Contributed abstracts

Overview
Home
French site
IPG History
Scientific council
Organizing committee
Sponsors
Previous meetings
IPG videos
IPG posters

Schedule
Invited talks
Contributed abstracts
Detailed schedule
Satellite meetings

How to submit?
Call for papers
Abstract submission

How to attend?
Registration
Local arrangements

Natural histone code polymorphisms in yeast

Muniyandi Nagarajan, Jean-Baptiste Veyrieras, Maud De Dieuleveult, Hélène Bottin, Lars M. Steinmetz, Xavier Gidrol, Gaël Yvert

Laboratoire de Biologie Moléculaire de la Cellule, Ecole Normale Supérieure de Lyon, CNRS, 46 Allée d'Italie, Lyon, F-69007, France; Université Lyon 1, Lyon, F-69003, France

Although the histone code is being systematically sequenced from various tissues, stages or disease, its intra-species natural variation has never been investigated. We describe here that the epigenomic sequence of histone H3 acetylation at Lysine 14 differs greatly between two unrelated strains of S. cerevisiae, with 5,303 of 62,824 interrogated nucleosomes differing at 1% False Discovery Rate. Density of these "Single-Nucleosomal EpiPolymorphisms" (SNEP) was negatively correlated to the density of DNA polymorphisms, and was significantly higher among nucleosomes covering regulatory sequences. SNEPs were not correlated with inter-strain differences in gene expression but with differences in transcriptional plasticity. Our observations raise essential questions on the origin of this diversity, its relevance to gene x environment interactions, and the need for personalized epigenomics and population epigenetics.

Footprints of inversions at present and past pseudoautosomal boundaries in human sex chromosomes

Claire Lemaitre, Marilia Braga, Christian Gautier, Marie-France Sagot, Eric Tannier and Gabriel AB Marais

Université de Lyon, F-69000, Lyon ; Université Lyon 1 ; CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France

Inversions are known to occur repeatedly in genome evolution but their evolutionary significance remains obscure. It has been suggested that five large inversions might have affected the human Y chromosome and have reduced the pseudoautosomal region (the recombination zone for X and Y) in five steps, which have produced the five evolutionary strata (chromosomal domains with similar X-Y level of divergence) that we can observe in humans. Here we tested this idea by focusing on a region (= XAR, which includes the recent strata) that has conserved gene order between human

X and its chicken homologue, which means that detected rearrangements between X and Y in this region must have occurred in Y. First, we evaluated the number of possible scenarios of Y inversions given the gene orders in XAR and in its Y homolog. We found that there are many scenarios in which Y inversions coincide with strata boundaries. Using simulations with randomly distributed inversions on Y, we suggest that it is unlikely that this pattern has emerged just by chance. Another set of simulations with Y inversions occurring only among strata seems to indicate that recent strata have arisen by inversions on the Y. We then looked for footprints of Y inversions at strata boundaries using a method to detect and analyse rearrangement breakpoints precisely. We found clear evidence of two Y inversions coinciding with strata 4 and 5 boundaries. Our results strongly support the view that in humans, the recent strata have arisen by inversion on the Y and suggest that inversions have played a major role in the differentiation of our sex chromosomes.

Developmental patterning by mechanical signals in Arabidopsis

Olivier Hamant1,6, Marcus Heisler2,6, Henrik Jönsson3,6, Pawel Krupinski3, Magalie Uyttewaal1, Plamen Bokov4,5, Francis Corson4, Patrik Sahlin3, Arezki Boudaoud4, Elliot Meyerowitz2, Yves Couder5, Jan Traas1

1Université de Lyon, INRA, CNRS, ENS, 46 Allée d'Italie, 69364 Lyon Cedex 07, France 2Division of Biology, California Institute of Technology, Pasadena, California 91125, USA 3Computational Biology and Biological Physics Group, Department of Theoretical

A central question in developmental biology is if and how mechanical forces experienced by cells in tissues serve as cues for cellular behavior, and thereby regulate morphogenesis. Here we show that mechanical properties of the Arabidopsis shoot apex depend on the microtubule cytoskeleton which, in turn, is regulated by mechanical stress. Through a combination of experiments and modeling we show that this feedback loop is sufficient to account for the coordinated patterns of microtubule arrays observed in epidermal cells, and for patterns of apical morphogenesis. These results reveal a mechanism by which plant tissues generate form through a feedback loop in which stress regulates microtubules, and microtubules (through controlling the oriented deposition of cellulose fibers) regulate response to stress.

TvMULEs, Trichomonas vaginalis Mutator-like elements, have been amplified by self-mobilization and recent local duplications

Fabrício R. Lopes 1, Joana C. Silva 2, Marlene Benchimol 3, Gustavo G. L. Costa 4, Gonçalo A. G. Pereira 4, Claudia M. A. Carareto 1

1 UNESP - São Paulo State University, São José do Rio Preto, São Paulo, Brazil; 2 University of Maryland School of Medicine, Baltimore, Maryland, USA; 3 USU - Santa

Ursula University, Rio de Janeiro, RJ, Brazil; 4 UNICAMP - State University of Campinas,

Mutator elements, firstly thought to be exclusive of plants, have been recently characterized in few organisms, such as in fungi and early-diverging amoebas. Here, we describe and characterize four families of Mutator-like elements in a new eukaryotic group, the Parabasalids. The draft genome sequence of the G3 strain of Trichomonas vaginalis was obtained from the website of The Institute for Genomic Research. This draft, based on ~7.2-fold coverage of the genome, consists of 17,290 scaffolds, representing ~160 Mbp (Carlton et al. 2007). Sequence similarity searches using the four consensus sequences of TvMULEs as query against the T. vaginalis genome were performed using BLASTN (Altschul et al. 1990), with parameters E=e-20, V=10,000 and B=10,000. Significant matches were required to be >200 bp long and display 80% identity. Phylogenetic analyses were performed using Mutator sequences of T. vaginalis and Candida albicans and of related TE families from a variety of taxa, including plants, fungi, protists and bacteria. Analysis of host distribution and transcriptional activity were done in seven Trichomonad species. The Trichomonas vaginalis Mutator-like elements, or TvMULEs, are active in T. vaginalis and patchily distributed among 12 trichomonad species and isolates. Despite their relatively distinctive amino acid composition, the inclusion of the repeats Tvmu1, Tvmu2, Tvmu3 and Tvmu4 into the Mutator superfamily is justified by sequence, structural and phylogenetic analyses. In addition, we identify three new TvMULE-related sequences in the genome sequence of Candida albicans. While Tvmu1 is a member of the MuDR clade, predominantly from plants, the other three TvMULEs, together with the C. albicans elements, represent a new and quite distinct Mutator lineage, which we named TvCaMULEs. Our analyses suggest that while local duplications of the host genome may have contributed to the expansion of the Tvmu1 and Tvmu2 families to several dozen copies, Tvmu3 and Tvmu4 most likely increased in copy number by self-mobilization. These findings expand the taxonomic distribution and the range of functional motif architectures of MULEs among eukaryotes. The remarkably recent common ancestry of each TE family in the T. vaginalis genome is attested to by the high copy number and nearly complete within-family sequence similarity of these TvMULEs, features that are shared with the other ~55 repeat families identified in the T. vaginalis genome. Moreover, given the recency of the TE expansion in T. vaginalis, their long-term effect on the survival of the species is as yet unclear. It is possible that, with each TE family expansion, this species is steadily proceeding to extinction. These features highlight this species' unmatched placement as an ideal system in which to characterize new transposable elements. Financial Support: CNPq

Forced periodic expression of G1 cyclins phase-locks the budding yeast cell cycle: implications for integration of cell size control

Gilles Charvin, Fred Cross, Eric Siggia

The Rockefeller University, NY, USA ; Laboratoire Joliot-Curie, Ecole Normale Superieure, Lyon, France

Phase-locking (frequency entrainment) of an oscillator, in which a periodic extrinsic signal drives oscillations at a frequency different from the unperturbed frequency, is a useful property for study of oscillator stability and structure. The cell cycle is frequently described as a biochemical oscillator; however, since this oscillator is tied to key biological events such as DNA replication and segregation, and to cell growth (cell mass

increase), it is unclear whether phase locking is possible for the cell cycle oscillator. We found that forced periodic expression of the G1 cyclin CLN2 phase locks the cell cycle of budding yeast over a range of extrinsic periods in an exponentially growing monolayer culture. We characterize the behavior of cells in a pedigree using a return map to determine the efficiency of entrainment to the externally controlled pulse. We quantify differences between mothers and daughters and how synchronization of an expanding population differs from synchronization of a single oscillator. Mothers only lock intermittently while daughters lock completely and in a different period range than mothers. We can explain quantitative features of phase locking locking in both cell types with an analytically solvable model based on cell size control and how mass is partitioned between mother and daughter cells. A key prediction of this model is that size control can occur not only in G1, but also later in the cell cycle under the appropriate conditions; this prediction is confirmed in our experimental data. Our results provide quantitative insight into how cell size is integrated with the cell cycle oscillator.

A multi-scale model of erythropoiesis

I. Demin, F. Crauste, O. Gandrillon and V. Volpert

Université Claude Bernard Lyon 1

A multi-scale mathematical model of erythropoiesis is proposed in which erythroid progenitors are supposed to be able to self-renew. Three cellular processes control erythropoiesis: self-renewal, differentiation and apoptosis. We describe these processes and the regulatory network that governs them. We give a detailed description of the regulatory network, through biochemical reactions involving the different proteins that compose this network. Two proteins (ERK and Fas) are considered as the basic proteins participating in this regulation. All erythroid progenitors are divided into several sub-populations depending on their maturity level. Feedback regulations by erythropoietin (Epo), glucocorticoids (GCs) and Fas ligand (FasL) are introduced in the model. The model consists of a system of ordinary differential equations describing intracellular protein concentration evolution and cell population dynamics. We study steady states and their stability. We carry out computer simulations of an anemia situation and analyze the results.

Discrete multi-clock modelling of biological systems

Nolwenn Le Meur, Michel Le Borgne, Jérémy Gruel, Nathalie Théret

INSERM U620 - IRISA Symbiose

Modelling biological processes is a step toward a better understanding of the complex behaviours of living systems. Among several formalisms, discrete modelling is especially suitable for the integration of qualitative biological data that constitute most of the regulatory interaction data currently available. Another advantage of these models is that the description of biological interactions is flexible and rather easy. Each entity (gene, protein, ...) is represented by a finite-state variable and their values

reflect the different biological properties of the molecule. For instance, in discrete logical models, the value indicates whether a molecule is active or not. The evolution equation between two states of a molecule is then conditioned by the state of other variables of the system. However, one disadvantage of these models is the lack of dynamic. Time is absent of the model or defined a priori: either all possible transitions are fired at the same time (synchronous) or only one transition at a time is allowed (asynchronous). Such crude modelling of time is not satisfactory to model biological systems. Indeed different biological reactions have different durations, which give a subtler sequencing of biological events. In an attempt to get finer descriptions of time, some authors introduced the notion of priorities among transitions [1]. Although, they are applied at the simulation steps and thus remain part of the interpretation. Here we propose a new formalism to include time in the model. This formalism is inspired by the formal models underlying real time programming languages such as Esterel, Lustre and Signal. The time is the logical time used in computer science: it does not correspond to the duration of events but to their relative sequencing. The absence of an event is always relative to the presence of (at least) another event. This allows the description of several biological signals with different clocks, i.e., multi-clock systems. In this formalism, we essentially use two concepts: signals and states. A signal is an event bearing a value (e.g., true or false). A state acts as a memory: it can be read at any time and be modified by a signal when present. Without constraints on clocks, all interlacing are allowed. We thus introduce constraints on clocks to specify temporal behaviours. We include a synchronous and an exclusive instruction, allowing synchronous and asynchronous transitions to be evaluated. At that time, we have designed a language, along with a compiler and a simulator, for specifying Boolean models with temporal constraints. To illustrate our approach we modelled the core network controlling the mammalian cell cycle using the model published by [1] as a benchmark to test and validate our approach. For instance, we verified that our model could mimic the same behaviours. Our next goal is to integrate the signalling pathways involved in the control of the mammalian cell cycle. To that aims, we will extend the proposed formalism, develop the specification of temporal constraints, and implement formal methods such as model checking. [1] Faure et al. (2006). Bioinformatics, 22: e124-e131.

Flexible oligonucleotide design for custom applications using tiling microarray

Laurent Jourdren1,2, Aurélie Duclos1,2, and Stéphane Le Crom1,2,3

1 INSERM, CNRS, IFR36, Plate-forme Transcriptome, 46 rue d'Ulm, 75230 Paris Cedex 05, France 2 École normale supérieure, 46 rue d'Ulm, 75230, Paris Cedex 05, France 3 INSERM, U784, 46 rue d'Ulm, 75230 Paris Cedex 05, France

The way we work with microarray has changed with the increase in feature resolution, the availability of multipack format and the flexibility of custom slide production from commercial providers. A wide range of applications is now accessible from the fine tiling analysis of a specific sub-region from larger genomes to whole genome analyses of microbial organisms. The main difficulty that must be faced for such custom microarray applications is the oligonucleotide design step. A lot of parameters have to be taken into account to ensure that the selected oligonucleotides offer the best specificity and sensitivity. In addition, with tiling microarrays, the even distribution of oligonucleotides has to be kept in mind. This is an optimisation problem between the selection of the most centred probe and the best sensitive and specific one in each genomic window selected. Several tools have been developed to help with this design step. Each program works with one specific goal either optimising the even distribution

of the tiling path or favouring the selection of the best specific probe. Finally, it is quite difficult to customize the parameters leading to the selection of one probe by window. To work around this limitation, we developed a new tiling oligonulceotide design solution based on an evaluation-selection procedure. For each target organism, we design all possible oligonucleotides all along the genome. Then, for each oligonucleotide, we calculated several parameters linked to probe sensitivity and specificity. Next, we apply filters in order to reject wrong probes. Finally, for each window, a position score is calculated, an oligonucleotide score is computed by applying weights to parameters, and a global score is obtained by merging the position and the oligonucleotide score. In each window we then select the probe with the best global score. We test this design algorithm on a custom tiling array example with the goal to design one probe every 150 b along a 33 Mb microorganism genome. Our solution achieves the best compromise between even distribution of the probes and parameter homogeneity compared to other available software. In addition, our methodology has two main advantages. First, calculated scores are associated to each selected probe. These scores are helpful during data analysis, where correlations can be done between the strength of the signal measured on the array and the oligonucleotide quality. Second, our design approach is versatile. We can choose between the best quality design by selected only probes with a good oligonucleotide score. On the contrary, we can advantage the even distribution of probes by favouring the position score. Finally, the implemented solution we offer is flexible. We designed the program around open modules, parameter calculation and filters, so that one can create is own module to add new property to the design program. With our solution, we offer a flexible design for tiling microarray. Its open organisation offers several ways to adapt the design to everyone needs with the goal to always achieve the selection of the best oligonucleotide.

The v-erbA oncogene induces variations in ribosomal protein gene expression.

Lefebvre, AT., Gonin-Giraud, S., Madjar, JJ., Gandrillon, O.

CGMC (Centre Génétique Moléculaire et Cellulaire)

The Avian Erythroblastosis Virus (AEV), which carries the two oncogenes v-erbA and v-erbB, induces erythroleukemia and sarcomas in chickens. The v-erbA oncogene, a mutated oncogenic form of the thyroid hormone receptor T3Rα, transforms erythroid progenitors by blocking their differentiation and freezing them in a state of self-renewal. In order to identify the v-erbA target genes involved in the transformation process, the transcriptome of avian erythroid progenitors cells (T2ECs) expressing either v-erbA or a non transforming point mutant form of v-erbA were compared by SAGE (Serial Analysis of Gene Expression) and qPCR (Quantitative Polymerase Chain Reaction). Among the genes the expression of which varies between cells expressing the wild-type and the mutant form of v-erbA, several encode ribosomal proteins (Bresson et al., 2007). Only some, but not all, mRNA encoding ribosomal proteins were shown to be affected. These results suggest that v-erbA could modulate the composition of ribosomes and / or modulate the extraribosomal functions of several ribosomal proteins by regulating their expression at the protein level. We therefore decided to complete our transcriptomic analyses with proteomic approaches. For this, we analyzed the level of ribosomal proteins associated to ribosomes by 2D-DIGE (Two Dimensional Difference Gel Electropheresis). Our preliminary results suggest that the level of several ribosomal proteins associated to ribosomes indeed display marked changes in T2ECs expressing the oncogenic form of v-erbA. Interestingly, we observed both some proteins being more abundant and some being less abundant in the cells expressing v-erbA, in contrast to mRNA level where only

downregulation had been observed. In order to confirm those 2D findings, we investigated the level of the ribosomal proteins S3 and S3a by western blot. We observed that v-erbA expression leads to a significant decrease in the level of S3a but not of S3. Altogether, our results suggest the involvement of ribosomal proteins in the transformation process induced by v-erbA. In the future, we will confirm our preliminary data on the differential composition of ribosomes, and analyze the role that this variations might play in the transformation process. Bresson, C., Keime, C., Faure, C., Letrillard, Y., Barbado, M., Sanfilippo, S., Benhra, N., Gandrillon, O. and Gonin-Giraud, S. (2007) Large-scale analysis by SAGE reveals new mechanisms of v-erbA oncogene action. BMC Genomics, 8, 390.

Structure and evolution of MER2-like DNA transposable elements in fish genomes

Frédéric BRUNET, Laurie NAVE, Audrey LOISEAU, Christina SCHULTHEIS, Jean-Nicolas VOLFF

Institut de Génomique Fonctionnelle de Lyon

DNA-based transposable elements (TEs) and retroelements are found in the genome of almost all organisms. In vertebrates, TEs have been particularly well described in mammals (human and mouse) but their impact on genome evolution, biodiversity and speciation in fish remains to be studied. Five fish genomes are available to date, the zebrafish (Danio rerio), a cypriniform that is about 280 MYA distant from the four percomorphs, the medaka (Oryzias latipes), the stickleback (Gasterosteus aculeatus) and the two tetraodontiforms fugu (Takifugu rubripes) and tetraodon (Tetraodon nigroviridis). Comparing transposable elements in these species, determining their genomic location, showing if they are conserved or not and if they play a role in the evolution are thus essential to understand the evolution of fish genomes. We have used RepeatMasker for establishing an initial catalogue of TEs in fish genomes. We focused our attention onto MER2 type DNA-based transposable elements, which present a high copy number in the genomes analyzed. MER2-type transposons are particularly composed of non-autonomous MER2 sequences, some miniature inverted TEs derived from autonomous Tigger transposons. We found that the distribution of MER2-type elements is different among species. We will describe these elements from different fish species, with evidence for recent transposition activity.

Sex determination and sex chromosome evolution in the platyfish Xiphophorus maculatus

Astrid Böhne, Christina Schultheis, Delphine Galiana-Arnoux, Marie Sémon, Amandine Darras, Frédéric Brunet, Béatrice Ségurens, Arnaud Couloux, Sylvie Bernard-Samain, Catherine Ozouf-Costaz, Manfred Schartl and Jean-Nicolas Volff

IGFL, ENS de Lyon

Sex-determining systems in birds and mammals are widely conserved, with strongly differentiated heteromorphic sex chromosomes at very advanced stages of evolution.

Such sex chromosomes are not informative concerning the initial steps leading to the molecular differentiation between different types of gonosomes. In contrast, most fish species possess homomorphic sex chromosomes at early stages of evolution, with polymorphic and fast-evolving sex-determining systems. Genetic sex determination in fish is possibly subject to frequent switch of master sex-determining genes linked with the creation of novel sex chromosomes. The reasons for this phenomenon remain unknown and the genes involved are still to be characterized. Comparative analysis of sex-determining systems and sex chromosomes in different fish species might help to understand the molecular and evolutionary mechanisms driving the diversity of sex determination and to obtain valuable information on primary steps in sex chromosome evolution. The platyfish Xiphophorus maculatus is an established model for the study of sex determination. BAC (bacterial artificial chromosome) contigs covering the sex-determining (SD) region on the X and the Y chromosomes have been assembled using sex-chromosomal genes and molecular markers as starting points for chromosome walking. Through the sequencing of more than 40 BAC clones, we have obtained new information on the molecular structure of the SD region and identified about 60 genes located on the X and Y chromosomes of the platyfish. Some of these genes present a gonad-specific expression and are candidates for the master sex-determining gene. Initial steps of molecular differentiation between the X and Y chromosome in the SD-region were detected as inversions, duplications, transpositions and Y-specific repeat amplification. No synteny was detected with sex chromosomes from other fish and vertebrates, suggesting independent evolutionary origins.

Development of the Acyrthosiphon pisum Cyc database (ApsCyc): from genome sequence to metabolic network analyses

Augusto Vellozo 1,2,3*, Stefano Colella 1,3*, Ludovic Cottret 2,3, Gérard Febvay 1,3, Federica Calevro 1,3, Yvan Rahbé 1,3, Marie France Sagot 2,3 and Hubert Charles 1,3

1 Université de Lyon, INRA, INSA-Lyon, IFR-41, UMR203 BF2I, Biologie Fonctionnelle Insectes et Interactions, 20 ave A. Einstein, F-69621 Villeurbanne, France; 2 Université de Lyon, F-69000, Lyon; Université Lyon 1; CNRS, UMR5558, Laboratoire de Biométrie

The pea aphid (Acyrthosiphon pisum) genome sequence was recently released and we generated a dedicated BioCyc database (ApsCyc). As the genome annotation will be evolving over time, we developed an automated annotation management system to allow the integration of the latest sequence annotation information into the metabolic network reconstruction and analysis. Our data management system is centred on an ad hoc modified BioSQL database, complemented by a set of Java scripts to import and export relevant information. Data from GeneBank and from different metabolic gene annotation tools (such as KAAS, PRIAM, Annot8r) are collected into the database and later extracted to generate a complete input file to build and/or update ApsCyc using the "Pathway tools" software (BioCyc). ApsCyc offers a framework for the analysis of the integrated metabolic network shared between the aphid and its symbiotic bacterium (Buchnera aphidicola), for which a BioCyc database already exists. Furthermore, the ApsCyc database is a key resource for computational systems biology research and its open platform formats (BioPAX, SBML) will also allow the integration of the data into other tools to perform complex genomics data analysis.

In silico investigation of ADAM12 effect on TGF-beta receptors trafficking.

J. Gruel, M. Le Borgne, N. Le Meur and N. Théret.

INSERM / IRISA

Dynamics and regulation of the TGF-beta signaling pathways are central in many complex cellular processes such as growth, apoptosis, differentiation and proliferation. This growth factor acts through activation of a heteromeric complex of two types of transmembrane Ser/Thr kinases, TGF-beta receptor I and TGF-beta receptor II, phosphorylates the receptor-regulated Smad that governs transcriptional response. We previously demonstrated that ADAM12, a member of the desintegrin and metalloproteinase family associated with liver fibrogenesis and cancer, modifies the TGF-beta signaling by modulating TGF-beta receptors trafficking [1]. Those experiments also showed that receptors keep trafficking among membranes, and keep being subject to the same degradation processes in presence of ADAM12. Our main hypothesis is: ADAM12 does not modify the structure of the receptor trafficking, but rather modulates the involved flows. Recently, two mathematical models demonstrated that TGF-beta receptors trafficking plays a critical in TGF-beta signaling regulation [2;3]. Both are based on differential equations. The standard procedure to compute a change of parameters in such models is to use some model fitting technique. This would need time series or dynamical observations while the data at hand is more of a qualitative nature. Previous investigations showed that the two models have a unique stable stationary state in the positive subspace. An analysis of the experimental protocols shows that almost all observations of the effect of ADAM12 are made when the cells have reached this equilibrium point. With qualitative data, we cannot expect to find a unique solution, seen as a vector of modified parameters from original models, explaining the effect of ADAM12 on TGF-beta signaling. Instead, we expected a whole subspace of parameters. We believe that the shape of this space gives valuable biological information. In order to investigate this subspace of parameters, we have designed an algorithm with two purposes. The first one is to reach the set of parameters compatible with the observations and the second to explore this subspace of parameters. The fitting of the parameters is measured by a cost function depending essentially on the equilibrium point. This cost function contains inequalities and consequently is not differentiable. For that reason, we decided to implement a kind of probabilistic hill-climbing algorithm for minimizing the cost function. In the process, the equilibrium point is computed by numerical integration with a standard Runge-Kutta method. The algorithm gives a satisfactory sampling of the space of solutions. As expected, from the shape of the solutions space arise interesting information. Our main prediction is an ADAM12 induced acceleration of TGF-beta receptors trafficking between endosomal and cell membranes. It also appears that ADAM12 might be a good candidate to change the TGF-beta signaling from a transient to a permanent response. This study appears to be an original methodology to extract predictions from differential models and qualitative observations together, that could be used to explore various systems in biology. [1] Atfi et al. - J. Cell Biol., 2007, 178:201-208 [2] Vilar et al. - PLoS Comput. Biol., 2006 [3] Zi et al. - PLoS ONE, 2007

Human promoters contain a likely nucleosome translational and rotational guide

Charles Hebert, Hugues Roest Crollius

Genome Organisation and Dynamics, CNRS UMR 8541, Ecole Normale Supérieure

Eukaryotic genomes are packaged as chromatin and nucleosomes represent the fundamental structural brick of this organisation. It is well known that the nucleosome positioning is a major contributor to chromatin accessibility, and hence contributes to the regulation of gene expression. However the dynamics of chromatin remodelling and nucleosome positioning around promoters is not fully explained and two models have been opposed in the last decade: the existence of a genomic code (DNA backbone as a key feature for DNA-nucleosome recognition), versus a more physical view (histone tails and DNA groove orientation). Both reflect the ability of DNA to wrap around the nucleosome. Our study provides an analysis of a collection of human promoters with experimentally well-defined transcription start sites (TSS). We asked if these sequences might contain a nucleosome-positioning signal near the TSS. The dbTSS database was used to extract 13.622 unique promoters. Signal processing techniques applied to the average base composition of this set of sequences aligned at the TSS reveal two antiphased and overlapping 10bp periodic patterns. These patterns, respectively composed by purine and pyrimidine dinucleotides, are positioned downstream of the TSS and are precisely co-localised with experimentally positioned nucleosome. CpC and GpA/ApG seem to be the most important antiphased contributors to the 10bp periodicity. These dinucleotides can be respectively defined as flexible and rigid steps in the DNA backbone, responsible for the DNA wrapping around the nucleosome. Nucleosome displacement modulates the accessibility of binding sites for transcription factors and so potentially influences gene expression level. Therefore, we wanted to know if our periodic signal might be related to changes in transcript level. We estimated for each promoter a gene expression plasticity score using GeneAtlas expression data. This score is computed from 73 tissues and is based on two measures: the mean expression level over all tissues and the number of tissues where the promoter expression is significant. We show that the periodicity pattern quality and amplitude is significantly correlated with a high expression level in a large number of tissues. This result indicates that our periodic signal is characteristic of broadly expressed "house-keeping" genes. We next sought to test the second model of nucleosome-DNA interaction. The periodicity signal can be considered as a rotational guide to expose histone tails to the post translational modifications machinery associated with the polymerase II. We selected promoters with the p300 binding site located between the TSS and the periodic signal, and show that the presence of p300 sites correlates with an increased periodic signal. Our results suggest that promoters binding p300 require a more specific nucleosome-positioning signal. Because p300 specifically acetylates histone H3 tails, this positional signal might be required to present histone tails to p300 in an optimal configuration to promote efficient transcription. Results show evidence for a nucleosome-positioning signal in human promoters. This signal also exists in mouse promoters and is thus likely to be conserved in mammals. This signal can be considered as a local translational and rotational guide for highly expressed genes.

Combinatorial usage of transcription factors in human tissues

Florence M G Cavalli 1, Juan M Vaquerizas 1, Nicholas M Luscombe 1,2

1 EMBL-European Bioinformatics Institute, 2 Gene Expression Unit, EMBL-Heidelberg

Precise regulation of gene expression is essential for an organism's survival. The first

step in gene expression is regulated by proteins known as transcription factors (TFs), which bind specific DNA sequences to regulate transcription. A key aspect of their function is their combinatorial activity with other TFs as well as with non-DNA-binding transcriptional regulators. The expression and combinatorial behaviour of TFs vary between different tissues, such that a unique subset of genes is expressed in each tissue. However, few human TFs are well characterised and even less is known about how they work in concert to induce tissue-specific gene expression. Here we characterise the combinatorial usage of human TFs by integrating complementary information. First, we use our own manually curated set of 1,368 human sequence-specific DNA-binding TFs, and gene expression for 33 healthy tissues from the SymAtlas dataset to statistically define tissue-specific TFs. Second, we extract protein-protein interaction (PPI) data from the Intact and HPRD databases to evaluate the extent of physical interactions between TFs. Third, we retrieve TF binding site (TFBS) predictions from phylogenetic footprinting analyses (Blanchette et al., 2006) and assess the degree of co-occurrence of binding sites within a regulatory region. These data are relevant to detect the combinations of TFs as they should be co-expressed, are likely to interact physically and/or bind in close locations. Integrating these datasets using a Bayesian model for the expression data with the PPIs and TFBS data as priors, we show that the regulatory network consists of combinatorial modules, ie, sets of interacting TFs, containing both ubiquitous and tissue-specific components. The tissues-specific TF detection enables us to evaluate the combinatorial changes between tissues. These results represent the key differences inside the TF regulatory mechanism, which are most likely to be responsible for tissue function.

Chromosome structures and transcription regulation networks

Ivan Junier 1, Olivier Martin 2, François Képès 1

1 : Programme d'Épigénomique, Génopole, 91034 Evry; 2 : LPTMS, 91405 Orsay

Transcription within the eukaryotic nucleus and within the bacterial nucleoid is expected to both depend on and determine the three-dimensional conformation of the chromosomes. Accordingly, transcribed genes, RNA polymerases and transcription factors (TF) have been shown to gather in vivo into discrete foci. The fine structure of the chromosome at the level of these factories remains however to be elucidated. A thermodynamic framework of DNA chains is presented, based on a worm like chain model where sparse sites along the DNA are able to interact. These sites aim at representing the genes that are to be co-localized. In the case of a periodic positioning of the sites, transcription factories spontaneously emerge according to a microphase separation where clusters of genes are separated by bundles of DNA chains free of genes. Within this scope, I will discuss the different chromosome conformations we have found with the help of Monte-Carlo simulations. In a second part, it will be shown that compared to a random location, a regular positioning of the genes along the DNA makes specialized co-localization more efficient. This corroborates recent results that have highlighted a periodic positioning of genes that are regulated by the same transcription factors in yeast, and the TF coding genes as well in E. coli. In this regard, I will present a method that aims to efficiently detect periodicities in a noisy signal. By applying the method to the positions of genes that are regulated by the same transcription factors in E. coli, we are able to reveal the periodic presence of 10 kbps domains all along the bacterial chromosome. Our results further suggest that these domains are spatially coordinated through the action of global transcription factors such as CRP. Overall, our results corroborate the existence of topologically independent domains that are roughly 10 kbps long and suggest new mechanisms for their coordination.

Comparative genomics of spermatogenesis in vertebrates: insights from the medaka fish Oryzias latipes

vincent martin, astrid böhne, christina schultheis, philippe durand, jean-nicolas volff and delphine galiana

IGFL

Sexual development into a male or a female is controlled through a gene cascade which is activated (or not) by a master sex-determining gene. The master sex-determining gene is conserved among mammals (SRY) but apparently changed very frequently during evolution in fish. So far, the DMY gene discovered in the medaka Oryzias latipes is the only master sex-determining gene known in fish. This gene is restricted to several species of the genus Oryzias. In contrast, genes implicated in sex differentiation downstream of sex determination are generally more conserved among vertebrates, even if reproductive genes tend to evolve rapidly. Of particular interest for fertility are genes expressed in the germ cell niche and expressed during the first wave of spermatogenesis and/or oogenesis. A comparative analysis of transcript profiles of pre- and post-puberty spermatogonia and Sertoli cells in Rattus norvegicus has revealed several markers of prepubertal spermatogonia cells. In order to approach conserved and divergent aspects of spermatogonial stem cell biology in vertebrates, genes identified in rat have been cloned and analyzed in the medaka fish Oryzias latipes, a species with sequenced genome and characterized master sex-determining gene. Interestingly, some spermatogonial genes present as a single copy in mammalian genome are duplicated in fish, probably as the result of an ancestral fish-specific genome duplication prior to the radiation of teleosts. Expressed sequence tag profiling revealed that many candidate genes are preferentially expressed in medaka gonads, suggesting a conservation of function over 450 million years of evolution. The expression, evolution and function of these genes is further studied to assess the evolutionary dynamics of spermatogenesis genes in vertebrates.

Three independent and recurrent proviral integrations contribute to a same transcriptional network involved in Friend murine erythroleukemia

G Juban1§, G Giraud1§, B Guyot1, S Belin1, JJ Diaz1, J Starck1, C Guillouf2, F Moreau-Gachelin2, F Morlé1*

1 : CGMC CNRS UMR 5534, Université Lyon1, Villeurbanne; 2 : INSERM U528, Institut Curie, Paris. § : first co-authors

The Friend virus complex is composed of a replication-defective Spleen Focus-Forming Virus (SFFV) and a replication-competent Friend Murine Leukemia Virus (F-MuLV) which both induce clonal erythroleukemia in susceptible mice. Most erythroleukemic clones induced by these retroviruses display recurrent proviral integrations that independently activate either one of three different genes Spi-1 (>90% of SFFV

integrations), Fli-1 (>75% of F-MuLV integrations) or Fli-3 (a minority of F-MuLV integrations). Spi-1 and Fli-1 encode two ETS family transcription factors (Spi-1/PU.1 and Fli-1, respectively) that recognize the same core DNA binding motif GGAA. Fli-3 encodes miRNA cluster miR17-92 already known to be involved in other cancers. We already showed that Fli-1 is a direct target gene activated by Spi-1/PU.1 in erythroleukemic cells displaying a Spi-1 activated gene. In the present study, we used siRNA mediated depletion of Spi-1/PU.1 and/or Fli-1 to demonstrate that Spi-1/PU.1 and Fli-1 additively contribute to the proliferation, survival and inhibition of differentiation of these erythroleukemic cells thus indicating that Fli-1 is not a simple effector of Spi-1/PU.1. This prompted us to perform whole transcriptome analyses to identify all gene transcripts that are regulated in an additive manner by Spi-1/PU.1 and Fli-1. Among these genes, we found a surprisingly high proportion of genes that are involved in different steps of ribosome biogenesis. All these genes display at least one highly conserved ETS binding site at less than a few hundred base pairs from the known transcription initiation site and chromatin immunoprecipitation revealed that their promoters are all bound by both Spi-1/PU.1 and Fli-1 in erythroleukemic cells. Surprisingly, we found that Fli-3 transcripts (pri-miRNA 17-92) are similarly regulated in an additive manner by Spi-1/PU.1 and Fli-1 and that Fli-3 promoter which also displays conserved ETS binding sites, is also bound by both Spi-1/PU.1 and Fli-1 in vivo. Furthermore, we found that all these genes identified as common direct targets of Spi-1/PU.1 and Fli-1 in erythroleukemic cells displaying a Spi-1 activated gene are also direct target genes of Fli-1 in erythroleukemic cells displaying a Fli-1 activated gene. Altogether, these results show that two ETS family transcription factors and one micro RNA gene cluster that are recurrently and independently activated in Friend erythroleukemia are surprisingly linked to a same transcriptional network. This in turn strongly suggest that, downstream of this network, the stimulation of ribosome biogenesis as well as some transcripts regulated by miRNA belonging to cluster miR17-92 are most probably critical determinants in the erythroleukemic process.

Yeast chromatin: nucleosome positioning via excluding energy barriers

Vaillant Cedric

Laboratoire de Physique- ENS de Lyon- CNRS

The genomewide analysis of nucleosome occupancy profiles measured by microarray Chip experiments reveals an heterogeneous pattern with alternance of regions of well positioned nucleosomes, fuzzy nucleosomes and small nucleosome free regions. We propose a simple physical modelling of sequence dependent dynamical nucleosome assembly that accounts very well for these specific chromatin features. Our modelling suggest that in vivo ``average'' chromatin is controlled to a large extend controlled by the underlying genomic sequence but subject to finite-range remodelling action. We further show how ``nucleosome energy barriers'' contribute significantly to the the long-range nucleosome ordering, and finally, we analyse the implication of such ``positioning via excluding'' mechanism on the structure and function of genes.

Modeling membrane micro-domain formation through inhomogeneous diffusion

H. Soula, G. Beslon

INSA RMND/ INSA LIRIS

The cellular membrane is a complex system mixing various lipids and proteins. It has been shown that these proteins undergo isotropic random movements, resulting in a 2D diffusion on the membrane surface. However, although such a diffusion should result in an homogeneous distribution of the components, plasmic membrane actually appears to be an highly structured system in which large areas seem protein-free whereas the so-called microdomains concentrate a wide collection of membrane proteins in higher concentration. In addition, some of these microdomains are sometimes called "lipid raft" due to a specific lipid composition. Various mechanisms have been proposed in the literature to reconcile both membrane properties and to explain the existence of heterogeneity in a diffusing medium: trapping systems, picket fences, phase transition... to cite but a few. However, none of these mechanisms are fully satisfactory, either because they rely on the existence of hypothetical membrane components that constraints protein diffusion or because they cannot explain the colocalization of lipid and proteins. Here we propose a simple mechanism that is able to explain the colocalization of protein and lipid rafts by simply taking into account the properties of the different lipid components of the membrane. Using very simple mathematical models and agent-based simulations, we show that the variation of membrane viscosity directly leads to variation of the local concentration of diffusive particles. Yet, it is known that the presence of specific lipid phases in the membrane (particularly cholesterol) modify the organization of the lipidic bi-layer, thus slowing down the diffusion by a factor of ten. Our model shows that, in such a situation, the freely diffusing protein still undergo a Brownian motion but that they concentrate in the areas where the diffusion coefficient is diminished. Moreover, we are able to calculate their concentration depending on the variation of the diffusion coefficient and on size of the slow compartment related to the size of the membrane. This very simple mechanism does not need any additional supposition or components. It is simply based on the known existence of different lipid components in the membrane. Moreover, it can be indirectly confirmed by some simple experiments. It is well known that cholesterol and phospholipids behave differently in different temperature: when the temperature increases, the cholesterol viscosity decreases while, for the other membrane components, it increases. Thus, at low temperatures, proteins should be more homogeneously distributed on the membrane, a property that has been observed in vitro. Our model is very general and can be applied to other cellular domain where passive diffusion must conciliate with domain constitution (e.g. the eukaryotic nucleus). However, it is probably not the unique structuring process in these systems. Using computational simulation, we have been able to show that, when combined with phase transition due to protein-protein interactions, the slow diffusion is able to trigger the transition resulting in a local phase transition in the vicinity of the cholesterol molecules. In silico experiments made on the model (such as virtual imaging or virtual FRAP) have shown that the conjunction of both mechanisms results in a behavior similar to real membranes

Interaction usage among contextualized tissular interactomes

Ouissem Souiai, Emmanuelle Becker and Christine Brun

TAGC Inserm U928, Parc Scientifique de Luminy

Tissue-specificity is usually perceived and understood through transcription regulation and differentiation programs. Once these programs are settled and established,

differentiated cells express sets of genes which allow them to perform their physiological role. This work aims at tackling tissue-specificity with a protein-protein interaction network perspective. For this, a high quality human protein-protein interaction network containing ~11000 interactions has been contextualized by integrating expression data. Assuming that all possible interactions between proteins expressed in a given tissue form the protein-protein interaction network of this particular tissue, we have inferred 21 possible tissular networks. We then defined the 'interaction usage' as the number of tissues in which each interaction is possible. A study of the distribution of the 'interaction usage' allowed us to characterize the 'largest common interactome network' (LCIN) as the set of interactions shared by all the investigated tissues. It currently encompasses 27% of the interactome. More broadly, we showed that 70% of the interactions are possible in the vast majority of the tissues (16 to 21 tissues). On the opposite, tissue-specific interactions are very few. Functionally, a GO term enrichment/depletion analysis demonstrated that the LCIN is formed by proteins and interactions contributing to the accomplishment of basic cellular processes. This suggests that the LCIN may constitute the 'functional core' of the network into which housekeeping proteins are interacting. When such analysis is undertaken according to the interaction usage, it reveals that certain biological processes are over-represented among the proteins involved into tissue-specific interactions whereas depleted among proteins involved in the LCIN and vice-versa, as expected. Interestingly, a fine tuning of some processes can be proposed based on our analysis according to the interaction usage. Finally, we are presently trying to understand how tissue-specific interactions are articulated with the LCIN and more generally how they are shaping the interactome.

Detection and evolutionary analysis of ancient mammalian pseudogenes

David Brawand, Walter Wahli, and Henrik Kaessmann

Center for Integrative Genomics, University of Lausanne

We developed an approach that allows systematic screening of genomes for ancient pseudogene remnants and analysis of their time of inactivation. The detection of pseudogene sequences in our approach is based on a sensitive sequence alignment algorithm, while the evolutionary history of detected pseudogenes is assessed using evolutionary simulations of neutral gene evolution. We will present details of this approach and illustrate its utility by showing detected cases of ancient mammalian pseudogenes that provide unique insights into the genomic and phenotypic evolution of mammals. For example, a screen for the major egg yolk genes - vitellogenins (which provide the bulk of nutritional reserves for non-mammalian embryos) - in mammals using our procedure, revealed that the emergence of lactation in the common mammalian ancestor and the development of placentation in eutherian and marsupial mammals allowed for the gradual loss of yolk-dependent nourishment during mammalian evolution.

Analysis of fine-scale mammalian evolutionary breakpoints provides new insight into their relations to genome organisation and open

chromatin

Lamia Zaghloul (1,2,*), Claire Lemaitre (1,3,4,*), Marie-France Sagot (1,3,4), Christian Gautier (1,3,4), Alain Arneodo (1,2), Eric Tannier (1,3,4) and Benjamin Audit (1,2)

(1) Université de Lyon, F-69000 Lyon, France; (2) Laboratoire Joliot-Curie et Laboratoire de Physique, CNRS, Ecole Normale Supérieure de Lyon, F-69007 Lyon, France; (3) Laboratoire Biométrie et Biologie Evolutive, CNRS, Université Lyon 1, F-69100 Villeurb

Rearrangements that shuffle a chromosome organisation happen because of double-strand breaks and of the joining of double-stranded DNA ends, resulting in new conformations of the molecules. If the molecular bases of DNA damage and repair mechanisms that control these genetic instabilities at the cellular level are well described, the biophysical forces and selection constraints at work on an evolutionary timescale still need to be fully understood. In fact, despite recent progress made in this field of research, it remains an open question simply to design a model that would account for the observed positions of rearrangement breakpoints. We join the debate and investigate the mechanisms governing the large-scale rearrangements responsible for chromosome evolution with respect to genome organisation into domains of different activity. To do so, we apply a novel pairwise comparison strategy between the genome of human and those of five sequenced eutherian mammals. This allows us to delineate evolutionary breakpoint regions along the human genome with a finer resolution (median size ~26.6kb) than obtained before. According to the Intergenic Breakage Model, which is the current up-to-date model of structural genome evolution, breakages happen with a uniform propensity along the genome but are counter-selected in genes, their regulatory regions, and in between. Correspondingly, we observe that breakpoints are under-represented in genes. Surprisingly however, the density of breakpoints in small intergenes (~1 per Mb) appears significantly higher than in gene deserts (~0.1 per Mb). Actually, we find that the distribution of breakpoints follows the organisation of the genome into isochores, and is highly correlated with hypo-methylation and DNA accessibility markers (namely CpG observed/expected ratio and experimental DNase~I hyper-sensitivity). We also observe a high breakpoint density and high small intergene abundance around a set of 1060 putative replication origins previously identified in the human genome. We propose a mechanistic explanation for this, based on the hypothesis that the regions of high transcriptional activity and replication initiation have open chromatin and are thus subject to a higher fragility. Natural selection doesn't seem sufficient to explain the heterogeneous distribution of evolutionary breakpoints along human chromosomes. We highlight the importance of also taking into account a mutational bias linked to local chromatin state and formulate a model combining the two.

Protein domain detection by co-occurence : application to Plasmodium falciparum

Nicolas TERRAPON, Olivier Gascuel, Laurent Bréhélin

LIRMM Université Montpellier 2

Hidden Markov Models (HMMs) have proved to be powerful for protein domain identification. Notably, Pfam provides a large collection of HMMs covering 73% of SWISSPROT/TrEMBL proteins. Each Pfam HMM is a probabilistic model characterizing a given domain. When analyzing a new protein sequence, a score is computed to

measure the similarity between the sequence and the domain at hand. This score is then compared to a stringent threshold (provided by Pfam) above which the domain presence in the protein is asserted. However, with highly divergent proteins this procedure may miss numerous domains. For example, with Plasmodium falciparum (the main causal agent of Malaria), 1300 distinct domains are detected, corresponding to ~50% of the proteome, while with Yeast these numbers are of 2100 and 73%, respectively. Although this observation could be explained by the existence of genes that are unique to P. falciparum, it is likely further exacerbated by high genome atypicity, with 80% A+T. This induces a strong bias in amino-acid composition, which makes homology detection particularly difficult. Decreasing Pfam thresholds might enable more domains detections, but at the expense of numerous false positive predictions. Here, we propose to filter these false positives using domain co-occurrence. Several studies show that most domains are solely associated with a few other favorite domains. Our approach combines this property with HMMs using loose thresholds. First, we established a list of domain pairs showing a strong co-occurrence, i.e. for which the presence of one domain is a strong clue of the presence of the other one. This list of Conditionally Dependent Pairs (CDP) is built from the domain compositions of well annotated proteins in SWISSPROT, using a statistical test of correlation. Then, we list for each protein of the query organism its known and potential domains. Known domains are those found using standard Pfam thresholds or curated by experts. Potential domains are predicted by Pfam HMMs after threshold lowering. The presence of a potential domain is then validated by a known domain if the domain pair appears in the CDP list. Importantly, our method includes a statistical procedure to estimate the false discovery rate among so validated domains. This procedure uses a shuffling algorithm to estimate the number of domains the approach would validate under the hypothesis that the potential domains were randomly predicted. Our method was assessed thanks to test experiments with Yeast and then applied to P. falciparum. On this organism, it allows detecting more than 290 additional domains in 270 proteins, with an estimated false discovery rate of ~10%. Moreover, it provides new Gene Ontology annotations for more than 70 proteins. Among the new domains, 96 domain types (involved in functions like apoptosis) were previously unknown in P. falciparum. Our experiments show that the approach has a high accuracy and the ability to structurally and functionally characterize numerous hypothetical genes, even with well characterized genomes. For example with Yeast, it allows discovering 280 domains in 256 proteins, and provides new GO annotations for 40 proteins.

IDENTIFICATION OF MOLECULAR MARKERS ASSOCIATED WITH HIGH METASTATIC RISK IN UVEAL MELANOMA PATIENTS

Cecile Laurent, E. Barillot, J. Couturier, C. Decraene, L. Desjardins, J.P. Meyniel, S. Piperno-Neumann, X. Sastre, J.P. Thiery, and S. Saule

Institut Curie

Uveal melanoma occurs to the detriment of uveal melanocytes (located in the iris, ciliary body and choroid) and is the most common intraocular malignancy in adults, with 500-600 new cases every year in France. Unlike the improved survival rates for the numerous cancers in which early detection and management have improved, the survival rate for uveal melanoma has not increased significantly in the last 20 years. Despite the common embryonic origin of their precursor neural-crest cells, there are clear differences between the clinical behaviour and molecular biology of uveal and

cutaneous melanoma, such as metastasis pattern and chromosomal imbalances. The etiological factors involved in the process of malignant transformation of uveal melanoma are poorly understood. Gene expression profiling from 63 tumours with a three years follow-up have been already performed, and genes selection using significant analysis of microarrays performed between primary tumours that have metastasized before thirty-six months (noted meta1) and those that do not have metastasised before thirty-six months (noted meta0) showed that about 1000 probe sets are differentially expressed using a FDR cut off of 5%. We focus on three genes (NEDD9 / FAK / PTP4A3) involved in cell migration. We used real-time PCR to validate our results, using a set of selected genes in the microarray expression values. The results obtained with 14 uveal melanoma showed a complete concordance with the Affymetrix microarrays. PTP4A3 is a phosphatase implicated in cell migration and proliferation, PTK2 is a focal-adhesion kinase, an important mediator of growth-factor signalling, cell proliferation, cell survival and cell migration and NEDD9 is a gene described by oncogenomic analysis as involved in melanoma metastasis. Gene expression profiling from 115 liver metastasis have also been performed and compared with the primary tumors. A part of the samples are also analysed by Comparative Genomic Hybridization (CGH) analysis and the chromosomal imbalances were correlated with the transcriptomal expression using a method developed in the Institut Curie bioinformatic team, in order to identify genes associated with the bad prognosis that could play a causal role in the disease. Biochemical pathways associated with an increased metastatic risk will be presented.

Exploring cell cycle regulation in alpha proteobacteria

Brilli M, Fondi M, Fani R, Mengoni A, Ferri L, Bazzicalupo M, Biondi EG

Department of Evolutionary Biology, University of Florence

Cell cycle regulation in bacteria is a poorly understood field. Recently a single, self-sufficient model has been proposed to explain cell cycle regulation in the alpha-proteobacterium Caulobacter crescentus. In this organism CtrA is the master regulator of the cell cycle progression. In fact, CtrA is a response regulator that binds specific DNA sequences and can repress or activate expression of genes involved in many aspects of the cell cycle or it can silence the origin of replication. CtrA ability to bind DNA depends on its phosphorylation state that is carried on by a phosphorelay composed by the hybrid histidine kinase CckA and the phosphotransferase ChpT. CtrA is also regulated at a transcriptional level by DnaA and GcrA and by methylation of its promoter by CcrM. Moreover CtrA inhibits its activity through transcriptional activation of DivK, which is involved, when phosphorylated, in down regulation of CckA activity. Finally it is also known that at least two histidine kinases, DivJ and PleC, control DivK phosphorylation level. CtrA is also controlled by a specific cell cycle-dependent proteolysis by the ClpPX protease, involving also other factors, CpdR and RcdA. Here we systematically analyzed the organization of factors involved in the regulation of CtrA in Caulobacter in the genomes of alpha proteobacteria. In particular we looked for orthologous genes of the Caulobacter ones and also we predicted the CtrA regulon in all alphas. By this systematic bioinformatic analysis we have reconstructed the cell cycle regulation architecture of the seven clusters (A to G) of alpha proteobacteria, shedding light on properties and logic of this fundamental regulation.

Impact of expression level on the fate of genes following whole-genome duplications in Paramecium tetraurelia.

Gout JF, Kahn D ans Duret L.

Laboratoire de Biologie et Biométrie Evolutive, UMR CNRS 5558, Université Lyon1

Whole Genome Duplications (WGDs) are dramatic events that result in doubling the number of genes in the genome. Most of gene pairs (named ohnologons in honor of the pioneering ideas of Susumu Ohno on the role of WGDs in genome evolution) originating from WGDs progressively lose one copy during a period of massive gene loss that typically follows WGDs. However, some pairs are retained through very long evolutionary times and represent unique opportunities of genomic innovation. Despite the fact that WGDs are now studied in many different eukaryotes, the mechanisms that explain the retention or loss of genes following WGDs are still under debate. Here, we take advantage of the 3 successive WGDs that occurred in the lineage of Paramecium tetraurelia to address this question. With the help of transcriptomic data obtained by microarrays we analyze the influence of expression patterns on the retention of ohnologons. We find that the average expression level is strongly (and positively) correlated to the retention rate. We explore two main models to explain this observation. 1. Considering that gene loss following WGD is a gradual process in which the dying gene can still express a truncated protein before complete pseudogenization, we investigate the deleterious effects caused by this expression. We postulate that this deleterious effect is proportional to the level of expression of these truncated proteins. This first model is somehow analogous to the model of 'mistranslation-induced protein misfolding' proposed by Drummond and colleagues to explain why highly expressed proteins evolve slowly (a general trend observed in all studied organisms). 2. In a second hypothesis, we explore a model that takes into account the cost of expression of proteins. Assuming that the expression level of genes at the very beginning of the massive gene loss period is at an optimum, this model predicts that the impact on fitness of halving the expression level (ie : losing one gene copy) increases with the original expression level. We take advantage of data available from the yeast deletion project to test the predictions of this model. With about 50% of the gene pairs still retained from a recent WGD, paramecium is the best model organism available to capture the dynamics of post-WGD evolution predicted by these models. Therefore, we will discuss the pros and cons of these two models in the light of genomic and transcriptomic data available for P. tetraurelia.

Gene expression and the dynamics of transcriptional regulation: a model for theoretical and experimental investigations

Antoine Coulon (1,2), Olivier Gandrillon (1) & Guillaume Beslon (2)

(1) Centre de Génétique Moléculaire et Cellulaire (CGMC) & (2) Laboratoire d'InfoRmatique en Image et Systèmes d'information (LIRIS)

Transcriptional regulation is an inherently dynamic process that rely on the stochastic molecular events occurring on the regulatory region of a gene. From prokaryotes to eukaryotes, it can take various forms and result from the interaction of a variety of

molecules and epigenetic factors. We propose a generic model for investigating the stochastic dynamics of any arbitrary regulatory structure and its impact on gene expression. It describes the stochastic molecular events and their mutual influence. In particular, it can represent the interplay between various epigenetic factors (chromatin opening/closing, nucleosome sliding, histone tails covalent modifications) and TFs molecules (including both cooperative/competitive aspects of their association/dissociation as well as their remodeling/enzymatic activity on epigenetic factors). Subsequent steps of gene expression (transcription, translation and degradation of transcripts and proteins) are modeled simply but explicitly to assess how promoter dynamics impacts on the expression level. Contrarily to classical thermodynamic approaches [Saiz et al. 2006, Mol Syst Biol], our kinetic one allows to represent systems with energy-dependant transitions (ie. ATP-dependent chromatin remodeling) and is therefore relevent for investigating eukaryotic regulation. This approach also makes it possible to predict dynamic aspects of the activity of the system, which turn out to be essential for the understanding of these structures. The study of this model shows that the spontaneous activity of a single gene can demonstrate much complexity due to the mere interplay between molecules. In particular, the regulatory structure can demonstrate a strongly periodic activity as it is observed in vivo on eukaryotic promoters. Indeed, it is known that regulation proceeds through cycles of recruitment of TFs and of modifications of epigenetic factors [Metivier et al. 2003, Cell]. How this property can arise and what it implies can be investigated theoretically with such a model. For instance, changing TFs concentration, as in classical regulation, can provide as a flexible way to tightly modulate several parameters of this dynamics (ie. frequency and coherence of periodicities). Moreover, an energetic cost can be precisely quantified for a given system. In addition to be necessary for any periodic activity, this cost appears to be linked to other aspects of the dynamics (such as flexibility or robustness). This have certainly important evolutionary outcomes and call for further investigations. This model can also be used to integrate various experimental approaches. Indeed, the activity of the system can be derived in terms of various indicators reproducing measures of experimental devices such as flux cytometry on clonal populations of cells, spectral analysis of single-cells gene expression timecourses, FRAP, FRET and time-resolved ChIP. This model appears as a unifying framework for confronting experimental evidences at various levels and time scales that provides different viewpoints on the same object. Thus it appears as central tool for systemic experimental investigations of regulatory structure dynamics and stochasticity in gene expression.

Confronting regulaion network models to CGH and microarray data

Sylvain Blachon 1, Carito Gudziolowski 1, Gautier Stoll 2, Andrei Zinoviev 2, Ovidiu Radulescu 1, Michel LeBorgne 1, Anne Siegel 1

1 IRISA/INRIA Rennes ; 2 Institut Curie Paris

High throughput technologies allow biologists to think in a system way. One of the current challenges is to use generated data to enrich our knowledge on molecular processes occurring in biological systems. One major issue concerns modeling regulation networks : there is no simple way to perform this task. Confronting models to real data may help to focus on relevant models. Here, we present a framework which uses biological data to support models of regulatory networks. The objective is to test the network coherence when confronted to experimental data. Given a

qualitative network, Bioquali is able to compute its coherence and make predictions on the node states which are unknown. In this formalism, the information is coded by + and -. A + on a node A (respectively a -) means that A variates positively (respectively negatively). A + (respectively -) on an edge from the node A to B means that A positively (respectively negatively) influences B. However simple, this formalism combines two advantages: first, the description is abstract and sticks to biological intuitions. This avoids the problem of choosing a network resolution and of speculating on biological constants. The second advantage is that it is based on a sound and robust mathematical formalism. It enables to compute consistencies and predict node states on large scale networks. The network model comprises mRNA nodes and active protein nodes. Microarray data give information on mRNA nodes. The variation sign between two conditions defines the node state. It is positive when the gene is overexpressed and negative when underexpressed. From CGH data, one derives the influence of a gene copy number on its expression level. A « CGH influence » node is thus added to the network. The influence sign depends on the variation of gene copy number between two conditions. It is positive when the gene copy number is higher and negative when lesser. Ewing tumors are the most frequent pediatric tumors. A translocation, producing the chimeric protein EWS-FLI1, deregulates the normal cell physiology, causing the tumor developpement. EWS-FLI1 effects may be enhanced by chromosomal aberrations. To investigate this, Olivier Delattre's team produced CGH and microarray data on 39 samples mostly derived from biopsies. 741 pairwise comparisons between pairs of patients were performed to to better understand the variability in tumor outcome. On this system, an interaction network involving 141 genes was modeled from litterature to understand the effect of EWS-FLI1 on key signal pathways (cell cycle, apoptosis and cell migration). Altogether, this produces a set of 741 networks enriched with biological data. The coherence of these networks was computed by Bioquali : they appear to be mostly coherent, supporting the soundness of the model. From those, predictions on nodes can be computed, to help biologists to formulate new hypotheses. The inconsistent networks will be thoroughly examined to understand the weaknesses in the modelization, in order to further refine the model.

Overview | Schedule | How to submit? | How to attend? | IPG 2008