The BioArte

XLIX Congresso Nazionale AMCLI
27. 02. 2022.

New developments of high-throughput sequencing methods in the clinical assessment of the human microbiota

Márton Szoboszlay, Manuele Biazzo

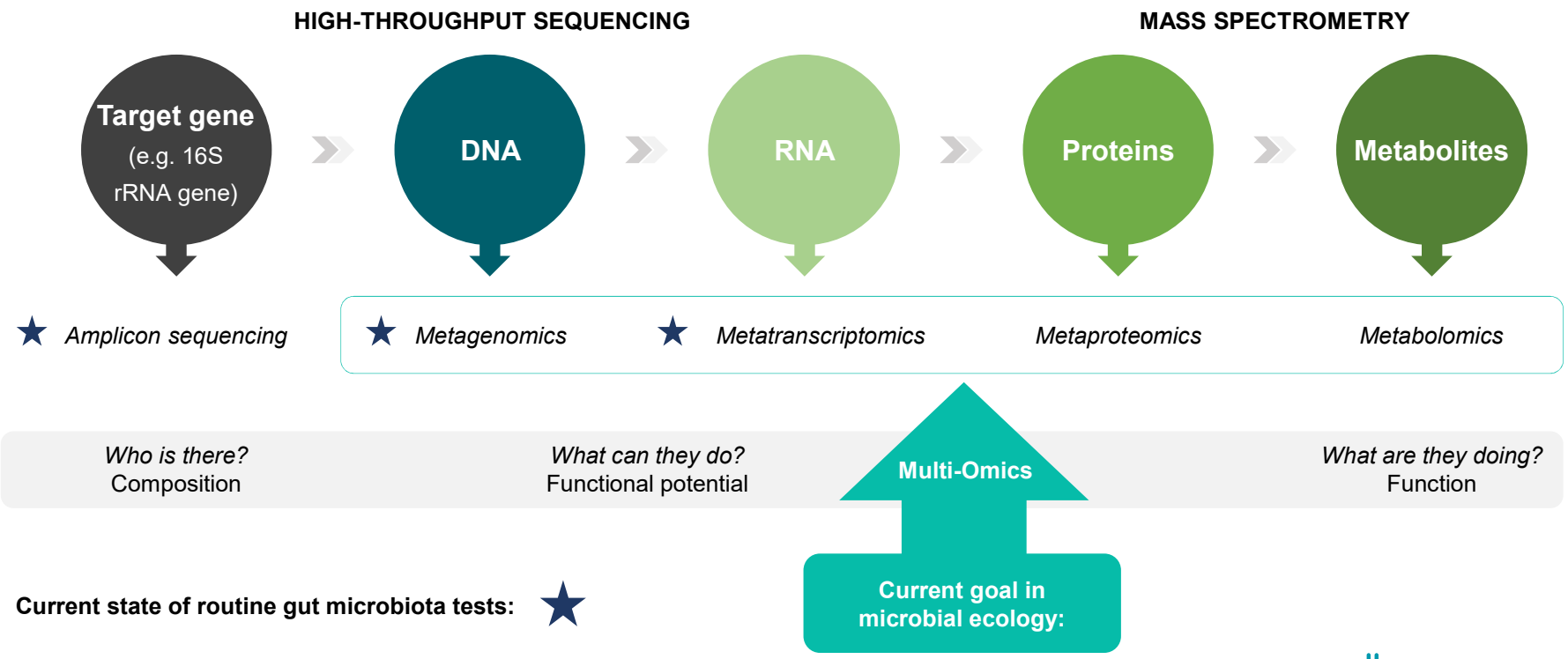
Assessing the microbial community composition in human samples (especially stool) is gaining popularity

- Part of a health check-up, or to get personalized advice on nutrition
- Several private companies including The BioArte
- Sequencing methods and data analysis tools in microbial ecology are rapidly developing
- How we handle and interpret the data from high-throughput sequencing is changing
- This will (and should) affect the clinical application

What is the current state of human microbiota testing?

How will it change in the near future?

Molecular methods in microbial ecology



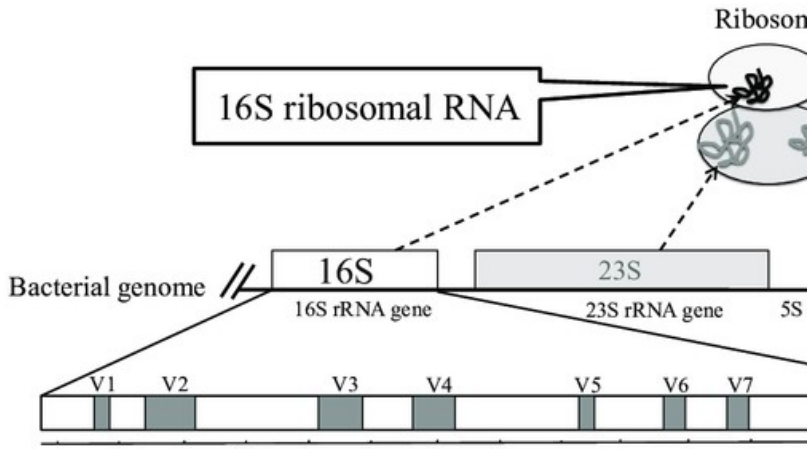
Current state of routine gut microbiota tests: ★

Current “microbiome test” services

Clinical strong evidence on microbiome / microbiota → only *Bacteria* (and *Archaea*)

Increasing work on biomarkers → Today’s test mainly descriptive, comparison to a reference, no hypothesis

01. PCR of the target gene: a region of the 16S rRNA gene
02. Sequencing: most often with Illumina
03. Data processing with a commonly used pipeline e.g. QIIME2
04. Reporting the relative abundances of taxa comparison to a reference dataset or previous results
05. **Interpretation**
    a. by a human microbiota expert in context of the background and other clinical results of the patient
    b. by an AI
    c. left to the physician who ordered the test
    d. left to the client



Fukuda et al. 2016 (DOI: 10.7888/juoeh.38.223)

01 The target gene / region

Amplicon sequencing will likely remain relevant for routine application

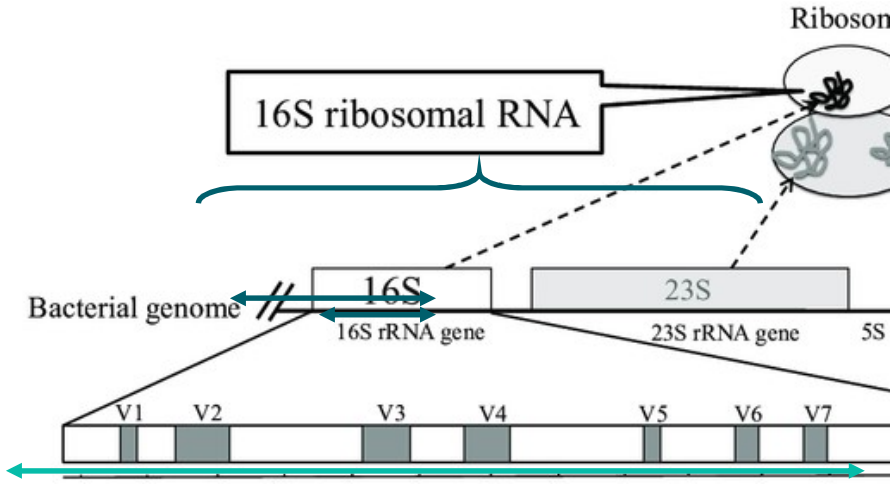
1. Metagenomics for low-diversity samples? (e.g. skin, vagina)
2. Metatranscriptomics ⟷ sample storage and transport

The 16S rRNA gene is the best single molecular marker to identify *Bacteria* and *Archaea*

The choice of target region has been mostly limited by the sequencing technology

Illumina

Nanopore, PacBio



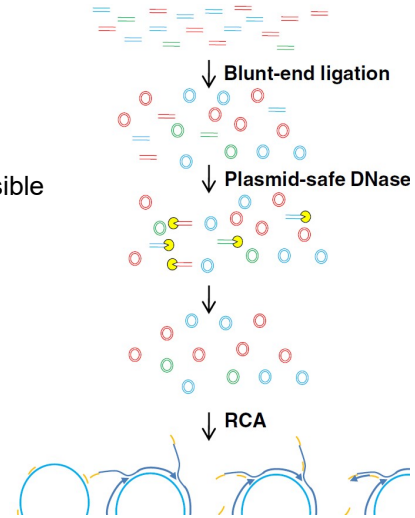
Fukuda et al. 2016 (DOI: 10.7888/juoeh.38.223)

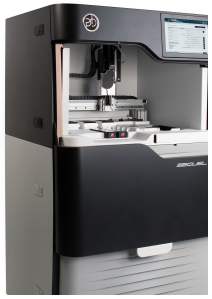
02 Sequencing methods

Illumina dominates the market
- no significant innovation since 2013, only larger scale
- limited information (short reads) but high quality

Long reads to get more information (full 16S rRNA gene, full rRNA operon)
- PacBio: relatively low throughput, very high upfront cost, not widely accessible
- Nanopore: ''cheap'', fast, accessible, BUT high error rate

Methods to improve the error rate increase the cost



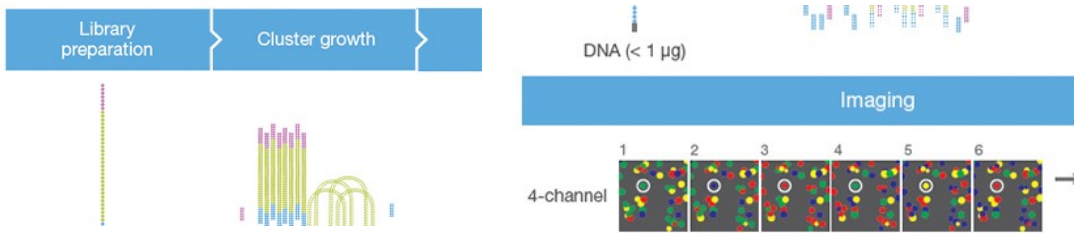


Does the high error rate outweigh the benefits of the long reads?

Li et al. 2016 (DOI: 10.1186/s13742-016-0140-7)

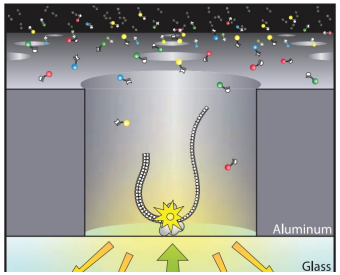
02 Sequencing methods

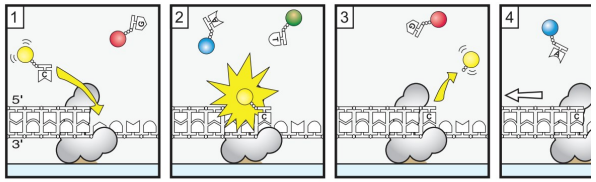
Illumina



www.illumina.com

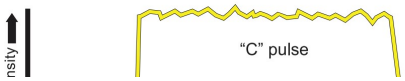
PacBio

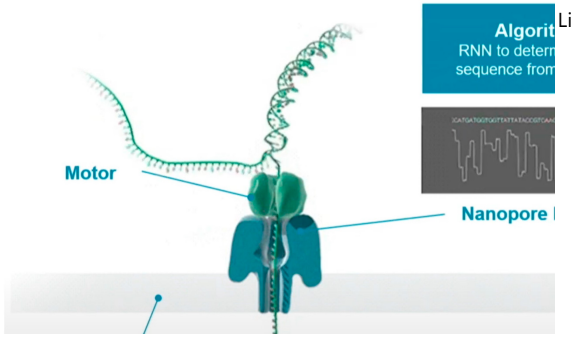




Eid et al. 2009 (DOI: 10.1126/science.1162986)

Nanopore





Lin et al. 2021 (DOI: 10.3390/bios11070214)

02 Sequencing methods

At genus level resolution, with its current quality, Nanopore is just as suitable for 16S rRNA gene amplicon sequencing as Illumina

Illumina – Nanopore comparison

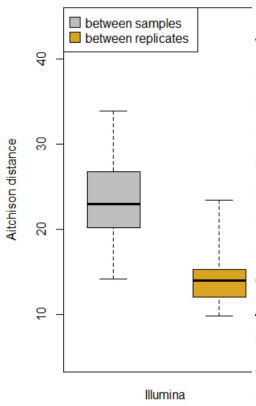
a. human faecal samples
b. mock community of 14 bacterial species

The Illumina and Nanopore results were not identical but had good agreement on the abundant genera

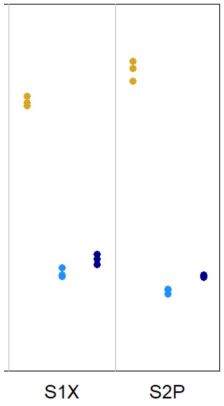
Nanopore had somewhat better precision in assessing the bacterial community structure

Illumina gave substantially higher estimation of richness

Nanopore outperformed Illumina in the analysis of the mock community



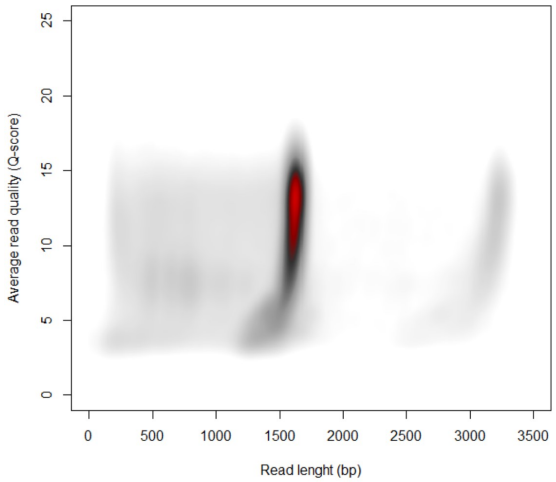
% relative abundance		Illumina			Nanopore		
Genus	Theoretical	Rep A	Rep B	Rep C	Rep A	Rep B	Rep C
Faecalibacterium	17.63	14.20	17.74	17.59	17.65	18.44	17.07
Veillonella	15.87	5.68	3.28	4.03	21.15	19.75	20.93
Escherichia	12.12	5.66	8.43	7.58	9.68	9.32	9.39
Bacteroides	9.94	9.51	7.85	13.95	10.96	13.01	12.85
Roseburia	9.89	4.38	5.58	4.91	7.29	7.36	7.26
Lactobacillus	9.63	2.28	3.10	2.25	5.01	5.36	4.81
Bifidobacterium	8.78	8.23	9.57	8.13	8.06	4.64	4.63
Fusobacterium	7.49	9.22	13.60	4.39	5.18	7.68	6.56
Prevotella	4.98	5.66	0.91	4.13	5.38	4.49	5.12
Clostridioides	2.62	1.69	2.60	2.10	4.59	5.53	4.89
Akkermansia	0.97	0.68	0.58	0.36	1.11	0.82	0.97
Methanobrevibacter	0.066	0.043	0.033	0.029	0.0013	0.004	ND
Salmonella	0.009	ND	ND	ND	0.0013	0.008	0.0065
Enterococcus	0.0009	0.081	0.0037	0.00086	ND	ND	ND
Clostridium	0.0002	0.232	0.012	0.019	ND	ND	ND
Total	**100.0**	**67.6**	**73.3**	**69.5**	**96.1**	**96.4**	**94.5**



02 Sequencing methods

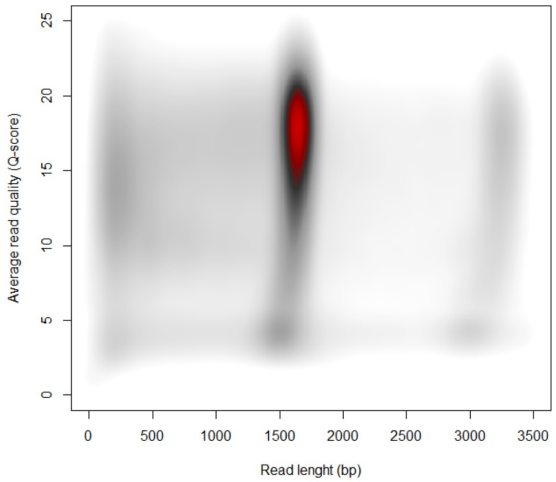
Nanopore improvements

Mid 2021: new basecalling algorithm → **Late 2021: new sequencing chemistry (Q20+)**



Potential direct RNAseq of rRNA without PCR

Karst et al. 2018 (DOI:10.1038/nbt.4045)



Half of the reads have 2% or fewer expected errors

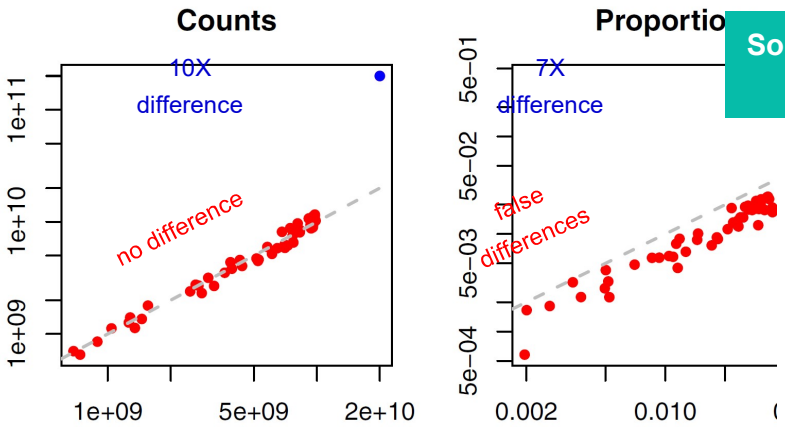
03 Reporting the results

Comparing **the relative abundance** of taxa to “normal” values and looking for **biomarkers** (taxa whose abundance correlates with a condition)

Relative abundance

- The problem starts when the results are compared to a reference dataset or previous results of the patient

324 G.B. Gloor et al. / Annals of Epidemiology 26 (2016) 322–329



Solution: logratios instead of relative abundances

Microbiome Datasets Are Compositional: And This Is Optional

03 Reporting the results

Comparing the **relative abundance** of taxa to "**normal**" values and looking for **biomarkers** (taxa whose abundance correlates with a condition)

Normal / healthy composition

- It's easier to define normal function than normal composition
- Problem: weak correlation between taxonomy and function

Gastroenterology
Volume 160, Issue 2, January 2021, Pages 483-494

Article

The Healthy Microbiome—What Is the I of a Healthy Gut Microbiome?

Fergus Shanahan [1, 2], Tarini S. Ghosh [2, 3], Paul W. O'Toole [2, 3]

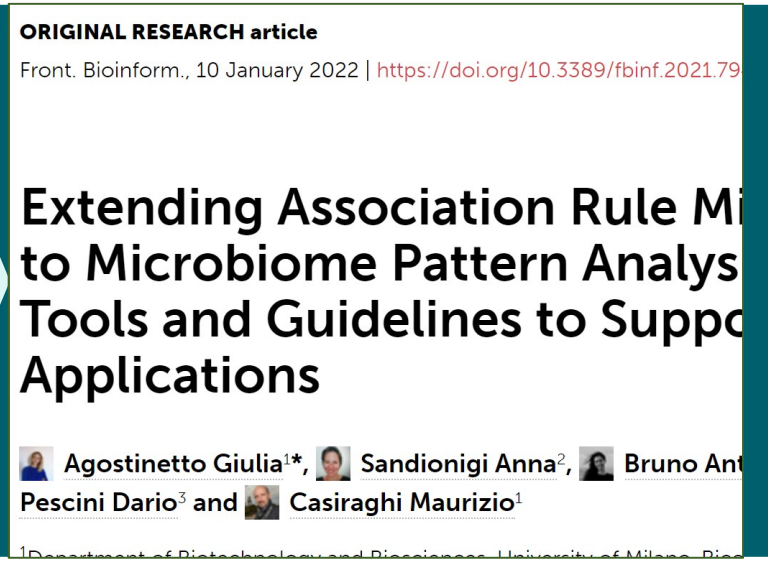
Show more

03 Reporting the results

Comparing the **relative abundance** of taxa to "**normal**" values and looking for **biomarkers** (taxa whose abundance correlates with a condition)

Species (taxa) as biomarkers

- Except for a few pathogens, there are no simply good or bad microbes
- The activity of a population depends on its interactions within the ecosystem and on the environmental properties of its habitat (i.e. the host)
  - The human microbiota is not a list of isolated populations but an ecosystem


ORIGINAL RESEARCH article

Front. Bioinform., 10 January 2022 | https://doi.org/10.3389/fbinf.2021.79

Extending Association Rule Mi
to Microbiome Pattern Analys
Tools and Guidelines to Suppo
Applications

Agostinetto Giulia[1]*, Sandionigi Anna[2], Bruno Ant
Pescini Dario[3] and Casiraghi Maurizio[1]

We need objective ways to identify patterns in the microbial community structure that can be associated with

- health conditions
- functions / ecosystem services

04 Interpreting the results

Interdisciplinary field between medicine and microbial ecology <– requires knowledge of both

- There are very few experts
- Utilisation of AI

Several articles written by gastroenterologists, GPs, nutritionists claim that currently available gut microbiota tests are difficult to interpret

Reasons:

- The issues and limitations discussed in part 4 (normal composition, biomarkers)
- Lack of good interpretation of the results

If this doesn't change, it doesn't matter how much the sequencing and data analysis tools improve

Communication between physicians, microbiologists, and bioinformaticians, workshops and courses to make sure that:

- The tests deliver information useful for the clinical practice
- Physicians have the necessary knowledge to interpret the results

The BioArte

XLIX Congresso Nazionale AMCLI
27. 02. 2022.

Thank you for your attention

© BioArte Limited

03 Data processing

Illumina: **Short reads** ❯ Not enough information for species- or even genus level classification

High quality ❯ Focus on resolving small sequence differences
- oligotyping
- ASVs (amplicon sequence variants)

Is this useful for a clinical tests?

- Current tools are designed for short reads, Illumina error profiles, and genus level taxonomic classification
- New tools are needed for long reads
  - focus on taxonomy and not on small sequence variations
  - especially problematic:
    - alignment
    - clustering
    - chimera detection
    - classifier that considers the basecalling quality

OTUs / ASVs with limited taxonomical information ❯ **Species**