

ECCB 2014 Accepted Posters with Abstracts

D: Computational systems biology

D03: Alicia Amadoz, Patricia Sebastián-León, Francisco Salavert and Joaquín Dopazo. PATHiPRED: prediction models using the activation status of stimulus-response signaling circuits.

Abstract: Signaling pathways provide a formal representation of the processes by which the cell triggers actions in response to particular stimulus through a network of intermediate gene products. Consequently, sub-networks corresponding to stimulus-response circuits can directly be related to cell functionalities. Some recent pathway topology based methods focus particularly on the estimation of the activity of stimulus-response signaling circuits from gene expression data [1,2]. Here, we present PATHiPRED web tool that takes advantage of changes detected in the activity of stimulus-response signaling circuits for predictive purposes.

The activation status of elementary components of signaling pathways are rich-informative biomarkers that can be linked to specific cell functionalities and provide mechanistic explanations for the molecular basis of complex traits. A major challenge in the diagnosis and treatment of complex diseases is to identify relevant alterations in the biological pathway activities and discover their relationship to the disease. Using stimulus-response signaling circuits to distinguish between two classes, such as control and disease samples, or continuous variables, such as drug activity, would improve our understanding of complex phenotypes. PATHiPRED uses the probability of activation of stimulus-response signaling circuits obtained with PATHiWAYS methodology [2] to compute a pathway-based classifier for either two classes or continuous variables. Prediction models are obtained with a method based on SVM with cross-validation and it can be used to predict new datasets within the web application (<http://pathiways.babelomics.org>).

We found that the performance of the classification method using mechanism-based biomarkers was accurate and also that the suggested molecular mechanisms were reported in previous studies.

References

- [1] Nam, S. and Park, T. (2012). Pathway-based evaluation in early onset colorectal cancer suggests focal adhesion and immunosuppression along with epithelial-mesenchymal transition, PLoS ONE, 7(4):e31685.
- [2] Sebastián-León, P., Carbonell, J., Salavert, F., Sanchez, R., Medina, I., and Dopazo, J. (2013). Inferring the functional effect of gene expression changes in signaling pathways. Nucleic Acids Research, 41(W1):W213-W217.

D04: Liliana Ironi and Diana X Tran. Model-based design of synthetic networks

Abstract: In designing gene regulatory networks (GRN), in silico approaches are a must before costly in vitro experiments. However, synthetic biology still lacks a reliable tool for computer-aided design of GRNs, one that can reveal the full range of nonlinear dynamic behaviors in a single run. Here, we propose a network design cycle that utilizes both a qualitative simulator of GRNs modeled by a class of ODE equations and the intrinsic stochasticity of regulation to ultimately design a network that exhibits a specified desired behavior with the highest probability.

Network design is made easier by our simulator as the responses to perturbations (eg.

modifications in gene number or connectivity) are quickly and rigorously calculated, yet clear and easy to interpret. Also, the initial state conditions and parameter space can be conveniently explored by simple declarations between model parameters, mathematically expressed by symbolic inequalities.

Thus, model-based network design becomes a development cycle that consists of the following phases:

1. Hypothesized networks. Plausible GRN network structures are conceptualized from preexisting functional modules or created ex novo with the goal of exhibiting a desired dynamic behavior.
2. Model construction. Formalization of a symbolic ODE model based on the regulatory interactions of one selected gene network.
3. Qualitative simulation. Prediction of all the potential qualitative behaviors from specific set of initial conditions and parameter constraints.
4. Hypothesis testing. A network that is unable to reproduce the desired behavior is eliminated; this restarts the design cycle. Otherwise, the simulated results contain rich information, such as all transient states before the final behavior and the parameter conditions to reach any state, which can be used to revise the original design goal. For example, a designer may specify requirements about the transient states before arriving at the desired behavior; and consequently, the designer is able to define an idealized temporal profile of gene network activity.
5. Stochastic parameters. Inclusion of stochasticity allows the probability optimization, so the parameter space can be refined to guarantee the highest occurrence of the desired behavior.
6. Network selection. The proposed networks are ranked according to the likelihood of exhibiting the desired behavior. Selection for in vitro implementation starts with the highest-ranking model and proceeds down the ranks.

Finally, in a case study, our method has been tested on a real-life benchmark gene network to get a synthetic oscillator.

D05: Isaac Crespo, Nicolas Guex, Sylvian Bron, Assia Ifticene-Treboux, Eveline Faes-Van'T Hull, Solange Kharoubi, Robin Liechti, Patricia Werffeli, Mark Ibberson, Francois Majo, Mich  el Nicolas, Julien Laurent, Abhishek Garg, Khalil Zaman, Hans-Anton Lehr, Brian J. Stevenson, Curzio R  egg, Jean-Fran  ois Delaloye, Ioannis Xenarios, George Coukos and Marie-Agn  s Doucey. Angiogenic activity of breast cancer patients' monocytes reverted by combined use of systems modeling and experimental approaches

Abstract: Angiogenesis plays a key role in tumor growth and cancer progression. TIE-2-expressing monocytes (TEM) have been reported to critically account for tumor vascularization and growth in mouse tumor experimental models, but the molecular basis of their pro-angiogenic activity are largely unknown. Moreover, differences in the pro-angiogenic activity between blood circulating and tumor infiltrated TEM in human patients has not been established to date, hindering the identification of specific targets for therapeutic intervention.

In this work, we investigated these differences and the phenotypic reversal of breast tumor pro-angiogenic TEM to a weak pro-angiogenic phenotype by combining Boolean modelling and experimental approaches.

Firstly, we show that, in breast cancer patients the pro-angiogenic activity of TEM increased drastically from blood to tumor suggesting that the tumor microenvironment shapes the highly pro-angiogenic phenotype of TEM. Secondly, we predicted in silico all minimal perturbations transitioning the highly pro-angiogenic phenotype of tumor TEM to the weak pro-angiogenic phenotype of blood TEM and vice versa. In silico predicted perturbations were validated experimentally using patient TEM. In addition, gene expression profiling of TEM transitioned

to a weak pro-angiogenic phenotype confirmed that TEM are plastic cells and can be reverted to immunological potent monocytes. Finally, the relapse free survival analysis showed a statistically significant difference between patients with tumors with high and low expression values for genes encoding transitioning proteins detected in silico and validated on patient TEM.

In conclusion, inferred TEM regulatory network accurately captured experimental TEM behavior and highlighted crosstalk between specific angiogenic and inflammatory signaling pathways of outstanding importance to control their pro-angiogenic activity. Results showed the successful in vitro reversion of such an activity by perturbation of in silico predicted target genes in tumor derived TEM, and indicated that targeting tumor TEM plasticity may constitute a novel valid therapeutic strategy in breast cancer.

D06: Gabor Beke, Matej Stano and Lubos Klucar. Modelling the interaction between bacteriophages and bacteria

Abstract: Bacteriophages, viruses infecting prokaryotic organisms, can be successfully applied in treatment of bacterial infections and elimination of undesirable bacterial populations in food and bio-fermenting processes. Successful practical application of phages (either in medicine or in food biotechnology) requires detailed knowledge of interactions between phages and their hosts. For the description of these interactions, mathematical models are used.

Our aim was to develop a mathematical model describing the dynamics of relation between phages and bacteria more accurately than already known models, employing more parameters. This model could be used in our further studies related to devitalisation of bacterial pathogens in food. The new mathematical model is based on existing models (Schrag and Mittler, 1996; Zwietering et al, 1996) and can simulate the dynamics of interaction between phage and its bacterial host under specific conditions. Our final model is a system of four delay differential equations. The first describes the change of glucose (R), the second number of bacteria (N), the third yield of infected bacteria (M) and the fourth describes the population of bacteriophages (P). We calculate specific growth rate of bacteria (μ) as a function of pH and temperature (based on experimental results). The model includes parameters of bacterial growth under specific pH and temperature, efficiency of bacteriophage infection, adsorption rate and burst size of bacteriophages. These parameters were obtained experimentally. We made computational simulations of this mathematical model and developed an interactive website (<http://dublin.embnet.sk:3838/model/>) where the users can run simulations based on their own parameters. The results are displayed both graphically (five graphs, one for each equation and a graph, which combines the number of bacteria and infected bacteria) and numerically in a table (each column represent a variable and its value in time), that can be exported as CSV spreadsheet document. For the computational simulations and the website development R and its additional packages were used – deSolve (to solve differential equations numerically), ggplot2 (to produce graphs), shiny and shinyIncubator (both for creating website within R).

(supported by APVV 0098-10)

Schrag, S.J. and Mittler, J. E. Host-Parasite Coexistence: The Role of Spatial Refuges in Stabilizing Bacteria-Phage Interactions. *The American Naturalist* 148(2): 348-377, 1996. <http://dx.doi.org/10.1086/285929>

Zwietering, M. H.; De Wit, J. C.; Notermans, S. Application of predictive microbiology to estimate the number of *Bacillus cereus* in pasteurized milk at the point of consumption. *International Journal of Food Microbiology* 30: 55-70, 1996. [http://dx.doi.org/10.1016/0168-1605\(96\)00991-9](http://dx.doi.org/10.1016/0168-1605(96)00991-9)

D07: Mathias Weyder, Marc Prudhomme, Patrice Polard and Gwennaele Fichant. Modeling competence regulation during bacterial transformation in *S. pneumoniae*

Abstract: Natural genetic transformation is a transient, regulated process that induces a change of the physiological state of the cell, named competence. It proceeds through the internalization, processing and homologous recombination of exogenous DNA. In *S. pneumoniae*, competence is the result of transcription waves of three groups of genes, named early, late, and delayed com genes respectively. In cultures, competence develops abruptly during exponential growth phase in response to a competence-stimulating peptide (CSP) encoded by comC, exported and matured by the ComAB exporter. At a critical concentration, extracellular CSP activates the two-component signal transduction system ComDE. ComE~P activates both comAB and comCDE operons (early com genes), establishing a positive feedback loop, which results in a sudden rise in extracellular CSP levels, rendering all cells in a culture simultaneously competent. ComE~P activates also the competence-specific sigma factor comX, which control late com genes coding for exogenous DNA uptake and transforming protein machinery. Recently, shut-off of pneumococcal competence has been shown to be dependent on two mechanisms. First, while ComE~P activates early competence genes, ComE antagonizes their expression. Second, DprA interacts with ComE~P to block ComE-driven transcription. Therefore, we undertook the modeling of the competence regulatory circuit through ordinary differential equations. The unknown model parameters have been estimated with the Copasi software. Parameters have been fitted against known protein concentrations during competence state. Constraints have been set to ensure that the relative affinities of regulatory proteins for their target regulatory sites were consistent with experimental data. Model was parameterized so that the system is at steady state when the partner stoichiometry is not disturbed. After addition of CSP into the system, competence is induced and the kinetic of the early and late competence genes is consistent with available experimental data. Simulations of gene knock out thought to impair competence are in agreement with experimental observations. However, the simulation of the network behavior when dprA was knocked out resulted in the competence shut-off while the opposite effect was expected. To explain these result further investigations are required. Finally, competence appears also to be repressed in cells over-expressing comCDE operon as shown in previous experimental results. The next development will be to include methods enabling the simulation of competence in a multi-cellular environment to understand how the competence that initiates at the level of individual cell extends to the whole population in liquid cultures and possibly to neighboring cells in biofilms.

D08: Rafael Björk, Patrik Rydén and Tatjana Pavlenko. Structure learning for improved classification accuracy for high-dimensional omics data

Abstract: Background: High throughput omics technologies in life science such as high-throughput DNA and RNA sequencing, expression arrays, and methylation arrays have allowed genome-wide measurements of complex cellular responses for a broad range of treatments and diseases. The technologies are powerful, but in order analyze them effectively new statistical tools is often required. A biological system may be represented by a graph, where nodes represent variables, and edges represent variable interactions, direct communication between variables. Typically, many variables are unobserved and the graph will be estimated by the conditional dependency graph where edges represent conditional dependency between observed variables. Under the assumption that the graph is approximately block-diagonal, variables from different blocks can be considered independently of each other. Furthermore, information about the block structure can improve interpretability as well as improve classification accuracy for various classifiers.

Objectives: The aim is to develop a method to cluster variables such that it approximates the conditional dependency graph as block-diagonal. Furthermore we intend to develop a classifier, block-LDA, that utilize the estimated block-structure to improve the classification accuracy of the linear discriminant analysis (LDA) method.

Methods: The conditional dependency graph was estimated using graphical lasso coupled with bootstrap on the estimated correlation matrix. The block-structure of the resulting graph was estimated by sorting the graph using distance sensitive ordering and employing dynamical programming to minimize the cost of partitioning the graph into blocks. The algorithm depend on two parameters: the penalty in the graphical lasso (λ) and the cost (η) describing the cost of edges outside blocks relative to non-edges within blocks. The estimated structure is incorporated in a modification of the LDA-classifier, block-LDA, which uses cross-validation to select λ and η , as well as the blocks to be included in the classifier. The method was evaluated on simulated and gene expression microarray data for Ovarian Adenocarcinoma and its performance was estimated by the adjusted rand metric and the observed misclassification rate (mcr).

Results: The method was successful in recovering the block structure, with adjusted rand values ranging from 0.45 to 0.858, depending on simulation parameters. Within simulation we noticed a high correlation between adjusted rand-values and prediction accuracy. For simulated data the block-LDA model had significantly lower mcr than the LDA. For the microarray data the block-LDA model had roughly 20% lower mcr than the LDA classifier.

Conclusions: The suggested approach can be used to estimate the block structure which in turn can be used to improve the LDA-classifier. We believe that this approach is general and can be adapted to a wide range of classifiers and improve their performance.

D09: Guillaume Brysbaert, Mélanie Tanchon, Ralf Blossey, Marc Aumercier and Marc Lensink. Targeting the interactions of the Ets-1 oncoprotein

Abstract: The ETS-1 oncoprotein is a transcription factor that promotes DNA expression in specific biological processes such that ETS-1 is expressed only in certain circumstances. Elevated levels of expression have been found in cancerous cells and ETS-1 plays a particularly important role in invasive tumors, however all attempts to inhibit it in a therapeutic purpose have failed. In this context it is essential to study the partners of ETS-1 targeting its interactions. We have recently purified and identified novel interaction partners of ETS-1 which are PARP-1 and DNA-PK, two DNA repair enzymes strongly implicated in cancer growth and we have determined their domains of interaction. In order to characterize the interactions at a molecular level, we have first identified the occurrence of homologs of the interacting domains in the human genome. After phylogenomic clustering, we have performed large-scale pair-wise protein docking studies of representative cluster members of the interacting domains as well as their homologs, using homology modeling for domains not represented in the Protein Data Bank. The cross-docking studies may identify further putative interaction partners and will lead to a consensus characterization of the molecular details of the binding site, allowing at a later stage the design of small inhibitory molecules. The homologs were also used in the building of ETS-1/PARP-1/DNA-PK protein interaction network and its analysis with gene expression data will strengthen the identification of putative partners. The position and role of these new partners in protein interaction networks and pathways will lead to a better understanding of the role of ETS-1 in cellular signaling, especially in light of the cellular cycle of invasive cancers. They may in fine constitute new targets for anti-cancerous therapies.

D10: Ganna Androsova, Sophie Rodius, Petr Nazarov, Arnaud Muller, François Bernardin, Céline Jeanty, Simone Niclou, Laurent Vallar and Francisco Azuaje. A comprehensive integrative analysis of the transcriptional network underlying the zebrafish heart regeneration

Abstract: Despite a notable reduction in incidence of acute myocardial infarction (MI), patients who experienced it remain at risk for premature death and cardiac malfunction. The human cardiomyocytes are not able to achieve extensive regeneration upon MI. Remarkably, the adult zebrafish is able to achieve complete heart regeneration following amputation, cryoinjury or genetic ablation. This raises new potential opportunities on how to boost heart healing capacity in humans. The objective of our research is to characterize the transcriptional network of the zebrafish heart regeneration and underlying regulatory mechanisms.

To conduct our investigation, we used microarray data from zebrafish at 6 post-cryoinjury time points (4 hours, and 1, 3, 7, 14 and 90 days) and control samples. We thereon looked for the gene co-expression patterns in the data and, based on that, constructed a weighted gene co-expression network. To detect candidate functional sub-networks (modules), we used two different network clustering approaches: a density-based (ClusterONE) and a topological overlap-based (Hybrid Dynamic Branch Cut) algorithms. The visualization of the expression changes of the candidate modules reflected the dynamics of the recovery process. Also we aimed to identify candidate “hub” genes that might regulate the behavior of the biological modules and drive the regeneration process.

We identified eighteen distinct modules associated with heart recovery upon cryoinjury. Functional enrichment analysis displayed that the modules are involved in different cellular processes crucial for heart regeneration, including: cell fate specification (p-value < 0.006) and migration (p-value < 0.047), ribosome biogenesis (p-value < 0.004), cardiac cell differentiation (p-value < 3E-04), and various signaling events (p-value < 0.037). The visualization of the modules’ expression profiles confirmed the relevance of these functional enrichments. For instance, the genes of the module involved in regulation of endodermal cell fate specification were up-regulated upon injury until 3 days. Among the candidate hub genes detected in the network, there are genes relevant to atherosclerosis treatment and inflammation during cardiac arrest. These and other findings are currently undergoing deeper computational analyses. The top promising targets will be independently validated using our zebrafish (in vivo) model.

In conclusion, our findings provide insights into the complex regulatory mechanisms involved during heart regeneration in the zebrafish. These data will be useful for modelling specific network-based responses to heart injury, and for finding sensitive network points that may trigger or boost heart regeneration.

D11: Wout Bittremieux, Dirk Valkenburg, Aida Mrzic, Hanny Willems, Bart Goethals and Kris Laukens. Pattern mining of mass spectrometry quality control data

Abstract: Mass spectrometry is widely used to identify proteins based on the mass distribution of their peptides. Unfortunately, because of its inherent complexity, the results of a mass spectrometry experiment can be subject to a large variability. As a means of quality control, recently several qualitative metrics have been defined. Initially these quality control metrics were evaluated independently in order to separately assess particular stages of a mass spectrometry experiment. However, this method is insufficient because the different stages of an experiment do not function in isolation, instead they will influence each other. As a result, subsequent work employed a multivariate statistics approach to assess the correlation structure of the different quality control metrics. However, by making use of some more advanced data mining techniques, additional useful information can be extracted from these quality control metrics.

Various pattern mining techniques can be employed to discover hidden patterns in this quality control data. Subspace clustering tries to detect clusters of items based on a restricted set of dimensions. This can be leveraged to for example detect aberrant experiments where only a few of the quality control metrics are outliers, but the experiment still behaved correctly in general.

In addition, specialized frequent itemset mining and association rule learning techniques can be used to discover relationships between the various stages of a mass spectrometry experiment, as they are exhibited by the different quality control metrics.

Finally, a major source of untapped information lies in the temporal aspect. Most often, problems in a mass spectrometry setup appear gradually, but are only observed after a critical juncture. As previous analyses have not used this temporal information directly, there remains a large potential to detect these problems as soon as they start to manifest by taking this additional dimension of information into account. Based on the previously discovered patterns, these can be evaluated over time by making use of sequential pattern mining techniques.

The awareness has risen that suitable quality control information is mandatory to assess the validity of a mass spectrometry experiment. Current efforts aim to standardize this quality control information, which will facilitate the dissemination of the data. This results in a large amount of as of yet untapped information, which can be leveraged by making use of specific data mining techniques in order to harness the full power of this new information.

D12: Valérie Sautron, Elena Terenina, Élodie Merlot, Pascal Martin, Yannick Lippi, Laurence Liaubet, Armelle Prunier, Pierre Mormede and Nathalie Villa-Vialaneix.
Longitudinal CCA to analyze stress responses in pigs

Abstract: Background: The increasing development of high-throughput techniques produces data that are characterized not only by their very large dimension but also by more and more complex experimental designs. In particular, it is now a common approach to collect data at different levels of the living organism (i.e., transcriptomic data, metabolomic data... together with more integrated phenotypes) at different time points. An important question faced by systems biology is thus to understand the relation existing between these data, whilst taking into account their longitudinal nature: this question should lead to understand the global evolution of the systems in the condition under study. Method: The present contribution will present a comparative study which aims at integrating two high dimensional longitudinal datasets. The method is an adaptation of Canonical Correlation Analysis (CCA) to cubic data: the correlations between two datasets that correspond to the observations of the same variables on the same individuals at the same multiple time points are simultaneously analyzed. Several approaches will be presented to simultaneously represent all the data on a same projection space. Most of these approaches are derived from the Double Principal Component Analysis (DPCA: Bouroche, 1975) approach which targets the analysis of a single dataset obtained from a similar design. In particular, a common representation subspace for all time points is built using different strategies: selecting the best projection among the time-dependent projections, using a combined criterion across time that takes into account the different levels of correlations at different time points. A regularizing penalty is also added to the method (as in, e.g., Gonzales et al, 2009) to handle the case where the number of variables exceed the number of individuals. Data: The results of the different strategies are illustrated to study the effect of stress responses in pigs through simultaneous analysis of transcriptomic data and blood composition (cortisol, glucose ...): transcriptomic data and phenotypes were collected at three different time points before and after ACTH (corticotropin) injection. This hormone is produced in response to stress. Its principal effects are the increased production and the release of corticosteroids. The work hypothesis of the project is that there is an

antagonism between production traits and robustness and that selecting pigs which are more resistant to stress may help to increase newborn survival, resistance to disease, animal welfare and thus production. Results: The proposed methods emphasize the main correlations between gene expression and blood composition that are conserved during the stress response process. A set of genes found to be highly correlated with blood components are also extracted from the results.

D13: Teppo Annala, Anantha-Barathi Muthukrishnan, Abhishekh Gupta, Ramakanth Neeli Venkata and Andre Ribeiro. Properties of the spatial organization of Tsr protein clusters in live *Escherichia coli* cells

Abstract: *Escherichia coli* have evolved several mechanisms for detecting and responding to external stimuli such as, e.g., chemotaxis. This process is based on clusters of chemoreceptors that can perform multiple tasks (e.g. thermosensing and aerotaxis). One identified transmembrane chemotaxis receptor protein is Tsr. It forms clusters, particularly at the cell poles. In this ongoing project we use Tsr tagged with Venus fluorescent proteins to study the number, size and spatial distribution of Tsr clusters in live, individual cells as a function of temperature. In particular, from our observations, we address the following questions: is the accumulation of Tsr-Venus proteins in the cell membrane symmetric along the major cell axis? Do cell divisions generate functional asymmetries between older and newer cell poles? Do the process of clustering and the spatial distribution of the clusters differ with temperature and numbers of Tsr-Venus proteins? Finally, what drives the clusters of Tsr-Venus to the cell poles? Answers to these questions should inform us on the robustness and plasticity of the cellular functions that these clusters participate in. So far, we have analyzed the spatial distributions of Tsr-Venus and we have found that, in all temperature conditions tested, these proteins distribute preferentially at the poles and symmetrically relative to the major cell axis. Also, cell divisions introduce asymmetries in numbers between old and new poles. In finally, the degree of polar segregation changes with temperature and, in all conditions, the correlation between cluster size and distance to midcell is positive and increases in time. The results should assist in finding the sources responsible for the observed behaviors.

D14: Ralph Patrick, Kim-Anh Le Cao, Bostjan Kobe and Mikael Boden. PhosphoPICK: Probabilistic Modelling of Cellular Context for Predicting Kinase-Substrate Phosphorylation Events

Abstract: The determinants of kinase-substrate binding can be found both in the substrate sequence, and the surrounding cellular context. Cell cycle progression, interactions with mediating proteins and even prior phosphorylation events are necessary for kinases to maintain substrate specificity. While much work has focussed on the use of sequence based methods to predict phosphorylation sites, there has been very little work invested into the application of systems biology to understanding phosphorylation. However, lack of specificity in many kinase binding motifs means that sequence methods for predicting kinase binding sites are susceptible to high false-positive rates. While context information is readily available in various databases, incomplete coverage and variable certainty means that the integration of context features into a model is non-trivial.

In this work we explore a probabilistic model to accommodate missing values, seamless combination of protein interactions and cell-cycle expression, and to provide flexible options for querying potential kinase substrates. The model we present here, named PhosphoPICK (Phosphorylation in a Protein Interaction Context for Kinases), integrates known kinase-substrate relationships, protein-protein interactions, and cell-cycle data to predict kinase substrates across a variety of kinase families. PhosphoPICK shows high prediction accuracy,

with a mean AUC of 0.85 across the kinases tested. When using the model to complement sequence based kinase-specific phosphorylation site prediction, we find that the additional information can greatly increase prediction performance at low false positive levels. Our results demonstrate that a model harnessing context data can account for the short-falls in sequence information and provide a robust description of the cellular events that regulate kinase-protein phosphorylation.

D15: Alejandro F. Villaverde, Federico Morán and Julio R. Banga. Computationally efficient network inference using information theory: fMIDER

Abstract: Network inference in computational biology is the task of recovering the interactions among a set of molecular entities (genes, transcription factors, proteins, metabolites...) that are present in a cellular network. It is possible to deduce relations among nodes from data, using statistical measures such as correlation, or information-theoretical concepts such as mutual information. Indirect interactions, i.e. when an entity A exerts an influence in C by means of an intermediate entity B (A—B—C), are difficult to detect, and frequently in that situation the existing methods will predict a spurious interaction (not only A—B and B—C, but also A—C). The difficulty increases when dealing with higher-order interactions, which may involve four or more entities. Although there are a few methods available that can cope with this issue [1,2], their application to large-scale problems can be computationally costly if one wants to explore high-order interactions. This limitation is especially problematic when dealing with time-series data. MIDER [1] (Mutual Information Distance and Entropy Reduction) is a general purpose tool for reverse engineering network structure. It calculates distances among variables using mutual information, and uses joint entropies of multiple variables to distinguish between direct and indirect interactions. It takes into account time delays, and assigns causality to the predicted links using transfer entropy. MIDER is available in Matlab, thus representing an alternative in a popular environment to other inference methods which are predominantly written in R. However, the use of Matlab has some drawbacks, such as (i) the need of buying commercial licenses, and (ii) low computational efficiency compared to other languages. Here we present fMIDER, an advanced implementation of MIDER that overcomes these issues. fMIDER is written in FORTRAN, allowing for more efficient computations than Matlab; it is free for academic use; it does not require any commercial software; it is provided both as source code and as an executable; and can be run in parallel environments, which allows for additional speed-ups in performance. Results obtained on different datasets show that fMIDER can be orders of magnitude faster than the Matlab implementation of MIDER.

D16: Elson Tomás, Alexandra M. Carvalho, Paulo Mateus and Susana Vinga. Unsupervised classification of pharmacokinetic responses using non-linear mixed effects models

Abstract: Pharmacokinetic (PK) is a branch of pharmacology dedicated to the study of the evolution of substances administered externally to a living organism and comprises processes from absorption to extraction. PK dynamic models are often based on homogeneous, multi-compartment assumptions, which allow to describe drugs concentration time-series parametrically.

In this work we present an algorithm for clustering patients based on their PK drug responses. To estimate the clusters, we use an adaptation of the Expectation-Maximization algorithm proposed by Azzimonti et al. (2012) that collapses clusters that are closer than a given threshold, and estimates its parameters iteratively. We initially consider that PK responses are well described through non-linear mixed effects models (NLME) but without population parameters, considering all parameters as random effects. For maximization of the likelihood,

numeric strategies such as Newton's method can be applied to estimate the NLME parameters. The experimental multivariate data validates the model obtained, which can be further used in the prediction of the PK responses and behavior of a new patient.

This work was partially supported by national funds through FCT, Portugal, under contract PEst-OE/EME/LA0022/2011, project IntelGen (PTDC/DTP-FTO/1747/2012). SV acknowledges support by Program Investigador FCT (IF/00653/2012) from FCT, co-funded by the European Social Fund (ESF) through the Operational Program Human Potential (POPH).

D17: Lujia Chen, Chunhui Cai, Vicky Chen and Xinghua Lu. Trans-species learning of cellular signaling systems with bimodal deep belief networks

Abstract: Motivation: Model organisms play critical roles in biomedical re-search of human diseases and drug development. An imperative task is to translate information/knowledge acquired from model or-ganisms to human. Here, we aim to predict human cell responses to diverse stimuli, based on the responses of rat cells treated with the same stimuli.

Results: We hypothesized that rat and human cells share a com-mon signal-encoding mechanism but employ different proteins to transmit signals, and we developed a bimodal deep belief network (bDBN) and a semi-restricted bimodal deep belief network (sbDBN) to represent the common encoding mechanism and perform trans-species learning. The models include hierarchically organized latent variables capable of capturing the statistical structures in the ob-served proteomic data in a distributed fashion. The results show that the models significantly outperform two current state-of-the-art classification algorithms. Our study demonstrated great potential in using deep hierarchical models to simulate cellular signaling sys-tems.

D18: Monica Golumbeanu, Pejman Mohammadi, Celine Hernandez, Manfredo Quadroni, Amalio Telenti, Angela Ciuffi and Niko Beerenwinkel. Characterizing the dynamics of cellular response to HIV-1 infection through clustering of time-series proteomics data

Abstract: HIV-1 goes through a 24-hour replication cycle, during which the pathogen enters the cell, integrates its genome into that of the host and exploits the cellular machinery in order to reproduce new virions. The transcriptional and proteomic profiles of the host cell are consistently influenced during this process. Describing the dynamics of the cellular response following viral invasion is of key importance to understanding the underlying mechanisms of HIV infection.

Our study focuses on characterizing the progression of proteomic and phosphoproteomic response of the host cell within 24h after presentation of the virus. We track the dynamics of protein synthesis through quantitative time-series SiLAC scanning of healthy and infected primary CD4+ T cells. Protein expression and phosphorylation levels in the cells are measured at six hours intervals.

First, we are interested to identify the proteins and phosphorylation sites which consistently suffer significant changes throughout the experiment. Subsequently, we aim to characterize and stratify the proteomic response of healthy versus infected cells by clustering proteins or phosphorylation sites with similar activity patterns in time. In order to do so, we design a maximum likelihood methodology for clustering time-series data.

D19: Arnau Montagud, Andrei Zinovyev and Emmanuel Barillot. Multiscale mathematical modelling of breast cancer invasion

Abstract: Metastasis is a process that starts with invasion of surrounding tissue by tumour cells. It requires remodelling of the extra-cellular matrix and Epithelial-to-Mesenchymal

Transition (EMT), loss of cell adhesion and polarity and increased motility. Understanding invasion mechanisms is crucial to improve prognosis and develop new cancer treatment strategies, but we still lack a detailed explanation of this process. In the past years several efforts have been done in systematising different mechanisms of cell migration, also termed invasion modes, and understanding their underlying causes (1,2).

We devised a mathematical model that would incorporate information of a series of traits, cellular and environmental, that would output in a set of invasion modes. For this, the model incorporates different pathways such as apoptosis, EMT determinants, cell cycle, tumour microenvironment-cell sensing, cell motility, extracellular matrix modification, etc. The resulting influence network is being translated into a mathematical model using discrete logical modelling (3,4). The model will be ready to be tuned by observed phenotypes on existing data from experimental results on tumours, cell lines and organoids.

Any realistic and useful mathematical model of tumour invasion must be multiscale as the process of invasion involves at least three levels of details: intracellular molecular processes determining individual cellular properties (5); interaction between a cell and its microenvironment affecting cell state and properties (6); and biochemical and biophysical interactions between cells in the context of tumour microenvironment, leading to various patterns of collective cell behaviour (7). Present work is part of a collaborative effort to model tumour invasion in order to identify treatment strategies and to understand underlying properties of metastasis.

Bibliography:

1. Friedl, P & Wolf, K. Plasticity of cell migration: a multiscale tuning model. *J Cell Biol.* 188, 11–19 (2010)
2. Friedl, P & Alexander, S. Cancer Invasion and the Microenvironment: Plasticity and Reciprocity. *Cell* 147, 992–1009 (2011)
3. Calzone, L et al. Mathematical modelling of cell-fate decision in response to death receptor engagement. *PLoS Comput Biol.* 6, e1000702 (2010)
4. Stoll, G, Viara, E, Barillot, E & Calzone, L. Continuous time Boolean modeling for biological signaling: application of Gillespie algorithm. *BMC Syst Biol.* 6, 116 (2012)
5. Mateescu, B et al. miR-141 and miR-200a act on ovarian tumorigenesis by controlling oxidative stress response. *Nat Med.* 17, 1627–1635 (2011)
6. Poincloux, R, Lizárraga, F & Chavrier, P. Matrix invasion by tumour cells: a focus on MT1-MMP trafficking to invadopodia. *J Cell Sci.* 122, 3015–3024 (2009)
7. Ramis-Conde, I & Drasdo, D. From genotypes to phenotypes: classification of the tumour profiles for different variants of the cadherin adhesion pathway. *Phys Biol.* 9, 036008 (2012)

D21: Mahsa Ghanbari, Julia Lasserre and Martin Vingron. Reconstruction of gene networks using prior knowledge

Abstract: Reconstructing gene regulatory networks (GRNs) from expression data is a challenging task that has become essential to the understanding of complex regulatory mechanisms in cells. The major issues are the usually very high ratio of number of genes to sample size, and the noise in the available data. Integrating biological prior knowledge to the learning process is a natural and promising way to partially compensate for the lack of reliable expression data and to increase the accuracy of network reconstruction algorithms.

In this poster, we describe PriorPC, a new algorithm based on the PC algorithm. Despite being one of the most popular methods for Bayesian network reconstruction, PC is known to depend strongly on the order in which nodes are presented, especially for large networks. PriorPC exploits this flaw to include prior knowledge. We show on both synthetic and real data that the structural accuracy of networks obtained with PriorPC is greatly improved compared to PC. We also show the robustness of PriorPC to the noise in the prior knowledge.

PriorPC is also fast and scales well for large networks which is important for its applicability to real data.

D22: Samuel Collombet, Morgane Thomas-Chollier, Touati Benoukraf, Annouck Luyten, Chris Van Oevelen, Daniel G. Tenen, Thomas Graf and Denis Thieffry. Logical modelling of immune cell specification and reprogramming

Abstract: Blood cells arise from a common set of hematopoietic stem cells that differentiate into more specific progenitors, ultimately leading to different functional lineages. This process relies on the activation and repression of different genes modules, controlled by transcription factors (TFs) that recognise specific DNA sequence in genomic elements (cis-regulatory elements) and regulate gene expression. Novel high-throughput technologies allow the characterisation of cell-specific regulatory elements by studying chromatin state and TFs binding sites (ChIP-seq), in conjunction with genes expression (RNAseq). Proper integration and analysis of these data enable the delineation of novel regulatory interactions, which can be modelled and analysed using formal methods, thereby fostering our understanding of the mechanisms controlling cell fate at a system level, and enabling the prediction of the effects of molecular perturbations in silico.

To reconstruct the regulatory network controlling hematopoietic specification, we combined information extracted from the literature with data from ChIP-seq experiments targeting TFs involved in myeloid and lymphoid specification. Additionally, we used histone modifications ChIP-seq data (H3K4me1/2/3, K3K27ac, H3K27me3) to identify others regulatory elements, and sequence analysis (motif discovery and pattern matching) to predict TF recruitments.

These experiments were performed at different stages of hematopoietic development (stem cells, restricted progenitors and differentiated cells), and during reprogramming of B-cells (lymphoid lineage) into macrophages (myeloid).

Using a multilevel logical framework, we built a dynamical model of the resulting network and fitted it to gene expression data from the FANTOM5 (<http://fantom.gsc.riken.jp/>) and IMMGEN consortia (<http://www.immgen.org/>). Dynamical simulations of this model further enabled us to predict the effect of perturbations of specific regulatory components or interactions (gain/loss-of-function, mutation of binding sites) at specific stages of development. These prediction are currently being assessed experimentally. Moreover, we are performing additional ChIP-seq experiments to delineate the targets of several TFs in cell types for which data are not yet available.

This study has already contributed to the identification of novel regulatory interactions, and ongoing experiments should further help us to refine our model. This will ultimately allow us to generate more precise predictions, in particular regarding efficient cell reprogramming protocols.

D23: Aristotelis Kittas, Amelie Barozet, Jekaterina Sereshti, Niels Grabe and Sophia Tsoka. CytoASP: A Cytoscape plug-in for logical modelling of signalling networks using BioASP

Abstract: The analysis and interpretation of regulatory networks lies at the core of Systems Biology challenges. The sign consistency modelling framework for analysis of interaction graphs allows for reasoning over the interaction graph and qualitative data. In this context, the BioASP library offers functions that allow implementation of Answer Set Programming (ASP), a declarative problem-solving paradigm in which a problem is encoded by a collection of rules such that its intended models (called answer sets) represent solutions to the problem [1]. BioASP, facilitates computational predictions and hypothesis-generation, using the sign consistency model [2]. Although BioASP is offered through the Python programming language, its use through popular network analysis and visualisation platforms, such as

Cytoscape [3], can enhance its integration with other widely used analysis protocols and facilitate its use by scientists with little programming experience.

We present CytoASP, a Cytoscape 3.x plugin for analyzing regulatory networks. It allows the use of BioASP for consistency checking, diagnosis, and repair of regulatory connections. BioASP confronts an influence graph with experimental data, checking the consistency of the observed variations on the underlying graph and calculating repair sets. Implementation as a Cytoscape plugin allows visualisation of the results and simultaneous analysis of multiple networks, as well as a wide selection of visualisation options, such as custom colouring of nodes and edges, allowing easy visualisation of the regulation state of nodes. Experimental observations are represented as solid coloured nodes, and predictions that hold under all prediction sets as border coloured. It also offers repair mode visualisation, making easy to spot what nodes or edges must be changed in order to achieve consistency in the network. CytoASP is provided as a stand-alone program and all functionality is packaged into a single file. It does not require Python, Python libraries, BioASP or BioASP solvers to be installed and will not interfere with any installed libraries.

References

- [1] C. Baral, Knowledge Representation, Reasoning and Declarative Problem Solving, vol. 2. Cambridge University Press, 2003.
- [2] M. Gebser, A. König, T. Schaub, S. Thiele, and P. Veber, “The BioASP Library: ASP Solutions for Systems Biology,” in 2010 22nd IEEE International Conference on Tools with Artificial Intelligence (ICTAI), 2010, vol. 1, pp. 383–389.
- [3] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, “Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks,” *Genome Res.*, vol. 13, no. 11, pp. 2498–2504, Nov. 2003.

D24: Sabeur Aridhi, Haitham Sghaier, Mondher Maddouri and Engelbert Mephu Nguifo. Domain knowledge-based model for phenotype prediction of ionizing-radiation-resistance in bacteria

Abstract: Abstract: The use of certain resistant to the treatment of radioactive waste microorganisms is determined by their surprising ability to adapt to stress caused by ionizing radiation (Daly, 2009) (Daly, 2012) (Sghaier et al., 2013). Only few works were interested to in silico methods for prediction of ionizing radiation in bacteria. In a recent work (Aridhi et al., 2013), we proposed a multiple-instance learning (MIL) approach for predicting ionizing-radiation-resistant bacteria (IRRB) using proteins implicated in basal DNA repair in IRRB. The experimental results of the proposed approach are satisfactory and provide a MIL-based prediction system that predicts the ionizing radiation in bacteria. However, the proposed system in (Aridhi et al., 2013) uses only structural information about basal DNA repair proteins. In this present work, we aim to combine machine learning techniques with biochemical properties of basal DNA repair proteins in order to enhance the developed system. The objective of this combination is to produce a multi-criteria prediction system that includes properties like the decimal reduction dose (D10) of the studied microorganisms, rate of amino acids and amino acid sites under positive selection in basal DNA repair protein structures. The proposed system aims to predict the ionizing radiation in bacteria.

References:

- Aridhi S., Sghaier H., Maddouri M. and Mephu Nguifo E. (2013) Computational phenotype prediction of ionizing-radiation-resistant bacteria with a multiple-instance learning model. In Proceedings of the 12th International Workshop on Data Mining in Bioinformatics (BioKDD '13). ACM, New York, NY, USA, 18-24.
- Daly M.J. (2009) A new perspective on radiation resistance based on *Deinococcus*

radiodurans. *Nat Rev Microbiol* 7(3):237–245

Daly M.J. (2012) Death by protein damage in irradiated cells. *DNA Repair* 11(1):12–21

Saidi R., Maddouri M., and Mephu Nguifo E. (2010) Protein sequences classification by means of feature extraction with substitution matrices, *BMC Bioinformatics* 2010, 11:175, ISSN 1471-2105.

Sghaier H., Thorvaldsen S., and Saied N. (2013) There are more small amino acids and fewer aromatic rings in proteins of ionizing radiation-resistant bacteria. *Annals of Microbiology*, pages 1–9.

D25: Konstantin Kozlov and Alexander Samsonov. Differential Evolution Entirely Parallel Method for Sequence-based Modeling of Gene Expression

Abstract: Computational systems biology relies on mathematical modeling as a tool that provides new insights into the mechanisms controlling the biological system behavior. An important task is to develop methods to find model parameters that empower a model to predict in silico the consequences of biological experiments. Here we consider the refined Differential Evolution Entirely Parallel (DEEP) method. The distinctive features of DEEP are flexible mechanism for handling multiple objective functions and substitution strategy that takes into account the age of population members defined as the number of generations in which an individual survived without changes. The method implementation is freely available at <http://urchin.spbcas.ru/trac/DEEP>. We investigate the dependence of its convergence on the control parameters with the set of test functions. The numerical results showed that, though the method still needs some control parameters to be set manually, the population size is the most important parameter. A sequence-based model of gap gene regulatory network controlling segment determination in the early *Drosophila* embryo was developed. The state variables of this model are the concentrations of mRNAs and proteins encoded by four gap genes *hb*, *Kr*, *gt*, and *kni*. The model implements the thermodynamic approach to calculate the expression of a target gene at the RNA level. This expression level is proportional to gene activation level also called promoter occupancy, and is determined by concentrations of eight transcription factors *Hb*, *Kr*, *Gt*, *Kni*, *Bcd*, *Tll*, *Cad* and *Hkb* taken from FlyEx database. Two sets of the reaction-diffusion differential equations for mRNA and protein concentrations describe the dynamics of the system. We added the delay parameter to account for the average time between events of transcription initiation and corresponding protein synthesis. The model spans the time period of cleavage cycles 13 and 14A and the interval of A-P axis from 0% to 100% of embryo length. The obtained parameters accurately describe the data and are able to predict the expression patterns of three gap genes *hb*, *gt*, and *kni* in the *Drosophila* embryo mutant for *Kr* gene. Though these data were not used for fitting the parameters, the model is able to reproduce the characteristic features of gap gene pattern in *Kr* mutants, namely, the decrease in the level of gap gene expression and a large shift of the *gt* posterior domain.

This study was supported by the "5-100-2020" Program of the Ministry of Education and Science of the Russian Federation and by RFBR Grant 14-01-00334.

D26: Nicole Radde, Karsten Kuritz, Caterina Thomaseth and Frank Allgöwer. The circuit-breaking algorithm for systems with order preserving flow

Abstract: Motivation: In earlier work we have developed the circuit-breaking algorithm (CBA), a method

that uses the circuit structure of the interaction graph of a biological regulatory network in order to construct a one-dimensional characteristic whose zeros correspond to the fixed points of the system. Since that time the usefulness

of this algorithm was demonstrated via application to many models for intracellular regulation mechanisms.

Results: Here we apply the CBA to systems whose flow preserves a partial order with respect to some cone. We consider relations between stability of the fixed points and the derivative of their corresponding zeros of the circuit-characteristic. In particular, we derive sufficient conditions for instability of a fixed point in case that the open loop system is globally asymptotically stable. We furthermore fully characterize stability of the fixed points if in addition the closed loop system is monotone.

D27: Alain Sewer and Florian Martin. Using data-biased random walks on signed graphs to quantify perturbations in causal biological network models

Abstract: Background: Analyzing high-throughput transcriptomics data in the context of large-scale gene networks has significantly improved the understanding of biological processes at the molecular level. Among others, this approach requires the ability to extract sub-networks that are most relevant to the input data. A method using data-biased random walks was recently proposed by Komurov et al. to address this task in the case of protein-protein interaction networks [1]. However, as it is based on graphs with unsigned edges, it cannot be directly applied to networks containing edges describing regulatory relationships such as activation or inhibition. Here, we present an extension of Komurov's method applicable to signed graphs based on data-biased random walks. It enables us to analyze input data in the context of refined network types such as causal network models and thereby more accurately quantifies the biological perturbations measured in the experiment.

Method: We used the concept of graph polarization to generate an unsigned graph from the original signed graph. We also redefined the Markov rules governing the data-driven bias of the random walk, which also includes specific conditions at the graph boundaries. After checking the irreducibility of the random walk defined on the unsigned graph, we computed its main asymptotic properties. They enabled the quantification of the importance of each node by comparing the obtained results to the case of an unbiased random walk. We used several metrics such as the "Markov centrality" to quantify the effects of the data-driven biases in the random walk properties. Our approach deals well with incomplete data or networks, representing a flexible means for integrating experimental data and the prior knowledge contained in the network models.

Results: The proposed method was first applied to an example in order to verify the correctness of the behavior of the random walk in typical configurations, which are more diverse than in the case of an unsigned graph considered by Komurov et al. We then considered a transcriptomics dataset describing a proof-of-principle experiment where Normal Human Bronchial Epithelial cells were exposed to varying concentrations of TNF α for various time periods. The results obtained for a network model describing the NF- κ B signaling pathway showed that the perturbed subnetwork was correctly identified and that the corresponding metrics displayed the expected time- and dose-dependences.

Conclusions: Our novel method provides a mathematically sound means for a flexible integration of transcriptomics data with prior biological knowledge encoded in causal network models, which explicitly takes into account the full information contained in the corresponding signed graph.

[1] Komurov K, White MA, Ram PT (2010) Use of Data-Biased Random Walks on Graphs for the Retrieval of Context-Specific Networks from Genomic Data. PLoS Comput Biol 6(8)

D28: David Cohen, Loredana Martignetti, Emmanuel Barillot, Andrei Zinovyev and Laurence Calzone. Modelling the intracellular molecular network of tumoural invasion

Abstract: Understanding the etiology of metastasis is very important in clinical perspective, since it is estimated that metastasis accounts for 90% of cancer mortality. Metastasis is a sequence of multiple steps: 1) infiltration of tumour cells into the adjacent tissue, 2) migration of tumour cells towards vessels, 3) intravasation of tumour cells by bridging the endothelial monolayer, 4) survival and travelling in the circulatory system (blood or lymphoid), 5) extravasation when circulating tumour cells re-enter distant tissue and 6) colonisation and proliferation in distant organs. The early stages of invasion are tightly controlled in normal cells and can be drastically affected by malignant mutations. They thus constitute the principal determinants of metastatic rate even if the later stages take long to occur. We introduce two mathematical models that recapitulate a number of published experimental results in molecular biology of early tumour cell invasion. The models are characterised by two levels of granularity at the level of individual genes and at the level of pathways, where each pathway is a module from the detailed model. The model has been validated on 16 previously described mutants and in addition on high-throughput data collected from tumour samples and cell lines. For this, we have established a simple method to compare the stable states of the logical model with gene expression. Our aim is to suggest a systematic mechanistic explanation for the majority of experimentally validated mutations on local invasion and migration processes, by describing the functioning of an intracellular network controlling them. Our analysis also predicts the effect of those mutations and their combinations on several cellular phenotypes that have not been yet performed in wet experiments.

D29: Andreas Troll. A new Approximation Approach for the Chemical Master Equation

Abstract: In chemical or biological reaction networks, reactions often happen only between a few individuals of the involved species. Examples from different areas are chemical reactions with only a few molecules involved, gene regulatory systems or the outbreak of a disease in a hamlet. Here randomness has a huge influence on how often which reaction fires (do two molecules react, how many proteins are produced or will a person get infected if he comes in contact with an ill one). The “chemical master equation” (CME), a system of ordinary differential equations, provides a good mathematical model for calculating the probability density function of the network at a specific time. In the cases described above the CME is better suited than for example rate equations, which are more accurate for problems with a very high number of individuals. Because the number of equations in the CME rises exponentially with the number of involved species, also known as “curse of dimensionality”, solving the CME directly is often impossible. Analytical solutions are known only in a few simple cases. There are different approaches for approximating the solution: Stochastic methods, like the stochastic simulation algorithm of Gillespie, use Monte-Carlo simulation to simulate possible pathways. For numerical deterministic methods, first the big (or even infinite) state space has to be truncated until it is small enough for computers to work with. After that, one can use several different approaches (for example low-rank-tensors or wavelets) to further reduce the number of equations. If the chosen state space is too small, it is possible that important states with a high probability are not in the area one works with, which leads to a poor approximation. To lower this truncation error, it is often recommended to select the state space so large, that “the truncation error can be neglected”, which produces a big space and so expensive calculations. Even if one is just interested in a small part, one will need to calculate a large area to get accurate results. We will introduce a new approach, based on a z-transform, which enables us to make the state space smaller without losing too much accuracy. The outer state space with low probabilities will be approximated and transformed. The so obtained partial differential equations are then coupled with the inner part, which is calculated with an arbitrary method.

D30: Anida Sarajlic, Vladimir Gligorijevic, Djordje Radak and Natasa Przulj. Network wiring of pleiotropic kinases yields insight into dissociation of diabetes and aneurysm

Abstract: Recent studies suggest a protective role of diabetes on the development of aneurysm, but the biological mechanisms behind this are still unknown [1,2]. Interestingly, this type of association is not present in case of diabetes and atherosclerosis despite similar risk factors for aneurysm and atherosclerosis. We use molecular interaction networks to examine the underlying molecular mechanisms responsible for these relationships. We postulate the existence of genes that disrupt the pathways needed for the onset of aneurysm in the presence of diabetes. Motivated by the significance of genetic interactions for understanding disease-disease associations [3], we suspect that a mutation of a gene on a pathway involved in diabetes is related to a functional change of a protein on an aneurysm-related pathway, explaining the protective role of diabetes on the development of aneurysm. We approach this problem by integrating protein-protein interaction and genetic interaction data. We create a protein-protein interaction sub-network that contains pathways related to the three diseases that contain genes involved in the following genetic interactions: one gene in the genetic interaction is part of a diabetes-related pathway and the other gene is part of an aneurysm-, or an atherosclerosis-related pathway. We examine the topology of this sub-network and use Simmelian brokerage measure [4] to identify proteins whose local topology could explain their high “destructiveness” for the pathways they are in. The identified set of proteins is enriched in biological functions including the cell-matrix adhesion, which facilitates mechanisms that have already been suggested as possible causes of diabetes-aneurysm dissociation. We further narrow the set down to 16 proteins that are on an aneurysm- or an atherosclerosis-related pathway and are encoded by genes participating in genetic interactions with a gene on a diabetes pathway. The set is enriched in kinases and phosphorylation processes, with two kinases that are both on aneurysm and atherosclerosis pathways being pleiotropic. Kinases can turn on or off proteins, explaining how functional changes of such proteins could result in disruption of pathways. So if on an aneurysm-related pathway a gene is turned off, this could prevent the onset of the disease. However, mutations of pleiotropic genes could have effects only on one of the traits, which explains why pleiotropic kinases that are both on aneurysm- and atherosclerosis-related pathways could disrupt aneurysm-related pathways resulting in dissociation of diabetes and aneurysm, but not affect the atherosclerosis-related pathways. We believe that this set of 16 proteins could guide future research on relationships between the three diseases.

[1] Prakash et.al., J Am Heart Assoc (2012),1(2): jah3-e000323.

[2] Rango et.al., J Vasc Surg (2012),56(6):1555-63

[3] Ashworth et.al., Cell (2011),145(1):30-8

[4] Latora et. al., J Stat Phys (2013),151:745–764.

D32: Isa Kirk, Søren Brunak and Kirstine Belling. The correlation between the human protein interactome and conserved mammalian syntenic blocks

Abstract: Synteny, the conservation of gene arrangements between different species, was first studied by cytological hybridizations of genomes from different species. Today, this has been replaced by studies on sequence conservation. The functional reason why genes in some chromosomal regions stay together through evolution is yet unknown. In this study, we investigated whether genes have stayed in close proximity during evolution because of the interactions of their gene products.

We defined syntenic blocks using orthologous genes from the following five species: human, chimpanzee, pig, dog and mouse. Bilateral syntenic blocks (BSBs) were defined as genomic

stretches of at least two orthologous genes between human and each of the other four species. Subsequently, we accounted for micro-rearrangements by collapsing neighboring blocks despite one was inverted. We defined few BSBs between human and chimpanzee, but we did not in general observe a correlation between the number of BSBs and the evolutionary closeness of the species to human.

We combined the BSBs from the four comparisons into 829 multilateral syntenic blocks (MSBs), defined as conserved gene stretches in all five mammals. We saw that half of the human chromosomes are fairly conserved with only few MSBs covering whole chromosomes, whereas others have considerable more MSBs due to more recombination events during speciation. We investigated the ratio of protein interactions within (internal) and outside (external) the blocks using protein-protein interactions (PPIs) from our in-house resource, InWeb, consisting of four million PPIs. Of the 232 collapsed MSBs where both internal- and external-PPIs have been detected, we found that MSBs with fewer genes ($n < \sim 30$) have much higher int-/ext PPI ratios than the MSBs with more genes, and half of the blocks had a significantly higher int/ext PPI ratio than one million permutations. To further investigate this tendency, we studied the interaction patterns of the non-collapsed and thereby overall shorter MSMs and saw an even stronger trend of many interactions between the proteins encoded in the same syntenic blocks compared to external interactions.

We conclude that the reason why some genes stay in close genomic proximity during evolution might be influenced by the need to facilitate the interactions of their gene products. The results of our study suggest a potential upper limit for the maximum number of genes in syntenic blocks, which is in agreement with previous studies estimating the maximum number of proteins present in functional protein complexes.

D33: Lingjian Yang, Aristotelis Kittas, Johnathan Watkins, Anita Grigoriadis, Sophia Tsoka and Lazaros Papageorgiou. An optimisation framework inferring module activity for breast cancer classification

Abstract: In complex diseases such as, breast cancer, current diagnostic and prognostic tools cannot accurately predict clinical outcomes, due to the heterogeneous nature of the disease, unknown mechanisms and measurement noise. High throughput characterisation of disease properties in patient samples suffers from dimensionality problem, i.e. the number of genes and gene products is far greater than the number of samples, making prediction of reliable and robust biomarkers difficult. Dimension reduction methods evaluate the discriminative power of all genes and select a subset of most differentially expressed genes before a classifier is trained using the subset of genes to predict the disease outcome of new samples. However, gene signatures derived from different datasets share very little overlap and offer inadequate prediction accuracy. Recently, the integration of microarray gene expression profiles with known molecular interactions through biochemical pathways or protein-protein interaction (PPI) networks have been proposed to improve prediction. It has been demonstrated that classifiers based on functional modules outperform traditional classifier based on single genes in terms of diagnosis and prognosis power.

We present a novel network-based computational protocol for disease classification. A list of seed genes with largest numbers of direct interactors in PPI network is made. We generate modules in this networks by selecting seed gene and their direct interactors. For each module, expression patterns of constituent genes are summarised into a new feature, termed module activity, which expresses a weighted linear combination of expression values across all constituent genes. Gene weights are determined by our proposed optimisation model so as to achieve optimal discriminative power among disease outcomes. The resultant module activities are then order-ranked by information gain feature ranking, and the top module activities are used by classifiers to predict phenotypic outcome in new samples. The optimal

number of module activities is determined using cross-validation. The proposed framework is applied to breast cancer datasets that represent both two-phenotype and multi-phenotype classification problems. Extensive comparative analyses will be discussed to evaluate model performance with regards to existing approaches in the literature.

D34: Otoniel Rodríguez Jorge, Linda Kempis Calanis, Denis Thieffry and Angélica Santana Calderón. Logical modelling of TLR5 signals helps unravel the mechanism of neonatal CD4 T cell activation by flagellin.

Abstract: Toll-Like Receptor 5 (TLR5) specifically recognizes the flagellin monomer as ligand. As several other TLR ligands, flagellin is being evaluated as a vaccine adjuvant given its ability to induce pro-inflammatory signaling cascades in a variety of cell types [1]. In T cells, TLR5 directly recognizes flagellin providing a co-stimulatory signal that synergises with the T cell receptor-mediated (TCR) signals [2]. However, neonatal CD4 T cells produce a defective response to infections as well as to several vaccines [3].

Proper integration of current data into a predictive dynamical model would constitute a very useful tool to assess the effect of TLR5 activation in neonatal versus adult T cells. Although models for TCR activation in adult exist [4], dynamic models for co-receptor molecules that could be potentially used as adjuvants have not been considered yet. This work precisely aims at building a predictive dynamic model for TLR5 signaling in response to flagellin in neonate and adult CD4 T cells.

We have used the software GINsim (<http://www.ginsim.org>) to (i) define and annotate a regulatory graph for the TLR5 pathway, (ii) assign logical rules to each component of this graph, (iii) define a dynamically consistent reduction of this model, and (iv) perform asynchronous simulations for several initial conditions and perturbations (wild-type, loss- or gain-of-functions, etc.), focusing primarily on situations reported in the literature.

In parallel with this modelling work, we are experimentally measuring the activation of AP-1 and NF-kappaB transcription factors in response to flagellin. Our preliminary results point to differences in NF-kappaB activation between neonatal and adult CD4 T cells in response to flagellin. Remarkably, neonatal CD4 T cells seem to mount a stronger response to flagellin than their adult counterparts. These results are currently being used to refine our model.

Our model should be useful to predict the effects of the adjuvant flagellin on neonatal and adult CD4 T cell activation. Furthermore, it could be used as a template to model the activation of different TLRs in a variety of cells.

References:

[1] Levy O, Goriely S, Kollmann TR (2013). Immune response to vaccine adjuvants during the first year of life. *Vaccine* 31: 2500-5.

[2] Crellin NK et al (2005). Human CD4+ T cells express TLR5 and its ligand flagellin enhances the suppressive capacity and expression of FOXP3 in CD4+CD25+ T regulatory cells. *J Immunol* 175: 8051-9.

[3] PrabhuDas M et al (2011) Challenges in infant immunity: implications for responses to infection and vaccines. *Nat Immunol* 12: 189-94.

[4] Saez-Rodriguez J et al (2007). A logical model provides insights into T cell receptor signaling. *PLoS Comput Biol* 3: e163.

D35: Djordje Djordjevic, Andrian Yang, Armella Zadoorian, Kevin Rungrueeecharoen and Joshua Ho. How difficult is inference of mammalian causal gene regulatory networks?

Abstract: Gene regulatory networks (GRNs) play a central role in systems biology, especially in the study of mammalian organ development. Many methods have been developed to infer GRNs from genome-wide expression data, but there is currently no gold standard mammalian

developmental GRNs for assessing these methods and their underlying assumptions. Two key questions remain largely unanswered: Is it possible to infer tissue-specific mammalian causal GRNs using gene expression profiles and other molecular network data? What experimental design should be used? We assembled two mouse GRN datasets (embryonic tooth and heart) and matching microarray gene expression profiles to systematically investigate the difficulties of mammalian causal GRN inference. The GRNs were assembled based on >2,000 pieces of experimental genetic perturbation evidence from manually reading >150 primary research articles. These data have thorough annotation of tissue types and embryonic stages, as well as the type of regulation (activation, inhibition and no effect), which uniquely allows us to estimate both sensitivity and specificity of the inference of GRN edges. The two microarray datasets together contain almost 200 gene expression profiles from embryonic developmental time-series and molecular perturbation experiments performed on in vivo tooth or heart tissues. Using these unprecedented datasets, we found that gene co-expression does not reliably distinguish true positive from false positive interactions, making inference of GRN in mammalian development very difficult. Taking into account perturbation experimental design greatly increases the sensitivity and specificity of the analysis. We showed that causal gene regulatory relationship can be highly cell type or developmental stage specific, suggesting the importance of employing expression profiles from homogeneous cell populations. This study provides essential datasets and empirical evidence to guide the development of new GRN inference methods for mammalian organ development.

D36: Adrien Fauré, Barbara Vreede, Élio Sucena and Claudine Chaouiya. A Discrete Model of *Drosophila* Eggshell Patterning Reveals Cell-Autonomous and Juxtacrine Effects

Abstract: The remarkable structure of the *Drosophila* eggshells, with their dorsal appendages, proceed from the two-dimensional patterning of the follicular epithelium that surrounds the oocyte during oogenesis. In the recent years, this system has been given particular attention from experimentalists and modelers with an interest in pattern formation and evolution, as the patterns shows significant variability across *Drosophila* species. Yet several key aspects remain to be clarified, as experimental evidence regarding the underlying genetic network is inconclusive or even contradictory in several respects.

Indeed, very similar experiments over the role of the BMP pathway have lead to opposite conclusions regarding the specification of the anterior competence region where dorsal appendages form (Shrivage et al., 2007; Yakoby et al., 2008). Moreover the putative mechanism proposed to control the formation, during stage 10b of oogenesis, of the roof / floor boundary within the appendage forming region (Simakov et al., 2012) fails to account for key observations.

We address these issues through logical modeling. Focusing on *D. melanogaster*, we propose that the specification of the anterior competence region results from the early influence of both the EGF and BMP pathways. Moreover, building on a phenomenological description, we introduce a mechanism of juxtacrine communication to account for boundary formation between the presumptive roof and floor domains.

In contrast with Simakov et al. (2012), we propose that this boundary forms at the interface with the roof domain, and results in the differentiation of a row of neighboring cells into floor cells. This mechanism requires the cessation of the Grk signal, which suggests that the concomitant formation of the vitelline membrane might be a major trigger of the stage 10a to 10b transition.

Extensive simulation of our model reproduces most experimental evidence, both in the wild-type case and a number of mutants, with unprecedented accuracy. Remarkably, our model thus reconciles apparently conflicting experimental observations regarding the role of the BMP pathway in the definition of the anterior competence region. Our simulations provide a

detailed time-line of events that can be easily tested, and we are now planning experimental validation of our predictions.

References:

Fauré et al. (2014) PLOS Comp. Biol. 10(3): e1003527.

Shrivage et al. (2007) Development 134: 2261–2271.

Simakov et al. (2012) Development 139: 2814–2820.

Yakoby et al. (2008) Development 135: 343–351.

D37: Vijayabaskar Ms, Nadine Obier, Monika Lichtinger, Debbie Goode, Michael Lie-A-Ling, Elli Marinopoulou, Josh Lilly, Constanze Bonifer, Valarie Kouskoff, Georges Lacaud, Berthold Göttgens and David Westhead. Understanding the mechanism of in vitro cellular differentiation in mouse through integrative analysis of genome-wide chromatin accessibility, chromatin modifications, transcription factor binding and gene expression data

Abstract: The regulation of spatiotemporal gene expression is crucial for determining the differentiation state of a cell and this regulatory mechanism in eukaryotes is a composite system involving interplay of complex processes like combinatorial transcription factor binding, chromatin remodelling, and epigenetic modifications. Recent advances in next-generation sequencing have enabled us to perform comprehensive genome-wide analyses of these individual systems gaining a massive wealth of information and data. However, effective integration of this information to gain a deeper understanding of transcriptional control has been a challenge. Here we have taken haemopoiesis as the model system where mouse embryonic stem cells differentiate progressively in vitro into a fully committed macrophage. To elucidate the system-wide interrelationship between epigenome and transcriptome we have systematically analysed gene expression (RNA-seq), chromatin accessibility (DNaseI-seq), histone modifications (ChIP-seq) H3K27ac, H3K9ac, H3K4me3, and H3K27me3 and binding of transcription factors (ChIP-seq) that act as master regulators for each of the six cell types in the differentiation pathway. Transcriptome data was used to obtain meta-clusters of genes with distinct expression patterns that are involved in lineage commitment and in the maintenance a particular differentiation state. By integrating chromatin accessibility and chromatin modifications data we were able to obtain both proximal and distal cis-regulatory elements that maintain the states of differentially expressed genes as active, poised or repressed along the differentiation pathway. The consolidated results from these analyses helped us identify transcription factors (TFs) that may be key regulators during differentiation. Through a detailed study of TF binding sites from ChIP-seq data for these master regulations along with comprehensive motif discovery in chromatin accessible regions, we derived probable combinatorial binding patterns of TFs that form the core regulatory system of transcription. By combining combinatorial binding events with chromatin events, we have focussed on a methodical reconstruction the cellular events that control gene expression and thereby help us follow the differentiation pathway at a molecular level.

D38: Wassim Abou-Jaoudé, Maximilien Grandclaoudon, Pedro T. Monteiro, Aurélien Naldi, Claudine Chaouiya, Vassili Soumelis and Denis Thieffry. Logical modeling of T-helper cell differentiation and plasticity

Abstract: T helper (CD4+) lymphocytes play a key role in the regulation of immune responses. Potentially faced with a large diversity of microbial pathogens, antigen-inexperienced (naïve) CD4+ T cells differentiate into various T helper (Th) subsets, which secrete distinct sets of cytokines. This differentiation process requires the integration of multiple signals, mainly produced by antigen presenting cells (APC), triggering specific

surface receptors, including the T cell receptor, co-stimulatory molecules, and cytokine receptors. Diverse combinations of these signals lead to the differentiation of naïve T cells into diverse Th subsets, among which Th1, Th2, Treg and Th17 subtypes, tailoring the immune system towards an adaptive response to the encountered pathogen. Noteworthy, recent experimental studies highlight the diversity and plasticity of Th lymphocytes. Our goal is to develop a comprehensive logical model of Th differentiation accounting for observed Th subsets and plasticity in response to cytokine environment. Our current model encompasses 20 signalling pathways, a dozen of transcription factors, and about 30 cytokines, amounting to a hundred of components in total. To cope with this large model size, novel computational methods (implemented in the logical modelling software GINsim) have been used, emphasizing the attractors, corresponding to Th subsets, along with the most important transitions underlying commitment. We further rely on model-checking tools (using the symbolic model checker NuSMV) to efficiently automate the verification of Th subtypes plasticity under specific environmental conditions. Here, we mainly focus on reachability properties between Th cellular subtypes in documented cytokine environments. We assess the consistency of our model by comparing its dynamical properties with published experimental observations dealing with Th differentiation and plasticity. The model reproduces the polarisation of naïve Th cells into various canonical subsets (Th0, Th1, Th2, Th17, iTreg, Th9, Th22, Tfh, in their activated, quiescent or anergic states), under relevant cytokine environments. Specific questions regarding the intracellular regulatory mechanisms underlying Th commitment, as well as Th polarisation properties under poorly studied environmental conditions are further addressed.

D39: Emilia Wysocka, James Snowden, Matthew Page and Ian Simpson. Towards a semi-automated framework of rule-base model creation for neuropsychiatric disease.

Abstract: Systems level approaches are often used to help understand the pathogenesis of complex neuropsychiatric disorders. Commonly, mechanistic modeling of the underlying dysfunction occurs at the level of critical signalling pathways which are popular targets for drug development. However, even with an abundance of information about these pathways, traditional equation-based models have become inadequate where the size, combinatorial complexity of reactions and the variety of post-translational modifications is large. These issues are being addressed by new methods of rule-based modelling, embodied by languages such as BioNetGen and Kappa. Although these approaches have allowed for a massive expansion in size and complexity of the models that can be built their construction remains prohibitively labour intensive, seriously limiting their application. Here we present the development of a tailored framework for the semi-automated construction of rule-based models designed to facilitate the process of rule-based model creation. We integrate and automate access to primary data such as post-translational modification sites, protein and domain interactions and translate these data into the syntax of rule-base language. In order to demonstrate the utility of the framework in a real-world example, we present reconstructed dynamic models of biochemical pathways and protein complexes relevant to Attention Deficit Hyperactivity Disorder (ADHD), many of which are shared with other neurological diseases including Autism and Parkinson's disease. Our targets are derived from static models constructed from protein-protein interaction (PPI) networks, homology relations, pathway analysis, expression data, gene partitions inferred by cluster analysis and mining of experimental literature for kinetic parameters and initial states. In future work, we plan to adopt a modular approach to model construction, which is already encapsulated in rule-based languages (e.g. context-free molecular species) and is independent of the choice of simulation tools used. We will establish repositories of model parts to

facilitate the assembly-like construction of large-scale models, using similar principles to those pioneered in Synthetic Biology.

D40: Azim Dehghani Amirabad and Marcel H. Schulz. Exploiting RNA-Seq data to the fullest: Models for miRNA-transcript target interactions

Abstract: MicroRNAs (miRNA) are small non-coding RNAs which play a critical role in a wide range of biological processes, via post-transcriptional gene regulation. Identifying miRNA targets is a critical step toward elucidating their functions in different diseases. In recent years, several computational methods based on miRNA-mRNA sequence complementarity information have been developed. However the expected false positive rate of sequence based predictions is still large. In addition many target relationships are context-specific. Therefore, most approaches incorporate miRNA-mRNA expression levels to improve prediction accuracy.

Next generation RNA-sequencing (RNA-seq) extends the possibilities of transcriptome profiling to quantitative analysis of expression levels of genes and their isoforms. In this study, we formulate different regression models for inferring miRNA-mRNA interaction networks in cancer using gene and transcript expression levels.

In principle transcript expression levels should allow better prediction of true miRNA binding events. In a thorough comparison we show that this is not always true and that the different formulations can lead to differences in predicted context-specific miRNA-mRNA targets.

D41: Léon-Charles Tranchevent, François-Olivier Desmet, Hussein Mortada, Emilie Chautard, Marion Dubarry, Clara Benoit-Pilven and Didier Auboeuf. A computational platform to predict the functional consequences of alternative splicing variations

Abstract: Alternative splicing allows the production of several protein isoforms from a single gene. These isoforms have different protein sequence, and thus different functional domains. The produced isoforms often have diverse or even antagonistic functions. Furthermore, there is increasing evidence demonstrating that alternative splicing variations contribute to diseases, including myopathies and numerous cancers. In particular, alternative splicing variations have been associated with the phenotypic plasticity of tumor cells during tumor progression and in response to treatment. Exon arrays and RNA-sequencing technologies are frequently used to characterize the alternative splicing events that are specific of a given condition. Here, we describe a computational platform to analyze the results of such experiments in order to detect which protein features are impacted by alternative splicing events.

For a given gene, the gene structure with alternative promoters, polyadenylation sites and exons is displayed. The structure of the associated transcripts is also shown and indicates which exons are included/excluded together with potential start/stop codons and NMD predictions. For each variant, the corresponding protein is displayed and the protein features are indicated. The functional domains covered in the current interface are mainly based on experimental data and include protein domains (e.g., RNA binding domains, catalytic sites), interaction domains (e.g., protein physical interactions), structural elements (e.g., coiled coils, unstructured), sub-cellular localization elements (e.g., NES/NLS, trans-membrane elements, signal peptides), and post-translational modifications (e.g., phosphorylation, acetylation). These data types have been selected based on evidence that protein isoforms have different protein domains, different interaction partners, different 3D structures, are localized in different cell compartments and are subject to different PTMs. It is thus possible for a user to evaluate the effect of an observed splicing event on domain content and therefore derive hypotheses about the impact on the protein function.

Second, analysis are performed on a global scale by considering a list of alternative splicing

events in a particular context. More specifically, we have developed a pipeline for the combinatorial, statistical and visual analysis of exon expression data in the context of networks and pathways. These analyses will allow to identify pathways/networks altered not only owing to global gene expression mis-regulation but also owing to gene splicing alterations. These approaches have been tested and experimentally validated using a breast cancer model of tumor progression. By analyzing 20 breast cancer cell lines, we have also set up and experimentally validated a strategy allowing to better predict the likelihood of targeted therapy resistance.

The tool is accessible at <http://fasterdb.lyon.unicancer.fr/>

D42: Priscila Da Silva Figueiredo Celestino Gomes, Isaure Chauvot de Beauchene, Nicolas Panel, Pedro Geraldo Pascutti, Eric Solary and Luba Tchertanov. Impact Of Oncogenic Mutations On Allosteric Regulation Of Receptor Tyrosine Kinases: Application To The Drugs Design

Abstract: Receptor tyrosine kinases (RTKs) control signal transduction pathways in cells through tightly regulated allosteric mechanisms. Their conformational plasticity enables them to recognize a great number of targets. In solution, RTKs are at equilibrium between various conformations and this equilibrium can be displaced by ligand binding, phosphorylation or by point mutations. In type III RTKs (KIT, FMS, PDGFRs, FLT3), activating mutations inducing oncogenic effects were identified in the juxtamembrane region (JMR) and in the tyrosine kinase (TK) domain. Mutations localized in the activation (A-) loop (D816V in KIT or D802V in FMS) confer in addition resistance to imatinib. We are using in silico approaches for an innovative strategy to identify conformational states specific to oncogenic and/or resistant mutated forms of RTKs and targeting these conformations by small molecules, typically allosteric modulators/inhibitors. First, we characterized the impact of point mutations on crucial regulation segments (JMR, A-loop and the loop proximal to the catalytic (C-) helix) on structure and dynamics of the TK domain in KIT. Two types of effects were observed: local, detected at proximity to the mutation sites, and long-range effects, manifested in regions distant from the point mutations [1]. Second, we studied the molecular determinants of allosteric regulation of KIT and FMS in the native and mutated forms [2]. We were able to describe and modulate the communication pathways between two remote regulatory segments (JMR and A-loop) [3]. A strong correlation between the communication and the structural/dynamical features was established. Further, the described in silico effects were correlated with the auto-activation rate of the mutants and its sensitivity to the drugs (in vitro and in vivo data) [4]. Finally, we studied the interaction of imatinib with KIT and CSF-1R in their native and mutant forms. Using the data obtained from the MD simulations, we were able to characterize the stability of the inhibitor-receptor complexes and detail imatinib affinity in terms of prevalence of hydrogen bonds involving critical active site residues and free energy of binding. We believe our work will open a way for innovative rational strategies for the design and development of novel efficient anti-cancer targeted treatments delivered from experimental evidence and theoretical modeling and simulations.

[1] Laine, E. ; Chauvot de Beauchêne, I. ; Perahia, D. ; Auclair, C. ; Tchertanov, L. PLoS Comput Biol. 2011,6, doi:10.1371/journal.pcbi.1002068.

[2] Da Silva Figueiredo Celestino Gomes, P.; Panel, N.; Laine, E.; Pascutti, P. G.; Solary, E.; Tchertanov, L. PLoS ONE 2014, 9, doi: 10.1371/journal.pone.0097519.

[3] Laine, E.; Auclair, C; Tchertanov, L . PLoS Comput Biol. 2012, 8, doi:10.1371/journal.pcbi.1002661.

[4] Chauvot de Beauchêne, I. ; Alain, A. ; Panel, N. ; Laine, E.; Dubreuil, P ; Tchertanov, L. PLoS Comput Biol. Accepted.

D43: Georgij Arapidi, Igor Fesenko, Konstantin Babalyan, Emile Zakiev, Anna Seredina, Regina Chazigaleeva, Elena Kostrukova, Sergey Kovalchuk, Nikolay Anikanov, Tatiana Semashko, Vadim Govorun and Vadim Ivanov. Identification of small open reading frames with high coding potential in moss *Physcomitrella patens*

Abstract: It has been revealed that small open reading frames (sORFs, up to 100 codons) have the potential to encode biologically active peptides that have regulatory roles in eukaryotic cells (Kastenmayer et al., 2006), (Kondo et al., 2010), (Andrews et al., 2014). In plants, a number of peptides encoded by sORFs play significant roles in various aspects of plant growth and development (Hanada et al., 2012). However, most *ab initio* gene prediction programs are not well suited for identifying sORFs with coding potential. Moreover, existing standard proteomic approaches poorly suited for the identification of proteins less than 10 kDa.

We used prediction program sORFinder (Hanada et al., 2012) to find intergenic regions with high coding potential in the genome of the model object moss *Physcomitrella patens*. High-throughput RNA-Seq by SOLiD 4 genetic analyzer (Life Technologies, Applied Biosystems) and identification of native peptides by TripleTOF 5600 LC-MS/MS (ABSciex) has been carried out on gametophore, protonema and protoplast cells of moss *Physcomitrella patens*. Optimal procedure for endogenous peptide extraction and identification has been worked out to demonstrate translation of sORFs.

Using sORFinder we distinguished 241,228 sORFs within intergenic region with high coding potential. RNA-Seq confirmed transcription of 8,450 sORFs from intergenic region and 16,928 previously known genes of *Physcomitrella patens*. Tandem mass-spectrometry analysis resulted in identification of 18 peptides derived from 12 sORFs within intergenic region, 52 peptides derived from 42 sORFs that were previously thought to be untranslated region of mRNAs and more than 100 peptides from about 100 alternative sORFs within previously known ORFs.

Comparative analyses of sORFs sequences distinguished in moss *Physcomitrella patens* with genomes of other plant species revealed high conservation in terms of synonymous/nonsynonymous substitutions. The report will be discussed further steps to validate the results: overexpression and knockout mutants of coding sORFs, functional categorization and expression under stress.

D44: Amhed Vargas-Velazquez, Pierre-Marie Chiaroni, Morgane Thomas-Chollier and Denis Thieffry. Modelling the interplay between transcriptional regulation and chromatin remodeling during cell differentiation in response to Retinoic Acid

Abstract: Cell differentiation is a highly orchestrated process relying upon the interplay of transcriptional regulators and chromatin remodelling factors. How is this dynamical process coordinated in time and space remains a central question in biology. In this respect, high-throughput datasets on gene expression (microarrays, mRNA-seq), transcription factors binding (ChIP-Seq), chromatin architecture (HiC), as well as on the genomic distribution of epigenetic marks, constitute novel and powerful means to address this question.

Here, we report the construction of a molecular map of the regulatory network controlling mouse embryonic stem cell differentiation in response to Retinoic Acid (RA) induction.

Based on an extensive analysis of relevant publications, this map encompasses the particular different components playing major roles in RA recognition and signalling, transcriptional regulation, and concomitant chromatin remodelling.

In parallel, we have gathered temporal ChIP-seq datasets for RXRa and RAR α , as well as for PolII binding and different chromatin marks, along with transcriptome profiles for the same conditions (five time points, in presence or not of ATRA and of different RAR agonists).

After proper processing (read filtering and mapping, normalisation and peak calling), the resulting peak sets are fed into a multivariate Hidden Markov Model (ChromHMM, <http://compbio.mit.edu/ChromHMM/>) to segment the genome into different regions characterised by specific chromatin state profiles. Regions with similar chromatin mark profiles (in particular those with enhancer or promoter signatures) are then further analysed using motif discovery and pattern matching tools (encompassed in RSAT suite, <http://www.rsat.eu>) to delineate cis-regulatory motifs and potential co-factors involved in RA-induce cell differentiation.

Using the resulting computational predictions along with the molecular map mentioned above, we ultimately aim at building a dynamical model (using GINsim software, <http://www.ginsim.org>) integrating transcriptional and epigenetic regulatory mechanisms, which will be systematically confronted to existing data. Potential discrepancies between model simulations and experimental observations will be exploited to revise the model, until a satisfactory consistency is reached. The resulting model will then be used to assess the effects of unreported perturbations and to design novel experiments.

In conclusion, this study combines different functional genomic datasets and computational approaches to enhance our understanding of mechanisms driving RA-induced cell differentiation. The inclusion of forthcoming chromatin architecture data should further improve the delineation of regulatory mechanisms and hence the model predictive power.

D45: Pauline Traynard, Adrien Fauré, François Fages and Denis Thieffry. Logical modeling of the mammalian cell cycle

Abstract: The molecular networks controlling cell cycle progression in various organisms have been previously modelled, predominantly using differential equations (see in particular the seminal studies by the groups of Tyson, Novák and Goldbeter). However, this approach meets various difficulties as one tries to include additional regulatory components and mechanisms: difficulty to formalise complex regulatory terms, poorly characterised kinetic parameters, numerical difficulties generated by complex, stiff systems, etc. This led to the development of qualitative dynamical models based on Boolean or multilevel frameworks, which are easier to define, simulate, analyse and compose (Fauré et al., Molecular Biosystems, 2009).

Here, we revisit a Boolean model for the core network controlling G/S transition in mammalian cell cycle (Fauré et al. Bioinformatics, 2006), taking into account recent advances in the characterisation of the underlying molecular networks to obtain a better qualitative consistency between model simulations and documented mutants features. In particular, we introduced Skp2, which targets cell cycle control elements, such as p27, and is repressed by the tumour suppressor protein Rb. Furthermore, to supersede the limitations inherent to the Boolean simplifications, we have considered the association of multilevel logical components with key cell cycle regulators, including the tumour suppressor protein Rb. Indeed, it is well established that differently phosphorylated forms of Rb result in different effects on other components of the network, which can be faithfully modelled using a multilevel rather than a Boolean variable.

To evaluate the dynamical properties of the resulting models, we perform synchronous and asynchronous simulations using the software GINsim (<http://www.ginsim.org>), for the wild-type case and documented perturbations (e.g. combinations of component loss- or gain-of-functions). In addition, we have designed a series of temporal logic queries (expressed in the CTL language), which enable an efficient and automatic verification of key dynamical properties (existence of a cyclic attractor or of a stable state, conditions on the order of changes of component levels, etc.), using the popular symbolic model checker NuSMV. This strategy greatly facilitates the dynamical analysis of increasingly detailed and complex cell

cycle models.

Our goal is to obtain a core cell cycle model consistent with the most relevant experimental results on mammalian cells, which will then be used as a module in more comprehensive cellular models, including cross-talks with the circadian clock network and key signalling pathways (e.g. MAPK pathways, see Grieco et al. PLoS Computational Biology, 2013, for a recent logical model), whose deregulation underlies the development of various cancers.

D46: Pauline Traynard and François Fages. A bi-directional coupled model of the cell cycle and the circadian clock

Abstract: Recent studies have put in evidence autonomous self-sustained circadian oscillators in individual fibroblasts, and proved the existence of several molecular links between the circadian clock and the cell cycle. All these interactions establish a control of the cell cycle by the circadian clock, and several models of these couplings have been studied to assess the conditions of entrainment of the cell cycle length by the circadian clock.

However, experimental observations have shown a possible entrainment of the circadian clock by cell divisions, particularly acceleration of the circadian clock by fast divisions. This entrainment cannot be explained by unidirectional models. Here we try to reproduce this entrainment with a differential model of a bi-directional coupling between the circadian clock and the cell cycle, and we investigate the conditions in which both cycles are mutually entrained.

We focus on the control of the cell cycle by the circadian clock through the kinase Wee1, while the reverse coupling appears through the inhibition of clock genes transcription during mitosis. In this respect we use a fully detailed model of the circadian clock (Leloup and Goldbeter, PNAS, 2003) and a simplified model of the mammalian cell cycle (Qu et al, Biophys. J., 2003) focusing on the G2/M transition under the control of Wee1 activity.

The choice of differential equations based modeling facilitates the fitting of quantitative properties of the system such as cycle length and phase shifts between cell divisions and the circadian clock. It is balanced with the classical difficulties associated to this approach brought by numerous and poorly characterized kinetic parameters. We address this issue by specifying the desired properties as constraints formalized with quantitative temporal logic. This formalism provides a flexible language to express complex yet imprecise dynamical properties. We exploit the continuous evaluation of the constraint satisfaction in collaboration with evolutionary algorithms for searching parameter values. This method is implemented in the modeling platform BIOCHAM (<http://contraintes.inria.fr/biocham/>).

Beyond curve fitting, this approach could handle complex conditions of phase shifts and cycle lengths under both wild and perturbed conditions, succeeding in estimating corrected values for up to 50 parameters in order to obtain a model consistent with experimental data. The model is then used to predict the effect of perturbations in the system.

D47: Johannes Barth and Christian Fufezan. Applying novel computational tools to dissect the interwoven light and oxygen effects in the ROS stress response network of *Chlamydomonas reinhardtii* by enhanced quantitative mass spectrometry

Abstract: Oxygenic photosynthesis requires the efficient conversion from light into chemical energy and as a by-product, reactive oxygen species (ROS) and oxygen are constantly produced. Besides their damaging aspect to proteins, nucleic acids and lipids, signaling functions of ROS has also been revealed. Since as much light as possible is harvested, ROS production is inevitably increased, which limits crop and algal biomass production yields. A comprehensive understanding of the ROS network is difficult because ROS, light and oxygen responses are entangled. Thus dissecting those responses is an important step in understanding

the ROS network. This is essential to satisfy the increased need for biofuels and more importantly to address and overcome global hunger. Therefore we applied a comprehensive LC-MS/MS based quantitative proteomic analysis to dissect the light, oxygen and ROS networks in *Chlamydomonas reinhardtii* employing a novel interlinked experimental setup. The analysis required that we develop a series of novel computational tools (Barth et al. 2014). These tools include the piqDB framework (the protein identification and quantification database framework), which is based on mongoDB and includes several tools for the analysis, evaluation and validation of experiments, e.g. a routine for high quality retention time alignment of LC-MS/MS runs. A significantly higher proteome and protein sequence coverage over all experimental condition was achieved by using the synergistic combination of all our tools. Furthermore, an our hierarchical clustering approach (pyGCluster, Jaeger et al. 2014) was able to reveal distinct groups of co-regulated proteins. From the obtained co-regulated communities, the most interesting were a) a combined light and oxygen dependent induction, e.g. related to ROS production or inactivation via O₂ as seen in the ~30% induction of all Calvin Bensons cycle enzymes, b) a regulation that reacts to light and oxygen in an additive effect, where light is the obligatory first step, e.g. GRX1 and PRX4, c) a light induced and oxygen independent regulation, e.g. LhcSR3 and carbon concentrating mechanisms and d) a down regulation of proteins in anaerobic high light conditions, e.g. PSI assembly factors or chlorophyll biosynthesis proteins (GUN4). ROS production and induced damage is low under such an anaerobic highlight condition. As a result this could be explained e.g. by a down regulation of proteins if their turnover is reduced via product feedback mechanisms. Further projects will focus on the influence of higher O₂ concentrations and characterization of key-players in the ROS response network. This and consecutive studies using *C. reinhardtii* contributes further to the understanding of the general ROS response network and the improves the knowledge about photosynthesis.

D48: Shelly Mahlab and Michal Linial. miRNA-mRNA interactions: Probabilistic and dynamic perspectives

Abstract: MicroRNAs (miRNAs) are short non-coding RNAs that negatively regulate gene expression post-transcriptionally in healthy and diseased tissues. In human, there are 20,000 coding mRNAs and as many as 2000 matured miRNAs. It is estimated that about 50% of the mRNAs are targeted by miRNAs. The mechanism for miRNA-target recognition is the base-pair complementarity where the elementary unit is a miRNA binding site (MBS) at the 3'-UTR of the mRNA. The identity of MBS is mostly from prediction programs that incorporate features of complementarity, evolutionary conservation, free energy and binding accessibility. Importantly, the interaction of miRNAs with their targets creates a complex network where one miRNA can recognize multiple MBS (in tens or hundreds different mRNAs), and each mRNA can be occupied by one of more miRNAs on multiple MBS. The dynamics of such complex network is poorly understood.

This study is our attempt to develop a generic probabilistic framework that takes into account the distributions of miRNAs and mRNAs in cells. In order to mimic the nature of the network, we assumed the following: (i) miRNAs are stable, therefore their total amount remains constant; (ii) following binding by miRNA the mRNA is subjected to degradation. Thus, over time, the total amount of mRNA is reduced; (iii) overlapping MBS cannot be occupied. Our simulation is composed of a series of steps that mimic transient cell states. Each step considers the availability of free miRNA that can engage into interaction with MBSs on the available targets. In each of the iterations, we randomly choose miRNA according to the distribution of all miRNAs. A probabilistic model based on experimental data is used to estimate the success of the interaction. Once a successful interaction occurs, the cell state is updated in order to reflect the changes in the free and occupied miRNA and mRNAs. All

together we performed 100,000 iteration steps for capturing the cell dynamics. The parameters used in the simulations derived from experimental data: (i) the distribution of mRNAs and miRNAs were obtained from HEK-293 cells deep sequencing measurements; (ii) initial cell state was based on data from the CLASH (crosslinking, ligation, and sequencing of hybrids) methodology. CLASH detects the physical hybrid of miRNA and its mRNAs; (iii) the numbers of miRNAs and mRNA molecules in cells were literature-based. The results of the simulations for a wide range of parameters are extremely robust. The results were almost insensitive to the initial states or the mRNA' degradation rate. We are currently repeating these simulations using less stringent probabilities for miRNA-mRNA interactions. Furthermore, we are comparing the simulation results with miRNA overexpression data. In summary, we present a flexible platform for capturing the combinatorial and stochastic nature of miRNA regulation in a cell.

D49: Monika Kurpas, Katarzyna Jonak and Krzysztof Puszynski. The novel mathematical model of ATR-p53-Wip1 signaling pathway: studies on prediction of cellular response to DNA damages

Abstract: Eukaryotic cells are daily exposed to the stress agents, like UVC radiation, resulting in DNA damages and development of several disorders, such as cancer. For a better understanding and prediction of the cell behavior, the mathematical model of ATR pathway, a detector of DNA breakages caused by UV, was developed. The ATR system was connected to p53 pathway, which plays a role during cancer development, and to Wip1 phosphatase. The ATR-p53-Wip1 mathematical model was built using Haseltine-Rawlings postulate, and the stochastic and deterministic simulations were performed, in order to analyze the response of a single cell and population of cells to UVC radiation. Our results show that the apoptotic threshold, where more than half of cells die, is equal to 23.5 J/m². The threshold shifts when the specified proteins involved in the pathway are blocked or reduced (especially Wip1 production blockade and ATR protein reduction). The results indicate that ATR is an effective system for detection of DNA breakages and results in strong amplification signal of p53 and Wip1. The absence of Wip1 module is very noticeable for the cell, what may make it an important future drug target used against cancer.

D50: Costas Bouyioukos, Ivan Junier and François Képès. Genome REgulatory Architecture Tools (GREAT). The SCAN suite for the detection of regular patterns along genomes. GREAT:SCAN

Abstract: Recent advances in genomics, transcriptomics and genome structural biology have revealed significant insights on the non-random arrangement of genes on one hand, and on the interplay between transcription, gene position and genome structure on the other. Here we present the first implementation of a software suite designed to perform a systematic and integrated analysis of regular patterns along genomes. The suite is based on an algorithm to detect periodicities and it provides an easy to use interface to execute complicated analyses of regular patterns and visualise results. The suite comprises two software tools. GREAT:SCAN:patterns, a package for systematic study of periodic patterns, clustering and visualisation and GREAT:SCAN:integrate, a novel computational process which integrates regularities along multiple transcription factors (TFs) and chromosomes. GREAT:SCAN:patterns systematically analyses every predicted period and calculates weighted (exact) p-values. The first step returns a rank of periods based on the exact p-value. On the second step a clustering algorithm detects clusters of genes that are in-phase on the modulo period coordinates and provides an insight of possible local spatial proximity of

genes. On the last step a more fine-tuned search for regularities is taking place based on a variable size sliding window which detects periods on specific domains of the chromosome. In this work, we present a complete analysis of the 6 major TFs of *E. coli* and report preliminary results that regions of periodic arrangement are associated with the macro-domain organisation of this bacterial genome.

The software is developed to detect periods on co-regulated genes however it can work with any gene set of interest as well as with any set of genomic positions of interest, including but not limited to chip-seq data.

GREAT:SCAN:integrate: is a computational process automatically performs a consolidated and integrated analysis of periodic patterns on multiple TFs and/or multiple chromosomes. It consists of a series of seven steps. Initially, periods are detected on all the groups of co-regulated genes and then a couple of integration steps on the TF and the chromosome level consolidates periods and extends overlapping extremities. Finally, the process is searching for integral multiple periods and collects them all together to form families of harmonics with their periodic extremities extended. The result of the final step is visualised as a set of periodic regions that span chromosomes and the results of each intermediate step are stored in a database for further analysis and/or visualisation.

The seven individual steps of the process are formally described in the poster. We also present initial evidence of an application of the GREAT:SCAN:integrate process on the yeast *Saccharomyces cerevisiae* TF network identifying common periods, harmonics and significant degree of overlap between the master transcription regulators of yeast.

D51: Wojciech Bensz and Krzysztof Puszynski. A stochastic model of the p53 ubiquitination system.

Abstract: p53 is one of the most widely investigated proteins known to science as it is transcription factor responsible for induction of processes crucial for cell fate such as DNA repair, cell cycle arrest and apoptosis, hence also one of the most crucial tumor suppressors. Mdm2 protein is the main negative regulator of p53, acting via mechanism of ubiquitination - a post-translational protein modification usually designating proteins for efficient, proteasome dependent degradation. Constant fast degradation of p53 protein maintains its low level in normal cells. Monoubiquitination of p53 was proven to mark the protein for nuclear export and translocation to mitochondria, where it stimulates apoptosis initiation by interacting directly with Bcl-2 family proteins[1].

New stochastic model of p53 protein ubiquitination process was proposed, basing on earlier p53/Mdm2 feedback loop model[2] and more recent discoveries. Degradation reactions of the original model were replaced by ubiquitination and autoubiquitination reactions catalysed by Mdm2, ultimately leading to degradation of polyubiquitinated proteins. Deubiquitination reactions catalysed by HAUSP protein were additionally incorporated. Stochastic nature of DNA double strand lesions generation, gene copy activation and deactivation was included in the model by applying hybrid approach based on Haseltine-Rawlings postulate. All other variables were modeled in deterministic fashion, using ODE. Incorporation of stochastic effects allowed for single cell response simulations and estimation of apoptotic and surviving fractions of cells under different conditions. Distinction between mono- and polyubiquitinated forms of p53 allowed to study the role of HAUSP dependent stoichiometry of the two for cell fate determination.

Simulation analyses of cells irradiated with different doses of ionizing radiation and expressing different levels of HAUSP deubiquitinase showed that control of the latter could be possibly useful therapeutic strategy. Results obtained for irradiation and HAUSP knockout or overexpression alone are in agreement with available knowledge derived from wet laboratory experiments[1,3,4] and also suggest that HAUSP overexpression might lead to

increased apoptotic fraction of cells irradiated with small IR doses.

This work was supported by National Science Centre – decision no. DEC-

2012/05/D/ST7/02072

References

1. K. Becker, N.D. Marchenko, G. Palacios, U.M. Moll, A role of HAUSP in tumor suppression in a human colon carcinoma xenograft model, *Cell Cycle*, 7(9), 2008, 1205-1213
2. K. Puszynski, B. Hat, T. Lipniacki, Oscillations and bistability in the stochastic model of p53 regulation, *Journal of Theoretical Biology* 254, 2008, 452– 465
3. D.R. Green, G. Kroemer, Cytoplasmic Functions of the Tumor Suppressor p53, *Nature*, 58(7242), 2009, 1127
4. N. Kon, Y. Kobayashi, M. Li, C.L. Brooks, T. Ludwig, W. Gu. Inactivation of HAUSP in vivo modulates p53 function. *Oncogene*, 29(9), 2010, 1270–1279

D52: Karsten Kuritz and Frank Allgöwer. Determining cell-cycle induced variations from snap-shot data sets

Abstract: We here present a method and the underlying theory which allows the determination of cell-cycle induced variations of proteins levels from single population snap-shot measurements like FACS or fluorescence microscopy.

The method is motivated by the observation that the response of cell populations to external stimuli is often not homogeneous.

One reason for a heterogeneous response may be the fact that the cells in a population are in different cell-cycle stages during the stimulation. Key-players of the stimulus response might be up- or down regulated in certain stages of the cell cycle, resulting in an increased or diminished reaction.

Assessing the variation of protein abundance along the cell cycle is a non-trivial task and typical experimental methods require cell synchronization or live cell tracking.

The here presented method is based on the assumption that the movement of a single cell through the cell-cycle follows a stochastic differential equation in one dimension. The time evolution of the number density of a cell population along the cell-cycle is then given by a partial differential equation, the Fokker-Planck equation. One can derive an equation for the cell speed along the cell-cycle which is parameterized by the population growth rate and the cell number density along the cell cycle. The number density can be extracted from FACS experiments which are typically used to determine the cell stage distribution in a cell population. Every measured cell is then mapped to its position in the cell-cycle and any additional readout can likewise be associated with a specific cell-cycle position.

A single FACS experiment is thereby sufficient to determine the time evolution of protein levels along the cell-cycle.

D53: Gabriella Sferra, Federica Fratini, Marta Ponzi and Elisabetta Pizzi. Dynamics of *P. falciparum* protein-protein interaction network: the membrane microdomain interactome.

Abstract: In recent years, several computational methods have been developed to predict protein-protein interaction at a genome-wide level. We applied a Bayesian approach, which integrates data from diverse sources, to reconstruct a probabilistic global interactome of *Plasmodium falciparum* – the causative agent of human malaria - using genomic, transcriptomic and proteomic data.

We generated novel genomic data (phylogenetic profiles and rosetta-stone) and processed transcriptomic ones. In particular, we performed a re-assessment of the phylogenetic profile method proposing a new strategy to select reference genomes and adopting the distance correlation as novel measure of similarity between phylogenetic profiles. We also produced a

new set of rosetta-stone fusion genes on the basis of a large set of genomes used as reference set. Furthermore, diverse transcriptomic data, have been organized to obtain a unique profile covering the entire erythrocytic Plasmodium life-cycle (asexual and sexual stages).

To gain insights on protein-protein interactions, which occur in specific subcellular compartments, we developed an experimental procedure to identify proteins associated with cholesterol-rich membrane microdomains. These specialized membrane compartments are extremely dynamic and play a key role in parasite development and invasion.

Proteomic data from these microdomains purified at different time-points of the parasite life-cycle have been produced and used to map protein components on the predicted global interactome. The resulting microdomain sub-networks were analyzed.