
Variable selection in transcriptomics data using knockoffs in a classification framework

Julie Cartier*^{†1,2,3}, Johanna Lagoas^{1,2,3}, Youmna Ayadi^{1,2,3}, Adeline Fermanian⁴,
Chloé-Agathe Azencott^{1,2,3}, and Florian Massip^{‡1,2,3}

¹Centre de Bioinformatique – Mines Paris - PSL (École nationale supérieure des mines de Paris) –
France

²Institut Curie – PSL - University – France

³Oncologie Computationnelle (U1331) – Institut National de la Santé et de la Recherche Médicale -
INSERM – France

⁴LOPF Califrains'Machine Learning Lab – Califrains – France

Abstract

The emergence of new sequencing technologies has facilitated the acquisition of large amounts of biological data, which has proven to be a useful tool for better understanding biological systems. One way to take advantage of the potential of sequencing data is to use them to identify the relationship between biological units (e.g. genes) and phenotypical characteristics (e.g. disease outcomes). This question, formulated as a variable selection problem, remains difficult because of the size of the data (n

Keywords: Variable Selection, Knockoffs, Transcriptomic

*Speaker

[†]Corresponding author: julie.cartier@minesparis.psl.eu

[‡]Corresponding author: florian.massip@minesparis.psl.eu