Profiling DNA modifications at base pair resolution reveals expanded low gene body methylation and associated cell specific epigenomic states

Nicolas Walker[2*], Marcela K. Sjöberg[1*], Catalina Vallejos[3], Hui Shi[2], Michael J. Booth[4], Kim Wong[1], Martin Del Castillo-Velasco Herrera[1], Martin Bachman[7], Anastasiya Kazachenka[2], Richard Gunning[1], Mamunur Rashid[1], Lira Mamanova[6], Alexandra Ashcroft[1], Laurent Vanhille[5], Salvatore Spicuglia[5], Shankar Balasubramanian[7], Sarah Teichmann[8], John Marioni[8], David J. Adams[1,+], Anne C. Ferguson-Smith[2,+]

[1] *Experimental Cancer Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.*
[2] *Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK*
[3] *European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK.*
[4] *Department of Chemistry, University of Cambridge, Cambridge, UK.*
[5] *Inserm U1090, Technological Advances for Genomics and Clinics (TAGC), F-13009 Marseille, France; Aix-Marseille University UMR-S 1090, TAGC, F-13009 Marseille, France*
[6] *The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.*
[7] *Department of Chemistry, University of Cambridge, Cambridge, UK; Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK.*
[8] *European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK; Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.*

Abstract

Dynamic changes in methylation can be identified in gene bodies however their extent and impact on gene expression is not well understood. As part of the BLUEPRINT epigenome project we performed genome-wide comparative epigenomics and transcriptomics on purified quiescent mouse naïve CD4+ T and B lymphocytes. Integrative analysis of single-nucleotide resolution 5-methylcytosine and 5-hydroxymethylcytosine (5hmC), histone marks and RNA-seq profiles reveal contrasting epigenetic states between cell types. We identified T-cell-specific 5hmC at overall levels comparable to ES cells. We used a computational strategy to identify long range T cell specific 5hmC Enriched Domains (HEDs) and demonstrate several spatial patterns associated with gene bodies, showing in particular that HEDs are not always constrained within gene boundaries, and are also found in intergenic regions. We further computed methylation differences between the two lymphocytes and identified remarkable expansion of low methylation in the gene bodies of T cells proximal to some 5hmC domains, and demonstrate these are also conserved between human and mouse. Repressed and active low methylated genes contain expanded domains of H3K27me3 and H3K27Ac/H3K4me3 respectively. We explore the relationship between gene expression and low methylated genes at both the population and single cell level. The latter providing new insights between epigenetic states and gene expression characteristics.