Titre du projet de thèse : Integrative analysis of gene expression regulation during *Listeria* infection

Spécialités : Bio-informatique, Génomique, Biologie des systèmes, (Biologie moléculaire*)

Résumé du projet de thèse *(1 page maximum, en anglais)*

The invasion of mammalian cells by intracellular bacterial pathogens reshuffles their gene expression and functions; typically host cells launch antibacterial defences while pathogens subvert various host pathways to their own benefit. We have substantial evidence that during the infection of epithelial cells by *Listeria monocytogenes* (*Lm*), the food-borne cause of listeriosis, host cell gene expression is reshaped quantitatively and qualitatively by a combination of regulations affecting RNA synthesis, maturation, decay and translation. We have recently gathered different sets of –omics data covering with high resolution the flux of host gene expression from transcription to translation during an infection time-course (0, 2, 5, 10 h) in intestinal human epithelial cells using the following technologies: RNA-Seq, 4sU-Seq of nascent RNA, Ribo-Seq of translome, and Nanopore MinION full-length cDNA sequencing. Our initial statistical analysis of these datasets, experiment by experiment, has confirmed their quality and reliability of each dataset, and further revealed groups of transcripts that are induced or repressed during infection, transcriptionally, and/or translationally.

To better understand how *host transcripts are co-regulated during infection*, we now aim at developing an integrative computational biology analysis of our datasets and classify individual transcripts into functional regulatory clusters. More precisely, we intend to highlight common regulatory features and functional annotations of groups of transcripts sharing a similar fate during infection, from transcription to protein production.

To this end, the PhD student will define groups of transcripts sharing regulation profiles along the infection time-course as well as across datasets. The resulting clusters will then be statistically tested for enrichment in genes coding for components of known molecular pathways using gene set enrichment analysis (GSEA). This will allow us to assess whether common regulatory profiles correlate with possible co-functions. The student will also search for enrichment of known or novel sequence motifs using the RSAT and MEME software suites, microRNA-binding sites from MirTarBase, or structural features using DotAligner, which could explain a co-regulation. This part of the project will be performed in collaboration with Morgane Thomas-Chollier (IBENS). The regulatory potential of such sequences would then be confirmed biologically by other members of the team*. If the enriched motifs correspond to binding sites for known RNA-binding proteins, their role in regulation will also be assessed. Potential miRNA-dependent regulation will be further challenged in silico, by intersecting the predicted targets of miRNAs previously found to be induced or repressed during the infection of epithelial cells by *Lm* with the transcripts consistently affected in their stability and/or translation. The most promising regulatory interactions will be validated experimentally.

*As an alternative, the student may ambition mixed wet lab/dry lab training, and may thus wish to perform his/her own experimental validations. The scope of the bioinformatics analysis would then be reduced accordingly; for instance, modelling of regulatory loops would not be attempted, to avoid dispersion and save experimental time.
Depending on the results for the previous steps, we will further initiate the **modelling of the most relevant gene circuits** coordinating the regulation of subsets of host targets genes during infection will be designed, from our data and that extracted from the literature, in collaboration with Denis Thieffry and Morgane Thomas Chollier. Predictions of the resulting model will then be experimentally tested by the student’s co-workers.

The PhD student will also investigate if **alternative isoforms** are expressed during infection, based on events of alternative splicing than we have detected in our paired-end RNA-Seq datasets. For instance, we have identified that differential splicing of one pre-mRNA leads to the use of a distal 3’-UTR during infection instead of a proximal 3’-UTR in non-infected cells. This likely affects the loading of regulatory RNA-binding proteins or miRNA-RISC complexes, and thus translation. By analysing **complete cDNA reads** produced with MinION sequencing, the student will identify alternative splicing isoforms, and test statistically if their occurrence varies during infection. Mapping will be performed using the Eoulsan framework developed in-house by IBENS Genomics facility. The dedicated analysis of processing changes will require benchmarking of existing tools, and further development in collaboration with the Genomics facility. The impact of such events on the translation of targets and on the outcome of infection will then be validated in our team.

Last, the student will further exploit our recent Ribo-Seq datasets to detect variations in ribosome occupancy within transcripts, indicative of events that could impact either quantitatively on gene expression, or on the function of the protein product (translation of novel open reading frames, or of alternative isoforms, frame-shifting, ribosome pausing, etc.).

The PhD student will benefit from the infrastructure and diverse expertise of the members of the **IBENS Computational Biology Centre**. This transversal structure aims at reinforcing synergies between individual computational scientists in each team, by mutualising a potent computing cluster, providing training alongside specialised workshops and seminars, and stimulating regular exchanges on methodological issues. Computational biology analyses will rely on Python and R scripts, taking advantage of existing R packages and other relevant software. Jupyter notebooks will be used to develop, annotate and share the code together with the result so as to ensure that scripts can be durably understood and used by different users. Any tool developed for this project will be implemented on a private GitHub, and made open upon publication.

Ultimately, this integrative computational approach of multiple –omics datasets will outline groups of transcripts sharing **common regulation patterns, similar sequence motifs, structure predictions and/or functional annotations**, and tentatively delineate control mechanism. It is anticipated to highlight candidate regulated mRNA targets, and possibly host regulatory factors (miRNAs or RNA-binding proteins) playing a role in the infectious process. Putative candidates critical for the control infection by *Lm* or other pathogenic bacteria could subsequently become promising druggable targets.

**Thèses actuellement en cours dans l’équipe**

*Tous les encadrements doivent être indiqués (y compris les co-directions avec un autre HDR, et les encadrements dans le cadre de programmes doctoraux tels qu’IPV, FDV...)*

<table>
<thead>
<tr>
<th>Nom et Prénom du doctorant</th>
<th>Directeur(s) de thèse</th>
<th>Année de 1ère inscription</th>
<th>ED</th>
<th>Financement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peron Cane Caroline (co-direction)</td>
<td>Nicolas Desprat (ENS, LPS) et Alice Lebreton</td>
<td>2015</td>
<td>ED564 – PIF</td>
<td>UPMC - Programme Interfaces pour le Vivant</td>
</tr>
</tbody>
</table>

**Trois publications récentes du directeur de thèse** (du co-directeur ou du co-encadrant s’il y a lieu). Mettre en gras le nom du directeur de thèse.


**Docteurs encadrés par le directeur de thèse** ayant soutenu après septembre 2012 et publications relatives à leur sujet de thèse. Mettre en gras le nom du directeur de thèse et celui du docteur.

Aucun