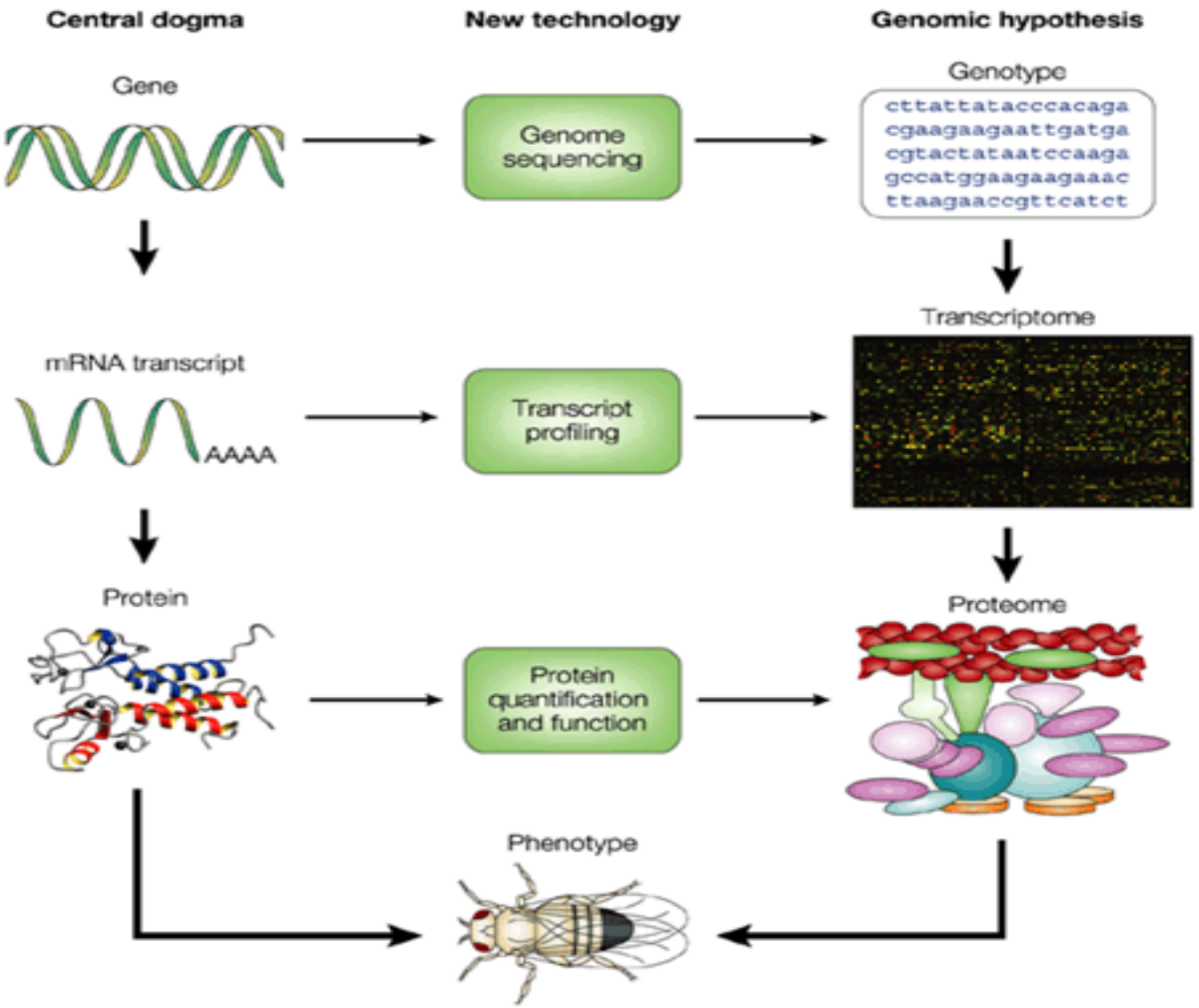
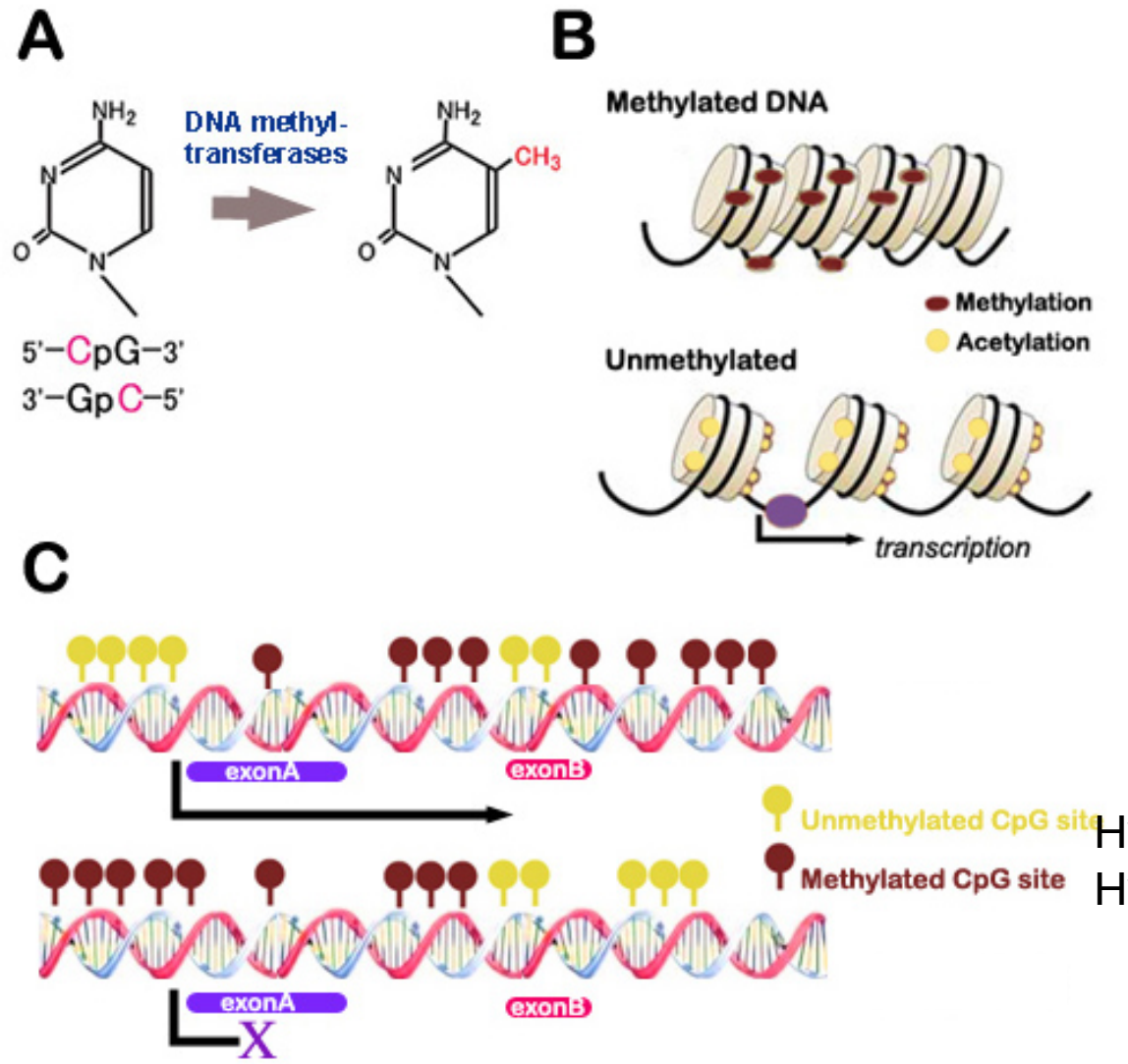
AI models to predict Microbiome related drug resistance

Uğur Sezerman

Acıbadem Üniversitesi

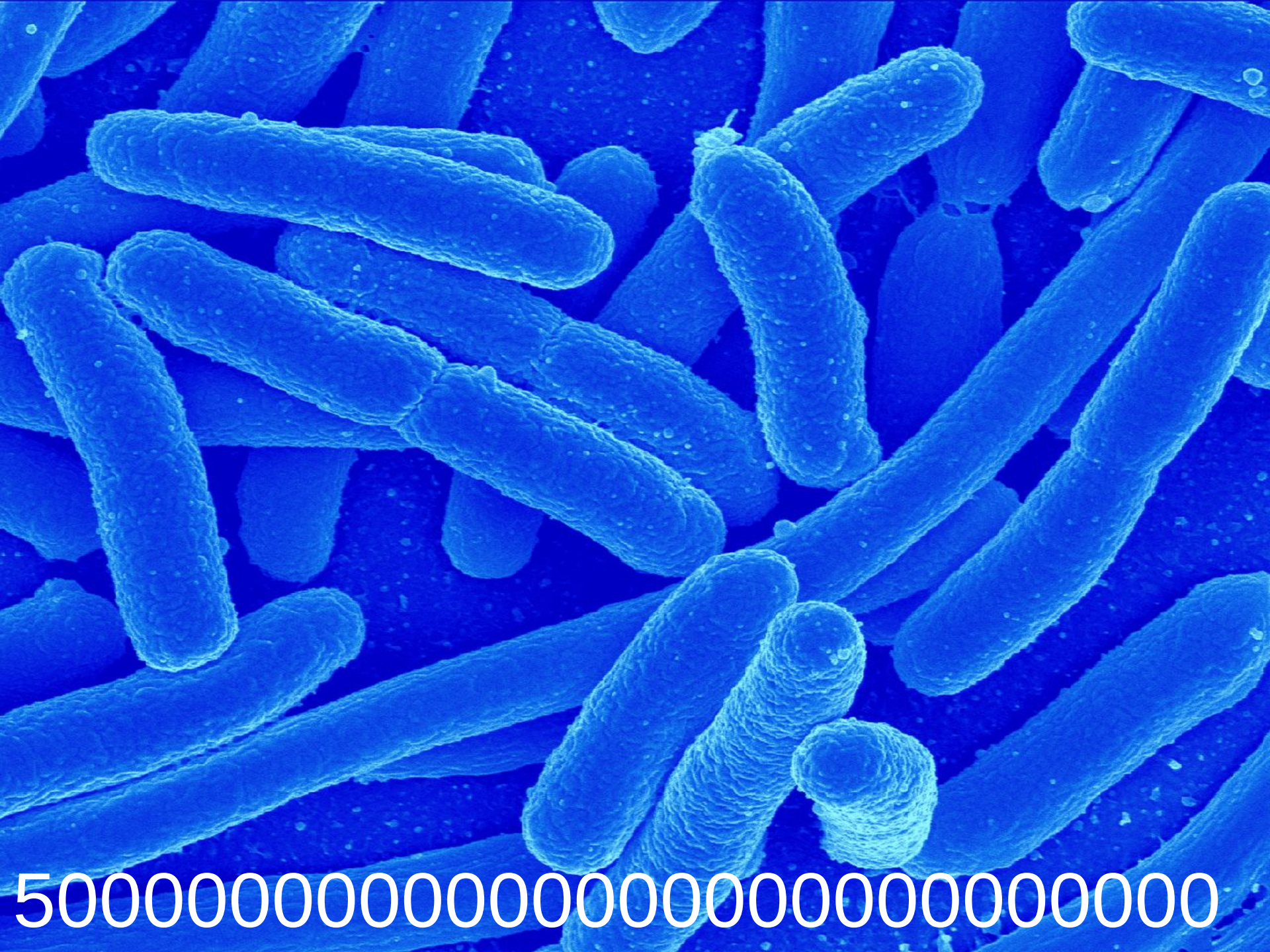
ACIBADEM
ÜNİVERSİTESİ

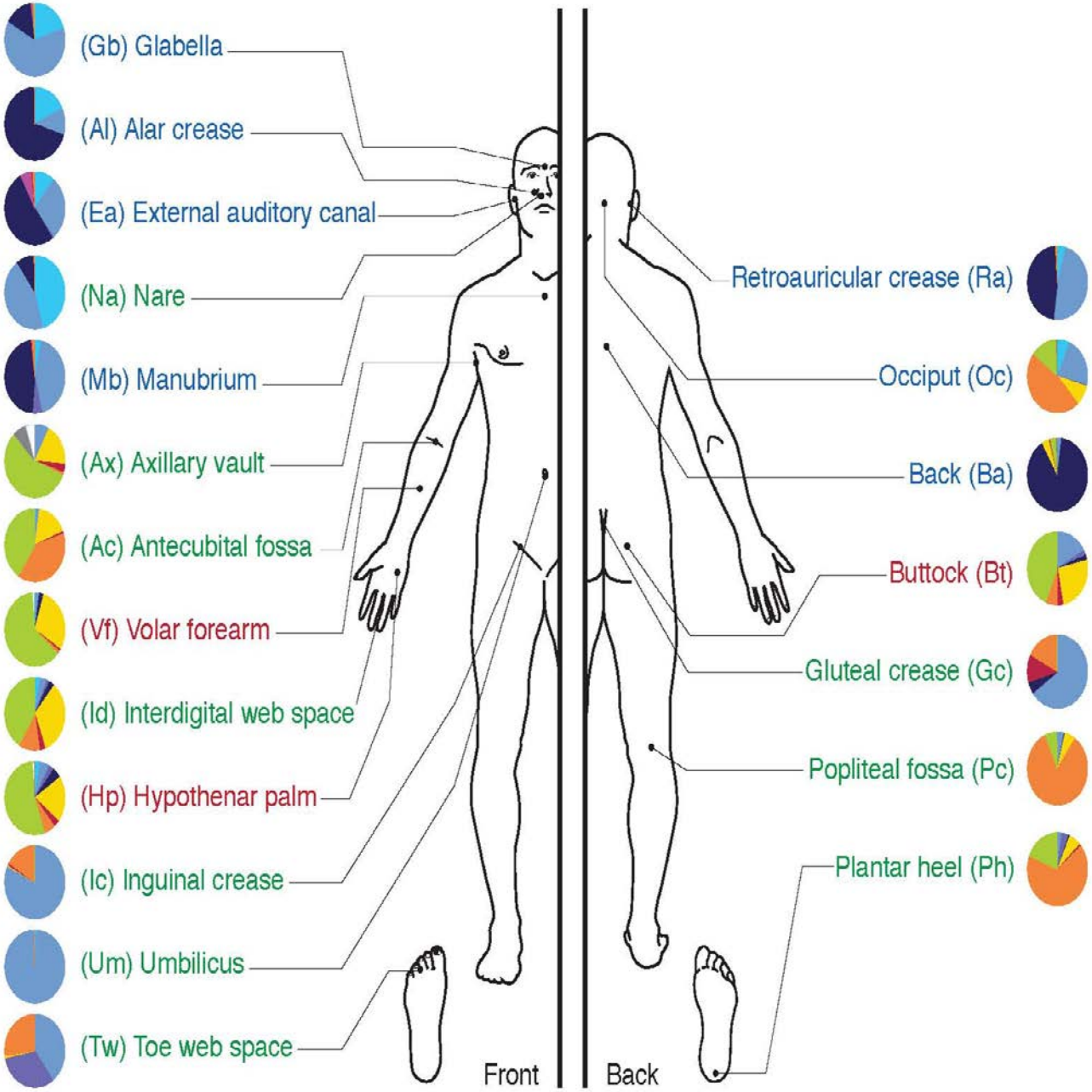

Central dogma
New technology
Genomic hypothesis
Gene
Genome sequencing
Genotype
cttattatacccacaga
cgaagaagaattgatga
cgtactataatccaaga
gccatggaagaagaaac
ttaagaaccgttcatct
mRNA transcript
AAAA
Transcript profiling
Transcriptome
Protein
Protein quantification and function
Proteome
Phenotype
Nature Reviews | Genetics

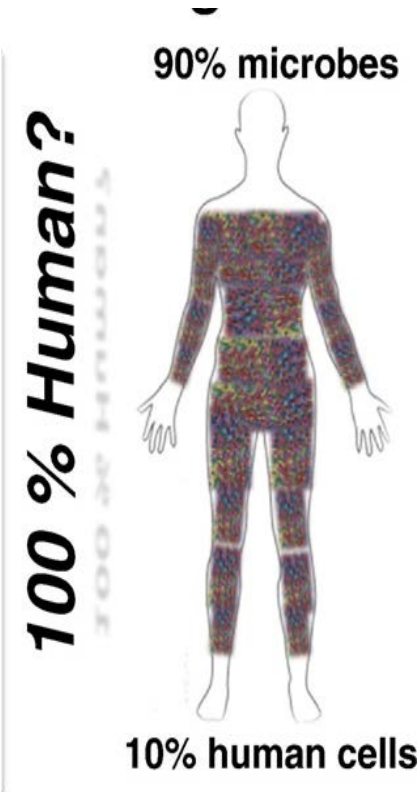


Hypomethylation
Hypermethylation

http://www.cellscience.com/reviews7/Taylor1.jpg


5000000000000000000000000000





We Are Really More Bug than

Microbiome based on16S Region Sequencing

- Being highly conserved among prokaryotes, it allows attachment of primers
- However, one organism, such as *Escherichia coli*, could have more than one 16S region, some of which are so similar to *Shigella spp.*'s that it becomes impossible to distinguish between the two, hence, could cause issues in annotation of the particular sequence.

Long and Short Reads

- A short sequence from 16s region may affect the annotation.
  - Below sequence with the length of 199 bp belongs to one of the 16s rRNA gene regions with 1550 bp of *Bacillus velezensis strain AD-3* (NCBI-complete genome)
    “CTTTATTGGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCG
    GCGTGCCTAATACATGCAAGTCGAGCGGACAGATGGGAGCTTG
    CTCCCTGATGTTAGCGGCGGACGGGTGAGTAACACGTGGGTAA
    CCTGCCTGTAAGACTGGGATAACTCCGGGAAACCGGGGCTAATA
    CCGGATGCTTGTTTGAACCGCATGGT”
  - However, in reference based mapping, it could be mapped to other Bacillus species, as well.

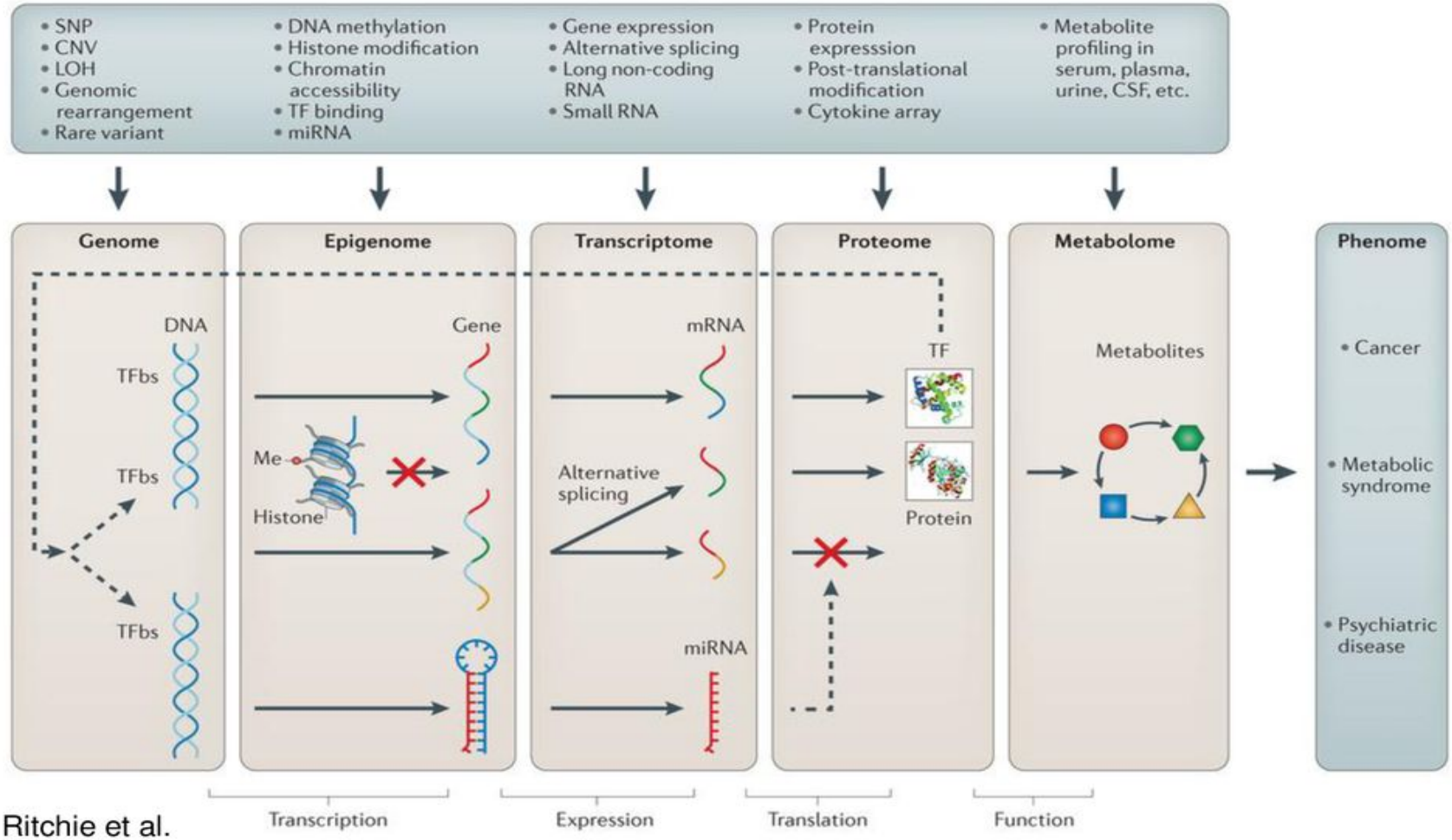
Annotation Database Inaccuracy

- The NCBI database, which is constantly being updated, like other databases such as SILVA, is set up relying on various phenotypic information like staining, appearance...
- However, the scientific community is trying to change this classification to genotype-based allocation of prokaryotes.

Change in the Database

- Currently, some genera, even whole families are being assiged to other upper taxonomic ranks. This results with assigning less relative abundance values to the upper taxonomic ranks before the change.
  - Clostridiales (order) is completely removed from the NCBI Entrez taxonomic database, having become a part of Eubacteriales (order)
  - Some Ruminococcus species (gnavus, lactaris and torques) moved under Lachnocpriceae (family) and Mediterraneibacter (genus)

Multi-omics: genotype -> phenotype mapping

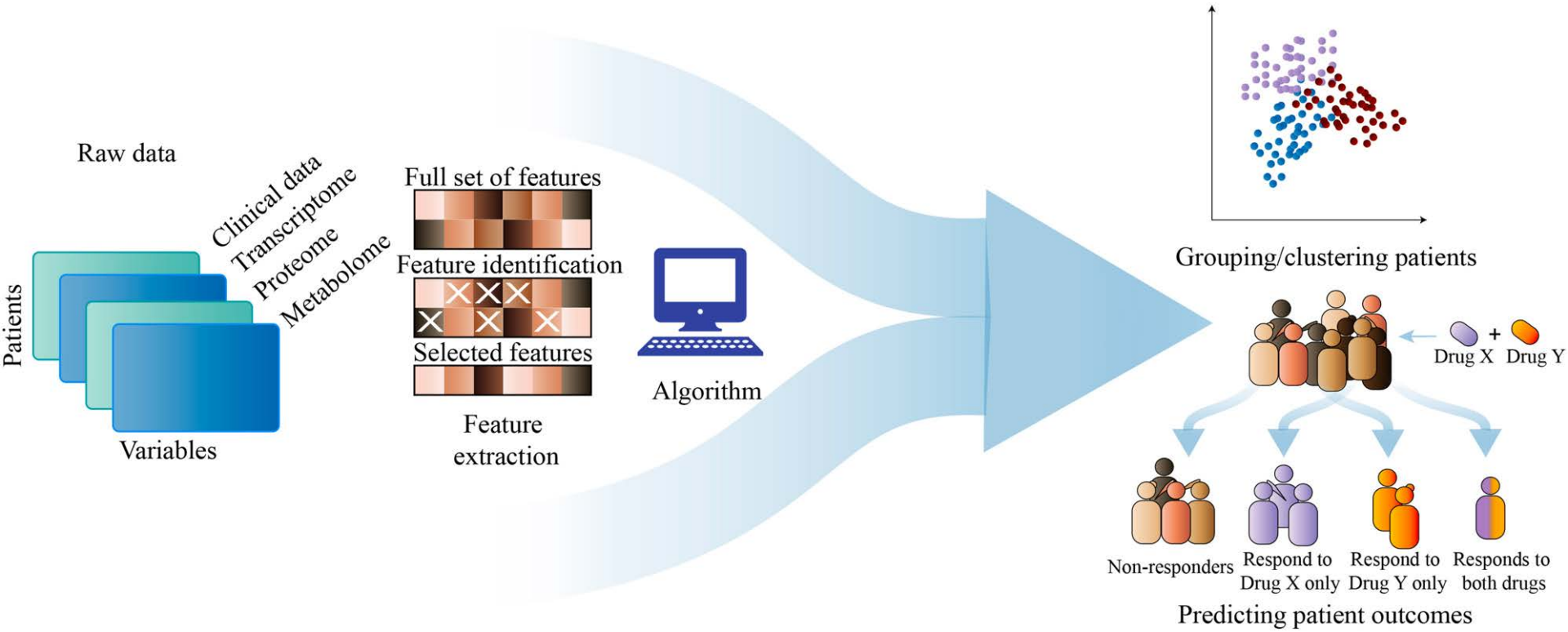


Ritchie et al.

Nature Rev Genet **16**, 85 (2015)

Nature Reviews | Genetics

AI/ML in Translational Medicine



Toh TS, Dondelinger F, Wang D. Looking beyond the hype: Applied AI and machine learning in translational medicine. EBioMedicine. 2019;47:607-615.

Example Applications

Unsupervised hierarchical clustering (part of ACME analysis)

- Identified associations between *BRAF* mutant cell lines of the skin lineage being sensitive to the MEK inhibitör

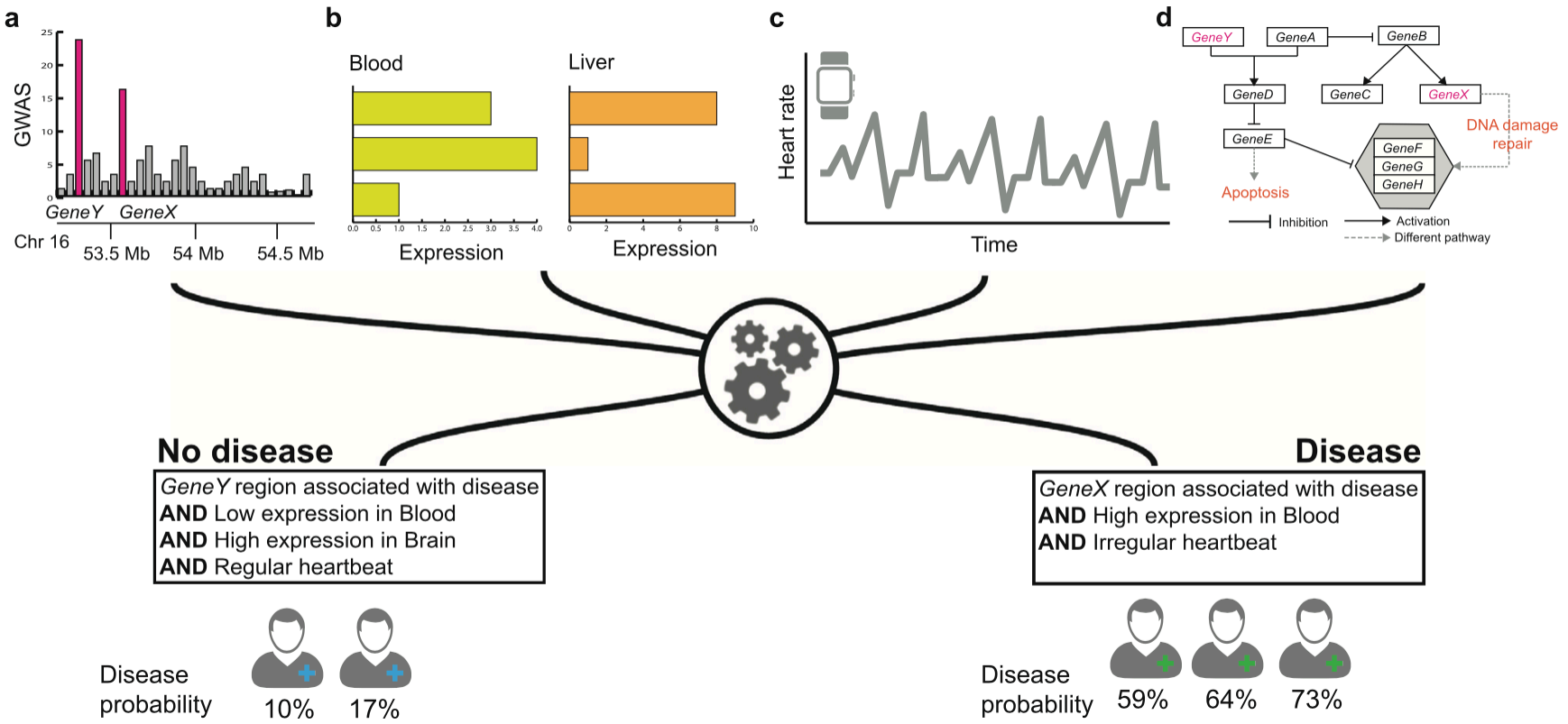
Seashore-ludlow B, Rees MG, Cheah JH, et al. Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. Cancer Discov. 2015;5(11):1210-23.

- Spectral clustering by SNF
  - Identification of new medulloblastoma subtypes

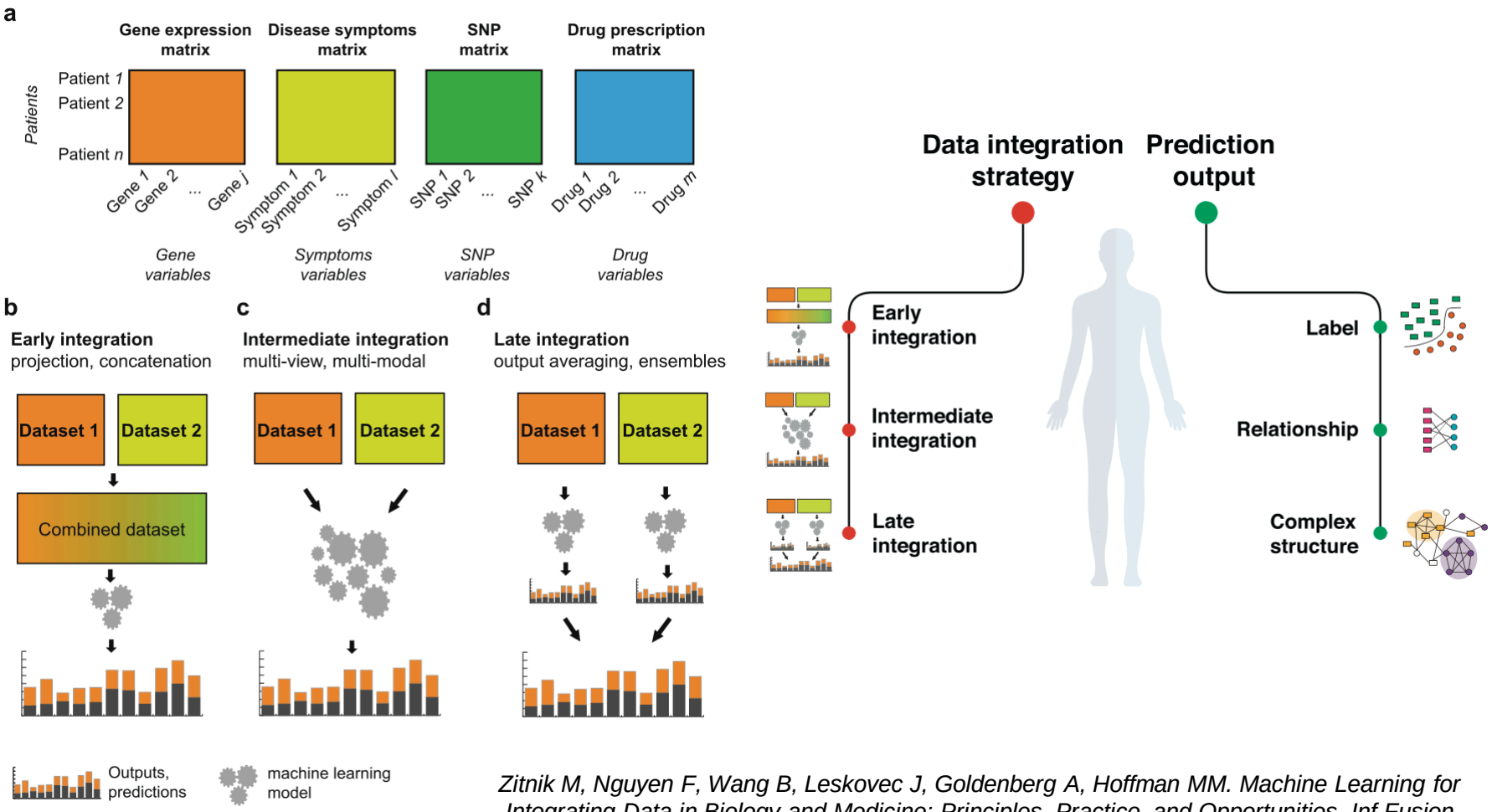
Cavalli FMG, Remke M, Rampasek L, et al. Intertumoral Heterogeneity within Medulloblastoma Subgroups. Cancer Cell. 2017;31(6):737-754.e6.

- Elastic net regression
  - Identification of *BRAF* and *NRAS* mutations in cell lines, were among the top predictors of drug sensitivity for a MEK inhibitor

Barretina J, Caponigro G, Stransky N, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012;483(7391):603-7.

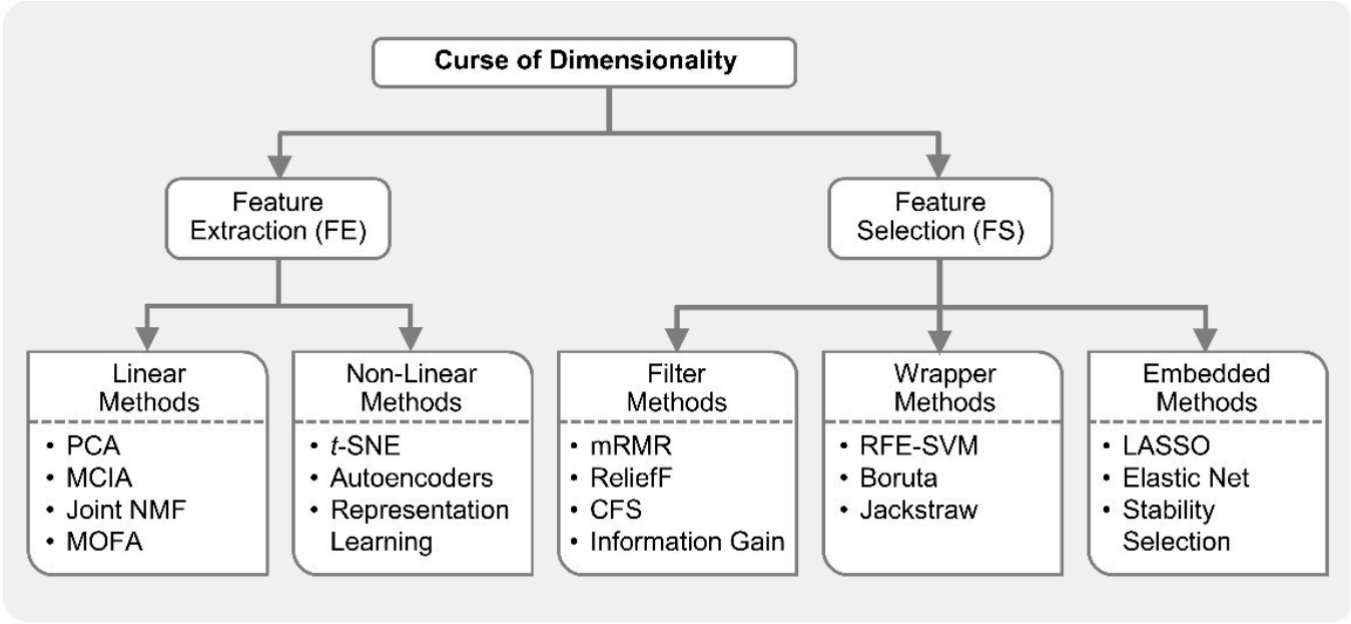


Zitnik M, Nguyen F, Wang B, Leskovec J, Goldenberg A, Hoffman MM. Machine Learning for Integrating Data in Biology and Medicine: Principles, Practice, and Opportunities. Inf Fusion. 2019;50:71-91.



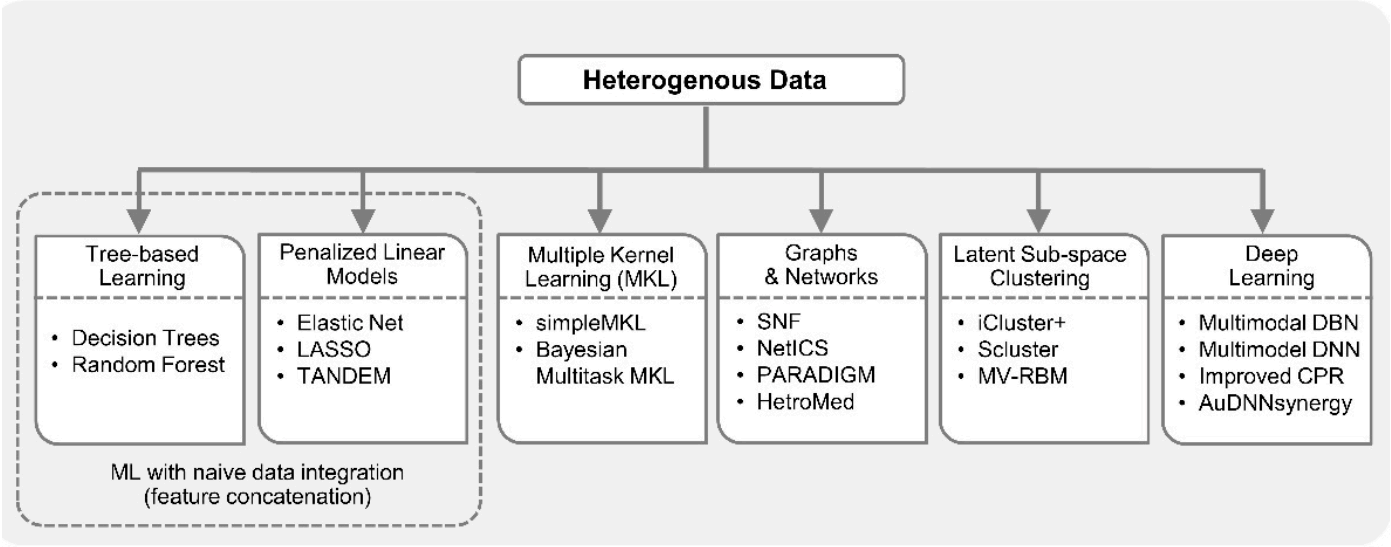
Zitnik M, Nguyen F, Wang B, Leskovec J, Goldenberg A, Hoffman MM. Machine Learning for Integrating Data in Biology and Medicine: Principles, Practice, and Opportunities. Inf Fusion. 2019;50:71-91.

Challenges



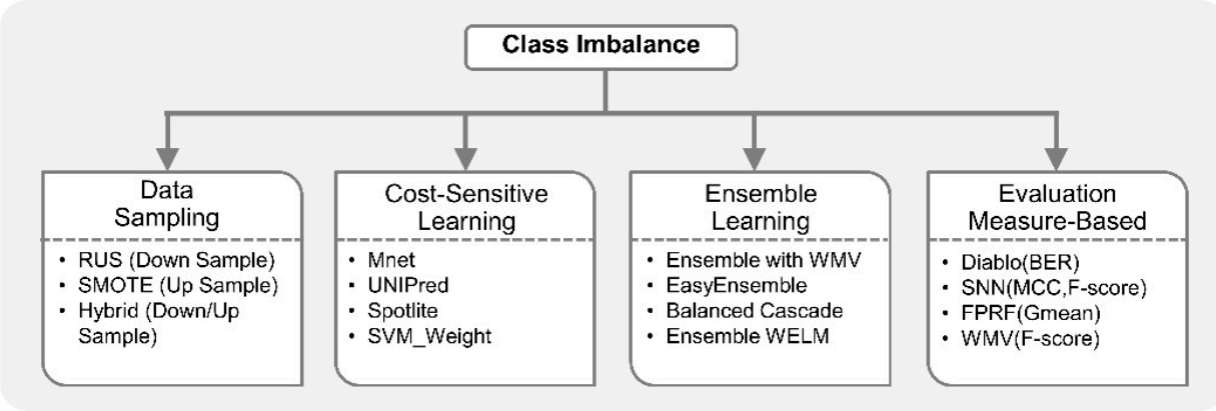
Mirza B, Wang W, Wang J, Choi H, Chung NC, Ping P. Machine Learning and Integrative Analysis of Biomedical Big Data. Genes (Basel). 2019;10(2)

Challenges



Mirza B, Wang W, Wang J, Choi H, Chung NC, Ping P. Machine Learning and Integrative Analysis of Biomedical Big Data. Genes (Basel). 2019;10(2)

Challenges

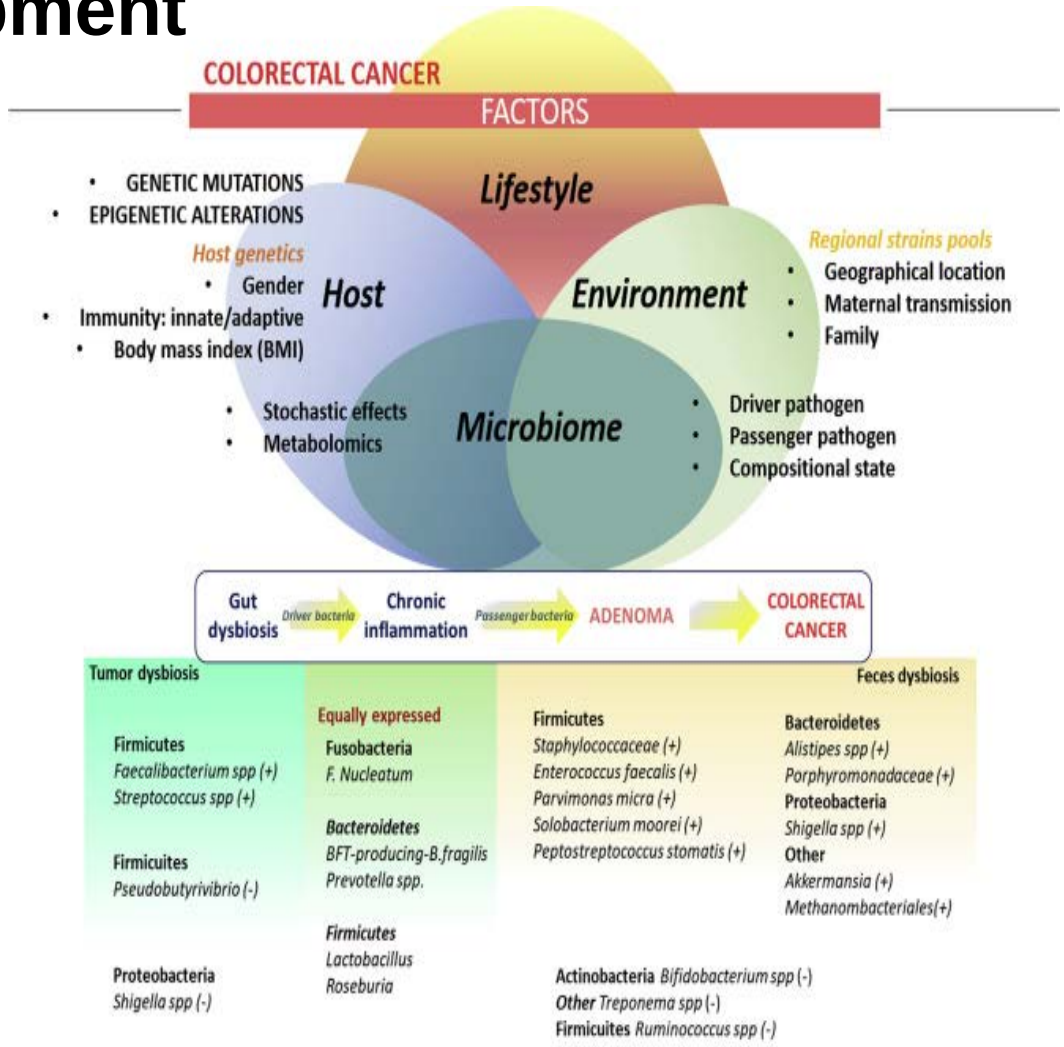


Mirza B, Wang W, Wang J, Choi H, Chung NC, Ping P. Machine Learning and Integrative Analysis of Biomedical Big Data. Genes (Basel). 2019;10(2)

Microbiome plays an important role in colorectal cancer (CRC) development

- Numerous studies have verified that gut microbiota can alter CRC susceptibility and progression since the gut microbiota can have an impact on colorectal carcinogenesis by inducing tumor proliferation.

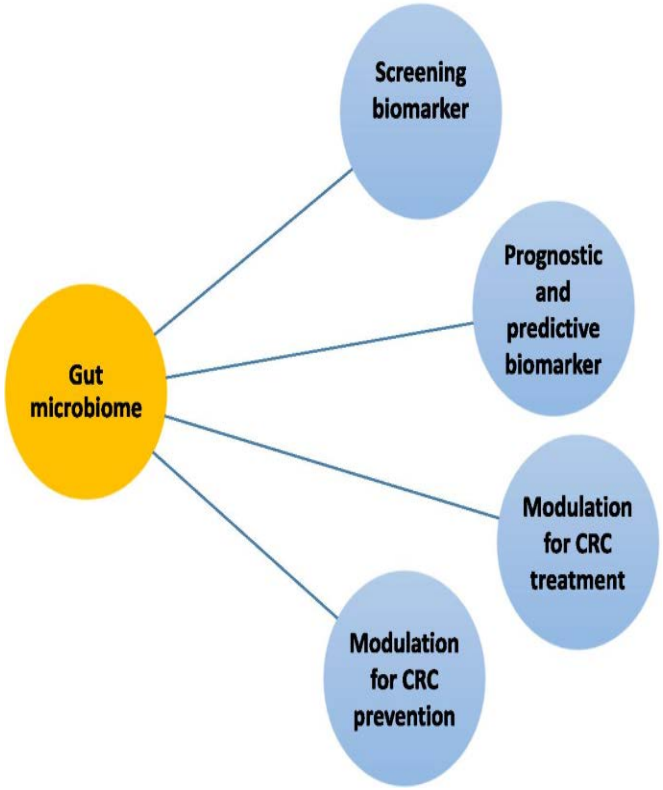
- Recent approaches for the treatment of colorectal cancer, various strategies are applied that consider the microbiome diversity in the patient—such as dietary interventions, antibiotic treatments, probiotics, prebiotics, and postbiotics



E. Saus et al. Microbiome and colorectal cancer: Roles in carcinogenesis and clinical potential, Molecular Aspects of Medicine, Volume 69, 2019

Using traditional biostatistical methods for targeting the microbiota for providing new opportunities involving tailored therapies for individual patients

Recently, it has been published that specific bacteria have been causing chemoresistance. The most commonchemotherapeutic drug given to patients with colorectal cancer is 5-fluorouracil, which dissolves with the presence of bacteria such as Fusobacterium nucleatum, Escherichia coli, or Bacteroides fragilis in the gut microbiome and thus it is not efficient. Lately, the treatment of colorectal cancer patients has been prolonged due to the usage of antibiotics such as ampicillin, colistin, and streptomycin to suppress pathogenic bacteria and promote immunotherapy outcomes.



Our study – Data preprocessing

Dataset article: “Gut microbiota in patients after surgical treatment”
Jin, Y et al. Gut Microbiota in Patients after Surgical Treatment for Colorectal Cancer. Environ. Microbiol. 2019, 21, 772–783.

We preprocessed the datased by removing the adapter and barcode sequences and the amplicon sequence primer sets (V3–V4). For this purpose, we used the BBMap (v.38.90) tool [23]. We applied this approach due to the errors that can occur when the primer sequences are accepted as amplicon ends. The aforementioned approach can produce incorrect consensus sequences and influence the taxonomic assignment.

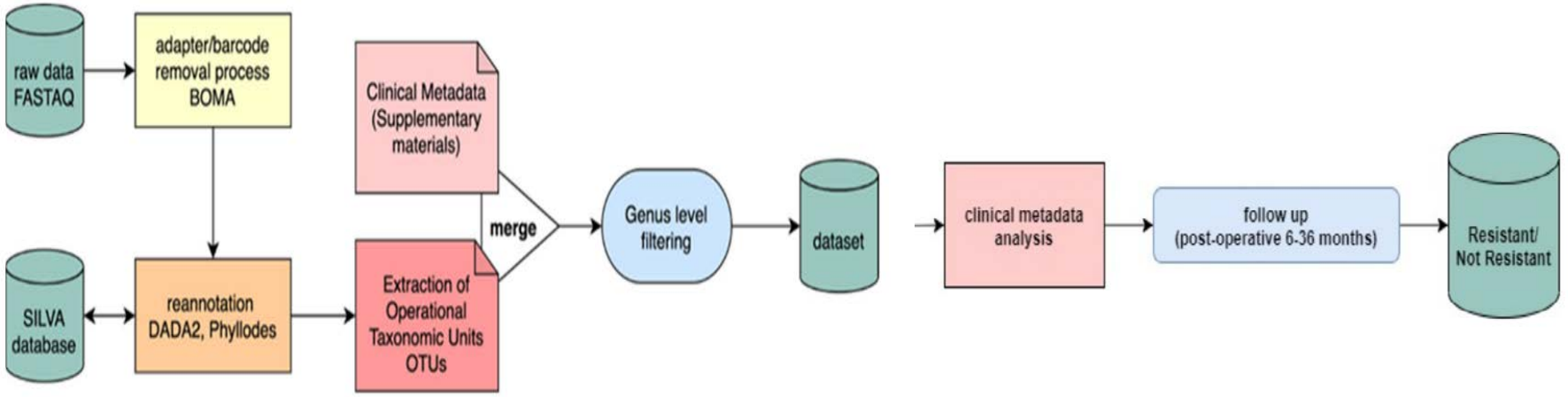
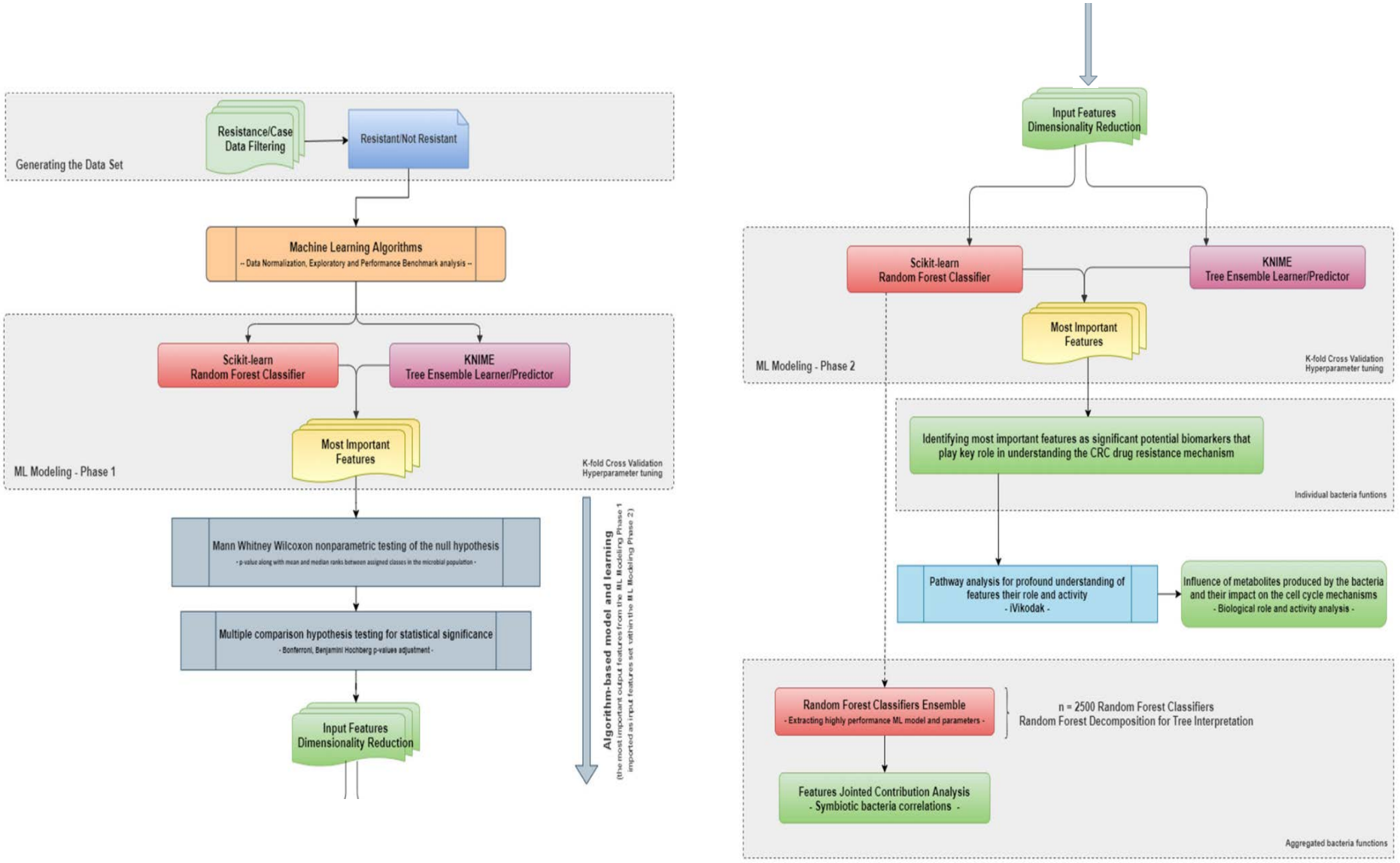


Figure 1. Data preprocessing and transformation.

Our study – Methodology



Our study – ML Modeling
Screening Phase

- We applied naïve Bayes, logistic regression, K-nearest neighbor, support vector machine with principal component analysis (PCA), and decision tree algorithms.

- Referring to the performance metrics of the decision tree approach, we proceed to explore the ensemble-based algorithms (Scikit-learn random forest classifier in Python and tree ensemble learner in KNIME), building multiple decision trees and taking advantage of the tree-related majority voting.

Table 1. Screening modelling phase algorithms overall accuracies

ML Algorithms	Overall Accuracy
Naive Bayes	0.429
Logistic Regression	0.425
K-Nearest Neighbors	0.325
Support Vector Machine	0.497
Decision Tree	0.764

* The algorithm overall accuracy was selected as the main algorithm selection indicator

Our study – ML Modeling Results

- We concluded that the tree-based algorithms accomplished the highest scores compared with the other techniques we applied according to the performance metrics.

- We also tried XGBoost and AdaBoost algorithms, which resulted in no significant improvements compared with the forest-based approach described above.

- We identified the second-phase Python-based random forest classifier as the most performant and selected the resulting most important features as a reference set for further statistical analysis.

Table 2. General ML modeling performance metrics for the resistant and non-resistant CRC post-operative individuals' group.

Environment	ML Algorithms	Normalization/Scaling	Accuracy	Sensitivity	Specificity
Python Scikit-learn	RFC (P1)	Standard Scaler	0.9	1.000	0.833
Python Scikit-learn	RFC (P1)	Z-Score Normalizer	0.9	1.0	0.75
KNIME	TEL (P1)	Z-Score Normalizer	0.833	0.778	1.0
Python Scikit-learn	RFC (P2)	Standard Scaler	0.917	1.000	0.833
KNIME	TEL (P2)	Z-Score Normalizer	0.9	1.000	0.8

RFC—Scikit-learn random forest classifier, TEL—Tree ensemble learner, P1—Phase 1 ML modeling, P2—Phase 2 ML modeling.

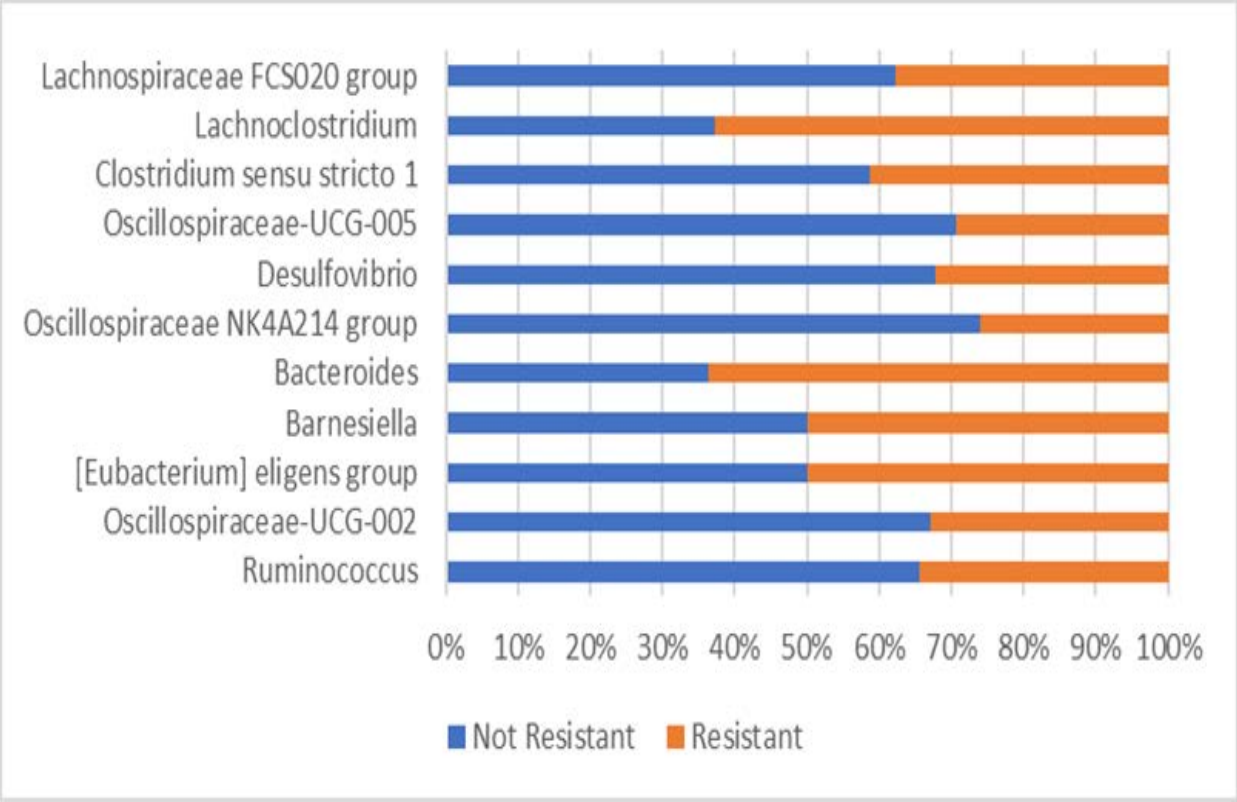
Table 3. Detailed ML modeling performance metrics for the resistant and non-resistant CRC post-operative individuals' group.

	Precision		Recall		F1-Score	
Environments and ML Algorithms	Resistant	Non-Resistant	Resistant	Non-Resistant	Resistant	Non-Resistant
Python Scikit-learn—RFC (P1)	0.83	1.00	1.00	0.80	0.91	0.89
Python Scikit-learn—RFC (P1)	0.75	1.00	1.00	0.86	0.86	0.92
KNIME—TEL (P1)	1	0.778	0.600	1.000	0.750	0.875
Python Scikit-learn—RFC (P2)	0.83	1.00	1.00	0.86	0.91	0.92
KNIME—TEL (P2)	0.800	1.000	1.000	0.833	0.889	0.909

RFC—Scikit-learn random forest classifier, TEL—Tree ensemble learner, P1—Phase 1 ML Modeling, P2—Phase 2 ML modeling.

Our study – Statistical Analysis Results

Our taxonomic analysis of the raw data, assuming the improved taxonomical precision since the bacterial references are constantly changing, resulted in 3603 different bacterial taxonomic units detected. Thus, the gut microbiome consisted of 20 unique phyla, 35 classes, 72 orders, 119 families, and 259 unique genera



Our study – Highly Contributing and Joint Features Contribution Analysis Results

This study points out the **different perspectives of treatment** since our aggregate analysis gives precise results for the **genera that are often found together** in a resistant group of patients, meaning that **resistance is not due to the presence of one pathogenic genus in the patient microbiome, but rather several bacterial genera that live in symbiosis**.

Table 4. Aggregated bacteria significance contributions to the resistant class.

Aggregated Bacteria	'Resistance' Contribution
['Escherichia-Shigella', 'Subdoligranulum', 'Gemella', 'Negativibacillus']	0.00770053
['Blautia', 'TM7x'] ['	0.0061875
['Escherichia-Shigella', 'Coprococcus', 'Lachnospiraceae UCG-010', 'Family XIII UCG-001']	0.00555556
['Terrisporobacter', 'Weissella', 'Slackia']	0.00538462
['Enterococcus', 'Haemophilus', 'UCG-005']	0.005
['Intestinibacter', 'Enterococcus', 'Lachnospiraceae NC2004 group', 'Lachnoclostridium']	0.0047138
['Coprococcus', 'Megasphaera', 'Parasutterella', 'UCG-002']	0.0045
['Streptococcus', 'Phascolarctobacterium', 'Paraprevotella', 'Dubosiella']	0.00403846
['Subdoligranulum', 'Blautia', 'Paraprevotella', 'Oxalobacter']	0.00317853
['Subdoligranulum', 'Butyrivibrio']	0.00307692
['Lachnospiraceae UCG-010', 'Barnesiella']	0.00235897
['Blautia', 'Oxalobacter'] ['	0.00231884
['Clostridium sensu stricto 1', 'Flavonifractor', 'Agathobacter', 'Butyricimonas']	0.00227193
['Flavonifractor', 'Agathobacter', 'Butyricimonas', 'Anaerofustis']	0.00222222
['[Eubacterium] ruminantium group', '[Eubacterium] eligens group', 'Moryella']	0.00198413
['Haemophilus', 'Alistipes']	0.00188889
['Clostridium sensu stricto 1', 'Blautia', 'TM7x', 'Butyricimonas']	0.00188235
['Ruminococcus', 'Enterococcus', 'Turicibacter', 'Leuconostoc']	0.00181818
['[Eubacterium] ruminantium group', 'Denitrobacterium']	0.00179724
['Turicibacter', 'Leuconostoc']	0.00171429
['Slackia', 'Eubacterium']	0.00162037
['Escherichia-Shigella', 'Subdoligranulum']	0.0013468
['Enterococcus', 'Weissella', 'Lachnoclostridium']	0.00133333
['Enterococcus', 'Coprococcus', 'Anaerococcus', 'Senegalimassilia']	0.00128205
['Ruminococcus', 'Weissella', '[Eubacterium] ruminantium group', 'Denitrobacterium']	0.00121212

On Going Microbiome Project

- Understanding the role of microbiome in Drug resistance mechanisms-Joint Project with Istituto Giannina Gaslini Genoa ITALY
- Multiomics Approaches for personalized therapy in CRC patients –S. Korea-Turkey Joint Project
- The role of Microbiome in SMA Patients-Joint Project with U. Milan
- Skin Microbiome –Joint project with UMASS

Istituto Giannina Gaslini

- THANKS to
- Miodrag Cekic
- Milena J. Ozdemir
- Orhan Özcan

Special THANX to