Microbial signatures of response to anti-PD1 therapy in metastatic melanoma patients:
Towards novel biomarker development with artificial intelligence algorithms

Krzysztof Odrzywołek[1,2], Paweł Biernat[1], Szymon Wojciechowski[1], Agata Szymanek[1], Jan Havránek[1,3], Zuzanna Karwowska[1], Jan Majta[1,4], Kaja Milanowska-Zabel[1]

1 Ardigen, Krakow, Poland, 2 Department of Computer Science, Faculty of Computer Science, Electronics and Telecommunications, AGH University of Science and Technology, Krakow, Poland, 3 Microbial Genomics Group, Malopolska Centre of Biotechnology, Jagiellonian University, Krakow, Poland, 4 Department of Computational Physics and Bioinformatics, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Krakow, Poland

BACKGROUND

- The body of evidence **linking gut microbiota with the response to cancer therapy** has been rapidly growing in the past few years.
- In 2018 works by groups of Wargo, Gajewski, and Zitvogel [1-3] have shaped the field, describing the correlation between the gut microbiota and the response to immune checkpoint blockade (ICB).
- Those studies also reported the **increase of the ICB response rate by gut microbiota modulation.**

CHALLENGE

- Two technologies for studying microbiome – 16S rRNA sequencing and Shotgun Metagenomic Sequencing (SMS) – are widely applied, but the taxonomy-based approach of 16S has shaped the way most projects are designed.
- 16S sequencing is a cost-effective way to analyze the taxonomic composition of a sample. However, it is limited to genus/species level description, not allowing for functional explanation.
- In crucial research [1-3] the SMS potential was not used to its fullest. **Certain bacterial taxa have been identified as associated with response to checkpoint blockade therapy.** On the other hand, no functional explanation was provided for these findings.
- Beneficial bacterial strains identified in each paper were inconsistent between studies, and the findings were **not statistically significant, when validated on an external cohort.**
- The work by Gharaibeh and Jobin [4] presented the machine learning model that utilizes SMS-derived functional data. However, this **model was not validated on an external cohort, and stratification potential was not demonstrated.**

Fig. 1. Challenges for the immune checkpoint blockade therapies that are linked with the role of gut microbiota.

DATASETS

Dataset	Matson	Gopalakrishnan
Responders	15	14
Non-Responders	24	11
Total	39	25

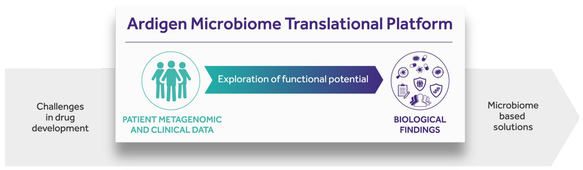

Fig. 2. Scheme of Ardigen's process of microbiome derived Dx/Tx development.

ANALYSIS OUTLINE

- Data analysis and interpretation were carried out using **Microbiome Translational Platform** (MTP) developed by Ardigen.
- Shotgun sequencing data of described cohorts were processed with **Microbiome Scout** tool that identified **Metagenomic Features** associated with the analyzed phenotype.
- Selection performance was measured by ROC AUC of the used classifier and by statistical significance (p-value) of subgroups separation in Time To Progression (TTP) analysis (where applicable).
- We used all SMS samples for building a study-specific feature space. Ardigen's Microbiome Scout builds metagenomic **Feature Space** with up to 30% of previously unknown metagenomic sequences (Fig. 3a).
- All the metagenomic features sequences were encoded into a functionally-valid space (Fig. 3b).
- For classification problems, Random Forest and Logistic Regression algorithms were used as implemented in Ardigen's Microbiome Scout (Fig. 3c).
- Each model was validated in 10x repeated 5-fold cross-validation.

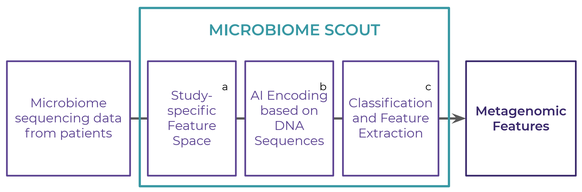

Fig. 3. Workflow of Microbiome Scout procedure with separate stages of building study-specific feature space (a), encoding obtained features into functionally-valid space (b) and machine learning classifiers (c).

IMPROVED SCORING SYSTEM

- Standard approach for functional analysis is to map metagenomic reads to the catalog of known genes. Than, genes are quantified by the number of mapped reads
- Ardigen encodes those reads to Metagenomic Features with AI-based algorithms and apply a proprietary scoring system (**BSScore**) to quantify their abundance

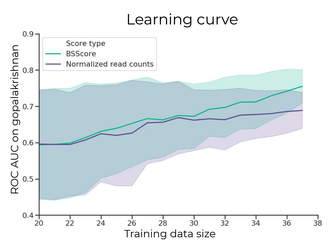

Fig. 4. Learning curve in function of training dataset size. We can observe saturation of NRC approach, and constant growth of BSScore

- BSScore algorithm aims at understanding the biological meaning of such mapped read and reference sequence, not only marking similarity
- The key principle of this approach is to train an algorithm to detect metagenomic sequences with similar biological information
- As the training dataset grows, BiomeScout is able to retrieve this signal, being insensitive to growing noise level

MICROBIOME-BASED STRATIFICATION

- The group predicted to respond to therapy had significantly longer TTPs (median=66) compared to those predicted not to respond (median=92).
- **This is the first reported stratification of checkpoint blockade patients with an external dataset used for validation.**
- Unprecedented cross-cohort performance of our model indicates the presence of the common gut microbiome' functional features influencing therapy outcomes that are **shared by patients across studies**.

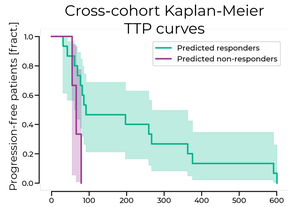

Fig. 5. Kaplan-Meier TTP curves of patients predicted to respond (green) and not to respond (pink) to anti-PD-1 treatment. Patients predicted to respond stayed progression-free significantly longer (p<0.05).

THE CHALLENGE OF THE UNKNOWN

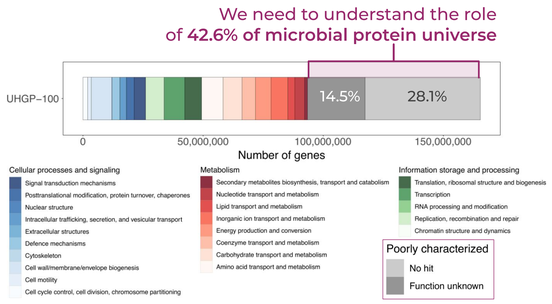

Fig. 6. Functional annotations of protein sequences stored in Unified Human Gastrointestinal Genome (UHGP [5]) - the largest gut microbiome protein collection currently available.

- Almost half of bacterial proteins are not understood in terms of its function (Fig 6.).
- Metagenomic sequencing studies detect more and more novel, unexplained proteins.
- Having this understanding gap growing, Ardigen aims at leveraging the information we already have.
- Advances in unsupervised learning methods enable the utilization of information stored in protein sequences catalogs, but requires functional discovery step to be included in the procedure.

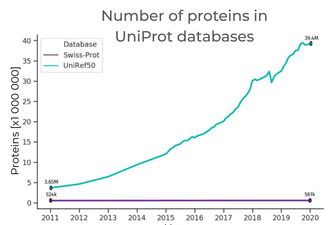

Fig. 7. Dynamic of growth of UniRef50 (unreviewed & automatically annotated) and Swiss-Prot (reviewed & manually-annotated) databases in 2010s

ARDIGEN'S FUNCTIONAL DISCOVERY ALGORITHM

- Ardigen develops a Functional Discovery algorithm for unannotated microbial proteins.
- We tested our approach on manually curated catalog of protein sequences: Swiss-Prot.
- The algorithm was trained to predict multiple functional annotations with high performance.
- Obtained results (Fig. 8) show that Functional Discovery algorithm properly assigns functions to unannotated sequences.

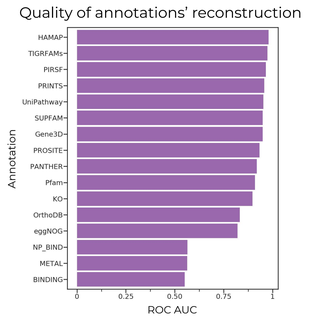

Fig. 8. Proportion of proteins from Swiss-Prot database accurately annotated with Ardigens algorithm

- Initial benchmark of Functional Discovery module's performance was done on protein sequences from UHGP catalog.
- We selected sequences that were annotated with standard tools: EggNOG mapper and InterProScan.
- Ardigen returns the functional annotation for 45-95% more proteins despite sequential dissimilarity.

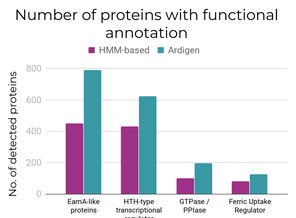

Fig. 9. Proportion of proteins from UHGP with assigned functional annotation with HMM-based, and Ardigen's method

UNDERSTANDING OF KEY FEATURES

- We used Functional Discovery algorithm for explanation of key Metagenomic Features allowing for stratification of patients.
- We compared our annotation with commonly used tools - PROKKA [6] and InterProScan [7].
- For a number of Metagenomic Features, Ardigen's algorithm provides functional interpretation of a significantly larger portion of underlying sequences.

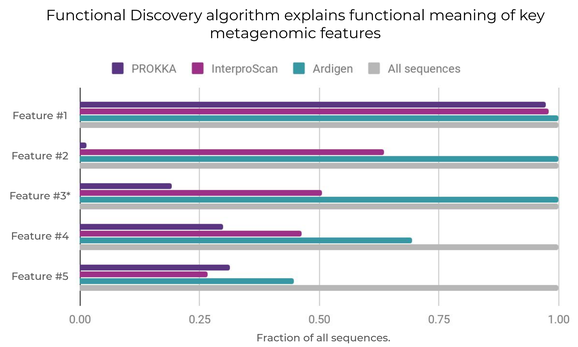

Fig. 10. Proportion of sequences behind Metagenomic Features that were functionally annotated with PROKKA, InterProScan, and Ardigen's Functional Discovery. For Feature #3, Ardigen provides less precise, yet broader understanding.

OUTLOOK AND DEVELOPMENT DIRECTIONS

- **Gut microbiota should be regarded as one of the crucial factors driving patients' response to checkpoint blockade therapy.**
- Functional traits in gut microbiota allow for **patients stratification in anti-PD-1 treatment of metastatic melanoma** with a mechanistic explanation.
- The small number of samples in cohorts (n=25 and n=39) is a significant limiting factor of presented work.
- Not only functional but also immunogenic features should be considered when microbiome role in immune checkpoint blockade is analyzed. This can be done with Ardigen's Epitope Discovery algorithm.

Fig. 11. Workflow of Epitope Discovery procedure that assessing immunogenicity of peptides derived from Metagenomic Features.

- Ardigen is running clinical studies **(NCT04136470, NCT04169867 and NCT04145232)** to gather stool along with blood (PBMC) and tumor (FFPE), to be analyzed with Translational Microbiome Platform.

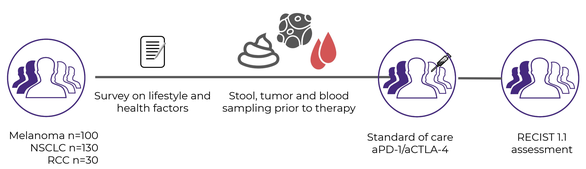

Fig. 12. Scheme of Ardigen's non-interventional clinical studies. We are gathering stool, blood (PBMC), and tumor (FFPE) samples from patients undergoing standard of care anti-PD-1 or anti-CTLA-4 therapy in NSCLC (n=130) and RCC (n=30). The response is measured with RECIST 1.1 criteria.

LITERATURE

1. Gopalakrishnan, Wargo, Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. Science. 2018 Jan 5;359(6371):97-103. doi: 10.1126/science.aan4236
2. Matson, Gajewski, The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. Science. 2018 Jan 5;359(6371):104-108. doi:10.1126/science.aao3290
3. Routy, Zitvogel Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors, Science 2018 Jan 5; 359(6371):91-97. doi:10.1126/science.aan3706
4. Gharaibeh, Jobin Microbiota and cancer immunotherapy: in search of microbial signals. Gut 2019 Mar; 68(3):385-388. doi:10.1136/gutjnl-2018-317220
5. Almeida, Finn, A unified sequence catalogue of over 280,000 genomes obtained from the human gut microbiome, bioRxiv, 2019 Oct
6. Seemann T., Prokka: rapid prokaryotic genome annotation,Bioinformatics 2014 Jul 15;30(14):2068-9
7. Jones, Hunter , InterProScan 5: genome-scale protein function classification, Bioinformatics 2014, May 1; 30(9): 1236–1240.

This work was funded by NCRD Fast Track Grant POIR.01.01.01-00-0347/17 awarded to Ardigen S.A.

contact: jan.majta@ardigen.com