

Mélanges gaussiens bidimensionnels pour la comparaison de deux échantillons de chromatine immunoprécipitée

Caroline Bérard¹, Marie-Laure Martin-Magniette^{1,2}, Alexandra To³, François Roudier³, Vincent Colot³ et Stéphane Robin¹.

¹ UMR AgroParisTech/INRA MIA 518, 16 rue Claude Bernard, PARIS Cedex 05.

² UMR INRA 1165 - CNRS 8114 - UEVE URGV, 2 rue Gaston Crémieux, EVRY.

³ UMR CNRS 8186, Département de Biologie, 46 rue d'Ulm, PARIS Cedex 05.

caroline.berard@agroparistech.fr , marie_laure.martin@agroparistech.fr
to@biologie.ens.fr , roudier@biologie.ens.fr
colot@biologie.ens.fr , stephane.robin@agroparistech.fr

Résumé L'immunoprécipitation de la chromatine (ChIP) permet d'étudier les interactions entre les protéines et l'ADN ainsi que différents états chromatinien. Le ChIP-chip est une technique combinant l'immunoprécipitation de la chromatine avec le principe des puces à ADN, ce qui permet une étude à l'échelle du génome. Nous nous intéressons ici à l'analyse des différences entre deux échantillons d'ADN immunoprécipité. Biologiquement, on s'attend à distinguer quatre groupes différents : un groupe d'ADN non-immunoprécipité, un groupe d'ADN immunoprécipité identiquement dans les deux échantillons et deux groupes dans lesquels l'ADN est immunoprécipité en quantités différentes. Nous modélisons ces données par un mélange de gaussiennes bidimensionnelles à quatre composants. Les matrices de variance sont contraintes afin d'intégrer des connaissances biologiques. Les paramètres sont estimés par l'algorithme EM. Nous appliquons cette méthode pour étudier la différence de méthylation d'une histone entre l'écotype sauvage de la plante modèle *Arabidopsis thaliana* et un mutant.

Mots-clés : Mélange gaussien, décomposition spectrale, algorithme EM, ChIP-chip.

Résumé Chromatin immunoprecipitation (ChIP) enables to investigate interactions between proteins and DNA and also various chromatin states. ChIP-chip is a well-established procedure combining chromatin immunoprecipitation with DNA microarrays, which allows a study of the whole genome. We are interested in the analyze of the differences between two immunoprecipitated DNA samples. From a biological point of view, we expect to distinguish four different groups : a group of non-immunoprecipitated DNA, a group of immunoprecipitated DNA in both samples, and then two groups in which DNA is differently immunoprecipitated. We propose to model these data with a mixture of two-dimensional Gaussians with four components. Biological knowledges are included as constraints on the variance matrices. The parameters are estimated by the EM algorithm. This method is applied to NimbleGen data in order to study the histone methylation difference between the wild ecotype of the model plant *Arabidopsis thaliana* and a mutant.

Keywords : Gaussian mixture, eigenvalue decomposition, EM algorithm, ChIP-chip.