Identification of the gastric microbiome from endoscopic biopsy samples using whole genome sequencing

Doron Betel
May 7, 2015
Hematology/Oncology
Institute for Computational Biomedicine
Weill Cornell Medical College
Our “second” genome
“Do you, Ashley, take Nesbitt and his genome to be your husband?”
Impact of microbiome on cancer

Commensal Bacteria Control Cancer Response to Therapy by Modulating the Tumor Microenvironment

Noriho Iida,1,* Amiran Dutzev,1!* C. Andrew Stewart,† Loretta Smith,1 Nicol Rebecca A. Weingarten,4 Daniel A. Molina,5 Rosalba Salcedo,6 Timothy Back,1 Ren-Ming Dai,1,2 Hiu Kiu,1 Marco Cardone,3 Shrutl Naik,3 Anil K. Patri,6 Ena Francesco M. Marincola,1,7 Karen M. Frank,4 Yasmine Belkaid,3 Giorgio Trinchieri,1,†† Romina S. Goldszmid1,††

Science 22 November 2013

The Intestinal Microbiota Modulates the Anticancer Immune Effects of Cyclophosphamide

Sophie Viaud,1,3 Fabiana Saccheri,1 Grégoire Mignot,3,5 Takahiro Yamazaki,1 Remon Daillère,1,3 Dalil Hannani,1 David P. Enot,7 Cristina Pfirsich,3 Camilla Engblom,8 Mikael J. Pittet,9 Andreas Schlüter,10 Florent Ginhoux,10 Lionel Apetoh,11 Elisabeth Chachaty,11 Paul-Louis Woerther,11 Gérard Eberl,12 Marion Bérard,13 Chantal Ecobichon,14,15 Dominique Clermont,16 Chantal Bizet,16 Valérie Gaboriau-Routhiau,17 Nadine Cerf-Bensussan,17,18 Paule Opolot,19,20 Nadia Yassaad,21,22,23,24 Eric Vivier,21,22,23,24 Bernhard Ryffel,25 Charles O. Elson,26 Joël Dore,27,28 Guido Kroemer,7,8,29,30 Patricia Lepage,17,27 Ivo Gomperts Boneca,34,15 François Ghiringhelli,5,*, Laurence Zitvogel1,3,†

Genomic analysis identifies association of Fusobacterium with colorectal carcinoma

Aleksandar D. Kostic,1,2 Dirk Gevers,1 Chandra Sekhar Pedamallu,1,3 Monia Michaud,4 Fujiko Duke,1,3 Ashlee M. Earl,1 Akinyemi I. Ojesina,1,3 Joonil Jung,1 Adam J. Bass,1,3 Josep Tabernero,5 José Baselga,5 Chen Liu,6 Ramesh A. Shvidasani,3 Shuji Ogino,2,3,7 Bruce W. Birren,1 Curtis Huttenhower,1,8 Wendy S. Garrett,1,3,4 and Matthew Meyerson1,2,3,9

1Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts 02142, USA; 2Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA; 3Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA; 4Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts 02115, USA; 5Department of Medical Oncology, Vall d’Hebron University Hospital, Universitat Autònoma de Barcelona, 08035 Barcelona, Spain; 6Departments of Pathology, Immunology, and Laboratory Medicine, University of Florida College of Medicine, Gainesville, Florida 32610, USA; 7Department of Pathology, Brigham and Women’s Hospital, Boston, Massachusetts 02115, USA; 8Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, USA

Genome Res. 2011.
Disruption of host-microbiota and onset of immune response is a possible mechanism of cancer development.
Helicobacter pylori

Barry Marshall
Robin Warren
**H. pylori fact sheet**

- Endemic infection present in almost half of all human population. Particularly prevalent in developing countries.

- Gram-negative bacteria that infects the antrum mucosa lining of the stomach leading to increase in pH levels.

- Causes chronic gastritis, an inflammation of the stomach lining, ulcer, and even atrophy of the stomach lining.

- World Health Organization class-I carcinogenic for development of gastric cancer although most infected individuals do not develop cancer. It is one of the strongest risk factors for gastric cancer although reasons for this increased risk are not fully understood.
Aims

• To establish the feasibility of microbiome detection by whole genome sequencing of stomach biopsies.

• To examine the impact of *H. pylori* infection on the gastric microbiome and changes in microbiome following treatment of *H. pylori* infection.

• To detect microbiome differences between normal mucosa and cancer samples.

• To profile the immune response of *H. pylori* infected samples and treated samples.
Study design

• Biopsies collected from patients with and without *H. pylori* infection that are undergoing upper endoscopy.

• Three locations from antrum and stomach body are sampled.

• 28 samples collected from 22 patients and sequenced as paired-end 50 at ~10x coverage.
Microbial identification by sequencing

16S rRNA

Variable areas
Conserved primer areas

Nature Reviews Microbiology 9, 244-253, 2011

Fred Matsen, http://matsen.fhcrc.org
Microbial identification by sequencing

Whole genome sequencing

http://bacteria.ensembl.org
Only a small fraction of reads are used for bacterial identification

<table>
<thead>
<tr>
<th>Sample</th>
<th># of reads</th>
<th>%hg19 mapped</th>
<th>%unmapped</th>
<th>#used for Microbiome identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>05-Antrum</td>
<td>205.8M</td>
<td>97.18</td>
<td>5.81M(2.82%)</td>
<td>0.262M(0.127%)</td>
</tr>
<tr>
<td>07-Antrum</td>
<td>201.9M</td>
<td>96.34</td>
<td>5.36M(2.92%)</td>
<td>0.242M(0.122%)</td>
</tr>
<tr>
<td>08-Antrum</td>
<td>195.4M</td>
<td>97.28</td>
<td>5.31M(2.72%)</td>
<td>0.212M(0.108%)</td>
</tr>
<tr>
<td>09-Antrum</td>
<td>196.2M</td>
<td>97.08</td>
<td>5.73M(2.92%)</td>
<td>0.239M(0.122%)</td>
</tr>
<tr>
<td>09-ProxB</td>
<td>175.7M</td>
<td>97</td>
<td>5.28M(3.0%)</td>
<td>0.236M(0.134%)</td>
</tr>
<tr>
<td>10-Antrum</td>
<td>185.1M</td>
<td>97.2</td>
<td>5.12M(2.77%)</td>
<td>0.288M(%0.156)</td>
</tr>
<tr>
<td>11-Antrum</td>
<td>187.2M</td>
<td>97.5</td>
<td>4.60M(2.46%)</td>
<td>0.191M(%0.102)</td>
</tr>
</tbody>
</table>
Analysis Pipeline

Map to the human reference genome hg19

unmapped

MetaPhlAn
(align to a database of unique bacterial sequences)
Metagenomics analysis results
Initial MetaPhlAn results contained large fraction of unknown bacteria

MetaPhlAn microbiome profile
Revised Analysis Pipeline

- Map to the human reference genome hg19
- Map to 50 additional assembled human genomes
- Filter repeat regions
- Map to all NCBI human sequences
- Map to all assembled bacteria genomes (~1400 species)
- Evaluate bacterial genomic coverage
- Normalize

- Filtering of human DNA using modified PathSeq protocol (Kostic et al. Nature Biotech, 2011)
- Map to all assembled bacteria genomes using bwa + pathoscope
- Evaluate coverage along bacterial genome
- Normalization and relative abundance calculation
Mapping steps remove false identification

Thermoanaerobacter wiegelii

Log_{10}(CPM+1E^{-6})

BWA(UCSC_hg19)  BWA(Other genomes)  RepeatMasker+MegaBLAST  BLASTN

Samples:
- 05-Antrum-B2
- 05-Antrum-B3
- 07-Animal-B2
- 08-Animal-B2
- 09-Animal-B2
- 09-Animal-B2
- 10-Animal-B1
- 11-Animal-B1
- 11-Animal-B1
- 12-Animal-B1
- 13-Animal-B1
- 14-Animal-B1
- 14-Animal-B1
- 15-Animal-B1
- 15-Animal-B1
Mapping steps remove false identification

Thermoanaerobacter wiegelii (10–Antrum–B2)
Helicobacter pylori

Log10(CPM+1E-6)

BWA(UCSC_hg19)  BWA(Other genomes)  RepeatMasker+MegaBLAST  BLASTN
Using Q-Q plot to measure coverage

QQ–plot of Helicobacter pylori (10–An)

QQ–plot of Candidatus Carsonella ruddii (10–An)
qPCR validation confirmed quantification of microbiome content by WGS

(A) Spearman rank cor. = 0.908

(B) Spearman rank cor. = 0.785

(C) Spearman rank cor. = 0.842
Mapping to bacteria genomes is consistent with Human Microbiome study.

- Bacteroides_vulgatus
- Bifidobacterium_breve
- Bifidobacterium_longum
- Bifidobacterium_undefined
- Burkholderia_cenocepacia
- Finegoldia_magna
- Gardnerella_vaginalis
- Lactobacillus_amylovorus
- Lactobacillus_brevis
- Lactobacillus_crispus
- Lactobacillus_fermentum
- Lactobacillus_gasseri
- Lactobacillus_iners
- Lactobacillus_jensenii
- Lactobacillus_johnsonii
- Lactobacillus_timonensis
- Prevotella_bivia
- Prevotella_bivia
- Propionibacterium_acnes
- Staphylococcus_aureus
- Staphylococcus_epidermidis
- Ureaplasma_parvum
- Ureaplasma_undefined

Methods: Metaphlan, WGS

<table>
<thead>
<tr>
<th>Species</th>
<th>Metaphlan</th>
<th>WGS</th>
<th>Metaphlan</th>
<th>WGS</th>
<th>Metaphlan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacteroides_vulgatus</td>
<td></td>
<td>0.0</td>
<td></td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>Bifidobacterium_breve</td>
<td></td>
<td>0.0</td>
<td></td>
<td>0.0</td>
<td>0.015</td>
</tr>
<tr>
<td>Bifidobacterium_longum</td>
<td></td>
<td>0.0</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Bifidobacterium_undefined</td>
<td>0.818</td>
<td>0.0</td>
<td></td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>Burkholderia_cenocepacia</td>
<td>0.038</td>
<td>0.0</td>
<td></td>
<td>0.0</td>
<td>0.013</td>
</tr>
<tr>
<td>Finegoldia_magna</td>
<td>0.045</td>
<td>0.0</td>
<td></td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>Gardnerella_vaginalis</td>
<td>28.647</td>
<td>9.276</td>
<td>0.002</td>
<td>0.002</td>
<td>94.315</td>
</tr>
<tr>
<td>Lactobacillus_amylovorus</td>
<td>0.001</td>
<td>0.0</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Lactobacillus_brevis</td>
<td></td>
<td>0.0</td>
<td></td>
<td>0.0</td>
<td>0.967</td>
</tr>
<tr>
<td>Lactobacillus_crispus</td>
<td></td>
<td>0.0</td>
<td></td>
<td>0.0</td>
<td>0.002</td>
</tr>
<tr>
<td>Lactobacillus_fermentum</td>
<td>0.091</td>
<td>0.127</td>
<td>0.002</td>
<td>0.002</td>
<td>96.704</td>
</tr>
<tr>
<td>Lactobacillus_gasseri</td>
<td>65.833</td>
<td>88.934</td>
<td>58.097</td>
<td>59.784</td>
<td>0.002</td>
</tr>
<tr>
<td>Lactobacillus_iners</td>
<td></td>
<td>0.0</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Lactobacillus_jensenii</td>
<td>0.174</td>
<td>0.0</td>
<td>0.002</td>
<td>0.164</td>
<td>3.754</td>
</tr>
<tr>
<td>Lactobacillus_johnsonii</td>
<td></td>
<td>0.0</td>
<td></td>
<td>0.534</td>
<td>0.020</td>
</tr>
<tr>
<td>Lactobacillus_timonensis</td>
<td>0.009</td>
<td>0.0</td>
<td></td>
<td>0.0</td>
<td>0.002</td>
</tr>
<tr>
<td>Prevotella_bivia</td>
<td></td>
<td>0.0</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Prevotella_bivia</td>
<td></td>
<td>0.0</td>
<td></td>
<td>0.0</td>
<td>0.002</td>
</tr>
<tr>
<td>Prevotella_bivia</td>
<td></td>
<td>0.007</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Propionibacterium_acnes</td>
<td>0.071</td>
<td>0.134</td>
<td>0.011</td>
<td>0.013</td>
<td>0.014</td>
</tr>
<tr>
<td>Staphylococcus_aureus</td>
<td></td>
<td>0.0</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Staphylococcus_epidermidis</td>
<td>3.339</td>
<td>1.27</td>
<td>0.0</td>
<td>0.0</td>
<td>0.015</td>
</tr>
<tr>
<td>Ureaplasma_parvum</td>
<td>0.969</td>
<td>0.0</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Ureaplasma_undefined</td>
<td></td>
<td>0.0</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Percentage

- 1e+01
- 1e-01
- 1e-03
- 1e-05
### Patient group with mixed *H. pylori* status

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Gender</th>
<th>Gastric Cancer</th>
<th><em>H. pylori</em> status</th>
<th>Prior <em>H. pylori</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>05</td>
<td>F</td>
<td></td>
<td>Negative</td>
<td>Y(2009)</td>
</tr>
<tr>
<td>07</td>
<td>F</td>
<td></td>
<td>Negative</td>
<td>Y(2010)</td>
</tr>
<tr>
<td>08</td>
<td>M</td>
<td></td>
<td>Active infection</td>
<td>N</td>
</tr>
<tr>
<td>09</td>
<td>M</td>
<td></td>
<td>Active infection</td>
<td>Y(2003)</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td></td>
<td>Active infection</td>
<td>Y(2009)</td>
</tr>
<tr>
<td>11</td>
<td>F</td>
<td>X</td>
<td>Negative</td>
<td>N</td>
</tr>
<tr>
<td>12</td>
<td>F</td>
<td></td>
<td>Negative</td>
<td>N</td>
</tr>
<tr>
<td>13</td>
<td>M</td>
<td></td>
<td>Negative</td>
<td>N</td>
</tr>
<tr>
<td>14</td>
<td>M</td>
<td>X</td>
<td>Negative</td>
<td>Y(2013)</td>
</tr>
<tr>
<td>15</td>
<td>F</td>
<td></td>
<td>Active infection</td>
<td>Y(2012)</td>
</tr>
<tr>
<td>16</td>
<td>M</td>
<td>X</td>
<td>Negative</td>
<td>N</td>
</tr>
<tr>
<td>17</td>
<td>F</td>
<td>X</td>
<td>Negative</td>
<td>N</td>
</tr>
<tr>
<td>18</td>
<td>M</td>
<td>X</td>
<td>Negative</td>
<td>N</td>
</tr>
<tr>
<td>19</td>
<td>F</td>
<td></td>
<td>Negative</td>
<td>Y</td>
</tr>
<tr>
<td>20</td>
<td>M</td>
<td></td>
<td>Negative</td>
<td>Y</td>
</tr>
<tr>
<td>21</td>
<td>F</td>
<td></td>
<td>Negative</td>
<td>Y</td>
</tr>
<tr>
<td>22</td>
<td>F</td>
<td></td>
<td>Active infection</td>
<td>N</td>
</tr>
<tr>
<td>23</td>
<td>F</td>
<td></td>
<td>Active infection</td>
<td>N</td>
</tr>
<tr>
<td>24</td>
<td>F</td>
<td></td>
<td>Active infection</td>
<td>N</td>
</tr>
<tr>
<td>25</td>
<td>F</td>
<td></td>
<td>Active infection</td>
<td>Y</td>
</tr>
<tr>
<td>26</td>
<td>F</td>
<td></td>
<td>Negative</td>
<td>Y</td>
</tr>
<tr>
<td>27</td>
<td>F</td>
<td></td>
<td>Negative</td>
<td>Y</td>
</tr>
<tr>
<td>28</td>
<td>F</td>
<td></td>
<td>Active infection</td>
<td>N</td>
</tr>
<tr>
<td>29</td>
<td>F</td>
<td></td>
<td>Active infection</td>
<td>N</td>
</tr>
</tbody>
</table>
Clinical Microbiome Findings

**Species**
- *Streptococcus parasanguinis*
- *Staphylococcus epidermidis*
- *Campylobacter concisus*
- *Enterobacter aerogenes*
- *Enterobacter cloacae*
- *Moraxella catarrhalis*
- *Pseudomonas poae*
- *Enterococcus 7L76*
- *Helicobacter pylori*
- *Lactococcus garvieae*
- *Helicobacter sulex*
- *Klebsiella pneumoniae*
- *Lactobacillus gasseri*
- *Lactobacillus paracasei*
- *Lactobacillus lactis*
- *Monaella catenulata*
- *Propionibacterium acnes*
- *Pseudomonas rosea*
- *Staphylococcus epidermidis*
- *Streptococcus parasanguinis*
- *Streptococcus thermophilus*

**Clinical Status**
- **Prox**
- **Tumor**
- **Antrum**
- **B1**

**Percentage**
- 1e+01
- 1e+00
- 1e+00
- 1e-00
- 1e-00

**Clinical Microbial Profiles**
Comprehensive molecular characterization of gastric adenocarcinoma

The Cancer Genome Atlas Research Network
Microbiome analysis of TCGA gastric samples identified *H. pylori* in 48% of the samples.
Microbiome analysis of TCGA gastric samples identified *H. pylori* in 48% of the samples.

*H. pylori* detected in RNA–seq

- Matching normal
- Tumor
Conclusions

• Detection of microbiome from gastric biopsies using WGS is possible.

• Species identification requires extensive filtering and is dependent on sequencing approach.

• Preliminary results: prior treatment of *H. pylori* infection may not fully clear the infection, or infection may recur. **Persistent infection may contribute to carcinogenesis.**
Acknowledgments

Funding:
DeGregorio Family Foundation
CADC Pilot Funding
Starr Cancer Consortium

Chao Zhang
Manish Shah

Kyle Cleveland
Parshant Thakkar
Felice Schnoll-Sussman
Bridget McClure
Michelle Bigg
Nikolaus Schultz (MSKCC)

Epigenomics Core facility at Weill Cornell