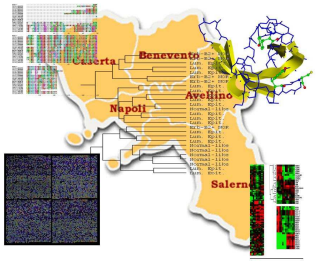

BBCC2011

Sixth Annual Meeting

Bioinformatica e Biologia Computazionale in Campania

November 4th, 2011
Avellino, Italy
Istituto di Scienze dell'Alimentazione
Consiglio Nazionale delle Ricerche

http://bioinformatica.isa.cnr.it/BBCC/BBCC2011/

BBCC2011 – Scientific Committee

Scientific Committee

Angelo Facchiano (Chair) - Istituto di Scienze dell'Alimentazione, CNR, Avellino, Italy

Claudia Angelini - Istituto per le Applicazioni del Calcolo "M. Picone", CNR, Napoli, Italy

Graham Ball - School of Science and Technology, Nottingham Trent University, UK

Maria Luisa Chiusano - DISSPAPA Dipartimento di Scienze del Suolo della Pianta dell'Ambiente e delle Produzioni Animali - Università degli Studi di Napoli “Federico II”, Italy

Antonio d'Acierno - Istituto di Scienze dell'Alimentazione, CNR, Avellino, Italy

Italia De Feis - Istituto per le Applicazioni del Calcolo "M. Picone", CNR, Italy

Pedro Fernandes - Instituto Gulbenkian de Ciencia, Oeiras, Portugal

Anna Marabotti - Istituto di Tecnologie Biomediche, CNR, Segrate (MI), Italy

Andrew C.R. Martin - University College London, UK

Mark Ragan - Institute for Molecular Bioscience, University of Queensland, Australia

Steffen B. Petersen - Department of Health Science and Technology, University of Aalborg, Denmark and the Biophotonics Group at International Iberian Nanotechnology laboratory

Thomas Simonson - Ecole Polytechnique and Centre National de la Recherche Scientifique, Palaiseau, France.

BBCC2011 is supported by the Bioinformatics Italian Society – www.bioinformatics.it

BBCC2011

Sixth Annual Meeting

Bioinformatica e Biologia Computazionale in Campania

November 4th, 2011

Avellino, Italy

Istituto di Scienze dell'Alimentazione

Consiglio Nazionale delle Ricerche

Program

9.30 - 10.00	**Registration**
10.00 - 10.30	**Welcome and Introduction** *Angelo Facchiano* ISA-CNR and BBCC Chair
10.30 – 11.00	**Gene Wiki: Crowdsourcing Knowledge Extraction From the Biomedical Literature** *Salvatore Loguercio* Biotechnology Center, Technische Universität Dresden, Germany
11.00 – 11.20	**Understanding, Communicating and Automatically Managing Bioinformatic Protocols** *Francesco Napolitano* Dipartimento di Informatica, Università di Salerno
11.20 – 11.40	**A correlation graph clustering algorithm for analyzing gene expression data** *Crescenzio Gallo* Università di Foggia
11.40 – 12.00	**Computational approaches for genome-wide mRNA and miRNA expression profiling in human breast cancer cell lines expressing (ERβ+) or lacking (ERβ-) estrogen receptor beta by microarray hybridization and massively parallel sequencing (miRNA-Seq)** *Maria Rosaria De Filippo* Lab. di Medicina Molecolare e Genetica, Fac. Medicina, Università di Salerno
12.00 – 12.20	**Functional effect of the intrinsic disorder in the family of the human chemokine membrane receptors** *Giovanni Colonna* Seconda Università di Napoli
12.20 – 13.00	**Discussion with all participants:** - **scientific topics from oral presentations** - **bioinformatics and computational biology perspectives**
13.00 – 14.20	**Lunch and Poster Session**

14.20 – 14.40 **Integer programming models for RNA tertiary structure reconstruction**
Maria De Cola
IASI-CNR, Roma

14.40 – 15.00 **Simulating gene-gene and gene-environment interactions in complex diseases: Gene-Environment iNteraction Simulator 2**
Giovanni Scala
Università di Napoli "Federico II"

15.00 – 15.20 **Molecular investigations of the interaction between the ligand-binding domain of the Human aryl-hydrocarbon receptor and its endogenous ligands and xenobiotics from dietary sources.**
Maria Salzano
ISA-CNR e Seconda Università di Napoli

15.20 – 15.40 **COCOMAPS and CONS-COCOMAPS: novel web tools for the analysis of crystallographic complexes and of multiple docking solutions**
Anna Vangone
Università di Salerno

15.40 – 16.00 **Solving Biclustering with a GRASP-like Metaheuristic: Two Case-Studies on Gene Expression Analysis**
Francesco Musacchia
Università di Napoli "Federico II"

16.00 – 16.20 **Gene profiling of human PBMCs ex vivo vaccinated with anti idiotypic vaccine for HCV-related lymphoproliferative disorders**
Annacarmen Petrizzo
Fondazione Pascale, Napoli

16.20 – 16.40 **Structural analysis of Antarctic teleost Toll-like receptor 2**
Maria Rosaria Coscia
IBP-CNR, Napoli

16.40 – 17.20 **Discussion with all participants:**

- **scientific topics from oral presentations and posters**
- **BBCC network**

17.20 **Meeting Closure**

Abstracts of oral presentations and posters

Gene Wiki: Crowdsourcing Knowledge Extraction From the Biomedical Literature

Salvatore Loguercio (1), Benjamin M. Good (2), Andrew I. Su (2)

(1) Biotechnology Center, Technische Universität Dresden, Germany
(2) Scripps Research Institute, La Jolla, California (US)

The goal of the Gene Wiki project is to create a continuously-updated, collaboratively-written, and community-reviewed review article for every gene in the human genome. The Gene Wiki currently takes the form of ~10,000 Wikipedia articles, each describing one human gene. It receives over four million page views and one thousand edits per month. Having attained a sizeable and growing volume of content, one major area of emphasis is improving the computability of this resource. Herein, we will describe a variety of strategies for mining, verifying, and sharing new gene annotations from the community-generated content of the Gene Wiki.

REFERENCES

Huss, J.W., 3rd, et al. (2008) A gene wiki for community annotation of gene function, PLoS Biol, 6, e175.
Huss, J.W., 3rd, et al. (2010) The Gene Wiki: community intelligence applied to human gene annotation, Nucleic Acids Res, 38, D633-639.
Good BM and Su AI (2011) Mining Gene Ontology Annotations From Hyperlinks in the Gene Wiki. Proceedings of the Translational Bioinformatics Conference (http://proceedings.amia.org/T2011)
Good BM, Loguercio S, Su AI (2011) Linking genes to diseases with a SNPedia-Gene Wiki mashup, Journal of Biomedical Semantics (accepted).

Understanding, Communicating and Automatically Managing Bioinformatic Protocols

Francesco Napolitano, Roberto Tagliaferri

D.I., University of Salerno, Fisciano (SA)

The data analysis phase in a bioinformatic study is usually made of ad hoc software tools developed in high level language with a prototypical approach. While the source code of such tools provides by itself a precious means to replicate the analysis, many details are often not included in it. In this presentation we try to formalize the concept of protocol in bioinformatic data analysis and introduce the Leaf System, including a graph design language to define the main steps of the data analysis and a Python engine to automatically and efficiently apply it with support for the management of data resources.

A correlation graph clustering algorithm for analyzing gene expression data

Crescenzio Gallo, Michelangelo De Bonis

Universita' di Foggia - Dip. Scienze Economiche/Dip. Scienze Biomediche

The possible applications of modeling and simulation in the field of bioinformatics are largely adopted in exploring genetic variability. Experimental results carried out with DNA microarrays allow researchers to measure expression levels for thousands of genes simultaneously, across different conditions and over time. A key step in the analysis of gene expression data is the detection of groups of genes that manifest similar expression patterns. In this paper we describe a mixed clustering algorithm for analyzing gene expression data, comparing its results with the classification deriving by the application of unsupervised neural networks. The algorithm is based on correlation graphs and has been implemented and tested on real gene expression data, demonstrating its flexibility and effectiveness.

REFERENCES

1 Alizadeh, A., Eisen, M., Davis, R., Lossos, I., Rosenwald, A., Boldrick, J., Sabet, H., Tran, T., Yu, X., & Powell, J. (2000). Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling. Nature, No. 403 (pp. 503–522).

2 Collet, P., & Eckmann, J. (2002). The number of large graphs with a positive density of triangles. J. Stat Phys, No. 109 (pp. 923–943).

3 Eckmann, J., & Moses, E. (2002). Curvature of co-links uncovers hidden thematic layers in the world wide web. Proc NatL Acad Sci USA, No. 99 (pp. 5825–5829).

4 Eisen, M., Spellman, P., Brown, P., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc NatL Acad Sci USA. No. 95 (pp. 14863–14828).

5 Furey, T., Cristianini, N., Duffy, N., Bednarski, D., Schummer, M., & Haussier, D. (2000). Support vector machine classification and validation of cancer tissue samples using microarray data. Bioinformatics, No. 16 (pp. 906–914).

6 Herrero, J., Valencia, A., & Dopazo, J. (2001). A hiererchical unsupervised growing neural network for clustering gene expression patterns. Bioinformatics, No. 17 (pp. 126–136).

7 Yanai, I., & De Lisi ,C. (2002). The society of genes: networks of functional links between genes from comparative genomics. Genome Biology, No. 3.

Computational approaches for genome-wide mRNA and miRNA expression profiling in human breast cancer cell lines expressing (ERβ+) or lacking (ERβ-) estrogen receptor beta by microarray hybridization and massively parallel sequencing (miRNA-Seq)

Maria Rosaria De Filippo (1), Giorgio Giurato (1), Roberta Tarallo (1), Maria Ravo (1), Francesca Rizzo (1), Concita Cantarella (1), Giovanni Nassa (1),, Ernesto Nola (2) and Alessandro Weisz (1,2)

(1) Laboratory of Molecular Medicine and Genomics, Faculty Medicine and Surgery of the University of Salerno, Baronissi (SA), Italy
(2) Department of General Pathology of the Second University of Naples, Napoli, Italy

microRNAs (miRNAs) are evolutionary conserved small non coding RNA that negatively regulate gene expression. Recent studies have demonstrated that mutations or aberrant expression of miRNAs are associated with cancer, suggesting that genes encoding these RNAs may act as oncogenes or tumor suppressors. Estrogen receptor alpha (ERα) and beta (ERβ) are transcriptional factors (TFs) that mediate estrogen signaling and define the hormone responsive phenotype of breast cancer. The two receptors can be found co-expressed, and play specific, often opposite roles, with ERβ being able to modulate the effect of ERα on gene transcription and cell proliferation. ChIP-Seq (Chromatin immunoprecipitation followed by sequencing) analysis of breast cancer cell lines (MCF7) showed that ERα and ERβ bind in close proximity of several miRNA genes, suggesting a direct involvement of these nuclear receptors in biogenesis of these small RNAs[1,2].Starting from these observations, I investigated miRNA expression patterns by miRNA-Seq (direct sequencing of small non coding RNA) and microarray hybridization in two ERα positive breast cancer cell lines: one lacking (ERβ-) and one expressing (ERβ+) estrogen receptor β. At first, I analyzed data obtained from microarray experiments (Agilent Human microRNA microarray). After normalization, student t-test was performed to identify differentially expressed miRNA between the two cell lines[3]. Subsequently, I focused on analyzing miRNA-seq data from the same cell lines. To this end, reads obtained from sequencing were analyzed with a specific bioinformatic tool, miRAnalyzer. The R Bioconductor's package, DeSeq, was then used to perform differential expression analysis for sequence count data between ERβ- and ERβ+ cells. Comparison of data obtained from microarray analyses and miRNA-Seq was carried out to evaluate the reliability, sensitivity and reproducibility of the two analytical tools. Finally, differentially expressed miRNAs was used to search for their putative target mRNAs through the use of dedicated bioinformatics tools. The results of this analysis were exploited to reveal biological functions and molecular processes in which these miRNA targets are involved and are therefore controlled at a post-transcriptional levels by ERβ *via* specific miRNAs.

REFERENCES

[1] Cicatiello L, Mutarelli M, Grober OM, Paris O, Ferraro L, Ravo M, Tarallo R, Luo S, Schroth GP, Seifert M, Zinser C, Chiusano ML, Traini A, De Bortoli M, Weisz A. Estrogen receptor alpha controls a gene network in luminal-like breast cancer cells comprising multiple transcription factors and microRNAs. *Am J Pathol. 2010 May;176(5):2113-30.*
[2] Grober OM, Mutarelli M, Giurato G, Ravo M, Cicatiello L, De Filippo MR, Ferraro L, Nassa G, Papa MF, Paris O, Tarallo R, Luo S, Schroth GP, Benes V, Weisz A. Global analysis of estrogen receptor beta binding to breast cancer cell genome reveals an extensive interplay with estrogen receptor alpha for target gene regulation. *BMC Genomics. 2011 Jan 14;12:36.*
[3] Paris O, Ferraro L, Grober OM, Ravo M, De Filippo MR, Giurato G, Nassa G, Tarallo R, Cantarella C, Rizzo F, Di Benedetto A, Mottolese M, Benes V, Ambrosino C, Nola E, Weisz A. Direct Regulation of microRNA Biogenesis and Expression by Estrogen Receptor Beta in Hormone-responsive Breast Cancer. *Oncogene 2011. In press.*

Functional effect of the intrinsic disorder in the family of the human chemokine membrane receptors.

Marco Miele, Raffaele Raucci, Susan Costantini and Giovanni Colonna

CRISCEB e Dipt Biochimica e Biofisica - SUN
CROM - Istituto Pascale (AV)

The structural characteristics of 19 human chemokines receptors were examined for the presence of intrinsically disordered regions. The results show that the N and C terminal arms possess structural characteristics such that they can be considered intrinsically disordered with a high structural flexibility and the presence of numerous charged patches. In addition, several putative phosphorylation sites have been found in these regions suggesting their possible involvement in regulating the binding of cytokines to different ligands thus explaining the phenomenon of pleiotropy. To have a physical model of the structural behavior of these receptors we have modeled and subjected to molecular dynamics the receptor CXCR3, quite similar to the receptor CXCR4 whose X-ray coordinates are known but with its N-terminal missing. The structural features of the CXCR3 model were then compared to those of CXCR4. A feature common to the most part of the humane chemokines, i.e., the presence of a Pro adjacent to the Cys involved in a bond with a Cys of the TM helix No 7, was carefully analyzed. The Pro residue, because of steric constraints dictated by the close rigid disulfide bridge, points out of the membrane thereby directing to the outside also the remaining very flexible polypeptide chain of the N-terminus. This feature is present in both the CXCR3 and CXCR4 and allowed us to model the structure of this segment drawing functional conclusion.

REFERENCES

1) Modelling of the membrane receptor CXCR3 and its complexes with CXCL9, CXCL10 and CXCL11 chemokines: Putative target for new drug design. Trotta T, Costantini S, Colonna G. Mol Immunol. 2009 Sep 30. [Epub ahead of print] PMID: 19800124.
2) S. Costantini, M. Miele, G. Colonna: "Intrinsically Unordered Proteins: structural properties, prediction and relevance", in Protein Folding, Chapt.6; (2010) - Editors: Eric C. Walters Series: Protein Biochemistry, Synthesis, Structure and Cellular Functions, Nova Science Publishers, Inc. Hauppauge NY – USA ISBN: 978-1-61761-259-6.
3) S. Costantini, M. Costantini, G. Colonna: "Frequencies of specific peptides in intrinsic disordered protein domains" Protein and Peptide Letters (2010) 17(11), 1398-402.
4) S. Costantini, G. Castello, G. Colonna: "Human Cytokinome: a new challenge for systems biology". Bioinformation (2010) 5(4), 166-167.
5) D. Evangelista, G. Colonna, M. Miele, F. Cutugno, G. Castello, S. Desantis and S. Costantini: "CDMS (Clinical Data Mining Software): a cytokinome data mining system for a predictive medicine of chronic inflammatory diseases" PEDS (2010) 123(12): 899-902.

Integer programming models for RNA tertiary structure reconstruction

Maria De Cola, Giovanni Felici

IASI-CNR, Roma

The Nuclear Magnetic Resonance Spectroscopy is an important technique which allows of study the structure of biomolecules. The correlation signals between the nuclei of a RNA molecule from NMR experiments can be used to reconstruct its three-dimensional shape. Very recently, a new approach that represents such signals as edges in a graph model G = (V;E) colored in according to the type of interaction between the peaks in 3D NMR spectrum has been proposed [2]-[3]. An alternating path between the vertices of G is the reconstruction of a magnetization transfer pathway between the cross-peaks of the spectral graph. The problem is to find a longest path in a graph where edges can have one out of c colors, under the constraints that the edges of the path must follow a specified order of colors, since in a RNA chain the interactions between the protons are alternated whit relation to the type of experiment. Such path is called assignment walk in [1], thus we refer to this problem as the AssignmentWalk Problem (AWP). In this talk we describe some Integer programming models for solving AWP transforming the edge-colored graph into a n-partite digraph, where n represents the number of partite sets of the vertex set.

REFERENCES

[1] M. Szachniuk, M. Popenda, R.W. Adamiak, J. Blazewicz, An Assignment Walk through 3D NMR Spectrum, In Proceedings of CIBCB'2009, 215-219 (2009).

[2] J. Blazewicz, M. Szachniuk, A. Wojtowicz, RNA tertiary structure determination: NOE pathways construction by tabu search, Bioinformatics, 21, 2356-2361 (2005).

[3] R.W. Adamiak, J. Blazewicz, P. Formanowicz, Z. Gdaniec, M. Kasprzak, M. Popenda, M. Szachniuk, An Algorithm for an Automatic NOE Pathways Analysis of "D NMR Spectra of RNA Dueplexes, Journal of Computational biology, 11, 163-179 (2004).

Simulating gene-gene and gene-environment interactions in complex diseases: Gene-Environment iNteraction Simulator 2

Michele Pinelli (1,2,*), Giovanni Scala (1,3,*), Roberto Amato (1,3), Sergio Cocozza (1,2), Gennaro Miele (1,3,5)

(1) Gruppo Interdipartimentale di Bioinformatica e Biologia Computazionale, Università di Napoli "Federico II" - Università di Salerno, Italy
(2) Dipartimento di Biologia e Patologia Cellulare e Molecolare "L. Califano", Università di Napoli "Federico II", Napoli, Italy
(3) Dipartimento di Scienze Fisiche, Università di Napoli "Federico II", Complesso Universitario di Monte S.Angelo, Napoli, Italy
(4) Dipartimento di Medicina Clinica e Sperimentale, Università di Napoli "Federico II", Napoli, Italy
(5) INFN, Italy
(*) these authors contributed equally to this work

Background: Analysis of complex diseases is a key question in human genetics. Despite its central role, still few data are available. Simulating large biologically realistic data set with known gene-gene and gene-environment interactions influencing a complex disease risk is a convenient way to assess performances of statistical methods.
Results: GENS2 simulates interactions among two genetic factors and one environmental factor also allowing for epistatic interactions. It is based on data with realistic patterns of linkage disequilibrium, has no limitations in terms of how many individuals can be simulated and genetic and environmental factors within a simulated data set. A large effort have been made to allow the input of parameters in terms of standard epidemiological measures in order to be immediately usable by the biomedical community. GENS2 is written in Python language and takes advantage of operators and modules provided by simuPOP framework. It can be used by a graphical or command line interface. It is freely available from http://sourceforge.net/projects/gensim
Conclusions: Data produced by GENS2 can be used as benchmark for statistical tools in the identification of GxG and GxE.

Molecular investigations of the interaction between the ligand-binding domain of the Human aryl-hydrocarbon receptor and its endogenous ligands and xenobiotics from dietary sources.

Maria Salzano (1), Anna Marabotti A (2), Luciano Milanesi (2), Angelo Facchiano (1)

(1) Institute of Food Science (ISA), CNR, Via Roma 64, 83100 Avellino, Italy.
(2) Institute of Biomedical Technologies (ITB), CNR, Via F.lli Cervi 93, 20090 Segrate (MI), Italy

Dioxins are chlorinated heteroaromatic polyciclic organic chemicals. 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD or dioxin), the most known congener of the group, with four chlorine atoms, is considered one of the most toxic man made compounds ever released into the environment. The effects on human health are mediated by the activation of aryl-hydrocarbon receptor (AhR), a cytosolic transcription factor that, in its latent unliganded state, forms complexes with HSP90, p23 and XAP2. Upon ligand binding, AhR translocates to the nucleus, where it complexes with its hetero-dimerization partner, the AhR Nuclear Translocator (ARNT), to modulate expression of AhR target genes containing specific DNA enhancer sequences, known as AhR responsive elements (AhREs). The human AhR (hAhR), accordingly to its documented physiological role, shows high affinity to a distinct subset of ligands, structurally divergent from typical exogenous AhR ligands like TCDD, including sterols, indigoids, heme metabolites, tetrapyrroles such as bilirubin, arachidonic acid metabolites and dietary components. We investigated the interaction of hAhR with TCDD and known ligands in order to obtain new insights about their effects on human health. The ligand binding domain (LBD) of hAhR has been modelled by homology modelling techniques and used for docking simulations with dioxin and eleven xenobiotics and endogenous ligands, in order to investigate the possible effects of food contaminants, as well as of natural food components, in terms of binding and interference with the physiological role of the receptor and its binding with endogenous ligands. The study has evidenced that different recognition sites may exist on the surface of LBD-hAhR. Preferences of the evaluated ligands for the different sites are reported and discussed in view of their functional role.

REFERENCES

Salzano M, Marabotti A, Milanesi L, Facchiano A. (2011) Human aryl-hydrocarbon receptor and its interaction with dioxin and physiological ligands investigated by molecular modelling and docking simulations. Biochem Biophys Res Commun. 2011 Sep 23;413(2):176-81.

COCOMAPS and CONS-COCOMAPS: novel web tools for the analysis of crystallographic complexes and of multiple docking solutions

Anna Vangone (1) Raffaele Spinelli (2) Vittorio Scarano (2) Luigi Cavallo (1) and Romina Oliva (3)

(1) Department of Chemistry and Biology, and (2) Dipartimento di Informatica ed Applicazioni, University of Salerno, Via ponte don Melillo 84084, Fisciano (SA), Italy.
(3) Department of Applied Sciences, University "Parthenope" of Naples, Centro Direzionale Isola C4 80143, Naples, Italy

Most important molecular processes in the cell rely on the interaction between biomolecules. Therefore, the availability of a 3D structure of a biological complex and the characterization of the interacting interface is a fundamental step for possible biomedical and biotechnological applications. Unfortunately, the 3D structure of a significant fraction of biomolecular complexes is difficult to solve experimentally. In this scenario, the development of accurate protein-protein docking softwares is making this kind of simulations an effective tool to predict the 3D structure and the surface of interaction between the molecular partners in macromolecular complexes. However, correctly scoring the obtained solutions to extract native-like ones is still an open problem, which is recently also object of assessment in CAPRI (Critical Assessment of PRedicted Interactions), a community-wide blind docking experiment. This requires the accurate and tedious screening of many docking models in the analysis step. It is therefore of timely interest, both for bioinformaticians and wet biologists, to have programs and tools able to automatically analyse features of a complex interface, and to easily and intuitively discriminate between similar and different binding solutions. However, no available web tool has been implemented to provide interactive contact maps from the 3D structure of a biomolecular complex. For these reasons, we developed COCOMAPS (bioCOmplexes COntact MAPS), a novel web tool to easily and effectively analyse and visualize the interface in protein-protein and protein-nucleic acid complexes. COCOMAPS combines in a single tool the traditional analysis and 3D visualization of biocomplexes with the effectiveness of the contact map view. It can straightforwardly be applied to the analysis of interfaces both in experimental and predicted 3D structures of biological complexes. Further, we used contact maps as the basis for a novel tool, CONS-COCOMAPS (CONSensus-COCOMAPS) developed to measure and visualize the conservation of inter-residue contacts in multiple docking solutions. The application of CONS-COCOMAPS to test-cases taken from recent CAPRI rounds has shown that it is very efficient in highlighting even a very weak consensus that often is biologically meaningful.
Availability:
COCOMAPS: http:/www.molnac.unisa.it/BioTools/cocompas
CONS-COCOMAPS: http:/www.molnac.unisa.it/BioTools/conscocomaps/

REFERENCES

Vangone A, Spinelli R, Scarano V, Cavallo L, Oliva R. (2011) COCOMAPS: a web application to analyze and visualize contacts at the interface of biomolecular complexes. Bioinformatics 27:2915-6.
Vangone A, Oliva R., Cavallo L. CONS-COCOMAPS: A novel tool to measure and visualize the conservation of inter-residue contacts in multiple docking solutions. BMC Bioinformatics, submitted.

Solving Biclustering with a GRASP-like Metaheuristic: Two Case-Studies on Gene Expression Analysis

Angelo Facchiano(1), Paola Festa(2), Anna Marabotti(3), Luciano Milanesi(3), Francesco Musacchia(2)

(1) Istituto di Scienze dell'Alimentazione - CNR, Italy. angelo.facchiano@isa.cnr.it
(2) University of Napoli "Federico II", Italy. {paola.festa,francesco.musacchia}@unina.it
(3) Istituto di Tecnologie Biomediche - CNR, Italy. {anna.marabotti,luciano.milanesi}@itb.cnr.it

The explosion of "omics" data over the past few decades has generated an increasing need of efficiently analyzing high-dimensional gene expression data in several different and heterogenous contexts, such as for example in information retrieval, knowledge discovery, and data mining. For this reason, biclustering, or simultaneous clustering of both genes and conditions has generated considerable interest over the past few decades. Unfortunately, the problem of locating the most signifi- cant bicluster has been shown to be NP-complete. We have designed and implemented a GRASP-like heuristic algorithm to efficiently find good solutions in reasonable running times, and to overcome the inner intractability of the problem from a computational point of view. Experimental results on two datasets of expression data are promising indicating that this algorithm is able to find significant biclusters, especially from a biological point of view.

REFERENCES

1. Facchiano A, Festa P, Marabotti A, Milanesi L, Musacchia F: "A new GRASP-like algorithm for biclustering: a case-study on gene expression analysis". Proceedings of "CIBB 2011 – Eight International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics" (2011) (ISBN: 978-889-0643-70-5).
2. Musacchia F, Marabotti A, Facchiano A, Milanesi L, and Festa P: "Biclustering of gene expression data based on GRASP-like algorithms". Abstract published in the abstract book edited by Edizioni ETS Pisa.

Gene profiling of human PBMCs ex vivo vaccinated with anti idiotypic vaccine for HCV-related lymphoproliferative disorders

Annacarmen Petrizzo, Marialina Tornesello, Franco Maria Buonaguro, Luigi Buonaguro.

Lab. of Molecular Biology and Viral Oncogenesis & AIDS Reference Center, Istituto Nazionale Tumori "Fond. G. Pascale", Naples, Italy

Hepatitis C virus (HCV) is one of the major risk factors for chronic hepatitis, which may progress to cirrhosis and hepatocellular carcinoma, as well as for type II mixed cryoglobulinemia (MC), which may further evolve into an overt B-cell non-Hodgkin's lymphoma (NHL). It has been previously shown that B-cell receptor (BCR) repertoire, expressed by clonal B-cells involved in type II MC as well as in HCV-associated NHL, is constrained to a limited number of variable heavy (VH)- and light (VL)-chain genes. Among these, the VK3-20 light chain idiotype has been selected as a target for passive as well as active immunization strategy. In the present study, we describe the results of a multiparametric analysis of the innate and early adaptive immune response after ex vivo stimulation of human immune cells with the VK3-20 protein. This objective has been pursued performing gene expression profiling analysis as well as multiplex analysis of cytokines.

Structural analysis of Antarctic teleost Toll-like receptor 2

S. Varriale (1), S. Ferraresso (2), M.R. Coscia (1), L. Bargelloni (2), U. Oreste (1)

(1) Institute of protein Biochemistry CNR Naples
(2) Dept of Public Health, Comparative Pathology, and Veterinary Hygiene Univ of Padova

Toll-like receptors are transmembrane proteins involved in the innate immune response, sensing invariant molecular patterns present in microorganisms. All TLRs share the same molecular architecture consisting of an N-terminal ectodomain including a various number of leucine-rich repeat (LRR) motifs, followed by a transmembrane domain and a cytoplasmic Toll/IL-1 R (TIR) domain. In particular, TLR2 recognizes lipoproteins.
We determined the nucleotide sequence of TLR2 from the Antarctic teleosts *Trematomus bernacchii* (Nototheniidae) and *Chionodraco hamatus* (Channichtyidae), respectively. Both sequences, encode 20 leucine-rich repeats (LRRs) in the extracellular region and a classical Toll/IL-1R (TIR) domain in the intracellular region. Nucleotide identity with orthologs from other teleost species ranged between 54.6 and 78.8 %.
Deduced amino acid sequences of the Antarctic TLR2 were aligned with sequences of TLR2 from temperate teleost species. In all the sequences analyzed the same number of LRRs was identified. In all sequences two loops exceeding the LRR sequences were identified at the same positions: L1 next to LRR7 and L2 next to LRR13.
Different methods were used to identify the positively and negatively selected positions by comparing the number of non-synonymous substitutions per non-synonymous site (dN) with the number of synonymous substitutions per synonymous site (dS). Eight codons underwent positive selection and twenty sites were found to be negatively selected.
Molecular models of the *C. hamatus* and *T. bernacchii* TLR2 ectodomains were built by homology modeling. The crystal structure of the ectodomain of *M. musculus* TLR2 (3A7C.pdb) at a resolution of 2.40 Å was used as template. The obtained models were validated using the PROCHECK program and minimized with the GROMAX3.2 package; MD simulations were performed in water for 15 nsec.
The sites of positive selection reside on the convex side of the solenoid, four out of six are in a 35-residue-long region including the central/N-terminal domain boundary: two (335 and 341) in the external loop of LRR11 and the other two (366 and 369) in LRR12 loop.
The region including the positively selected sites has been demonstrated in human TLR2 structure (2Z7X) to be the functional site of ligand interaction, suggesting that the selection pressure has shaped this molecule in ways that increase its activity in the peculiar Antarctic environment.

A five years experience in Agriculture genomics: some comments

Maria Luisa Chiusano

Department of Soil, Plant, Environmental and Animal Production Sciences, University of Naples "Federico II", Via Università 100, 80055 Portici (Italy)

It is almost five years since we began our activity in food science and agriculture genomics being involved in the tomato genome sequencing effort. In this frame we implemented the Italian SOLanaceae Integrated Platform (ISOL@), (Chiusano et al., 2008). In its early life, ISOL@ was exclusively focused on the analysis of the BAC-based tomato genome sequence, but it was constantly updated and evolved to adapt to novel technologies and to include other Solanaceae species. At present, ISOL@ includes tomato (Solanum lycopersicum) and potato (Solanum tuberosum) genome sequences and the publically available Solanaceae transcriptomes, collected in the SolEST database (D'Agostino et al., 2009), offering accessory applications to improve cross referencing among data and profiting from the integration of various, heterogeneous collections from different species. The gathering and convergence of data generated by high-throughput technologies, the effective integration of different collections and the analysis of the information content based on comparative approaches represent the challenges that ISOL@ attempts to solve.
Besides this main target, we had also the opportunity to set up collaborative efforts mainly dedicated to species of agriculture interest, therefore we were basically involved in organizing collections from "omics" approaches which permitted us to exploit the usefulness of large scale analyses for specific applications.
This experience also permitted us to face new challenges, to live the beginning of the next generation sequencing era and to experience the new face of bioinformatics which cannot be anymore considered a low demanding science. What we did not expect is that this effort could require getting back to the beginning of the story: we are now working on A. thaliana since the reference plant genome, completely sequenced in 2000, still requires our attention.

Classification of cancer cell death with generalized eigenvalues

Mario R. Guarracino (1), Petros Xanthopoulos (2), Georgios Pyrgiotakis (3) , Vera Tomaino (2,4), Brij M. Moudgil (3), Panos M. Pardalos (2)

(1) High Performance Computing and Networking Institute - National Research Council of Italy (ICAR-CNR), Via P. Castellino, 111 - 80131 Naples, Italy
(2) Center for Applied Optimization, Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611-6595, USA
(3) Particle Engineering Research Center, University of Florida, Gainesville, FL 32611, USA
(4) Department of Experimental Medicine and Clinic, University Magna Græcia of Catanzaro, Italy

Objective: Accurate cell death discrimination is a time consuming and expensive process that can only be performed in biological laboratories. Nevertheless, it is very useful and arises in many biological and medical applications.
Methods and material: Raman spectra are collected for 84 samples of A549 cell line (human lung cancer epithelia cells) that has been exposed to toxins to simulate the necrotic and apoptotic death. The proposed data mining approach for the multiclass cell death discrimination problem uses a multiclass regularized generalized eigenvalue algorithm for classification (multiReGEC), together with a dimensionality reduction algorithm based on spectral clustering.
Results: The proposed algorithmic scheme can classify A549 lung cancer cells from three different classes (apoptotic death, necrotic death and control cells) with 97.78% ±0.047 accuracy versus 92.22±0.095 without the proposed feature selection preprocessing. The spectrum areas depicted by the algorithm corresponds to the >C O bond from the lipids and the lipid bilayer. This chemical structure undergoes different change of state based on cell death type. Further evidence of the validity of the technique is obtained through the successful classification of 7 cell spectra that undergo hyperthermic treatment.
Conclusions: In this study we propose a fast and automated way of processing Raman spectra for cell death discrimination, using a feature selection algorithm that not only enhances the classification accuracy, but also gives more insight in the undergoing cell death process.

REFERENCES

Mario R. Guarracino, Petros Xanthopoulos, Georgios Pyrgiotakis, Vera Tomaino, Brij M. Moudgil, Panos M. Pardalos. Classification of cancer cell death with spectral dimensionality reduction and generalized eigenvalues. Artificial Intelligence in Medicine 53 (2011) 119– 125.

Biomarker discovery by integrating MALDI-TOF mass spectrometry and bioinformatics techniques: analysis of peritoneal fluid from women affected by endometriosis.

Valentina Spada (1,2), Maria Fiorella Mazzeo (1), Angelo Facchiano (1), Antonio d'Acierno (1), Matteo Stocchero (3), Nadia Miraglia (2), Nicola Sannolo (2), Rosa Anna Siciliano (1)

(1) Istituto di Science dell'Alimentazione – CNR, Avellino
(2) Dipartimento di Medicina Sperimentale, Seconda Università degli Studi di Napoli
(3) S-IN Soluzioni Informatiche Srl, Vicenza

Strategies based on MALDI-TOF mass spectrometry have been recently applied in biomarker discovery for several pathologies by directly analyzing the peptide/protein pattern of biological fluids such as urine, plasma and serum.
In pathological conditions, the proteins or peptides could be abnormally expressed, secreted or processed as a result of the disease: this could modify the peptide/protein profile and consequently the mass spectra, leading to the identification of specific biomarkers. In these years, for instances, this approach has been successfully applied to discover biomarkers of breast and ovarian cancer and Alzheimer disease in serum or plasma samples (1-3).
MALDI-TOF-MS does not require complex pre-analytical steps, thus allowing the rapid analysis of a large set of samples with high reproducibility, specificity and sensitivity. However, suitable bioinformatics tools for the processing and the statistical analysis of MS data are mandatory to identify specific biomarkers.
In the present study this methodology has been applied to the analysis of the peptidic component of peritoneal fluid (PF) samples from women with endometriosis in order to correlate particular features of mass spectra with the different stages of the disease (I-II minimal-mild, III-IV moderate-severe). Endometriosis is a common, benign, oestrogen-dependent, chronic gynecological disorder associated with pelvic pain and infertility. It causes the increase of PF volume, an ovarian exudate, as a result of an enhanced vascular permeability probably due to the high local oestrogen concentration (4). The precise aetiology of endometriosis remains unknown, however it has been suggested a correlation between the onset of the disease and women exposure to endocrine disrupters (5-6).
PF samples, depleted from high molecular weight proteins, have been directly analysed by MALDI-TOF-MS and mass spectra (12 replicates for each sample), acquired in the m/z range 750-4000, were processed with Mascot Wizard program. The obtained peak lists have been aligned along the m/z axis, using the Neapolis software and submitted to clustering procedures with Cluster and TreeView software, and multivariate analysis with Umetrics software to assess correlations between mass spectral features and stage of disease.
Preliminary results highlight that, although the individual variability highly affects the peptidic profile, a few but significant biomarkers are present with a higher frequency in the mass spectra of PF from women with endometriosis in III-IV stages.

REFERENCES

1- Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC, Liotta LA. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet*. 2002, 359 (9306), 572-577.
2- Conrads TP, Zhou M, Petricoin EF 3rd, Liotta L, Veenstra TD. Cancer diagnosis using proteomic patterns. *Expert Rev Mol Diagn*. 2003, 3 (4), 411-420.
3- Lopez FM *et al* High-resolution serum proteomic profiling of Alzheimer disease samples reveals disease-specific. Carrier-protein-bound mass signatures. *Clinical chemistry.* 2005, 51:10, 1946-1954.
4- Koninckx PR, Kennedy SH, Barlow DH. Endometriotic disease: the role of peritoneal fluid. *Human Reproduction Update*. 1998, 4(5), 741-751.
5- Cobellis L, Latini G, De Felice C, Razzi S, Paris I, Ruggieri F, Mazzeo P, Petraglia F. High plasma concentrations of di-(2-ethylhexyl)-phthalate in women with endometriosis.Hum Reprod. 2003, 18(7), 1512-1515.
6- Cobellis L, Colacurci N, Trabucco E, Carpentiero C, Grumetto L. Measurement of bisphenol A and bisphenol B levels in human blood sera from healthy and endometriotic women. *Biomed Chromatogr.* 2009,23(11), 1186-1190.

Interaction between a peptide derived from glycoprotein gp36 of feline immunodeficiency virus and a phospholipid bilayer: a molecular dynamics study

Giuseppe Vitiello, Gerardino D'Errico, Luigi Paduano, Filomena Sica, Antonello Merlino

Dipartimento di Chimica "Paolo Corradini" Università di Napoli "Federico II"

Viral fusion glycoproteins present a membrane-proximal external region (MPER) which is usually rich in aromatic residues and presents a marked tendency to stably reside at the membrane interfaces, leading, through unknown mechanisms, to a destabilization of the bilayer structure. This step has been proposed to be fundamental for the fusion process between target membrane and viral envelope. Both the human immunodeficiency virus (HIV) and the feline analogue (FIV) effect cell entry via a mechanism that involves surface glycoproteins named gp41 and gp36, respectively. In present work, we investigate the interaction between an octapeptide (C8) deriving from the MPER domain of gp36 and POPC bilayers by Molecular Dynamics simulations. The results are compared with experimental data obtained by Neutron Reflectivity, Electron Spin Resonance, Circular Dichroism and Fluorescence Spectroscopy. C8 binds to the lipid bilayer adsorbing onto the membrane surface without deep penetration. As a consequence of this interaction, the bilayer thickness decreases. The association of the peptide with the lipid membrane is driven by hydrogen bonds as well as hydrophobic interactions that the Trp side chains form with the lipid headgroups. Upon peptide-bilayer interaction, C8 forms transient secondary structures ranging from 310 helices to turn conformations, while acyl chains of the peptide-exposed POPC molecules assume a more ordered packing. This asymmetric lipid bilayer perturbation could play a fundamental role in favouring the membrane fusion process.

CpG Islands Undermethylation in Human Genomic Regions under Selective Pressure

Sergio Cocozza (1,2), Most. Mauluda Akhtar (1,2), Gennaro Miele (1,3,4), Antonella Monticelli (5)

(1) Gruppo Interdipartimentale di Bioinformatica e Biologia Computazionale, Universita` di Napoli ‘‘Federico II’’ - Universita` di Salerno, Naples, Italy
(2) Dipartimento di Biologia e Patologia Cellulare e Molecolare ‘‘L. Califano’’, Universita` degli Studi di Napoli ‘‘Federico II’’, Naples, Italy
(3) Dipartimento di Scienze Fisiche, Universita` degli Studi di Napoli ‘‘Federico II’’, Naples, Italy
(4) Istituto Nazionale di Fisica Nucleare – Sezione di Napoli, Naples, Italy
(5) Istituto di Endocrinologia ed Oncologia Sperimentale, CNR Napoli, Naples, Italy

DNA methylation at CpG islands (CGIs) is one of the most intensively studied epigenetic mechanisms. It is fundamental for cellular differentiation and control of transcriptional potential. DNA methylation is involved also in several processes that are central to evolutionary biology, including phenotypic plasticity and evolvability. In this study, we explored the relationship between CpG islands methylation and signatures of selective pressure in Homo Sapiens, using a computational biology approach. By analyzing methylation data of 25 cell lines from the Encyclopedia of DNA Elements (ENCODE) Consortium, we compared the DNA methylation of CpG islands in genomic regions under selective pressure with the methylation of CpG islands in the remaining part of the genome. To define genomic regions under selective pressure, we used three different methods, each oriented to provide distinct information about selective events. Independently of the method and of the cell type used, we found evidences of undermethylation of CGIs in human genomic regions under selective pressure. Additionally, by analyzing SNP frequency in CpG islands, we demonstrated that CpG islands in regions under selective pressure show lower genetic variation. Our findings suggest that the CpG islands in regions under selective pressure seem to be somehow more ‘‘protected’’ from methylation when compared with other regions of the genome.

REFERENCES

Cocozza S, Akhtar MM, Miele G, Monticelli A, 2011 CpG Islands Undermethylation in Human Genomic Regions under Selective Pressure. PLoS ONE 6(8)

The molecular model of Act d 11, a new kiwifruit allergen belonging to the Ripening-related Protein family, shares the same fold of Bet v 1-like allergens

Maria Antonietta Ciardiello (1), Lisa Tuppo (1,2), Laura Camardella (1), Vito Carratore (1), Adriano Mari (2) and Rossana D'Avino (1)

(1) Institute of Protein Biochemistry, CNR, Naples, Italy
(2) Center for Clinical and Experimental Allergology, IDI-IRCCS, Rome, Italy

The primary structure of kirola, the new kiwi allergen named Act d11, has been obtained by direct sequencing of the purified protein. Following the analysis of the primary structure, kirola has been included in the Major Latex Protein/Ripening-related Protein family (ref). The capacity of Act d 11 to inhibit, at least partially, the IgE binding to Bet v 1 and to the homologous allergens, such as Cor a 1, Dau c 1, and Mal d 1(ref), suggests an epitope sharing higher than that inferable from the conservation of specific amino acid residues. Variability in such IgE co-recognition of the different Bet v 1-like molecules suggests that few epitopes could be involved. We have built the molecular model of Act d11 in order to analyse the position of specific amino acid residues and of structural motifs, reported to play an important role for the IgE binding to Bet v 1. Act d 11 model was built using as a template the crystal structures of the birch pollen allergen Bet v 1 (PDB code 1BV1) and of the major celery allergen Api g 1 (PDB code 2BK0:A). Protein fold recognition server Phyre recognized Act d 11 to share the same fold of the two proteins, with highly confident E-values of 1.8 e-18 and 9.8 e-19, respectively. Structure-based sequence alignment of the three proteins submitted to Modeler indicated Act d 11 to share 17% and 15 % identity with Bet v 1 and Api g 1, respectively. Despite the low sequence identity, the model of the three-dimensional structure indicates that the amino acid sequence of Act d 11 fits well into the overall fold of Bet v 1, which is shared with the members of this superfamily and consists of a seven-stranded antiparallel β-sheet and three α-helices. Comparative analysis among available structures of Bet v 1 related proteins indicates that major differences are observed within the α3 helix and the loop regions, whose variability affects the entrance and the shape of the internal cavity. In Act d 11, the α3 helix is shorter, and the loop connecting the helix to the rest of the structure is longer, due to the unwinding of the two N-terminal turns of the α3 helix. The possible conservation of specific amino acid residues/positions playing an important role for the IgE binding to Bet v 1 and to homologous allergens has been analysed. Most of these residues are conserved in Act d 11 and cluster mainly in the regions comprising the p-loop motif and the protein C-terminal domain.

REFERENCES

D'Avino R, Bernardi ML, Wallner M, Palazzo P, Camardella L, Tuppo L, Alessandri C, Breiteneder H, Ferreira F, Ciardiello MA, Mari A (2011) Kiwifruit Act d 11 is the first member of the Ripening-related Protein family identified as an allergen. Allergy 66(7):870-7

microRNA Sequence Analysis in Hormone-Responsive Breast Cancer Cells by Massively Parallel Sequencing

Giorgio Giurato(1), Maria Rosaria De Filippo(1), Concita Cantarella(1), Francesca Rizzo(1), Roberta Tarallo(1), Maria Ravo(1), Giovanni Nassa(1), Ernesto Nola(2) and Alessandro Weisz(1,2)

(1)Laboratory of Molecular Medicine and Genomics, Faculty Medicine and Surgery of the University of Salerno, Baronissi (SA), Italy
(2)Department of General Pathology of the Second University of Naples, Napoli, Italy

microRNAs (miRNAs) are master regulators of gene expression for their ability to influence mRNA concentration and activity by post-transcriptional mechanisms. Several lines of evidence indicate extensive miRNA deregulation in breast cancer (BC) and recent studies highlighted the role of miRNAs in BC cells response to estrogens. In particular, our previous genome-wide miRNA expression studies[1,2] have shown a widespread miRNA responsiveness to estrogens, suggesting the possibility that estrogens exert a specific control upon miRNA gene activity in BC cells. Indeed chromatin imuno-precipitation/massively parallel sequencing (ChIP-Seq) showed that in MCF-7 cells[1,3] estrogen receptors bind in close proximity of several miRNA genes, suggesting involvement of these transcription factors in small non-coding RNA biogenesis. Starting from these observation we have applied a Next-Generation-Sequencing technology approach (miRNA-Seq) for miRNome profiling in hormone-responsive BC cells model to investigate the responsiveness of miRNA to estrogens and to identify new miRNA targeted by these receptors. Our attention, have been also focused on analysis of single nucleotide variants (SNVs) involving miRNAs and the corresponding genes. The rationale for this study was that polymorfisms in miRNA genes can affect the expression of many downstream-regulated genes and SNVs can influence the structure of miRNA precursor, the efficiency of miRNA biogenesis and miRNA-target recognition. Through bioinformatics tools, we are evaluating potential polymorfisms on pre-miRNA and mature miRNA sequences. Also, comparing RNA and DNA sequences in the same cell line we are trying to understand whether a class of SNVs we identified result from inherited SNPs (single nucleotide polymorphisms), somatic mutations that occurred during the carcinogenic process or cell-type specific RNA editing.

REFERENCES

[1] Cicatiello et al, Am J Pathol. 2010; 176(5):2113-30
[2] Paris O. et al, Oncogene 2011. In press
[3] Grober et al, BMC Genomics. 2011; 12(1):36

Second-generation genomic sequencing and a bioinformatic approach to the study of the human microbiome in selected diseases of the human gut

Giorgio Casaburi (1,2), Valeria D'Argenio (1,2), Vincenza Precone (1,2), Lucia Sacchetti (1,2), Francesco Salvatore (1,2)

(1) CEINGE-Biotecnologie Avanzate, Naples, Italy
(2) Department of Biochemistry and Medical Biotechnologies, University of Naples Federico II, Naples, Italy

The advent of next-generation sequencing technologies has greatly impacted on various fields of biological research. Metagenomics is a discipline that enables the study of all the genomes of a microbial community directly recovered from specific environments. Understanding the dynamic and variable nature of human microbial communities is important because these communities seem to be related to disease onset and/or development. We plan to study the human gut microbiome to assess its associations with the genes of the human intestinal microbiota, and with health and disease. The aim of this Ph.D. project is to set up and validate an efficient pipeline for the management and the analysis of metagenomic data obtained with next-generation sequencing methodologies. A problem in metagenomic data analysis is to determine the taxonomic composition of a given dataset. A metagenomic dataset consists of thousands of short DNA or RNA fragments called 'reads'. Efficient handling of this large amount of data is paramount to our project. We will evaluate different tools and softwares to develop an in-house pipeline for fast and accurate analysis. The pipeline and the related tools will be used to study groups of patients with different intestinal diseases (e.g., celiac and inflammatory bowel diseases) and matched controls that will be analyzed by NGS in our lab. To provide a bioinformatic tool for the quick analysis of metagenomic sequencing data will assess the role of the human gut microbiome in different disease. The results of this study will be easily applicable to other research areas.

The lag time before the secondary response to antigen: insights from a mathematical interpretation of experimental data

Francesca Mantile (1), Carla Basile (1), Antonella Caivano (2), Mario Nicodemi (3), Piergiuseppe De Berardinis (2), Antonella Prisco (1)

(1) Istituto di Genetica e Biofisica, CNR, Naples, Italy.
(2) Istituto di Biochimica delle Proteine, CNR, Naples, Italy.
(3) Dipartimento di Fisica, Università degli Studi di Napoli Federico II, Naples, Italy.

The magnitude of the antibody response to a vaccine can usually be increased by multiple administrations. We set out to interpret the contribution of two subsequent doses of the same vaccine to the antibody titer. We analysed the time course of the antibody titer to a model B epitope elicited by vaccination with two different non-replicating vaccines antigen delivery systems, namely filamentous bacteriophage fd and self-assembling multimeric protein E2. Data obtained from mice that received one dose of vaccine were used to generate a mathematical prediction of the antibody titer elicited by two doses, under the hypothesis of a merely addictive effect. Comparing predictions with experimental observations, we observed that, when the second dose was administered 6 months after the first dose, observed antibody titers were significantly higher than titers predicted under the hypothesis of an addictive effect: a booster effect / immunological memory could be observed. Conversely, depending on antigen "context", namely the different antigen delivery system, the effect of a day 15 dose was either quantitatively identical to the effect of the first dose (with an addictive effect) or quantitatively different (with a booster effect). Thus, a lag time exists before a booster effect can be evoked; during lag time, the effect of a second dose of vaccine on antibody titers is merely addictive. Lag time can significantly vary, depending, for instance, on the antigen delivery system.

Docking studies for simulating the complexes between Sirt-1 and some activators

Vasu Gautam (1), Susan Costantini (2), Antonello Mai (3) and Giovanni Colonna (1)

(1) Dottorato di Biologia Computazionale, CRISCEB – Seconda Università degli Studi di Napoli
(2) INT Pascale - CROM, via Ammiraglio Bianco, Mercogliano (AV)
(3) Istituto Pasteur - Fondazione Cenci Bolognetti, Dipartimento di Chimica e Tecnologie del Farmaco "Sapienza", Università di Roma

Sirt-1 is a NAD-dependent deacetylase that acts as a regulator of mammalian mitochondrial biogenesis. Through its deacetylase activity, Sirt-1 can activate and control some important metabolic cascades, such as glucose metabolism and levels of insulin secretion. Therefore it is involved in diseases such as diabetes and metabolic syndrome and is suspected to have a key role in the mechanisms of longevity induced by caloric restriction. Recently our research group has predicted three-dimensional structure of Sirt-1 and its complex with AROS, that is its endogenous activator. Sirt-1 presents structural and functional properties typical of disordered proteins. In fact, it is composed of 4 different regions indicated as N-terminal region, allosteric site, catalytic core and C-terminal region. The catalytic core has an alpha/beta structure of Rossman-fold type, while the allosteric site, located between the N-terminal region and the catalytic core, has a structure consisting of 4 short alpha helices. On the other hand the two terminal regions consist of short helices and beta-strands and very long irregular regions in agreement with the disorder predictions made by various specific programs. In this work we have modeled complexes between Sirt-1 and various organic molecules, which have experimentally shown an high capacity to activate this protein. In particular, Autodock 4 program has been used to model the complexes using docking methods and considering the flexible ligands. For the analysis of complexes we have used different programs for the evaluation of the interaction residues, hydrogen bonds, salt bridges, etc. The complexes obtained show that the different activators bind Sirt-1 using all the same localized region near the allosteric site. Comparing these complexes with that between AROS and Sirt-1, already published, it appeared that the binding region is quite conserved residues and consists of charged and aromatic compounds. Since in literature the experimental EC50 data of Sirt-1 activity are reported, we correlated these data with the values of binding energy between the activators and Sirt-1. Further studies will regard the modeling of other complexes between Sirt-1 and new activators or inhibitors to highlight the molecular basis of these interactions and to understand how this protein is activated or inhibited by different molecules.

jFKD: a Java Tool for Developing Diagnostic Decision Support Systems

Antonio d'Acierno

Institute of Food Sciences - Italian National Research Council. Via Roma 64, Avellino, Italy. dacierno.a@isa.cnr.it

Recently, we proposed and validated a six-steps methodology for extracting fuzzy knowledge from data; our current aim is to implement this solution in java, since this simplifies the porting of solutions available in the public domain avoiding reinventing each time the wheel. Moreover, variables involved, for example, in a diagnostic decision support systems need to be carefully described using one of the apposite (already available) ontologies and java offers a plethora of methods to handle formally represented knowledge. Then, sure enough, java code is the easiest to be distributed and ported on whatever hardware platform.
Using an *iterative and incremental development* cycle, we created a base version of the system that is currently under test. To simplify the fuzzification step, we assumed that the crisp knowledge is expressed as a set of rule each of which is in the form of a conjunction of conditions. Then, we do not yet allow modifications and additions of crisp rules, an important feature to be added having in mind to allow physicians to drive the extraction process. Concerning the FIS, we assume a simplified model with a single binary output (to be interpreted, in diagnostics domains, as disease/not disease) and, moreover, the possibility of using more FISs in parallel has not yet considered. We also adapt our fuzzy systems using straight mean square errors, while interpretability and confidence need to be also considered. Ontology based facilities are not yet included in our system while, last, our system lacks an (even minimal) graphical interface.
We have clearly organized our system in packages. The *crispKnowledge* package mainly contains the classes used to represent, to extract and to use the crisp knowledge (perhaps) available in our data while the *fuzzyKnowledge* package is used to handle and represent the fuzzy knowledge; its entry point is the *Fuzzifier* that, starting from the extracted crisp rules, creates the *FuzzyKnowledge* object that in turn is, roughly specking, a set of fuzzy rules.
The obtained fuzzy knowledge is used to instantiate a *FIS* object, belonging to an abstract class whose implementation has at least to provide an *adapt()* method, used in our methodology to adapt membership functions, and an *evaluate()* method that has in input a *DataPattern*.
It is last worth to be cited the *DataSet* class used to handle data to be used. Data are organized as a set of *DataPattern* objects, each storing a pattern to be analyzed (values and output class).
Presently, there is a running version of our tool; first of all, the *CrispExtractor* is implemented wrapping a MATLAB decision tree. For what concerns the *FIS*, we use *jFuzzyLogic*, a java package that implements the standard Fuzzy Control Language specification. *DataPattern* has been implemented for some well known datasets (WBCD and PIMA) that so can be used without any further coding activity.

Paralog gene networks and gene expression profiles in *Arabidopsis thaliana*

Pasquale Di Salle*, Maria Luisa Chiusano

Department of Soil, Plant, Environment and Animal Production Sciences.
**Phd student in Computational Biology and Bioinformatics, XXVI cycle, University of Naples Federico II*

One of the most important model organism for molecular biology in plants is *Arabidopsis thaliana.* After the release of the whole genome sequence by "The Arabidopsis Genome Initiative" in 2000, this plant showed interesting and amazing features: despite easy cultivation properties and a small diploid genome, the unexpected complexity of the latter in terms of intragenome duplications and the lack of a fulfilled gene annotation, underline that deeper analyses are still required, in order to understand the molecular biology and the evolution of this plant. Many resources are currently available for this reference species. Among these the "Expression Atlas of Arabidopsis Development" (Nascarrays) release in 2004, collects the expression values for a wide collection of genes and from 80 biologically different samples in triplicate, representing several tissues and developmental stages.
We considered the collection of results from the Nascarray concerning different parts of the plant in physiological conditions, and performed a set of preliminary investigations to define classes of highly and low expressed genes among all those represented in the chip. Pearson correlation tests permitted to define genes correlated by their expression and we highlighted that i) about 83% of the genes have at least one correlation with another gene and that ii) inverse correlations are less frequent than direct ones. We also noticed that genes with high or low number of correlations belong to specific gene ontology classes and that, interestingly, they show peculiarities in specific tissues as in root and in flower.
The preliminary results here presented are in the frame of our effort to analyse expression patterns in *A. thaliana* at the light of the duplication events that shaped the genome and consequently the gene organization of this plant.