

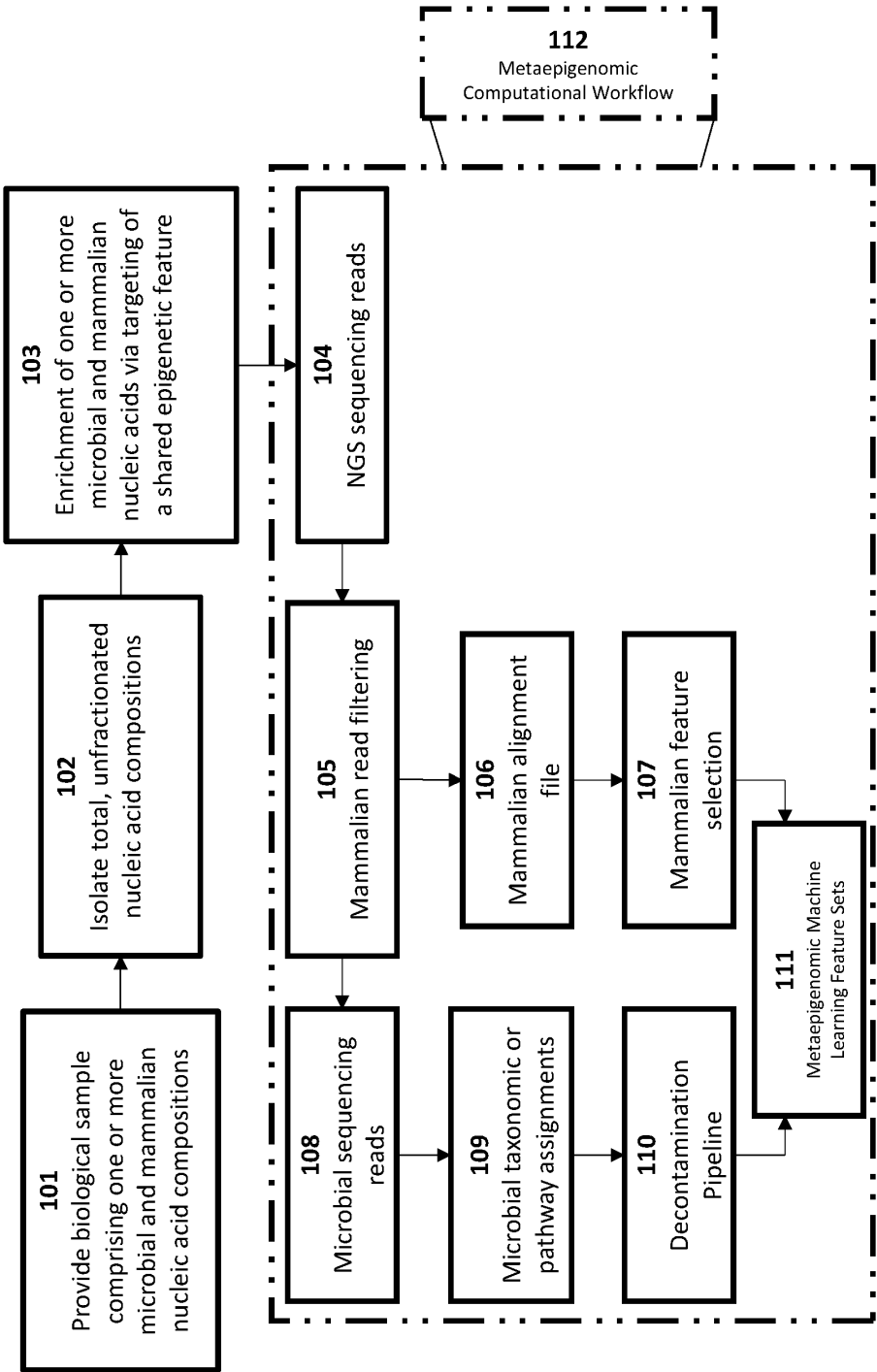


FIG. 1A

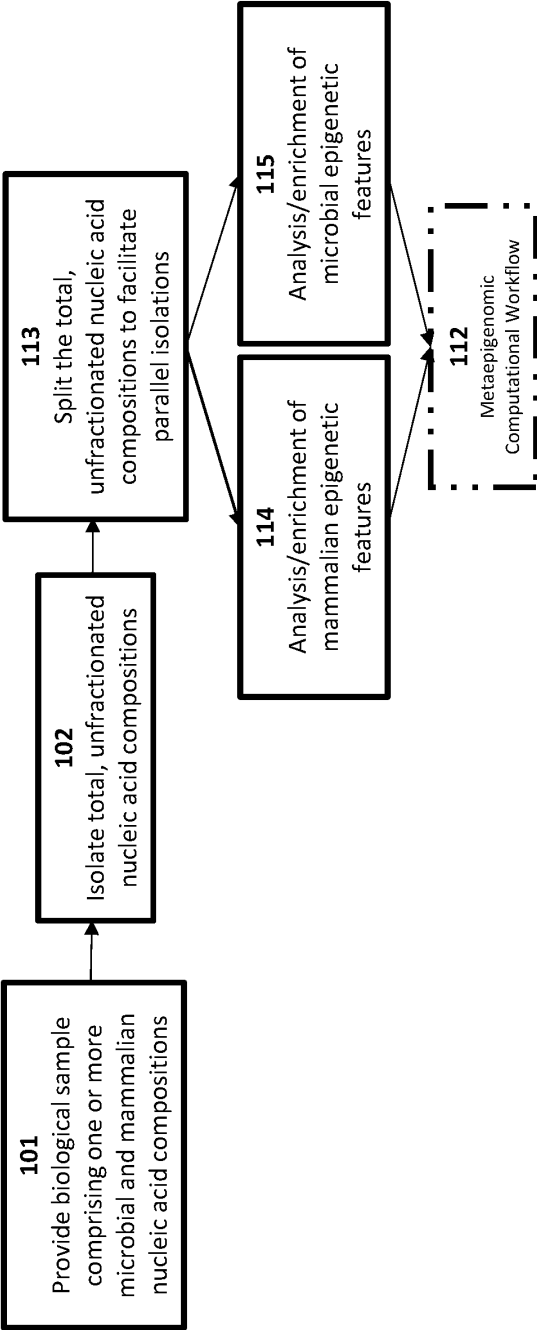


FIG. 1B

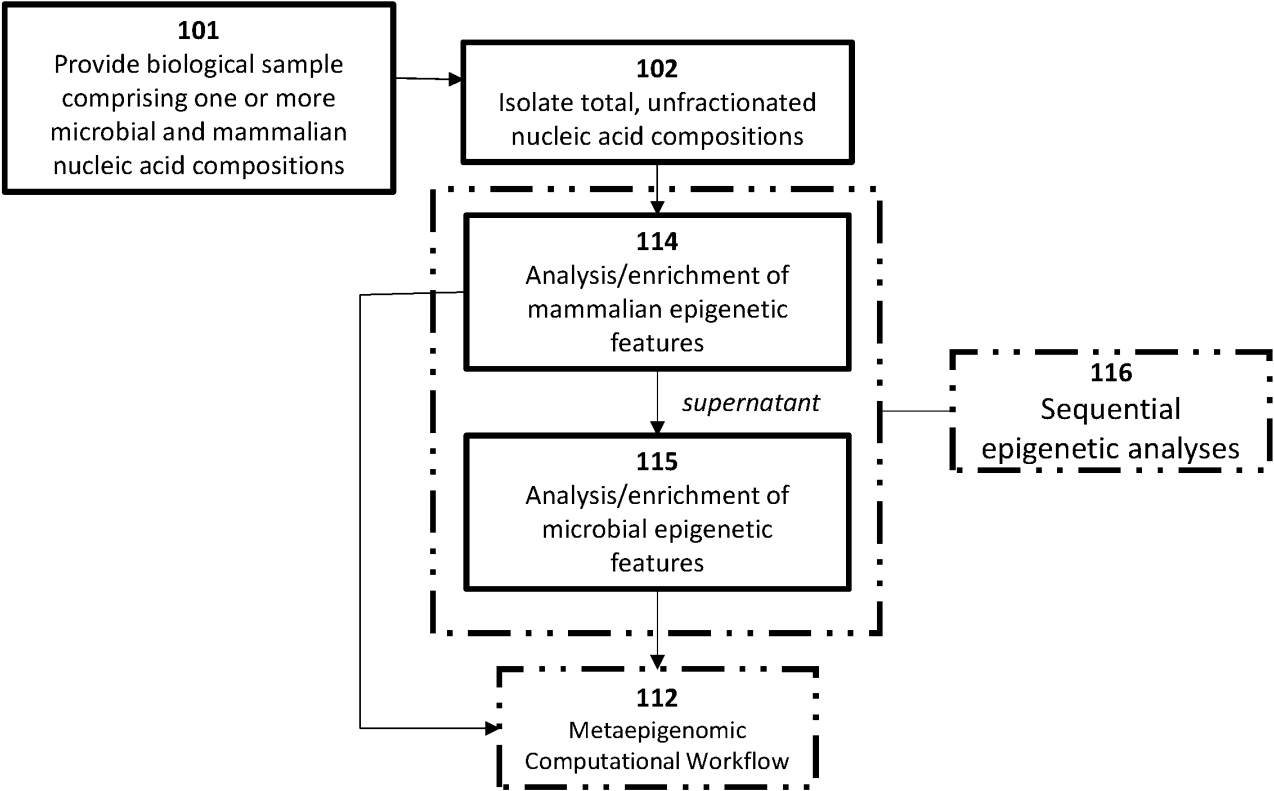


FIG. 1C

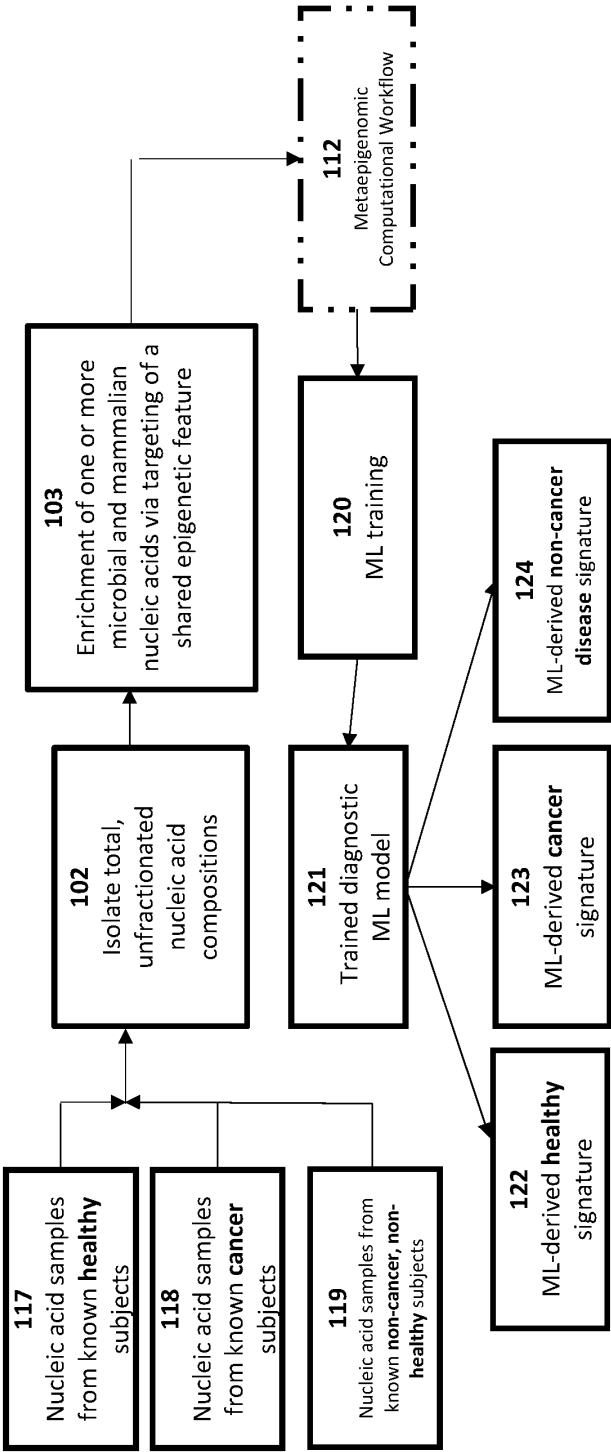


FIG. 1D

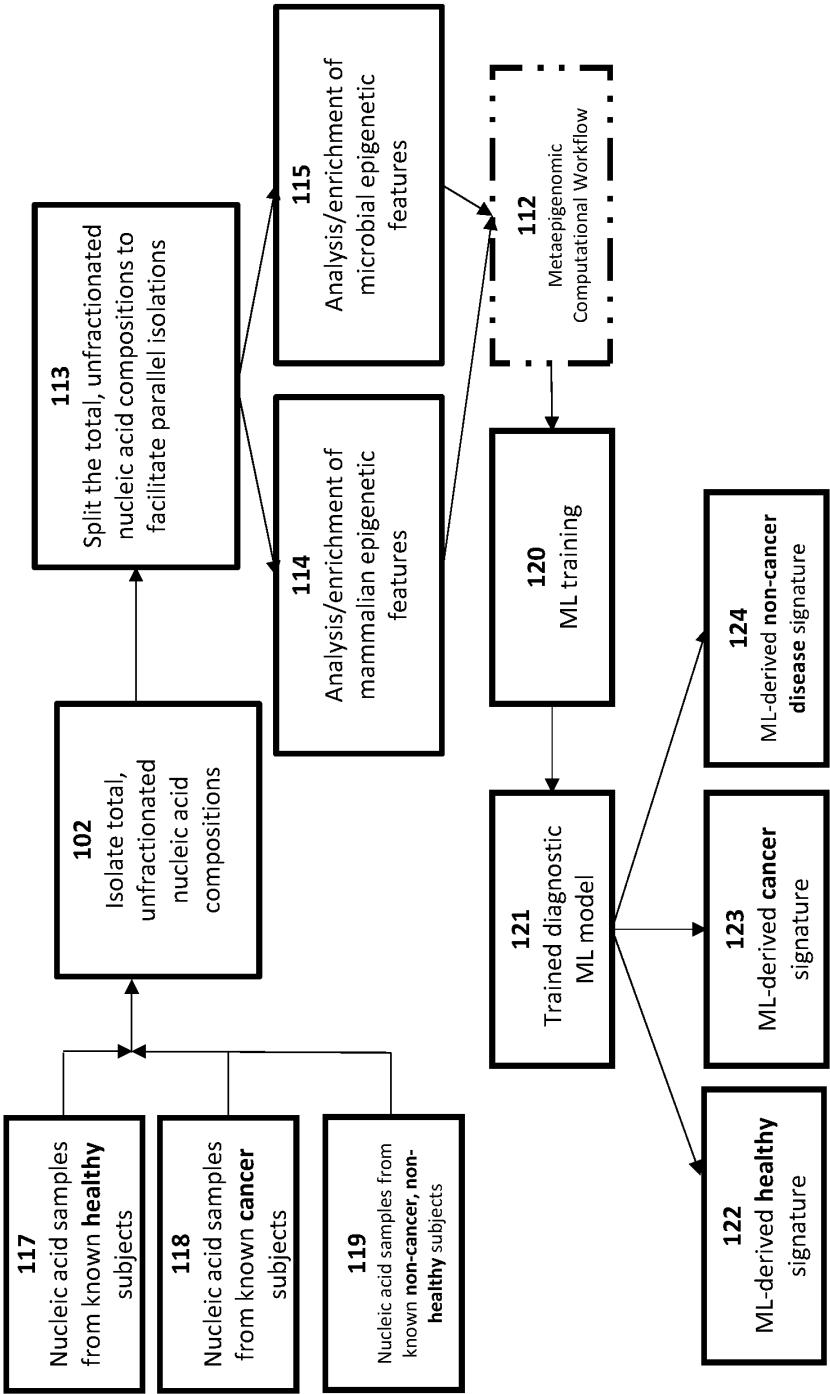


FIG. 1E

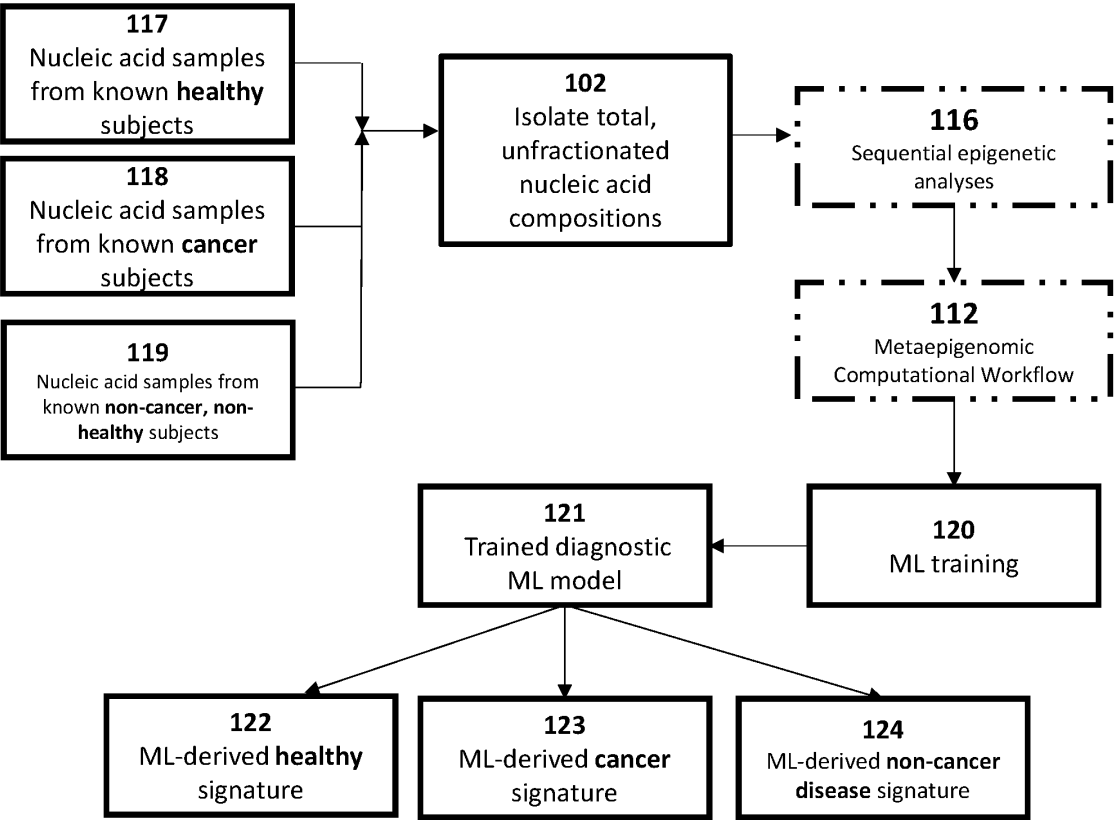
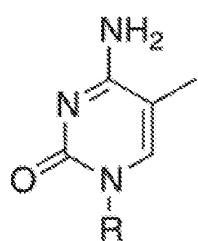


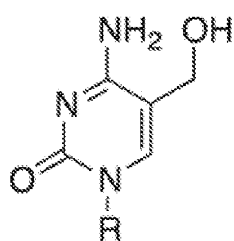
FIG. 1F

7/20



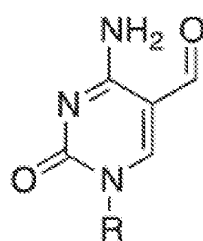
5mC

FIG. 2A



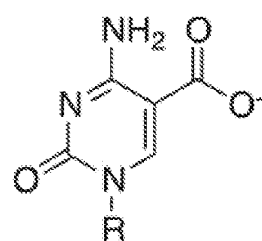
5hmC

FIG. 2B



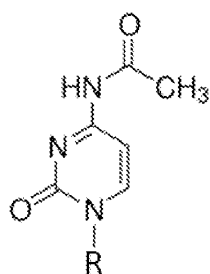
5fC

FIG. 2C



5caC

FIG. 2D



N4AcC

FIG. 2E

8/20

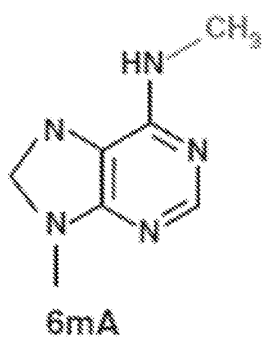


FIG. 3A

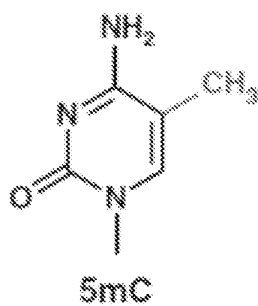


FIG. 3B

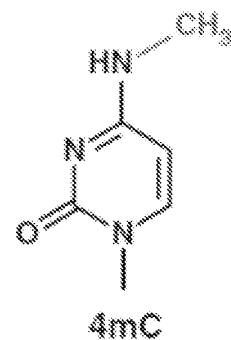


FIG. 3C

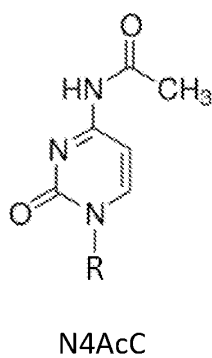


FIG. 3D

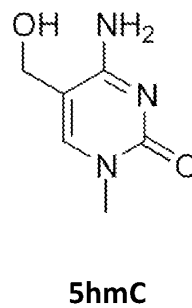


FIG. 3E

9/20

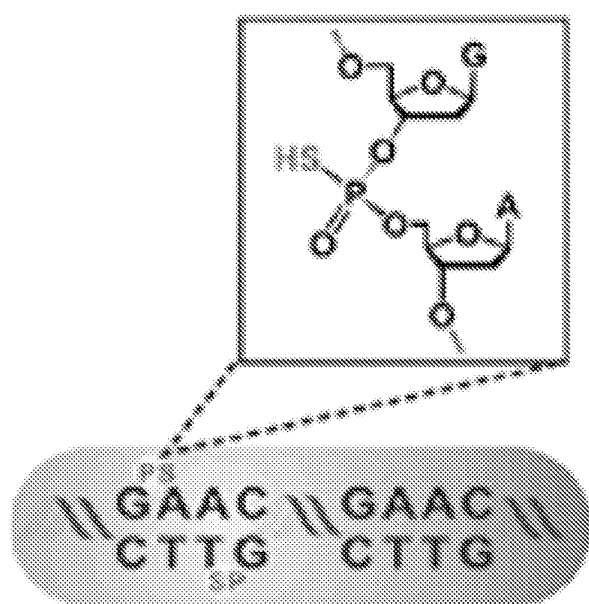


FIG. 4

Dataset

<https://www.nature.com/articles/cr2017106>

5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages

Chen, Wen, ^{1,15,*} Scott, Yu, ^{2,3,*} Li, He, ^{4,5} Amanda, Wheeler, ⁶ Yu, Chen, ⁷ Yan, Zhao, ⁸ Hu, Li, ^{9,7} Jie, Xiong, ⁸ Xue, Chen, ⁹ Jie, Xu, ⁹ Zhen, Chen, ⁹ Bao, Chen, ² Zhi, Han, ¹⁰ Robert, S. Jeffrey, ⁶ Mei, Shi, ¹¹ Samuel, Li, ^{4,5} Yanna, Li, ¹¹ Xue, Yi, ² Jie, Chen, ¹⁰ Dan, Xie, ^{2,9,11,*} and Stephen, B. Quake, ^{1,12,*}

Sequencing Method:

5hmC sequencing of cfDNA based on selective chemical labeling and pulldown of 5hmC-containing DNA fragments

input: unenriched input cfDNA

5hmC: 5hmC enriched cfDNA

Sample info

5hmC samples

- Lung Cancer: 15
- Pancreatic Cancer: 7
- Healthy: 10

Input:

- Lung Cancer: 15
- Pancreatic Cancer: 7
- Healthy: 4

FIG. 5A

Hits to rep206 database

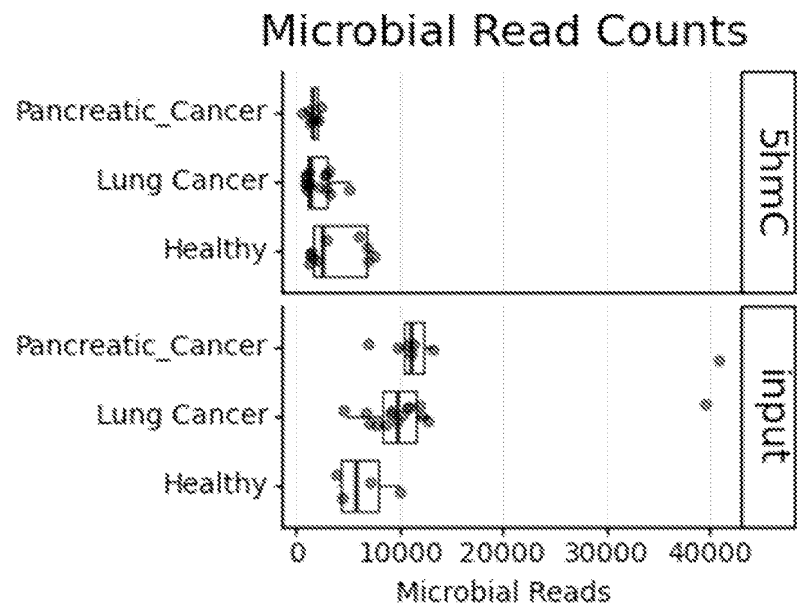


FIG. 5B

12/20

Table S5 Clinical information for pancreatic cancer samples.

sample ID	TNM	stage	metastasis to	gender	age
pancreatic9	T3N0M1	IV	liver	male	76
pancreatic15	T1N0M0	IA	-	male	64
pancreatic22	T4N1M0	III	-	female	71
pancreatic27	T4N1M1	IV	abdominal wall, omentum	male	55
pancreatic68	T3N0M1	IV	liver	male	63
pancreatic69	T3N0M0	IIA	-	male	66
pancreatic75	T3N0M0	IIA	-	male	54

FIG. 5C

Pancreatic Cancer vs Healthy cfDNA

Machine Learning info
Data: rep206 bacteria + fungi + viruses
Model: randomForest
Method: LOO-cross-validation

Sample info	
<u>5hmC</u>	<u>input</u>
7 pancreatic cancer	7 pancreatic cancer
VS	VS
10 Healthy individuals	4 Healthy individuals

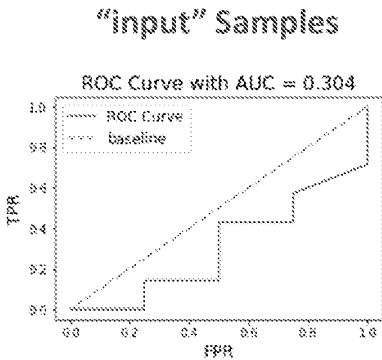
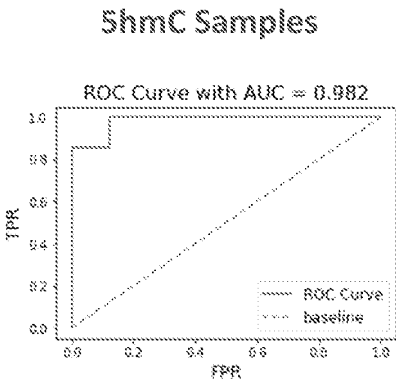


FIG. 5D

Lung Cancer Samples – Clinical Details

Table S3 Clinical information for lung cancer samples.

sample ID	category	TNM	stage	gender	age
lung395	non-metastatic lung cancer	T4N2Mx	III	female	62
lung419	non-metastatic lung cancer	T1N2M0G2	IIIa	female	53
lung492	non-metastatic lung cancer	T2N0M0	I	male	55
lung493	non-metastatic lung cancer	T1N3M0	IV	female	66
lung496	non-metastatic lung cancer	T3N1M0	IIIa	male	68
lung512	non-metastatic lung cancer	-	-	female	67
lung513	non-metastatic lung cancer	T2N1M0	I-II	male	47
lung514	non-metastatic lung cancer	T2N0M0	I-II	female	57
lung515	non-metastatic lung cancer	cT3N1M0	IIIa	male	52
lung293	metastatic lung cancer	cT4N3M1a	IV	female	52
lung323	metastatic lung cancer	TxN2M1	IV	female	68
lung324	metastatic lung cancer	TxNxM1	IV	male	56
lung417 §	metastatic lung cancer	-	-	male	62
lung418	metastatic lung cancer	TxN3Mx	IIIb-IV	male	59
lung517	metastatic lung cancer	cT4N2M1b	IV	male	68

All are non-small cell lung cancer samples unless otherwise noted.

§ Small cell lung cancer.

FIG. 5E

Lung Cancer vs Healthy cfDNA

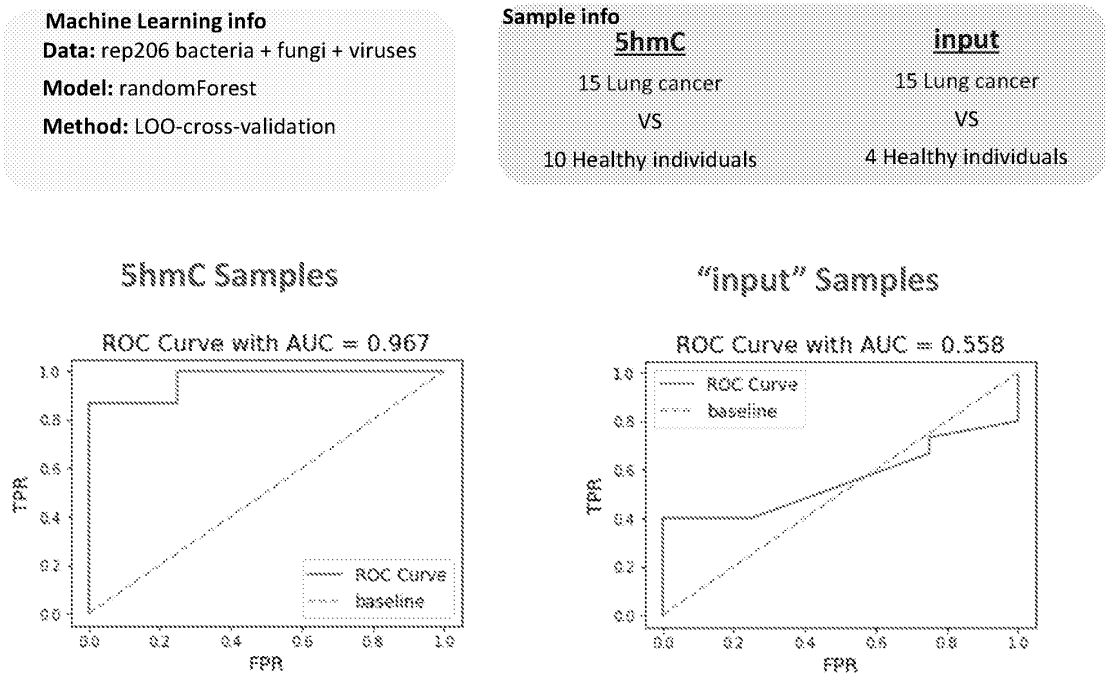


FIG. 5F

Dataset

<https://www.nature.com/articles/cr2017121>



5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers

Yanmei Li^{1,2}, Xu Zhang^{2,3}, Xinyu Lu^{3,4,5}, Lei Yan^{5,6}, Yuesun Song⁴, Zhongqiang Luo⁵, Jun Zhang⁵, Jin Ni⁵, Yanwei Zhou⁷, Qiannan Xu⁷, Yaping Wang⁵, Yuansheng Fang⁵, Shulin Yu⁷, Jun Hong⁸, Jianqiang Shi⁵, Hantao Hao⁵, Fan Luo⁵, Luchun Hua⁵, Peng Wang⁷, Xiaoping Qian⁵, Liang Guan^{10,11}, Jianhuan Xie¹¹, Ming Cui⁵, Tingting Zhao⁵, Guan-Liao⁵, Menghui Dai⁵, Zhen Lu⁵, Ge Chen⁵, Katherine Rachel¹², Sarvani Adhikari¹², Guilong Jia^{11,13}, Marc S. Blotnick¹², Xiaoxiao Zhang^{10,11}, Yuesi Zhao⁵, Wei Zhang^{14,5}, Chuan He^{5,10,7} and Jin Lu^{1,15,5}

- 5hmC profiling of cell free DNA and tissue

Blood

	Cancer	Benign
colon	76	49
liver	25	
pancreas	34	
stomach	61	22
thyroid	46	

Tissue

	Tumor	Adjacent
colon	44	44
stomach	52	52

healthy blood

FIG. 6A

Healthy vs Cancer Types

Sample info

Machine Learning info

Data: rep206

Model: randomForest

Method: 20x repetitions 70:30
train-test split

242 cancer cfDNA

•76 Colon **VS**

•61 Stomach

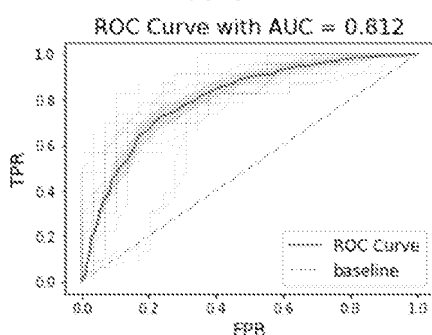
•46 Thyroid

•34 Pancreas

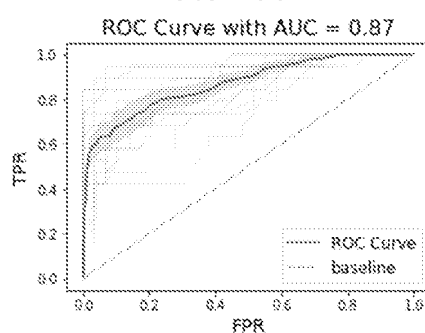
•25 Liver

96 Healthy cfDNA

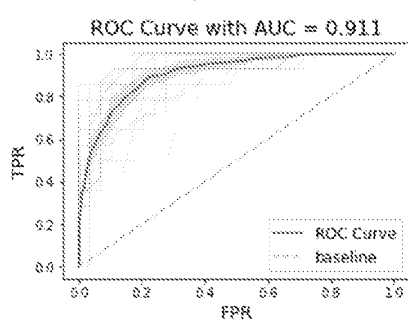
Colon



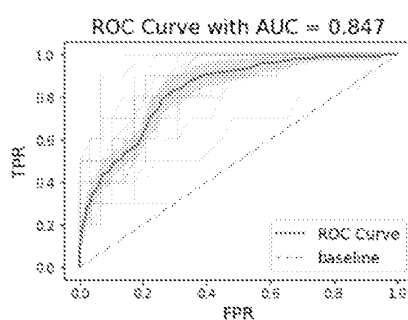
Stomach



Thyroid



Pancreas



Liver

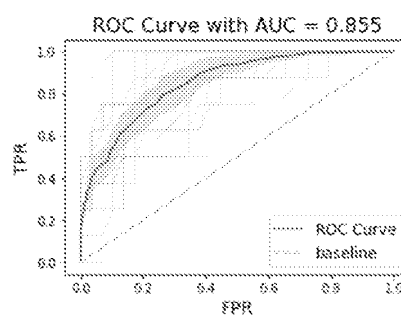


FIG. 6B

Cancer vs Benign tumors

Machine Learning info

Data: rep206

Model: randomForest

Method: 20x repetitions 70:30
train-test split

Sample info

139 Cancer cfDNA 78 Benign tumor cfDNA

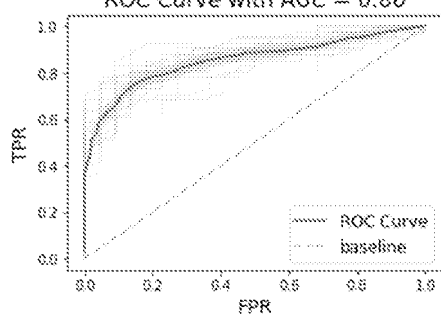
•76 Colon **VS** •49 Colon
•61 Stomach •22 Stomach

Colon + Stomach cancer

vs

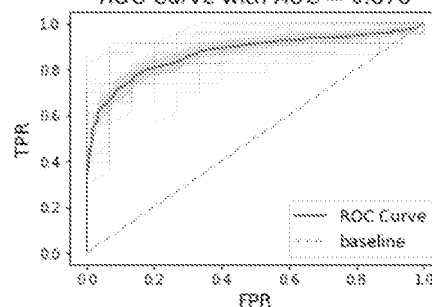
colon + stomach benign

ROC Curve with AUC = 0.86



Colon cancer vs Benign colon

ROC Curve with AUC = 0.876



Stomach cancer vs Benign stomach

ROC Curve with AUC = 0.92

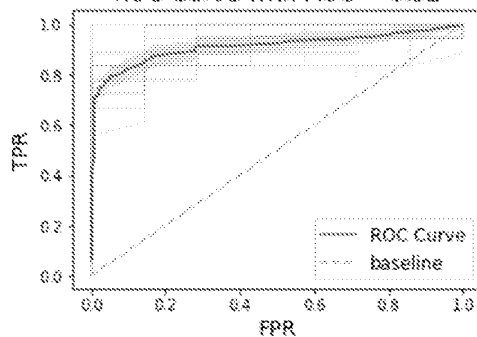


FIG. 6C

Cancer vs Benign blood by Kingdom

Machine Learning info

Data: rep206
Model: randomForest
Method: 20x repetitions 70:30 train-test split

Sample info

139 Cancer cfDNA	VS	78 Benign tumor cfDNA
•76 Colon		•49 Colon
•61 Stomach		•22 Stomach

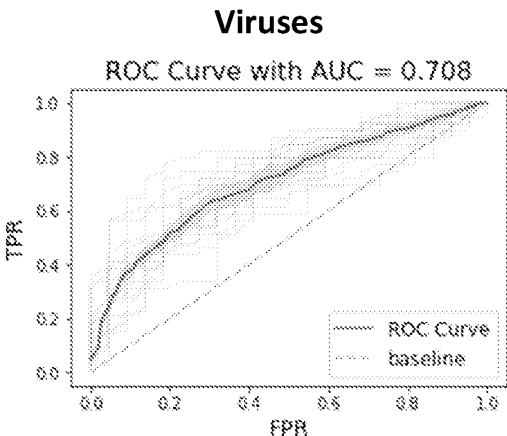
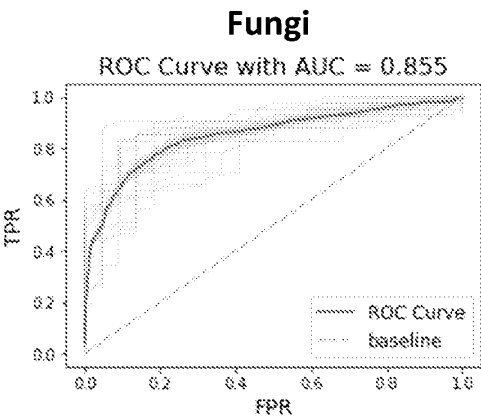
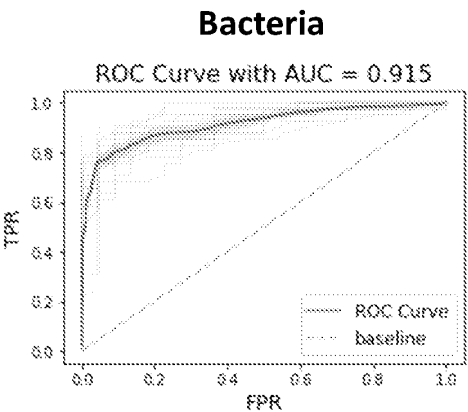


FIG. 6D

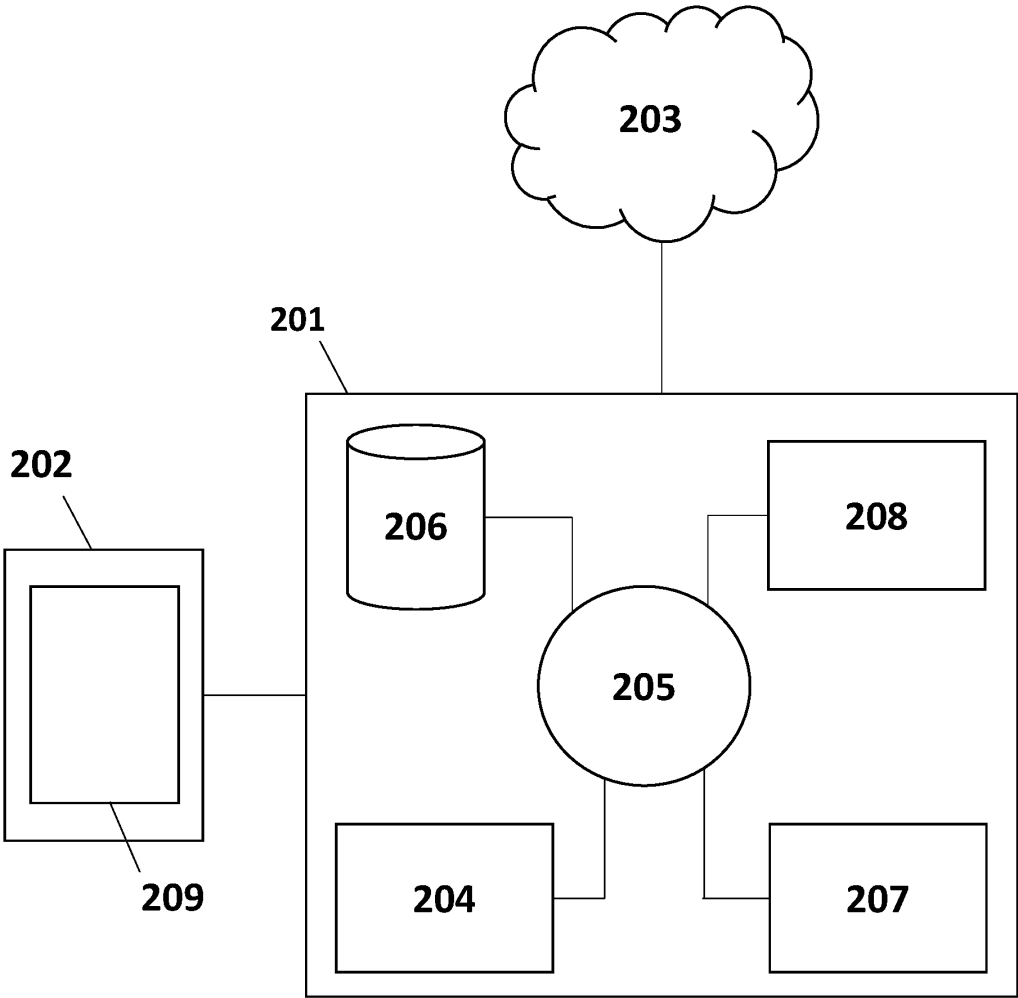


FIG. 7