Bio-IT task with *meth-atlas*

# Introduction

*Meth-atlas* is a bioinformatic tool that performs deconvolution analysis on epigenomic data, specifically on methylome data. It takes a *csv* file with samples data to be analyzed and a reference *csv* file, and returns a new *csv* file with the deconvoluted samples and a stacked bar chart.

Rows in both input files correspond to CpG sites, and their values in each column to the level of DNA methylation in that site (for each sample or for each cell type, depending on the file). The output consists of the proportion of each cell type in each provided sample.

# Objectives

To perform deconvolution analysis on a provided file containing bulk DNA methylation beta values for a series of samples, using a reference that consists of seven blood cell types methylation profiles, using *meth-atlas*. To determine possible sample clusters and compute differential methylation analysis between them.

# Methods

*Meth-atlas* was installed by cloning the corresponding github repository. In order to run it, a conda environment containing all necessary *Python3* packages was installed using a provided *yaml* file. The program was runned via command line, providing the necessary files as arguments.

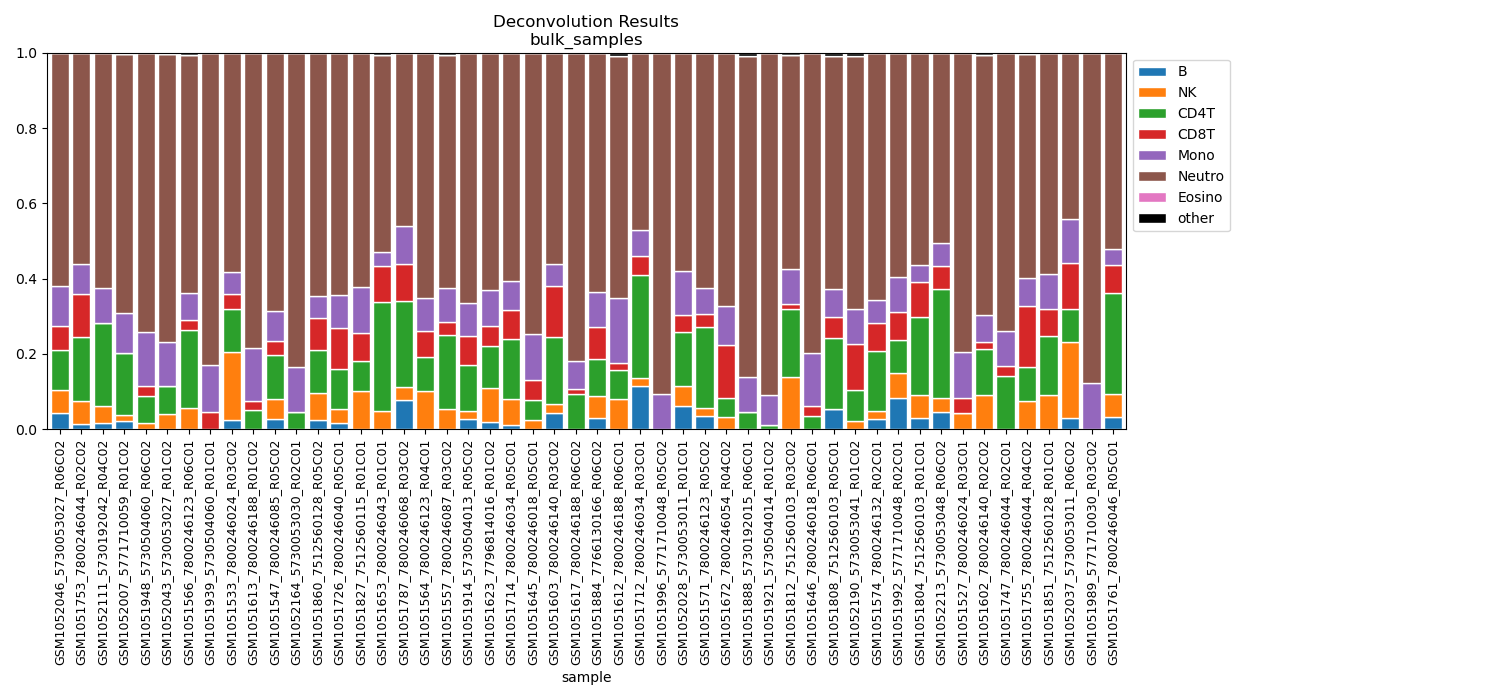
In order to determine possible clusters in the analyzed samples and CpGs, Pearson correlation analysis and PCA were carried out using *scikit-learn*. Next, the potential clusters were used to search for differentially methylated CpGs (DMGs) using *R-limma*. Furthermore, the array’s annotation package in R was used to determine genes associated with those CpGs.

Finally, correlation analysis between the DMGs and the previously determined CpG clusters was done; and enrichment analysis for the DMGs was also carried out to better understand the biology behind.

All scripts, outputs and data are available in the github repository: <https://github.com/CDSchuster/Bio-IT-VIB>.

# Results

After running *Meth-atlas*, two files were obtained. One is a *csv* file containing the proportions of each blood cell type for the 50 samples that were provided to analyze. The second file is a stacked bar chart representing the same results that are in the *csv* file, allowing for an easier and more visual way to discuss the data.



**Fig 1.** Deconvolution results from Meth-atlas for the given DNA methylation data

# Discussion

As Fig. 1 shows, the vast majority of cells in each sample are neutrophils. Furthermore, a considerable number of samples have a significant number of T helper cells and monocytes. On the other hand, cell types like B cells, eosinophils and natural killer cells (NK cells) are negligible in most analyzed samples. Finally, it must be noted that meth-atlas detected small percentages of other cell types that were not present in the reference data.