Fig | We performed a motif matching algorithm within the genomic region in which the CpG sites were tested for differential methylation between CML and control. First, we obtained position frequency matrices from Jaspar [1], Uniprobe [2] and Hocomoco [3] repositories. Then, we used Biopython [4] to evaluate the corresponding position weight matrix [5] and bit-score threshold [6] of each position frequency matrix. The motif matching procedure was performed using MOODS [7]. The threshold corresponded to a p-value of 10-4, and a regularizing factor of 0.1 was added to all position frequency matrices. We post-process all putative binding sites by removing the matches with bit-score log-odds ratio (with regard to the corresponding bit-score threshold) less than 4. This procedure has shown to provide reliable results [8]. In this Figure, we exhibit the three motifs (REST, ZBTB33 and ZBT7B) that matched close to the region with the highest difference between the methylation levels of CML and control. All motifs matched to the reverse strand, in the promoter region of MTSS1 (also in reverse strand). ZBTB33 is associated with the recruitment the N-CoR repressor complex to promote histone deacethylation (formation of repressive chromatin structure). ZBT7B has been linked with the repression of certain genes such as type I collagen. REST represses neuronal genes in non-neuronal tissues by associating with two distinct corepressors: mSin3 and CoREST. As evidence of repressive chromatin formation, we were able to observe a CoREST ChIP-seq peak in this particular locus in HepG2 cell type.

[1] Mathelier, A. et al. (2014). JASPAR 2014: an extensively expanded and updated open access database of transcription factor binding profiles. Nucleic Acids Research, 42(D1), D142-D147.

[2] Robasky, K. and Bulyk, M. L. (2011). UniPROBE, update 2011: expanded content and search tools in the online database of protein-binding microarray data on protein-DNA interactions. Nucleic acids research, 39 (Database issue).

[3] Kulakovskiy, I. V. et al. (2013). HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. Nucleic acids research, 41 (Database issue).

[4] Cock, P. J. A. et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics, 25(11), 1422-1423.

[5] Stormo, G. D. (2000). DNA binding sites: representation and discovery. Bioinformatics, 16(1), 16-23.

[6] Wilczynski, B. et al. (2009). Finding evolutionarily conserved cis-regulatory modules with a universal set of motifs. BMC bioinformatics, 10(1), 82+.

[7] Korhonen, J. et al. (2009). [MOODS: fast search for position weight matrix matches in DNA sequences.](http://bioinformatics.oxfordjournals.org/cgi/content/abstract/25/23/3181) Bioinformatics 25(23), 3181-3182.

[8] Gusmao, E.G. et al. (2014). Detection of active transcription factor binding sites with the combination of DNase hypersensitivity and histone modifications. Bioinformatics, 30(22), 3143-3151.