

Introduction into machine learning and analysis of Breast Cancer Proteomes

Theme09 - Introduction to Machine Learning

Mats Slik

344216

BFV3

October 4, 2022

Dave Langers (LADR)

Bart Barnard (BABA)

Introduction into machine learning and analysis of Breast Cancer Proteomes

Theme09 - Introduction to Machine Learning

Mats Slik

344216

Bioinformatics

Institute for Life Science & Technology

Hanze University of Applied Sciences

Dave Langers (LADR)

Bart Barnard (BABA)

October 4, 2022

Abstract

the data were used to assess how the mutations in the DNA are affecting the protein expression landscape in breast cancer. Genes in our DNA are first transcribed into RNA molecules which then are translated into proteins. Changing the information content of DNA has impact on the behavior of the proteome, which is the main functional unit of cells, taking care of cell division, DNA repair, enzymatic reactions and signaling etc. my question is: Are there different ways to categorize breast cancer based on protein expression data, with machine learning being used to classify them without using the pam50 proteins?

Table of Contents

Abstract	i
List of Abbreviations	iii
List of Figures	iii
List of Tables	iii
1 Introduction	1
2 Methods	2
3 Results	3
4 Discussion and Conclusion	7
5 References	8
6 Appendices	9

List of Abbreviations

EDA	Exploratory Data Analysis
TCGA	The cancer Genome Atlas Program
CPTAC	Clinical Proteomic Tumor Analysis Consortium
DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid

List of Figures

List of Tables

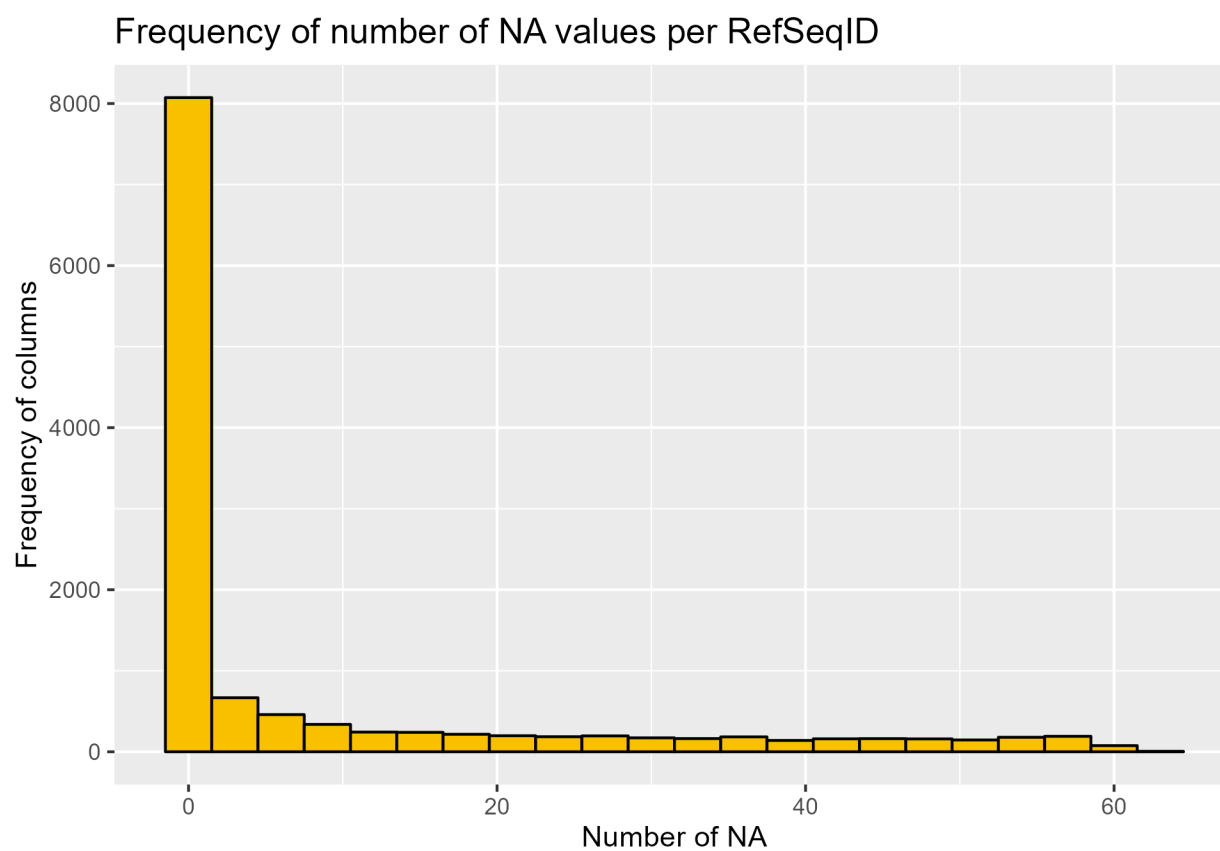
1 Introduction

??

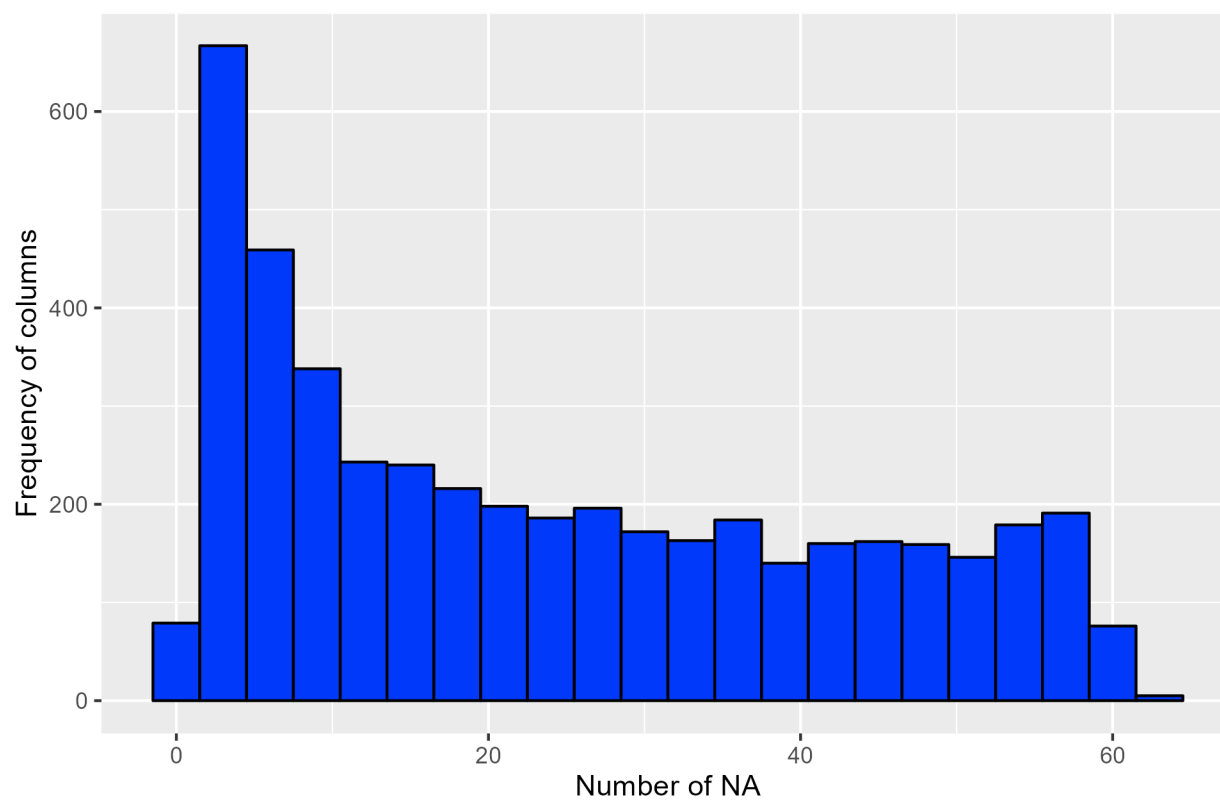
2 Methods

3 Results

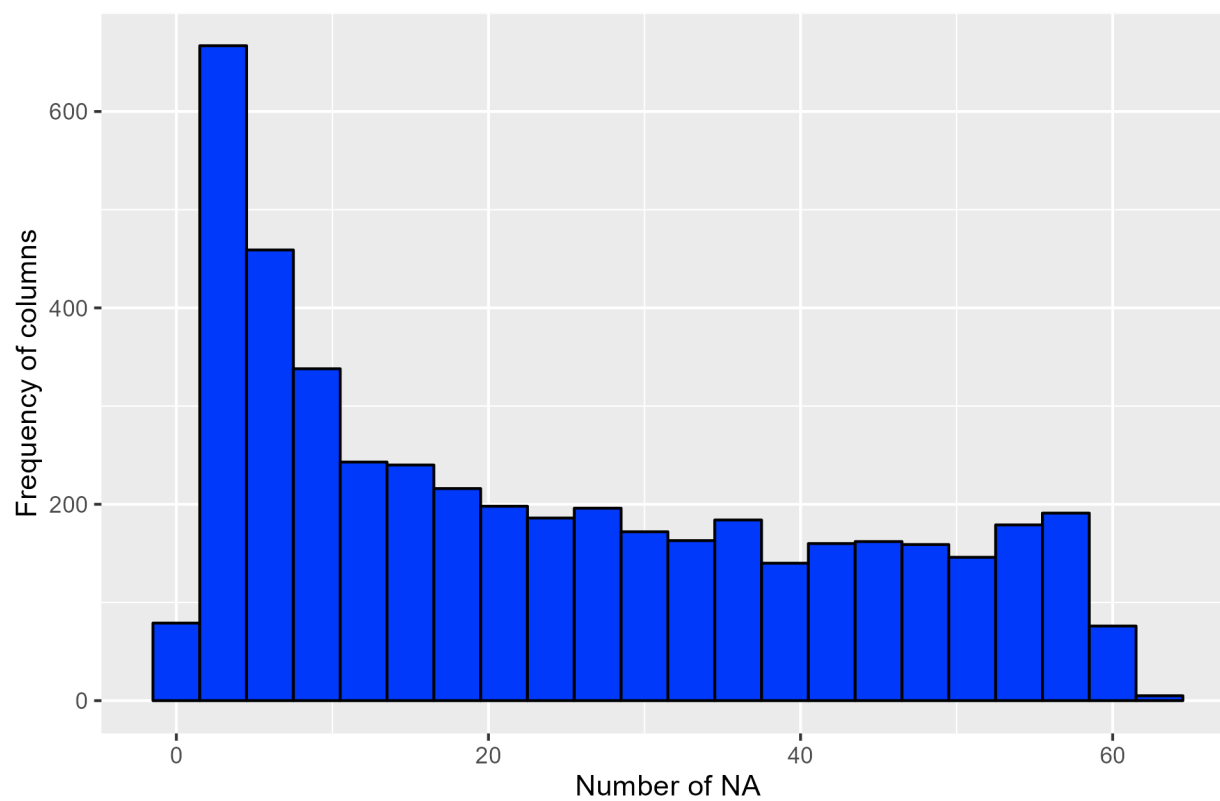
When looking at the dimensions of the data set we can see there are allot of proteins



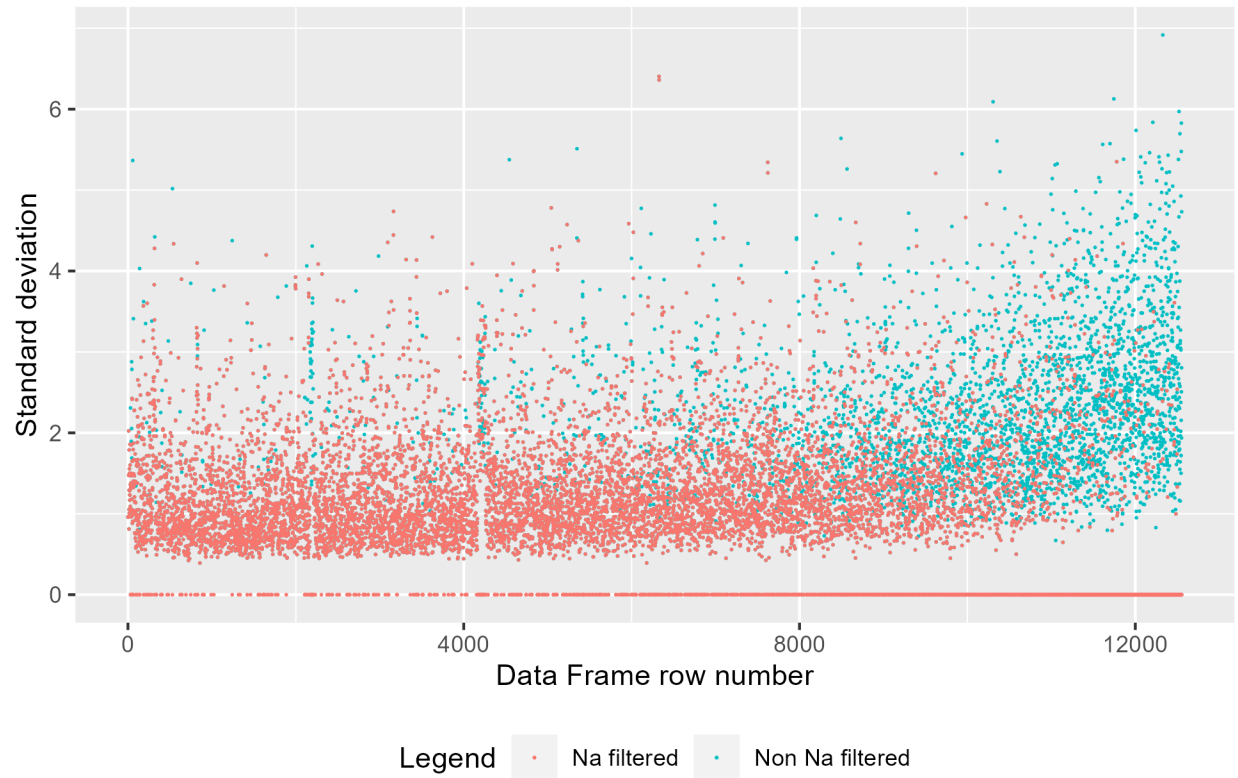
Frequency of number of NA values per RefSeqID with 0 omitted



Frequency of number of NA values per RefSeqID with 0 omitted



Density plot for standard deviation of protein expression



4 Discussion and Conclusion

5 References

Mertins, Philipp, D R Mani, Kelly Ruggles, Michael Gillette, Karl Clauser, Pei Wang, Xianlong Wang, et al. 2016. “Proteogenomics Connects Somatic Mutations to Signaling in Breast Cancer.” *Nature* 534 (May). <https://doi.org/10.1038/nature18003>.

6 Appendices