

BCB 731:

Defense Against the Dark Arts



Optimist: Microbiome analyses of blood and tissues suggest cancer diagnostic approach

November 27th, 2023



[nature](#) > [articles](#) > article

Article | [Published: 11 March 2020](#)

Microbiome analyses of blood and tissues suggest cancer diagnostic approach

[Gregory D. Poore](#), [Evgenia Kopylova](#), [Qiyun Zhu](#), [Carolina Carpenter](#), [Serena Fraraccio](#), [Stephen Wandro](#), [Tomasz Kosciolek](#), [Stefan Janssen](#), [Jessica Metcalf](#), [Se Jin Song](#), [Jad Kanbar](#), [Sandrine Miller-Montgomery](#), [Robert Heaton](#), [Rana Mckay](#), [Sandip Pravin Patel](#), [Austin D. Swafford](#) & [Rob Knight](#) 

[Nature](#) **579**, 567–574 (2020) | [Cite this article](#)

82k Accesses | **526** Citations | **923** Altmetric | [Metrics](#)

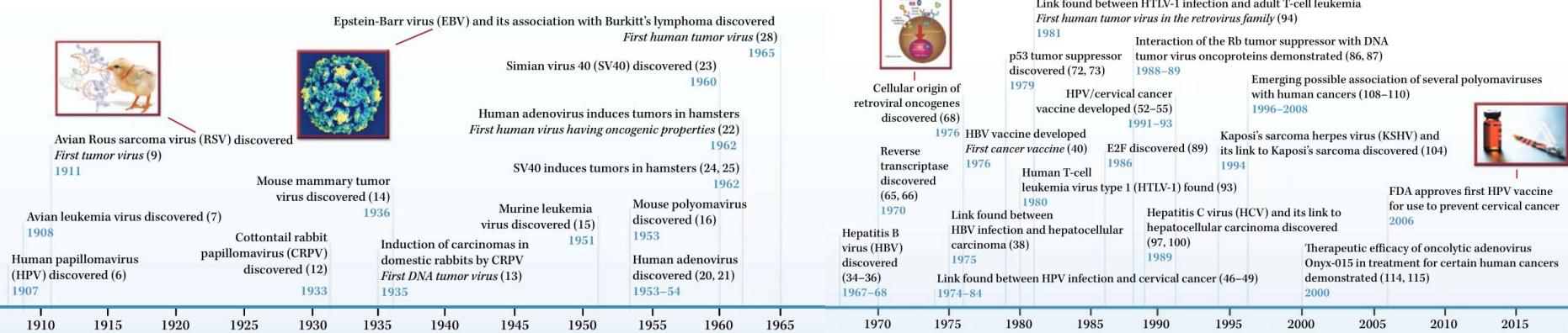
Background

A History of Cancer Research: Tumor Viruses

Joseph Lipsick



Figure 2. The origins of Rous sarcoma virus. (Top) A sarcoma caused by injection of fragments of the transmissible tumor. (Bottom) Histopathologic evidence of a sarcoma invading into muscle. (Reprinted from Rous P. 1910. *J Exp Med* 12: 696–705.)



Cancer and bacteria (1909 & 1918)

Surgical Section.

July 13, 1909.

Mr. J. WARRINGTON HAWARD, President of the Section, in the Chair.

The Treatment of Inoperable Sarcoma by Bacterial Toxins
(the Mixed Toxins of the Streptococcus erysipelas and
the Bacillus prodigiosus).

By WILLIAM B. COLEY, M.D. (New York.)¹

THE subject upon which I have been asked to address you is one upon which I have been working constantly the last seventeen years, and one which has grown more interesting to me with each succeeding year. While the results have not been as satisfactory as one who is seeking perfection could wish, they have been sufficiently real and tangible, I think, to be entitled to more careful consideration than they have yet received. Furthermore, they may have an important bearing upon the whole cancer problem, since, if by the administration of certain bacterial toxins we can cause the degeneration, death, and absorption of living tumour cells of one variety of cancer—sarcoma—it is not unreasonable to suppose that by the use of some other forms of bacterial toxins we may succeed in destroying or inhibiting the growth of the other and more common variety—carcinoma.

THE CANCER PROBLEM

BY C. H. MAYO, M.D.

Rochester, Minnesota

mice and the lower types of life. In this connexion, Erwin Smith's work on cancer in plant life is of great interest. He has shown many plant tumors as being due to bacteria and insect irritation, and he has been able to reproduce and transplant certain tumors which compare in malignancy with those of animal cells. He also predicts almost the localizing of the cancer by use of certain irritants injected with the cancer cells.

Ulceration, under proper conditions, may permit cancer to develop. The ulcerative process itself is usually of bacterial origin, and the bacteria are carried by the circulation, local conditions of infection being developed through capillary infarctions. It is very probable that the essential factor in the development of carcinoma is a derangement inside of the single cell, in an acid field and that

Presented before the Ontario Medical Association, May 29th, 30th and 31st, 1918, at Hamilton, Ont.

Helicobacter pylori & gastric cancer (1983)

UNIDENTIFIED CURVED BACILLI ON GASTRIC EPITHELIUM IN ACTIVE CHRONIC GASTRITIS

SIR.—Gastric microbiology has been sadly neglected. Half the patients coming to gastroscopy and biopsy show bacterial colonisation of their stomachs, a colonisation remarkable for the constancy of both the bacteria involved and the associated histological changes. During the past three years I have observed small curved and S-shaped bacilli in 135 gastric biopsy specimens. The bacteria were closely associated with the surface epithelium, both within and between the gastric pits. Distribution was continuous, patchy, or focal. They were difficult to see with haematoxylin and eosin stain, but stained well by the Warthin-Starry silver method (figure).

I have classified gastric biopsy findings according to the type of inflammation, regardless of other features, as "no inflammation", "chronic gastritis" (CG), or "active chronic gastritis" (ACG). CG shows more small round cells than normal while ACG is characterised by an increase in polymorphonuclear neutrophil leucocytes, besides the features of CG. It was unusual to find no inflammation. CG usually showed superficial oedema of the mucosa. The leucocytes in ACG were usually focal and superficial, in and near the surface epithelium. In many cases they only infiltrated the necks of occasional gastric glands. The superficial epithelium was often irregular, with reduced mucinogenesis and a cobblestone surface.

When there was no inflammation bacteria were rare. Bacteria were often found in CG, but were rarely numerous. The curved bacilli were almost always present in ACG, often in large numbers and often growing between the cells of the surface epithelium (figure). The constant morphology of these bacteria and their intimate relationship with the mucosal architecture contrasted with the heterogeneous bacteria often seen in the surface debris. There was normally a layer of mucus secretion on the surface of the mucosa. When this layer was intact, the debris was spread over it, while the curved bacilli were on the epithelium beneath, closely spaced over the surface (figure).

The curved bacilli and the associated histological changes may be present in any part of the stomach, but they were seen most consistently in the gastric antrum. Inflammation, with no bacteria, occurred in mucosa near focal lesions such as carcinoma or peptic ulcer. In such cases, the leucocytes were spread through the full thickness of the nearby mucosa, in contrast to the superficial infiltration associated with the bacteria. Both the bacteria and the typical histological changes were commonly found in mucosa unaffected by the focal lesion.



Curved bacilli on gastric epithelium.

Section is cut at acute angle to show bacteria on surface, forming network between epithelial cells. (Warthin-Starry silver stain; bar = 10 µm.)

Determinative Bacteriology. The stomach must not be viewed as a sterile organ with no permanent flora. Bacteria in numbers sufficient to see by light microscopy are closely associated with an active form of gastritis, a cause of considerable morbidity (dyspeptic disease). These organisms should be recognised and their significance investigated.

Department of Pathology,
Royal Perth Hospital,
Perth, Western Australia 6001

J. ROBIN WARREN

Helicobacter pylori & gastric cancer

Helicobacter pylori and Gastric Cancer: The Causal Relationship

Review article: exploring the link between *Helicobacter pylori* and gastric cancer

Shajan Peter Christoph Beglinger

Department of Gastroenterology, University Hospital, Basel, Switzerland

Helicobacter Pylori is the Cause of Gastric Cancer

Ilija Barukčić¹

The role of *Helicobacter pylori* infection in the web of gastric cancer causation

Bárbara Peleteiro^{a,b}, Carlo La Vecchia^{c,d} and Nuno Lunet^{a,b}

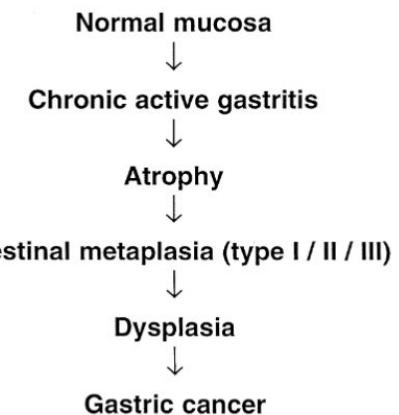


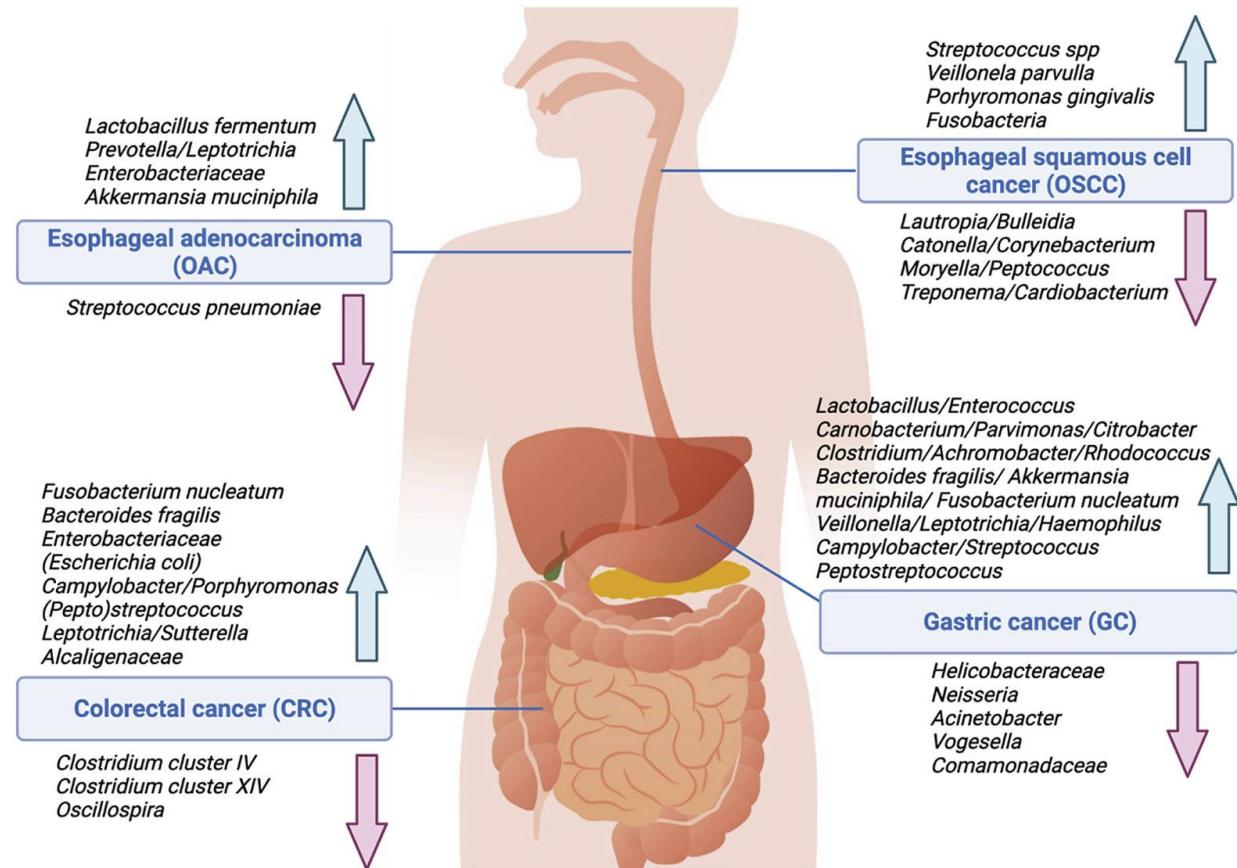
Figure 1. The histological cascade leading to intestinal type gastric cancer.

Other bacteria / other cancers

The Role of Microbiota in Gastrointestinal Cancer and Cancer Treatment: Chance or Curse?



Annemieke Smet,^{1,2} Juozas Kupcinskas,³ Alexander Link,⁴ Georgina L. Hold,^{5,§} and Jan Bornschein^{6,§}
¹Laboratory of Experimental Medicine and Paediatrics, Faculty of Medicine and Health Sciences, ²Infla-Med Research Consortium of Excellence, University of Antwerp, Antwerp, Belgium, ³Institute for Digestive Research, Department of Gastroenterology, Lithuanian University of Health Sciences, Kaunas, Lithuania, ⁴Department of Gastroenterology, Hepatology and Infectious Diseases, Otto-von-Guericke University, Magdeburg, Germany, ⁵Microbiome Research Centre, St George and Sutherland Clinical School, University of New South Wales, Sydney, Australia, and ⁶Translational Gastroenterology Unit, Nuffield Department of Experimental Medicine, University of Oxford, John Radcliffe Hospital, Oxford, United Kingdom

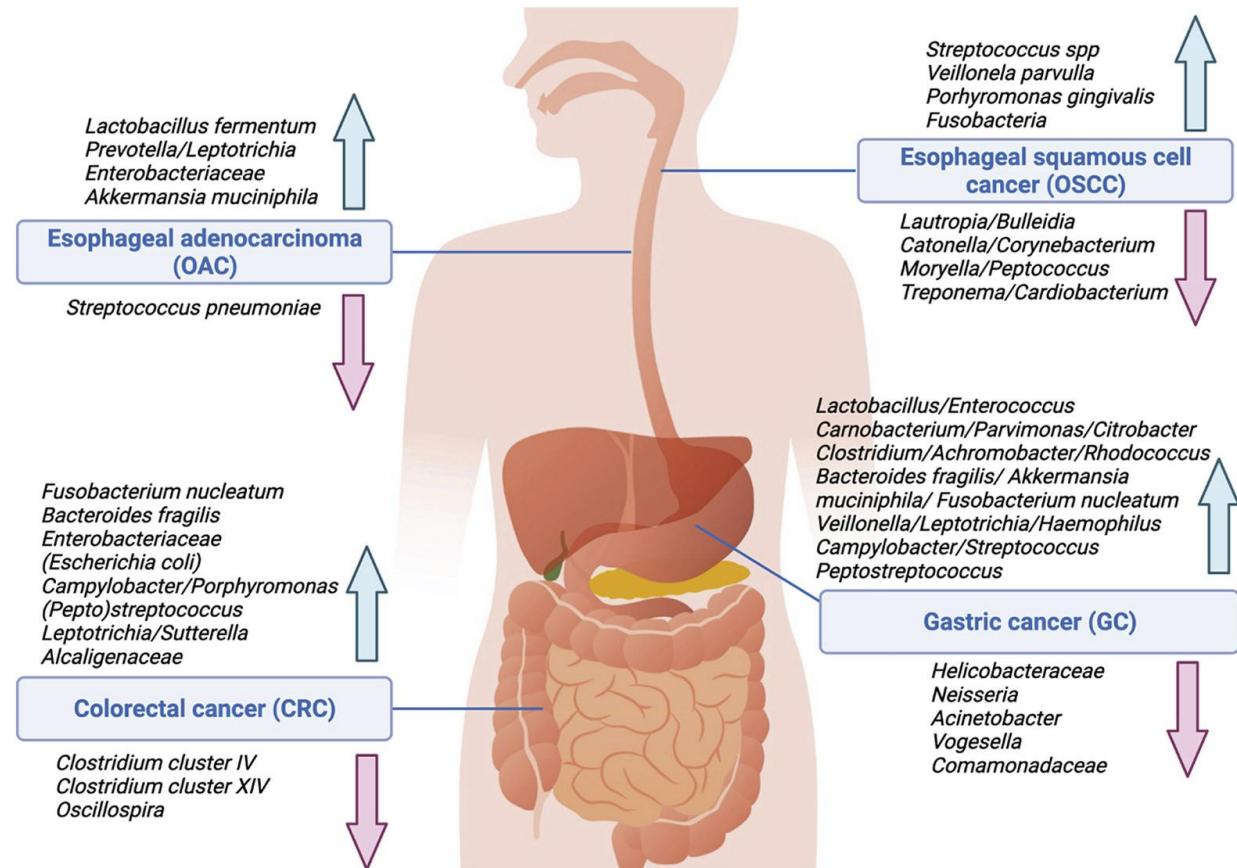


Other bacteria / other cancers

The Role of Microbiota in Gastrointestinal Cancer and Cancer Treatment: Chance or Curse?



Annemieke Smet,^{1,2} Juozas Kupcinskas,³ Alexander Link,⁴ Georgina L. Hold,^{5,§} and Jan Bornschein^{6,§}
¹Laboratory of Experimental Medicine and Paediatrics, Faculty of Medicine and Health Sciences, ²Infla-Med Research Consortium of Excellence, University of Antwerp, Antwerp, Belgium, ³Institute for Digestive Research, Department of Gastroenterology, Lithuanian University of Health Sciences, Kaunas, Lithuania, ⁴Department of Gastroenterology, Hepatology and Infectious Diseases, Otto-von-Guericke University, Magdeburg, Germany, ⁵Microbiome Research Centre, St George and Sutherland Clinical School, University of New South Wales, Sydney, Australia, and ⁶Translational Gastroenterology Unit, Nuffield Department of Experimental Medicine, University of Oxford, John Radcliffe Hospital, Oxford, United Kingdom



Distal effects of bacteria on cancer immunity?

Diet-driven microbial ecology underpins associations between cancer immunotherapy outcomes and the gut microbiome

Rebecca C. Simpson^{1,2,3,19}, Erin R. Shanahan^{3,4,19}, Marcel Batten^{1,2,5,19}, Irene L. M. Reijers⁶, Mark Read^{1,3,7,8}, Ines P. Silva^{1,9}, Judith M. Versluis⁶, Rosilene Ribeiro^{3,4}, Alexandra S. Angelatos^{3,10}, Jian Tan^{3,10}, Chandra Adhikari¹, Alexander M. Menzies^{1,2,11}, Robyn P. M. Saw^{1,2,12}, Maria Gonzalez¹, Kerwin F. Shannon^{1,12}, Andrew J. Spillane^{1,4,13}, Rebecca Velickovic^{1,2}, Alexander J. Lazar^{1,14,15}, Ashish V. Damania¹⁵, Aditya K. Mishra¹⁵, Manoj Chevanambi¹⁶, Anik Banerjee¹⁷, Nadim J. Ajami¹⁵, Jennifer A. Wargo^{15,16}, Laurence Macia^{1,3,10}, Andrew J. Holmes^{3,4}, James S. Wilmott^{1,2,3}, Christian U. Blank^{1,6}, Richard A. Scolyer^{1,2,3,18} and Georgina V. Long^{1,2,3,11,12}

The gut microbiota shapes the response to immune checkpoint inhibitors (ICIs) in cancer, however dietary and geographic influences have not been well-studied in prospective trials. To address this, we prospectively profiled baseline gut (fecal) microbiota signatures and dietary patterns of 103 trial patients from Australia and the Netherlands treated with neoadjuvant ICIs for high risk resectable metastatic melanoma and performed an integrated analysis with data from 115 patients with melanoma treated with ICIs in the United States. We observed geographically distinct microbial signatures of response and immune-related adverse events (irAEs). Overall, response rates were higher in *Ruminococcaceae*-dominated microbiomes than in *Bacteroidaceae*-dominated microbiomes. Poor response was associated with lower fiber and omega 3 fatty acid consumption and elevated levels of C-reactive protein in the peripheral circulation at baseline. Together, these data provide insight into the relevance of native gut microbiota signatures, dietary intake and systemic inflammation in shaping the response to and toxicity from ICIs, prompting the need for further studies in this area.

nature medicine

Article

<https://doi.org/10.1038/s41591-023-02453-x>

Fecal microbiota transplantation plus anti-PD-1 immunotherapy in advanced melanoma: a phase I trial

JEM
Journal of Experimental Medicine

REVIEW

Microbiome influencers of checkpoint blockade-associated toxicity

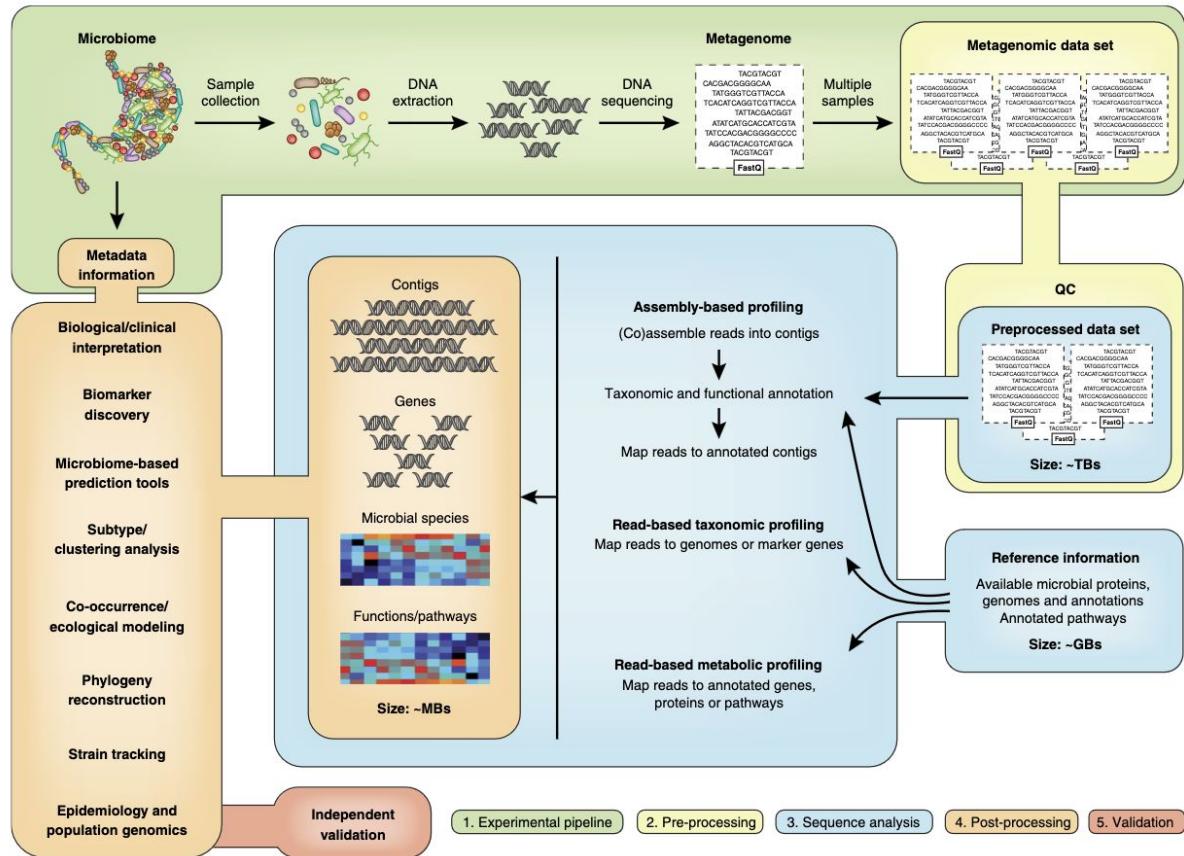
Yinghong Wang^{1,*}, Robert R. Jenq^{2,3*}, Jennifer A. Wargo^{2,3,4}, and Stephanie S. Watowich^{3,5}

Immunotherapy has greatly improved cancer outcomes, yet variability in response and off-target tissue damage can occur with these treatments, including immune checkpoint inhibitors (ICIs). Multiple lines of evidence indicate the host microbiome influences ICI response and risk of immune-related adverse events (irAEs). As the microbiome is modifiable, these advances indicate the potential to manipulate microbiome components to increase ICI success. We discuss microbiome features associated with ICI response, with focus on bacterial taxa and potential immune mechanisms involved in irAEs, and the overall goal of driving novel approaches to manipulate the microbiome to improve ICI efficacy while avoiding irAE risk.

Metagenomics

Shotgun metagenomics, from sampling to analysis

Christopher Quince^{1,7}, Alan W Walker^{2,7} , Jared T Simpson^{3,4}, Nicholas J Loman⁵ & Nicola Segata⁶ 



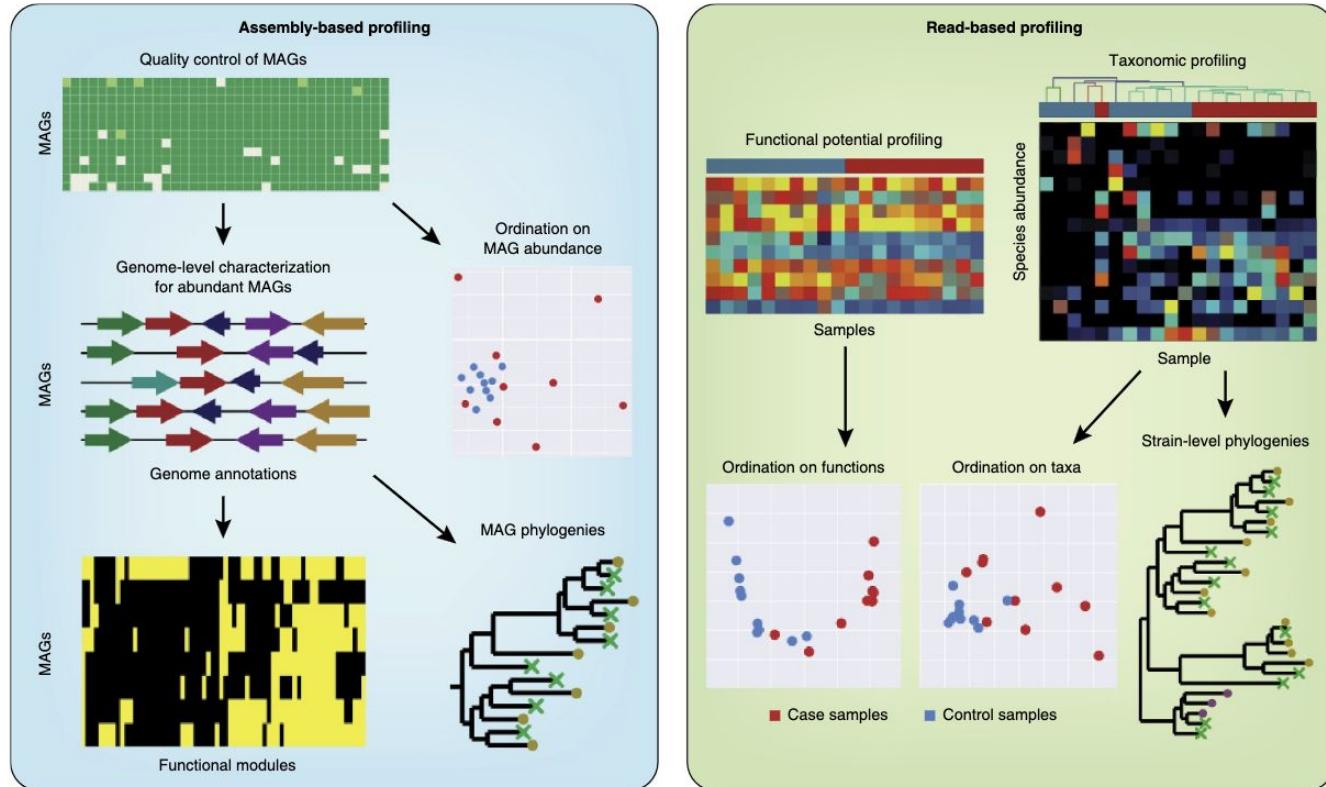
Metagenomics with & without assembly

Binning contigs

Metagenome assemblies are highly fragmented, comprising thousands of contigs (Table 2), and researchers do not know *a priori* which contig derives from which genome, or even know how many genomes are present. The aim of contig 'binning' is to group contigs into species. Supervised binning methods use databases of already sequenced genomes to label contigs into taxonomic classes. Unsupervised (clustering) methods look for natural groups in the data.

Assembly-free metagenomic profiling

Taxonomic profiling of metagenomes identifies which microbial species are present in a metagenome and estimates their abundance. This can be carried out without assembly, through external sequence data resources, such as publicly available reference genomes. This approach can mitigate assembly problems, speed up computation and enable profiling of low-abundance organisms that cannot be assembled *de novo* (Supplementary Box 1). Its main limitation is that previously uncharacterized microbes are difficult to profile (Supplementary Box 1). However, the number of reference genomes available is increasing rapidly, with thousands of genomes produced each year, including some derived from difficult-to-grow species targeted by new cultivation methods⁷⁰, single-cell sequencing approaches⁷¹ or metagenomic assembly. The diversity of reference genomes available for some sample types, such as the human gut⁷², is now extensive enough to make assembly-free taxonomic profiling efficient and successful, including for comparatively low-abundance microbes that lack sufficient sequence coverage and depth to enable assembly of their genome. Analysis of more diverse environments, including soil and oceans, is hampered by a lack of representative reference genomes. As a result, it is generally advisable to use assembly when analyzing metagenomes from these environments.



The Tree(s) of Life

SILVA, RDP, Greengenes, NCBI and OTT —
how do these taxonomies compare?

Monika Balvočiūtė* and Daniel H. Huson

◦ **Bacteria** (eubacteria) Click on organism name to get more information.

- **Acidobacteriota**

- **Blastocatellia**

- [Blastocatellales](#)
- [Candidatus Frugalibacteriales](#)
- [Chloracidobacterium](#)
- [unclassified Blastocatellia](#)
- [environmental samples](#)

- **Candidatus Polarisedimenticola**

- [Candidatus Polarisedimenticolales](#)
- [unclassified Candidatus Polarisedimenticola](#)

- **Holophagae**

- [Acanthopleuribacterales](#)
- [Holophagales](#)
- [Thermotomaculales](#)

Review

NCBI Taxonomy: a comprehensive update on curation, resources and tools

Conrad L. Schoch*, Stacy Ciufo, Mikhail Domrachev, Carol L. Hotton,
Sivakumar Kannan, Rogneda Khovanskaya, Detlef Leipe,
Richard Mcveigh, Kathleen O'Neill, Barbara Robbertse,
Shobha Sharma, Vladimir Sossov, John P. Sullivan, Lu Sun,
Seán Turner and Ilene Karsch-Mizrachi

National Center of Biotechnology Information, National Library of Medicine, National Institutes of Health,
9600 Rockville Pike, Bethesda, MD 20892, USA

Thermotomaculum hydrothermale

Taxonomy ID: [981385](#) (for references in articles please use NCBI:txid981385)
current name

Thermotomaculum hydrothermale Izumi et al. 2017
type strain of *Thermotomaculum hydrothermale*: personal::AC55, [JCM:17643](#), [DSM:24660](#),
[NBRC:107904](#)
includes: *Acidobacteria* bacterium AC55

NCBI BLAST name: **bacteria**

Rank: **species**

Genetic code: [Translation table 11 \(Bacterial, Archaeal and Plant Plastid\)](#)

[Lineage](#)(full)

[cellular organisms](#); [Bacteria](#); [Acidobacteriota](#); [Holophagae](#); [Thermotomaculales](#); [Thermotomaculaceae](#); [Thermotomaculum](#)

Entrez records		
Database name	Direct links	Links from type
Nucleotide	4	4
Protein	4,088	-
Genome	1	-
PubMed Central	6	-
SRA Experiments	19	-
Identical Protein Groups	2,246	-
BioProject	3	-
BioSample	3	2
Assembly	1	1
Taxonomy	1	-

NCBI taxonomy + GenBank = all the genomes

FASTA ▾

Send to: ▾

Thermotomaculum hydrothermale strain AC55 chromosome, complete genome

NCBI Reference Sequence: NZ_AP017470.1

[GenBank](#) [Graphics](#)

>NZ_AP017470.1 Thermotomaculum hydrothermale strain AC55 chromosome, complete genome

```
TTTTTTTTGTGGTGTATTATTGATTAGCAAAATGAATAAAGAATATTAACAGATTGGAAAAATT  
GTGAAAATAATGTGGAAATTGATTAAGACAAATTAAAAACCTATTCCTGAGGAATCATTAAAG  
AATGGTTTGACATACAAAAGAGGTAGATTCATCAAACGGGTAATCAAAGTGGTGTCTGATTCTGA  
GGTTGCTGATTATAAAAGAGTTTATTGGGATTTTATACAAAAGCAATGCTGAAACTGGTGTAAAC  
AAACCGATTGAATTTCATCTAACATGTGCTTAAATGGGGATGTTAAAGCAGCACGGCATTA  
AAACGCCAACATGTCTAACATTGACAAGCTATTTGAATAGCAGTACATACCTTGAACACTTGT  
TGTTGGCCAAGCAATGAACTTGCATTGCAACCTGCAAAGCAGTTGCAGAAAGCCCCAACCGTATAC  
AATCCACTTTATATTACGGAGGGACAGGATTAGGAAAACACATTGTCATGCTGTTGCCACAGGT  
TTTGCAGACTTACCTGAGAAAAAAATTCTTTAATTCAGCTGAAAGGTTTTTACGATTATGTTCT  
TGCTGTGAATCAAATAATTCTTCTTTCAGAGAAAAGTATAGAGAGGTTGATATACTCCTTGTGAT  
GATATTCAATACCTGCAAAATAAGTGGACAAAAGAAGAGTTTCTAGACTTCAATGAATTGTTG  
ACAATAAAAACAATTGTAATATCTTCAAGATTCTTCAACGATATATATCAGTTCTAGGAAAGGCT  
TGTTTCAAGGTTAAATGGGTATTGTCAGATATCCAGGCCACCTGAAATTAGAGACAAGAATTGCTATT  
TTAACAAAAAAAGCAAAAGATTTTAATTTGAGATTGACGATAGCGATTACTTTATTGCAACCAAAT  
TGAGAAAAAAATGTAAGGGATTGGAGGTGCTTAAAGAAATGTGGACTTAATGCCAGGTTGAAAATTC  
AAAAATACTGTTGAAATAGCAAAATGCCCTAAAGATGTTAGATGATAATGATGTTATAACCCCT  
GACAGAATTATGAAATTGTCAGATTACTATAAAACCTGCCGAAAATTAAAGGAAAAAAACAATT  
CAAGAAATATGTTTACCCAGACAAATTGCAATGTACCTTGCACAGTTAACAAATCTATCTTCC
```

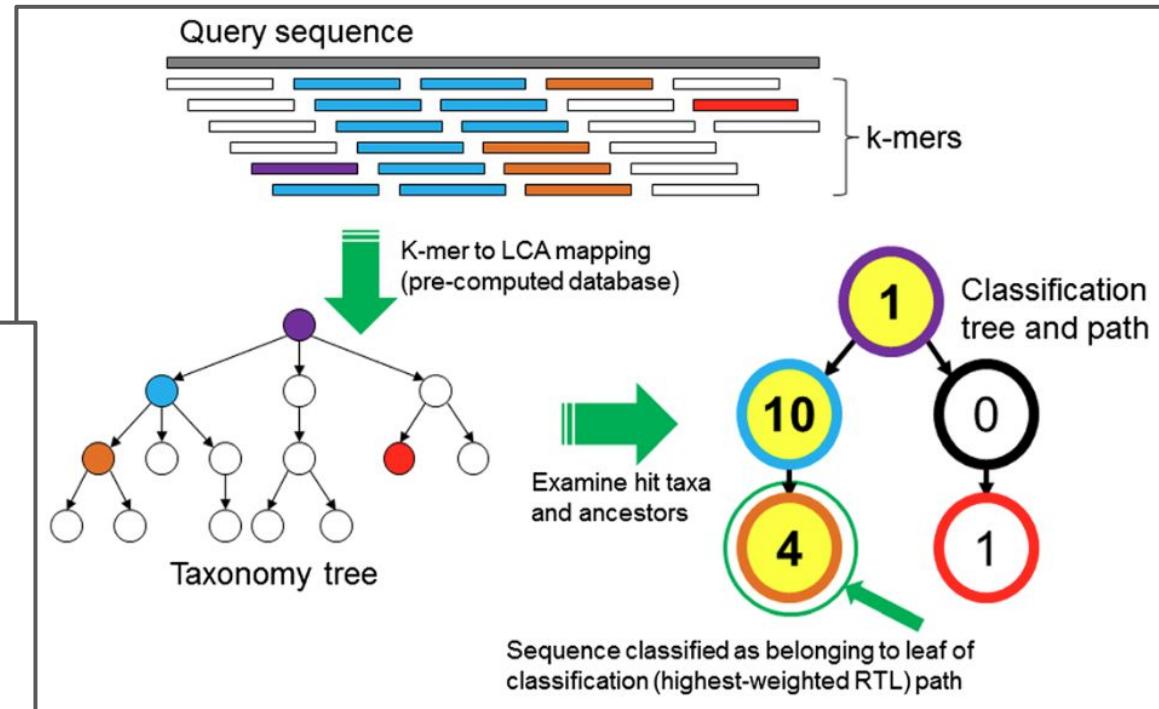
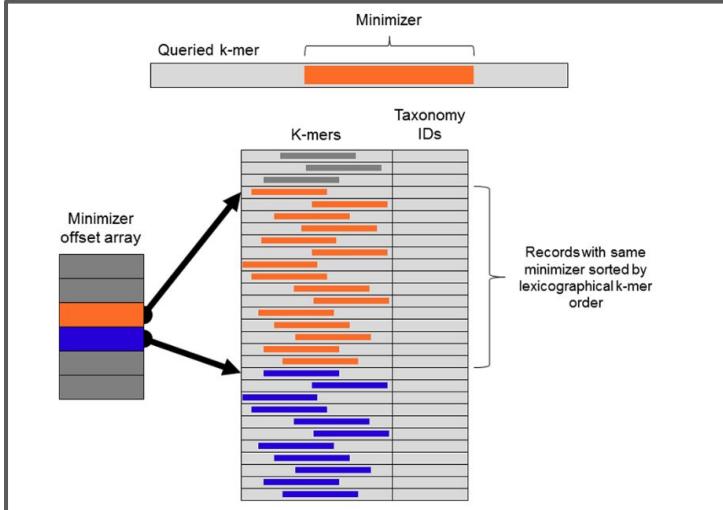


Downloading Large Sequence: 0.89MB

Assembly-free metagenomics w/ Kraken

Kraken: ultrafast metagenomic sequence classification using exact alignments

Derrick E Wood^{1,2*} and Steven L Salzberg^{2,3}



The Cancer Genome Atlas (TCGA)

The TCGA Research Network: The Process of Genomic Discovery

The TCGA program includes a broad cross-section of the cancer research community. The TCGA Research Network includes scientists, bioinformaticians, bioethicists, doctors, nurses, advocates, and many others.

The TCGA Network consists of four basic components:

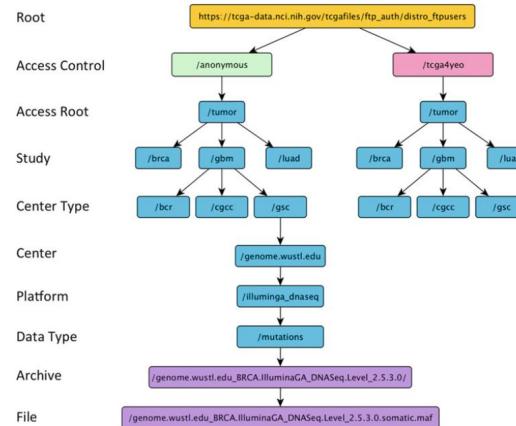
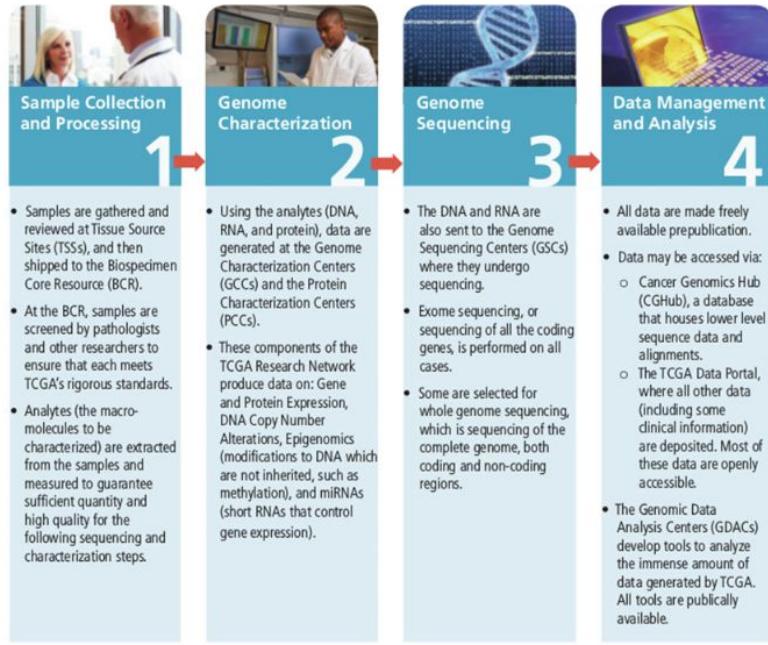
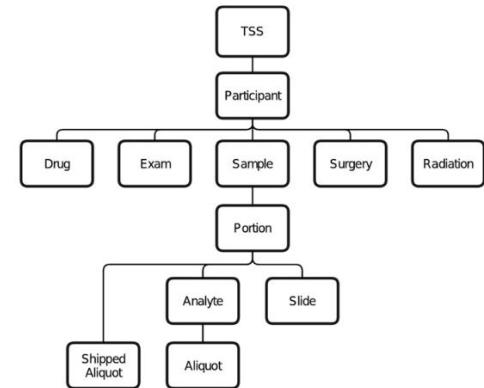
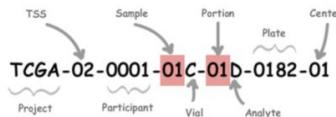
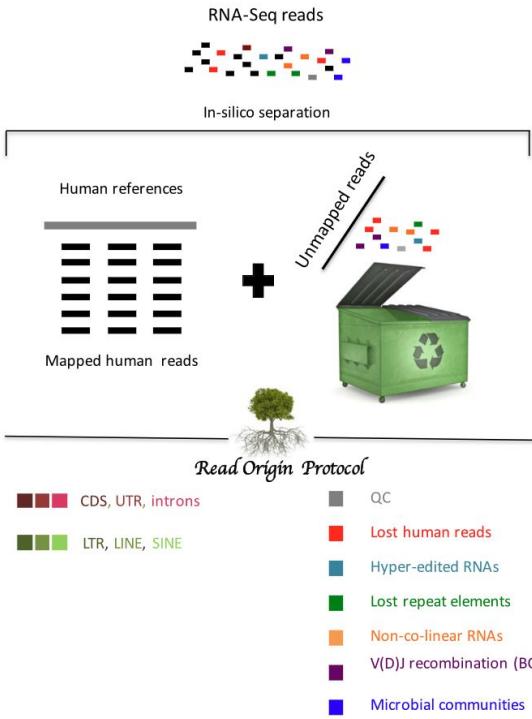


Table 1
TCGA DCC data levels

Data level	Level type	Description	Example
1	Raw	Low-level or instrument data for single aliquot not normalized	Affymetrix CEL file
2	Processed	Normalized single aliquot data	Signal of a probe or probe set for a sample; VCF file; MAF file containing germline mutation calls
3	Segmented or interpreted	Aggregate of processed data from single sample; tumor/normal sample normalized differentials	Expression signal or differential expression of all genes; MAF file containing only somatic mutation calls

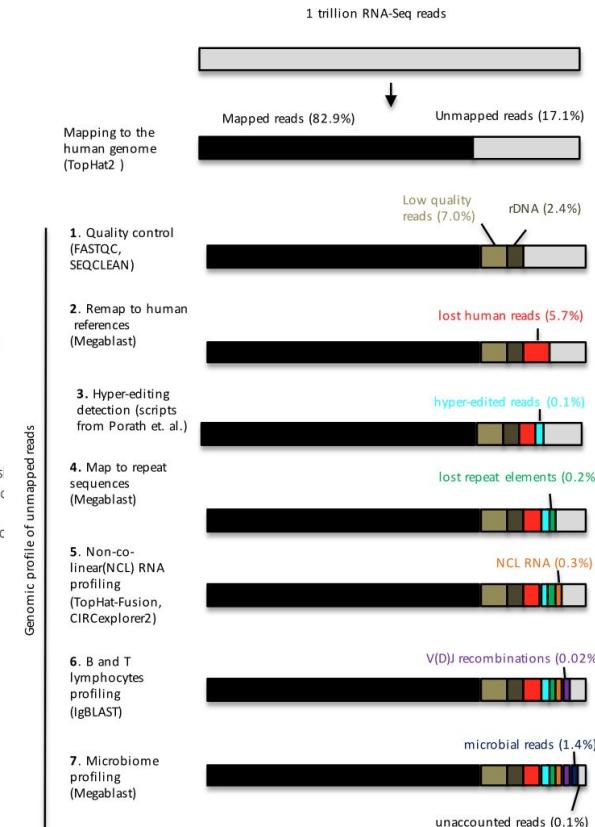


Unmapped Reads (can be a gold mine)



ROP: dumpster diving in RNA-sequencing to find the source of 1 trillion reads across diverse adult human tissues

Serghei Mangul^{1,2*}, Harry Taegyun Yang¹, Nicolas Strauli³, Franziska Gruhl^{4,5}, Hagit T. Porath⁶, Kevin HS Linus Chen⁷, Timothy Daley⁸, Stephanie Christenson⁹, Agata Wesolowska-Andersen¹⁰, Roberto Sporea¹¹, Cydney Rios¹⁰, Celeste Eng¹¹, Andrew D. Smith⁸, Ryan D. Hernandez^{12,13,14}, Roel A. Ophoff^{15,16,17}, Jos Rodriguez Santana¹⁸, Erez Y. Levitan⁹, Prescott G. Woodruff⁹, Esteban Burchard²², Max A. Seibold^{21†}, Sagiv Shifman^{21†}, Eleazar Eskin^{1,16†} and Noah Zaitlen^{9*†}



Now, the paper

[nature](#) > [articles](#) > article

Article | [Published: 11 March 2020](#)

Microbiome analyses of blood and tissues suggest cancer diagnostic approach

[Gregory D. Poore](#), [Evgenia Kopylova](#), [Qiyun Zhu](#), [Carolina Carpenter](#), [Serena Fraraccio](#), [Stephen Wandro](#), [Tomasz Kosciolek](#), [Stefan Janssen](#), [Jessica Metcalf](#), [Se Jin Song](#), [Jad Kanbar](#), [Sandrine Miller-Montgomery](#), [Robert Heaton](#), [Rana Mckay](#), [Sandip Pravin Patel](#), [Austin D. Swafford](#) & [Rob Knight](#) 

[Nature](#) **579**, 567–574 (2020) | [Cite this article](#)

82k Accesses | **526** Citations | **923** Altmetric | [Metrics](#)

a

Disease Type	TCGA Abbreviation
Adrenocortical Carcinoma	ACC
Acute Myeloid Leukemia	AML
Bladder Urothelial Carcinoma	BLCA
Brain Lower Grade Glioma	LGG
Breast Invasive Carcinoma	BRCA
Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma	CESC
Cholangiocarcinoma	CHOL
Colon Adenocarcinoma	COAD
Esophageal Carcinoma	ESCA
Glioblastoma Multiforme	GBM
Head and Neck Squamous Cell Carcinoma	HNSC
Kidney Chromophobe	KICH
Kidney Renal Clear Cell Carcinoma	KIRC
Kidney Renal Papillary Cell Carcinoma	KIRP
Liver Hepatocellular Carcinoma	LIHC
Lung Adenocarcinoma	LUAD
Lung Squamous Cell Carcinoma	LUSC
Lymphoid Neoplasm Diffuse Large B-cell Lymphoma	DLBC
Mesothelioma	MESO
Ovarian Serous Cystadenocarcinoma	OV
Pancreatic Adenocarcinoma	PAAD
Pheochromocytoma and Paraganglioma	PCPG
Prostate Adenocarcinoma	PRAD
Rectum Adenocarcinoma	READ
Sarcoma	SARC
Skin Cutaneous Melanoma	SKCM
Stomach Adenocarcinoma	STAD
Testicular Germ Cell Tumors	TGCT
Thymoma	THYM
Thyroid Carcinoma	THCA
Uterine Carcinosarcoma	UCS
Uterine Corpus Endometrial Carcinoma	UCEC
Uveal Melanoma	UVM

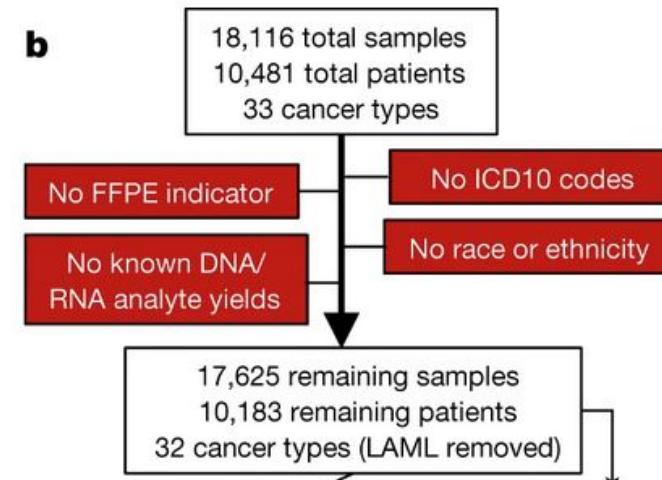
Let's take all the TCGA seq data

Experimental Strategy

- RNA-Seq
- WGS

Sequencing Platform

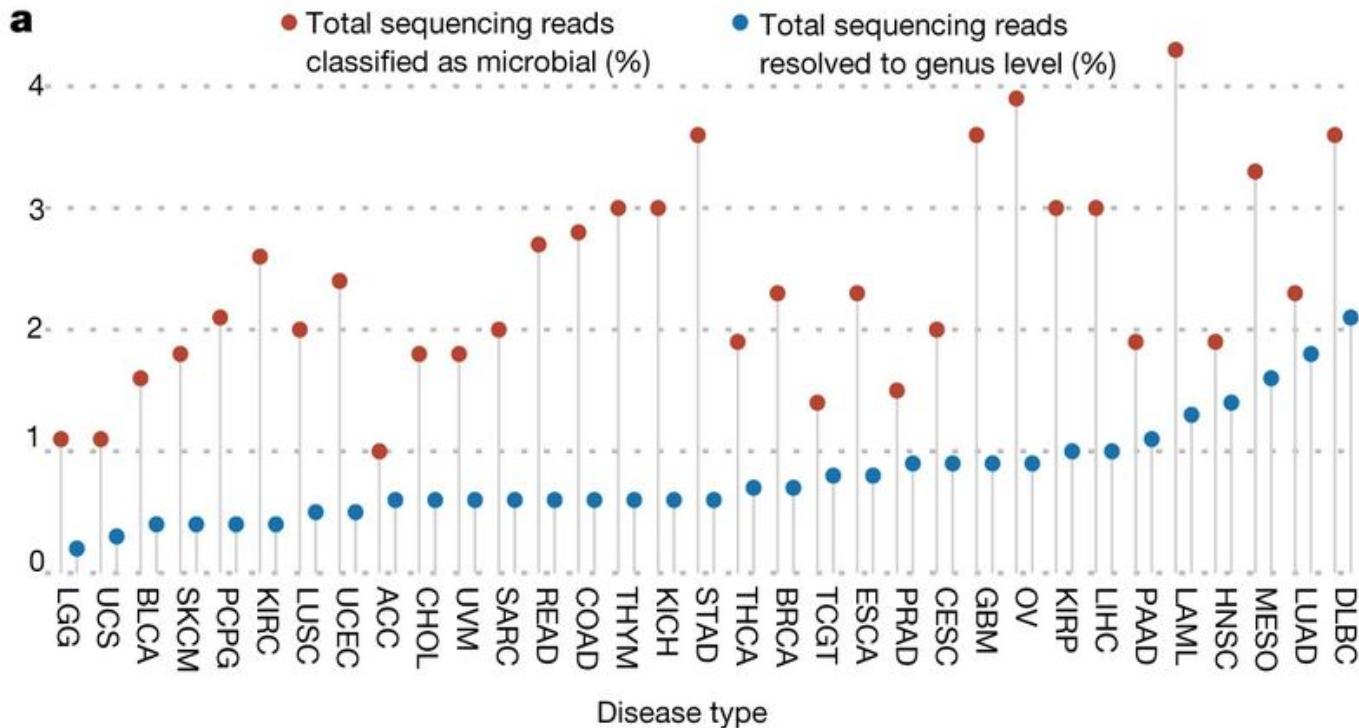
- ABI SOLiD
- HiSeq X Ten
- Illumina GA
- Illumina HiSeq
- Illumina MiSeq
- LS 454

b

...and throw unmapped reads into Kraken

Sequencing reads that did not align to known human reference genomes (based on mapping information in the raw BAM files) were mapped against all known bacterial, archaeal, and viral microbial genomes using the ultrafast Kraken algorithm²³. A total of 71,782 microbial genomes were downloaded using RepoPhlan (<https://bitbucket.org/nsegata/repophlan>) on 14 June 2016, of which 5,503 were viral and 66,279 were bacterial or archaeal. On the basis of prior literature, bacterial and archaeal genomes were filtered for quality scores of 0.8 or better⁵⁸, which left 54,471 of them for subsequent analysis, or a total of 59,974 microbial genomes.

Lots of microbial reads! (~1-4%)



Expected pathogens (HPV & H. pylori) are detected in cervical & stomach cancers

level). The four types of cancer that were directly aligned included CESC as a putative positive viral control (for HPV), STAD as a putative positive bacterial control (for *H. pylori*), and two others (LUAD, OV) based on microbial signatures in the literature and/or available mass-spectrometry proteomic information (data not shown)^{5,24,60,61,62}. We found that 98.91% of

“Pathogen” reads still align to something microbial with BWA

level). The four types of cancer that were directly aligned included CESC as a putative positive viral control (for HPV), STAD as a putative positive bacterial control (for *H. pylori*), and two others (LUAD, OV) based on microbial signatures in the literature and/or available mass-spectrometry proteomic information (data not shown)^{5,24,60,61,62}. We found that 98.91% of reads that were classified to genus level or lower by Kraken (on which our main findings are based) also aligned with BWA to the microbial database (bacteria, archaea, viruses; see Supplementary Table 3), or a false-positive rate of 1.09%, suggesting that the genus-level, Kraken-labelled, pan-cancer microbial reads were sufficiently usable for further analyses.

Beware the manifold of contamination

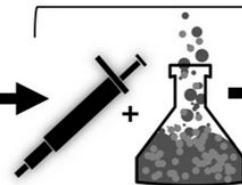
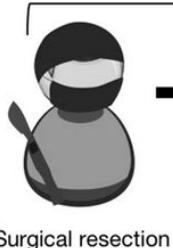
a

Possible sources of contamination and methods to mitigate or measure these effects

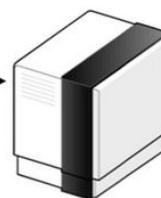
SourceTracker (Knights et al., 2011)
Decontam (Davis et al., 2018) using 17,625
DNA/RNA concentrations

Reagent “negative blank” genera removal (Salter et al., 2014)
Supervised batch correction (SNM; Mecham et al., 2010)
Decontam (Davis et al., 2018)

SNM (Mecham et al., 2010)
Simulated contaminants



Lab reagents and
library prep



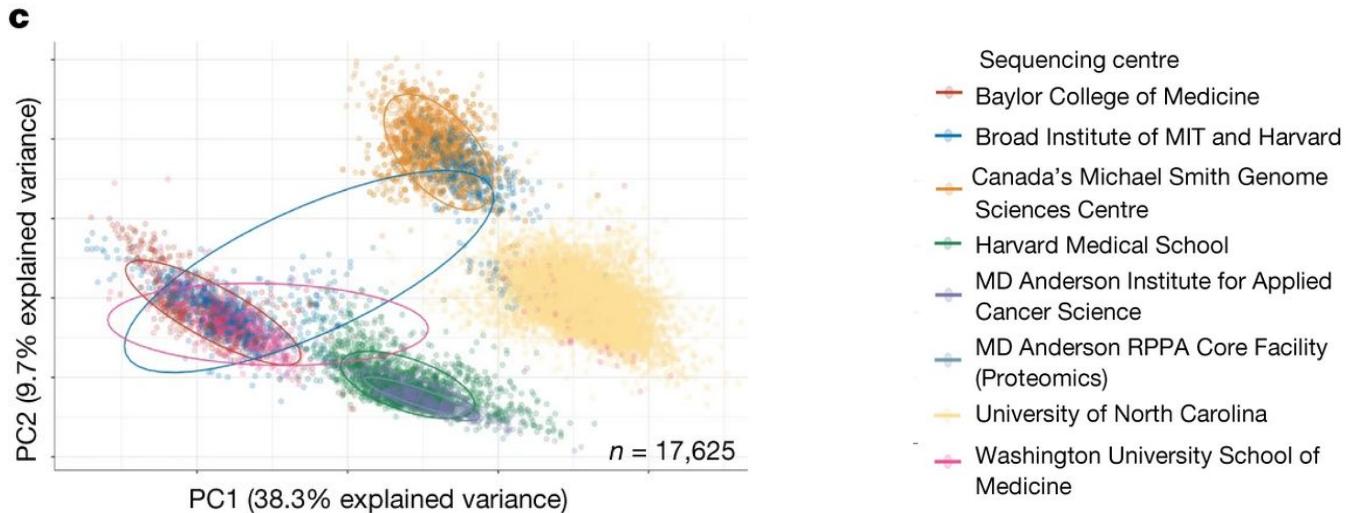
Sequencing platform
cross-contamination



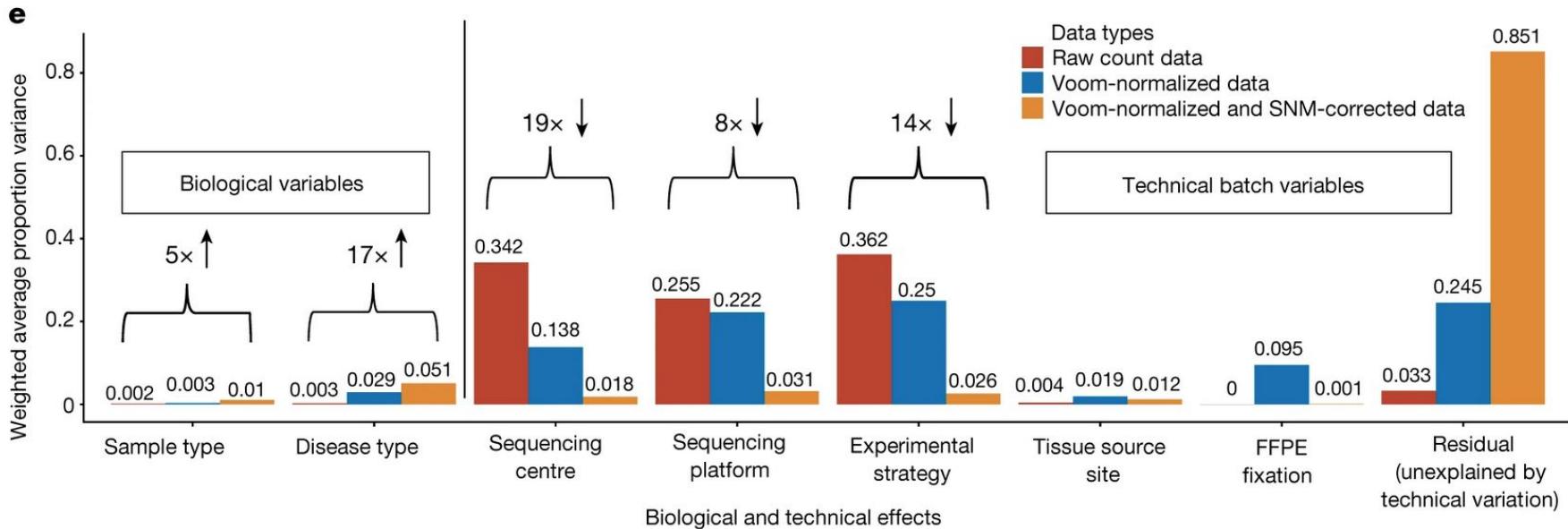
Institutional batch
effects

- Simulated Pseudo-contaminants
- 1) Uniform low level abundance across all samples from one sequencing center
 - 2) Uniform low level abundance across all samples from four sequencing centers
 - 3) Uniform low level abundance across all samples from all sequencing centers
 - 4) Spikes of high abundance across 100 randomly selected samples from one sequencing center
 - 5) Spikes of high abundance across 1000 randomly selected samples from all sequencing centers

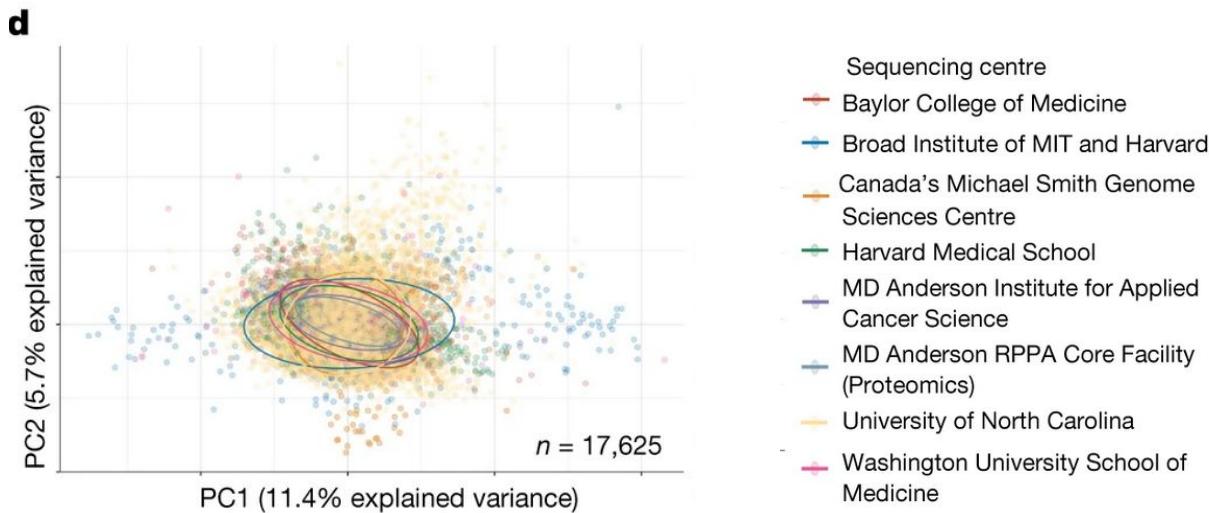
...indeed they see lots of technical artifacts by sequencing center



Let's control for everything!

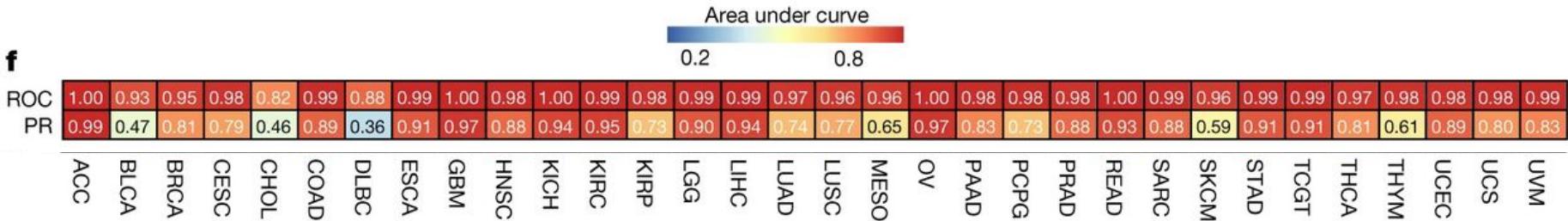


...and it's fixed!



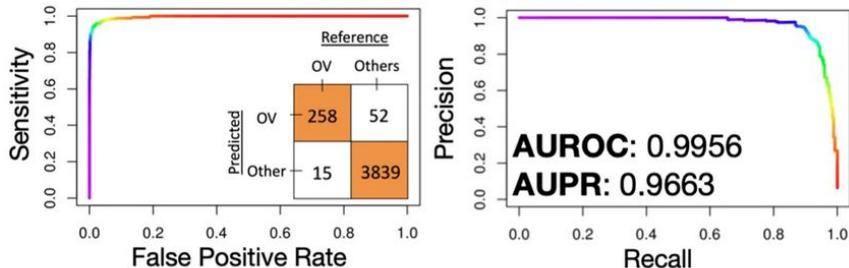
Now we can predict cancer types

f



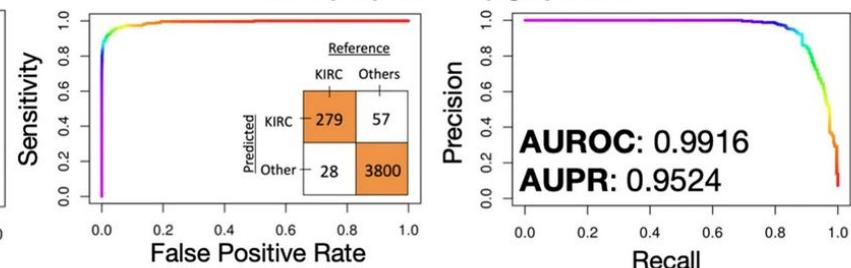
a

TCGA OV vs All Other Types (Primary tumor):
ROC (left) and PR (right) curves



b

TCGA KIRC vs All Other Types (Primary tumor):
ROC (left) and PR (right) curves



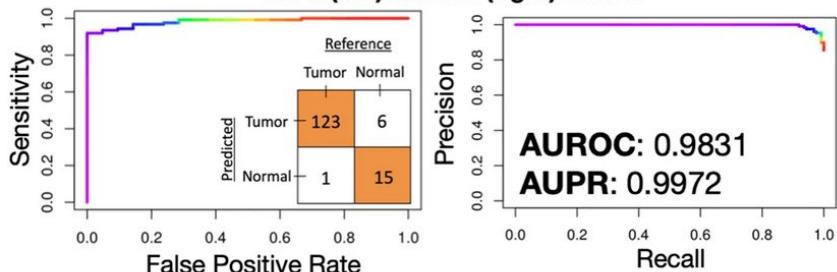
Or tumor vs. normal

g

ROC	NA	0.87	0.97	NA	NA	0.96	NA	0.87	NA	0.91	0.98	0.99	0.84	NA	0.98	0.86	0.96	NA	NA	NA	NA	NA	NA	NA	0.93	NA	0.85	NA	0.99	NA	NA	NA
PR	NA	0.99	1.00	NA	NA	1.00	NA	0.98	NA	0.99	0.99	1.00	0.97	NA	1.00	0.95	0.99	NA	NA	NA	NA	NA	NA	NA	0.99	NA	0.98	NA	1.00	NA	NA	NA
ACC	BLCA	BRCA	CESC	CHOL	COAD	DLBC	ESCA	GBM	HNSC	KICH	KIRC	KIRP	LGG	LIHC	LUAD	LUSC	MESO	OV	PAAD	PRAD	PCPG	READ	SARC	SKCM	STAD	TCGT	THCA	THYM	UCEC	UCS	UVM	

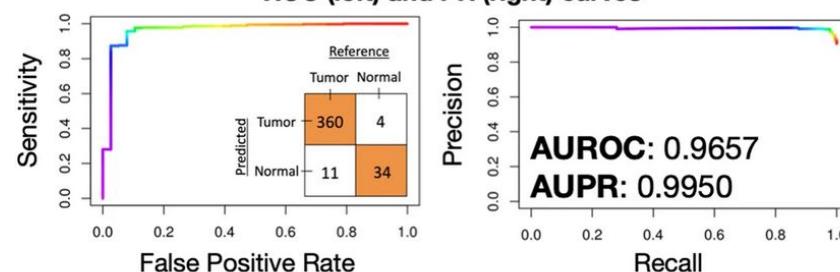
c

TCGA LIHC Tumor vs Normal:
ROC (left) and PR (right) curves



d

TCGA BRCA Tumor vs Normal:
ROC (left) and PR (right) curves



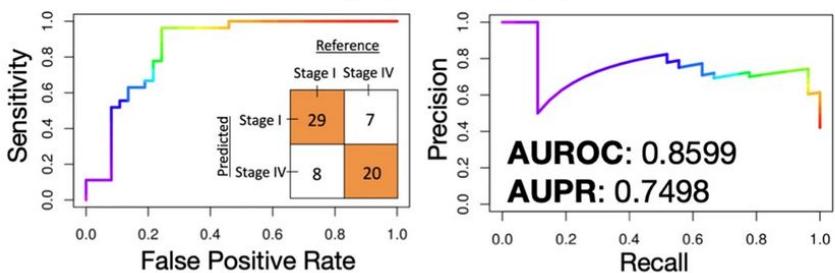
Or tumor stage

h

ROC	NA	NA	0.54	NA	NA	0.80	NA	NA	NA	0.66	NA	0.88	NA	NA	0.48	NA	NA	NA	0.75	NA	NA	0.86	NA	0.61	NA	NA	NA		
PR	NA	NA	0.09	NA	NA	0.81	NA	NA	NA	0.95	NA	0.74	NA	NA	0.11	NA	NA	NA	0.75	NA	NA	0.75	NA	0.20	NA	NA	NA		
ACC	BLCA	BRCA	CESC	CHOL	COAD	DLBC	ESCA	GBM	HNSC	KICH	KIRC	KIRP	LGG	LIHC	LUAD	LUSC	MESO	OV	PAAD	PCPG	PRAD	READ	SARC	SKCM	STAD	TCGT	THYM	UVM	UCS

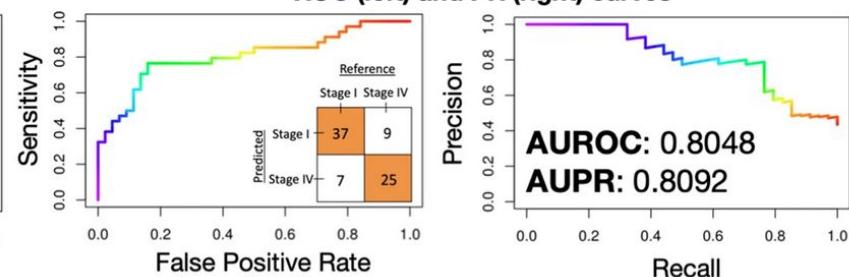
e

TCGA STAD Stage I vs Stage IV Tumors:
ROC (left) and PR (right) curves

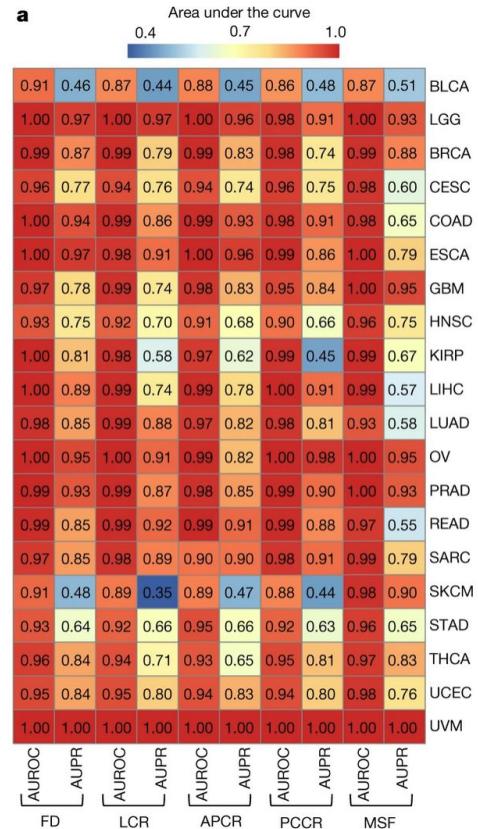


f

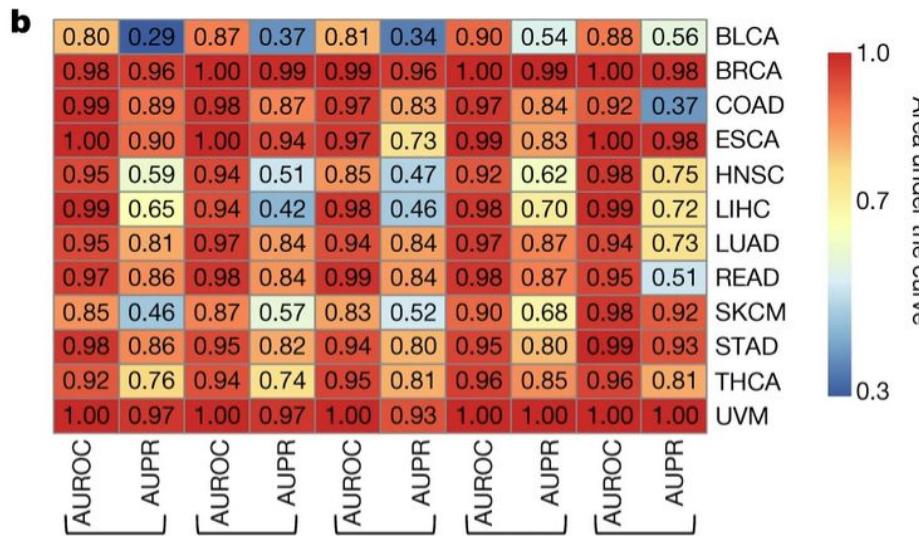
TCGA COAD Stage I vs Stage IV Tumors:
ROC (left) and PR (right) curves



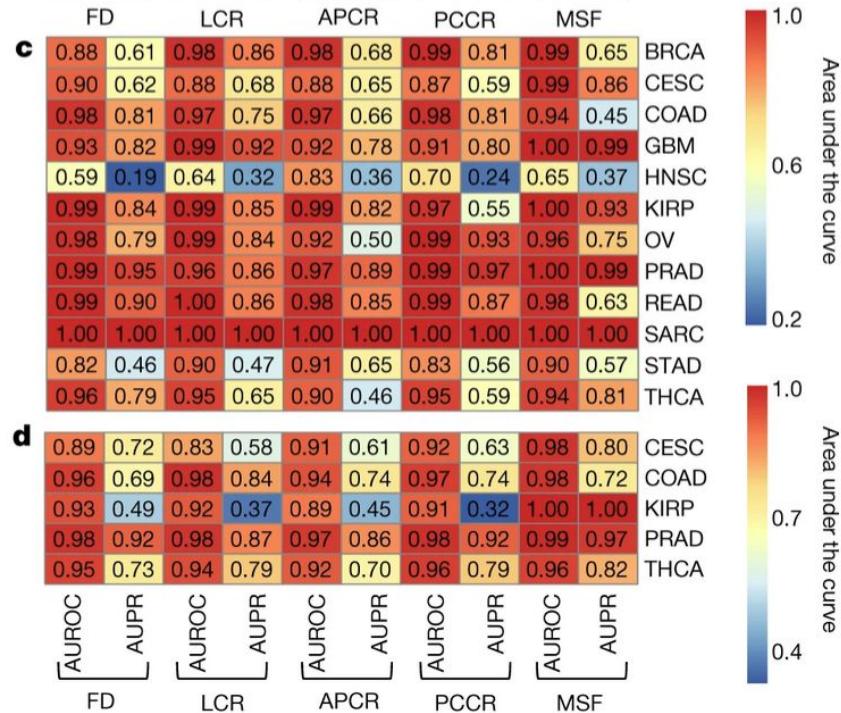
Also, tumor type works from blood!



Even for early stage cancers!



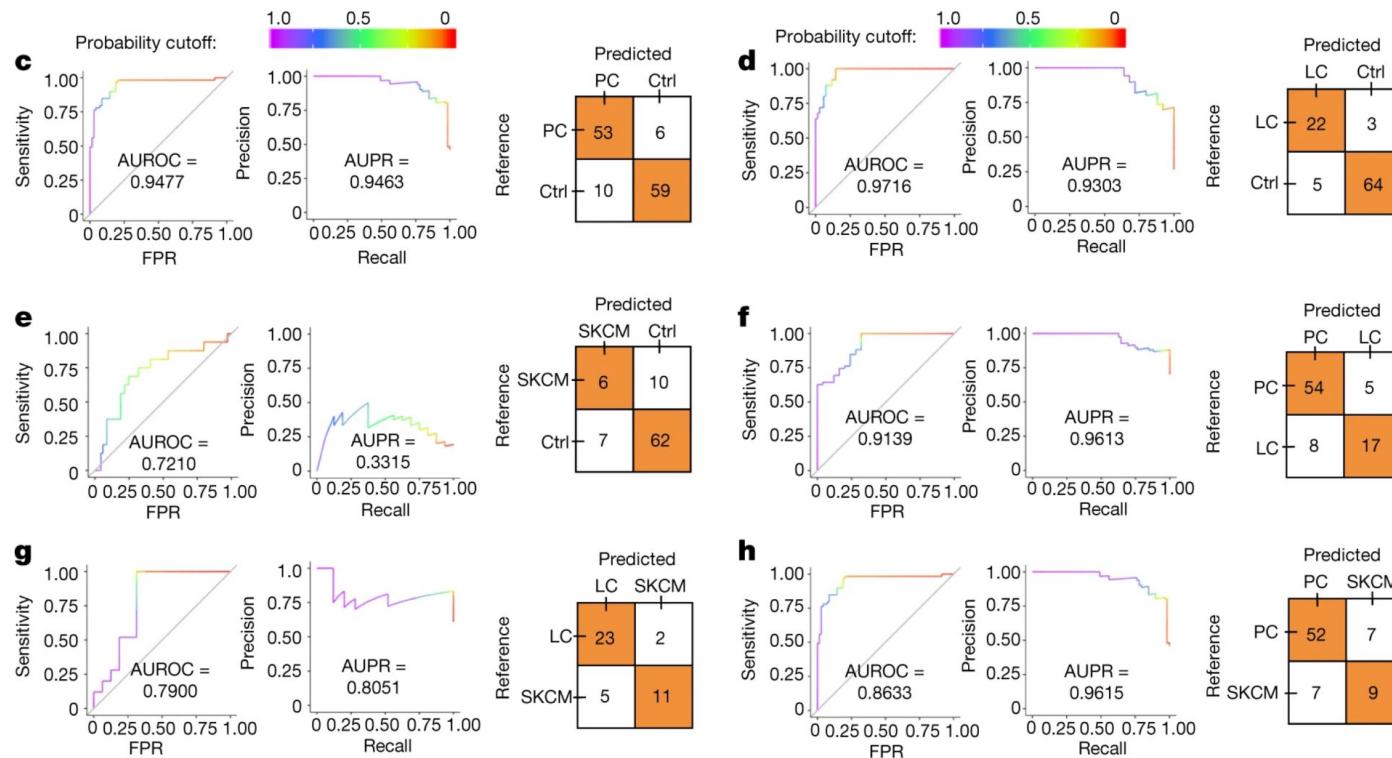
Even when tumor & ctDNA assays detected no mutations!



Validation study

Variable	Control	PC	LC	SKCM
Sample size (n)	69	59	25	16
Age (years) (mean ± s.d.)	45.00 ± 12.80	69.70 ± 9.63	69.50 ± 9.88	58.50 ± 13.00
Sex (% female)	24.6%	0%	68%	12.5%
Known condition(s)/ subtype(s) (n)	HIV-free (69)	HSPC (32) CRPC (27)	LUAD (15) LUSC (5) Sarcomatoid (1)	SKCM (16)
			Large cell (1)	
			NOS (3)	

Validation performance



 *Fin*