

Statistical modeling and inference to identify DNA sequence elements involved in transcription regulation

Sophie Lèbre

IMAG UMR 5149 - Equipe Probabilités et Statistique, Université Paul Valéry Montpellier 3,
Montpellier, France

Gene expression is orchestrated by distinct regulatory regions of the DNA to ensure a wide variety of cell types and functions. Identifying DNA determinants involved in the regulation of gene expression will accelerate genetic analyses and precision medicine approaches. As part of the overall effort to achieve this goal, we propose several approaches dedicated to the identification of DNA features associated with the process of transcription regulation.

Through a pluridisciplinary research collaboration (IGMM/IMAG/LIRMM), we develop simple and explanatory models to explain gene expression level or transcription factor binding, and attempt for robust variable selection.

In this talk, I will focus on feature engineering and statistical modeling dedicated to biological hypothesis testing from publicly available data as well as specific datasets (Bejjani et al., NAR 2021). Statistical inference includes various regularization methods (lasso, group-lasso, elastic net).

References

- *A systematic analysis of the genomic features involved in the binding specificity of transcription factors*

R. Roméro, C. Menichelli, J.-M. Marin, S. Lèbre, C.-H. Lecellier & L. Bréhélin, Preprint in progress.

- *Identification of long regulatory elements in the genome of Plasmodium falciparum and other eukaryotes*

C. Menichelli, V. Guitard, R. M. Martins, S. Lèbre, J.-J. Lopez-Rubio, C.-H. Lecellier, L. Bréhélin, Plos Computational Biology, 2021.

- *Debiasing the Elastic Net for models with interactions*

F. Bascou, S. Lèbre and J. Salmon, Journées de Statistique de la SFDS, 2020.

- *Probing transcription factor combinatorics in different promoter classes and in enhancers*

J. Vandel, O. Cassan, S. Lèbre, C.-H. Lecellier and L. Bréhélin, BMC Genomics, 2019.

- *Probing instructions for expression regulation in gene nucleotide compositions*

C. Bessière, M. Taha M., F. Petitprez, J. Vandel, J.-M. Marin, L. Bréhélin, S. Lèbre and C.-H. Lecellier, PLoS Computational Biology, 2018.