....	11
G-9. Reverse-engineering gene regulatory networks for the abiotic stress response in Arabidopsis thaliana.....	12
G-10. Predicting protein-protein interaction using mirror tree	13
G-11. Comparison of Bayesian networks and its extensions applied to the inference of regulatory networks.....	14
G-12. Insight into the mechanism of specific regulation of STAT proteins by atypical dual-specificity phosphatases.....	15
G-13. VoteDock: The consensus docking method for prediction of protein-ligand interactions	16
G-14. Inferring genetic regulatory networks with a hierarchical Bayesian model and a parallel sampling algorithm	17
G-15. Prediction of protein-protein interactions in the apoptosis pathway.....	18
G-16. Systematic network analysis.....	19
G-17. Model of trisporic acid synthesis in Mucorales shows bistable behaviour.....	20
G-18. Simulation of immune response to Mycobacterium tuberculosis using an agent-based model	21
G-19. Design, optimization and predictions of a coupled model of the cell cycle, circadian clock, DNA repair system, irinotecan metabolism and exposure control under temporal logic constraints22	
G-20. Adding structural information to the von Hippel-Lindau (VHL) tumor suppressor interaction network	23

G-21. Computational multiscale modeling of brain tumor growth.....	24
G-22. SLIDER: a generic metaheuristic for the discovery of correlated motifs in protein-protein interaction networks.....	25
G-23. The effective interactions between proteins via coarse-grained modeling.....	26
G-24. Predicting cellular perturbation effects and identifying causal perturbations using a message-passing based machine learning method.....	27
G-25. Towards better receptor-ligand prediction.....	28
G-26. Extracting and visualising protein-protein interfaces for data-driven docking.....	29
G-27. Visualizing spatiotemporal information in heterogeneous biological networks with Arena3D.....	30
G-28. Efficient learning of signaling networks from perturbation time series via dynamic nested effects models.....	31
G-29. Discovering patterns of differentially regulated enzymes in metabolic pathways of tumors.....	32
G-30. Extending the yeast metabolic network using a data integration approach.....	33
G-31. Predicting genetic interactions by quantifying redundancy in biochemical pathways.....	34
G-32. Hub protein interfaces and hot region organization.....	35
G-33. A functional-genomics fermentation platform to identify and optimize industrial-relevant properties of <i>Lactobacillus plantarum</i>	36
G-34. DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks.....	37
G-35. Combination of network topology and pathway analysis to reveal functional modules in human disease.....	38
G-36. Attacking interface & interaction networks.....	40
G-37. Pathway Projector: web-based zoomable pathway browser using KEGG Atlas and Google Maps API.....	41
G-38. Mapping the "Farnesylome": structure-based prediction of farnesyltransferase targets.....	42
G-39. Integrative analysis of gene expression and copy number variation data to elucidate a reference human gene network.....	43
G-40. Unknown player in modelling of signal transduction pathway.....	44
G-41. Assembly of protein complexes by integrating graph clusterings.....	45
G-42. How natural host species avoid CD4+ T cell depletion during SIV infection.....	46
G-43. Kinase-specific phosphorylation site prediction in <i>Arabidopsis thaliana</i>	47
G-44. BiologicalNetworks: enabling systems-level studies of host-pathogen interactions.....	48
G-45. Genome-wide identification of disease-related genes.....	49
G-46. Bioinformatics tools for the investigation of proteomics aspects involved in cardiovascular diseases: from biomarker discovery to protein-protein interaction networks.....	50
G-47. ABC database: the Analysing Biomolecular Contacts database.....	51
G-48. The core and pan metabolism in the <i>Escherichia coli</i> species.....	52

G-49. BioGraph: discovering biomedical relations by unsupervised hypothesis generation.....	53
G-50. Tuning noise propagation in a two-step series enzymatic cascade.....	54
G-51. The effect of interactome evolution on network alignment.....	55
G-52. Analyzing compounds' mode of action by biological relatedness of proteins.....	56
G-53. Logical modelling of the regulatory network controlling the formation of the egg appendages in <i>Drosophila</i>	57
G-54. A curated database of microRNA mediated feed-forward loops involving MYC as master regulator.....	58
G-55. The Biological Connection Markup Language (BCML): a SBGN compliant data format for visualization, filtering and analysis of biological pathways.....	59
G-56. Estimating the size of the <i>S.cerevisiae</i> interactome.....	60
G-57. Laplacian eigenmaps, penalized principal component regression on graphs, and analysis of biological pathways.....	61
G-58. Predicting metabolic pathways from bacterial operons and regulons.....	62
G-59. Inferring translationally active RNA binding proteins: mRNA interactions from polysomal profiling data with a Bayesian inference approach.....	63
G-60. DASS-GUI: a new data mining framework.....	64
G-61. EnrichNet: network-based gene set enrichment analysis.....	65
G-62. Interacting copy number alterations in breast cancer.....	66
G-63. Compression of mass spectral imaging data using discrete wavelet transform guided by spatial information.....	67
G-64. Learning ancestral polytrees for HIV-1 mutation pathways against nelfinavir.....	68
G-65. Ranking of genes from RNAi perturbation data.....	70
G-66. A network centered approach for genome wide data integration applied to Alzheimers disease prediction.....	71
G-67. How to efficiently hunt the disease causing gene.....	72
G-68. Study of the ultrasensitivity of the BCL-2 apoptotic switch.....	73
G-69. Improved genome-wide protein-protein interaction prediction and analysis of biological process coordination in <i>Escherichia coli</i>	74
G-70. A generic gene regulatory network reconstruction method: application to <i>Lactococcus lactis</i> MG1363.....	75
G-71. Dissecting the specificity of E2-E3 interaction in the ubiquitination pathway by a molecular docking approach.....	76
Author Index.....	77

G-1. Phenotypic effects of network rewiring in regulatory hierarchies

Bhardwaj N (1,*), Kim PM (2,3,4,5), Gerstein MB (1,6,7)

Tinkering with transcriptional regulatory networks has been used as a tool in the post-genome era to measure the tolerance of the cell in response to different kinds of perturbations. This tampering may include constructing synthetic model networks, knocking out and over-expression of genes. To obtain further insights into the organizational architecture of these networks, they have been rearranged into more intuitive structures like pyramidal hierarchies. In this study, we study the phenotypic effects of various kinds of network rewiring events in regulatory hierarchies in *E. coli* and Yeast.

Materials and Methods

To study first order effects, we build intuitive pyramidal hierarchies with the ‘chain-of-command’ pointing downward and superimpose the phenotypic effects of tampering with nodes and edges related to various levels of the hierarchy. We find that rewiring events that affect the upper levels have more dramatic effect on cell growth than those affecting lower level ones.

Results

Extending it further, second order effects involve allowing the hierarchies to change upon deletions and insertions of nodes/edges in the network. To study these effects, we reconstruct modified hierarchies in response to various perturbations that change the arrangement of genes in different levels. From this analysis, we find that more than the absolute number of changes in the altered hierarchies, it is the location and type of change that more accurately reflects the phenotypic effect of rewiring; upper level changes more adversely affect the cell fitness than lower level ones.

Discussion

Since, connectivity of regulators is not linearly related with its position in the hierarchy, these results show that approaches based on the position of regulators in hierarchies can give better predictions about their importance for cell fitness than those just based on their in- or out-degrees.

Presenting Author

Nitin Bhardwaj (nitin.bhardwaj@yale.edu)

Molecular Biophysics and Biochemistry, Yale University

Author Affiliations

1 Program in Computational Biology and Bioinformatics, 6 Department of Molecular Biophysics and Biochemistry, and 7 Department of Computer Science, Yale University, Bass 432, 266 Whitney Avenue, New Haven, CT 06520. 2 Terrence Donnelly Centre for Cellular and Biomolecular Research, 3 Banting and Best Department of Medical Research, 4 Department of Molecular Genetics, 5 Department of Computer Science, University of Toronto, Toronto, ON M5S 3E1, Canada.

G-2. Bringing order to disorder: comparative genomics and genetic interactions uncover three biologically distinct forms of protein disorder

Bellay J (1,=), Han S (2,3,=), Michaut M (2,3,=,), Costanzo M (2,3), Andrews BJ (2,3,4), Boone C (2,3,4), Bader GD (2,3,4,5), Myers CL (1), Kim PM (2,3,4,5)*

Intrinsically disordered regions are common in many proteins, especially in higher eukaryotes. Intrinsically disordered proteins, which have a large fraction of disordered residues, have been associated with a large variety of functions and many diseases. However, a detailed understanding of the role of disordered regions has remained elusive. In this study, we help to shine some light on this by systematically distinguishing different types of intrinsic disorder using genetic interactions and comparative genomics.

Materials and Methods

To investigate disordered proteins in the context of genetic interactions, we used the most comprehensive genetic interaction network for *S. cerevisiae* (Costanzo et al. 2010). To examine their evolutionary properties we investigated which disordered regions were also disordered in orthologous proteins across the yeast clade. We defined three distinct classes: regions of conserved disorder with quickly evolving sequences (“flexible disorder”), regions where disorder is conserved with highly constrained amino acid sequence (“constrained disorder”) and, lastly, non-conserved disorder.

Results

We observed that genes that have numerous genetic interactions often tend to encode proteins that have a higher percentage of disordered residues. We found that regions of conserved disorder are strongly predictive for harboring linear motifs and phosphorylation sites. Flexible disorder is closest to the canonical notion of protein disorder and is responsible for its association with signaling pathways and multi-functionality. Conversely, proteins high in constrained disorder are involved in RNA binding and protein folding. Non-conserved disorder appears to be largely non-functional sequence.

Discussion

We thus conclude that by analyzing evolutionary signatures of disordered regions we can distinguish three functionally distinct subdivisions that also correspond to biophysically distinct phenomena.

Presenting Author

Magali Michaut (magali.michaut@utoronto.ca)

University of Toronto, CCBR

Author Affiliations

(1) Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA (2) Terrence Donnelly Centre for Cellular and Biomolecular Research (3) Banting and Best Department of Medical Research (4) Department of Molecular Genetics (5) Department of Computer Science University of Toronto, Toronto, ON M5S 3E1, Canada (=) These authors contributed equally to this work

G-3. Prediction of genetic interactions in yeast using machine learning

Schrynemackers M (1,), Geurts P (1), Wehenkel L (1), Madan Babu M (2)*

The inference of the genetic interaction network of an organism is an important challenge in systems biology. The knowledge of these interactions is very useful to understand the functions of the genes and their products. In yeast *S.cerevisiae*, interactions subnetworks (E-MAPs) on four subsets of genes have been measured. For the time being, it remains however impossible to test experimentally the 18 millions potential interactions between the 6000 genes. In this work, we propose to use computational techniques based on machine learning to complete the experimentally confirmed interactions.

Materials and Methods

We proposed several strategies to transform this problem into one or several standard classification problems and we exploited two families of supervised learning algorithms: tree-based ensemble methods and support vector machines. We considered as inputs various feature sets, including chemo-genomic profiles, expression data, and morphological data. We validated the approach by using cross-validation on four available E-MAPs. We experimented with several protocols, including the completion of missing values in a given E-MAP and the prediction of interactions in one E-MAP from the others.

Results

Globally, the best results are obtained with support vector machines. Cross-validation shows that we are able to predict new interactions with a reasonable accuracy. As expected, predictions of interactions between genes from the training E-MAPs are more accurate than predictions of interactions between genes not present in the training set. Some E-MAPs are also much easier to predict than others. Among input feature sets, the chemo-genomic profiles are the most predictive followed by the morphological data, while we found that expression profiles are not informative.

Discussion

We have mostly focused on the prediction of negative interactions. Positive interactions are less frequent, which renders their prediction by machine learning techniques more challenging. We will now focus on these interactions. Future work will also consider the addition of other input features (e.g., interaction networks) or further methodological developments. Our ultimate goal is to make genome-wide predictions with our algorithms and to prioritise these predictions for an experimental validation.

Presenting Author

Marie Schrynemackers (Marie.Schrynemackers@ulg.ac.be)

Department of EE and CS & GIGA-R, University of Liège, Belgium

Author Affiliations

(1) Department of EE and CS & GIGA-R, University of Liège, Belgium (2) MRC Laboratory of Molecular Biology, Cambridge, UK

Acknowledgements

This poster presents research results of the Belgian Network BIOMAGNET (Bioinformatics and Modeling: from Genomes to Networks), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. MS is recipient of a F.R.I.A. fellowship and PG is a Research Associate of the F.R.S.-FNRS.

G-4. Prioritizing candidate genes by network analysis of differential expression using machine learning approaches

Nitsch D (1,), Moreau Y(1)*

Discovering novel disease genes is still challenging for diseases for which no prior knowledge - such as known disease genes or disease-related pathways - is available. Performing genetic studies frequently result in large lists of candidate genes of which only few can be followed up for further investigation. We have recently developed a computational method that identifies the most promising candidate genes by replacing prior knowledge by experimental data of differential gene expression between affected and healthy individuals (Nitsch et al. 2009, PlosOne 4(5): e5526).

Materials and Methods

To improve the performance of our prioritization strategy, we have extended our previous work by applying different machine learning approaches that identify promising candidate genes by determining whether a gene is surrounded by highly differentially expressed genes in a functional association or protein-protein interaction network. Our novel ranking strategies, scoring disease candidate genes, rely on network-based machine learning approaches, such as kernel ridge regression, heat kernel, Arnoldi kernel approximation, and a local measure based on a direct neighborhood measure.

Results

We have used a standard procedure in genetics that ranks candidate genes based solely on their differential expression level as a baseline. Our results showed that our four strategies could outperform this standard procedure and that the best results were obtained using the heat kernel approach leading to an average ranking position of 8 out of 100 genes, an AUC value of 92.3% and an error reduction of 52.8% relative to a standard genetic procedure which ranked the knockout gene in average at position 17 with an AUC value of 83.7%.

Discussion

In this study we could identify promising candidate genes using network-based machine learning approaches even if no knowledge is available about the disease or phenotype.

Presenting Author

Daniela Nitsch (Daniela.Nitsch@esat.kuleuven.be)

K.U.Leuven, ESAT-SCD

Author Affiliations

(1) K.U.Leuven, ESAT-SCD

G-5. The human E3 ubiquitin ligase enzyme protein interaction network

Kar G (1,), Keskin O (1), Nussinov R (2,3), Gursoy A (1)*

Ubiquitination is crucial for protein degradation in eukaryotic cells. It is achieved by a sequential cascade of ubiquitin-activating (E1), ubiquitin-conjugating (E2) and ubiquitin-ligating (E3) enzymes. E3 ligases mediate ubiquitin transfer from E2s to substrates and as such confer substrate specificity. Despite their essential role, current knowledge of their distinct biological functions and interaction partners is limited. Here, using structural data, efficient structural comparison algorithms and appropriate filters, we construct human E3 ubiquitin ligase enzyme protein interaction network.

Materials and Methods

We first compile the available structures for E2 and E3 proteins in the human ubiquitination pathway. Second, we apply our efficient protein-protein interaction prediction algorithm PRISM, which uses experimental (X-ray, NMR) protein-protein interface templates to model the interactions of E3 and E2 proteins in a large, proteome-scale docking strategy based on interface structural motifs. Then, we include flexibility and energetic considerations in our modeling using FiberDock, a flexible docking refinement server, to obtain more physical and biologically relevant interactions.

Results

Analysis of the human E3 ubiquitin ligase enzyme protein interaction network reveals important functional features and uncovers an a priori unknown E3-E2 and E3-E3 interactions. Our results show that E3 proteins such as Mdm2 and Huwe1 share E2 partners, which may explain how both Mdm2 and Huwe1 ubiquitinate p53 tumor suppressor protein for degradation. In addition, we discover the mode of E3-E3 interactions such as Mdm2-Siah1, which are known to enhance the degradation of the Numb protein.

Discussion

Here, for the first time, we constructed a structural human E3 ubiquitin ligase enzyme protein interaction network. Our strategy allows elucidation of both which E3s interact with which E2s in the human ubiquitination pathway and how they interact. In addition to identifying E3-E2 interactions, our strategy also reveals functionally-relevant E3-E3 interactions in the human ubiquitination pathway that were hitherto unknown.

Presenting Author

Gozde Kar (gkar@ku.edu.tr)
Koc University

Author Affiliations

(1) Koc University, Center for Computational Biology and Bioinformatics, and College of Engineering, Rumelifeneri Yolu, 34450 Sariyer Istanbul, Turkey (2) Basic Research Program, SAIC-Frederick, Inc., Center for Cancer Research Nanobiology Program, NCI-Frederick, Frederick, MD 21702, USA (3) Sackler Inst. of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

G-6. Fast motif enumeration and clustering in integrated networks

Audenaert P (1,), Van Parys T (2,3), Pickavet M (1), Demeester P (1), Van de Peer Y (2,3), Michoel T (2,3)*

Network motifs, small frequently occurring subgraphs, are the basic building blocks of complex networks of interactions between DNA, RNA, proteins and metabolites. Network motifs aggregate into larger, self-contained units reflecting the organization of a cell into discrete modules carrying out specific biological functions.

Materials and Methods

All algorithms were validated on an integrated network of more than 50,000 transcriptional regulatory, posttranslational regulatory, and physical protein-protein interactions in the yeast *S. cerevisiae*. Our algorithms were developed in a Java environment, running on stock hardware (2Gb of RAM and a dual core 2.4GHz CPU).

Results

Enumerating all motif instances is computationally intensive. We achieved a significant speedup by optimized branch-pruning allowing us to enumerate larger motifs than was previously possible. Motif clusters are defined as sets of nodes with a high number of motif instances between them, relative to their size. To find high-scoring clusters, we first defined PageRank-like cluster membership weights. Optimal weights are found by solving a multilinear set of equations and converted to motif clusters by taking a provably optimal weight cutoff. User interfaces are provided through Cytoscape.

Discussion

We developed novel algorithms and a Cytoscape user interface for identifying clusters of overlapping network motifs in large-scale networks composed of multiple interaction types spanning all levels of regulation in a cell.

Presenting Author

Pieter Audenaert (Pieter.Audenaert@intec.ugent.be)
Universiteit Gent

Author Affiliations

(1) Ghent University, Faculty of Engineering, Dept. of Information Technology (2) VIB, Dept. of Plant Systems Biology (3) Ghent University, Faculty of Sciences, Dept. of Plant Biotechnology and Genetics

Acknowledgements

We acknowledge funding of the IWT (SBO-BioFrame), IUAP P6/25 (BioMagnet), Ghent University (Multidisciplinary Research Partnership "Bioinformatics: from nucleotides to networks") and IBBT.

G-7. Dynamic Deterministic Effects Propagation Networks: learning signalling pathways from longitudinal protein array data

Bender C (1,*), Henjes F (1), Fröhlich H (2), Wiemann S (1), Korf U (1), Beißbarth T (3)

Network modelling has become an important tool to study cancer related molecular interactions, which is the basis for the development of novel drugs and therapies. We present a new signalling network reconstruction method using longitudinal protein array data after external perturbation, measured on Reverse Phase Protein Arrays, and infer a network among ERBB-signalling related proteins in a human breast cancer cell line.

Materials and Methods

Our method models the signalling dynamics by a boolean signal propagation mechanism that defines a sequence of state transitions for a given network structure. A likelihood score is proposed that describes the probability of our measurements given a particular state transition matrix. We identify the optimal sequence of state transitions via a Hidden Markov Model. Network structure search is performed by a genetic algorithm that optimises the overall likelihood of a population of candidate networks.

Results

We test our method on simulated networks and data and show its increased performance in comparison to other Dynamical Bayesian Network approaches. The reconstruction of a network in our real data results in several known signalling chains from the ERBB network, showing the validity and usefulness of our approach.

Discussion

Our method was able to reconstruct several signalling chains from the ERBB network known in the literature. Due to restrictions in the data, all edges must be carefully assessed and validated to proof their existence. Given that a major part of the inferred edges in our network are described in previous publications, it is a valuable tool for the generation of hypotheses on the underlying biological signalling processes in the biological system which was studied.

URL

<http://www.dkfz.de/mga2/ddepn>

Presenting Author

Christian Bender (c.bender@dkfz.de)
German Cancer Research Center

Author Affiliations

(1) German Cancer Research Center, Department of Molecular Genome Analysis, 69120 Heidelberg, Germany (2) Bonn-Aachen International Center for IT, Department of Algorithmic Bioinformatics, 53113 Bonn, Germany (3) University of Göttingen, Department of Medical Statistics, 37099 Göttingen, Germany

Acknowledgements

This project was supported through the Helmholtz Alliance on Systems Biology network SB-Cancer, through the German Federal Ministry of Education and Science (BMBF) project BreastSys in the platform Medical Systems Biology and the network IG Cellular Systems Genomics (01GS0864) in the platform NGFNplus. Further by the Clinical Research Group 179 through the DFG.

G-8. Systematic mapping of multiple specificity in peptide recognition modules reveals new binding modes of protein domains

Gfeller D (1,), Ernst A (1), Verschueren E (2), Vanhee P (2), Dar N (1), Serrano L (2), Sidhu SS (1), Bader GD (1), Kim PM (1)*

Accurately modeling binding specificity is a promising way to understand the properties of protein interaction domains such as SH3 or PDZ domains and predict new protein interactions, which can then lead to better understanding of signaling pathways. Current computational models in general disregard correlations between ligand residues. Using phage display results, we observe that this assumption is unrealistic. We develop a new model that provides a more detailed view of binding specificity and enable uncovering multiple specificity in protein interaction domains.

Materials and Methods

Experimental datasets: Phage display experiments for PDZ and SH3 domains (Tonikian et al, 2008, Tonikian et al 2009), as well as next-generation sequencing phage display data for Erbin PDZ mutants. Computational tools: Correlations between residues within an alignment of peptide interacting with a domain are measured using Mutual Information. Multiple specificity was detected by clustering the peptides and running a Maximum-Likelihood algorithm based on mixture models.

Results

We find that peptides interacting with a domain often exhibited highly significant correlations among their residues, which can be accurately modeled by considering different specificities for a domain. Our results reveal widespread occurrence of multiple specificity and enable us to identify structurally distinct binding modes. In particular, we find a new binding mode of PDZ domains, where an additional amino acid can be accommodated at the C-terminus of the ligand. This non-canonical binding mode predicts new and unexpected protein interactions.

Discussion

Identifying multiple specificity in peptide recognition modules both yields more accurate protein interaction predictions and provides new structural insights. While this kind of approaches have often been hampered by the limited number of experimental data, new sequencing technologies, combined with techniques such as phage display, are revolutionizing our capacities to experimentally detect protein interactions. Efficient computational tools, such as the one presented here, will be crucial to analyze these data and accurately model the binding specificity of protein interaction domains.

Presenting Author

David Gfeller (david.gfeller@isb-sib.ch)
Swiss Bioinformatics Institute

Author Affiliations

(1) Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario M5S 3E1, Canada (2) EMBL-CRG Systems Biology Unit, CRG-Centre de Regulacio Genomica, Dr. Aiguader 88, 08003 Barcelona, Spain

Acknowledgements

Canadian Institute of Health Research Swiss National Science Foundation

G-9. Reverse-engineering gene regulatory networks for the abiotic stress response in *Arabidopsis thaliana*

Vermeirssen V (), De Clercq I, Van Parys T, Van Breusegem F, Van de Peer Y*

In order to understand how the plant responds to its changing environment, we study gene regulatory networks under stress conditions in *Arabidopsis thaliana*. Gene regulatory networks describe gene expression as a function of regulatory inputs specified by interactions between regulators and DNA. Often these data are not available yet on a genome-wide scale and need to be predicted from more abundant transcription expression profiles. This is often a high-dimensional, underdetermined problem, further complicated by noise and the indirect regulator-gene relations present in the data.

Materials and Methods

In this study we used a probabilistic model to “reverse engineer” microarray expression profile data into biologically relevant gene regulatory networks. LeMoNe (learning module networks) uses gene expression profiles to extract ensemble gene regulatory networks of coexpression modules and their prioritized regulators. Through post-processing data integration we demonstrate the value of our method.

Results

We compiled gene expression profiles for 283 stress conditions in *Arabidopsis thaliana* and inferred through LeMoNe the stress gene regulatory networks. Through GO and AraCyc enrichment analysis, comparison with the gene-gene association network AraNet, cis-regulatory motif analysis and integration of other biological data, we show that LeMoNe identifies functionally coherent coexpression modules and prioritizes regulators that relate to similar biological processes as the module genes. Furthermore, we predict functional relationships for uncharacterized genes and regulators.

Discussion

LeMoNe is a highly qualified reverse-engineering algorithm for generating hypotheses on differential gene expression. In an era of functional genomics data abundance, it provides a means to put forward selected follow-up experiments for experimental validation. The knowledge gained in this systems biology study might lead to the identification of genes that are potential candidates for innovative molecular breeding strategies to develop stress-tolerant crops.

Presenting Author

Vanessa Vermeirssen (vanessa.vermeirssen@psb.vib-ugent.be)
VIB-UGent

Author Affiliations

Department of Plant Systems Biology, VIB, Ghent University, Technologiepark 927, 9052 Gent, Belgium

G-10. Predicting protein-protein interaction using mirror tree

Esmailbeiki R () , Nebel J-C*

Protein-protein interactions (PPI) are important for every process that takes place within a living cell. Although many experimental techniques can estimate PPI, their cost and inaccuracy have led to the development of computational methods. Among them, the popular Mirror Tree (MT) is based on the intuitive assumption that proteins which interact are under co-evolution pressure. Latest improvements attempt to remove the species evolution signal (Tree of Life) from co-evolution measurements. Since many enhancements of the MT have been proposed a comparative study is of great interest.

Materials and Methods

MT relies on measuring co-evolution between two protein chains. This is achieved by comparing their phylogenetic trees which are built from the multiple alignment of sequence homologues. We propose to use mutual information to quantify their similarity. Experiments were conducted using a dataset containing 179 E.coli proteins where 281 pairs are known to interact. For each protein, 65 orthologues were extracted to produce multiple alignments using ClustalW. Quantitative comparison is presented for recent MT implementations including ours.

Results

Evaluation of the different methods shows that usually removing Tree of Life (TOL) information produces more accurate results. Although the MI metric generates high sensitivity, its specificity is very low. Finally, our ClustalW based MTs outperform all other implementations.

Discussion

As a whole, our TOL-TM achieves the best compromise between sensitivity, specificity, and AUROC. Results suggest that ClustalW produces more accurate phylogenetic trees and Mutual Information is not a better measure for the comparison of these trees. In the future, we plan to extract Tree of Life information using advanced models from information theory and attempt to detect the specific amino acids which are involved in the protein-protein interaction.

Presenting Author

Reyhaneh Esmailbeiki (k0854421@kingston.ac.uk)

Kingston University London

Author Affiliations

Kingston University London

G-11. Comparison of Bayesian networks and its extensions applied to the inference of regulatory networks

Werhli A V(1,*)

The rapid increase in the availability and diversity of molecular biology data has instigated the investigation of the viability of learning biological network's structure from these data. Among the various methods available for the inference of the network's structure, Bayesian networks (BNs) are very attractive due to its flexibility and probabilistic nature. For instance, BNs permit interventional experiments and/or extra knowledge from diverse data sources to be properly included in the inference. The main aim of this work is to compare how these approaches compare with each other.

Materials and Methods

The main aim is to devise a BN structure from a set of data. In the first setting a simple BN is applied. The second investigated approach explicitly uses a set of interventions in the data, BN_i. In the third method an extra source of prior knowledge is included in the inference through an hierarchical Bayesian model, BN_e. In all the approaches the networks are sampled with a Markov Chain Monte Carlo. For comparison purposes both synthetic data and real data (obtained from flow cytometry experiments) are employed. The network investigated is the raf signalling pathway Sachs et al. 2005.

Results

In order to compare the methods the area under the ROC curve (AUC) is computed with two distinct settings. In one setting the directions are considered (DGE) and in the other they are all discarded (UGE). For all types of data, namely Gaussian, Netbuilder and flow cytometry, the BN_i and BN_e approaches significantly outperform the simple BN when considering the DGE metric. When the directions are discarded there is no significant differences among the methods in simulated data but there is a difference in flow cytometry data. The results are obtained from 5 different data sets.

Discussion

The methods BN_i and BN_e clearly outperform the simple BN specially when considering the edge directions, DGE. This suggests that the inclusion of extra-knowledge or interventions breaks up symmetries and resolve the direction of some edges that can not be resolved only on basis of the data due to the existence of equivalence classes. Therefore the use of more refined BN methods i.e. BN_i and BN_e is completely justified. As future research we plan to combine the addition of interventions and the extra knowledge and also test the methods in different networks' structures.

Presenting Author

Adriano V. Werhli (werhli@gmail.com)
Universidade federal do Rio Grande

Author Affiliations

(1) Centro de Ciências Computacionais, Universidade Federal do Rio Grande

Acknowledgements

This project is supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPQ.

G-12. Insight into the mechanism of specific regulation of STAT proteins by atypical dual-specificity phosphatases

Jardin C (1,), Sticht H (1)*

Protein phosphorylation at tyrosine residues is regulated by the coordinated activity of protein tyrosine kinases and tyrosine phosphatases. Dephosphorylation of Signal Transducer and Activator of Transcription (STAT) proteins by the atypical dual-specificity protein-tyrosine phosphatases (DSP) is highly specific. To date however, there is no experimental data that explain this specificity. Our goal was to get an explanation at the molecular level of the mechanism that regulates the phosphatase activity of these DSPs toward specific STATs using computational methods.

Materials and Methods

An approach combining docking and molecular dynamic (MD) simulations was used to investigate the binding between the PTP domain of the phosphatases and the SH2 domain of STATs, starting from the crystal structures. Docking conformations between VHR and STAT5 and between VH1 and STAT1 were generated systematically. For those docking solutions that agree with the mutagenesis data MD simulations were done. The interactions at the binding interfaces were compared in sequence alignments of the STAT aminoacid sequences to understand the origins of the specificity.

Results

In agreement with the mutagenesis data, we found a conformation for VHR interacting with STAT5 that involves a phosphorylated tyrosine of VHR, located apart from the catalytic site, and an arginine of the STAT5 SH2 domain that is conserved in all STAT members. A similar binding site was found for VH1 with STAT1. The binding interfaces are stable in MD simulations and present an intensive network of interactions between the phosphatases and STATs. The analysis of this network reveals interactions that are specific of each DSP-STAT complex.

Discussion

Our simulations reveal that the site of dephosphorylation at the phosphatase PTP domain is, at least alone, not enough to explain the specificity. The PTP domain exhibits another site, located apart from the catalytic one, which is responsible both for the binding to STAT SH2 domains and for the specificity. Thus, atypical DSP phosphatases possess both the catalytic and recognition sites in the same polypeptide chain that forms the PTP domain. We thus suggest that the two sites play complementary roles in a two-sites/two-steps mechanism for the phosphatase activity of atypical DSPs on STATs.

Presenting Author

Christophe Jardin (Christophe.Jardin@biochem.uni-erlangen.de)
Bioinformatics, Institute for Biochemistry, University of Erlangen-Nuremberg

Author Affiliations

Bioinformatics Institute for Biochemistry Emil-Fischer Center University of Erlangen-Nuremberg

Acknowledgements

DFG Sonderforschungsbereich 796 Rechenzentrum Erlangen (RRZE)

G-13. VoteDock: The consensus docking method for prediction of protein-ligand interactions

Plewczynski D (1,), Łaźniewski M (1), von Grothuss M (2), Rychlewski L (3), Ginalski K (1)*

Molecular recognition plays a fundamental role in all biological processes that is why great efforts have been made to understand and predict protein-ligand interactions. Finding a molecule that can potentially bind to a target protein is particularly essential in drug discovery but experimental techniques are still expensive and time-consuming. Thus in silico tools are frequently used to screen molecular libraries in order to identify new lead compounds [Lee2009]. If information about protein structure is known, also various protein-ligand docking programs can be used [Perola2004][Li2010].

Materials and Methods

The aim of docking procedure is to predict correct poses of ligand in the binding site of the protein, as well as to score them according to the strength of interaction in a reasonable time frame. The purpose of our studies was to present the novel consensus approach to predict both protein-ligand complex structure and its binding affinity. Our method, called VoteDock, uses as the input the results from seven docking programs (Surflex, LigandFit, Glide, GOLD, FlexX, eHiTS and AutoDock) that are widely used by community, as most of them are part of popular modeling software packages.

Results

Those programs were evaluated on the extensive dataset of 1300 protein-ligands pairs from refined set of PDBbind database [Wang2004], for which the structural and binding affinity data are available. We compared independently ability of proper scoring by calculating Pearson correlation between docking score and experimental binding affinities, and posing by measuring RMSD of obtained conformations with native structure. That procedure allows us to compare performance of individual programs with that of VoteDock.

Discussion

In most cases our consensus-based method was able to dock properly about 20% of pairs more than docking methods in average, and more than 10 % of pairs more than single best program. Also drop in RMSD of top scored conformation can be observed, with value 0.5Å lower than that of best individual program, namely GOLD. Similar increase in overall docking accuracy can be also observed for subsets created based on PDBbind, that explore various physico-chemical properties of ligands, their size or hydrophobic potential. Finally, we are able to boost the Pearson correlation of the predicted binding a

URL

<http://dock.bioinfo.pl>

Presenting Author

Dariusz Plewczynski (darman@icm.edu.pl)

Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw

Author Affiliations

(1) Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, Pawińskiego 5a Street, 02-106 Warsaw, Poland, (2) Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA, (3) BioInfoBank Institute, Poznan, Poland

Acknowledgements

Calculations were performed at the Interdisciplinary Center for Mathematical and Computational Modelling. This work was supported by Polish Ministry of Science and Higher Education N301 159735 grant.

G-14. Inferring genetic regulatory networks with a hierarchical Bayesian model and a parallel sampling algorithm

Mendoza M R (1), Werhli A V (2,)*

Bayesian Networks (BNs) are a particularly promising tool for inferring the structure of regulatory networks. It is usually assumed that a molecular biological system may be represented by a single regulatory network. In a previous work a method for reconstructing the regulatory structure of a network considering that its active parts can differ under different experimental conditions was proposed. Unfortunately the proposed method when sampled with a MCMC suffered problems of mixing and convergence. In the present work we propose a Parallel Sampling Algorithm to overcome this problem.

Materials and Methods

In order to integrate information from various different data sets obtained under distinct experimental conditions we use an hierarchical Bayesian model. The model is composed of data sets, a hypernetwork which is not directly associated with the data but encourages the graphs to be similar and the hyperparameters which associate each data set with the hypernetwork. All components of the model are sampled with a Metropolis Coupled Markov Chain Monte Carlo which involves the parallel execution of multiple Markov chains, some of which are heated, and a state swap proposal between chains.

Results

The method was tested using data simulated from a linear Gaussian distribution and with Netbuilder tool, as well as with real data collected in cytometry flow experiments. The approach was applied in the integration of five data sets: one composed of totally random values, representing an unsuccessful experiment, and four others of the same nature, either simulated or measured data. For synthetic data the method correctly identified the random data and properly converged. It also outperformed other options of using the various data sets. The convergence was not observed in real data though.

Discussion

The method was tested using data simulated from a linear Gaussian distribution and with Netbuilder tool, as well as with real data collected in cytometry flow experiments. The approach was applied in the integration of five data sets: one composed of totally random values, representing an unsuccessful experiment, and four others of the same nature, either simulated or measured data. For synthetic data the method correctly identified the random data and properly converged. It also outperformed other options of using the various data sets. The convergence was not observed in real data though.

Presenting Author

Adriano V. Werhli (werhli@gmail.com)
Universidade federal do Rio Grande

Author Affiliations

(1) Instituto de Informática, Universidade Federal do Rio Grande do Sul. (2) Centro de Ciências Computacionais, Universidade Federal do Rio Grande.

Acknowledgements

This research is supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ).

G-15. Prediction of protein-protein interactions in the apoptosis pathway

Acuner Ozbabacan SE (1,), Gursoy A (1), Keskin O (1), Nussinov R (2,3)*

Prediction methods of protein-protein interactions have been intensively developed by scientists for the past years. From the structural point of view, proteins interact through their interfaces and the structure of these regions play crucial roles in the determination of interaction affinities. Our group constructed a server named PRISM (PProtein Interactions by Structural Matching), which can predict the protein-protein interactions in a given target set by structurally matching the target structures with a template interface set. In this work, the aim is to predict new protein-protein interactions in the apoptosis pathway by using PRISM server and the default template set of this server.

Materials and Methods

The template set includes the interfaces obtained from the complexes that are known to interact. It was prepared by collecting all of the interfaces in the PDB (in 2006) and clustering them structurally. The homologous and nonbiological interfaces in the clusters were eliminated and 158 interfaces were left as the nonredundant representative template set. The target set is composed of the proteins/genes in the apoptosis pathway with available nonhomologous PDB structures. Protein-protein interactions, which are thought to complement the apoptosis pathway, are predicted and corresponding interaction energies are calculated by using the PRISM server and FiberDock server, respectively.

Results

Many interactions were predicted by PRISM but we concentrated only on around 2000 interactions, which are redundant and contain interactions that are already in the apoptosis pathway, having energy values lower than -10 kcal/mol. One of the lowest interaction energies, therefore one of the most favorable interactions, was found for CASP7-CASP8 interaction, which is already in the apoptosis pathway. These interactions were found by using 85 template interfaces out of 158, for structural matching.

Discussion

The prediction algorithm of PRISM server is based on the structural resemblance of the potential pairs in the target set to the templates which are interfaces of the interacting pairs (complexes) taken from the PDB. Hence, in addition to the prediction and analysis of candidate new interactions in the apoptosis pathway, a structural dimension has been introduced since the interactions are found by structural matching.

Presenting Author

Ece S. Acuner Ozbabacan (sozbabacan@ku.edu.tr)
Koc University

Author Affiliations

(1)Center for Computational Biology and Bioinformatics and College of Engineering, Koc University Rumelifeneri Yolu, 34450 Sariyer Istanbul, Turkey (2)Basic Science Program, SAIC-Frederick, Inc. Center for Cancer Research Nanobiology Program NCI-Frederick, Frederick, MD 21702 (3)Sackler Inst. of Molecular Medicine Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

G-16. Systematic network analysis

*Agarwal S (1, *), Villar G (1), Jones N (1)*

Many real-world systems, including a number of biological ones, are naturally represented as networks, and a variety of measures exist for network analysis. However, studies of networks typically employ only a small, largely arbitrary subset of these, and the lack of a systematic comparison makes it unclear which metrics are most appropriate for a given task.

Materials and Methods

We present a framework for systematic analysis of networks and network metrics, and use it to analyse a large set of real networks from a wide variety of domains, utilising several hundred network metrics or summary statistics thereof.

Results

We demonstrate the utility of the framework for finding redundant metrics, classifying networks, and selecting relevant features of networks, particularly in the context of evolutionary phylogenies on biological networks.

Discussion

Our approach provides a generic, data-driven way of thinking about networks of different kinds and relating structural and topological features of networks to phenotypic properties. In particular, it is useful in inferring what properties of biological networks such as metabolic or protein interaction networks are most relevant in the context of evolutionary pressures, and thus in helping to motivate models for network evolution and function.

Presenting Author

Sumeet Agarwal (sumeet.agarwal@physics.ox.ac.uk)

University of Oxford

Author Affiliations

1 University of Oxford

Acknowledgements

Systems Biology Doctoral Training Centre Oxford, BBSRC, EPSRC, Clarendon Fund

G-17. Model of trisporic acid synthesis in Mucorales shows bistable behaviour

Werner S (), Vlais S, Schroeter A, Schuster S*

The synthesis pathway of trisporic acid involved in sexual interactions in Mucorales is very interesting because it includes an exchange of intermediates between the two mating types. Trisporic acid itself enhances its synthesis rate by a positive feedback loop. It is experimentally observed that a switch from the production of a low-level concentration of trisporic acid to a high-level state occurs if the fungi approach the opposite mating type. This bistable behaviour is modelled in this work.

Materials and Methods

For the ODE model the pathway of trisporic acid production is shrunk to 19 essential species, subdivided into 13 metabolites and 6 enzymes, as well as 41 reactions (conversion, diffusion, degradation and inflow of metabolites). The reactions obey different kinetics: mass action kinetics, Michaelis-Menten kinetics, constant flux, and Hill kinetics. To test the model regarding bistable behaviour we defined an artificial inflow and outflow of trisporic acid into the growth medium represented by a constant addition resp. constant removal of trisporic acid over a fixed period.

Results

In response to mimicking the approach of fungi of the opposite mating type by an artificial inflow and outflow of trisporic acid, the model reveals bistable behaviour. Above a threshold value of trisporic acid constantly fed to the growth medium over a certain period of time the system switches to the second stable state.

Discussion

Due to lacking experimental data concerning metabolite concentrations and (kinetic) parameter values in the trisporic acid production pathway these have been chosen qualitatively. A more quantitative approach by integrating measured data is pending.

URL

<http://pinguin.biologie.uni-jena.de/bioinformatik/>

Presenting Author

Sarah Werner (sarah.werner@uni-jena.de)
Friedrich-Schiller-University Jena

Author Affiliations

Friedrich-Schiller-University Jena, Department of Bioinformatics, Germany

Acknowledgements

This project is funded by the Jena School for Microbial Communication (JSMC) and the German Research Foundation (DFG)

G-18. Simulation of immune response to *Mycobacterium tuberculosis* using an agent-based model

Galvão V (1,), Miranda JGV(1), Andrade RFS (1)*

The knowledge about the progression and control of tuberculosis is important to identify the factors responsible for the migration of immune cells to the affected tissue and the organization of these cells into a granuloma. To clarify these mechanisms, an experiment identified the kinetics of different cell types during the infectious process. Based on the results of this experimental study, we have developed an agent-based model to verify some hypothesis of the immune response to *Mycobacterium tuberculosis*.

Materials and Methods

Our model includes six different cell types: *M. tuberculosis*, macrophage, CD4 T cell, B cell, pneumocyte, and necrosis. CD4 T and B cells can be found in two states: inactivated and activated. Additionally, macrophages can be found in four states: inactivated, infected, chronically infected, and activated. It was constructed on a three-dimensional cubic lattice. The lattice consists of discrete sites, such that each one of them can be void or occupied by one type of autonomous agent.

Results

This computational model can reproduce some cellular properties including migration, activation, phagocytosis and death. Also, there are the bacillus replication and necrosis formation. This way, it reproduces the kinetics of *M. tuberculosis*, macrophages, CD4 T cells, B cell, and necrosis formation. Our results were quantitatively and qualitatively compared with experimental data, leading to a good agreement.

Discussion

Our results suggest that a multi-agent-based approach is a suitable instrument for investigating and modeling the cellular interaction in the progression of disease. Additionally, our model rules can simulate the immune response to *Mycobacterium tuberculosis*.

Presenting Author

Viviane Galvão (vivgalvao@gmail.com)

Universidade Federal da Bahia

Author Affiliations

(1) Universidade Federal da Bahia

Acknowledgements

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB).

G-19. Design, optimization and predictions of a coupled model of the cell cycle, circadian clock, DNA repair system, irinotecan metabolism and exposure control under temporal logic constraints

De Maria E (1,), Fages F (1), Rizk A (1), Soliman S (1)*

In systems biology, the number of available models of cellular processes increases rapidly, but re-using models in different contexts or for different questions remains a challenging issue. In this work, we study the coupling of the mammalian cell cycle, the circadian clock, the p53/Mdm2 DNA-damage repair system, the metabolism of irinotecan and the control of cell exposure to it. We show how the formalization of experimental observations in temporal logic with numerical constraints can be used to compute the unknown coupling kinetics parameter values agreeing with experimental data.

Materials and Methods

We consider a model of the mammalian cell cycle proposed by Novak and Tyson (2004), a model of the circadian clock developed by Leloup and Golbeter (2003), a model of proteins p53/ Mdm2 introduced by Ciliberto et al. (2005), and a model of the irinotecan metabolism by Dimitrio (2007). After encoding the models in the rule-based language of Biocham, we add suitable linking rules and we use an original method based on temporal logic constraint solving and optimization techniques to find parameter values for the new rules so that some expected properties are verified by the coupled model.

Results

The coupling of the composite models has been achieved and irinotecan exposure times and maximum amount that maintain toxicity low for healthy cells have been found (in Biocham). The predictive power of the coupled model was tested with respect to a limited set of mutants of the circadian clock genes. In the case of genes knock outs, we succeeded in considering temporal logic constraints over different traces corresponding to the mutations of different genes.

Discussion

Although preliminary, the results obtained are very encouraging for our coupling method. They pointed out that mass-entrained models of the cell-cycle have a limited possibility of entrainment by the circadian clock and that non mass-entrained models should be preferred in future studies. They also showed that the p53/Mdm2 model of Ciliberto et al. should be improved to introduce a threshold above which the DNA is no longer repaired. Finally, a PK/PD model of irinotecan in the body should be added to link the injection law to the cell exposure model and optimize the injection law directly.

URL

http://contraintes.inria.fr/supplementary_material/TCS-CMSB09/

Presenting Author

Elisabetta De Maria (elisabetta.de_maria@inria.fr)

INRIA Paris-Rocquencourt

Author Affiliations

INRIA Paris-Rocquencourt

Acknowledgements

This work was supported by the EU FP6 STREP project TEMPO on cancer chronotherapies and is now supported by the ERASysBio project C5Sys concerning circadian and cell cycle clock systems in cancer. We acknowledge fruitful discussions with the partners of this project, in particular with Francis Levi, Jean Clairambault and Annabelle Ballesta.

G-20. Adding structural information to the von Hippel-Lindau (VHL) tumor suppressor interaction network

*Leonardi E (1, 2, *), Murgia A (2), Tosatto S (1)*

The von Hippel-Lindau (VHL) tumor suppressor gene is a protein interaction hub, controlling numerous genes implicated in tumor progression. Here we focus on structural aspects of protein interactions for a list of 35 experimentally verified protein VHL (pVHL) interactors.

Materials and Methods

Regions of pVHL sequence where several proteins interact were inferred to contain a binding interface. For each class of interactors, the Pfam and CATH classification codes were compared to search for common domains or architectures. For each interactor, the GO molecular function and cellular localization allowed us to determine functions mediated by each interface region with different space and time patterns. A potential linear motif which could mediate the interaction with pVHL was searched on the protein sequences starting from the HIF1a peptide.

Results

The modular nature of pVHL becomes apparent from the subdivision in three interaction interfaces corresponding to processing, substrate recognition and localization. These highlight various protein interaction types, namely domain-domain (interface A) and domain-peptide (interface B), with interface C being less clear. Structural characterization of the putative interface B interaction peptides yielded both a complete list of hypothetical interaction motifs and the intriguing possibility for the pVHL N-terminus to auto-inhibit substrate recognition after phosphorylation.

Discussion

Our findings, rationalizing pVHL function at the molecular level through experimental data from the literature, can serve as a gold standard for the analysis of other putative interaction partners. They can be useful to design and interpret specific experimental interaction assays for the 200 or more remaining putative pVHL interaction partners. Addressing these in different time and space points will provide the ultimate validation for this complex protein.

Presenting Author

Emanuela Leonardi (emanuela.leonardi@unipd.it)
Dept. of Biology, University of Padua

Author Affiliations

Department of Biology, University of Padua, 35121 Padua, Italy Department of Pediatrics, University of Padua, 35128 Padua, Italy

G-21. Computational multiscale modeling of brain tumor growth

Schütz T A (1,2,), Toma A (1), Becker S (1), Mang A (1), Buzug T M (1)*

The prognosis for patients with malignant brain tumors remains poor. Consequently, a better understanding of the complex mechanisms underlying tumor progression is of key interest to design better treatment strategies. A powerful tool to test hypotheses on tumor evolution for individual patients and thus, to improve the understanding of the disease, is the mathematical modeling of tumor growth. A novel multi-scale approach for coupling our existing hybrid model for brain tumor progression on a microscopic scale with a molecular model based on ordinary differential equations (ODEs) is proposed.

Materials and Methods

The present work introduces a mathematical model describing the avascular tumor progression on a microscopic and molecular scale. More precisely, a hybrid model is used, which additionally considers the nutrient concentration described in terms of reaction-diffusion equations. The solution is computed using the finite element method (FEM) and itself influences a molecular network on the subcellular level. This network is defined by a system of ODEs whose solution reflects the tumor cell processes. The cellular automata method is used to simulate chemotactic motility and necrosis.

Results

Visual inspection of the results demonstrates the plausibility of the implemented model. The computed dynamics of cancerous cells follows the well known exponential increase in cell population during the early stage of cancer progression. For the subsequent cancer growth the model displays structures exhibiting a necrotic core, a rim of quiescent cells surrounded by dividing cells.

Discussion

We report first results for a multiscale simulation framework of brain cancer dynamics. The model depicts the expected characteristic spatial patterns. The central aspect of this work is the coupling of the cellular model including the nutrient concentration with a molecular network on the subcellular scale. Using the FEM to solve for the nutrient concentration it is possible to provide an efficient solution. The standard processes for cancerous cells are controlled by molecular reactions and protein concentrations. The model furthermore incorporates chemotaxis.

Presenting Author

Tina Anne Schütz (schuetz@imt.uni-luebeck.de)

Institute of Medical Engineering, University of Lübeck; Graduate School for Computing in Medicine and Life Sciences, University of Lübeck

Author Affiliations

1 Institute of Medical Engineering, University of Lübeck 2 Graduate School for Computing in Medicine and Life Sciences, University of Lübeck

Acknowledgements

Tina Anne Schütz is supported by the Graduate School for Computing in Medicine and Life Sciences funded by Germany's Excellence Initiative [DFG GSC 235/1]. Alina Toma and Stefan Becker are financially supported by the European Union and the State Schleswig-Holstein (Competence Center for Technology and Engineering in Medicine (TANDEM); grant no. 122-09-024).

G-22. SLIDER: a generic metaheuristic for the discovery of correlated motifs in protein-protein interaction networks

Boyen P (1,), Van Dyck D (1), Neven F (1), van Ham RCHJ (2), van Dijk ADJ (2)*

Correlated motif mining (CMM) is the problem of finding overrepresented pairs of patterns, called motifs, in sequences of interacting proteins. Algorithmic solutions for CMM thereby provide a computational method for predicting binding sites for protein interaction. Large-scale biological networks describing interactions between proteins are available now for several organisms. Such data show which proteins interact, but provides no insight into how interactions are encoded in protein sequences.

Materials and Methods

We adopt a motif-driven approach where the support of candidate motif pairs is evaluated in the network. We present the generic metaheuristic SLIDER which uses steepest ascent with a neighborhood function based on sliding motifs and employs the Chi-square-based support measure. We implemented these methods using Java. We validated our method on simulated data as well as the protein-protein interaction networks for yeast and human. The SLIDER-implementation and the data used in the experiments are available on <http://bioinformatics.uhasselt.be>.

Results

We experimentally establish the superiority of the Chi-square-based support measure over other support measures. Furthermore, we obtain that CMM is an NP-hard problem for a large class of support measures (including Chi-square) and reformulate the search for correlated motifs as a combinatorial optimization problem. We show that SLIDER outperforms existing motif-driven CMM methods and scales to large protein-protein interaction networks. For our results on the human PPI-network, we find significant overlap with the interface (the physically interacting surface areas of the proteins).

Discussion

This work establishes an adequate support measure and determines the complexity of the motif-driven CMM problem. SLIDER outperforms other motif-driven CMM algorithms and shows promising behavior on real-world PPI-networks. Directions for future work include investigating candidate generation for motif pairs. A detailed comparison with interaction-driven approaches should be done. Maybe ideas from both paradigms can be combined into a hybrid method. We used the simple model of (l,d)-motifs and our results suggest that this model suffers from false positives caused by indirect interactions.

URL

<http://bioinformatics.uhasselt.be>

Presenting Author

Peter Boyen (peter.boyen@uhasselt.be)
Hasselt University

Author Affiliations

(1) Hasselt University and Transnational University of Limburg (2) Applied Bioinformatics - Plant Research International, Wageningen UR

Acknowledgements

Peter Boyen is funded by a Ph.D grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). A. D. J. van Dijk is supported by an NWO (Netherlands Organisation for Scientific Research) VENI grant (863.08.027). This research is supported by the BioRange programme (SP 2.3.1) of the Netherlands Bioinformatics Centre (NBIC), which is supported through the Netherlands Genomics Initiative (NGI) and the Research Programme of the Research Foundation Flanders (FWO) (G030607). This work was also sponsored by the BiG Grid project for the use of computing and storage facilities, with financial support from NWO.

G-23. The effective interactions between proteins via coarse-grained modeling

Pool R (1,), Feenstra KA (1), Heringa J (1)*

The knowledge on protein-protein interactions (PPIs) is important since they play an important role in many cellular processes. Aggregation in solution can occur by driven processes or via self-assembly. Here the effective interaction between the monomers as a function of molecular separation (potentials of mean force (PMFs)) is of vital importance. Currently, a systematic exploration of a large number of PPIs is practically unfeasible due to the large amount of time needed for experiments or computations. Here we propose a coarse-grained approach that yields PMFs within reasonable times.

Materials and Methods

We use state-of-the-art molecular simulation techniques (hybrid Molecular Dynamics-Monte Carlo), applied to coarse-grained systems using the MARTINI force field. Sampling efficiency is further accomplished by reducing the system size to the volume including the interacting interfaces as well as the volume between them. All interacting species are simulated explicitly. Forces acting between the interfaces are averaged. The resulting force profile is then integrated to yield a PMF.

Results

Using our coarse-grained and reduced system approach we can successfully determine PMFs for given protein pairs. As expected the PMFs are attractive, The obtained PMFs agree with simulation results from complete coarse-grained systems that include the completely dissolved protein structures. In addition, the coarse-grained simulation results show a considerable agreement with more realistic and more expensive full atom simulations.

Discussion

The results obtained so far are promising, although we should be aware of physical phenomena we simply cannot capture using a coarse-grained model. The combination of the coarse-grained molecular model and the minimal system size is expected to enable large scale exploration of the possible existence of protein-protein interactions in currently available databases.

Presenting Author

René Pool (r.pool@few.vu.nl)

Centre for Integrative Bioinformatics Vrije Universiteit (IBIVU)

Author Affiliations

Centre for Integrative Bioinformatics Vrije Universiteit Amsterdam (IBIVU)

Acknowledgements

This work is funded by the Netherlands Bioinformatics Centre (NBIC): BioRange 2.3.1.

G-24. Predicting cellular perturbation effects and identifying causal perturbations using a message-passing based machine learning method

Hulsman M (1,), Reinders MJT*

Perturbations in a cell, as a result of a changing environment or a gene knockout, are propagated through different physical links within the cellular network. Changes in expression levels of genes give an impression of the resulting effects of a given perturbation. Together with direct and indirect information about the physical links, these observations can be used to unravel the cellular network and the network information flow. We present a method that simultaneously reconstructs and simulates a cellular network capable of propagating perturbations throughout the network.

Materials and Methods

The network reconstruction is based on a machine learning method that takes various data sources, such as protein-protein interaction and protein-DNA interaction, into account to predict the probability of an edge propagating the perturbation from its parent node to its child node. The network simulation is realized by a novel propagation model that passes messages throughout the network. The combined message passing scheme and learned network generates probabilities expressing whether a gene shows a change in expression given a perturbation.

Results

The proposed method has been applied to learn from gene knock-out data. It generalizes to unseen knockouts, i.e. one is able to predict genes that change expression levels after a hypothetical gene knockout as well as propose a hypothetical gene knockout that maximally explains an observed set of changed gene expression levels. For most knockouts, up to 50% of the affected genes are correctly predicted. Important data sources included literature and chip-chip data. Approximately, 12% of the knockout genes were accurately predicted from perturbed profiles.

Discussion

Predicting knockout effects is complicated as they disturb the normal functioning of the cell, leading to many side-effects. Often, this leads to growth defects, stress responses, etc. In our analysis, we found that 40% of the effects consisted of genes that were affected in more than 10 knockouts. HSP30, a known stress-response gene, was even affected in 95 of the 179 knockouts. GO analysis relates these genes to metabolic processes, which would be expected if growth defects occur. This observation suggests that predictions should be related to each other to characterize the core effects.

Presenting Author

Marc Hulsman (m.hulsman@tudelft.nl)
Delft University of Technology

Author Affiliations

1) Delft University of Technology Faculty of Electrical Engineering, Mathematics, and Computer Science Delft Bioinformatics Lab

G-25. Towards better receptor-ligand prediction

*Iacucci E (1, *), Ojeda F (1), De Moor B (1), Moreau Y (1)*

Regulation of cellular events is often initiated via extracellular signaling. Extracellular signaling occurs when a circulating ligand interacts with one or more membrane-bound receptors. Identification of receptor-ligand pairs is thus an important and specific form of PPI prediction. Given a set of disparate data sources (expression data, domain content, and phylogenetic profile) we seek to predict new receptor-ligand pairs. We create a combined kernel classifier and assess its performance with respect to the DLRP ‘golden standard’ as well as the method proposed by Gertz et al. (2003).

Materials and Methods

Our objective is to predict candidate receptor ligand pairs; more specifically, seeking to create a better method than what is currently available (Gertz et al. 2003) to identify known pairs as well as to determine putative pairs for further research. Our method involves taking multiple data sources and producing separate kernels for each data type, creating LS-SVM classifiers and combining the results to predict receptor ligand pairs.

Results

Among our findings, we discover that our predictions for the *tgfb* family accurately reconstructs over 95% of the supported edges (0.95 recall and 0.71 precision) of the receptor-ligand bipartite graph defined by the DLRP “golden standard”. In addition, the combined kernel classifier is able to relatively out-performed the Gertz et al. (2003) work by a factor of approximately two as the Gertz et al. (2003) work reconstruct 44% of the supported edges (0.44 recall and 0.53 precision) of the receptor-ligand bipartite graph defined by the DLRP “golden standard”.

Discussion

The prediction of receptor ligand pairings is a difficult and complex task. We have demonstrated that using multiple data sources provide a clear advantage over single data sources in solving this task. As more high through-put data becomes available, we expect to extend the current methodology to accommodate it.

Presenting Author

Ernesto Iacucci (ernesto.iacucci@esat.kuleuven.be)

K. U. Leuven

Author Affiliations

(1) Department of ESAT-SCD, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium

G-26. Extracting and visualising protein-protein interfaces for data-driven docking

Ghoorah AW (1,*), Devignes M-D (2), Smail-Tabbone M (3), Ritchie DW (1)

PPIs play a key role in many cellular processes. In order to understand these processes, we need to be able to analyse and model them at the molecular level. There now exist several databases which contain almost all of the available experimentally-determined PPIs. Our aim is to search such databases for potentially useful information to identify key interface residues automatically in order to help predict the structures of unknown complexes using protein docking simulations.

Materials and Methods

Given a query domain, we retrieve from the 3DID database all hetero domain-domain interactions involving the query. Biologically relevant interfaces are distinguished from other crystal contacts using calculated interface areas, and duplicates are removed using a sequence similarity threshold. We then annotate and store the 3DID interface residues according to their location in the interface (i.e. “core” or “rim”). We also calculate and store overall residue conservation. We use PFAM consensus sequence to perform 3D superposition of the PDB complexes which contain the query domain.

Results

For any given query domain, our approach retrieves automatically a set of non-redundant hetero interactions. For each interaction, we use VMD to show the core and rim residues at the interface and to highlight any conserved residues. In addition we show a superposition of the protein complexes in the coordinate frame of the query. This provides a convenient way to view and analyse all of the interactions involving the domain of interest. Using this approach we modelled CAPRI target T40, a complex between an API-A and two trypsins, and we obtained 9 out of 10 acceptable predictions.

Discussion

We have developed an automated approach for extracting and analysing relevant domain-domain interactions involving a query domain. We aim to make this approach available as a web server. We are also extending our approach to annotate PPIs according to their biological function and to distinguish permanent and transient interactions. In the longer term, we aim to apply machine learning techniques to discover symbolic rules which can identify key interface residues automatically.

Presenting Author

Anisah W Ghoorah (anisah.ghoorah@loria.fr)
INRIA Nancy Grand-Est

Author Affiliations

1. INRIA Nancy; 2. CNRS Nancy; 3. Nancy Université, LORIA, Vandoeuvre-lès-Nancy 54506, France

Acknowledgements

This project is funded by the Agence Nationale de la Recherche, France. Grant reference ANR-08-CEXC-017-01.

G-27. Visualizing spatiotemporal information in heterogeneous biological networks with Arena3D

Secrier M (1,*), Pavlopoulos GA (2), Schneider R (1)

Biological systems are complex, dynamic entities, the behavior of which is difficult to capture in a holistic manner over space and time. Visualization tools enhance the ability to explore the properties of such systems by exploiting man's capacity to perceive hidden features in the data. Several tools for the visualization of biological networks exist, but integrating the spatial and temporal information remains a major bottleneck in systems biology. We introduce Arena3D, a framework for biological visualization and analysis, with focus on network linking and dynamic pattern identification.

Materials and Methods

Arena3D employs a 2.5D framework with separate layers to distinguish different levels of biological information (genes, proteins, chemicals, pathways, diseases etc). The different entity layers are arranged via layout algorithms and the connections between the layers are visualized. The core graphics visualization in Arena3D is implemented in Java 3D (1.5.1 API). All other parts of Arena3D, including clustering methods and the GUI, are done in Java (JDK 1.6). Parsing is enabled for SMBL and KEGG file formats. Networks can be exported in Medusa, Pajek and VRML format.

Results

Visualizing biochemical reaction networks highlights the most prominent reactions at different time points and the changes in enzyme concentrations. The reaction profile of the cell cycle is described. The succession of phenotypic events upon siRNA knockdown of genes essential for the cell cycle (www.mitocheck.org) has been traced based on scoring and clustering of events. Distinct patterns are pinpointed. We observe many positive correlations between gene knockdowns for the binuclear and polylobed phenotypes and highly uncorrelated behavior in the large and dynamic phenotypes.

Discussion

The Arena3D software enables the investigation of a wide variety of networks and connections between systems at different levels, as well as their evolution in time, by emphasizing changes in gene expression and highlighting pathways that are activated at different stages. The 3D framework allows the user to easily understand the connections that appear between different biological levels. Providing an enriched informational content through spatiotemporal integration can trigger the discovery of new relationships between biomolecules, links between processes and patterns in evolution.

URL

<http://arena3d.org/>

Presenting Author

Maria Secrier (secrier@embl.de)

European Molecular Biology Laboratory, Heidelberg, Germany

Author Affiliations

(1) Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany (2) Department of Computer Science and Biomedical Informatics, University of Central Greece, Lamia, Greece

G-28. Efficient learning of signaling networks from perturbation time series via dynamic nested effects models

Praveen P (1,), Tresch A (2), Froehlich H (1)*

Reverse engineering of biological networks is a key to understand the biological systems. The exact knowledge of interdependencies among proteins in the living cell is crucial for the drug targets identification in various diseases. The advent of techniques, like RNAi, opens new aspects in network reconstruction. NEMs (Markowitz et al, 2005) allow inferring network hypothesis from high-dimensional, indirect downstream perturbation effects. The original approach works only with static data. However, time-resolved data can give better insight into the signal behavior and resolve feedback loops.

Materials and Methods

We propose an extension of the NEM approach for perturbation time series measurements, thus complementing the attempt of Anchang et al. (2009). It allows resolving the feedback loops in the signaling cascade and discriminates direct & indirect signaling therein. Our method works by unrolling the signal flow in the network over time. The signaling is modeled via a Boolean network with synchronous state updates. The likelihood of a network hypothesis is determined efficiently in a closed expression. A network structure prior limits the search space for structure learning via greedy hill climber.

Results

Realistic simulations show a high specificity and sensitivity of our proposed approach. Moreover, it compares favorable against standard Dynamic Bayesian Networks (DBNs), which do not model the propagation of a perturbation effects in a signaling network. In contrast to the approach by Anchang et al. our method does not use time consuming Gibbs sampling. Moreover, our model can estimate feedback loops from data as we allow genes to change their perturbation state over time. Application of our approach to a dataset for murine stem cell development leads to biologically interpretable results.

Discussion

We proposed an efficient statistical model to reconstruct features of signaling cascades from perturbation time series. It allows feedback loop estimation since genes are allowed to change its activation state over time. We suggested an appropriate prior for network structure learning via a greedy hill climbing algorithm. Realistic simulations evince the accuracy of our method, subsequently applied on data for murine stem cell development. A non-parametric bootstrap verifies the confidence of the learned network. Our results are in-line with the data and with previous results by Anchang et al.

URL

<http://www.abi.bit.uni-bonn.de>

Presenting Author

Paurush Praveen (praveen@bit.uni-bonn.de)
Algorithmic Bioinformatics, B-IT, University of Bonn

Author Affiliations

1. Algorithmic Bioinformatics, B-IT, University of Bonn 2. Gene Center Munich, Ludwig-Maximilians-University Munich

Acknowledgements

This work was partially supported by the state of NRW via the B-IT Research School

G-29. Discovering patterns of differentially regulated enzymes in metabolic pathways of tumors

Schramm G (1), Wiesberg S (2), Diessl N (1), Kranz A-J (1), Sagulenko V (3), Oswald M (2), Reinelt G (2), Westermann F (3), Eils R (1), Konig R (1,)*

Gene expression profiling by microarrays or transcript sequencing enables observing the pathogenic function of tumors on a mesoscopic level.

Materials and Methods

We investigated gene expression profiles of neuroblastoma and breast cancer tumors. In contrast to common enrichment tests, we took network topology into account by applying adjusted wavelet transforms on an elaborated and new 2D grid representation of curated pathway maps from the Kyoto Encyclopedia of Genes and Genomes.

Results

The aggressive form of the neuroblastoma tumors showed regulatory shifts for purine and pyrimidine biosynthesis as well as folate-mediated metabolism of the one carbon pool in respect to increased nucleotide production. We spotted an oncogenic regulatory switch in glutamate metabolism for which we provided experimental validation, being the first steps towards new possible drug therapy. For the breast cancer tumors, we found a regulatory switch at the bile acid biosynthesis pathway which may inhibit cholesterol degradation and therefore induction of estrogen synthesis.

Discussion

The pattern recognition method we used complements normal enrichment tests to detect such functionally related regulation patterns.

Presenting Author

Rainer Konig (r.koenig@dkfz.de)
Bioquant, IPMB

Author Affiliations

1 Department of Bioinformatics and Functional Genomics, Institute of Pharmacy and Molecular Biotechnology, and Bioquant, University of Heidelberg, Im Neuenheimer Feld 267, 2 Institute of Computer Science, 3 Interdisciplinary Center for Scientific Computing, University of Heidelberg 4 Department of Tumor Genetics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

Acknowledgements

Funding: BMBF-FORSYS Consortium Viroquant (#0313923); the Helmholtz Alliance on Systems Biology and the Nationales Genom- Forschungs-Netz (NGFN+) for the neuroblastoma project, ENGINE (#01GS0898).

G-30. Extending the yeast metabolic network using a data integration approach

Fisher P, Dobson P, Brenninkmeijer C, Canevet C, Taubert J, Rawlings C, Stevens R*

We are working towards the extension of the yeast metabolic network [PMID: 18846089]. However manual curation is expensive and therefore a combination of bioinformatics tools is required to generate a semi-automatic approach which can assist in the curation of the metabolic network, providing domain experts with relevant publications, and subsequently reducing curation time. We have adopted an integrated systems biology approach to exploit a wide range of datasets that are useful to the curators and bioinformaticians on the Yeast Jamboree Project.

Materials and Methods

The Yeast metabolic network was imported into Ondex and a filter applied to remove highly connected metabolites. Unconnected (orphan) metabolites were identified and used to filter the KEGG Yeast database. A Taverna workflow was created and run to find these orphans in the KEGG Yeast database, retrieve enzyme names, and assemble a MEDLINE query (restricted to the yeast literature) based on those names. Literature identified by this workflow was re-integrated back into the KEGG Yeast network stored within Ondex, as a means of identifying novel metabolic reactions for each orphan metabolite.

Results

The Taverna and Ondex tools were successfully combined into a single data integration platform. As a result, we developed a mechanism in which researchers are able to identify gaps within the yeast metabolism network by providing them with publications that may be relevant to close such gaps. New Taverna workflows were created to obtain a set of MEDLINE publications which allowed us to annotate a subset of the KEGG database relating to gaps in the Yeast metabolism network. Identified orphans were linked to publications via a Taverna workflow and re-integrated into the network.

Discussion

Through the combination of Ondex and Taverna, we were able to successfully construct an explicit methodology that provides a powerful data analysis platform for systems biology research. We have developed a system that considerably reduces the time needed by the curators of the Yeast Jamboree to identify evidence that links new metabolites into the yeast metabolic network. This has led to significant progress in extending the Jamboree model. Publications added to the yeast metabolic network are currently being reviewed by curators for their suitability to fill further gaps in the network.

URL

<http://www.ondex.org>

Presenting Author

Paul R Fisher (paul.fisher@manchester.ac.uk)

University of Manchester

Author Affiliations

The University of Manchester Rothamsted Research

Acknowledgements

The Ondex SABR project is funded by BBSRC Grants BBS/B/13640 and BB/F006039/1.

G-31. Predicting genetic interactions by quantifying redundancy in biochemical pathways

Delgado-Eckert E (1,), Gill S (1), Merdes G (1), Beerenwinkel N (1)*

Biochemical networks consist of interacting proteins that carry out a function. Most of the proteins in such a network are not essential, a property called robustness. Systematic double-knockout screens in *S. Cerevisiae* have started to shed light on the nature of this robustness by quantifying the degree of genetic redundancy. A pair of genes is said to genetically interact if the phenotype of the double-knockout mutant notably differs from the single-knockout mutants. Screening all pairs represents a huge effort. Thus, computational methods that help predict genetic interactions are required.

Materials and Methods

We devised a computational methodology that uses the connectivity information contained in the protein interactome to predict the likelihood of genetic interactions. To account for the uncertainty attached to experimental conditions, we represent the interactome using probabilistic networks, i.e., probabilities are assigned to the network's edges. Within this framework we use the concept of network reliability to address the stability of the network under simultaneous perturbation of node pairs.

Results

We are currently testing our methodology using the vast amount of protein interaction data and the recently established genetic interaction data available for *S. Cerevisiae*. We are focusing on the study of transcription pathways in *S. Cerevisiae* due to the large body of knowledge and experimental results regarding transcription mechanisms.

Discussion

Our first explorations show that one of the biggest challenges in this approach is to define the connectivity properties of the biochemical network that are required for the functionality of the pathway. This step requires deep knowledge about the molecular mechanisms that underlie a particular cellular function. This insight brings up the question as to what extent one could explore by means of simulation what connectivity properties of a network best explain the experimentally observed pattern of genetic interactions and pairwise redundancies.

Presenting Author

Edgar W. Delgado-Eckert (edgar.delgado-eckert@bsse.ethz.ch)
ETH Zurich

Author Affiliations

(1) Department of Biosystems Science and Engineering, ETH Zürich, Mattenstrasse 26, 4058 Basel, Switzerland.

Acknowledgements

The authors acknowledge support from the SystemsX initiative in systems biology of the Swiss National Science Foundation.

G-32. Hub protein interfaces and hot region organization

Cukuroglu E (1,*), Gursoy A (1), Keskin O (1)

Binary interactions of the proteins constitute the protein interaction networks. Some proteins are highly connected to the others (called as hub proteins), whereas some others have a few interactions (called as non hub proteins). Here, we concentrate on hub proteins from a structural point of view: protein-protein interfaces. Our proposed method analyzes the connection between organization of hot spots (hot regions) and being hub proteins. We define interfaces as the ones between two date hubs (DD), and two party hubs (PP).

Materials and Methods

A Hub protein dataset generated from PPIN dataset is used in this study. Date, party and non hub proteins which have complex structured PDB file are extracted and interfaces of complexes are fetched from the interface dataset to eliminate redundancy. After elimination of the structurally redundant complexes, there are 38 PP and 26 DD interfaces. Hot spots of the interfaces are defined and the groups of hot spots which have at least two contacting hotspot neighbors in the interface are called hot regions. Significant features of the interface are defined with ANOVA test.

Results

Results reveal that there are significant differences between DD and PP interfaces. More of the hot spots are organized into the hot regions in DD interfaces compared to PP ones. A high fraction of the interfaces are covered by hot regions in DD interfaces. There are more distinct hot regions in DDs. Since the same (or overlapping) DD interfaces should be used repeatedly, different hot regions can be used to bind to different partners. Further, these hot region characteristics can be used to predict whether a given hub interface is involved in a DD or a PP interface type with 80% accuracy.

Discussion

In this work, we conclude that there is a relationship between organization of hot spots (hot regions) and the status of hub proteins. The hot region characteristics (Hot spot ratio, average hot region size, average hot region ASA to interface ASA ratio, Polar amino acid (aa) frequencies of interfaces, Polar aa frequencies of hot spots, Polar aa frequencies of hot regions) can be used to predict whether an interface is formed between a DD or PP type of an interface with 80% accuracy.

Presenting Author

Engin Cukuroglu (ecukuroglu@ku.edu.tr)

KOC University

Author Affiliations

(1)Koc University, Center for Computational Biology and Bioinformatics and College of Engineering, Rumelifeneri Yolu, 34450 Sariyer, Istanbul, Turkey

G-33. A functional-genomics fermentation platform to identify and optimize industrial-relevant properties of *Lactobacillus plantarum*

Wels M (1,2,4,), Bron PA (1,2,), Wiersma A (1,2,4), Marco M (1,2,§), Kleerebezem M (1,2,3,4)*

Probiotics constitute an important growth market for the food industry. However, further development of the probiotics market is presently constrained by a lack of knowledge on probiotic cellular components important for health-promoting effects. We have developed a functional-genomics fermentation platform approach for the identification and optimization of expression of specific probiotic functionality parameters. The platform employs *Lactobacillus plantarum* WCFS1 as an extensively studied model microbe for which advanced molecular tools are available.

Materials and Methods

In the first set of fermentations, *L. plantarum* was grown in chemically defined medium according to a combinatorial fermentation scheme that included variations in medium composition and mild stress conditions (NaCl, pH, oxygen, amino acids and temperature). The molecular characteristics of the bacteria in samples harvested from these fermentations were investigated by transcriptome analysis, while in parallel their functional parameters were assessed using specific probiotic functionality assays, but also flavor profile changes that might influence product taste.

Results

Advanced bioinformatics analyses were performed to correlate transcriptomics results with the functionality parameters obtained, enabling the identification of genes involved in specific functions that are relevant for performance. The function of the identified genes and their relevance for performance will be further studied by genetic engineering (KO, overexpression).

Discussion

This approach will allow optimization strategies for the improvement of pre-selected target genes through specific modulation of fermentation conditions. Although developed with the use of a model organism, this approach is applicable to other bacteria, and the correlation of desired functional properties to “omics” datasets may assist identification of the underlying molecular mechanism, which opens avenues towards the design of fermentation strategies geared to improve the functional properties of bacterial (health-impact) cultures.

Presenting Author

Michiel Wels (michiel.wels@nizo.nl)

Top Institute Food and Nutrition

Author Affiliations

1TIFN/2NIZO food research B.V., the Netherlands; 3Wageningen UR, the Netherlands; 4 Kluiver Centre for Industrial Fermentations, Netherlands. § Present address UC Davis, USA

G-34. DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks

*Bauer-Mehren A (1, *), Rautschka M (1), Sanz F (1), Furlong LI (1)*

Most human diseases arise due to interactions among multiple genetic variants and environmental factors. To get a comprehensive view of the full spectrum of human diseases with genetic origin including monogenic, complex and environmental diseases, we developed a comprehensive database comprising gene-disease associations from expert-curated repositories and a text-mining source. Network representation allows studying global properties of genetic origin of human diseases. Hence, we developed DisGeNET, a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks.

Materials and Methods

We developed DisGeNET, a plugin for Cytoscape to access and explore our gene-disease association database. Gene-disease associations are displayed as bipartite-graphs with multiple edges between nodes, each representing a unique association found in the original data sources. Moreover, we color edges according to the association type following our gene-disease association ontology. Network projections are accessible presenting gene and disease centric view of the data. Diseases were classified into disease classes and represented with multiple node colors.

Results

DisGeNET allows user-friendly access to our database including queries restricted to the (i) original source, (ii) association type, (iii) disorder class of interest and (iv) specific disease/gene or set of diseases/genes. It represents gene-disease associations as bipartite graphs and additionally provides gene centric and disease centric views of the data. It assists the user in the interpretation and exploration of human diseases with respect to their genetic origin by a variety of built-in functions. DisGeNET is compatible with Cytoscape 2.x.

Discussion

DisGeNET is a valuable source for biomedical research with a variety of possible applications including finding new candidate disease genes, understanding mechanisms underlying diseases, studying influence of environmental factors such as drugs on human health. In addition, DisGeNET provides a user-friendly framework and the possibility to make use of the variety of available Cytoscape functionalities.

URL

<http://ibi.imim.es/DisGeNET/DisGeNETweb.html>

Presenting Author

Anna Bauer-Mehren (abauer-mehren@imim.es)

Integrative Biomedical Informatics Laboratory Research Group on Biomedical Informatics (GRIB) - IMIM/UPF

Author Affiliations

(1) Research Unit on Biomedical Informatics (GRIB) IMIM/UPF, C/Dr. Aiguader 88, 08003 Barcelona, Spain

Acknowledgements

This work was generated in the framework the EU-ADR project co-financed by the European Commission through the contract no. ICT-215847 and the e-TOX project from the European Community's Seventh Framework Programme (FP7/2007-2013) for the Innovative Medicine Initiative under grant agreement n° 115002. The Research Unit on Biomedical Informatics (GRIB) is a node of the Spanish National Institute of Bioinformatics (INB) and member of the COMBIOMED network. We thank the Departament d'Innovació, Universitat i Empresa (Generalitat de Catalunya) for a grant to author ABM.

G-35. Combination of network topology and pathway analysis to reveal functional modules in human disease

Bauer-Mehren A (1, *), Bundschuh M (2,3), Rautschka M (1), Mayer MA (1), Sanz F (1), Furlong LI (1)

Most human diseases arise due to interactions among multiple genetic variants and environmental factors. In the last years, several gene-disease association databases have been developed; nevertheless much information is still locked in the literature. To get a comprehensive view of the full spectrum of human diseases with genetic origin including monogenic, complex and environmental diseases, data integration is required. Network representation combined with functional analysis can be used to elucidate functional modules related to human diseases and to understand their underlying mechanisms.

Materials and Methods

A comprehensive database comprising gene-disease associations from expert-curated repositories and a text-mining derived network was developed. The integration required disease and gene vocabulary mapping and development of a gene-disease association ontology. We represent gene-disease associations as bipartite-graphs and generate disease and gene centric views of the data by means of network projection. Network analysis, graph-clustering approaches and functional annotation to GO terms and biological pathways were used to explore global properties of the data or to focus on specific diseases.

Results

Integration of data from diverse sources highlighted small overlap of genes, diseases and their associations. Most diseases have more associated genes than reported in a single database, even for diseases regarded as monogenic. We identified phenotypically derived gene clusters with varying degree of functional homogeneity. For most clusters more than one biological process was found suggesting a putative role of cross-talks of pathways in disease development. In addition, gene products of clusters characterized by single biological process were involved in direct physical interactions.

Discussion

Network analysis confirmed the need of integrating gene-disease associations from diverse sources to obtain a comprehensive picture of genetic basis of human diseases. Clustering and pathway analysis of disease-related processes can give insights into the underlying mechanisms which can help in finding new candidate disease genes, explanations for drug toxicity and eventually support development of new treatment strategies. Based on cluster and pathway enrichment analysis, a possible explanation for the toxicity on skeletal muscle of the drug Perhexilin is proposed.

URL

<http://ibi.imim.es/DisGeNET/DisGeNETweb.html>

Presenting Author

Anna Bauer-Mehren (abauer-mehren@imim.es)

Integrative Biomedical Informatics Laboratory Research Group on Biomedical Informatics (GRIB) - IMIM/UPF

Author Affiliations

(1) Research Unit on Biomedical Informatics (GRIB) IMIM/UPF, C/Dr. Aiguader 88, 08003 Barcelona, Spain (2) Institute for Computer Science, Ludwig-Maximilians-University Munich, Oettingenstr. 67, 80538 Munich, Germany (3) Siemens AG, Corporate Technology, Information and Communications, Otto-Hahn-Ring 6, 81739 Munich, Germany

Acknowledgements

This work was generated in the framework the EU-ADR project co-financed by the European Commission through the contract no. ICT-215847 and the e-TOX project from the European Community's Seventh Framework Programme (FP7/2007-2013) for the Innovative Medicine Initiative under grant agreement n° 115002. The Research Unit on Biomedical Informatics (GRIB) is a node of the Spanish National Institute of Bioinformatics (INB) and member of the COMBIOMED

network. We thank the Departament d'Innovació, Universitat i Empresa (Generalitat de Catalunya) for a grant to author ABM.

G-36. Attacking interface & interaction networks

*Engin B (1, *), Gursoy A(1), Keskin O(1)*

Protein-protein interactions (PPIs) are crucial in almost all biological processes such as cell regulation, signal transduction, gene replication, or translation. Understanding interactions between proteins is very important to explore biological processes, as well as solving the mysteries of disease mechanisms in organisms. Representation of interaction data by PPI networks enables us to utilize well-known and well-studied graph theoretical approaches to analyze protein-protein interaction networks where nodes represent proteins and the edges between them represent the interactions.

Materials and Methods

“Interface & Interaction Networks” result from integration of binding site information into PPI networks. This representation depicts proteins as nodes, interactions as edges and interfaces as a different kind of node. Protein interactions are carried over interface nodes. Here attacks are performed to interfaces, which mean deleting all similar interfaces to a chosen target interface (in order to mimic drug mechanism) and associated interactions from the network.

Results

We performed our analysis on p53 pathway and observed that among various target selection strategies the most effective one is the selection of the interfaces by the number of their occurrences. This is logical because more the interface is observed, the more the other interfaces will be affected.

Discussion

Complex Networks is an important paradigm in biology since protein-protein interactions are generally abstracted with graph representation. In order to improve ways of defining multiple targets for partial attacks we introduced structural information to interaction networks. Resulting network model details the interaction information, it also includes the interface(binding site) information which are used by interacting proteins.

Presenting Author

Billur Engin (bengin@ku.edu.tr)

Koc University

Author Affiliations

(1) Koc University, Istanbul, Turkey

G-37. Pathway Projector: web-based zoomable pathway browser using KEGG Atlas and Google Maps API

Kono N (1*), Arakawa K (1), Ogawa R (1), Kido N (1), Oshita K (1), Ikegami K (1), Tamaki S (2), Tomita M (1)

Biochemical pathways provide an essential context for understanding the experimental data and the systematic workings of a cell. Therefore, the availability of online pathway browsers will facilitate post-genomic research, just as genome browsers have contributed to genomics. Many pathway maps have been provided online as part of public pathway databases. Most of these maps, however, function as the gateway interface to a specific database, and the comprehensiveness of their represented entities, data mapping capabilities, and user interfaces are not always sufficient for generic usage.

Materials and Methods

The pathway map of Pathway Projector is based on the KEGG Atlas map, for the familiarity of its layout and for the availability of various analysis tools. However, because the KEGG Atlas only represents metabolite nodes, we added all gene and enzyme nodes semi-automatically on the reference pathway map. The software was implemented using AJAX (Asynchronous JavaScript + XML) programming paradigm and the main interface framework was built with Ext JS library. For the representation of the global pathway map, we adopted zoomable user interface (ZUI) using Google Maps API through G-language GAE.

Results

Here we developed a novel web-based pathway browser named the Pathway Projector, which allows browsing of the integrated pathway map with intuitive ZUI. Pathway Projector is based on the integrated metabolic pathway map of KEGG Atlas by adding gene and enzyme nodes, and uses Google Maps API for intuitive zooming. Pathway Projector has four types of search functionalities, including those by keywords and identifiers, by molecular mass, by possible routes between two metabolites using PathComp, and by sequence similarity using BLAST. It also has functionality for data mapping and manual editing.

Discussion

The understanding of omics layers is important for systems biology, and biochemical pathways provides a necessary context for this purpose. Since pathways do not exist independently, but are rather interconnected in vivo, use of an integrated map is desirable, especially for the mapping of comprehensive experimental data. Pathway Projector has an intuitive interface for such integrated global pathway maps. Moreover, capabilities of this software such as searching, editing, annotation, mapping and links to various databases, will be a useful gateway for pathway analysis.

URL

<http://www.g-language.org/PathwayProjector/>

Presenting Author

Nobuaki Kono (ciconia@sfc.keio.ac.jp)
Institute for Advanced Biosciences, Keio University

Author Affiliations

1. Institute for Advanced Biosciences, Keio University 2. Nara Institute of Science and Technology

Acknowledgements

This research was supported in part by the Grant-in-Aid for JSPS Fellows and the Grant-in-Aid for Young Scientists (A), No.222681029, 2010, from the Japan Society for the Promotion of Science (JSPS), as well as by funds from the Yamagata Prefectural Government and Tsuruoka City. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

G-38. Mapping the "Farnesylome": structure-based prediction of farnesyltransferase targets

London N (1), Schueler-Furman N (1)*

Farnesylation by farnesyltransferase (FTase) targets proteins to the membrane. We propose a structural modeling scheme to account for peptide specificity. We obtain a very good discrimination (AUC = 0.88/0.91 on training/test sets), and identify 85% of known targets. A genomic scan reveals 77 novel putative substrates undetected by sequence-based methods. These are currently validated by experiment. Our approach can easily be adapted to additional systems, and is expected to contribute significantly to the elucidation of the cellular network of peptide-mediated interactions.

Materials and Methods

Given a Cxxx sequence, we wish to predict whether it would bind FTase, using the following, Rosetta-based protocol: Starting from the solved structure, we find the optimal rotameric state within the binding pocket for a given peptide sequence. We minimize all of the peptide degrees of freedom while constraining two conserved hydrogen bonds and the ZN coordination. The energy contributions of the peptide residues is used as score for its classification as a binder or non-binder. To identify potential new targets, we then screen genomes for Cxxx peptides and score these.

Results

We chose a set of peptide sequences tested for FTase activity to devise our prediction protocol (AUC=0.88). We defined a threshold which yields 69% sensitivity with a FPR of only 8%. On an independent test set, we obtain an even better AUC (0.91), and the learnt threshold provides a TPR of 86% and FPR of 12%. This result gives us confidence in predicting novel sequences. Using a stringent score threshold, we scanned the Human genome and identified 167 potential targets, and those not identified by sequence-based approaches (e.g. PrePS) are now tested experimentally in the Fierke Lab.

Discussion

Elucidation of FTase targets has significant implications for signal transduction research and new therapeutics. We present a generic, robust and accurate protocol for peptide binding specificity prediction, which complements sequence based approaches and detects novel targets. Our structure based approach can be easily adapted to other systems and will expand our knowledge of the cellular peptide-protein interaction network.

Presenting Author

Ora Schueler-Furman (oraf@ekmd.huji.ac.il)

The Hebrew University, Institute for Medical Research IMRIC

Author Affiliations

Department of Microbiology and Molecular Genetics, Institute for Medical Research IMRIC, The Hebrew University.

Acknowledgements

Israel Science Foundation Grant No. 306/6

G-39. Integrative analysis of gene expression and copy number variation data to elucidate a reference human gene network

Orsini M (1,), Pinna A (1), De Leo V (1), de la Fuente A (1)*

Understanding diseases requires identifying the differences between healthy and affected tissues also in terms of regulatory networks that are dysfunctional in a given disease state. Gene expression and CNV data are currently available in public DBs. From gene-expression data alone, it is possible to compile Gene Co-expression Networks for both the healthy and affected samples, and then compare them to identify subnetworks that potentially underlie the disease phenotype. Gene Co-expression Networks are undirected networks, and thus lack the causal directions for the effects between genes

Materials and Methods

Raw measurement, sample properties and clinical information (healthy, disease state, tumor, biopsy, stadiation, treatment, etc) were obtained from public repositories. To limit dis-homogeneity, only data from Affy platforms were recovered. We explicitly considered only those sample were both CNV and gene expression measurement are available simultaneously. We stored these data in a database (IntegromicsDB) for integrative biology studies. Simulated gene expression data were generated with MATLAB using Ordinary Differential Equations based on biochemical kinetics.

Results

We drawn algorithms to elucidate Gene Networks thoroughly evaluated them with simulation studies, and applied these to human data to compile a “healthy” network. Although correlation does not always implies causation, if there is significant correlation between the CNV of a gene *i* and the gene expression level of gene *j*, then we can interpret such correlation as a causal effect of gene *i* on gene *j*, as it is unlikely that the gene expression level of gene *j* caused the CNV of gene *i*. They might be correlated due to confounding by a third factor, but our inference approach recognized such cases.

Discussion

The availability of a centralized expression and CNV data-resource is useful to investigate, common gene expression and CNV profiles among patients having same disease with similar characteristics. Comparison of Gene Networks among healthy and sick patients helps to identify genes and subnetworks involved in a given pathology. Integrative analysis of gene expression and CNV data allowed us to obtain directed Gene Networks, where edges represent causal effects between genes. This approach can represent a starting point to design new integrative and multi-dimensional genetic strategies of study.

Presenting Author

Massimiliano Orsini (orsini@crs4.it)
CRS4 bioinformatics

Author Affiliations

CRS4 Bioinformatica, c/o Parco Tecnologico della Sardegna, Edificio 3, Loc. Piscina Manna, 09010 PULA (CA), ITALY

G-40. Unknown player in modelling of signal transduction pathway*Sadeh M (1,*), Anchang B (1), Spang R (1)*

Nested Effects Models are a class of probabilistic models introduced to analyze the effects of gene perturbation screens visible in high-dimensional phenotypes like microarrays. NEMs require expression profiles of perturbation experiments of all signalling genes in a pathway. In many cases one does not know all these genes or one does not even know whether one knows them all. There is virtually always the possibility that unknown pathway players exist and that important links of the pathway are missing in our current knowledge. The propose method is useful to handle this situation.

Materials and Methods

In signalling pathway, there is virtually always the possibility that unknown pathway players exist and that important links of the pathway are missing in our current knowledge. This means that we might model too many genes including some that do not belong to the pathway. We hypothesize by considering an additional but unknown signalling Silencing gene in the model for which no data is available, we can still obtain better predictive performance of the cells response to treatment compared with the optimal model without this gene.

Results

We started the investigation of this idea in the context of a simulated random five gene network. In order to have an comprehensive conclusion, this method has been applied for 100 random networks (graph) with 5 genes (nodes). Nested Effects Models was applied with for noisy dataset which are generated from these networks. The results show, after removing the information of one random gene in networks, we would obtain better prediction performance by considering an additional but unknown signalling gene in the model for which no data is available.

Discussion

We introduced a method to address one specific problems which are central the analysis of interventional data in biology: Does the expression data used for Nested Effects Models contain the information that unknown players must exist? The proposal method tries to find a reasonable pattern in simulation study in order to use it in a real data scenario. So far we have investigated this idea in the context of simulated five genes networks, where one gene is not known. This investigation suggest a couple of new ideas in signalling network reconstruction.

Presenting Author

Mohammad Sadeh (mohammed.sadeh@klinik.uni-regensburg.de)

Institute for Functional Genomics Computational Diagnostics Group University of Regensburg

Author Affiliations

Institute for Functional Genomics Computational Diagnostics Group University of Regensburg

G-41. Assembly of protein complexes by integrating graph clusterings

Chin C-H (1, 4, *), Chen S-H (1), Chen C-Y (2), Hsiung C (2), Ho C-W (3), Ko M-T (1), Lin C-Y (1, 2, 3, 5)

Proteins in a complex are organized for particular functions, such as synthesis of biomolecules or protein degradation. The complex is similar to the idea of “community” of a network, in which vertices are joined tightly in between, while loosely out-connected. There exist many community detection methods. However, there is no suitable method for integrating them at present. For grasping more comprehensive community structures, we propose a framework to integrate different clustering methods' results to generate a better one.

Materials and Methods

There are two types of graph clustering algorithms: global clustering and local clustering. A clustering method is a global clustering if each vertex of the input graph is assigned a cluster in the output of the method; and a clustering method is a local clustering if the cluster assignments are only performed for a certain subset of vertices. Naturally, each type has both strengths and weaknesses. Here, we design a measure to judge the quality of a cluster as a community, and propose a framework to integrate different clustering results.

Results

First, we applied six clustering methods on a yeast's protein-protein interaction network downloaded from DIP, and computed their integration. Then, we validated these results by comparing them with Gene Ontology annotations and two reference sets from MIPS and Aloy et al. The validation shows that our framework works. After checking the biological significance of the integration, we found that the result can successfully identify undiscovered components and decipher common submodules inside these complexes like RNA polymerases I, II, III on yeast interactome.

Discussion

In local clustering methods, some protein components may not be classified into the resulting clusters due to the incomplete input PPI network structure. Therefore, the recall of a local clustering method is usually lower than a global clustering method. However, the precision of a local clustering method is often higher than a global clustering because a global clustering method does not classify "intermediate" vertices well. We show that a better result could be derived when we integrate the two types of methods properly. A user-friendly web service for this research is also online.

URL

<http://hub.iis.sinica.edu.tw/clustering/>

Presenting Author

Chia-Hao Chin (jovice@iis.sinica.edu.tw)
Academia Sinica, Taiwan

Author Affiliations

(1) Institute of Information Science, Academia Sinica, No. 128 Yan-Chiu-Yuan Rd., Sec. 2, Taipei 115, Taiwan. (2) Division of Biostatistics and Bioinformatics, National Health Research Institutes, No. 35 Keyan Rd. Zhunan, Miaoli County 350, Taiwan. (3) Institute of Fisheries Science, College of Life Science, National Taiwan University, No. 1, Roosevelt Rd. Sec 4, Taipei, Taiwan. (4) Department of Computer Science and Information Engineering, National Central University, No.300, Jung-da Rd, Chung-li, Tao-yuan 320, Taiwan. (5) Research Center of Information Technology Innovation, Academia Sinica, No. 128 Yan-Chiu-Yuan Rd., Sec. 2, Taipei 115, Taiwan.

Acknowledgements

The authors would like to thank National Science Council (NSC), Taiwan, for financially supporting this research through NSC 98-2221-E-008-079 to Ho C-W, 98-3112-B-400-010- and 98-2221-E-001-018- to Lin C-Y.

G-42. How natural host species avoid CD4+ T cell depletion during SIV infection

Chan ML (1,2*), Petravic J (1), Ortiz AM (3), Engram J (3), Paiardini M (3), Cromer D (4), Silvestri G (3,5), Davenport MP (1)

Natural hosts of SIV such as Sooty Mangabeys (SM) remain largely asymptomatic when infected. This is in sharp contrast with non-natural hosts, such as Rhesus Macaques (RM) and HIV-infected individuals, who experience a progressive decline of CD4+ T cells in the blood and lymphoid tissues, and develop AIDS. Currently, the lack of disease progression in natural hosts is still not clearly understood. A better understanding of why SIV is not pathogenic in natural hosts will in turn provide valuable insights into the pathogenesis of AIDS in HIV-infected individuals.

Materials and Methods

Experimental data show that as CD4+ T cells are depleted during infection, SM show only a small increase in the proportion of proliferating CD4+ T cells compared to larger increases in RM and HIV-infected individuals. This suggests that disease progression is associated with marked differences in the proliferation level of CD4+ T cells in response to CD4+ T cell depletion. We have developed a model that demonstrates the relationship between proliferation and disease outcome to help us understand why SIV is not pathogenic in SM, and otherwise in RM and HIV-infected individuals.

Results

We modelled the disease course in a SM- and a human-like host, which only differed in their maximal CD4+ T cell proliferation level. Since dividing cells are preferentially infected, the decrease in the proliferative response to CD4+ T cell depletion in the SM-like host led to a large increase in the total number of uninfected CD4+ T cells present in chronic infection. The model therefore suggests that a lower proliferative response to CD4+ T cell depletion in non-pathogenic SIV infection paradoxically leads to the preservation of CD4+ T cell counts during chronic infection.

Discussion

CD4+ T cell proliferation plays a dual role during infection: Increasing it will act to replace CD4+ T cells and restore immunity, but this also generates more cells that are susceptible to infection. The rapid CD4+ T cell proliferation in RM fuels the fire of infection by producing a large proportion of proliferating cells that are susceptible to infection. Conversely, the moderate homeostatic proliferation in SM is optimal for the preservation of CD4+ T cells during chronic infection. Hence SMs may have adapted to infection by attenuating their proliferative response to CD4+ T cell depletion

Presenting Author

Ming Liang Chan (daniel_chan@student.unsw.edu.au)
School of Medical Sciences, University of New South Wales

Author Affiliations

1. Complex Systems in Biology Group, Centre for Vascular Research, University of New South Wales, Kensington, NSW, Australia 2. School of Medical Sciences, Faculty of Medicine, University of New South Wales, Kensington, NSW, Australia 3. Department of Pathology and Laboratory Medicine, University of Pennsylvania, PA, USA 4. Department of Mathematics and Centre for Integrative Systems Biology, Imperial College London, London, UK 5. Yerkes National Primate Research Center, Emory University, Atlanta, GA, USA

Acknowledgements

NIH-AI-66998, HL-75766, NHMRC (Australia), ARC (Australia), UNSW-UIPA (Australia)

G-43. Kinase-specific phosphorylation site prediction in *Arabidopsis thaliana*

Dang T-H (1,), Verschoren A (1), Laukens K (1)*

Reversible phosphorylation of proteins plays critical roles in a diverse range of signaling pathways. Phosphorylated sites are usually experimentally identified by Mass Spectrometry based techniques which are often time-consuming, labor-intensive and expensive. Although *Arabidopsis thaliana* is a model plant and thus a significant focus of international plant research, the knowledge about phosphorylation processes in *Arabidopsis thaliana* is still limited. The *Arabidopsis* genome encodes at least two times more protein kinases than the human genome (Manning et al, 2002).

Materials and Methods

The method uses the InParanoid algorithm (O'Brien et al., 2005) to find kinases in *Arabidopsis thaliana* which are orthologous to extracted kinases belonging to kinase groups in the Phospho.ELM database. It then uses Pfam (Finn et al., 2010) to discard candidates that do not contain any kinase catalytical domains. Subsets of known residues phosphorylated by kinase groups in other species were extracted. From those the CRPhos models (Dang et al., 2008) were trained and used to make a prediction for all protein sequences in the *Arabidopsis thaliana*'s genome.

Results

The prediction results were compared with those from previous studies (Sugiyama et al. 2008, Heazlewood et al. 2008). They both do not provide any information about responsible kinases, and they overlap only partially. A statistically significant overlap between predicted sites and those from the two databases was found, and it increases proportionally to the increased FPR, as expected. Applying the method results in a predicted phosphorylation network for which is further analysed in detail.

Discussion

To our knowledge, the proposed method is the first attempt to predict the genome-wide phosphorylation network in *Arabidopsis thaliana*. The first results demonstrate that the approach based on orthologous protein kinases across multiple species is useful, in particular for organisms for which existing data on phosphorylations is limited, like in *Arabidopsis thaliana*. The predicted results are a solid basis for further computational analysis and experimental validations.

Presenting Author

Thanh Hai Dang (thanhhai.dang@ua.ac.be)

Intelligent Systems Laboratory (ISLab), Department of Mathematics and Computer Science, University of Antwerp

Author Affiliations

(1) Intelligent Systems Laboratory (ISLab), Department of Mathematics and Computer Science, University of Antwerp

Acknowledgements

SBO grant (IWT-600450, Bioframe) of the Flemish Institute supporting Scientific-Technological Research in Industry

G-44. BiologicalNetworks: enabling systems-level studies of host-pathogen interactions

Kozhenkov S (1), Sedova M (1), Dubinina Y (1), Ponomarenko J (1,2,), Baitaluk M (1)*

Understanding of immune response mechanisms of pathogen-infected host requires multi-scale analysis of genome-wide data. Data integration methods have proved useful to the study of biological processes in model organisms, but their systematic application to the study of host immune system response to a pathogen and human disease is still in the initial stage. To study host-pathogen interaction on the systems biology level, an extension to the previously described BiologicalNetworks (1) and IntegromeDB (2) is proposed.

Materials and Methods

Data integration and mapping to the internal database is fully automated and based on Semantic Web technologies and Web Ontology Language. The system represents a general-purpose graph warehouse with its own data definition and query language, augmented with data types for biological entities. The list of integrated databases is at <http://www.biologicalnetworks.net/Database/tut5.php>. Bioinformatics methods were implemented that allow to reconstruct/modify phylogenetic trees, get multiple sequence alignments, identify phylogenetically conserved transcription factor binding sites, and other.

Results

The developed system has been applied to the systems-level analysis of the influenza virus-host interactions, including host molecular pathways that are induced/repressed during the infections, co-expressed genes, and conserved transcription factor binding sites. Previously unknown to be associated with the influenza infection genes were identified and suggested for further investigation as potential drug targets.

Discussion

The developed methods and data integration and querying tools allow simplifying and streamlining the process of integration of diverse experimental data types, including molecular interactions and phylogenetic classifications, genomic sequences and protein structure information, gene expression and virulence data for pathogen-related studies. The data can be integrated from the databases and user's files for public use. References: 1) Baitaluk M, et al. (2006) *Nucleic Acids Res.* 34: W466. 2) Baitaluk M, Ponomarenko J (2010) *Bioinformatics*, doi:10.1093/bioinformatics/btq231.

Presenting Author

Julia Ponomarenko (jpon@sdsc.edu)
University of California San Diego

Author Affiliations

(1) San Diego Supercomputer Center, (2) Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, 9500 Gilman Drive, La Jolla, CA, 92093, USA.

Acknowledgements

The US National Institute of Health grants R01GM084881 and R01GM085325.

G-45. Genome-wide identification of disease-related genes

Jaeger S (1,*), Leser U (1)

For many human diseases it is not yet known which genes are involved in the pathogenesis of the diseases. Elucidating the underlying disease mechanisms is crucial for understanding the onset of diseases, and the development of specific diagnostic and therapeutic approaches. Several computational methods have been devised for identifying disease-related genes. However, the majority focuses on prioritizing candidates within defined chromosomal regions. These approaches fail short in appropriately handling diseases where the associated loci are still unknown and diseases without genetic origin.

Materials and Methods

We present a computational framework to identify novel disease genes in a genome-wide setting. For a given disease, we extract all genes that are known to be associated with this disease. We compile disease-specific networks by integrating directly and indirectly linked gene products using protein interaction as well as manually curated and predicted functional data. Network centrality analysis is applied to rank genes according to their relevance for the given disease. In addition, a simple normalization strategy is employed to adjust the ranking for a bias toward global hubs in the networks.

Results

We validate our method using disease-gene association data from OMIM. We show that using predicted functions allows for an improved ranking of proteins that are uncharacterized in first place by increasing the recovery rate of disease proteins from 80% to 86%. Cross-validation is used to simulate a genome-wide candidate search for diseases where no associated chromosomal regions are known. We show that using predicted functional data and including indirectly linked genes in combination with a proper score normalization improves the genome-wide identification of disease genes from 29% to 51%.

Discussion

We developed a linkage interval-independent framework for identifying disease genes using protein-protein interaction data and functional annotations in combination with network centrality analysis. Our approach does not depend on the availability of associated chromosomal regions, which makes it applicable to a much wider range of diseases than previous algorithms, such as disorders with very few or even only a single known disease protein, diseases with multiple, very large, or no associated loci, and also diseases without genetic origin.

Presenting Author

Samira Jaeger (sjaeger@informatik.hu-berlin.de)
Humboldt-Universität zu Berlin

Author Affiliations

(1) Knowledge Management in Bioinformatics, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

Acknowledgements

This work is funded by an Elsa-Neumann scholarship.

G-46. Bioinformatics tools for the investigation of proteomics aspects involved in cardiovascular diseases: from biomarker discovery to protein-protein interaction networks

Di Silvestre D (1,), Brambilla F (1), Brunetti P (1), Lionetti V (2), Agostini S (2), Cavallini C (3), Mauri PL (1)*

An approach based on systems biology implies the examination of a set of multiple elements. In this context, proteins, which represent the effector molecules for all cell functions, play a central role. In the last few years, mass spectrometry-based (MS) proteomics, such as MudPIT (Multidimensional Protein Identification Technology), have become the leading approach employed in high-throughput analysis. At the same time, the great amount of data produced by MS experiments, requires the development of computational tools for data processing and handling.

Materials and Methods

The MudPIT approach has been applied to Sus Scrofa cardiac tissues affected by myocardial infarction. Protein profiles of analyzed samples have been achieved using the SEQUEST algorithm in combination with the Homo sapiens protein database. In order to identify biomarkers and proteotypic peptides, the MAProMa and EPPI software have been employed, respectively. In addition, protein profiles and biomarkers were further processed by means of cluster analysis and to investigate protein-protein interaction networks using platforms such as Cytoscape.

Results

The combination of the data obtained from all MudPIT experiments allowed the identification of 1500 different proteins and 4800 peptides. The protein profiles identified were found to be useful in cluster analysis of analyzed tissues, which grouped according to their specific biological condition. Finally, the investigation of the protein-protein interaction network allowed the identification of sub-networks potentially involved in disease state and, at the same time, the visualization of a large amount of data.

Discussion

Employment of bioinformatics tools to handle and process the vast amount of data produced by “omics” technologies is fundamental to increase our knowledge of the biological systems investigated. In particular, the identification of biomarkers and the characterization of biological systems as scale free networks proved to be useful in the elucidation of mechanisms indicative of diseased phenotypes.

Presenting Author

Dario Di Silvestre (dario.disilvestre@itb.cnr.it)

Institute for Biomedical Technologies - National Research Council

Author Affiliations

(1)- Institute for Biomedical Technologies - CNR, Segrate (Milan), Italy. (2)- Scuola Superiore Sant'Anna - Sector of Medicine, Pisa, Italy. (3)- University Hospital S. Orsola Malpighi - Cardiology Department, Bologna, Italy.

G-47. ABC database: the Analysing Biomolecular Contacts database

Walter P (1), Metzger J (1,), Helms V (1)*

Protein-protein interactions are essential for most cellular processes such as signal transduction and immune response which makes them attractive targets for a wide range of pharmaceutical interventions. Thus, it is highly desirable to understand the molecular details of protein-protein contacts as well as contacts between proteins and small molecules. Studying such interactions is greatly facilitated by convenient simultaneous access to as much data as possible. For this reason we have designed the ABC database that provides a broad range of different precomputed interface features.

Materials and Methods

The protein complexes stored in our database are retrieved from the RCSB Protein Data Bank. Besides, we included some additional information such as CATH and SCOP classification, downloaded from the corresponding websites. The data import was done using a JAVA routine. The database and the web interface were created with non-commercial or freely available tools. The program logic was written in JAVA using Servlets and JSP. As web server we use Tomcat for the management of Servlets and JSP pages as well as Apache for the static web content. The relational database was realised with MySQL.

Results

The database can be queried by various types of parameters including criteria such as CATH denotation, polarity and size. After performing a query the user obtains a list of interfaces that fit the constraints. For each entry found further information can be displayed including statistics of the amino acid composition and a list of the residue-residue contacts. It is also possible to export the retrieved data into text files. We used the database to compare protein-protein interfaces with protein-small molecule interfaces according to features like amino acid composition and interface size.

Discussion

There exist numerous freely available meta-databases that provide convenient access to either data on protein-protein complexes or on protein-small molecule complexes taken from the PDB. To our knowledge, no meta-database so far combines both types of interfaces into one resource. Our ABC database is therefore a powerful tool allowing novel statistical analyses. For example, researchers interested in small-molecule modulators or inhibitors of protein-protein interactions may wish to analyze shared parts of protein-protein and protein-ligand interfaces.

URL

<http://service.bioinformatik.uni-saarland.de/abc>

Presenting Author

Jennifer Metzger (jmetzger@bioinformatik.uni-saarland.de)

Center for Bioinformatics, Saarland University

Author Affiliations

(1) Center for Bioinformatics, Saarland University

G-48. The core and pan metabolism in the *Escherichia coli* species

Vieira G (1,*), Sabarly V (2,3), Bourguignon P-Y (4), Durot M (1), Le Fèvre F (1), Mornico D (1), Vallenet D (1), Bouvet O (2), Denamur E (2), Schachter V (5), Médigue C (1)

Escherichia coli has a highly dynamic genome, a plasticity that is an important factor for evolution and pathogenicity. It is difficult to link genomic elements to phenotype, due to different layers of interaction and to evolutive convergence. Since most shared genomic elements are linked to metabolism, is it possible to use metabolism as an integrative layer? Unfortunately, no frameworks for metabolic network reconstruction of closely related strains exist yet. Could we provide highly detailed networks, highlighting links between evolution and metabolism?

Materials and Methods

We reconstructed the metabolic networks of 23 *E. coli* (8 commensal and 15 pathogenic strains) and 6 *Shigella* (pathogens). After having reannotated all genomes, we built all networks using EcoCyc as a primary high-quality pivot and MetaCyc. We computed and compared the core/pan genome and metabolism for all strains. We reconstructed the phylogenetic (metabolic) tree by maximum of likelihood (neighbour joining). We computed the intra- and inter-group phylogenetic distance. We conducted a Multiple Correspondence Analysis and a Classification And Regression Tree on the reactions.

Results

The quality of the annotation (i) and the introduction of a pivot (ii) increased the network quality: i) Increasing the number of reactions by 24%. ii) Decreasing the number of reactions without associated genes by 34 %. On average the networks possessed 1491 reactions, half of them were common to all the strains, biosynthesis processes being the most conserved. The different analyses showed a strong correlation between phylogenetic groups and metabolic networks. We also found reactions linked to pathogenicity, some already known and others requiring further investigation.

Discussion

By propagating specific knowledge from annotation and benefiting from expert curation made on the pivot, we were able to reconstruct detailed networks. Those networks closely mirror the phylogeny. Metabolism is thus a way to link genome and phenotype; we found metabolic processes specific to each phylogenetic group. Among those candidates, some are putative reactions. Currently, the Pan metabolism is underestimated and it will be more interesting to focus on filling metabolic gaps than adding new networks. These networks work will serve as a basis to reconstruct whole cell metabolic models.

URL

<http://www.genoscope.cns.fr/agc/metacoli/>

Presenting Author

Gilles Vieira (gvieira@genoscope.cns.fr)

CNRS-UMR 8030, Laboratoire d'Analyse Bioinformatique en Génomique et Métabolisme, Commissariat à l'Energie Atomique (CEA), Direction des Sciences du Vivant, Institut de Génomique, Genoscope

Author Affiliations

1: CNRS-UMR 8030, Laboratoire d'Analyse Bioinformatique en Génomique et Métabolisme, Commissariat à l'Energie Atomique (CEA), Direction des Sciences du Vivant, Institut de Génomique, Genoscope, 2 rue Gaston Crémieux, 91057 Evry Cedex, Evry cedex F-91006 2: INSERM U722 and Université Paris 7, 16 rue Henri Huchard, 75018 Paris, France 3: INRA, UMR de Génétique Végétale, INRA/CNRS/Univ Paris-Sud/AgroParistech, Ferme du Moulon, F-91190 Gif sur Yvette, France: 4: Max Planck Institute for Mathematics in the Sciences, Inselstr. 22, D-04103 Leipzig, Germany 5: TOTAL Gas & Power 2, place Jean Miller - La Défense 6 92078 Paris La Défense Cedex - France

Acknowledgements

This work is supported by a grant from the French National Research Agency (ANR): METACOLI project, ref. ANR-08-SYSC-011.

G-49. BioGraph: discovering biomedical relations by unsupervised hypothesis generation

Liekens A (1,*), De Knijf J (2), Daelemans W (3), Goethals B (2), De Rijk P (1), Del-Favero J (1)

For the computational identification of suitable targets among candidate genes in a biomedical context, the intelligence and intelligibility of the method are of vital importance for evaluating the prioritizations. Protein-protein interaction networks are often adopted, but are limited in functional expressivity. The integration with multiple types of biomedical knowledge can enhance the quality of automatically generated functional hypotheses relating contexts, e.g., a disease, and target sets, e.g., a set of candidate genes.

Materials and Methods

We propose a data mining framework that allows for the automated formulation of comprehensible functional hypotheses relating a context to targets. The method is based on the integration of heterogeneous biomedical knowledge bases and yields intelligible and literature-supported indirect functional relations. By assessing the plausibility and specificity of these hypothetical functional paths within a user-provided research context, the unsupervised methodology is capable of appraising and ranking of research targets, without requiring prior domain knowledge from the user.

Results

We highlight this methodology's application in the prioritization of susceptibility genes for hereditary diseases. We show that the proposed framework outperforms leading technologies on published benchmarks (AUC 91.31% on Endeavour benchmark) and is capable of robustly predicting recently discovered susceptibility genes for a range of hereditary psychiatric disorders and for suggesting new genome-wide putative candidates. The data mining method is publicly accessible via a web service at <http://www.biograph.be>

Discussion

Our proposed methodology offers a range of significant improvements over leading bioinformatics platforms for in silico identification of susceptibility genes: highly ranked targets are grounded in intelligible putative functional hypotheses with rich semantics, verifiable by their references in the literature. The method is unsupervised and does not require prior domain knowledge from the user. Beyond disease-gene applications, the method is applicable in various biological research settings requiring the intelligent and intelligible identification of promising research targets

URL

<http://biograph.be>

Presenting Author

Anthony M.L. Liekens (anthony@liekens.net)

VIB Department of Molecular Genetics, University of Antwerp, Belgium

Author Affiliations

(1) Applied Molecular Genomics group, VIB Department of Molecular Genetics, Universiteit Antwerpen, Antwerpen, Belgium, (2) Advanced Database Research and Modelling group, Department of Mathematics and Computer Science, Universiteit Antwerpen, Antwerpen, Belgium, (3) Computational Linguistics and Psycholinguistics Research Center, Universiteit Antwerpen, Antwerpen Belgium

G-50. Tuning noise propagation in a two-step series enzymatic cascade

Dhananjayulu V (1), Sagar PVN (1), Kumar G (1), Viswanathan GA (1,)*

Noise or fluctuations propagate and distribute via signaling pathways in cells, and thereby affect cellular response. While noise might have to be minimized in certain systems, it may be used as part of cellular decision-making process. Characterization of noise propagation in signaling pathways can provide useful insights into the functioning of cells in the presence of noise. Signaling pathways consists of several enzymatic building blocks such as Mitogen Activated Protein Kinase (MAPK) cascades. In this study, we characterize noise propagation in a two-step series MAPK enzymatic cascade.

Materials and Methods

The stochastic dynamics of the proteins involved in the enzymatic cascade is captured using chemical master equations. Using Gillespie simulations, we obtained noise dynamics and estimate (intrinsic) noise propagation in the cascade due to the stochastic nature of the biochemical reactions. We detected extrinsic noise propagation in the cascade using linear stability analysis of the stochastic differential equation (SDE) of the Langevin type and by solving the full SDE. We used global sensitivity analysis to study the effect of various system parameters on noise propagation in the cascade.

Results

Stochastic simulations suggest that intrinsic noise propagation dominates over the extrinsic noise. Extrinsic noise estimates obtained using steady state perturbation analysis are comparable to those obtained using stochastic simulations. Simulations indicate that there exists a critical upstream enzyme concentration below which noise propagation in the cascade is amplified. Sensitivity analysis shows that the extent of noise propagation is mostly controlled by the total concentrations of upstream enzyme and substrate of the downstream enzymatic reaction cycle.

Discussion

Steady state perturbation analysis, which is computationally less intensive compared to the full stochastic simulations, provide a good extrinsic noise propagation estimate. However, for prediction of intrinsic noise propagation, stochastic simulations, though tedious and computationally intensive, has to be performed. Noise propagation in the enzymatic cascade can be tuned by modulating the upstream enzyme concentration and/or the substrate of the downstream enzymatic reaction.

Presenting Author

Ganesh A. Viswanathan (ganesh@che.iitb.ac.in)

Department of Chemical Engineering, Indian Institute of Technology Bombay

Author Affiliations

Department of Chemical Engineering Indian Institute of Technology Bombay Powai, Mumbai 400076 India

Acknowledgements

Department of Science and Technology, Ministry of Science and Technology, India

G-51. The effect of interactome evolution on network alignment

Ali W (1,), Deane CM (1)*

Local alignments of protein interaction networks have found little conservation among several species. While this could be a consequence of the incompleteness of interaction data-sets and presence of error, an intriguing prospect is that the process of network evolution is sufficient to erase any evidence of conservation. Here, we aim to test this hypothesis using models of network evolution and also investigate the role of error in the results of network alignment.

Materials and Methods

We used the duplication-divergence and geometric models of network evolution to grow pairs of networks to the size of experimental datasets. The network evolution was supplemented with a process of ortholog evolution to model the distribution of orthologous relationships in real proteomes. Network alignments were then carried out on pairs of real and modelled data-sets and a distance metric based on graph statistics was used to quantify the agreement between the alignments. Error was finally introduced in the modelled networks to explain the discrepancy between real and modelled alignments.

Results

Our results indicate that evolution alone is unlikely to account for poor quality alignments given by real data. Alignments of modelled networks undergoing evolution are at least 4-5 times larger than real alignments. We compare several error models in their ability to explain this discrepancy. Estimates of false negative rates vary from 20 to 60% dependent on the inclusion of the effect of incomplete proteome sampling. We find that false positives affect network alignments little compared to false negatives indicating that incompleteness is the major challenge for interactome comparisons.

Discussion

While network evolution does play a considerable role in network alignment quality, the effect of error cannot be discounted. This is especially true for false negatives which seems to mostly account for the extremely low conservation observed in real interactomes. Moreover, using a quantitative comparison of real and simulated alignments makes it possible to arrive at independent estimates of network error and evaluate likely error models.

Presenting Author

Waqar Ali (ali@stats.ox.ac.uk)

Department of Statistics, University of Oxford

Author Affiliations

(1) Department of Statistics, University of Oxford, United Kingdom

Acknowledgements

Department of Statistics, University of Oxford

G-52. Analyzing compounds' mode of action by biological relatedness of proteins

Schmid R (1,*), Baum P (1), Itrich C (1), Fundel-Clemens K (1), Lämmle B (1), Birzele F (1), Weith A (1), Brors B (2), Eils R (2,3), Mennerich D (4), Quast K (1)

Identification of on- as well as off-target effects of drugs remains a common objective in scientific research. These effects are mediated by binding of compounds to proteins and subsequently influencing related regulatory networks within the organism. Discovering key proteins as well as their meaning in a broader biological context seems to be the most promising way to answer questions with respect to mode of action and cellular processes. As more and more data from diverse sources is available, the integration of this knowledge is an important step to get a deeper insight into biology.

Materials and Methods

Protein interactions are taken from iRefIndex and are enriched by literature derived confidence scores, information about transcription factor binding sites, Gene Ontology annotation as well as gene expression data. This accumulated information is translated to edge weighted graphs. Finally, subnets are extracted based on interactions exhibiting high biological relatedness with respect to the assigned weights. Results are quantitatively evaluated by a comparison to randomized networks. To elucidate the biological context of the subnets Fisher's exact test is conducted on predefined gene sets.

Results

We are able to derive protein interactions that are meaningful in the biological context under consideration. The results were biologically validated using gene expression measured in TGF- β stimulated cells. First, an enrichment of TGF- β related gene sets within the extracted nets could be detected. Second, we were able to detect on- as well as off-target effects of TGF- β receptor 1 kinase inhibiting compounds. On-target effects are detected by the direction of deregulation in the extracted nets, off-target effects are revealed by enrichment of genes related to off-target signaling cascades.

Discussion

Despite the extreme flexibility and simplicity of our approach, we achieve very good and traceable results. Based on the research objective the scores can be individually weighted and optimized and additional (experimental) data can be added. These are the main advantages towards other, more abstract approaches. As next steps, we plan to compare results based on our network to STRING based results and contrast our method with similar approaches. Finally, we plan to integrate phenotypic information of genes to gain even more value for biological analyses and publish our method as an R-package.

Presenting Author

Ramona Schmid (ramona.schmid@boehringer-ingenelheim.com)
Boehringer Ingelheim Pharma GmbH & Co. KG

Author Affiliations

1 Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach/Riss, Germany, 2 Theoretical Bioinformatics Department, German Cancer Research Center (DKFZ), Heidelberg, Germany, 3 Department of Bioinformatics and Functional Genomics, Institute of Pharmacy and Molecular Biotechnology (IPMB) and BioQuant, University of Heidelberg, Heidelberg, Germany, 4 Boehringer Ingelheim Pharmaceuticals Inc., Ridgefield, CT 06877, USA.

G-53. Logical modelling of the regulatory network controlling the formation of the egg appendages in *Drosophila*

*Fauré A (1, *), Vreede B (1), Sucena É (1), Chaouiya C (1)*

Formation of the dorsal appendages in *Drosophila* provides an interesting model to study epithelial morphogenesis. Experimental studies have identified Gurken and Decapentaplegic as the two main regulatory signals that are responsible for the patterning of the follicle cell epithelium that will secrete the eggshell later in oogenesis. However, the pattern of expression of the markers that characterise the different regions is dynamic in both space and time, and different groups report contradictory results. We use a modelling approach to improve our understanding of this system.

Materials and Methods

The logical formalism offers a simple framework to study regulatory networks. Briefly, regulatory components are represented by variables that can take a finite number of discrete values representing their level of activity with respect to a given threshold. Logical rules determine the target level of each variable depending on those of its regulators. Beyond topological properties of the wiring diagram, the logical formalism offers tools to study dynamic properties such as attractors, and logical models can be used to simulate experiments *in silico* and make new predictions.

Results

We review the literature to extract a map of the regulatory network controlling dorsal appendages formation in *Drosophila melanogaster*. Using the logical modelling software GINsim, we further simplify this map to build a dynamic model of the intracellular pathway, with Gurken and Decapentaplegic as inputs, and Rhomboid, Broad and Fasciclin III as reading outputs. Our preliminary results support the hypothesis that Broad inhibits Fas3 in the roof cells.

Discussion

In spite of several modelling efforts, the precise wiring of the regulatory network remains elusive. Our map synthesises in an intuitive format the current knowledge of the control of dorsal appendage patterning in *Drosophila melanogaster*. We have yet to confirm experimentally our modelling predictions. In the near future we also plan to use our cellular model as a module to build a multicellular model of the antero-dorsal follicle cell epithelium.

Presenting Author

Adrien Fauré (afaure@igc.gulbenkian.pt)
Instituto Gulbenkian de Ciência

Author Affiliations

(1) Instituto Gulbenkian de Ciência, Oeiras, Portugal

Acknowledgements

A. Fauré acknowledges funding from the Fundação para a Ciência e a Tecnologia, Portugal (project PTDC/EIACCO/099229/2008).

G-54. A curated database of microRNA mediated feed-forward loops involving MYC as master regulator

El Baroudi M (1,2), Corà D (2,3,), Bosia C (1,2), Osella M (1,2), Caselle M (1,2)*

The MYC Transcription Factors (TFs) are one of the most important family of regulators in human. They may act both as activators or repressors of their target genes and are involved in a host of key biological processes, ranging from cell cycle progression, to apoptosis and cellular transformation. They are also known to be involved in the biology of many human cancer types. Do date, little is known about the MYC / microRNAs cooperation in the regulation of genes at the transcriptional and post-transcriptional level.

Materials and Methods

In order to investigate the interplay between MYC and microRNAs, we concentrated our attention on the study of mixed microRNA / TF Feed-Forward Regulatory Loops (FFLs), i.e. elementary regulatory circuits in which a master TF regulates an microRNA and together with it a set of joint target protein-coding genes. Employing independent databases with experimentally validated data, we identified several mixed FFLs regulated by MYC and characterized completely by experimentally supported regulatory interactions, in human.

Results

We were able to identify a total of 130 mixed FFLs having MYC as master regulator, involving 36 MYC regulated microRNAs and 82 joint target protein-coding genes. Out of these 130 FFLs, 28 could be classified as incoherent (type_I) and 42 as coherent (type_II). We then study the statistical and functional properties of these circuits, showing that this class of loops is over-represented in the human regulatory network, it is functionally related to cancer and shows a remarkable redundancy of the microRNA branch. Finally, we discuss a few examples involving E2F1, PTEN, RB1 and VEGF.

Discussion

It has become by now clear that the interplay between transcriptional and post-transcriptional (microRNA mediated) regulation plays a crucial role in the modulation of gene expression. We report the assembly and characterization of a catalogue of human mixed TF / microRNA Feed-Forward Loops, having MYC as master regulator and completely defined by experimentally verified regulatory interactions. This study allowed us to draw a new and rich picture of the MYC centred network, revealing complex interplays with potential important role in defining accurate expression levels for many key genes.

Presenting Author

Davide Cora' (davide.cora@ircc.it)

Systems Biology Lab - Institute for Cancer Research and Treatment (IRCC), School of Medicine, University of Torino.

Author Affiliations

1) Department of Theoretical Physics, University of Torino and INFN, Via P. Giuria, 1. I-10125 Torino, Italy. 2) Center for Molecular Systems Biology, University of Torino, Via Accademia Albertina 13, I-10100 Torino, Italy 3) Systems Biology Lab, Institute for Cancer Research and Treatment (IRCC), School of Medicine, University of Torino, Str. Prov. 142 Km. 3.95, I-10060 Candiolo, Torino, Italy

G-55. The Biological Connection Markup Language (BCML): a SBGN compliant data format for visualization, filtering and analysis of biological pathways

Calura E (1,2,=), Beltrame L, (2,=), Popovici R (3,4), Rizzetto L (2), Rivero Guedez D (2), Donato M (3), Romualdi C (1), Draghici S (3), Cavalieri D (2)*

Throughout the years, many data models for pathways have been proposed, although successful, they did not tackle the underlying biological problem, especially for signaling pathways. These have interconnected, finely regulated structures and are affected by environment-dependent changes. Current models represent pathways regardless of these features. However, an imprecise model affects the power of the analyses carried out with it. BCML allows an unambiguous and fully dynamic representation of signaling pathways that is useful for both biologists and bioinformaticians.

Materials and Methods

An XML schema describing the complete SBGN Process Description version 1.1 was implemented according to the SBGN reference documentation. An additional schema, was written as an extension of the main SBGN schema. A set of utilities to validate, filter and extract entities from the schema was also developed in the Java programming language. In particular, the utilities support consistency checking, export to analysis formats, annotation, pathway filtering and visualization of BCML compliant files.

Results

The model, defined as an XML Schema, incorporates all the entities and interactions defined by the SBGN PD specification, including the rules and restraints that apply to them. The model can store additional information on the entities that compose the network: each entity can have identifiers linked to public databases and information can be stored for multiple species in the same entity definition. Data stored in a BCML compliant file can be manipulated mainly in four ways: discriminative selection of the pathway features, graphical representation, incorporation of experimental measurements.

Discussion

Most pathways present in public databases, although described for different organisms, lack information on the biological environment where they have been described. Basing on that we believe that improvements in the analysis of pathway will come from a better use of the knowledge present in scientific literature. BCML, which stores deeper descriptions of biological knowledge, turns out to be a format extremely suitable for advanced pathway analysis methods and its dynamic nature makes it an important tool for the dissection of complex, highly specific biological problems.

Presenting Author

Enrica Calura (enrica.calura@gmail.com)
University of Padova

Author Affiliations

1- Department of Biology, University of Padova, Padova, Italy 2- Department of Pharmacology, University of Firenze, Firenze, Italy 3- Department of Computer Science, Wayne State University, Detroit, USA 4- Miravtech ± These authors contributed equally to this work.

Acknowledgements

This work was financially supported by the Network of Excellence DC-THERA, Dendritic cells for novel immunotherapies (www.dc-thera.org) and by the Network of Excellence SYBARIS (Grant Agreement 242220), University of Firenze and University of Padova. We are grateful to the DC-Thera curation team specially to Maria Cristina Gauzzi and Sandra Gessani, Sonja Bushow and Carl Figdor, Philippe Pierre, Walter Reith, Andrea Splendiani and Jon Austin. We thank Gabriele Sales for helpful discussions.

G-56. Estimating the size of the *S.cerevisiae* interactome

Sambourg L (1,), Thierry-Mieg N (1)*

As protein interactions mediate most cellular mechanisms, protein-protein interaction networks are essential in the study of cellular processes. Consequently, many large scale interactome mapping projects have been undertaken, and protein-protein interactions are being distilled into databases through literature curation; yet protein-protein interaction data are still far from comprehensive, even in the model organism *S.cerevisiae*. Estimating the interactome size is important for evaluating the completeness of current datasets, in order to measure the remaining efforts that are required.

Materials and Methods

We show that literature curation has reported many more interactions for highly studied proteins than for poorly studied ones. So literature-curated data allows us to precisely assess the size of the 'well-studied' sub-network (sub-network containing proteins cited in more than 125 papers). We then compute the proportion of interactions that HT experiments detect for this sub-network. Assuming that the HT experiments coverage is independent of how thoroughly a protein has been studied allows to apply this coverage to the entire network, and gives an estimate of the interactome size.

Results

Several estimates of the size of the *S. cerevisiae* interactome have been proposed, but none of them directly take into account information from both literature-curated and high-throughput data. We propose here a simple and reliable method for estimating the size of an interactome, combining these two data sources. Our method yields an estimate of at least ~ 35,000 direct physical protein-protein interactions in *S.cerevisiae*.

Discussion

Our method allows to free oneself from the necessity of a gold standard and takes advantage of almost all available data, leading to a more reliable estimate. Contrary to several other methods, our estimate is very robust with respect to the choice of the high-throughput dataset used. Finally, combining literature-curated and high-throughput data leads to higher estimates of the *S. cerevisiae* interactome size. This confirms the complementarity of these two data sources, and provides a sobering view of the coverage of current interactomes: extensive efforts and/or new methods are sorely needed.

Presenting Author

Laure Sambourg (laure.sambourg@imag.fr)
CNRS UMR5525

Author Affiliations

CNRS UMR5525

G-57. Laplacian eigenmaps, penalized principal component regression on graphs, and analysis of biological pathways

Shojaie A (1,), Michailidis G (1,2)*

Gene, protein and metabolite networks provide valuable information on how components of biological systems interact with each other to carry out vital cell functions. The behavior of complex biological systems can only be understood by incorporating information on interactions among components of the system and by analyzing the effect of biological pathways, rather than individual components. The proposed method incorporates the available network information and provides a flexible framework for analysis of pathway effects.

Materials and Methods

In this poster, we establish a connection between Laplacian eigenmaps and principal components of the covariance matrix and propose a dimension reduction method that directly incorporates the network information. Using this framework, the significance of biological pathways can then be analyzed by solving an eigenvalue problem on the graph with Neumann boundary conditions. We hence reformulate the problem of analysis of biological pathways as a principal component regression problem on the graph, a group-lasso penalty to determine the significance of each subnetwork.

Results

We evaluate the performance of the proposed method using simulated data, as well as real gene expression data on yeast Galactose Utilization. The findings of this analysis suggest that the proposed method offers better efficiency compared to Gene Set Enrichment Analysis (GSEA) and is a computationally efficient alternative to full network models for analysis of biological pathways, such as our previously proposed NetGSA method.

Discussion

The proposed method offers a systematic approach for dimension reduction in networks, with a priori defined subnetworks of interest. It can also incorporate both weighted and unweighted adjacency matrices and can be easily extended for analysis of complex experimental conditions using the framework of generalized linear models. This method can also be used to assess the effect of biological pathways in longitudinal and time-course studies. Our simulation studies, and the real data examples suggest that the method offers significant improvements over methods of gene set enrichment analysis.

Presenting Author

Ali Shojaie (shojaie@umich.edu)
University of Michigan

Author Affiliations

(1) Department of Statistics, University of Michigan (2) Department of Electrical Engineering and Computer Science, University of Michigan

Acknowledgements

This work was partially supported by grant 1RC1CA145444-0110 from the US National Institute of Health (NIH).

G-58. Predicting metabolic pathways from bacterial operons and regulons

Faust K (1,*), Croes D (2), Dupont P (3), van Helden J (2)

The pathway extraction tool predicts metabolic pathways from sets of functionally related enzyme-coding genes [Faust, et al., *Bioinformatics* 2010, 26:1211]. In contrast to pathway projection approaches, pathway discovery does not rely on any assumption of pathway conservation. This approach can detect variants, super-pathways, and “cross-map” paths. It can be applied to organisms whose metabolism is unknown, but for which enzyme-coding genes have been identified in the genome, and some information is available about their functional grouping (co-expression, operons, gene fusion, ...).

Materials and Methods

To extract a pathway, we connect query compounds or reactions in the metabolic network using various algorithms: the random walk based kWalks algorithm, three approaches based on k-shortest path finding and combinations of kWalks with the latter. When predicting pathways from genes, we have to link enzyme-coding genes to reactions. In case of broad-specificity enzymes, this yields a large number of reactions, only a few of which are relevant in the pathway. To address this problem, reactions catalyzed by the same EC number are merged into equivalence groups.

Results

We evaluated the pathway extraction algorithms on 71 MetaCyc pathways and found that a combination of kWalks with a shortest-paths based approach yields the highest accuracy (77%). The pathway extraction tool has been integrated in the Network Analysis Tools (NeAT, <http://rsat.ulb.ac.be/neat/>), and can be accessed via a web interface, as stand-alone application or as SOAP/WSDL Web services. The seed nodes for subgraph extraction can be provided as reactions, (partial) compound names as well as EC numbers or genes. Pathways can be extracted from KEGG, MetaCyc or custom networks.

Discussion

We will present a selection of study cases illustrating the way to combine operon prediction, phylogenetic footprint discovery and pathway extraction in order to infer metabolic pathways from bacterial genomes. In future, this strategy will be systematically applied on bacterial genomes in the framework of the MICROME project (EU FP7).

URL

<http://rsat.ulb.ac.be/neat/>

Presenting Author

Jacques van Helden (Jacques.van.Helden@ulb.ac.be)

Université Libre de Bruxelles

Author Affiliations

1. Bioinformatics and (Eco-)Systems Biology (BSB). Vrije Universiteit Brussel, Pleinlaan 2. B-1050 Brussels, Belgium. Email: kfaust@vub.ac.be 2. Laboratoire de Bioinformatique des Génomes et des Réseaux (BiGRé). Université Libre de Bruxelles, Campus Plaine, CP 263. Bld du Triomphe. B-1050 Bruxelles, Belgium. Email: Jacques.van.Helden@ulb.ac.be 3. UCL Machine Learning Group, B-1348 Louvain-la-Neuve, Belgium. Pierre.Dupont@uclouvain.be

Acknowledgements

This work is supported by the MICROME Collaborative Project funded by the European Commission within its FP7 Programme, under the thematic area "BIO-INFORMATICS - Microbial genomics and bio-informatics", contract number 222886-2. The BiGRé laboratory is also supported by the Belgian Program on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office, project P6/25 (BioMaGNet). The contribution of Karoline Faust was supported by the Actions de Recherches Concertées de la Communauté Française de Belgique (ARC grant number 04/09-307).

G-59. Inferring translationally active RNA binding proteins: mRNA interactions from polysomal profiling data with a Bayesian inference approach

Tebaldi T (1,), Sanguinetti G (2), Niranjan M (3), Quattrone A (1)*

Post-transcriptional control of gene expression is strongly mediated by the action of RNA binding proteins (RBPs), capable of uncoupling changes in the abundances of mRNAs from changes in their polysomal access and therefore their availability to translation. Despite this, the binding preferences of the majority of RBPs and the specific interactions between individual RBPs and their mRNA targets are still poorly known.

Materials and Methods

We consider a linear model relating changes in the polysomal levels of RNA binding proteins to changes in the polysomal loading ratios of their putative target mRNAs. A Bayesian network is used to extract putative RBP-mRNA connections. The joint posterior probability of the network is estimated with a Gibbs Sampler. Clustering of RBPs takes place contemporaneously in order to solve dimensionality problems.

Results

The method has shown learning capability with synthetic data and it has been applied to the prediction of post transcriptional networks in yeast, starting from six experiments where both transcriptome and translome profiling microarray raw data were available.

Discussion

Uncoupling between transcriptome and translome changes is a general and measurable phenomenon, but little is known about the underlying molecular mechanisms which affect the fate of every mRNA molecule once it is transcribed. The presented approach is a first step to untangle the complexity of this problem and make reasonable functional predictions from existing data.

Presenting Author

Toma Tebaldi (tebaldi@science.unitn.it)

Centre for Integrative Biology

Author Affiliations

(1) Centre for Integrative Biology, University of Trento. (2) Institute for Adaptive and Neural Computation, University of Edinburgh (3) School of Electronics and Computer Science, University of Southampton

G-60. DASS-GUI: a new data mining framework

Hollunder J (1,), Van de Peer Y (1), Wilhelm T (2)*

“Data mining is the process of extracting patterns from data. Data mining is becoming an increasingly important tool to transform these data into information” (wikipedia). The growing importance of data mining is especially evident in biology, where modern ‘omics’ technologies are producing increasingly complex data. Traditional data mining techniques such as APRIORI and CHARM are predominantly applied outside the realm of biology. However, the closely related concept of biclustering has already found many biological applications, such as gene expression analysis.

Materials and Methods

DASS-GUI processes several steps including preparing the data, applying data mining algorithms, and evaluating the results. DASS-GUI runs in two modes namely calculation (pattern identification) and analysis mode (evaluation of the identified patterns). The identification step can be done by either our DASS-cs or one of two other implemented state-of-the-art algorithms (LCM and FPclose). DASS-cs has three outstanding features (a) analyzing multi-sets, (b) similarity pruning, and (c) significance calculation.

Results

The power and versatility of the DASS approach was already demonstrated in a number of different biological applications, for instance for the analysis of protein complexes and corresponding protein annotations (Hollunder et al., 2005; Hollunder et al., 2007), the identification of transcription factor binding site modules (Beyer et al., 2006), and the analysis of multi-domain proteins as well as the identification of conserved subnetworks in different species (Hollunder et al., 2007).

Discussion

Only DASS-cs allows similarity pruning during closed set calculation to avoid redundant clusters. Next straightforward steps are to integrate a pipeline preparing the data for directly working with expression data as well as to apply the extracted information for predictions and classifications. Generally, the identified patterns and rules can guide decision making and forecast effects of interactions.

URL

<http://www.ifr.ac.uk/dass/gui/>

Presenting Author

Jens Hollunder (Jens.Hollunder@psb.vib-ugent.be)

Department of Plant Systems Biology, VIB, Ghent University

Author Affiliations

(1) Department of Plant Systems Biology, VIB, Ghent University, Technologiepark 927, 9052 Gent, Belgium (2) Theoretical Systems Biology, Institute of Food Research, Norwich Research Park, Conley Norwich NR4 7UH, U.K.

Acknowledgements

We acknowledge the support of Ghent University (Multidisciplinary Research Partnership “Bioinformatics: from nucleotides to networks”) and Biotechnology and Biological Sciences Research Council Core Strategic Grant for the Institute of Food Research.

G-61. EnrichNet: network-based gene set enrichment analysis

Glaab E (1,), Baudot A (2), Krasnogor N (1), Valencia A (2)*

Classical overlap-based gene set enrichment analysis is often used to examine the significance of association between gene sets of interest and cellular processes. However, functional pathway/process annotations are still missing for many genes. Molecular interaction data might help to uncover some of this missing information. Here, we propose to measure the association between gene sets and functional pathways/processes based on the distances between corresponding proteins in a protein-protein interaction network, instead of only using the overlap of gene sets to estimate their association.

Materials and Methods

We present a new network-based enrichment analysis method. The approach maps gene and protein sets of interest onto a protein interaction network and computes the distribution of shortest path distances between all protein pairs from different sets. This distribution is then compared against the distribution across all pathway/process protein sets from a chosen database (KEGG, BioCarta, Reactome, etc.) using the Xd-distance (Olmea et al., 1999), which assigns higher weights to smaller distances, ensuring that large-distance outliers cannot distort the results.

Results

When applying our network-based gene set association scoring method to search for pathways altered in cancer (calculating network distances between known cancer genes, Futreal et al., 2004, and KEGG signalling pathways) a high correlation of pathway ranks to classical enrichment ranks is obtained, but our method additionally enables the ranking of non-overlapping gene sets. Moreover, the approach provides a visual analysis of sub-networks corresponding to the input gene/protein sets to identify interactions between proteins of interest and potential linking proteins.

Discussion

The network-based enrichment analysis method enables the combination of experimental evidence from multiple sources, different interaction networks (protein-protein, genetic interactions, etc.) and experimental data on genes and proteins (DNA microarrays, Western blots, RNA sequencing, gene mutation and methylation, etc.), helping to fill knowledge gaps in single experiments and increasing the robustness and interpretability of enrichment analysis.

URL

<http://www.infobiotics.net/enrichnet>

Presenting Author

Enrico Glaab (enrico.glaab@cs.nott.ac.uk)
Nottingham University

Author Affiliations

1. School of Computer Science, Nottingham University, Jubilee Campus, NG8 1BB Nottingham, UK 2. Structural Biology and Biocomputing Program, CNIO, E-28029 Madrid, Spain

Acknowledgements

We acknowledge support by the Biotechnology and Biological Sciences Research Council (BB/F01855X/1), the Spanish Science and Innovation (MICINN) grant (BIO2007-66855, "functions for gene sets") and Instituto de Salud Carlos III RTIC COMBIOMED network (RD07/0067/0014). AB is supported by the Juan de la Cierva postdoctoral fellowship.

G-62. Interacting copy number alterations in breast cancer

Canisius S (1,), Klijn C (1,2), Smid M (3), Martens J (3), Foekens J (3), Wessels L (1,2)*

One of the challenges faced by cancer research is finding out which genes play a role in the formation and growth of tumours. While many such genes have been found and described individually, it is known that deregulation of a single gene does not cause cancer. It is therefore equally important to reveal the interactions between genes that drive tumour development. Since DNA copy number alterations are an important cause of gene deregulation in cancer, we hypothesise that interactions between cancer genes can be found by detecting co-occurring copy number changes in a set of tumours.

Materials and Methods

For a set of 216 breast tumours, we obtained copy number data, generated with a high-density SNP array. An association matrix is constructed that scores the dependency between pairs of genomic loci. The estimates in this matrix are smoothed to account for the fact that copy number alterations tend to involve continuous stretches of DNA rather than isolated nucleotides. Given such a smoothed association matrix, regions of significant association are detected by a genome-wide permutation test that corrects for the massive amount of multiple testing.

Results

Analyses for co-occurring gains, losses, and gain/loss combinations revealed several genomic regions between which significant dependencies are observed. A closer inspection of these co-occurrences showed interesting links with known subtypes of breast cancer. With mRNA expression for the same set of tumours, we confirmed a positive relation between gene dosage and mRNA abundance. Moreover, protein interaction databases confirmed the existence of interacting genes in the co-occurring regions. These findings provide suggestions for the possible targets of the alterations.

Discussion

We present an overview of co-occurring copy number alterations in breast cancer. We show that such co-occurrences are a frequent phenomenon, and provide additional analyses that suggest an important regulatory role in tumorigenic processes. The strong association we found between copy numbers of genomic regions located on different chromosomes illustrate that tumour development not only selects for individual genes, but also for gene interactions. These interactions may be found to be the weak links in the regulatory networks that drive tumour development.

Presenting Author

[Sander V.M. Canisius \(S.Canisius@nki.nl\)](mailto:S.Canisius@nki.nl)

Netherlands Cancer Institute

Author Affiliations

(1) Division of Molecular Biology, The Netherlands Cancer Institute, Amsterdam, The Netherlands (2) Information and Communication Theory Group, Delft University of Technology, Delft, The Netherlands (3) Department of Medical Oncology, Erasmus Medical Center Rotterdam, Josephine Nefkens Institute and Cancer Genomics Centre, Rotterdam, The Netherlands

G-63. Compression of mass spectral imaging data using discrete wavelet transform guided by spatial information

Verbeeck N (1,*), Van de Plas R (1,3), De Moor B (1,3), Waelkens E (2,3)

Mass Spectral Imaging is a relatively new molecular imaging technology that makes it possible to detect thousands of molecules throughout tissue simultaneously, ranging from low-mass metabolites to high-mass proteins. This technology is of prime interest for the molecular characterization of tissue in biomedical studies. In recent years, MSI data sets have grown in size to such extent that it becomes more and more infeasible to computationally analyze them in their raw form due to both memory and calculation time constraints.

Materials and Methods

Previous research at ESAT by Van de Plas et al. has shown solid results using Discrete Wavelet Transform (DWT) on mass spectra to perform feature selection, thus reducing data size, dimensionality and noise. Our newest method further improves on this approach by incorporating one of the key aspects of MSI, spatial information, to better understand what part of the data can truly be considered noise. By using this information, we can selectively remove only those details that do not exhibit a spatial structure.

Results

We demonstrate the performance of this new compression method on a sagittal section of mouse brain and compare the results to the Van de Plas et al. method and to direct analysis of the raw measurements. The presented study focuses on neurodegenerative diseases that show spatially specific behaviour. Examples of such diseases include Parkinson's disease, where dopamine producing brain nuclei such as the amygdala are affected, and amyotrophic lateral sclerosis, where motor neuron regions in the brain are affected.

Discussion

By retaining a small number of detail coefficients that express spatial structure we can strongly improve reconstruction of the mass spectra while still achieving considerable compression.

Presenting Author

Nico Verbeeck (nico.verbeeck@esat.kuleuven.be)

Katholieke Universiteit Leuven

Author Affiliations

1. Katholieke Universiteit Leuven, Dept. of Electrical Engineering (ESAT), SCD-SISTA (BIOI), Kasteelpark Arenberg 10, B-3001 Leuven (Heverlee), Belgium. 2. Katholieke Universiteit Leuven, Dept. of Molecular Cell Biology, O&N, Herestraat 49 - bus 901, B-3000 Leuven, Belgium. 3. Katholieke Universiteit Leuven, ProMeta, Interfaculty Centre for Proteomics and Metabolomics, O&N2, Herestraat 49, B-3000 Leuven, Belgium.

Acknowledgements

* Research Council KUL: ProMeta, GOA Ambiorics, GOA MaNet, CoE EF/05/007 SymBioSys, START 1, several PhD/postdoc & fellow grants * Flemish Government: o FWO: PhD/postdoc grants, projects, G.0318.05 (subfunctionalization), G.0553.06 (VitamineD), G.0302.07 (SVM/Kernel), research communities (ICCoS, ANMMM, MLDM); G.0733.09 (3UTR); G.082409 (EGFR); o IWT: PhD Grants, Silicos; SBO-BioFrame, SBO-MoKa, TBM-IOTA3 o FOD:Cancer plans * Belgian Federal Science Policy Office: IUAP P6/25 (BioMaGNet, Bioinformatics and Modeling: from Genomes to Networks, 2007-2011 * EU-RTD: ERNSI: European Research Network on System Identification; FP7-HEALTH CHartED

G-64. Learning ancestral polytrees for HIV-1 mutation pathways against nelfinavir

Li GD (1,*), Beheydt G (1), Bielza C (2), Larrañaga P (2), Camacho RJ (3,4), Grossman Z (5,6), Torti C (7), Zazzi M (8), Prosperi M (9,10), Kaiser R (11), Van Laethem K (1), De Maeyer M (12), Jansen M (13), Vandamme AM (1,3)

Probabilistic graphical models like Bayesian networks or dependency trees reveal HIV-1 drug resistance pathways. However, the training data usually originate from different geographic regions with varied viral subtypes, resulting in sampling biases. Moreover, factors not being encoded in graphical models (e.g. interactions between HIV inhibitors and host proteins) might induce confounding effects. To take into account sampling biases and confounding effects, we use ancestral polytrees to test their robust learning capacity in the analysis of HIV-1 resistance against NFV, a well studied drug.

Materials and Methods

We extracted sequences from 3000 PI naïve and 1637 NFV experienced patients from a database containing data from Belgium, Germany, Israel, Italy, Portugal, Stanford and Sweden. The most prevalent amino acid was considered the wild type. Amino acids whose prevalence was above 1% were selected. Wild type was only included if it appeared simultaneously with two other amino acids. We trained ancestral polytrees using a mutual information criterion, orientation principles and dependency analysis (Fisher's exact test). Robustness was assessed by a non-parametric bootstrap method (100 replicates).

Results

The subtype distribution was: B (62.5%), G (8.2%), C (7.3%), A1 (4.1%) and others. Among the selected 105 amino acids (40 wild types), 87 pairs had a high mutual information score (>0.01), while NFV had high scores only with two mutations, 30N (0.105) and 90M (0.029). NFV ancestral polytree included 106 variables and 105 edges, of which 76 arcs had bootstrap supports over 65%, and 47 over 90%. Only 30N (83%) and 90M (69%) were directly linked to NFV, whereas other known resistance mutations like 10F/I, 46I/L, 71V/T and 88D/S were indirectly linked within 5 arcs away from NFV.

Discussion

NFV ancestral polytree found 30N and 90M as only mutations directly linked with NFV, implying their importance. Besides, we found that 10F/I, 46I/L, 71V/T and 88D/S were close to NFV in the ancestral polytrees. Mutations at position 19, 36 and 41 do not cluster together while other positions were highly clustered (bootstrap $>65\%$). Interestingly, mutations close to NFV in the ancestral polytree were also found to be located around the NFV binding pocket within 10 angstroms. Although there were major agreements with mutagenetic trees and Bayesian networks, the differences will need further study

Presenting Author

Guangdi Li (guangdi.li@rega.kuleuven.be)

Laboratory for Clinical and Evolutionary Virology, Rega Institute for Medical Research, Katholieke Universiteit Leuven, Leuven, Belgium

Author Affiliations

1 Laboratory for Clinical and Evolutionary Virology, Rega Institute for Medical Research, Katholieke Universiteit Leuven, Leuven, Belgium, 2 Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid, Madrid, Spain, 3 Centro de Malária e Outras Doenças Tropicais, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisboa, Portugal, 4 Laboratório de Biologia Molecular, Centro Hospitalar de Lisboa Ocidental, Lisboa, Portugal, 5 National HIV Reference Lab, Central Virology, Public Health Laboratories, MOH, Sheba Medical Center, Ramat-gan, Israel, 6 School of Public Health, Sackler Faculty of Medicine, Tel Aviv University, Tel-Aviv, Israel, 7 Università degli Studi di Brescia, p.le Spedali Civili 1, 25123 Brescia, Italy, 8 Department of Molecular Biology, University of Siena, Siena, Italy, 9 Catholic University of Sacred Heart, Clinic of Infectious Diseases, Rome, Italy, 10 Informa SRL, Rome, Italy, 11 Institute of Virology, University of Cologne, Cologne, Germany, 12 Laboratory for Pharmaceutical Biology, Faculty of Pharmaceutical Sciences, Katholieke Universiteit Leuven, Leuven, 13 Department of Mathematics, Katholieke Universiteit Leuven, Leuven, Belgium.

Acknowledgements

Li GD is supported by CSC and K.U.Leuven. This study was partially supported by the Interuniversity Attraction Poles Programme, Belgian State, Belgian Science Policy (IAP-VI P6/41). The research leading to the results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under the project "Collaborative HIV and Anti-HIV Drug Resistance Network (CHAIN)" - grant agreement n° 223131.

G-65. Ranking of genes from RNAi perturbation data

Siebourg J (1,*), Beerenwinkel N (1)

Using the technique of RNA interference, genome wide knock down screens have been performed. Finding the genes whose knock down has the greatest impact on the inspected situation, leads to building an importance ranking of genes according to their experimental values. Creating such a ranking is a difficult task because several issues have to be faced. Often one faces the problem of low numbers of replicates in the data that don't allow for statistical robustness.

Materials and Methods

There are a lot of different ranking techniques available, such as RSA (König 2010) for example, but all lead to different results. Another group of methods aim to assemble different rankings into a single more robust one like the rank aggregation (Pihur 2010). We compare different ranking methods in accessing their stability. For this we use stability selection (Meinshausen 2009). We also investigate their differences using distance based approaches (Fagin 2003). The results are validated using primary and secondary RNAi screening data.

Results

Single rankings can differ a lot in their individual results. Especially when only a low number of replicates is available in the data. We show that nevertheless some methods prove to be more stable than others and not all are suitable for RNAi data. Since for a secondary validation screen only a set of the most significant genes is important rather than the absolute rank values of the genes, we retrieve stable subsets of the data containing those genes irrespective of small changes in the ranks.

Discussion

Gene rankings are needed in various biological fields and there is not yet a single perfect method available to find a best ranking. We show which methods perform better when dealing with RNAi knock down screens and give an estimation in how trustworthy the resulting ranked gene lists really are.

Presenting Author

Juliane R. E. Siebourg (juliane.siebourg@bsse.ethz.ch)
ETH Zurich

Author Affiliations

(1) D-BSSE, ETH Zurich

G-66. A network centered approach for genome wide data integration applied to Alzheimers disease prediction

van den Akker E (1,2,), Heijmans B (1), Kok J, (1,3), Slagboom P (1), Reinders M (2)*

Medical research has shifted its focus from mono-factorial diseases to the more challenging field of multi-factorial diseases. Often, multiple types of genome wide data sources are created in search for candidates underlying the complex etiology of disease phenotypes. Furthermore, an increasing amount of knowledge on molecular interactions could be employed to help interpret the results of these genome-wide screens. All of this requires novel ways to analyze the available data that allows a joint interpretation of multiple data sources, while taking known molecular interactions into account.

Materials and Methods

We revisited SNP and expression data generated in a case-control setting in the Late Onset Alzheimers Disease (LOAD) cohort. In the methodology presented, we aim to identify subsets of the gene-gene interaction network predictive of Alzheimers disease using both data sources. We seed a subnetwork once at every node of the gene-gene interaction network and use a wrapper approach to extent the network to network neighbors until no improvement in classification performance is observed. Resulting subnetworks are validated on an independent set by employing a double loop cross-validation.

Results

When analyzed separately, the network centered analysis of gene expression data as well as the SNP data already yields biologically relevant signatures for Late Onset Alzheimers Disease. Furthermore, we present a comparison of predictive performance and biological interpretability of several ways on how to combine the different data sources: an early integration scheme that concatenates features, an intermediate scheme that combines distances in the separate data domains, and a late integration scheme that overlays the individually found networks.

Discussion

The work presented here illustrates the benefits of a network centered approach. By employing the data for identification of subnetworks in the gene-gene interaction network, the biological interpretation of genome wide data sources are not confined by arbitrarily set boundaries as seen in the classical grouped data analysis approaches. The presented framework is applicable to any type of genome-wide data source and depending on the biological question asked, offers a great flexibility in different settings for genome wide data integration.

Presenting Author

Erik B. van den Akker (erikvandenakker@gmail.com)

The Delft BioInformatics Lab, Delft University of Technology, Delft, The Netherlands

Author Affiliations

(1) Molecular Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands (2) The Delft Bioinformatics Lab, Delft University of Technology, Delft, The Netherlands (3) Algorithms, Leiden Institute of Advanced Computer Science, University Leiden, Leiden, The Netherlands

Acknowledgements

This project was funded by the Medical Delta.

G-67. How to efficiently hunt the disease causing gene

Tranchevent L-C (1,), Aerts S (2,4), Van Loo P (1,3,4), Hassan BA (2,4), Moreau Y (1)*

The development of advanced high-throughput technologies such as array CGH has led to many breakthrough in human genetics by deciphering disease-loci associations. One typical feature of array CGH is that the loci found to be altered usually contain dozens of genes of which only a few are really causative for the disease / phenotype of interest. The validation of all the candidate genes is expensive and time consuming, therefore there is a need for a method that identifies the most promising candidate genes to assay first.

Materials and Methods

ENDEAVOUR, a web resource for the prioritization of genes, indicates which genes are the most promising candidates [1]. It relies on evidence that suggest that functionally related genes often cause similar phenotypes [2]. ENDEAVOUR uses various data sources that describe the function of the genes, their expression profiles, their genetic interactors, and their regulation processes and integrate them to make one global prediction.

Results

Our approach has been successfully validated by mean of an extensive cross-validation on OMIM and MetaCore genetic disorders, and on GO, MetaCore, and Ingenuity pathways. Furthermore, we have experimentally validated it with a detailed study on DiGeorge syndrome [1], on neural development in *Drosophila* [3], and on congenital heart defects [4].

Discussion

In conclusion, ENDEAVOUR is a gene prioritization resource that was extensively validated and is publicly available at <http://www.esat.kuleuven.be/endeavour>. 1- Aerts S et al. Gene prioritization through genomic data fusion. *Nat. Biotechnol.* 2006. 2- Jimenez-Sanchez G et al. Human disease genes. *Nature.* 2001 3- Aerts et al. Integrating Computational Biology and Forward Genetics in *Drosophila*. *Plos Genetics.* 2009. 4- Thienpont et al. Haploinsufficiency of TAB2 causes congenital heart defects in humans. *Am J Hum Genet.* 2010.

URL

<http://www.esat.kuleuven.be/endeavour>

Presenting Author

Léon-Charles Tranchevent (Leon-Charles.Tranchevent@esat.kuleuven.be)

Department of Electrical Engineering ESAT-SCD, Katholieke Universiteit Leuven, Leuven, Belgium

Author Affiliations

1 Department of Electrical Engineering ESAT-SCD, Katholieke Universiteit Leuven, Leuven, Belgium. 2 Laboratory of Neurogenetics, Department of Molecular and Developmental Genetics, VIB, Leuven, Belgium. 3 Human Genome Laboratory, Department of Molecular and Developmental Genetics, VIB, Leuven, Belgium. 4 Department of Human Genetics, Katholieke Universiteit Leuven School of Medicine, Leuven, Belgium.

Acknowledgements

This work was supported by the Research Council KUL [GOA AMBioRICS, CoE EF/05/007 SymBioSys, PROMETA]; the Flemish Government [G.0241.04, G.0499.04, G.0232.05, G.0318.05, G.0553.06, G.0302.07, ICCoS, ANMMM, MLDM, G.0733.09, G.082409, GBOU-McKnow-E, GBOU-ANA, TAD-BioScope-IT, Silicos, SBO-BioFrame, SBO-MoKa, TBM-Endometriosis, TBM-IOTA3, O&O-Dsquare]; the Belgian Federal Science Policy Office [IUAP P6/25]; and the European Research Network on System Identification (ERNSI) [FP6-NoE, FP6-IP, FP6-MC-EST, FP6-STREP, FP7-HEALTH].

G-68. Study of the ultrasensitivity of the BCL-2 apoptotic switch

Tokár T (), Uličný J*

Bcl-2 apoptotic switch is important control point of apoptosis since regulating permeabilization of outer mitochondrial membrane. Bcl-2 apoptotic switch is molecular mechanisms converting continuous incoming signals to two mutually distinct outputs, ensuring transitions between different cellular states. Necessary requirement to generate such behaviour is ultrasensitive reaction mechanism in which the output response is more sensitive to change in stimulus than hyperbolic response. There are two competing hypotheses regarding functioning of Bcl-2 apoptotic switch, the indirect and direct model.

Materials and Methods

Based on hypotheses of direct and indirect activation, we constructed corresponding mathematical models. We introduced third, hybrid model involving controversial interactions from both, models. In this work we utilized ultrasensitivity of the Bcl-2 apoptotic switch models as a criterion, against which we judge plausibility of its models. For each model we analysed its robustness with respect to its ability to preserve ultrasensitivity against variations of its parameters around estimated reference values. We have analysed influence of the most debatable reactions on the behaviour of our models

Results

Results of robustness analysis show that it is very improbable that the indirect model could act as biological switch mechanism. Moreover, we have found that, while reaction specific for the direct model has particularly beneficial effect on ultrasensitivity of the Bcl-2 apoptotic switch, reactions proposed by the indirect model seems to reduce sensitivity in very strong manner.

Discussion

The direct model as we proposed in this work, can act as biological switch, under wide range of parameter settings. We found alternative variants as inappropriate, since unable to resemble requested behaviour.

Presenting Author

Tomas Tokar (tomastokar@gmail.com)

PhD. student

Author Affiliations

Department of Biophysics, Faculty of Science, University of P. J. Safarik, Kosice, Slovakia

Acknowledgements

This work was financially supported by grants VEGA-1/4019/07; APVV-0449-07 and VVGS PF 12/2009/F. This work could not be done without scholarships granted to Tomas Tokar from National Scholarship Programme of the Slovak Republic and "Hlavicka" scholarship from Slovensky plynarensky priemysel a.s., we wish to thank for this support. Tomas Tokar appreciate many valuable discussions with colleagues from Institute for Systems Theory and Automatic Control, University of Stuttgart, Germany.

G-69. Improved genome-wide protein-protein interaction prediction and analysis of biological process coordination in Escherichia coli

Muley VY (), Ranjan A*

Protein-protein interactions (PPI) play crucial role in executing physiological function and to maintain the structural integrity of cell in diverse environments. A number of computational methods have been proposed for PPI prediction using genomic context. In this study, we check the possibility of integration of gene expression, correlated mutation and genomic context based method to enhance the reliability of PPI prediction. Furthermore, We use panel of seven machine learning classifiers (MLCs) and compare their performance.

Materials and Methods

PPI prediction methods used for analysis are 1. Phylogenetic profile 2. Minimum Distance Method 3. Gene Expression Similarity 4. Improved Mirrortree Method 5. Gene Cluster Method 6. Gene Order Conservation We compute the association score using above methods and feed them to seven MLCs are as follows 1. Random Forest 2. Decision Tree 3. Naive Bayes 4. Bayesian Network 5. Logistic Regression 6. Neural Network 7. Support Vector Machine The gold standard positive dataset is derived from EcoCyc dataset whereas negative dataset generated by indirect evidence

Results

In this study, we build a genome-wide interaction map for E. coli which has two main aspects that could be of potential use to researchers. First, this work provides a interaction map that is reliable than previous reports with 10-fold cross-validation balanced accuracy of 93%. Second, this interaction map makes certain potentially interesting biological predictions: an antagonistic link between purine catabolism and pyrimidine biosynthesis and a coupling between cell division, lipopolysaccharide biosynthesis, replication initiation, ATP synthesis, and colanic acid biosynthesis.

Discussion

In the present study, we have introduced a consensus approach for predicting genome-wide functional interactions by a combination of seven MLCs. Functional analysis of consensus PPI network showed that gene expression pattern of the predicted non-interacting protein pairs is not similar even though they belong to the same biological pathway. Topological properties of networks predicted by various MLCs are not similar. Analysis of various biological pathways in context of gene expression shows novel connection between various biological pathways.

URL

<http://www.cdfd.org.in/ecofunppi/>

Presenting Author

Vijaykumar Y Muley (vijay.muley@gmail.com)

Center for DNA Fingerprinting and Diagnostics

Author Affiliations

Computational and Functional Genomics Group, Center for DNA Fingerprinting and Diagnostics, Bldg. 7, Gruhakalpa, 5-4-399 / B, Nampally, Hyderabad - 500 001

Acknowledgements

CSIR-UGC Fellowship CDFD

G-70. A generic gene regulatory network reconstruction method: application to *Lactococcus lactis* MG1363

Theunissen D (1,2,3,), Brouwer R (4,5), Kuipers O (2,4,5), Hugenholtz J (2,3), Siezen R (1,2,3,5,6), Van Hijum S (1,2,3,5,6)*

Gene regulatory networks (GRN) are used by prokaryotes to adapt to changing environment by regulating gene expression levels. The basic building blocks of GRN are regulons, which consist of a transcriptional regulator and target genes. In prokaryotes these target genes are organized into transcriptional units (TUs). We are developing a generic method to predict gene regulatory networks for any prokaryote. As a model system for predicting the GRN the dairy bacterium *Lactococcus lactis* MG1363 was chosen.

Materials and Methods

Our method consists of 4 steps: 1) training a classification model on known regulatory interactions of *E. coli* and *B. subtilis*. Any interaction occurring between genes part of the same regulon is used. Features used to predict interactions between two genes include operons, correlated expression, gene ontologies, metabolic pathways; 2) classifying interactions between any two genes; 3) identifying quasi-cliques, or dense sub-graphs of putative co-regulated genes; 4) identifying putative transcription regulators that govern the expression of the genes part of a quasi-clique.

Results

The result of applying the four steps onto gene-pairs of *L. lactis* MG1363 is a putative gene-regulatory network. Preliminary results indicate favorable performance of our method: in general, 60% of the known *L. lactis* regulatory interactions are correctly identified. When considering the smaller regulons, with member sizes of 6 to 14 members, accuracy increases to over 80%.

Discussion

In future, we will further develop the method, one way to do this is in implementing new predictive features. These new features can help improve the accuracy of our method. Finally we will implement our method into user-friendly software to accurately determine the GRN for any prokaryote.

Presenting Author

Daniel H.J. Theunissen (dtheunis@cmbi.ru.nl)
NIZO food research B.V.

Author Affiliations

1 Radboud University Medical Centre, Center for Molecular and Biomolecular Informatics, Nijmegen Center for Molecular Life Sciences, P.O. Box 9101, 6500 HB Nijmegen, the Netherlands 2 Kluyver Centre for Genomics of Industrial Fermentation, P.O. Box 5057, 2600 GA Delft, the Netherlands 3 NIZO food research, P.O. Box 20, 6710 BA Ede, the Netherlands 4 Molecular Genetics group, GBB, RUG, Kerklaan 30, Haren. 5 Netherlands Bioinformatics Centre, 260 NBIC, P.O. Box 9101, 6500 HB Nijmegen, the Netherlands 6 TI Food and Nutrition, P.O. Box 557, 6700 AN Wageningen, the Netherlands

G-71. Dissecting the specificity of E2-E3 interaction in the ubiquitination pathway by a molecular docking approach

Melquiond ASJ (1,), Rodrigues J (1), Bonvin AMJJ (1)*

Kinetic and thermodynamical properties of complexes account for basic regulatory properties. The affinity with which a group of proteins interacts allows for a certain degree of modulation of these interactions. Enzyme-inhibitor complexes usually interact strongly, while proteins involved in regulation or signaling, such as G-coupled receptors, often interact more dynamically. Starting from proteomics data on ~8000 E2-E3 interactions, we used molecular modeling tools to dissect the forces which drive the specificity of their interactions.

Materials and Methods

Up to now, only 7 structures of E2-E3 complexes are solved either by Xray crystallography or NMR spectroscopy. Starting from a set of Yeast Two-Hybrid assays between 37 E2s and 227 E3s, we used our molecular docking software HADDOCK to model more than 500 potential complexes. Structural alignment of E2 and E3 models on the bound structures of the existent complexes, followed by refinement in water and scoring with HADDOCK's force field and scoring function has been shown to discriminate between biologically relevant complexes and artefacts.

Results

The result reveal a high correlation between the HADDOCK score of the predicted complexes and the proteomics data. HADDOCK is equally fit to predict interactions starring any E3, whatever differences they may have. Specific complexes show the importance of electrostatic interactions between charged residues on the loop 1 of E2 and the ubiquitin ligase E3 protein

Discussion

Some new findings on the role of potential dynamical intermolecular salt-bridges were considered interesting enough to encourage a detailed inspection. They led to unexpected findings and proved that this method can provide priceless insights on the molecular mechanisms that drive the interaction.

Presenting Author

Adrien S.J. Melquiond (a.s.j.melquiond@uu.nl)
Utrecht University

Author Affiliations

(1) NMR Spectroscopy research group, Bijvoet Center for Biomolecular Research, Utrecht University

Acknowledgements

NWO VICI to A.M.J.J Bonvin (grant no. 700.96.442)

AUTHOR INDEX

Acuner Ozbabacan SE 18	Chen C-Y 45	Fröhlich H 10
Aerts S..... 72	Chen S-H..... 45	Fundel-Clemens K..... 56
Agarwal S 19	Chin C-H..... 45	Furlong LI 37, 38
Agostini S 50	Constanzo M 5	Galvão V 21
Ali W..... 55	Corà D..... 58	Gerstein M..... 4
Anchang B 44	Croes D 62	Gerstein MB 4
Andrade RFS..... 21	Cromer D..... 46	Geurts P 6
Andrews BJ 5	Cukuroglu E..... 35	Gfeller D..... 11
Arakawa K 41	Daelemans W 53	Ghoorah AW 29
Audenaert P..... 9	Dang T-H 47	Gill S 34
Bader GD 5, 11	Dar N..... 11	Ginalski K 16
Baitaluk M 48	Davenport MP 46	Glaab E..... 65
Baudot A 65	De Clercq I..... 12	Goethals B..... 53
Bauer-Mehren A 37, 38	De Knijf J 53	Grossman Z..... 68
Baum P..... 56	de la Fuente A 43	Gursoy A 8, 18, 35, 40
Becker S..... 24	De Leo V 43	Han S..... 5
Beerenwinkel N 34, 70	De Maeyer M..... 68	Hassan BA..... 72
Beheydt G 68	De Maria E..... 22	Heijmans B..... 71
Beißbarth T 10	De Moor B 28, 67	Helms V 51
Bellay J 5	De Rijk P 53	Henjes F 10
Beltrame L 59	Deane C..... 55	Heringa J 26
Bender C 10	Deane CM 55	Ho C-W 45
Bhardwaj N..... 4	Del-Favero J..... 53	Hollunder J..... 64
Bielza C..... 68	Delgado-Eckert E..... 34	Hsiung C..... 45
Birzele F..... 56	Demeester P 9	Hugenholtz J..... 75
Bonvin AMJJ 76	Denamur E 52	Hulsman M..... 27
Boone C 5	Devignes M-D..... 29	Iacucci E..... 28
Bosia C..... 58	Dhananjaneyulu V..... 54	Ikegami K..... 41
Bourguignon P-Y 52	Di Silvestre D..... 50	Ittrich C 56
Bouvet O 52	Diessl N..... 32	Jaeger S 49
Boyen P..... 25	Dobson P..... 33	Jansen M..... 68
Brambilla F 50	Donato M 59	Jardin C 15
Brenninkmeijer C..... 33	Draghici S 59	Jones N..... 19
Bron PA 36	Dubinina Y..... 48	Kaiser R..... 68
Brors B..... 56	Dupont P 62	Kar G..... 8
Brouwer R..... 75	Durot M..... 52	Keskin O..... 8, 18, 35, 40
Brunetti P 50	Eils R..... 32, 56	Kido N..... 41
Bundsclus M 38	El Baroudi M..... 58	Kim PM..... 4, 5, 11
Buzug T M 24	Engin B 40	Kleerebezem M 36
Calura E 59	Engram J 46	Klijn C..... 66
Camacho RJ 68	Ernst A 11	Ko M-T..... 45
Canevet C..... 33	Esmailbeiki R..... 13	Kok J 71
Canisius S 66	Fages F..... 22	Konig R 32
Caselle M..... 58	Fauré A..... 57	Kono N..... 41
Cavalieri D..... 59	Faust K..... 62	Korf U 10
Cavallini C 50	Feenstra KA 26	Kozhenkov S 48
Chan ML..... 46	Fisher P 33	Kranz A-J 32
Chaouiya C..... 57	Foekens J..... 66	Krasnogor N 65
Chen C 45	Froehlich H 31	Kuipers O 75

Kumar G	54	Pool R.....	26	Tebaldi T	63
Lämmle B	56	Popovici R.....	59	Theunissen D.....	75
Larrañaga P	68	Praveen P.....	31	Thierry-Mieg N	60
Laukens K.....	47	Prosperi M.....	68	Tokár T.....	73
Łązniewski M.....	16	Quast K	56	Toma A.....	24
Le Fèvre F.....	52	Quattrone A.....	63	Tomita M.....	41
Leonardi E.....	23	Ranjan A	74	Torti C	68
Leser U.....	49	Rautschka M	37, 38	Tosatto S.....	23
Li GD	68, 69	Rawlings C.....	33	Tranchevent L-C	72
Liekens A.....	53	Reinders M.....	27, 71	Tresch A	31
Lin C-Y	45	Reinders MJT.....	27	Uličný J	73
Lionetti V	50	Reinelt G	32	Valencia A.....	65
London N.....	42	Ritchie DW	29	Vallenet D	52
Madan Babu M	6	Rivero Guedez D.....	59	Van Breusegem F.....	12
Mang A	24	Rizk A	22	Van de Peer Y	9, 12, 64
Marco M	36	Rizzetto L.....	59	Van de Plas R.....	67
Martens J.....	66	Rodrigues J.....	76	van den Akker E.....	71
Mauri PL.....	50	Romualdi C	59	van Dijk ADJ.....	25
Mayer MA.....	38	Rychlewski L	16	Van Dyck D.....	25
Médigue C.....	52	Sabarly V.....	52	van Ham RCHJ	25
Melquiond ASJ	76	Sadeh M	44	van Helden J	62
Mendoza M R	17	Sagar PVN.....	54	Van Hijum S.....	75
Mennerich D	56	Sagulenko V.....	32	Van Laethem K	68
Merdes G.....	34	Sambourg L.....	60	Van Loo P	72
Metzger J.....	51	Sanguinetti G.....	63	Van Parys T.....	9, 12
Michailidis G	61	Sanz F.....	37, 38	Vandamme AM	68
Michaut M.....	5	Schachter V	52	Vanhee P	11
Michoel T.....	9	Schmid R.....	56	Verbeeck N.....	67
Miranda JGV.....	21	Schneider R.....	30	Vermeirssen V	12
Moreau Y	7, 28, 72	Schramm G	32	Verschoren A	47
Mornico D.....	52	Schroeter A	20	Verschueren E	11
Muley VY	74	Schrynemackers M.....	6	Via A	58
Murgia A.....	23	Schueler-Furman N.....	42	Vieira G.....	52
Myers CL.....	5	Schuster S.....	20	Villar G.....	19
Nebel J-C	13	Schütz T A	24	Viswanathan GA	54
Neven F.....	25	Secrier M.....	30	Vlaic S	20
Niranjan M.....	63	Sedova M	48	von Grotthuss M.....	16
Nitsch D	7	Serrano L.....	11	Vreede B.....	57
Nussinov R.....	8, 18	Shojaie A.....	61	Waelkens E.....	67
Ogawa R	41	Sidhu SS.....	11	Walter P.....	51
Ojeda F.....	28	Siebourg J.....	70	Wehenkel L	6
Orsini M.....	43	Siezen R	75	Weith A	56
Ortiz AM.....	46	Silvestri G	46	Wels M	36
Osella M.....	58	Slagboom P	71	Werhli A V.....	14, 17
Oshita K	41	Smail-Tabbone M	29	Werner S.....	20
Oswald M.....	32	Smid M.....	66	Wessels L	66
Paiardini M.....	46	Soliman S	22	Westermann F	32
Pavlopoulos GA	30	Spang R.....	44	Wiemann S	10
Petravic J.....	46	Stevens R.....	33	Wiersma A	36
Pickavet M	9	Sticht H	15	Wiesberg S	32
Pinna A	43	Sucena É.....	57	Wilhelm T	64
Plewczynski D	16	Tamaki S	41	Zazzi M	68
Ponomarenko J.....	48	Taubert J.....	33		