Network Biology SIG Program — Friday, July 8th

Accepted Oral Presentations

Detection of gene communities in multi-networks reveals cancer drivers

Laura Cantini[1,2,3,4], Enzo Medico[5,6], Santo Fortunato[7], Michele Caselle[8]

[1]Computational and Systems Biology of Cancer, Institut Curie, Paris, France, [2]PSL Research University, Paris, France, [3]Institut National de la Santé et de la Recherche Médicale U900, Paris, France, [4]Mines ParisTech, Paris, France, [5]Università di Torino, Department of Oncology, Candiolo, Italy, [6]Candiolo Cancer Institute, FPO IRCCS, Candiolo, Italy, [7]Department of Computer Science, Aalto University School of Science, Aalto, Finland, [8]Università di Torino, Department of Physics and INFN, Torino, Italy

In the past years the advent of high-throughput experimental technologies provided biologists with a flood of molecular data. This huge amount of information requires the design of efficient methodologies to be interpreted. Among them, network analysis proved to be very effective to capture the molecular complexity of human diseases. Thus far, network-based computational methods were primarily focused on the analysis of single biological networks. However, such approach turned out to be insufficient to unveil functional regulatory patterns originating from complex interactions across multiple layers of biological relationships. Therefore, a new pressing request in molecular biology is to design network-based methods allowing combined use of multiple levels of genomic information. Many solutions have been proposed in the last few years. Among them a special role has been played by multiplex networks, which emerged recently as one of the major contemporary topics in network theory. Some relevant applications in biology already exist: Li and colleagues studied a multilayer structure composed of 130 co-expression networks, in which each layer represents a different experimental condition. Subsequently, they also constructed two-layer networks, composed of a standard co-expression network and an exon co-splicing network. More recently, Bennett and co-workers identified communities on the multiplex network of physical, genetic and co-expression interactions, in yeast, using mathematical programming with the modularity by Newman and Girvan as objective function. Following this line we propose a multi-network-based approach for the identification of candidate driving genes in cancer. We use the expression multi-networks instead of multiplex because we will not consider couplings between the layers.

Cancer is a complex disease caused by a progressive accumulation of dysfunctions in neoplastic cells. During the last decade, technological advancements enabled laboratories to quantitatively monitor these alterations. Efficient methodologies were designed to interpret these data and identify the genes driving the neoplastic growth. However these approaches are classically applied to study separately biological measurements that are clearly not independent. For this reason, we consider the identification of driver cancer genes as perfectly suited for a multi-network-type analysis. To address this problem, we combined, in a single multi-network, four different gene networks: (i) Transcription Factor (TF) co-targeting network, (ii) microRNA co-targeting network, (iii) Protein-Protein Interaction (PPI) network and (iv) gene co-expression network. The rationale behind this choice is that the insurgence of cancer is typically due to a dysregulation of the signaling and/or of the regulatory network of the cell. These regulatory pathways are tightly controlled in the cell both at the transcriptional and at the post-transcriptional levels and their alteration very often involves modification in the expression levels of genes which are at the same time partners in a protein-protein interaction and targeted by the same set of transcription factors and miRNAs. These are exactly the events which are selected and prioritized in the Multi-network-based analysis that we propose. Following the construction of the multi-network, we proceed with the identification of communities, that is, of groups of nodes that are densely connected to each other, but sparsely connected to the other nodes of the network. This is achieved by detection of gene communities within each multi-network layer and subsequent identification of communities via consensus clustering across the four layers. It is well known that community detection within a network is an open and difficult problem, for this reason we tested some well-known community detection algorithms all of which can be run in our multi-network analysis package “Gene4x” described in Fig. 1 and available at https://github.com/lcan88/Gene4x.git. To check whether the multi-network communities are more biologically relevant than the communities obtained in the expression network alone, we applied the analysis to human gastric, lung, pancreas and colon cancer datasets, and tested the resulting multi-network or co-expression network communities for functional enrichment, or differential expression between tumor and normal tissues. In all four cancer types, the multi-network communities highlighted new relevant tumor-specific functional enrichments (including chromosomal aberrations, candidate markers and driver genes) not detected by the co-expression network alone, providing evidence of the power of Multi-network-based approaches in extracting knowledge from complex, multidimensional molecular data.

Integrating Prior Pathway Knowledge into Methods for Network Reconstruction

Frank Kramer[1], Tim Beissbarth[1]

[1]University Medical Center Goettingen

MOTIVATION

The integration of prior knowledge into methods for network reconstruction translates into using the accumulated pathway knowledge of the last decades as building blocks for future discoveries. A plethora of databases (Bader et al, 2006) offer vast amounts of literature knowledge about biological signaling networks, including the well known Reactome.org (Croft et al, 2010) and the Pathway Interaction Database of NCI (Schaefer et al, 2009). Over years literature knowledge of molecular interactions has been collected and manually curated into hundreds of pathways.

RESULTS

Recent advances ease the access to pathway knowledge encoded in a standardized way (Demir et al, 2010) and offer new approaches in working with pathway knowledge.

Based on our software package rBiopaxParser (Kramer et al 2013) we started by generating interactomes of complete pathway databases. These interactomes are the programmatically merged totality of interactions encoded in the pathway definitions of the specific databases.

We used graph theoretical approaches in order to compile merged consensus networks for specific molecules from these interactomes. Based on the available paths within these interactomes we reconstruct network from gene expressions data of intervention experiments on human breast cancer cells.

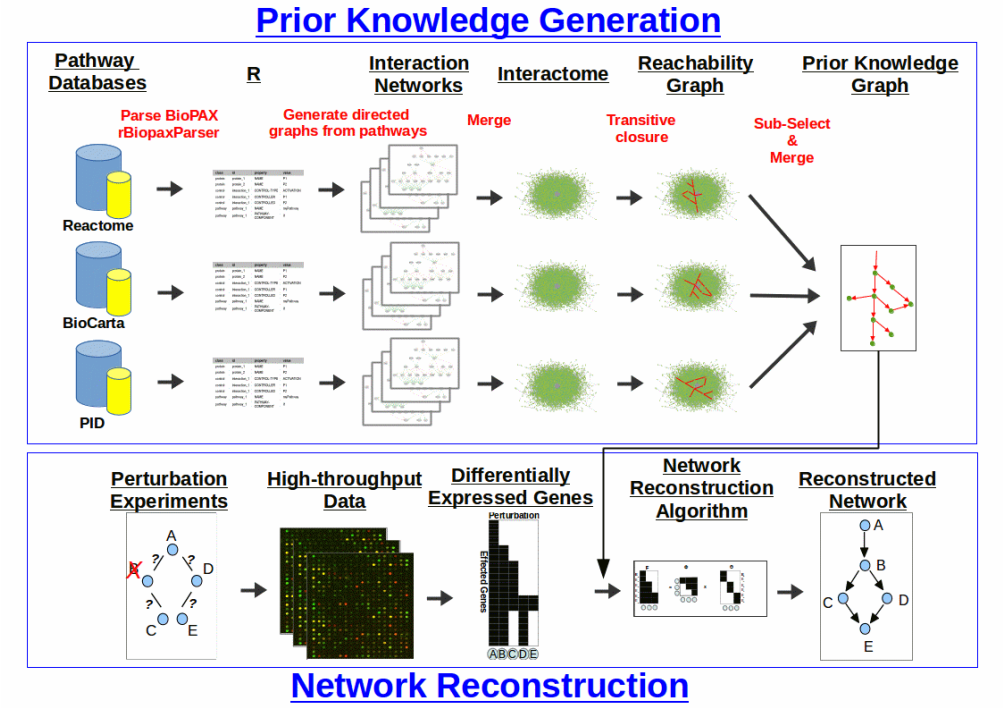

Figure 1: This figure demonstrates the workflow on how to integrate pathway knowledge into methods for network reconstruction.

PRESENTATION

In our talk we show that integration of prior knowledge networks increases the performance of network reconstruction methods. Furthermore, the concordance of database-specific interactomes, the prior knowledge graphs and the results of network reconstruction is evaluted. Impact of the integrated prior knowledge on the reconstruction results can be re-traced to specific paths or connections stemming from specific pathway databases.

Our talk will combine insights into prior knowledge generation based on a consensus network of different pathway databases and describe how literature knowledge can be reconstructed using methods for network reconstruction.

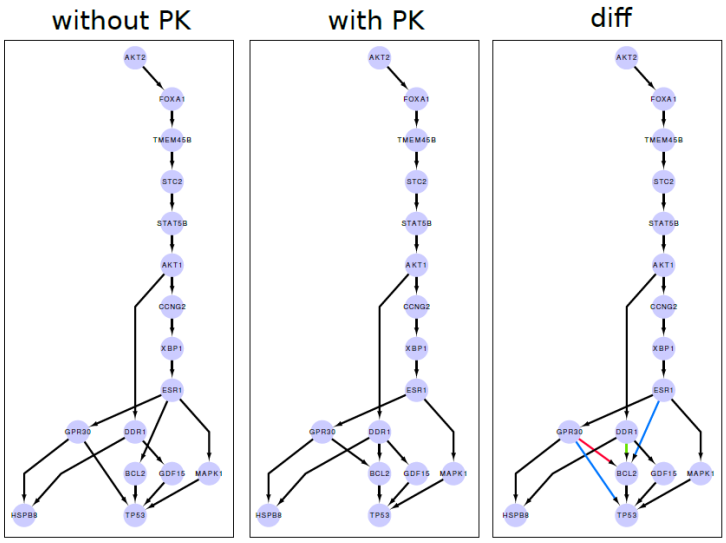

Figure 2: Transitively reduced visualization of the overlaps and differences of reconstructed breast cancer networks with and without integrated prior knowledge.

References

Bader,G.D., Cary,M.P. and Sander,C. (2006) Pathguide: a pathway resource list. Nucleic Acids Res. 34, D504–D506.

Croft,D. et al. (2010) Reactome: a database of reactions, pathways and biological processes. Nucl. Acids Res. 39:D691-7.

Demir,E. et al. (2010) The BioPAX community standard for pathway data sharing. Nature Biotechnol. 28(9):935-42.

Fröhlich,H. et al. (2009) Nested effects models for learning signaling networks from perturbation data. Biometrical Journal. 51(2):30423.

Kramer,F et al. (2013) rBiopaxParser -an R package to parse, modify and visualize BioPAX data. Bioinformatics 29(4):520-522.

Single-cell gene networks from co-expression

Megan Crow[1], Anirban Paul[1], Sara Ballouz[1], Z. Josh Huang[1], Jesse Gillis[1]

[1]Cold Spring Harbor Laboratory

Motivation: Cellular variation is critical in almost any complex biological system, making the desirability of assessing gene networks in a cell-type specific fashion self-evident. Previously, this has only been possible in either low-throughput assays or by inference from combined data where cell-type properties filter a universal network (often protein-protein interaction). With the rise of single-cell RNA-sequencing (scRNA-seq) we are now able to construct high-throughput gene networks derived from data at cellular resolution, enabling the discovery of pathways that are active in individual cells.

Background: Biology has increasingly looked to relationships between genes to explain phenotypic variability. One way to determine these groupings is from transcriptional data; genes with similar expression patterns are thought to be involved in the same cellular pathway or function. This can be generalized into co-expression networks, which are built from the assessment of gene-gene relationships across sources of variation. Co-expression networks have become an important resource in the interpretation of gene function and disease, and the comparison of networks built from diverse datasets enables characterization of condition-specific gene modules. Until recently, this type of analysis could only be performed by using bulk tissue or by pooling thousands, or even millions of cells, meaning that individual cell variation could not be assessed. This has changed with the increasing availability and cost-effectiveness of scRNA-seq. Using this technology, researchers have begun to grapple with the heterogeneity of gene expression in single cells, primarily leveraging this information to define previously unknown cell subtypes. Co-expression remains relatively uncharted territory.

Meta-analysis: We have performed the first major analysis of single-cell co-expression using data from 31 separate studies comprising 163 individual cell-types [1]. Sampling from more than 200 bulk RNA-seq datasets as a control, we assessed the functional connectivity of both bulk and single cell networks by measuring how well we could predict gene ontology (GO) functions using neighbor voting in cross-validation. We will present results demonstrating that single-cell network connectivity is less likely to overlap with known functions than co-expression derived from bulk data, and that outlier functionality is never observed in any of the single cell networks. Most interestingly, assessing single-cell data in which cell-type was held constant in each network showed little decrease in performance, suggesting that gene sets varying from cell-to-cell within a cell-type are similar to those that vary from cell-type to cell-type.

Wet-lab validation: To complement this analysis, we performed our own technically controlled scRNA-seq experiment using genetically targeted mouse interneuron classes to determine features and analysis practices that contribute to functional connectivity. Chandelier cells and parvalbumin-positive fast-spiking basket cells were prepared in known batches to generate co-expression networks for each. This allowed us to take the same meta-analytic approach we took to cross-laboratory comparison to characterization of technical properties within our data. We will present our results, showing that the use of raw or batch-corrected UMI data and post-co-expression network standardization provided the highest degree of network replicability, semantic similarity and overall functional connectivity. Importantly, we found that gene expression levels can be highly predictive of network topology, and that this can influence functional connectivity. To demonstrate this, we will walk through a re-analysis of the BrainSpan RNA-seq data, indicating that previously reported age-related variation of disease gene connectivity shows this dependency.

Conclusions: Single-cell RNA-seq data offers both promise and perils. Technical properties create confounds which can only be avoided with care and specific controls. Comparison with previous data in meta-analysis offers particular benefit for single-cell co-expression and we have made our data readily accessible. This can be used for both improving co-expression performance and for cross-laboratory comparison within a common framework.

References

[1] Crow M, Paul A, Ballouz S, Huang Z J, Gillis J. Exploiting single-cell expression to characterize co-expression replicability. Genome Biology. 2016; Accepted.

Figure legend: For our study, we profiled 128 Chandelier and Parvalbumin (Pv) basket cells using single cell RNA-sequencing. These two interneuron cell-types are both inhibitory; however, their laminar distributions within the cortex and their morphologies are markedly distinct, as demonstrated here in the immunofluorescence images (top two panels) and the Camera Lucida reconstructions (bottom two panels). L1/2/3/6 = laminae. Axons are colored blue and dendritic arbors are shown in red.

Systematic functional annotation of compound libraries through large-scale analysis of chemical-genetic interactions

Scott W. Simpkins[1,2], Justin Nelson[1,2], Jeff S. Piotrowski[3], Sheena C. Li[4], Raamesh Deshpande[2], Hamid Safizadeh[2,5], Jacqueline M. Barber[4], Reika Okamoto[4], Mami Yoshimura[4], Hiroki Okada[6], Karen Kubo[7], Tamio Saito[4], Yoshi Ohya[7], Hiroyuki Osada[4], Minoru Yoshida[4], Charles M. Boone[4,8], Chad L. Myers[1,2]

[1]University of Minnesota, Bioinformatics and Computational Biology Graduate Program, [2]University of Minnesota, Department of Computer Science and Engineering, [3]Yumanity Therapeutics, [4]RIKEN Center for Sustainable Resource Science, [5]University of Minnesota, Department of Electrical and Computer Engineering, [6]University of Pennsylvania, Department of Cell and Developmental Biology, [7]University of Tokyo, Department of Integrated Biosciences, [8]University of Toronto, Donnelly Centre for Cellular and Biomolecular Research

Current technologies allow rapid isolation or synthesis of many thousands of compounds with potentially interesting and useful biological properties. Chemical genomics is an unbiased method to discover the modes-of-action of these compounds that is complementary to current "target-centric" approaches. Target-centric approaches are inherently limited in the scope of targets they can discover, which makes genome-wide, compound-centric methods appealing for jump-starting progress in such areas as therapeutic development, toxicology screening, and biological probe discovery. To characterize large compound libraries that remain largely functionally unannotated, we developed an optimized yeast chemical genomics pipeline for high-throughput generation of chemical-genetic interaction profiles as well as data processing and interpretation. To date, we have screened more than 18,000 compounds to generate more than 15 million chemical-genetic interaction scores.

Our high-throughput yeast chemical-genomics assay leverages the yeast genetic interaction network, the power of which has remained largely untapped in the field of chemical genomics. For example, we used the genetic interaction network to effect an order of magnitude improvement in throughput by selecting a diagnostic set of 300 strains that recovers the functional information contained in the whole genome deletion collection (~5000 strains). In addition, our method for predicting a compound's modes-of-action utilizes genetic interaction profiles to interpret the chemical-genetic interaction profile[1] and aggregates the genetic interaction-based predictions into biological process predictions. Other improvements in throughput were made possible by hypersensitizing the mutant strains to compounds (less compound required) and using massively parallel sequencing instead of microarrays to measure barcode abundance.

We performed chemical-genomic screens for more than 18,000 compounds from a diversity of sources, including natural products and derivatives, compounds from combinatorial syntheses, and several hundred approved drugs.[2,3] Our large-scale chemical-genomic screen revealed insights into the biological function of compounds, at both the individual compound and compound collection levels. Interestingly, the larger and least biased collections were depleted for compounds affecting nuclear processes (e.g. DNA repair, transcription) – suggesting these targets tend to be relatively less accessible – and enriched for compounds affecting processes related to the periphery of the cell (e.g. glycosylation) (Figure 1). Additionally, we discovered molecular substructures that are enriched in sets of compounds with similar biological effects, which shows chemical genomics' ability to relate chemical structure to function. After identifying compounds that were likely to possess multiple modes-of-action, with predictions driven by distinct subsets of their chemical-genetic interaction profiles, we discovered a compound that perturbs DNA and disrupts the yeast cell wall. We confirmed the accuracy of our predictions by comparing them against known modes-of-action and performing large-scale validation experiments using orthogonal assays. Overall, we estimate that high confidence target predictions across all biological processes can be made for over 10% of the screened compounds.

References

1. Parsons, A. B. et al. Nature Biotechnology, 22, 62–9, 2004.
2. Kato, N. et al. Current Opinion in Chemical Biology, 16, 101–8, 2012
3. Drewry, D. H.; Willson, T. M.; Zuercher W. J. Current Topics in Medicinal Chemistry, 14, 340-2, 2014.

Figure 1. Functional characterization of compound collections

Distribution of functional neighborhoods perturbed by compounds across diverse collections. Heatmap pixels are yellow if the compounds in a collection predict process targets in a neighborhood at a significantly higher rate than the background, and blue for a lower rate than the background. Full names and descriptions of compound classes are as follows: "NPDepo" – RIKEN Natural Product Depository, all compounds; "NP" – NPDepo natural products; "NPD" – NPDepo natural product derivatives; "Clinical+" – clinically-relevant and other miscellaneous compounds from the NPDepo; "NCI NP" – NCI Open Chemical Repository (NCI-OCR) natural product collection; "NCI ONC" – NCI-OCR approved oncology drugs; "NCI MECH DIV" – NCI-OCR mechanistic diversity collection; "NCI STRUCT DIV" – NCI-OCR structural diversity collection; "NIHCC" – NIH Clinical Collection; "GSK KI" – GlaxoSmithKline published kinase inhibitor collection.

Cytoscape Cyberinfrastructure –Quality Network Analysis Done Quicker and Cheaper

Barry Demchak[1]

[1]UC San Diego

Summary

The Cytoscape Cyberinfrastructure (CI) combines Cytoscape and a microservices architecture to empower network biologists to create and share novel and reproducible web-based workflows quickly and cheaply. Using highly reusable, evolvable, and scalable computational and visualization services and user interface elements, it extends Cytoscape and its successful development model to Internet scale.

Abstract

Cytoscape's primary mission is to support the analysis and visualization of networks that describe biological processes and structure. As networks have become more useful and relevant over the last decade, Cytoscape usage has mushroomed – 3,000 startups per day and 14,000 downloads per month worldwide. Predictably, as network analysis has become more common, we have seen more demand for novel and reproducible network-based workflows, network sharing and collaboration, and analytics and capacity to manipulate ever larger, more complex, and more hyperlinked and multiscale networks. As a standalone workstation-based desktop application, Cytoscape is limited in its ability to address these opportunities due to limited memory and CPU resources (e.g., processing Network-based GWAS and Stratification), and is often difficult and time consuming to customize.

In the past year, we have answered this challenge by developing the Cytoscape Cyberinfrastructure (CI), an Internet-scale distributed system based on Microservices [1]. The CI's mission is to enable and encourage network biologists to create and deploy high quality, innovative and scalable network-based computation, collaboration and visualization services using the tool chain of their choice. The CI links services via a light weight REST-based aspect-oriented interchange protocol (called CX), which enables tailored data streams while supporting service innovation via evolvable standards. The CI service broker ("Elsa") enables biologists to write and publish biologically relevant services (e.g., enrichment calculations or heat propagation) comparatively easily and without concern for scaling or tracking job status. It provides the CyREST [2] interface layer to expose Cytoscape's rich functionality as services, and it provides a framework ("CytoWidgets") to enable the creation, delivery, reuse, and evolution of Javascript-based user interface modules that package service functionality within both web-based and desktop-based applications.

We demonstrate the CI in action by showing how applications (written in Java, IPython, R, and Javascript) call useful production services (written in the Java, Go, and Python

languages) to execute real world workflows. Particularly, we preview the NDEx Valet CytoWidget embedded within Cytoscape to provide a state of the art chooser and importer for shared NDEx-based networks [3]. We also show how to quickly and cheaply new and novel web-based, network-oriented workflows can be built using NDEx Valet and network display and manipulation widgets. Finally, we demonstrate how the CI brokers long-running, CPU and memory intensive services.

Future work includes adding clustering, analysis, layout, publishing and display microservices, interfaces to Galaxy and Taverna workflows, and support for the Publishing of the Future initiative.

References

[1] http://martinfowler.com/articles/microservices.html
[2] K. Ono et al. CyREST: Turbocharging Cytoscape Access for External Tools via a RESTful API. F1000 Research. 2015 Aug 5;4:478.
[3] D. Pratt et al. NDEx, the Network Data Exchange. Cell Systems 1, no. 4 (n.d.): 302–5

Systematic Functional Annotation of the 2016 Yeast Genetic Interaction Network

Anastasia Baryshnikova[1]

[1]Princeton University

Large-scale biological networks represent relationships between genes, but our understanding of how networks are functionally organized is limited. Here, I describe Spatial Analysis of Functional Enrichment (SAFE), a systematic method for annotating biological networks and examining their functional organization. SAFE visualizes the network in two-dimensional space and measures the continuous distribution of functional enrichment across local neighborhoods, producing a list of the associated functions and a map of their relative positioning. I applied SAFE to annotate the most recent genetic interaction network from budding yeast Saccharomyces cerevisiae, which was derived from the quantitative growth analysis of ~20 million double mutants. By annotating the network with GO biological process terms, SAFE proved to be accurate, robust to noise and sensitive to biological signal, while taking less than 1% of the time compared to alternative approaches. Furthermore, integration of genetic interaction and chemical genomics data using SAFE revealed a link between vesicle-mediate transport and resistance to the anti-cancer drug bortezomib. These results demonstrate the utility of SAFE for examining biological networks and understanding their functional organization.

Preprint available at http://biorxiv.org/content/early/2016/02/11/030551.

netDx: An algorithm to classify patients based on similarity networks

Shraddha Pai[1], Shirley Hui[1], Ruth Isserlin[1], Hussam Kaka[1], Gary D Bader[1]

[1]University of Toronto

Patient classification has widespread biomedical applications in clinical management, including tumour subtyping in cancer and the pharmacogenetic prediction of drug response. A general purpose, clinically useful prediction algorithm should be accurate, generalizable, be able to integrate diverse data types (e.g. clinical, genomic, metabolomic, imaging), handle sparse data and be intuitive to interpret. We describe netDx, a supervised patient classification framework based on patient similarity networks, that meets the above criteria. netDx models data as patient networks and uses the GeneMANIA machine learning algorithm[1] for network integration and feature selection. The results of the classifier are clinically and biologically interpretable, and can be visualized as an integrated patient similarity network. We will present two applications of netDx. In the first, we use a published gene expression dataset of medulloblastoma subtypes (N=103 tumours) and build a 4-class predictor for tumour subtype[2]. In the second, we predict case/control status in autism spectrum disorders based on the occurrence of rare copy number deletions in cellular pathways (N=3,291 patients)[3]; this predictor achieved better performance than previously published methods. Software will be made available as open-source R packages. Near-term extensions involve networks based on genotype similarity from dense SNP genotyping.

References: 1. Mostafavi S et al. (2008). Genome Biol. 9:S4 2. Northcott PA et al (2011) J Clin Oncol. 29 (11):1408. 3. Pinto et al. (2014). Am J Hum Gen. 94 (5):677.

Accepted Poster Presentations

Maximizing accuracy in pairwise and multiple network alignment via both node and edge conservation

Vipin Vijayan[1], Vikram Saraph[2], Tijana Milenkovic[1]

[1]University of Notre Dame, [2]Brown University

Analogous to genomic sequence alignment, biological network alignment (NA) aims to find regions of similarities between molecular networks of different species, in order to allow for transfer of biological knowledge from well- to poorly-studied species between conserved (aligned) network regions. Many existing NA methods aim to maximize total similarity over all aligned nodes (or node conservation). Then, they evaluate alignment quality by measuring the amount of conserved edges (or equivalently, the size of the common subgraph), but only after the alignment is constructed. Thus, we recently introduced MAGNA to directly optimize edge conservation while the alignment is constructed, without decreasing the quality of node mapping. In systematic evaluations against state-of-the-art methods MAGNA outperformed the existing methods on multiple data sets, in terms of both node and edge conservation as well as both topological and functional alignment accuracy. Even more recently, we developed a new MAGNA++ framework to further improve alignment quality by simultaneously optimizing both edge conservation and node conservation during the alignment constructions, which indeed yielded better alignments compared to maximizing only node conservation (as the existing methods do) or only edge conservation (as MAGNA does). The MAGNA++ framework is publically available as open source, with a friendly graphical user interface (http://nd.edu/~cone/MAGNA++/).

MAGNA++ is a state-of-the-art pairwise NA (PNA) method. PNA finds similar regions between two networks while multiple NA (MNA) can align more than two networks. MNA is believed to be more useful since it can capture at once biological knowledge common to multiple networks. Therefore, we introduce multiMAGNA++, an extension of MAGNA++ from PNA to MNA. MultiMAGNA++ generally outperforms or is on par with the existing MNA methods (Figure 1), while often completing faster than the existing methods. During method evaluation, we also introduce new topological and functional MNA quality measures to allow for more complete alignment characterization and more fair MNA method evaluation compared to using only the existing alignment quality measures.

Modification of human adipocyte metabolism for the prevention of type 2 diabetes.

Thierry Chénard[1], André Carpentier[1], André Tchernof[2], Rafael Najmanovich[1]

[1]Université de Sherbrooke, [2]Université Laval

Insulin resistance and alteration of glucose- induced insulin secretion are the two main physiopathologic pathophysiologic characteristics of type two diabetes (T2D). Insulin resistance precedes the development of type 2 diabetes and deficiency in glucose-induced insulin secretion is considered a sine qua non condition for the presence of T2D. Indeed, normal pancreatic β cell response can compensate for insulin resistance enabling the maintenance of normal blood glucose levels. Experimental data in both humans and animal models support that over-exposition exposure of lean tissues to fatty acids (FA) plays an important role in the development of insulin resistance and pancreatic dysfunction. Adipose tissues play a central role in the regulation of circulating FA fluxes. FA storage dysfunction in adipose tissues leads to elevated levels of circulating FA fluxes early in the development of T2D. Two general mechanisms govern adipose tissue expansion needed for the storage of excess FA: 1) hyperplasia (augmentation of pre-adipocyte recruitment, leading to more numerousincreased numbers of smaller adipocytes); 2) hypertrophy (augmentation increasedof adipocyte size leading to fewer, bigger larger adipocytes). A default in hyperplasia, diminished pre-adipocyte recruitment and differentiation into mature adipocytes leads to a pathologic hypertrophic expansion contributing to the development of T2D.

To analyse these mechanisms, which could influence adipose tissue expansion and T2D development, we developed a human adipocyte metabolic network by correcting errors present inimproving a published metabolic network. We used this updated network to identify genes which have the potential to influence expansion mechanisms of adipose tissues (hypertrophy and/or hyperplasia).

We did performed an in-silico gene deletion analysis on our human adipocyte network, iTC1390adip, using flux balance analysis. This analysis enabled us to predict the effect of gene deletion on the optimal production of biomass and of lipid droplets as a representation of hypertrophy and hyperplasia and hypertrophy respectively. We compared our predictions on the effect of gene deletions on the optimal biomass production of biomass to published experimental gene deletions realized in various different cancerous cell lines to evaluate the predictive capacities of our network as well asand establish a threshold for the production of biomass that givesgenerating the best concordance between our simulations and the experimental results. Following the gene deletion analysis, we predicted identified 26 genes as having the potential to reduce lipid droplet production with a limited effect on the production of biomass and which could have a positive effect on adipocyte hypertrophy. Some of the identified genes, such as LCAT and DGAT1, possess experimental results supporting our predictions while others, such as FAR2 and HSD17B12, could serve as new potential targets in the remodelling

of adipose tissues and in the treatment of T2D if their role is experimentally validated. We also used gene expression data from subcutaneous and visceral adipose tissues to identify metabolic pathways where gene expression is significatively significantly different between the two depots and which could be related to the varying importance of those depots in the development of T2D. These results will guide new in-vivo and in-cellulo studies to validate new therapeutic targets which could affect T2D development.

Figure 1: The effect of the deletion of each individual gene in the network on both biomass and lipid droplet production with each cross representing a separate gene. The imbedded graph represents the specificity and selectivity of our predictions compared to experimental data at varying thresholds of biomass productions. We selected the 80% of Wild type (WT) biomass production as the minimal value for which we consider genes as being interesting targets.

The Open PHACTS Discovery Platform from a network biology perspective.

Chris Evelo[1]

[1]Maastricht University and Open PHACTS

Introduction. Typically network biologists will think of a biological system as being composed of biological entity nodes connected over known relationships forming the edges. Semantic web scientists on the other hand, try to describe graphs of linked statements typically consisting of object-predicate-object relationships. Of course objects in such graphs can be considered nodes in a network and predicates can be considered edges, the two fields are thus very much related. Yet, the technologies applied are different and the two communities only show small overlap in persons and activities. The Open PHACTS Discovery Platform is a semantic web based knowledge resource for drug discovery, and is thought to be potentially useful for network biology applications.

The Open PHACTS project (Williams et al., 2012) developed guidelines for the production of RDF with explicit descriptors and provenance, it also supported the development of RDF by collaborating data providers. The RDF was loaded into a triple store that can be accessed via an API (Gray et al., 2014). All the loaded data is also available for download. In the project it became clear that well-formed RDF makes data linkable, but not necessarily linked. This was solved by adding three mapping services: for identifiers (based on BridgeDb (van Iersel et al., 2010)), for textual synonyms for concepts and one chemical (sub)structure resolution. This is only in part because different resources use different identifiers, terms from different ontologies or different synonyms. In principle that problem could be solved by hard coding the mappings in the graph itself. That would allow storing the whole mapped graph in a network resource like Neo4J. That approach is in fact used by some companies. However, it turned out that the mappings depend on the scientific questions asked. Stereoisomers matter, but not always, proteins are not genes, but for some applications they can be considered the same. That led to the concept of scientific lenses (Batchelor et al., 2014). The value of the Open PHACTS Discovery Platform was best shown by solving real drug discovery research questions using a.o. pipeline tools (Ratnam et al., 2014) and by its use for in silico target validation of cellular phenotypic screens (Digles et al., 2016).

Opportunities. You could say that any application of the Open PHACTS discovery platform is a network biology application. However, for this NetBioSIG it is interesting to know how you could use it in common network biology tools like Cytoscape. Creating a full, hardcoded graph like mentioned above, would be one option. Such a graph could for instance be uploaded to NDEX. This would use the mapping tools but not their flexibility. Using the available RDF in a triple store with the SPARQL app for Cytoscape is another option. But this needs separate mapping solutions, e.g. by using the Cytoscape BridgeDb app, which should then need to be updated to the new code base that allows

mapping of semantic URI's. Both approaches do not optimally benefit from the Open PHACTS architecture. Addition of drug-target relationships to networks is possible using CyTargetLinker. Interaction networks for CyTargetLinker can be created by using the Open PHACTS API with specific lens settings. Alternatively the Open PHACTS API could be accesses directly by a new Cytoscape app. Maastricht Science students did some development work on such a plugin and even applied it to some specific use cases.

Importance. The Open PHACTS discovery platform helped improve the quality of the underlying data. This wealth of data itself would be a big asset when easily available for network tools. Open PHACTS also helped us learn important lessons about dynamic mapping. Connecting the platform to network tools might help to connect the two scientific communities, which would allow us to more regularly combine the high quality data and data descriptors from the semantic web community with the flexible and powerful analysis approaches common in the network biology field.

Acknowledgements. I would like to thank the many partners and associated partners of the Open PHACTS project and from its successor the Open PHACTS foundation, without whom this resource would not be real. The Open PHACTS project received funding form Innovative Medicines Initiative Joint Undertaking under grant agreement n° 115191.

References

Batchelor, C. et al. (2014) The Semantic Web–ISWC 2014, 98. http://dx.doi.org/10.1007/978-3-319-11964-9_7
Digles, D. et al. (2016) MedChemComm In press.
Gray, AJG. et al. (2014) Semantic Web Journal 5:2. 101. http://dx.doi.org/10.3233/SW-2012-0088
Ratnam, J. et al. (2014) PLOS ONE 18:9. e115460. http://dx.doi.org/10.1371/journal.pone.0115460
van Iersel, MP. et al. (2010) BMC Bioinformatics 11:5. http://dx.doi.org/10.1186/1471-2105-11-5
Williams, AJ. et al. (2012) Drug Discovery Today 71:21/22. 1188. http://dx.doi.org/10.1016/j.drudis.2012.05.016

An enhanced map of the human Methyl Arginine Proteome

erik verschueren[1], rebecca pferdehirt[1], florian gnad[1], don kirkpatrick[1], jennie lill[1]

[1]Genentech

We present a systematic and comprehensive study to identify all enzymatic and regulatory protein-protein interactions of human Protein Arginine Methyl Transferases (PRMT). Methyl-Arginine is understudied as a protein post-translational modification but has been shown to be nevertheless widely occurring. From a disease perspective, deregulation of methylated Arginines has been primarily associated with cancer but is likely implicated in the pathogenesis of several different maladies. To compile a complete intracellular interaction network of the human PRMT family we systematically performed affinity purification and proximity labeling assays using multiple cell lines and affinity tags for all 9 human PRMTs, identified their binding partners through mass spectrometry and employed specialized scoring algorithms to prioritize biologically relevant interacting partners. We then integrated our protein interaction network with a vast catalogue of methylated Arginines comprising all three types: mono-methylated as well as symmetric and asymmetric di-methylated. Our network model provides the first comprehensive map of the human Arginine methylome to date and describes a number of modular sub-networks enriched for binding partners with oncogenic associations that could be therapeutically relevant.

Integrative approach in drugs discovery pipeline applied to Clostridium difficile.

Mathieu Larocque[1], Louis-Charles Fortier[1], Rafael Najmanovich[1]

[1]Université de Sherbrooke

Introduction:

Clostridium difficile is a nosocomial anaerobic pathogen of the intestinal flora. This pathogen produces two toxins which can cause diarrhea, pseudomembranous colitis and even death in 10-15% of the cases. In order to infect, the pathogen requires a perturbation of the intestinal flora which is usually linked to antibiotic treatment. It is the leading cause of diarrhea associated with this kind of medication. C. difficile is highly problematic in health care services where the prevalence of bacterial spores and the number of susceptible hosts is high. The spores allow C. difficile to survive commonly used disinfection and cleaning agents, making patient to patient transmission easier.

This pathogen is associated with high relapse rates: 24% for vancomycin and 12% for fidaxomycin, the two most common antibiotics used to treat C. difficile infections. This is highly preoccupying as each relapse brings about more severe symptoms. Due to all these factors, C. difficile is, as mentioned in the latest report of the CDC (Centers for Disease Control and prevention of the United States), « an immediate public health threat that requires urgent and aggressive action».

Hypothesis:

To answer this threat, we elaborated a new drug discovery pipeline solely revolving around the issues and problematics of treating C.difficile (figure 1). This integrative approach uses system biology (to identify potential targets based on the metabolism of the bacterium), molecular docking (to find potential binders of these targets) and experimental work (to validate both targets and binders) in order to identify pharmaceutical leads with higher specificity and targeting important steps of the pathogenesis of C. difficile.

Results:

In this work we created the first manually curated and validated metabolic network of C. difficile, called iMLTC806cdf [1]. This publicly available tool (BioMODELID:1407050000) contains 806 genes and enzymes catalyzing 911 reactions involving 600 unique metabolites and is the most adequate approximation of the metabolism of the bacterium currently available. The network was used to identify 163 essential genes under different conditions, resulting in the same number of potential therapeutic targets. Part of these results were validated as we achieved a precision of 89% for prediction of gene essentiality on rich media when compared to an experimental high-throughput mutagenesis study performed in similar conditions by Dembek et al. [2].

The pharmaceutical potential of all the targets was evaluated. Based on this potential, NadA, an enzyme involved in the biosynthesis of NAD+, an important energetic cofactor with potential implication in sporulation and low cross-reactivity probability, was selected as target of higher interest. This gene is predicted to be essential only in the absence of NAD+, which facilitate its study experimentally and, we think, reflects the condition most likely found during the infection. The conditional essentiality of NadA was validated experimentally by growth curve experiments in vitro, and validation of its importance in-vivo is currently underway using a mouse model of C. difficile infection.

An iterative high-throughput virtual screening method is currently being used on a homology model of the protein NadA in order to identify potential binders. This molecular docking protocol starts with the EGFL fragments library and gets refined with more complex molecules as compounds similar to the best scoring chemical groups are identified in the ZINC databases and become the starting points of future iterations. We will determine experimentally the binding affinity and effect on C. difficile and Escherichia coli (to evaluate the specificity) of the molecules with an adequate binding mode (determined arbitrarily).

Methods:

Information from the KEGG, Bio-cyc and Transport DB databases was used as a starting point of iMLTC806cdf. Validations of the network included removal of essential metabolites, simulation of different growth media and the utilization of different carbon sources. Flux balance analysis (FBA) and synthetic accessibility (SA) were used to analyze the network, considering the two approaches as complementary.

A wide array of data such as differential expression under different conditions (sporulation, germination, in-vivo, etc.), essentiality for sporulation on rich medium, sequence, function and domain homology was used to determine pharmaceutical potential of each target. Mutant strains of C. difficile were generated using the insertion of a type II intron protocol (ClosTron). I-tasser was used to create the homology model of NadA. Molecular docking was performed using FlexAid [3]; an algorithm developed inside our group.

References:

[1] Larocque M. et al. BMC Syst Biol. (2014) 8:117
[2] Dembek M. et al. MBio. (2015) 6(2):e02383.
[3] Gaudreault F et al. J Chem Inf Model. (2015) 55(7):1323-36.

A Minimal Vocabulary for Inter-Network Hyperlinks

Dexter Pratt[1]

[1]UCSD

We propose for a simple, flexible scheme for the hyperlinking of network models in a way that leverages the NDEx and Cytoscape infrastructure and demonstrate example use cases. The expressive power gained when network representations of knowledge can be organized in larger structures has long been understood. Global schemes for hyperlinked knowledge have been proposed as far back as the Xanadu project and have been broadly deployed with the rise of semantic web technologies. Tactical schemes are embodied in many network systems, such as Cytoscape, that enable networks to contain and organize subnetworks. There are, however, unmet needs in the biological community to more effectively and easily organize networks and ontologies, especially in cases where multiple individuals and algorithms may need to share and reuse content. We propose an approach in which biological networks with stable URIs (such as those stored in NDEx) may reference each other using a minimal vocabulary of network, node, and edge attributes, extending standard document relationship ontologies. As with the universally familiar html links between web pages, loose coupling of networks and their elements will be easy to understand and easy to use in applications, especially for browsing and visualization. Authors of diverse networks, ranging from mechanistic models, gene association networks, or data-driven ontologies, can reference each other, linking a node in one network to a pathway network or linking an edge representing a cellular transformation to an associated transcriptional network. We present examples in which networks stored in NDEx are linked to each other and where a user can browse between them in a visualization web application. In summary, we assert that a minimal scheme that is accessible to non-computational users and which encourages modular re-use of existing resources will be effective and will serve an important role alongside more sophisticated mechanisms of structuring inter-network relationships.

Extending the rBiopaxParser

Frank Kramer[1], Florian Auer[1], Tim Beissbarth[1]

[1]University Medical Center Goettingen

Keywords: BioPAX, Pathways, Ontologies

In the past years ontologies have been the tool of choice to represent and allow the sharing of knowledge of this biological reality. BioPAX is a commonly used ontology for the encoding of regulatory pathways. The R Project for Statistical Computing is the standard environment for statistical analyses of high-dimensional data and networks. A number of packages are available that provide the pathway data of databases like KEGG[1], the Pathway Interaction Database[2] or Reactome[3] as.

Our open-source package rBiopaxParser[4] parses BioPAX-Ontologies and represents them in R, allowing merging, editing and exporting of BioPAX-based pathway data within R. Class definitions, properties and restrictions of the ontology are mapped on a 1:1 basis, with respect to the limitations of object-orientation of R. The user is able to parse arbitrary BioPAX OWL files, for example the exports of popular online pathway databases like PID, Reactome or KEGG. Instances of BioPAX-Classes can be programatically added or removed. The interactions within pathways can be visualized using R plotting functions.

Here, we present our new release of the rBiopaxParser, which extends its functionality regarding visualization and interoperability. Pathways can now be displayed in a mechanistic fashion as well as an interaction graph. Furthermore, new interfaces to Cytoscape and the browser-based Cytoscape.js allow for a better visualization and editing of the pathway representations.

The rBiopaxParser package is available at Bioconductor:
https://bioconductor.org/packages/release/bioc/html/rBiopaxParser.html

This work is funded by the German Ministry of Education and Research (BMBF 01ZX1508, BMBF 031L0024A).

References

[1] Kanehisa M, Goto S, Kawashima S, et al (2004) The KEGG resource for deciphering the genome. Nucl Acids Res 32:D277–D280. doi: 10.1093/nar/gkh063
[2] Schaefer CF, Anthony K, Krupa S, et al (2009) PID: the Pathway Interaction Database. Nucl Acids Res 37:D674–D679. doi: 10.1093/nar/gkn653
[3] Joshi-Tope G, Gillespie M, Vastrik I, et al (2005) Reactome: a knowledgebase of biological pathways. Nucl Acids Res 33:D428–D432. doi: 10.1093/nar/gki072
[4] Kramer F, Bayerlová M, Klemm F, et al (2013) rBiopaxParser—an R package to parse, modify and visualize BioPAX data. Bioinformatics 29:520–522. doi: 10.1093/bioinformatics/bts710
[5] Shannon P, Markiel A, Ozier O, et al (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res 13:2498–2504. doi: 10.1101/gr.1239303
[6] Shannon P, Grimes M, Kutlu B, et al (2013) RCytoscape: tools for exploratory network analysis. BMC Bioinformatics 14:217. doi: 10.1186/1471-2105-14-217

The Noctua Modeling Tool

Seth Carbon[1], Heiko Dietze[1], Chris Mungall[1]

[1]Lawrence Berkeley National Laboratory

Website: http://noctua.berkeleybop.org
Repository: https://github.com/geneontology/noctua

[Detail screenshot of activity editing in Noctua.]

We present Noctua, a modern web application and stack for modeling complex biology. Noctua directly models information as a graph, escaping many of the pitfalls of more "tabular" modeling. Noctua also presents a rich, interactive, and collaborative user interface, as well as a complete set of tooling for data extraction and integration.

The Gene Ontology [1] project aims to create a comprehensive and up-to-date model of biological systems based on annotating knowledge graphs (ontologies) with curated and generated information. Over its existence, the project has aimed to capture this information with ever increasing richness and specificity. The traditional storage format for GO annotations, the tabular GAF, has undergone several iterations, but can no longer support the types of annotations required by current use cases. To move forward, the project has adopted LEGO, a graph-base abstraction for modeling biology. The Noctua application and stack was created to annotate information with the LEGO abstraction, but has the built-in flexibility to be used in any number of pathway or workflow models.

The current main end user client application for Noctua presents a real-time collaborative graph editing environment, allowing users to assemble graphs representing biological knowledge, including aspects such as references and evidence. The user environment uses typed inputs combined with a click, drag, and connect interface for easy and intuitive graph editing. As multiple users work on the same model, no matter the client, it updates the common environment in real time, allowing for easy discussion, presentation, and collaboration.

The Noctua stack is composed of three layers: the client, written in JavaScript using the jsPlumb and AmiGO/BBOP libraries; the communication layer, written in JavaScript and providing client-to-client and client-to-server communication; and the graph engine, a Java server that uses OWL to model and OWL universe tools for reasoning. This stack is strongly separated, respecting protocol and common patterns. For example, while the main user interface is a graph editor, it could be easily replaced by a different client that could speak the same wire protocol–the flexibility of the framework allows for the easy creation of alternate clients, such as the form or REPL based ones.

The Noctua stack is being actively used by the Gene Ontology, with the produced annotations finding their way back into the pipelines of several model organism databases. As well, Noctua data produced for the GO is loaded into AmiGO [2], where it

is available for exploration and made available to be consumed in clients via our JavaScript API. The data modeling and retrieval systems for the client are available as separate packages, allowing third parties to create their own clients or embed annotation widgets in their own resource pages.

References

[1] Gene Ontology Consortium: going forward. Nucl. Acids Res. 43, D1049D1056.
[2] AmiGO: online access to ontology and annotation data. Bioinformatics 25, 288289.

A high-performance pipeline for genome-wide network reconstruction from gene expression data

SON LE[2], Alberto Riva[1], David Tran[2]

[1]University of Florida - ICBR Bioinformatics, [2]Department of Neurosurgery, University of Florida, [3]University of Florida

Understanding the global biological implications of genomic changes requires large-scale gene expression profiling, as it provides information on cellular behaviors caused by driver mutations and their interactions with the microenvironment. Gene expression profiles can then be used to infer the global and local regulatory networks that regulate these behaviors, using reverse engineering tools designed to scale up to the complexity of mammalian cells. One example of such tools is ARACNe, a program to reconstruct regulatory networks on the basis of the Mutual Information (MI) between gene expression profiles [1]. ARACNe can be applied to a small set of genes (e.g., to find all possible targets of a transcription factor) or in genome-wide mode, to identify upstream master regulators in an unbiased way. In this case, its authors recommend running ARACNe on a cluster and employing a bootstrap procedure to make the analysis computationally tractable.

We have developed a collection of computational tools to facilitate and streamline the use of ARACNe in this fashion, making it suitable for large-scale work in a high-performance computing setting. The tools, implemented in a Python program called APPLE (ARACNe Processing PipeLine Extensions), consist of the following nine commands: random, bootstrap, consensus, filter, histogram, extract, convert, stats, translate. The analysis process is outlined in Figure 1. Given an initial database of expression values, the random command generates a number of randomized datasets by shuffling the expressions values within each row. All datasets (the single real one and the shuffled ones) then undergo a bootstrap procedure, implemented by the bootstrap command, with a user-specified number of bootstrap rounds and sample size.

The bootstrap files are then processed by ARACNe (in parallel in a cluster environment), producing one ADJ file for each. An ADJ file consists of one line for each hub gene, and lists all hub-gene connections with the respective MI. The resulting ADJ files are combined into a single ADJ file using the consensus command. This command writes an edge to the combined file if it was observed in more than S bootstrap files (where S is a user-specified minimum value) and on the basis of its False Positive rate (FPR). This command also writes two additional files, useful for subsequent analysis steps: a counts file (recording the number of edges having each support level) and a statistics file, that records for each edge its support, FPR, and sum of the mutual information of the supporting edges.

Using the counts and statistics files, the user can filter the resulting ADJ file for the real datasets in order to retain only the edges that show support and MI significantly higher

than those in the shuffled datasets. This is accomplished using the histogram command, which produces a histogram of all the MI values in an ADJ file. By comparing the histograms for the real and the shuffled datasets, an optimal MI value that maximizes the separation between the real and the shuffled datasets can be determined. This MI value can then be used to generate a new ADJ file containing only the edges with an MI value over this threshold, using the filter command. The process can then be repeated using the sum of the MI values for all edges connected to a hub gene.

The remaining commands provide utilities to print general statistics on one or more ADJ files, extract the edges for a specified set of genes, translate gene identifiers (e.g., from Ensembl to NCBI identifiers), and convert ADJ files to different formats for visualization, including Cytoscape format.

We applied the above pipeline to the publicly available datasets for breast cancer and glioblastoma from the Cancer Genome Atlas (TCGA) to establish reference gene networks in these two cancer types. To identify master network drivers in brain cancer stem cells, we compared the gene networks of brain cancer stem cells with brain cancer differentiated cells, using published data. Similar to recent reports, we found that OLIG2, MYT1L, ASCL1, SOX2 are core master drivers of brain cancer stem cells. For instance, mouse fibroblasts were shown to be converted into neuronal cells by the overexpression of BRN2, ASCL1 and MYT1L, two of which (ASCL1, MYT1L) are in our master driver lists. In addition, differentiated GBM cells were reported to be reprogrammed into stem-like tumor propagating cells by forced expression of POU3F2, SOX2, SALL2 and OLIG2. OLIG2 and SOX2 are in our list of master drivers. In mammary stem cells, we determined master stem cell drivers by comparing gene networks of normal mammary stem cells with differentiated mammary cells and identified, among others, beta-catenin, a well-known master driver of mammary stem cells.

References

[1] Margolin AA et al. ARACNe: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics. 2006 Suppl 1:S7.

Modelling temporal transcriptional regulatory programs activation over reconstructed gene regulatory networks describing cell fate transitions

marco mendoza[1], Julien Moehlin[1], Pierre-etienne Cholley[1], Hinrich Gronemeyer[1]

[1]IGBMC, [2]Université de Sherbrooke

Studying living organisms as an ensemble of components in which the whole is the consequence of the complexity of their interactions represents the biggest challenge of the current “big-data omics” era. Specifically since the release of the first draft of the human genome in 2001, followed by the rapid development of the massive parallel sequencing technologies, the avenue towards the analysis of genome functions from a holistic point of view has been opened. Importantly, the combination of a multiplicity of genomic readouts will provide means to describe living systems through the reconstitution of their genomic-regulatory functions which are at the basis of their defined state. Moreover, understanding the reorganisation of their regulatory wires – as a consequence of external/internal cues – represents a new approach to interpret the acquisition of novel physiological or aberrant system states. In a cellular context, the detailed comprehension of these reorganisations, known as cell fate transitions, is a major component of the novel therapeutic developments in regenerative medicine.

In our laboratory, we take advantage of retinoic acid (RA)-driven cell differentiation model systems as a systematic approach for enhancing our understanding of cell fate transitions. Specifically it is based on (i) the integration of temporal functional genomic readouts for the reconstruction of gene regulatory networks (GRNs) describing the regulatory principles taken place during the cell fate transition process; (ii) the use of computational approaches for modelling signal transduction propagation over the reconstructed GRN; and finally (iii) the validation of the prediction readouts by taken advantage of the current genome editing approaches (CRISPR).

In this context, we will present in this meeting our implemented signal transduction model able to first verify the coherence of the reconstructed GRN with the temporal transcriptional information describing the cell fate transition; then to allow to predict the capacity of any nodes composing the GRN to reconstitute the expected cell-fate transition behaviour. Specifically this methodology mimics signalling propagation over multiple temporal transcriptional response layers only in cases in which the interconnected nodes complies with (i) a change in the transcriptional state of the evaluated time-point; (ii) a coherent temporal directionality in the context of TF-TG interactions; and (iii) a signal propagation interconnections derived only from the initial signalling cue. Importantly, the presented approach has been challenged over large number of nodes, and complex interconnected GRNs, such that its performance is at the level of its expectative. Following the strong prediction power - verified by the use of CRISPR-dCas9 activation assays - observed in our RA-driven cell differentiation studies,

we are currently working in a second version able to incorporate in-silico knock-out modelling, but also combinatorial nodes activation assays in order to predict cooperative situations able to enhance signal transduction performance. Taken in consideration the interest of the scientific community to access to this type of signal transduction modelling instruments, we are currently preparing a Cytoscape app, such that users could benefit of the multiple options available in such environment.

Network Visualization Tool as a Collection of Components and Services

Keiichiro Ono[1], Barry Demchak[1]

[1]UC, San Diego

Creating an effective biological network visualization requires a set of sophisticated computing tools to integrate, analyze and visualize complex data sets. Biologists must integrate heterogeneous types of data from high-throughput omics experiments and use multiple visualization techniques to create visualizations to gain insights into the nature of biological systems.

There are numerous popular tools that help perform this task and Cytoscape is one of the de-facto standard platforms in network biology community. It is used in all of the basic steps in network analysis and visualization. In addition, researchers can implement their own algorithms as extensions, called Apps, on top of Cytoscape's core components if they need custom data visualization tools to fulfill their needs.

However, the challenge is the gap between the research code development and the production-level bioinformatics software for general audience, which includes biologists with average computing skills. In most cases, implementations of new bioinformatics algorithms and methods are focusing on the accuracy and completeness of the feature, not on their usability for general biologists. If a tool's deployment procedure is too difficult for non-professional software developers, it ends up with an excellent algorithm or methods no biologists can use except the authors. Also, traditional network visualization applications, including Cytoscape, are written in programming languages such as Java or C++, which are not the most productive tools for researchers to implement their new methods. Researchers prefer to use high-level programming languages such as R or Python that are popular in scientific computing community. For example, if they need to make an App for Cytoscape, they have to rewrite or wrap their existing code to be available as a Java application. To solve these issues, we need a new software architecture to deliver state-of-the-art bioinformatics methods written by researchers to other biologists who want to use them in their research projects.

In this presentation, we will demonstrate how the new application architecture, called Cytoscape Cyberinfrastructure (CI), changes the software development process of complex network visualization tools for modern computing environment, and shows how researchers can publish their existing research code as reusable services to make them building blocks for complex data visualization applications. Cytoscape CI is a collection of small tools, or services written by researchers in their choice of programming languages and visualization components for both web browsers and desktop applications, and we will use our existing web application, NeXO/AtgO ontology browser as an example target for future updates adopting this new architecture.