Microbiome analyses of blood and tissues suggest cancer diagnostic approach

2020.11.19

Becca

catalogue

Background (15 mins)

- tissue samples
- blood samples
- microbiome

Paper reading (25 mins)

Discussion

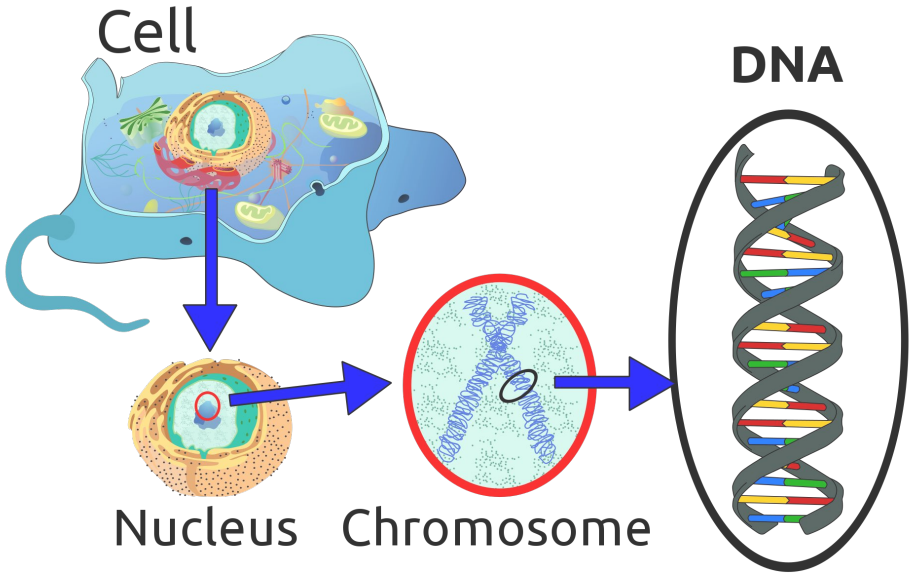
Article | Published: 11 March 2020

Microbiome analyses of blood and tissues suggest cancer diagnostic approach

Gregory D. Poore, Evguenia 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

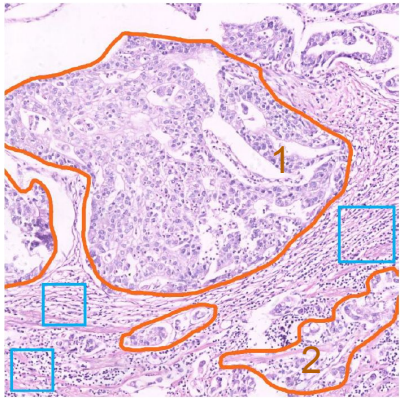
Nature **579**, 567–574(2020) | Cite this article

Microbiome analyses of blood and tissues suggest cancer diagnostic approach



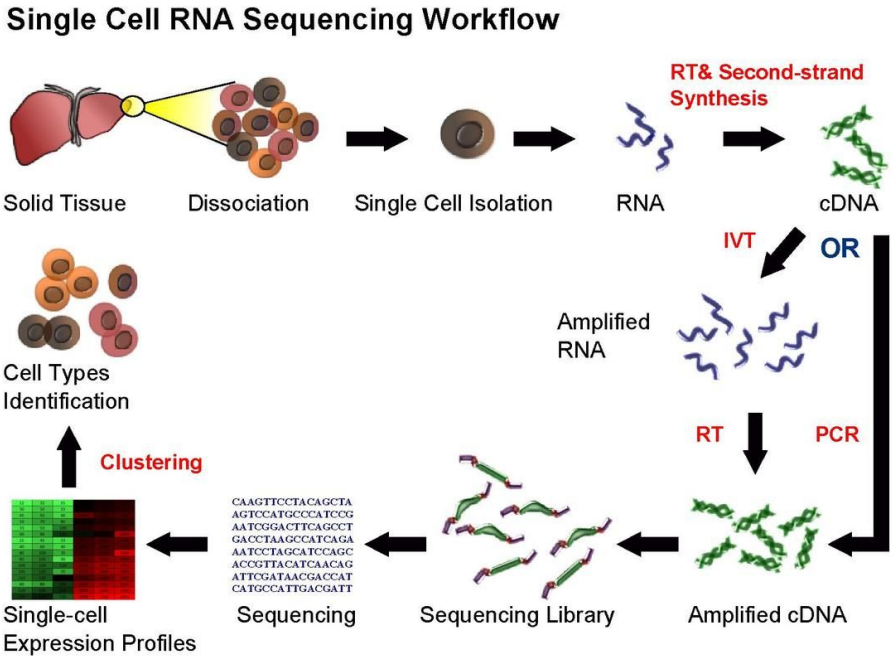
tissue sample

- large populations of mixed cells
- tumour tissue VS solid tissue normal (TCGA)
- heterogeneity



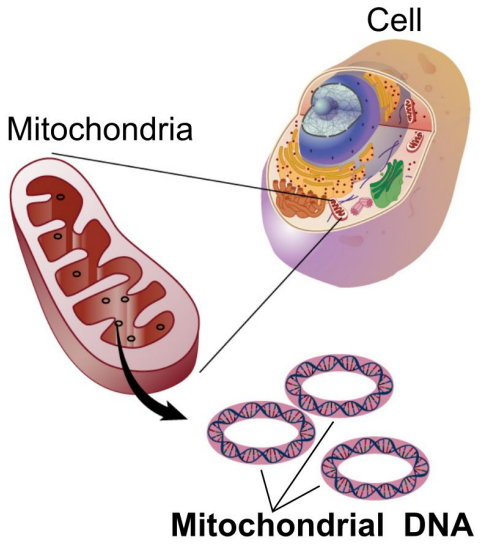
Extend Knowledge

single cell sequencing



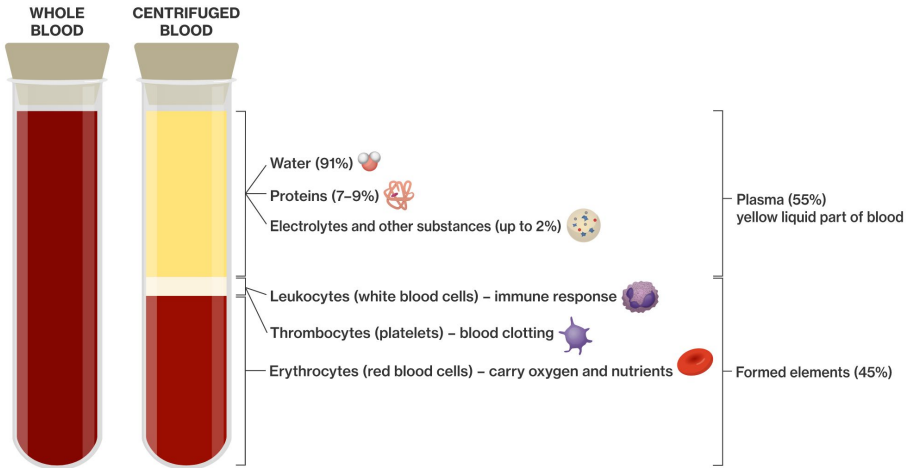
- Fluorescence-activated cell sorting (FACS) & microfluidics isolate cells
- At least 50 single cells to achieve this minimum CV value（3~8K cells）

Mitochondrial DNA

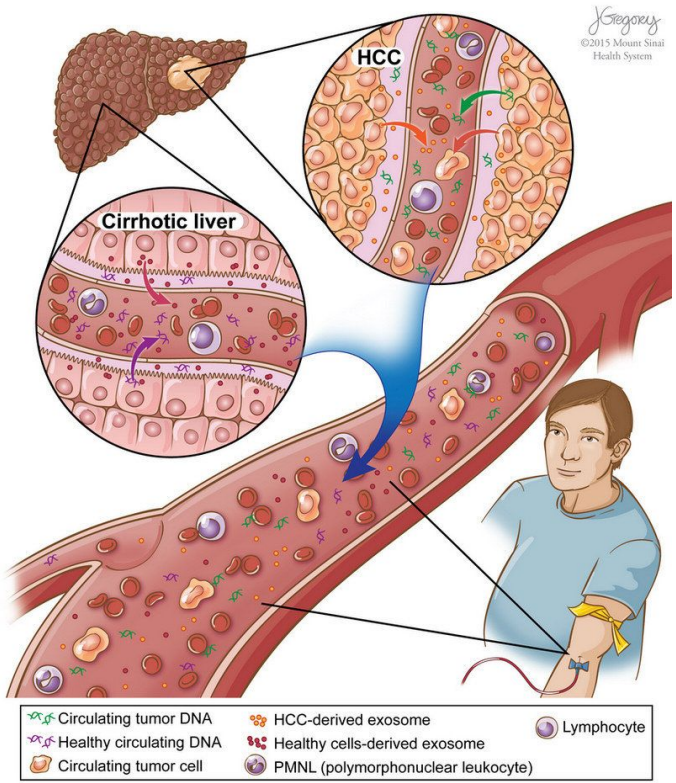


- In sexual reproduction, maternally inherited ,evolutionary analysis.
- human identification, disease diagnosis,relationship with aging and life spans.

Microbiome analyses of blood and tissues suggest cancer diagnostic approach



blood sample

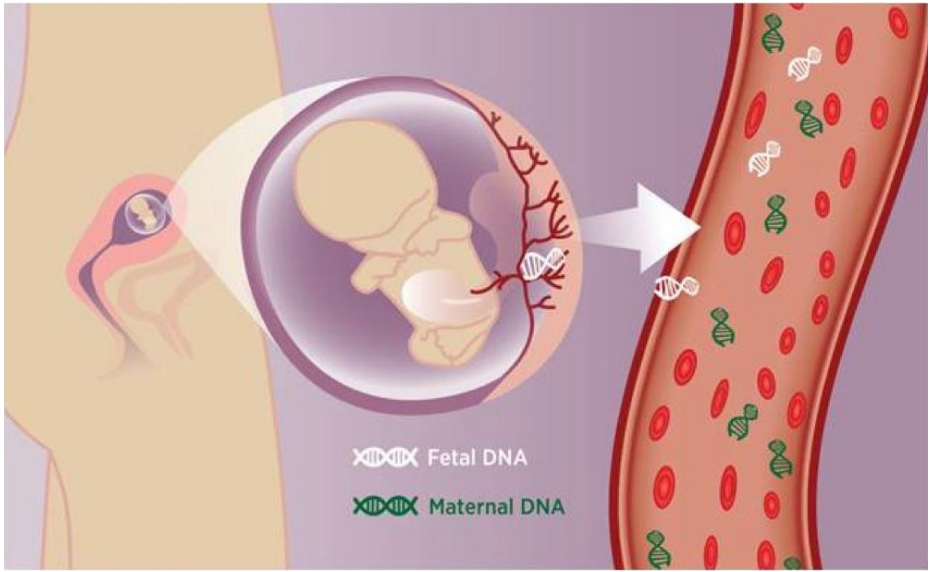


- Circulating free DNA (**cfDNA**): degraded DNA fragments released to the blood plasma.
- RNA DE / Somatic variant calling：blood samples VS cancer samples
- Liquid Biopsy: detect **ctDNA** , **CTCs**.
- Non-invasive Prenatal Testing (NIPT) : cffDNA (cell-free fetus DNA)

https://www.coursehero.com/sg/anatomy-and-physiology/blood-composition/

Extend Knowledge

cffDNA



- originates from placental trophoblasts,fragmented
- size difference (cffDNA ~200bp, shorter than maternal DNA; isolated by centrifuge)
- 11 - 13.4 % of the cfDNA in maternal blood samples
- detectable after 5-7 weeks gestation, till two hours after delivery.

DNA can be detected in blood samples and tissue samples

Microbiome analyses of blood and tissues suggest cancer diagnostic approach

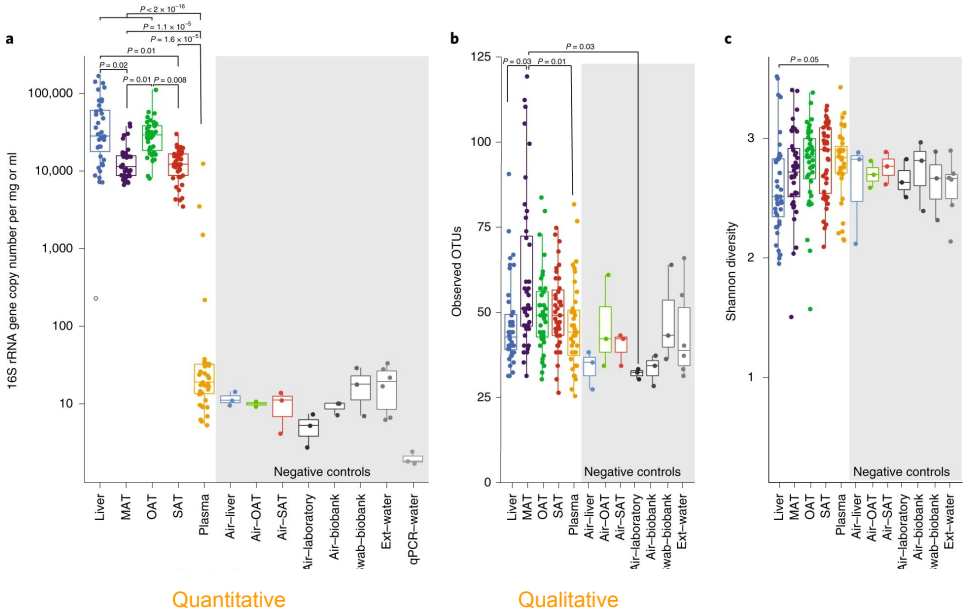
The human microbiome is a complex aggregate of the microbes residing at various sites in the human body and consisting of communities of a variety of microorganisms including Eukaryotes, Archaea, Bacteria, and the virus that reside in the different body habitat including the skin, the oral cavity, respiratory tract, gastrointestinal tract, urinary tract, reproductive tract etc. *(Sender et al. 2016; Shreiner et al. 2015)*

microbial population differ in early neonatal state, childhood and further changes throughout the lifespan depending upon the life style and diseased condition *(Johnson and Versalovic 2012)*

microbiome varies enormously at different sites within an organ system *(Dethlefsen et al. 2007)*

bacterial species population vary significantly among individuals

Bacterial distribution across body sites



cite from Anhê, F.F,2020
Negative controls were tested to control for environmental sample contamination at major steps in the analysis

Microbiome exist and vary among blood and tissue samples

Dekaboruah, E., Suryavanshi, M., Chettri, D. et al. Human microbiome: an academic update on human body site specific surveillance and its possible role. Arch Microbiol 202, 2147–2167 (2020)
Anhê, F.F., Jensen, B.A.H., Varin, T.V. et al. Type 2 diabetes influences bacterial tissue compartmentalisation in human obesity. Nat Metab 2, 233–242 (2020).

Abstract

Systematic characterization of the cancer microbiome provides the opportunity to develop techniques that exploit non-human, microorganism-derived molecules in the diagnosis of a major human disease. Following recent demonstrations that some types of cancer show substantial microbial contributions[1–10], we re-examined whole-genome and whole-transcriptome sequencing studies in The Cancer Genome Atlas[11] (TCGA) of 33 types of cancer from treatment-naive patients (a total of 18,116 samples) for microbial reads, and found unique microbial signatures in tissue and blood within and between most major types of cancer. These TCGA blood signatures remained predictive when applied to patients with stage Ia–IIc cancer and cancers lacking any genomic alterations currently measured on two commercial-grade cell-free tumour DNA platforms, despite the use of very stringent decontamination analyses that discarded up to 92.3% of total sequence data. In addition, we could discriminate among samples from healthy, cancer-free individuals ($n = 69$) and those from patients with multiple types of cancer (prostate, lung, and melanoma; 100 samples in total) solely using plasma-derived, cell-free microbial nucleic acids. This potential microbiome-based oncology diagnostic tool warrants further exploration.

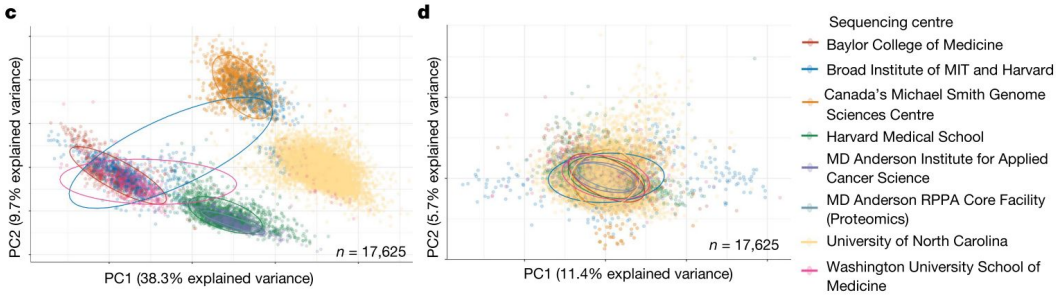
INPUT DATASET

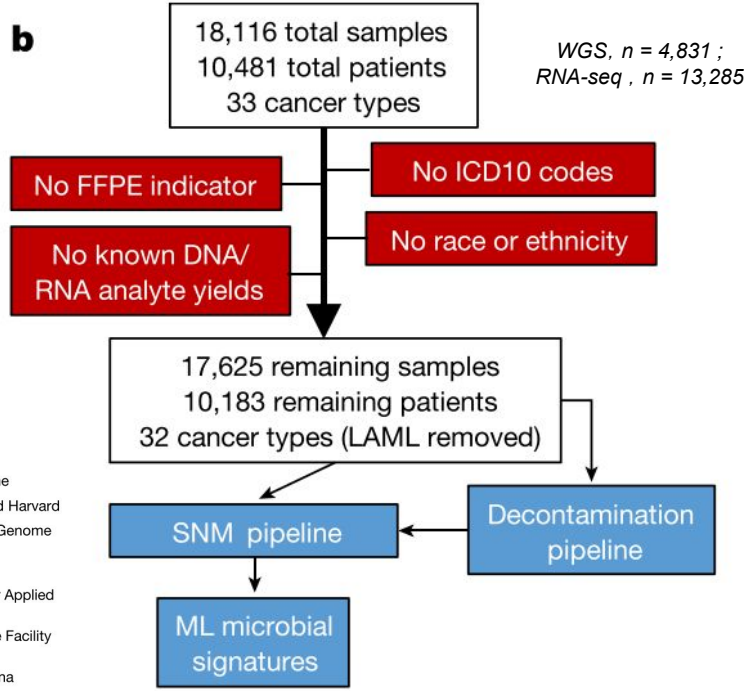
Analysis / Model training

OUTPUT / Contribution

Quality control & normalization

SNM Normalization:
R package limma, edgeR, and snm





Supervised normalization of microarrays Brigham H. Mecham, Peter S. Nelson, John D. Storey Bioinformatics. 2010 May 15; 26(10): 1308–1315. Published online 2010 Mar 31

Microbiome analyses of blood and tissues suggest cancer diagnostic approach

17,625 samples
10,183 patients ; 32 types of cancer

microbial genomes from tissue samples

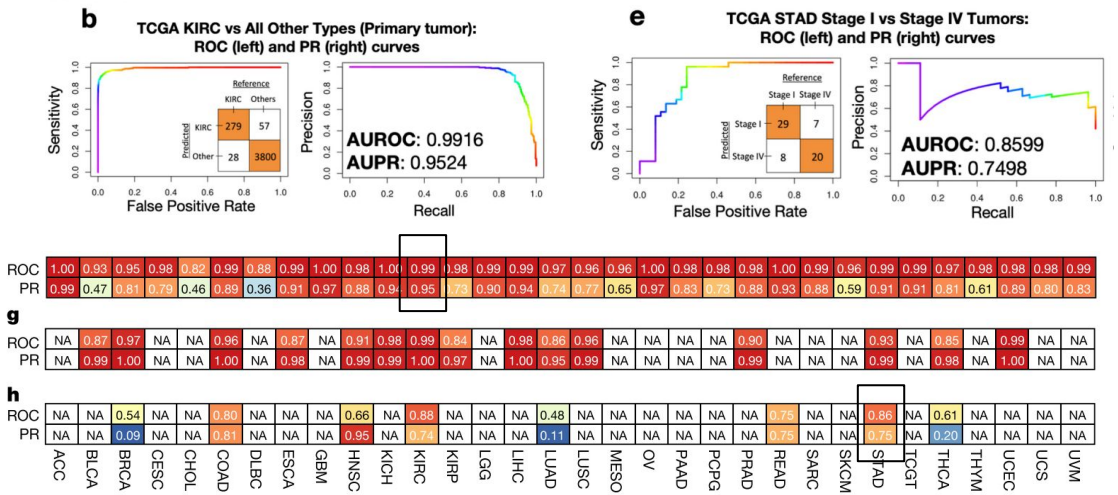
↓

Taxonomical counts (log-cpm)
[Kraken]*-positive, genus-level microbial reads

↓

Model training and performance#
stochastic gradient-boosting ML models
(i) 1 VS all other (n = 32 types of cancer)
(ii) Tumour VS normal (n = 15 types of cancer)
(iii) Tissue-based microbial models discriminate tumours stage (I and IV) (n = 8 types of cancer)

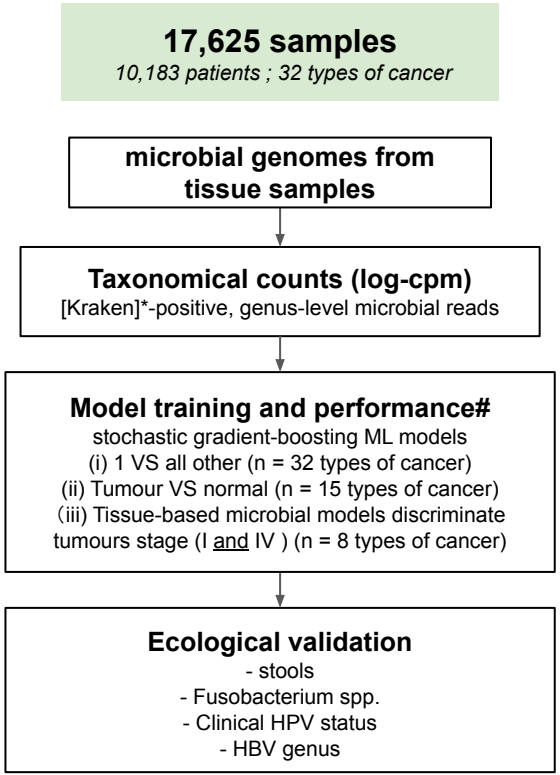
not perform well in other types of cancer and discriminate intermediate stages, suggest that microbial community structure dynamics may not correlate with cancer stages as defined by host tissue for all types of cancer



strong discrimination of microbial signatures —— microbial communities are unique to each cancer type

*kraken: matches short genomic substrings (k-mers) to taxa in a reference database
#AUROC: area under the receiver operating characteristic curve; AUPR: area under the precision–recall curve

Biological relevance of microorganism profiles

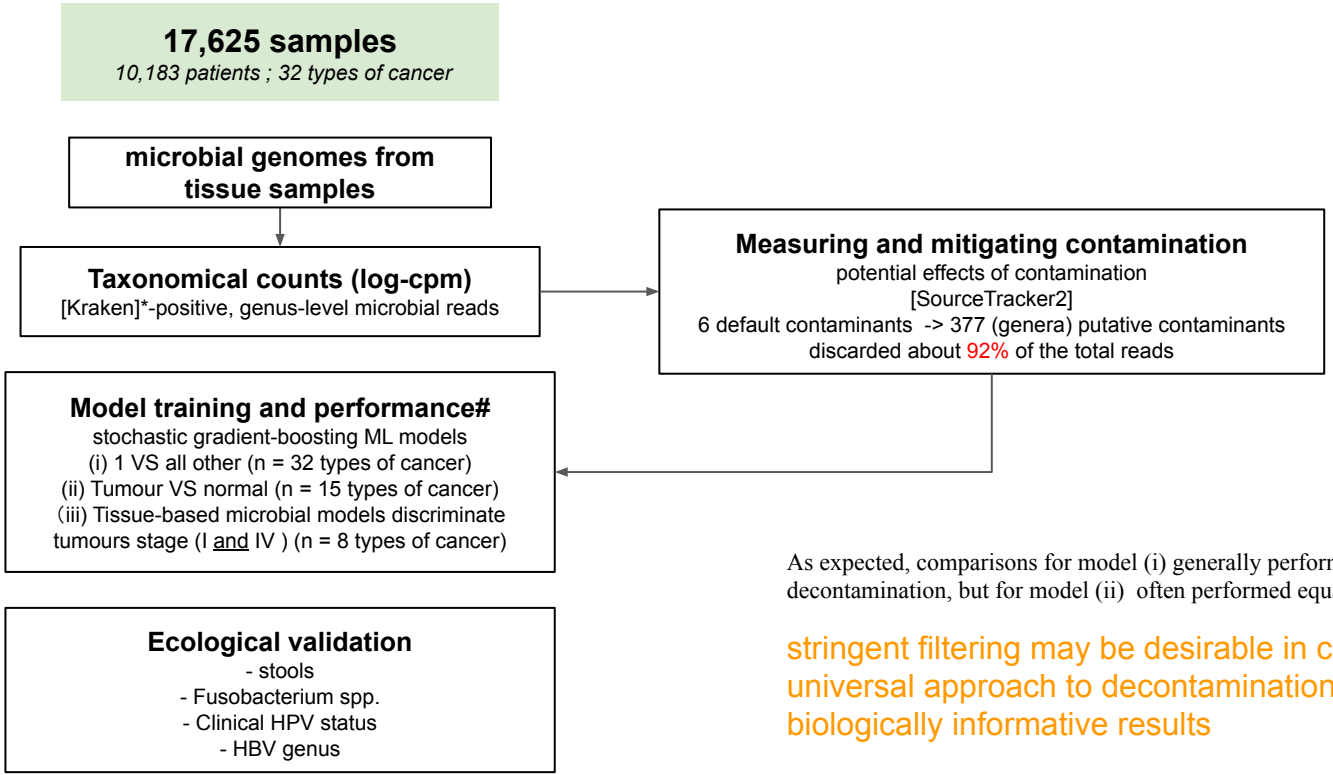


Namely, cancers with known microbial 'drivers' or 'commensals' provided initial evidence that the models were ecologically relevant; for example, **Alphapapillomavirus genus** was the most important feature for identifying **CESC tumours**; for **COAD tumours**, the **Faecalibacterium genus**; for **LIHC tumours**, the **Orthohepadnavirus genus** was the second most important feature.(after the **hepatotoxic Microcystis genus**)

Analysis result consistent with information about feature importance provided by trained models. Prove models reliability

interactive website:
exploration of normalized microbial abundances in TCGA cancers and major sample types
http://cancermicrobiome.ucsd.edu/cancermicrobiome_DataBrowser

Measuring and mitigating contamination



As expected, comparisons for model (i) generally performed less well with stringent decontamination, but for model (ii) often performed equally well or better.

stringent filtering may be desirable in certain comparisons, but a universal approach to decontamination may preclude biologically informative results

Microbiome analyses of blood and tissues suggest cancer diagnostic approach

17,625 samples
10,183 patients ; 32 types of cancer

microbial genomes from tissue samples

Taxonomical counts (log-cpm)
[Kraken]*-positive, genus-level microbial reads

Model training and performance#
stochastic gradient-boosting ML models
(i) 1 VS all other (n = 32 types of cancer)
(ii) Tumour VS normal (n = 15 types of cancer)
(iii) Tissue-based microbial models discriminate tumours stage (I and IV) (n = 8 types of cancer)

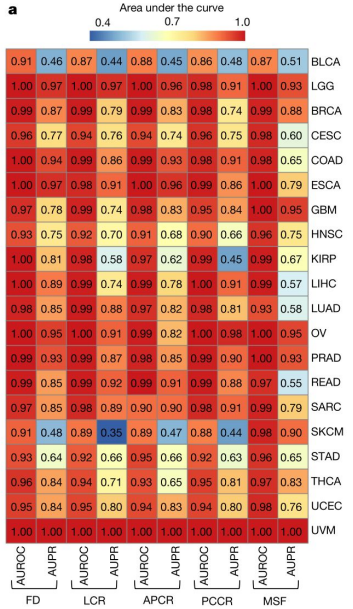
microbial genomes from WGS blood samples
ad hoc analyses (EBV & HPV)
viromes n= 4,433 bacteriomes n=1,880

blood-based microbial DNA (mbDNA)
【gold-standard microbiology protocols】
case - patients with prostate, lung, or skin cancers (n = 100)
control - healthy, cancer- and HIV-free participants (n = 69)

Taxonomical counts (log-cpm)
genomic-alignment-filtered Kraken data

Model training and performance#
stochastic gradient-boosting ML models
(i) 1 VS all other (n = 20 types of cancer)

mbDNA in blood can use for cancer discrimination



(i) model performances predicting one cancer type versus all others using blood mbDNA

FD, full data; LCR, likely contaminants removed by sequencing centre; APCR, all putative contaminants removed by sequencing centre; PCCR, plate–centre contaminants removed; MSF, most stringent filtering by sequencing centre.

Microbiome analyses of blood and tissues suggest cancer diagnostic approach

* Cell-free DNA extracted from plasma samples - whole metagenomic sequencing - Kraken

17,625 samples
10,183 patients ; 32 types of cancer

microbial genomes from tissue samples

Taxonomical counts (log-cpm)
[Kraken]*-positive, genus-level microbial reads

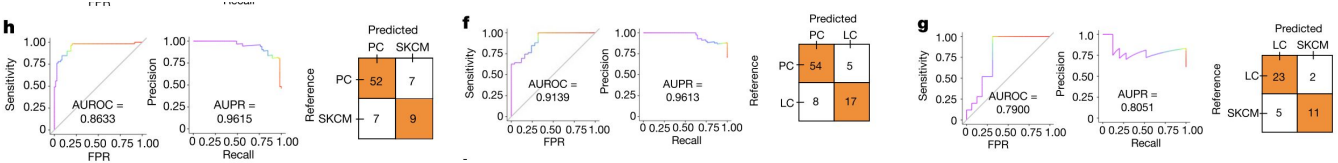
Model training and performance#
stochastic gradient-boosting ML models
(i) 1 VS all other (n = 32 types of cancer)
(ii) Tumour VS normal (n = 15 types of cancer)
(iii) Tissue-based microbial models discriminate tumours stage (I and IV) (n = 8 types of cancer)

microbial genomes from WGS blood samples
ad hoc analyses (EBV & HPV)
viromes n= 4,433 ;bacteriomes n=1,880 ;

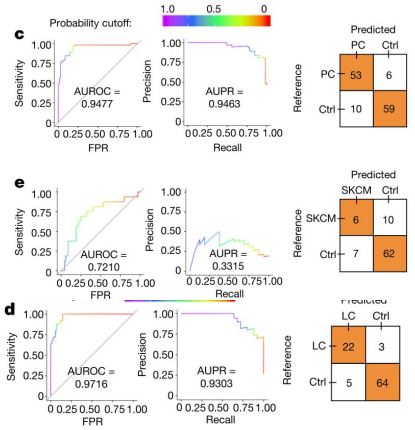
blood-based microbial DNA (mbDNA)
gold-standard microbiology protocols*
case - patients with prostate, lung, or skin cancers (n = 100)
control - healthy, cancer- and HIV-free participants (n = 69)

Taxonomical counts (log-cpm)
genomic-alignment-filtered Kraken data

Model training and performance#
stochastic gradient-boosting ML models
(i) 1 VS all other (n = 3 types of cancer)
(ii) Tumour VS normal (n = 3 types of cancer)

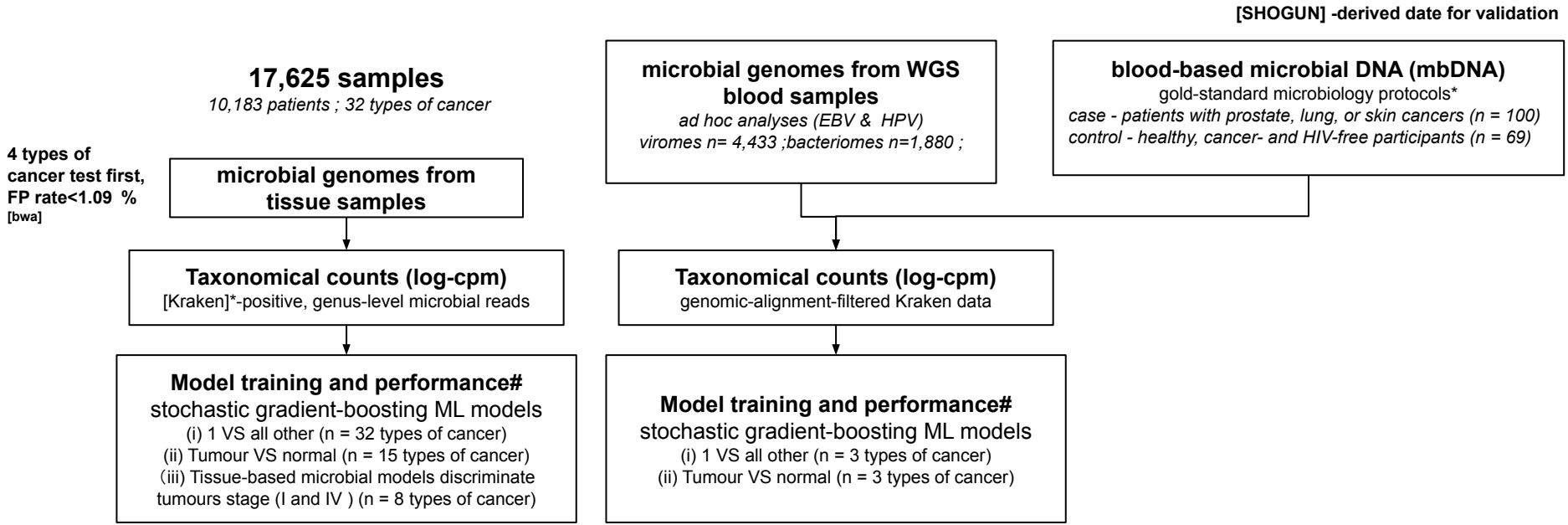


(i) discriminate type of cancer using mbDNA



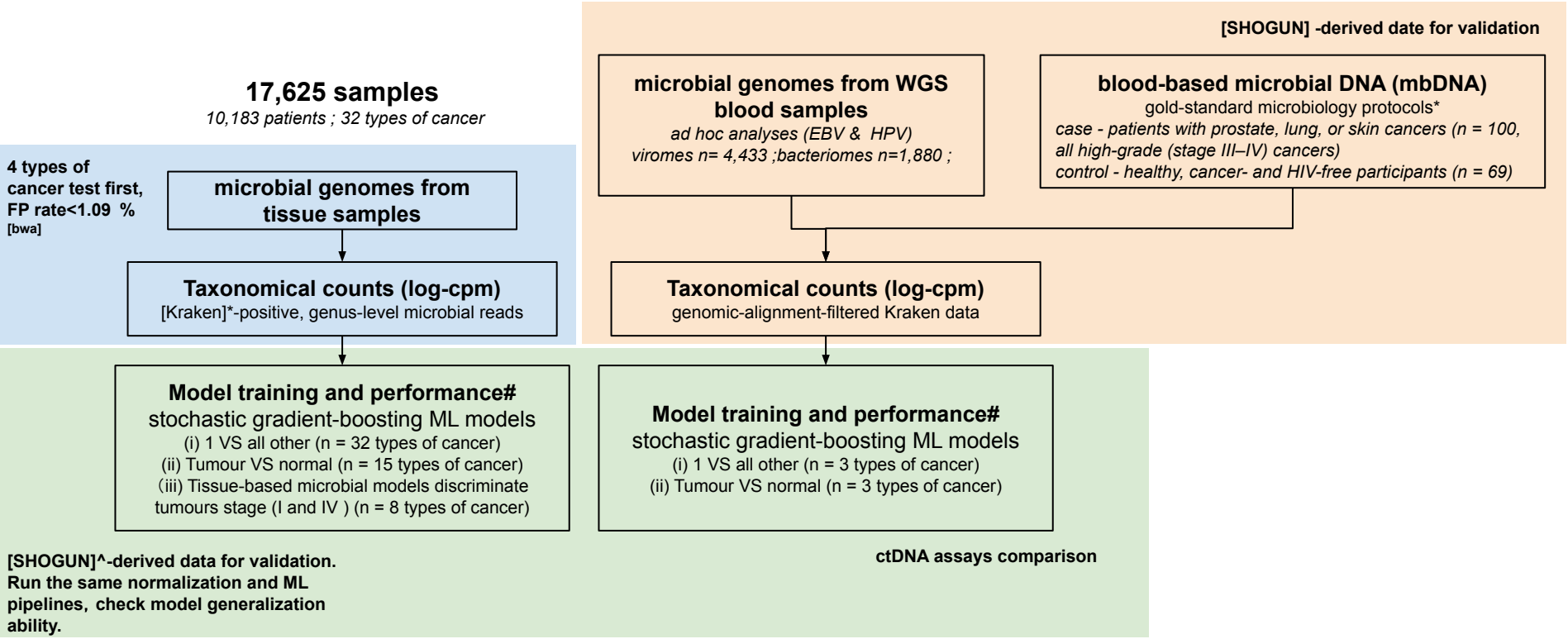
(ii) discriminate cancer and healthy controls

Microbiome analyses of blood and tissues suggest cancer diagnostic approach



*kraken: matches short genomic substrings (k-mers) to taxa in a reference database
^SHOGUN:an alignment-based microbial taxonomic pipeline

Microbiome analyses of blood and tissues suggest cancer diagnostic approach



Highlight

Verify carefully in most steps

Facing limitation/shortcoming honestly

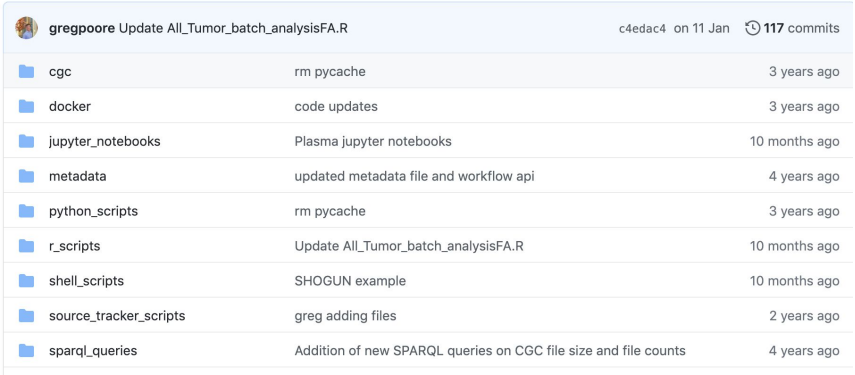
A lot of meticulous work

Clear figures

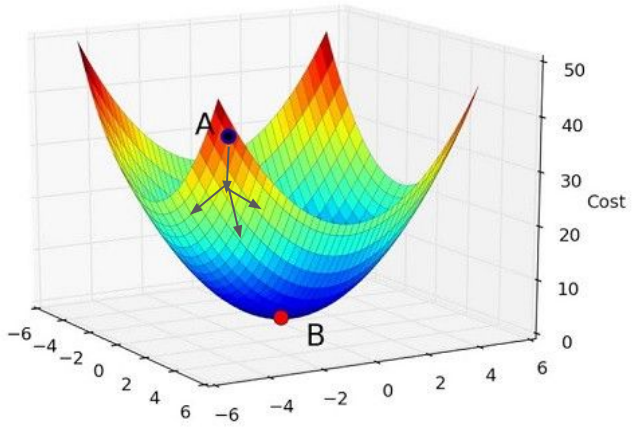
Data availability: ftp.microbio.me/pub/can

Code availability: github.com/biocore/tcga

mia)[8,10]. There is likely to be a continuum of ideal decontamination, as the effect of decontamination on model performance varied across types of cancer, but our filtering was limited by (i) not having access to the primary specimens, (ii) genus-level taxonomic resolution, and (iii) not knowing which non-TCGA samples were concurrently processed.



gradient boosting machine algorithm (GBM)



Start at point A, how we can arrived at the lowest point (B) ?

following the most steepest negative gradient direction

calculate GBM

training set

	x	y
sample1	1	2
sample2	2	4

model
y = θx, θ is the parameter

using model
when we know x=3, y=3*θ = 3*2=6

prediction:
regression: y=6
known when y>3, group = A
classification: y=6 > 3, A