The human gut microbiome and its association to immunotherapy response in metastatic melanoma patients

Master's Thesis in Bioinformatics, 30 ECTS

Rikke Myrhøj Jensen, 201605249

Supervised by

Associate Professor Palle Villesen
&
Ph.D. and CEO Louise Bruun Thingholm

Bioinformatics Research Centre

Aarhus University

June 3, 2022

AARHUS
UNIVERSITY

Acknowledgements

I would like to thank my external supervisor Ph.D and CEO Louise Bruun Thingholm, not only for for providing the dataset and allowing me the opportunity to conduct this thesis in collaboration with Biomcare, which has been an inspiring and motivating setting for day-to-day work, but also for providing valuable input and feedback throughout the process.

Also a huge thanks to Anne Mathiesen, who has laid the groundwork for my data analysis in her own thesis, been a great sparring partner and has kept me grounded, sane and focused from start to finish.

A special thanks to associate prof. Palle Villesen for expert guidance, for providing a safe space to express doubts and frustrations and for keeping me on track during this process.

Finally a thanks to my fellow master students Stine Katrine Rye Østergaard and Rikke Møller Larsen for rewarding scientific discussions of microbial data analysis and for the continued emotional support.

Abstract

Metastatic melanoma is associated with notoriously high mortality rates. Immunotherapy treatment of metastatic melanoma has increased overall survival rates. However, less than 50% of patients respond to immunotherapy, and treatment can in some cases have severe adverse side effects. For this reason, it is of great interest to elucidate the underlying reasons for why some patients respond to immunotherapy treatment while other do not.

Multiple studies have linked immunotherapy response to the composition of the gut microbiome, however attempts at identifying specific bacteria that consistently associate with the response have so far been futile.

This thesis conducts a meta-analysis of 'pre-immunotherapy treatment' faecal shotgun metagenomic data from metastatic melanoma patients, with the aim of clarifying whether it is possible to detect any clear signals in the data, that can predict if a patient will respond to immunotherapy or not. This includes an exploration of whether considering the community-structure of the microbiome rather than the isolated effect of individual microbial taxa can improve results.

This was achieved by the application and evaluation of machine learning methods for classification of immunotherapy response based on microbial relative abundances.

Results indicate patterns in the data related to immunotherapy response are not general, but instead a signal might be found within subgroups of similar individuals. However, as the study is severely underpowered, no conclusive results can be drawn.

It is suggested that future efforts should be put into increasing the number and quality of human microbiome studies, including consistent, standardized collection of patient information. Furthermore the field would benefit from the creation of standards for development and application of machine learning methods that allow trustworthy interpretation of microbiome data, even for non-experts.

List of Abbreviations

AMP	Antimicrobial peptide
ANI	Average Nucleotide Identity
CLR	Centered Log Ratio
CTLA-4	Cytotoxic T-Lymphocyte-associated Antigen 4
CV	Cross Validation
EDA	Exploratory Data Analysis
GLM	Generalized Linear Model
ICI	Immune Checkpoint Inhibitor
IgA	Immunoglobulin A
KNN	K-Nearest Neighbours
MAG	Metagenome-Assembled Genome
NR	Non-Responder
PCA	Principle Component Analysis
PD-1	Programmed Cell Death Protein 1
R	Responder

Contents

Acknowledgements **i**

Abstract **ii**

List of Abbreviations **iii**

1 Introduction **1**
1.1 Metastatic melanoma and immunotherapy 1
1.2 Characterisation of microbial assemblages 2
1.3 The human gut microbiome 4
1.4 The human gut microbiome & immunotherapy response 5
1.5 Detection of microbial biomarkers 6
1.6 Aim 7

2 Materials **8**
2.1 The dataset 8
2.1.1 Patient stratification 8
2.1.2 Metadata 8

3 Methods **9**
3.1 Preprocessing 9
3.1.1 Quality control 10
3.1.2 Rarefaction 11
3.1.3 Genome database creation 11
3.1.4 Mapping & Profiling 12
3.1.5 Relative abundances 12
3.1.6 Prefiltering 15
3.2 Machine Learning 15
3.2.1 Principle component analysis 17
3.2.2 Classification 18
3.2.3 Model comparison 20

4 Results & Discussion **22**
4.1 Overview 22
4.2 Principle component analysis 22
4.3 Model performance 24
4.4 Variable importance 28

5 Conclusion **31**

References **33**

Supplementary - Missing **40**