Gene-specific identification of spatial associations between epigenomic marks

Gabriele Schweikert[*][†][1] and Guido Sanguinetti[‡][2]

[1]University of Edinburgh – United Kingdom
[2]University of Edinburgh – University of Edinburgh, University of Edinburgh – France

Abstract

High throughput assays such as ChIP-Seq and CLIP/CRAC-Seq return complex data sets documenting the interactions between proteins and DNA or RNA respectively. These data often display rich spatial patterns, detailing complex local variations in binding which can in some cases be interpreted mechanistically in terms of local biological factors, such as local chromatin accessiblity, and binding of DNA/ RNA binding proteins which may act as biological cofactors. Such spatial patterns have been recently exploited to define powerful statistical testing procedures (Schweikert et al, BMC Genomics 2013), yet, to our knowledge, no methods attempt to relate spatial patterns across different marks. While some papers attempt to relate the global correlative behaviour of different epigenetic marks in a global network of histone modifications (Lasserre et al, PLoS CompBio 2013, Mitra et al, JASA 2013), local associations may be more reflective of functional interactions between different marks, and therefore identifying such associations could be an important step towards integrative studies of larger data sets. Here we proposed a non-parametric, kernel-based method to detect associations between spatial patterns which utilizes the Hilbert-Schmidt independence criterion (HSIC) to capture non-linear interactions between different data distributions (Gretton et al , 2007). The HSIC relies on embedding the distribution of reads in a mark within a reproducing kernel Hilbert space; the independence between the two distributions can then be quantified as a trace of products of kernel matrices.
We illustrate the approach on a large scale case study involving ChIP-seq data from tens of different histone modifications in human embryonic stem cells, harvested as part of the roadmap epigenome project. Our results indicate that the HSIC is effective at capturing non-trivial associations in post-genomic data, and could therefore be a valuable tool for exploratory analyses in integrative biology.

References:

Schweikert, G.B. Cseke, B, Clouaire, T, Bird, A. and Sanguinetti, G (2013) MMDiff: quantitative testing for shape changes in ChIP-seq data sets, BMC Genomics 14:826

Mitra, R, Mueller, P, Liang, S, Yue, L and Ji, Y (2013), A Bayesian Graphical Model for ChIP-seq data on Histone Modifications, J. Am. Stat. Ass. 108 (501), 69-80.

[*]Speaker
[†]Corresponding author: G.Schweikert@ed.ac.uk
[‡]Corresponding author: gsanguin@staffmail.ed.ac.uk

Lasserre, J. Chung, H-R, and Vingron, M (2013) Finding associations among histone modifications using sparse partial correlation networks, PLoS Comp. Biol. 9(9): e1003168

Gretton, A, Fukumizu, K, Teo, C-H, Song, L, Scholkopf, B and Smola, A, A Kernel Statistical Test of Independence, NIPS 2007.
Roadmap Epigenomics Consortium, et al (2015) Integrative analysis of 111 reference human epigenomes, Nature 518, 317–330