HMCan – a tool to detect chromatin modifications in cancer samples using ChIP-seq data

Haitham Ashoor[1], Aurélie Hérault[2,3], François Radvanyi[2,3], Vladimir B. Bajic[1], Emmanuel Barillot[2,4,5,6], ValenDna Boeva[2,4,5,6]

[1] King Abdullah University of Science and Technology (KAUST), Computational Bioscience Research Center (CBR), Computer, Electrical and Mathematical Sciences and Engineering Division, Thuwal, Saudi Arabia; [2] Institut Curie, Paris, France; 3UMR 144 CNRS; [4] INSERM, U900; [5] Mines Paris Tech, Fontainebleau, France; [6] Equiqe Laebllisee Ligue, Nationale Contre le Cancer

Introduction: Epigenetic changes often play an important role in cancer development. By introducing local or regional epigenetic silencing (RES), cancer cells may limit expression of tumor suppressor genes. One way to study epigenetic silencing in cancer is to explore patterns of tri-methylation of lysine 27 of histone 3 (H3K27me3), associated with Polycomb-mediated Repression. Though several tools were created to enable detection of histone marks in ChIP-seq data from *normal* samples, it was unclear whether these tools can be applied to ChIP-seq data generated from *cancer* samples. The challenge comes form the fact that cancer genomes are often characterized by frequent copy number alterations: gains and losses of large regions of chromosomal material. Copy number alteration may create a substantial statistical bias in the evaluation of histone mark signal enrichment and result in underdetection of the signal in the regions of loss and overdetection of the signal in regions of gain.

Results: We present HMCAN (Histone Modification in Cancer), a tool specially developed to analyze histone modification ChIP-seq data produced from cancer genomes. HMCan corrects for the GC- and copy number bias and then applies Hidden Markov Models (HMMs) to detect the signal in the normalized profile. We showed that HMCan provided a significantly better accuracy of predictions than commonly used tools such as CCAT, MACS, SICER. Of note, MACS and SICER were biased towards regions of copy number gain, while CCAT and HMCan did not demonstrate such a bias.We also generated a ChIP-seq dataset for the repressive histone mark H3K27me3 in a bladder cancer cell line. On this dataset, HMCan was able to recover the expected shape of the H3K27me profile in the vicinity of gene transcription start sites. HMCan predictions included, for example, previously detected H3K27me3 marks on the *DLEC1* gene, which is commonly inactivated in various carcinomas, as well as the *HOXD* gene cluster, which plays a crucial role in normal cell development and proliferation.

Availability: C++ source code is available at http://sourceforge.net/p/hmcan