ALESSANDRO VASCIAVEO

COMPUTATIONAL MODELS AND ALGORITHMS TO SOLVE LARGE-SCALE PROBLEMS IN NETWORK BIOLOGY
PRESENTATION OUTLINE

PRESENTATION OUTLINE - BY COLLABORATIONS

- Ph.D Studies
  1. Gene Regulatory Networks Simulations
  2. Network Models for Gene Therapy studies
  3. Network Models for Cancer Research
The emerging tools of network medicine offer a platform to explore systematically [...] complexity of a particular disease [...] advances in this direction are essential

Biologists are joining the big-data club - As they grapple with increasingly large data sets, biologists and computer scientists uncork new bottlenecks.


Investigation of mathematical and Information Theory models that are well suited for the representation of biological phenomena using network approaches applied to large datasets

Making models understandable to elucidate the mechanisms of cell regulation

New models and software tools have been the results of these studies, and either research advances about new conceptual frameworks or their implementation have been published on several international conferences and journals.
Transcriptome of non-coding regions produces functional RNAs

- Long noncoding RNAs (IncRNAs) are emerging as important regulators of tissue physiology and disease processes including cancer

- Functional studies have confirmed that miRNA dysregulation is causal in many cases of cancer


Gene Regulatory Networks - Research Rationale & Technological State-of-Art

- State-of-Art tools focus mainly on protein-coding genes.
- Increasing need of systematic procedures to analyze regulatory networks with different semantic layers (data integration).
- Leak of tools for the simulations of regulatory network dynamics with thousands of nodes.
Gene Regulatory Networks - Outline of the Approach

- Design and development of a network dynamics simulator for GRN by using a Boolean Network formalism (EBNT Simulator)

- Inclusion of post-transcriptional activity in the model

- Validation of the framework by qualitative analysis using MicroRNA

- On Cloud implementation for large networks

- Proposed methodology for a quantitative analysis
Network Attractors <-> Stable Cell’s States

Huang, BioEssay 2012
SIMULATION OF GENE REGULATORY NETWORKS – RESULTS

- Qualitative validation of the EBNT on biological pathways
  - mTOR pathway (KEGG hsa04150) / knock-out of miR-1976
  - Study of the role miR-7 in Drosophila on two case studies
    - Photoreceptor determination network
    - Sensory organ precursor (SOP) determination network
  - The central role of the miRNA for increasing the network stability has been highlighted in both the networks, confirming the cooperative stabilizing role of miR-7


GRN DYNAMICS SIMULATION – ON CLOUD COMPUTING – WHY?

- Drawbacks of state of art technologies
  - Simulation of tens or few hundreds or nodes
  - Limitation in the degree connectivity per node
  - Few tools let to analyze the State Space of GRN network dynamics
1. Pathway download from KEGG/Reactome

2. Network Enhancing with ReNE

3. Generation of Input file for EBNT

4. EBNT Simulation via multithreaded processes

5. MapReduce implementation for network dynamics state space generation and attractor collection

6. Output on XML custom format compatible with Cytoscape for State Space Network Visualization
Simulation of mTOR Pathway (KEGG hsa04150)

After enhancing process it has 1668 nodes and 12603 edges

Input Generation

MapReduce
Vasciaveo, Alessandro; Benso, Alfredo; Di Carlo, Stefano; Politano, Gianfranco; Savino, Alessandro; Bertone, Fabrizio; Caragnano, Giuseppe; Terzo, Olivier (2015). A cloud-based approach for Gene Regulatory Networks dynamics simulations. In: IEEE 4th Mediterranean Conference on Embedded Computing (MECO), Budva, Montenegro, 14-18 June 2015. pp. 72-76

GENE EXPRESSION VS. NETWORK ATTRACTORS

- How well a GRN is modeling the actual regulatory mechanisms in the cell?
- Possibility of correlating GRN Dynamics with Gene Expression Profiles
- Formalize a methodology
QUANTITATIVE FRAMEWORK - MOTIVATIONS

GENE EXPRESSION VS. NETWORK ATTRACTORS

PROBLEM STATEMENT

Experimental Assays

- qualitative information
- small/medium amount of genes
- model regulatory relationships
- model the regulatory network dynamics of a single cell

In Silico Assays

- quantitative information
- large amount of genes
- no causal relationship
- average information of a cell population

Gene Regulatory Networks Models

RNA-Seq

NGS Technologies

Microarray Assays
GENE EXPRESSION VS. NETWORK ATTRACTORS – THE PROPOSED METHODOLOGY
Design and development of a GRN simulator (EBNT) with post-transcriptional activity modeling capability.

In Silico validation by means of a qualitative analysis have shown results that seem compatible with the data available in literature.

The proposed framework Network Attractors Vs. GEP takes into account the different aspects of both in silico and in vitro/in vivo assays using a quantitative approach.

Further studies may assess the biological compatibility between a pathway and a GRN model.
A NoSQL DBMS (e.g. MongoDB) could increase the performance of the proposed on cloud architecture for the simulation of GRN by means of Boolean Networks.

Use of single-cell HTS technologies could increase the reliability of the outcome.
GENE THERAPY
WHAT IS GENE THERAPY

GENE THERAPY – WHAT IS?

- Gene Therapy is an experimental technique to treat or prevent diseases.
- Viral vector integration is a process exploited in GT to correct defective cells of an individual and to drive the health status from the pathological condition to a normal one.
- Large insertional mutagenesis screenings are used to:
  - Integration site analysis.
  - Assess the safety of the treatment in clinical GT, and design safer GT protocols.
  - Detection of diseases related genomic features (e.g. oncogenes).
HOW GENE THERAPY WORKS?

GENE THERAPY – HOW DOES IT WORK?

**Gene augmentation therapy**
- **Cell with non-functioning gene**
- **Functioning gene**
- **Cell functioning normally**

**Gene inhibition therapy**
- **Cell containing faulty gene**
- **Blocking gene**
- **New gene product blocks faulty gene**
- **Cell functioning normally**

As consequence of this **perturbation** the affected cell may step from the primary illness state to a secondary state.

Gene therapy is a **promising treatment** option for a number of diseases (including inherited disorders, some types of cancer, and certain viral infections),

GT technique **remains risky** and is still under study to make sure that it will be safe and effective.

Gene therapy is currently only being tested for the treatment of diseases that have no other cures.
**CIS DETECTION – STATE OF ART**

**CIS Identification**
- Kernel Convolution Frameworks
- Standard Fixed Windows Methods
- Thresholds of 10, 50, 100, 200 Kbp

**Novel Cancer Genes Identification**
- NGA (Nearest Gene Approach)
- One-to-many Annotation
Network approach to model CIS embedded in a statistical framework called *Graph-Based Framework (GBF)*

Graph theory can be used to infer characteristics and properties of the integration process

- Scale-free Networks as a model of viral integration processes

Novel algorithm for CIS identification
THE GBF METHODOLOGY

THE GRAPH BASED FRAMEWORK METHODOLOGY

- Using the Shannon Entropy as a measure of CIS heterogeneity
- Using a likelihood ratio as confidence test of CIS validity
- Gene Atmosphere (GA) for cancer gene discovery (many-to-many annotation)
- BioMart data service integration using a REST client
- Cytoscape Application GTAnalyzer
OVERVIEW OF THE RESULTS

GENE THERAPY - RESULTS

- Building of an integrate Systems Biology framework by using a network approach
- Network analysis using topological proprieties (i.e. scale-free networks)
- Using the Retroviral Tagged Cancer Gene Database (RTCGD) as dataset for the validation of the GBF
THE SOFTWARE IMPLEMENTATION

GENE THERAPY – GTANALYZER
Gene Therapy - Publications

- Vasciaveo, Alessandro; Velevska, Ivana; Politano, Gianfranco; Savino, Alessandro; Schmidt, Manfred; Fronza, Raffaele (2016). *Common integration sites of published datasets identified using a graph-based framework*. In: COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL, vol. 14, pp. 87-90. - ISSN 2001-0370

GENE THERAPY - CONCLUSIONS AND FUTURE WORKS

- The CIS disposition over a plane allows the researcher to identify the different macro component of the analysis in a qualitative way

- CIS 3D Chromatin Modeling by using Hi-C datasets
CANCER NETWORK AND SYSTEMS BIOLOGY
CANCER NETWORK BIOLOGY – RATIONALE

- Cancers are a leading cause of morbidity and mortality worldwide, with over 14 million new cases and 8 million deaths in 2012
  

- Cancer is an heterogeneous disease

- Cancers with a high level of genetic instability and high levels of clonal heterogeneity are also likely to have a poor prognosis
  
  LA Loeb, Nature Review Cancer 2011
Master Regulator Analysis – What is a MR?

- Master Regulator
- Tumor Checkpoint Hypothesis
- Reverse Engineering of GRN
- Protein Activity Inference and Interactome Building
- Reversion of signatures - GES/PAS

The rational behind is that if the drug is effective then its signature should be inversely correlated to the one implemented by the disease phenotype
ANALYSIS PIPELINE

- Building of a Context-Specific Network
- GES Generation
- PAS Generation
- Drug Perturbation Analysis
- Drug Reprioritization
- Synergy Analysis
NEUROBLASTOMA - NBL STATE-OF-ART

- NLB accounts for 50 percent of all cancers in infants, making it the most common tumor in infants younger than 1 year (15% of pediatric cancer death)

- This cancer is characterized by multiple tumor subtypes that vary in their aggressiveness and clinical prognosis

- Previous analysis identified MYCN submodules on different NBL subtypes (Califano Lab, under review)

- Inhibitors that disrupt these regulatory modules will likely stand a greater chance of success in vivo, and lead to tumor collapse
92 compounds with selectivity for MYCNA cells were screened for the ability to revert the MR signature that drives this tumor

These compounds also destabilize oncogenic MYCN protein

Using TARGET and NRC to build a GRN for the inference of a PAS signature

Analyze the drug perturbation assays (multiplexed shallow sequencing) to look for a reversion in the disease signature
MULTIPLEXED RNA-SEQ

DRUG PERTURBATION SCREENS BY SHALLOW SEQUENCING

- 96-well plate
- 4-12 controls (DMSO/MOCK)
- drug perturbations
  - 2-3 time points
  - 2-3 drug concentrations
NBL PIPELINE

NEUROBLASTOMA - NBL ANALYSIS

- Transcript isolation
- Barcoding
- cDNA synthesis in well

Determined IC80 for 92 bioactive molecules

Percent Viability

Drug Concentration (μM)

0 0.25 0.5 1 1.5 2 2.5

0 25 50 75 100

SK-N-BE(2) Cell Line

96-well format

pooled cDNA

Sequencing and Analysis
NBL - OVERVIEW OF RESULTS

NBL - ANALYSIS - RESULTS

Protein Activity (low to high)

Sertraline_15_24

Sirolimus_0.25_24

Good Outcome

Poor Outcome
NBL CONCLUSIONS

NBL ANALYSIS - ISSUES

▸ Used IC80 as drug dosage
  ▸ too strong dosage
  ▸ low reads affected outcome/outliers

NBL ANALYSIS - SOLUTIONS

▸ Use IC20 as drug dosage
▸ Use more cells in the culture
▸ Look for cell lines that are more representative and compute molecular (GES, PAS) against other cell lines
NBL - FUTURE WORKS

- Elucidate the mechanisms through which MYCNA selective compounds revert the MR signature

- It is hypothesized that compounds which are lethal by reverting the MR signature make the promising drug candidates

- Test selected drugs *in vivo* using on mouse PDX models
The developing of new drugs is a very long and expensive process that takes several years and costs hundreds of millions of dollars.

The risk that a drug is not approved by regulations because a failure in clinical phases is high.

The repurposing of existing drugs by computational analysis is a promising approach to overcome the aforementioned issues.
Many diseases are driven by complex molecular interactions.

The target of a single actionable component may not be sufficient for the disruption of the aberrant activity.

From single-agent compound to multi-target inhibitors.

Perform of Synergy Analysis

- Network analysis approaches
- Use GEO, TCGA, CMap, LINCS, DrugBank and other DB of drug perturbation signatures
OUTCOME OF THE SYNERGY ANALYSIS

COMPUTATIONAL SYNERGY ANALYSIS (II)

(sorted) PDX GBM Synergy Analysis – Drug pairs NES > 1.96 shown

Clomertinib_1.5578_24 x Melphalan_0.81866916_24
Valproic Acid_2.362847446_24 x Melphalan_0.81866916_24
Sonbplidin_0.24881724_24 x Melphalan_0.81866916_24
Clomertinib_1.5578_24 x Tamoxifen_1.960191772_24
Bosalidinib_0.377073906_24 x Leucovorin_2.10501837_24
Cabazitaxel_1.104376143_24 x Valproic Acid_2.362847446_24
Clomertinib_1.5578_24 x Docetaxel_0.4211280_24
Tonmelina_1.586110035_24 x Clomertinib_1.5578_24
Palbociclib_0.114987325_24 x Bosulinenb_0.377073906_24
Vismodegib_10_24 x Tamoxifen_0.960191772_24
Gliparib_3.9541_24 x Tamoxifen_0.960191772_24
Bosalidinib_0.377073906_24 x Melphalan_0.81866916_24
Nilotinib_2.284232304_24 x Bosulinenb_0.377073906_24
Tamoxifen_0.68068_24 x Gliparib_3.9541_24
Mitotane_0.63709_24 x Melphalan_0.81866916_24
Docetaxel_0.4211280_24 x Clomertinib_1.5578_24
Tamoxifen_0.960191772_24 x Clomertinib_1.5578_24
Nilotinib_2.284232304_24 x Tamoxifen_0.960191772_24
Abrpin_0.186854344_24 x Tamoxifen_0.960191772_24

Leucovorin_2_1.0501837_24 x Vandetanib_7.2336_24
Vandetanib_6.493_24 x Leucovorin_2_1.0501837_24
Bosalidinib_0.377073906_24 x Hydroxyurea_10_24
Cabazitaxel_0.309596389_24 x Vitamin A_2.11235579_24
Cytarabine_1.3871_24 x Valproic Acid_2.362847446_24
Irinotecan_1.4843_24 x Leucovorin_2.10501837_24
Prednisone_0.844577018_24 x Palbociclib_0.114987325_24
Decitabine_0.641413552_24 x Valproic Acid_2.362847446_24
Bevaxafen_1.593932353_24 x Leucovorin_2.10501837_24
Sonabegli_5.8941_24 x Paroxetine_0_01161_24
Busulfan_0.295994335_24 x Docetaxel_0.4211280_24
Vandetanib_7.2336_24 x Docetaxel_0.4211280_24
Tamoxifen_0.960191772_24 x Cytarabine_1.3871_24
Vitamin A_2.11235579_24 x Cabazitaxel_0.309596389_24
The analysis performed using these pipelines seem **very promising**.

The *elimination of bias* due to recent technology implementation can increase the reliability of the molecular signatures.

The integration of **epigenetic** layers (ATAC-Seq) into the pipeline could reduce the False Positive interactions in the interactome.
THANKS

ACKNOWLEDGMENTS

Systems Biology Research Group
Department of Control and Computer Engineering
Politecnico di Torino, Turin, Italy

- Prof. Alfredo Benso
- Prof. Stefano Di Carlo
- Dr. Gianfranco Politano
- Dr. Alessandro Savino

NCT (National Center for Tumor Diseases) - DKFZ (German Cancer Research Center)
Department of Translational Oncology
Section of Molecular and Gene Therapy, Heidelberg, Germany

- Prof. Manfred Schmidt
- Dr. Raffaele Fronza

Prezdborski Lab - Departments of Neurology, Pathology, and Cell Biology
Columbia University, New York, U.S.

- Prof. Serge Przedborski
- Kristin Politi

Califano Lab
Department of Systems Biology
Columbia University, New York, U.S.

- Prof. Andrea Califano
- Dr. Mariano Alvarez
- Dr. Eugene Douglass
- Dr. Prab Mundi
- Ding Hongxu
- Jing He
- Nicholas Giangreco

ISMB - Istituto Superiore Mario Boella, Turin, Italy

- Dr. Giuseppe Carignano
- Fabrizio Bertone

Stockwell Lab
Department of Biological Science
Columbia University, New York, U.S.

- Prof. Brent Stockwell
- Dr. Michael Stokes
REFERENCES


