Temporal evolution of gut microbiota during chemotherapy in acute leukemia pediatric patients

Alexander V. Tyakht[1], Anna S. Popenko[1], Alexander S. Zelenin[1], Boris A. Kovarsky[1], Dmitry G. Alexeev[1,2], Irina Y. Karpova[1], Andrei K. Larin[1], Galina G. Solopova[3], Dzhamilya T. Dzhandarova[3], Sergei A. Roumiantsev[3], Elena S. Kostryukova[1], Vadim M. Govorun[1,2]

[1] *Research Institute of Physico-Chemical Medicine (RIPCM), Malaya Pirogovskaya 1a, 119435 Moscow, Russia*
[2] *Moscow Institute of Physics and Technology (MIPT), Institutskii per. 9, 141700 Dolgoprudny, Moscow Region, Russia*
[3] *Federal Scientific Clinical Center of Pediatric Hematology, Oncology and Immunology named by Dmitry Rogachev, Samora Mashela 1, 117997 Moscow, Russia*

Introduction

Gut microbiota plays an important role in human homeostasis. During anticancer chemotherapy course, its balance is disrupted, resulting in less efficient treatment and worse prognosis: immunity is decreased, opportunistic pathogens are activated. Additional negative impact on microbiota is caused by antibiotics. Monitoring of the microbiota composition using metagenomic sequencing allows to assess the degree of dysbiosis, including quantitative profiling of a wide spectrum of pathogens. Analysis of microbiota temporal variation during the course and recovery period will clarify the mechanism of microbiota restoration and discover the ways to modulate microbiota to improve the recovery rate.
Here we present the results of a pilot project of metagenomic analysis of gut microbiota in several pediatric patients undergoing chemotherapy treatment.

Project overview

- Patient selection, sample and meta-data collection (Rogachev Center)
- Sample preparation, DNA sequencing (Genomic Center, RIPCM)
- Bioinformatic analysis: reads processing, statistical analysis and visualization (Laboratory of Bioinformatics, RIPCM)

Materials and methods

Objects of the study

- Pediatric patients with cancer or autoimmunity diseases undergoing chemotherapy in Rogachev Center (4 subjects, age 2-5 y.o.)
- Stool samples (n=15) collected at 3-4 time points for each patient during the treatment course (throughout several weeks)
- Clinical records with a detailed description of the treatment

DNA sequencing and bioinformatic analysis

- Data: whole-genome reads from high-throughput SOLiD 4 sequencer (Life Technologies): read length 50 bp, 1.5 ± 0.4 Gbp per sample.
- Database input and parallel data processing using *jsub* software pipeline on the RIPCM compute cluster.
- Reads preprocessing and mapping (Bowtie) to:
  - human genome;
  - reference set of 444 human gut microbial genomes.
- Quantitative assessment of taxonomic composition at the genus level.
- Complementary methods for detection of genomes not present in the reference set:
  - search for clade-specific gene sequence markers - MetaPhlAn;
  - assembly *de novo* (Velvet), similarity search (BLAST) against NCBI nr database.

Results

Varying proportions of the human and microbial DNA

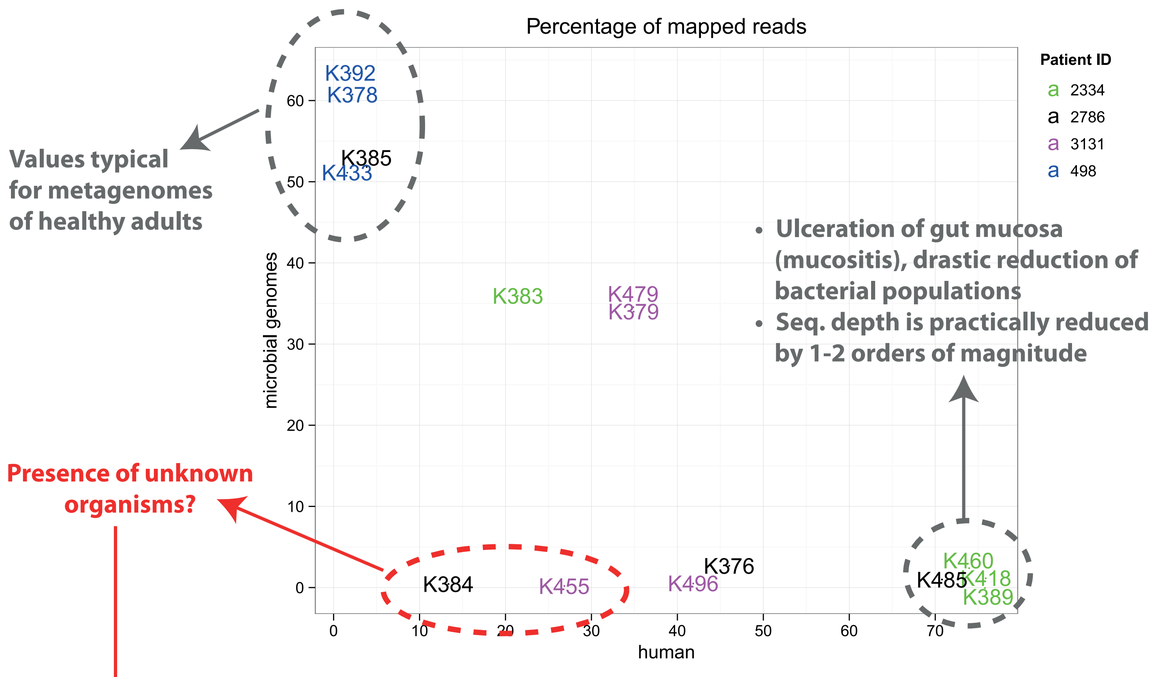


Fig. 1. Fraction of the reads mapped to human vs. microbial genomes.
Each text tag represents a single sample, while its colour denotes the patient.

Additional identifications using the complementary methods

- **MetaPhlAn: new bacterial components which genomes absent in the reference set**
  - *Pseudomonas* spp.
  - *Streptococcus vestibularis*
- **Assembly de novo + BLAST: eukaryotic components**
  - Yeast: contigs of sample K_384 were found to have high similarity to the genome of *Candidas albicans* a cause of morbidity in immunocompromised patients
  - Undigested food traces: sequences of *Triticum aestivum* (wheat), *Avena sativa* (oat), *Pisum sativum* (pea).

Update of the reference genome set

Bacterial composition: prevalence of potential pathogens

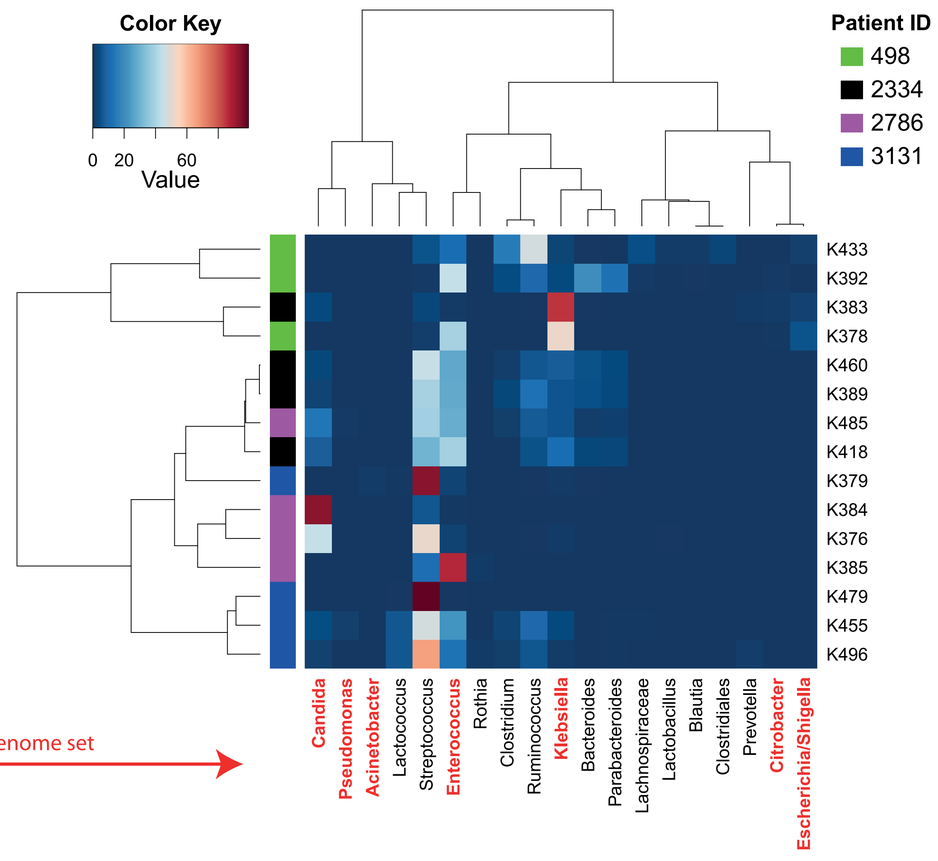


Fig. 2. Hierarchical heatmap of relative microbial abundance (in percents of the total amount of mapped reads for each sample). Potentially pathogenic bacteria and yeast are highlighted. Most identifications are in concordance with the results of the culture studies.
Among other *Streptococci*, there is a high proportion of a species close to *Streptococcus thermophilus*, a bacterium from fermented milk products. As such products and probiotics are prohibited during chemotherapy, this fact is interesting and is being examined in details.

Detection of identical strains in the samples using SNP signatures

- The reads mapped to a reference genome can be used for consensus SNP calling
- For a given sample, each sufficiently covered genome possesses a specific SNP set (signature)
- Same bacterial strains (genomes) in different samples => similar SNP signatures

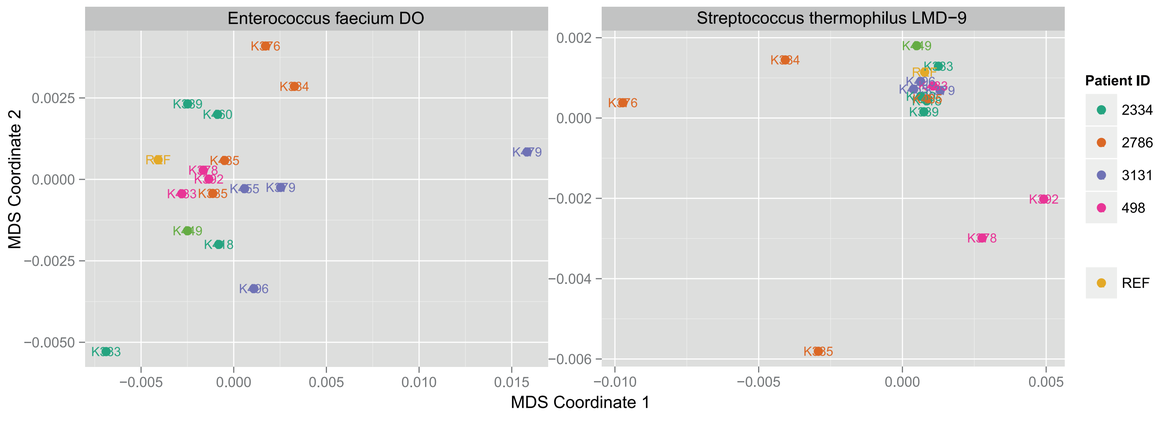
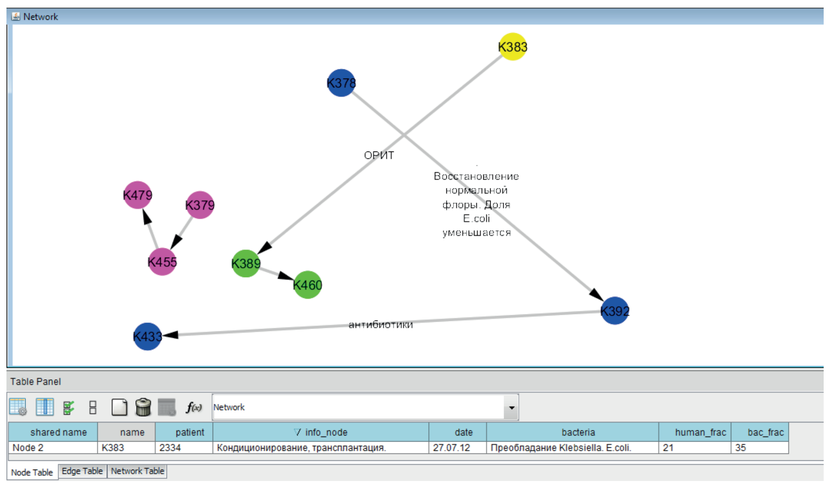


Fig. 3. Non-metric MDS plot of the SNPs signatures of the samples. Close points with same colour reflect the temporal stability of the strains inhabiting the gut of a single patient. Intra-patient similarity of the samples possibly indicate that the strain was acquired from the same source (i.e. hospital pathogens, probiotics).

Interactive visualization of biomedical and metagenomic data

? Taxonomic composition, temporal evolution, events in the clinical records – how to combine diverse data visually?
! Graph network (i.e. Cytoscape)



Nodes = samples

Node proximity = similarity of bacterial composition (i.e. based on MDS by UniFrac metric)

Directed edges = steps in time for each sample

Prospects

- Large-scale microbiota metagenomic study in Rogachev Center including the patients with acute leukemia and control group
- Functional analysis of metabolic potential
- 16S rRNA sequencing as a cheaper alternative that is free of depth reduction due to human DNA

Funding

This work was supported by State Contracts 16.512.11.2111, 16.552.11.7034.

Contacts

at@niifhm.ru
http://www.metagenome.ru

metagenome.ru