



Searching for the chromatin determinants of human hematopoiesis

Enrique Carrillo-de-Santa-Pau^{1,3}, David Juan^{1,3}, Vera Pancaldi¹, Felipe Were¹, Ignacio Martin-Subero², Daniel Rico^{1,4} & Alfonso Valencia^{1,4}

1. *Structural Biology and BioComputing Programme, Spanish National Cancer Research Centre (CNIO)*

2. *Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Department of Anatomic Pathology, Pharmacology and Microbiology, University of Barcelona, Barcelona, Spain.*

3. *Co-first authors*

4. *Corresponding authors*

Abstract

As part of the BLUEPRINT Consortium, we are characterizing the epigenomes of blood cells to understand how changes in chromatin are connected with the different lineage differentiation options. In this work, we present our analyses using hematopoietic stem cells (HSCs), monocytes, macrophages, neutrophils, B-cells (naive from venous blood and tonsil-derived germinal center B-cells) and T-cells (CD4 and CD8), combining hematopoietic samples from BLUEPRINT, ENCODE and NIH Epigenomic Roadmap. We have developed a bioinformatics pipeline to generate a 'chromatin space' where the different cell types are clustered by epigenetic similarity.

Our analysis is based on Multiple Correspondence Analysis (MCA), the analog of Principal Component Analysis when working with categorical data. We used our previous approach to deal with protein multiple alignments (Rausell, Juan et al PNAS, 2010) with critical enhancements to deal with millions of regions in the same analysis.

The analysis of the orthogonal dimension of the space allows us to identify chromatin determinant regions (CDRs), genomic regions with different epigenomic characteristics between the different groups. Functional enrichment analysis of the neighbouring genes suggests that the chromatin state in this regions could be directly linked with the different cell identities. Our analytical approach allows to combine samples from different sources and identify the regions for which chromatin status associates with cell lineage determination or disease conditions.