

Data report on the methylation of cell-free DNA

Daniel, Thomas & Hans-Henrik

March 6, 2024

Contents

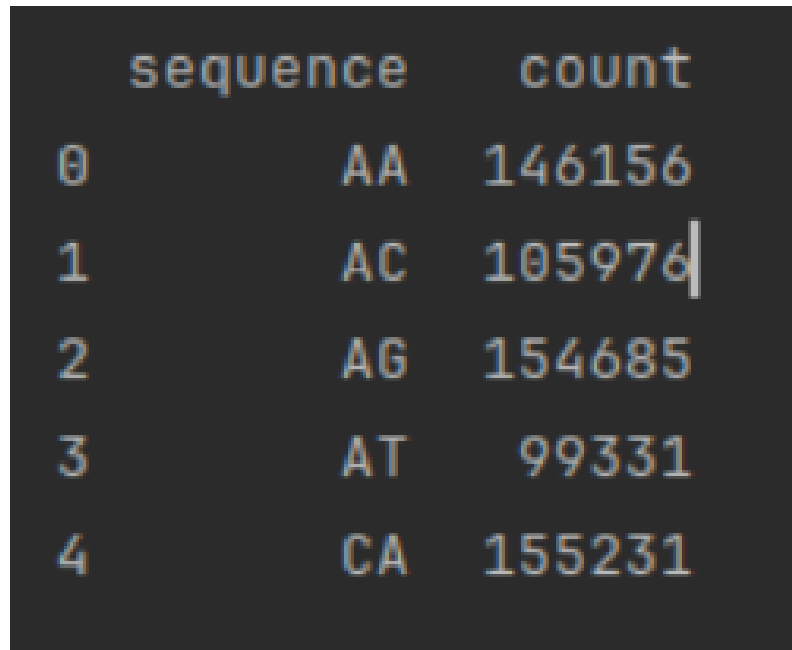
1	Introduction	2
2	The data	3

1 Introduction

In this project we will be looking at attempting to implement a model to predict, based on data from blood analysis, whether a given person potentially has cancer. The method in which we will be doing this, is by looking at methylized/unmethylized cell-free dna in peoples blood, specifically the fragmentation patterns. This is preferable to other methods, since it is both less invasive, and less expensive.

The data we will be looking at in this project, has been provided to us by Søren Bessenbacher, and contains information regarding fragmentation patterns for approximately 230 people diagnosed with cancer, and 230 people as a control group.

2 The data



	sequence	count
0	AA	146156
1	AC	105976
2	AG	154685
3	AT	99331
4	CA	155231

Figure 1: An example of the data

In the first column we have the k-mer, in this specific figure it is a 2-mer, and in the second column we have the count of that specific sequence. In addition to the control and the test samples, we also have a background file, which details the total number of combinations of sequences.