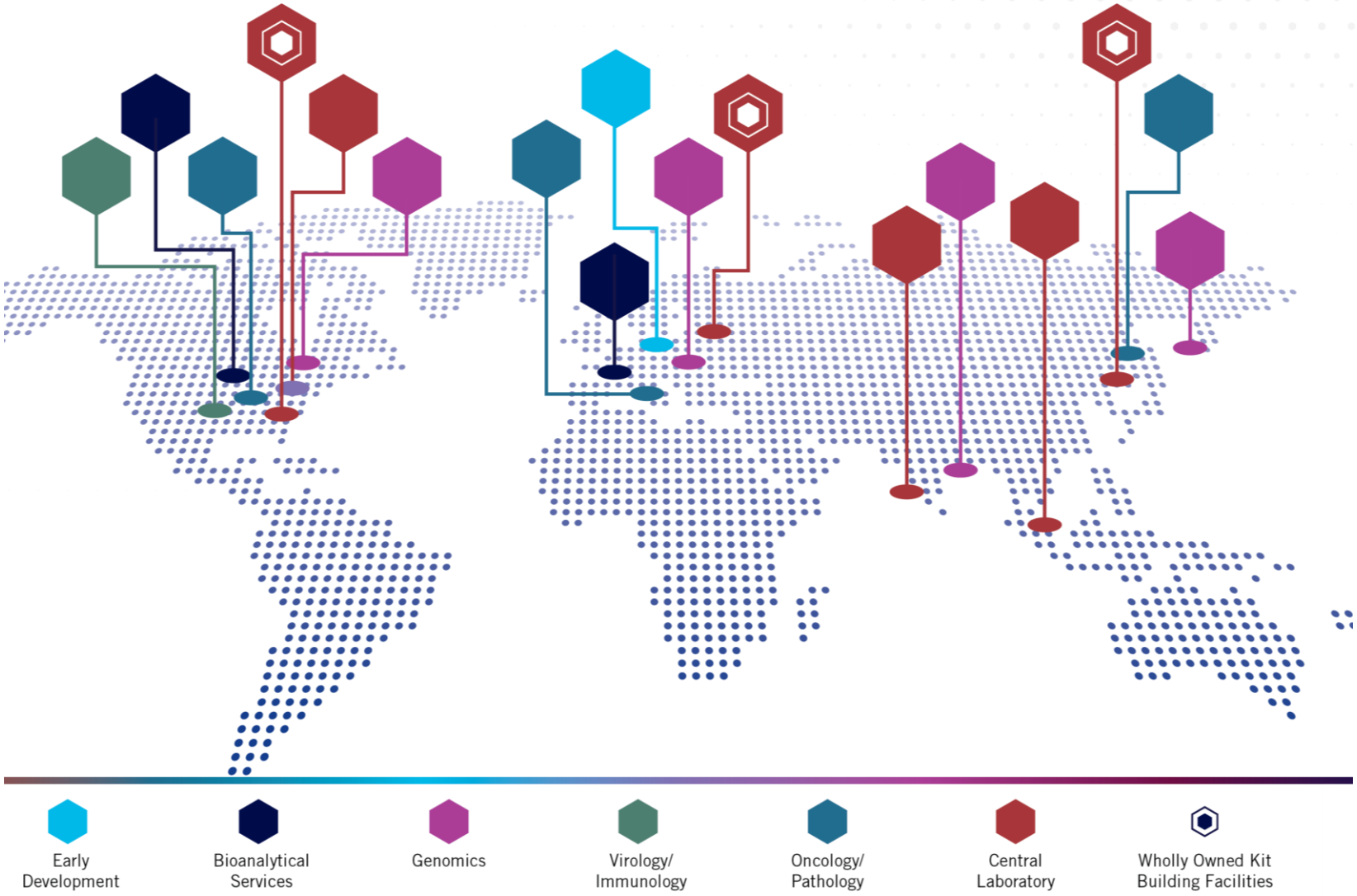
Explore the World of End-to-End Integrated Laboratory Services

eurofins | BioPharma Services

Clinical Testing Locations

eurofins | BioPharma Services



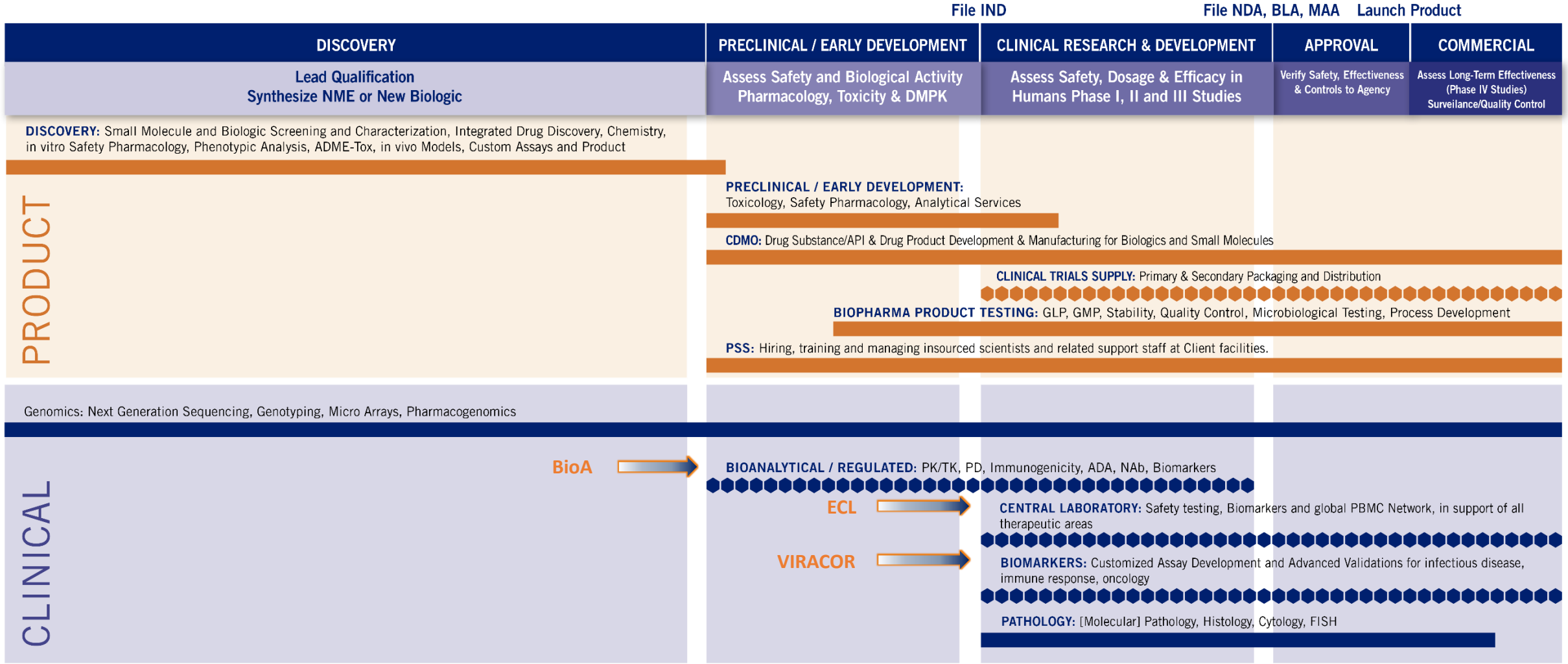
Eurofins Scientific

eurofins | BioPharma Services

- FOUNDED IN 1987 WITH 4 EMPLOYEES
- 61,000 STAFF IN 940 LABORATORIES ACROSS 59 COUNTRIES
- EURO 6.7 BILLION IN ANNUAL REVENUE IN 2021
- OVER 200,000 VALIDATED ANALYTICAL METHODS
- 450,000,000 ASSAYS PERFORMED ANNUALLY
- OVER 40 MILLION COVID-19 PCR TESTS CARRIED OUT SINCE START OF THE PANDEMIC

eurofins | BioPharma Services

End-to-End Testing Solution



Eurofins Scientific

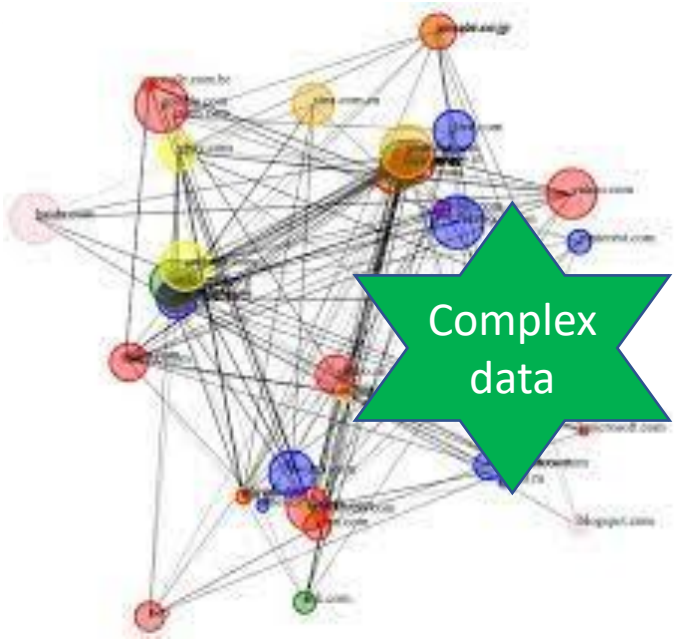
eurofins | BioPharma Services

Bioinformatics/AI to mine the public-data & boost the Lab of the Future

Rohita Sinha, PhD
Director, CoE for BI & AI, Eurofins, US

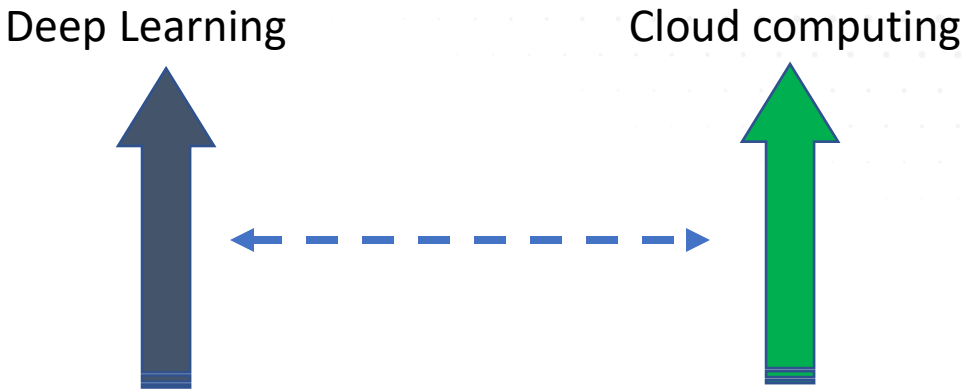
Data, more & more data points !!

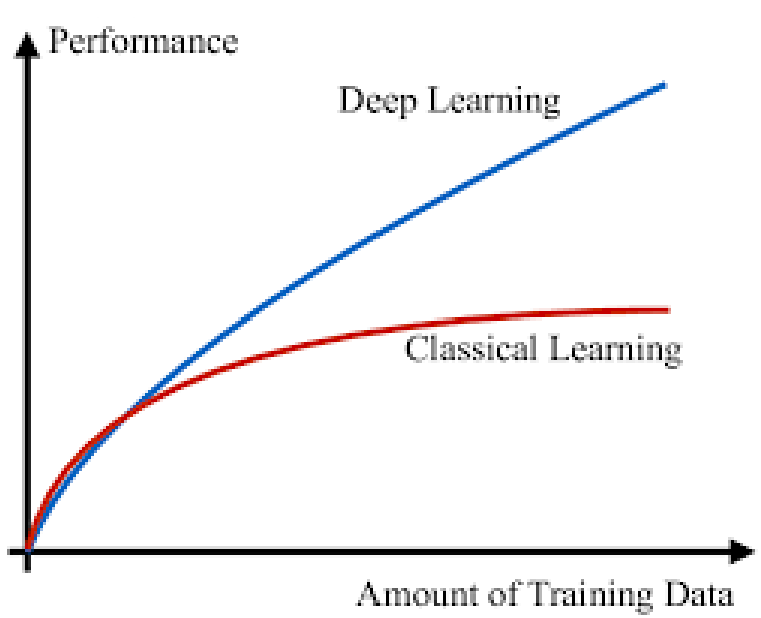
eurofins | BioPharma Services

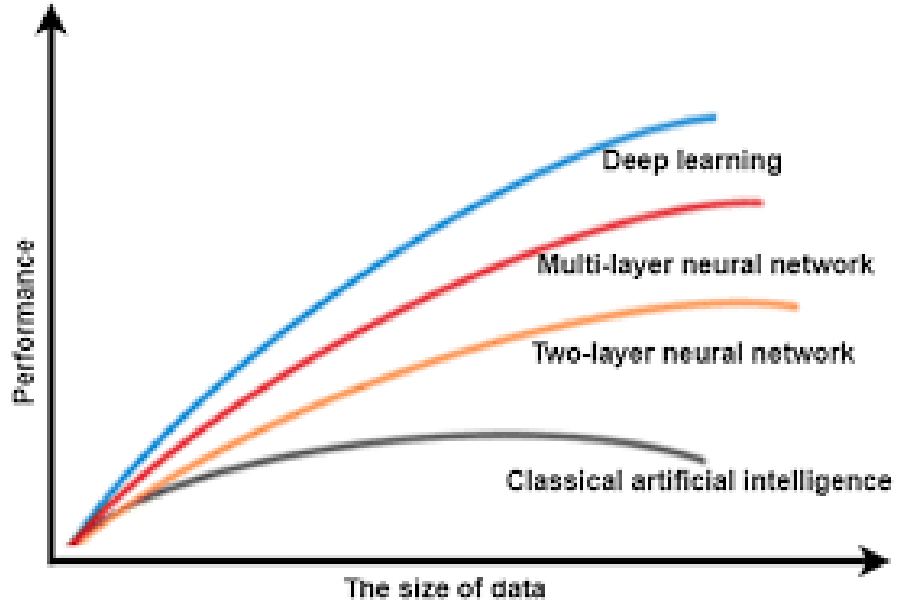


Why do we need so many data points?

eurofins | BioPharma Services







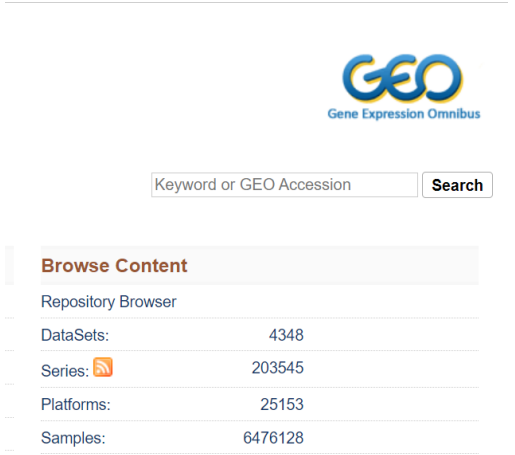
Public databases are a great source of data ?

eurofins | BioPharma Services

We (healthcare industry) need to either generate or collaborate with hospitals/other-centers to get enough data to validate our products.

For model-development: we can certainly source the data from the public-databases

Public databases offered by NIH (NCBI, GEO, PDB, TCGA, PubChem) and other groups such as GISAID (a global repository of Covid-2 genome sequences)

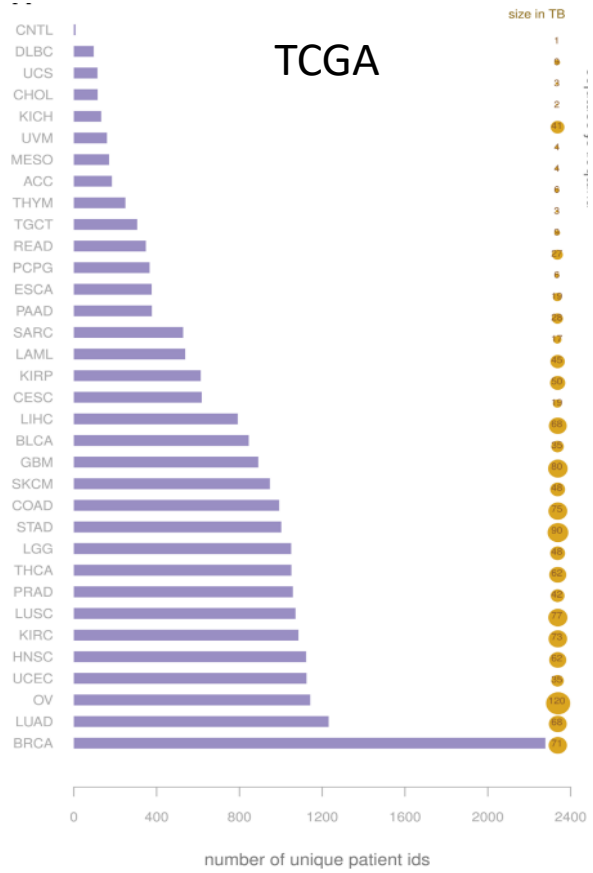


GTEx Portal

V8 Release	# Tissues	# Donors	# Samples
Total	54	948	17382
With Genotype	54	838	15253
Has eQTL Analysis*	49	838	15201

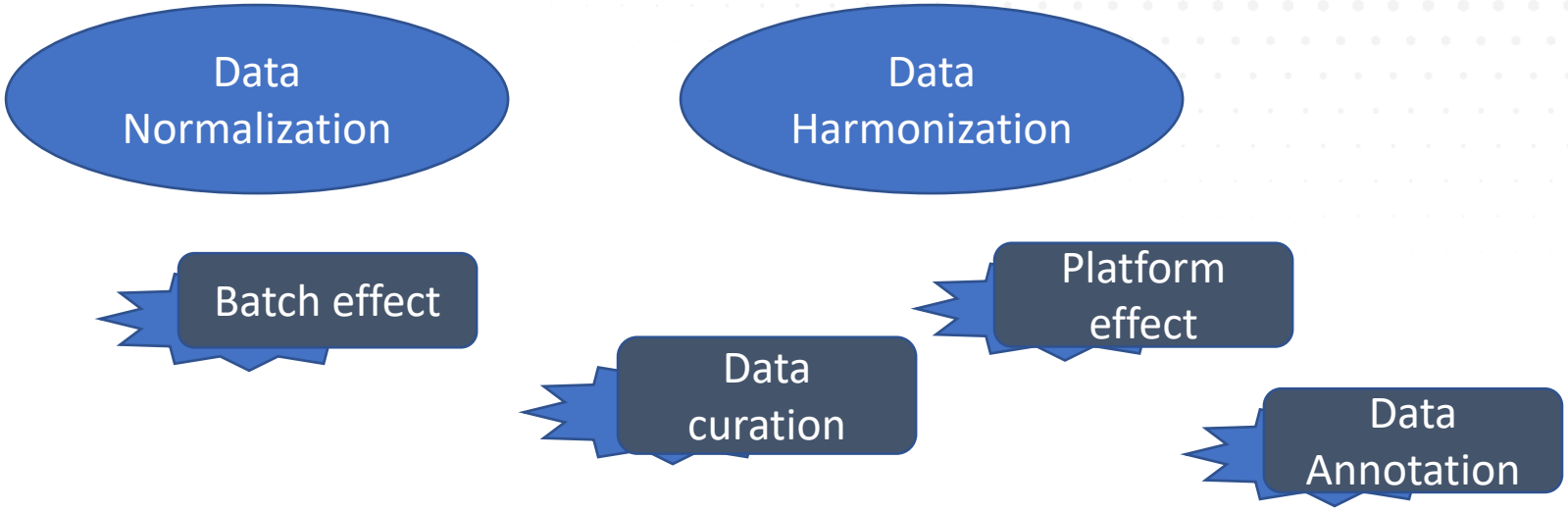
* Number of samples with genotype >= 70

TCGA



Challenges with the public data

eurofins | BioPharma Services



It's an important but iterative and complex task.

We created a computational tool to automate the detection and correction of many of these issues.

Challenges with the public data

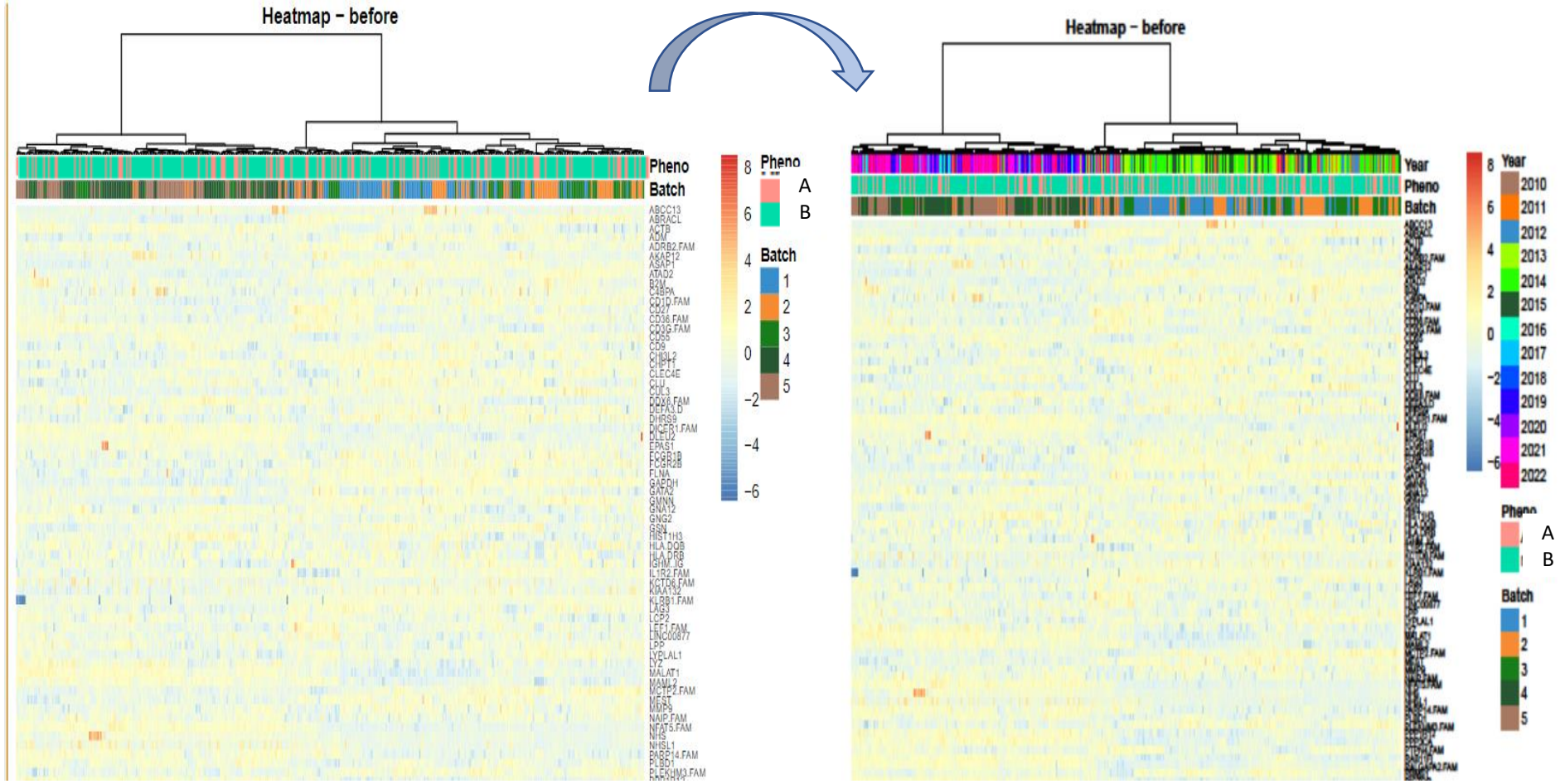
eurofins | BioPharma Services

Batch effect

Data curation

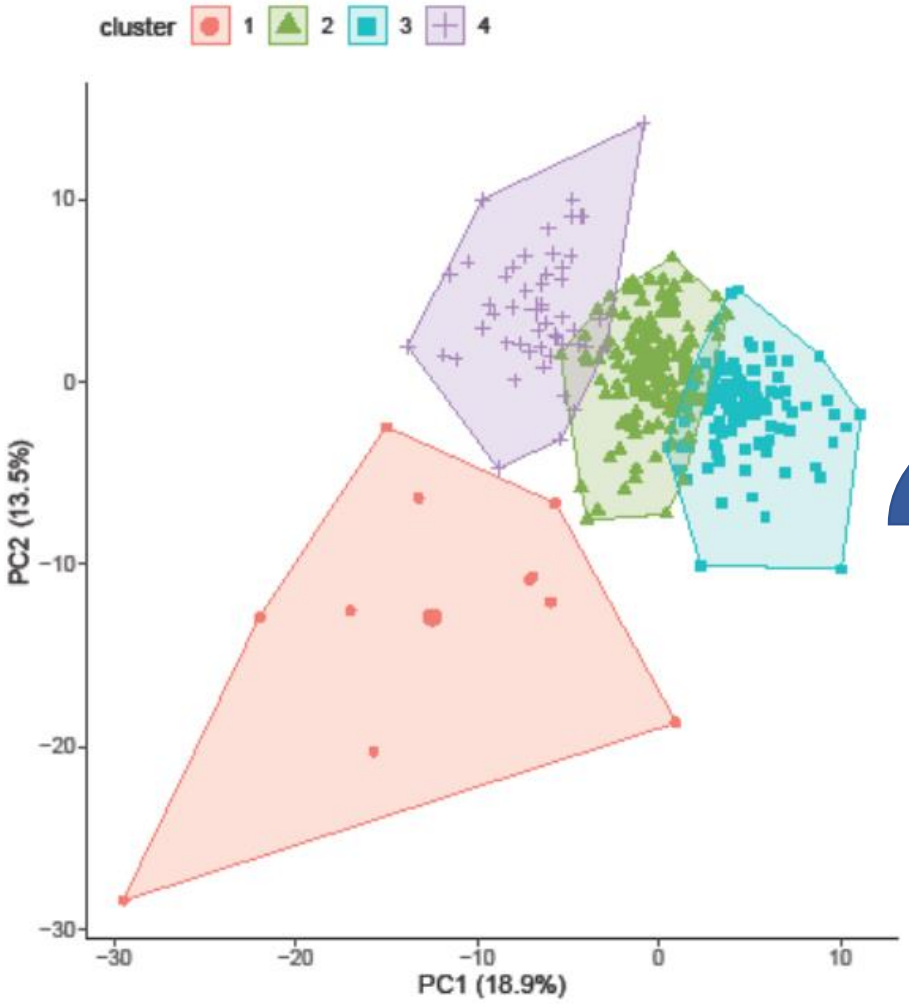
Platform effect

Data Annotation



Learn to deal with the heterogeneity of the human population

eurofins | BioPharma Services



1- Transplant patient's gene-expression profiles.
2- Phenotype: no-rejection & sub-clinical rejection.
3- No-rejection patients/samples are not healthy individuals.
4- Observed 4 clusters, none were unique to a phenotype.

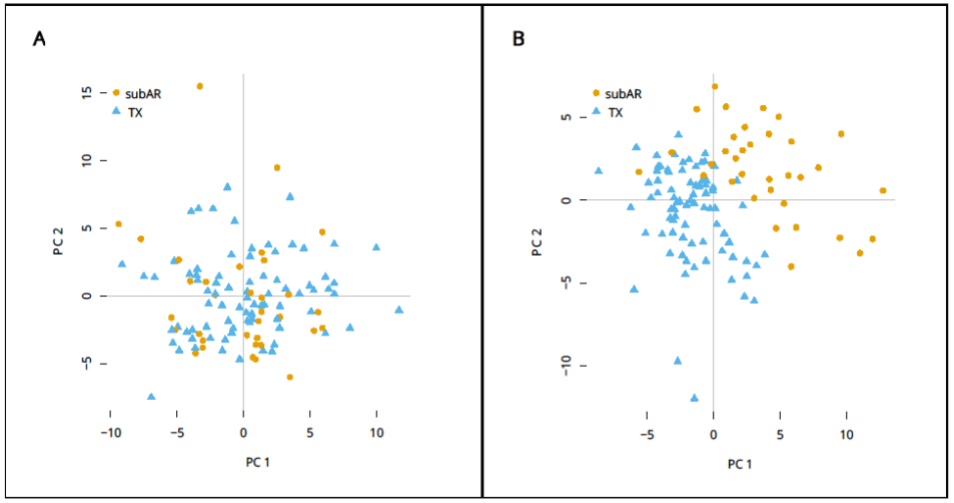


Figure 3: Panel (A)- PCA plot of the distribution of TX and subAR samples using all the features (16,600 genes); Panel (B)- PCA plot of the distribution of TX and subAR samples using top 1% features (166 genes).

I must share an old (2017) story….

eurofins | BioPharma Services

Metagenomic Images and Convolutional Neural Networks Establish the Association Between Gastrointestinal Microbiomes in Beef Cattle and Pathogen Shedding

Rohita Sinha[1], Andy Benson[1]; Steve Kachman[2]; Etsuko Moriyama[3]; Jennifer Clarke[1,2]; Jim Bono[4]; Jim Wells[4]; Larry Kuehn[4]

1- Department of Food Science and Technology, University of Nebraska, Lincoln, NE, 68588, USA; 2-Department of Statistics, University of Nebraska, Lincoln, NE, 68583, USA; 3-School of Biological Sciences, University of Nebraska, Lincoln, NE, 68588,4- US MARC

INTRODUCTION

- Shiga toxin producing Escherichia coli (STEC) are responsible for significant illness[1] and beef cattle are a major animal reservoir of these pathogens. Little is known about the relationship between the colonic microbiota and STEC shedding profiles[1].
- Shotgun metagenomic data and STEC shedding profile data was generated from >1,300 animals in five different cohorts. Each animals was sampled 7 times during peak shedding seasons and metagenomic data was generated from a single composite of all time points per animal while shedding was measured in each individual time point.
- Although the data set was designed to be statistically-powered, the unique biological and ecological features (sparse STEC shedding, unknown environmental attributes, macro-ecology of the MARC herd, and composited samples) led to a high signal to noise ratio. Consequently, we used a new approach discover associations between colonic microbiome profiles and STEC shedding in beef cattle.
- By converting microbial abundances from metagenome data to RGB images, we successfully retrained a Convolutional Neural Network-based image classifier (Inception V3) to classify Shedders and Non-shedders on the basis of the colonic microbiota.

Experimental Design & Metagenomic Data Processing

Fecal Sample Collection:
Total 1328 samples from 5 different cohorts over a 5-year period

Seven samples from each animal during peak shedding period were composited into a single sample per animal for shotgun Metagenome sequencing: (HiSeq: 2 x 150 paired end reads)

Quality check

1- Cutadapt: removes terminal low Phred-score bases
2- Trimmed reads shorter than 100bp were discarded

Microbial Profiling

Assembly free approach: Short NGS reads were taxonomically annotated in the protein-space using KAIJU (a metagenome taxonomy annotation package).

Metagenomic Images

1- Phylum, Family and Genus level abundance data was merged into a single table.
2- Rearrange the table:
- Rows: total 26 rows based on the alphabet A-Z.
- Each row had abundance of taxa starting with that alphabet.
- Each row had different effective number of columns. Rows were padded with zero-value columns to have same size.

Retraining CNN (Inception V3) Classifer

Figure1: Conceptual description of how Convolutional Neural Network works. Filters (colored boxes) are (N, M, D) matrices which are traversed over the input images and each one extracts/highlights different features of an image. In the current work we retrained Inception V3, an existing CNN based image classifier[2].

Figure2: Example of metagenomic images which we used to retrain the Inception V3..

Training set: 40 metagenomic images of Non-shedders (CFU == 0) and 40 images of Shedders (CFU > 0)

Test set: 100 Non-shedder images and 100 Shedder images

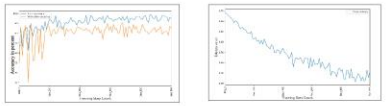


Figure3: Left image shows the progress in the test & validation accuracy with number of iterations of minimization step. Right panel shows the minimization of the cost-function with iterations.

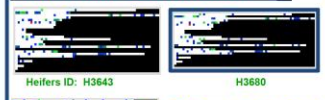
Minimal retraining was sufficient for classification

Validation accuracy: 92% validation and test accuracy was achieved on the training set.

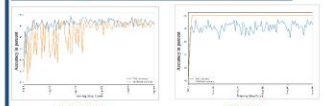
Test Accuracy: Classifier was further tested on additional Shedder (100) & Non-shedder (100) images, and the corresponding success rates were 98% and 88% for Shedder and Non-shedders, respectively.

- We observed a significant effect of the year of sample-collection on the metagenomic profile and the corresponding images. Models trained on specific year's samples had shown better success rates for samples from that year.

Metagenomic Images of Shedders (Sires & Heifers) are Gender Dependent



Retraining CNN: Gender Based Classifiers



Training set 1: 40 metagenomic images of Non-shedders (CFU == 0) and 40 images of Shedders (CFU > 0, Heifer Only)
Training set 1: 40 metagenomic images of Non-shedders (CFU == 0) and 40 images of Shedders (CFU > 0, Sire Only)
Test set:100 Non-shedder images and 63 Sire & 84 Heifer Shedder images
Shedder Prediction Accuracy: Sire Only: 62/63 and Hei
Low Cross Accuracy: Sire only classifier tested on Heifs
Sire Only: 42/63 and Heifer Only: 8/84

Conclusions

1. Groups of colonic microbes, rather than individual much more strongly associated with the STEC she
2. Microbial abundances across multiple taxonom Family and Genus) are informative and can classifiable Metagenomic images.
3. The CNN-based image processing has been u other complex systems[3,4] and should have bro MWAS with other complex features (e.g. feed intak

REFERENCES
1- "Colonization of Beef Cattle by Shiga Toxin-Producing *Escherichia coli* durin Study," PLoS ONE11(2): e0148518, Feb. 2016.
2- "Rethinking the Inception Architecture for Computer Vision," Computer Visio
3- "Dermatologist-level classification of skin cancer with deep neural networks," February 2017)
4- "When technology meets technology: Retrained 'Inception V3' classifier for t IEEE, BIBM 2017

Conferences > 2017 IEEE International Confe...

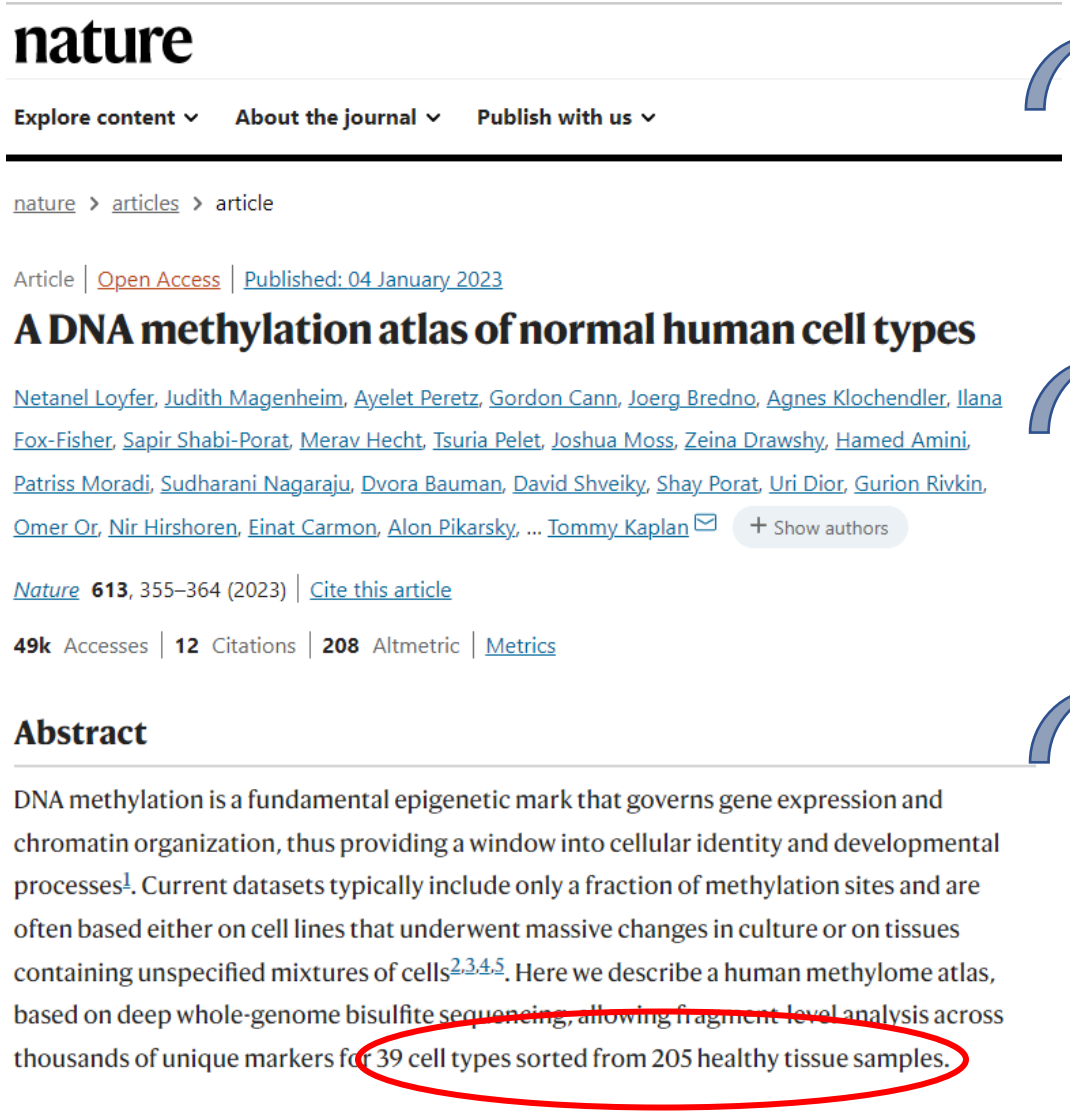
When technology meets technology: Retrained 'Inception V3' classifier for NGS based pathogen detection

Publisher: IEEE Cite This PDF

Rohita Sinha ; Jennifer Clarke **All Authors**

Let's discuss one of the epigenetic omics data – DNA Methylation

eurofins | BioPharma Services



nature

Explore content ∨ About the journal ∨ Publish with us ∨

nature › articles › article

Article | Open Access | Published: 04 January 2023

A DNA methylation atlas of normal human cell types

Netanel Loyfer, Judith Magenheim, Ayelet Peretz, Gordon Cann, Joerg Bredno, Agnes Klochendler, Ilana Fox-Fisher, Sapir Shabi-Porat, Merav Hecht, Tsuria Pelet, Joshua Moss, Zeina Drawshy, Hamed Amini, Patriss Moradi, Sudharani Nagaraju, Dvora Bauman, David Shveiky, Shay Porat, Uri Dior, Gurion Rivkin, Omer Or, Nir Hirshoren, Einat Carmon, Alon Pikarsky, ... Tommy Kaplan + Show authors

Nature **613**, 355–364 (2023) | Cite this article

49k Accesses | **12** Citations | **208** Altmetric | Metrics

Abstract

DNA methylation is a fundamental epigenetic mark that governs gene expression and chromatin organization, thus providing a window into cellular identity and developmental processes[1]. Current datasets typically include only a fraction of methylation sites and are often based either on cell lines that underwent massive changes in culture or on tissues containing unspecified mixtures of cells[2,3,4,5]. Here we describe a human methylome atlas, based on deep whole-genome bisulfite sequencing, allowing fragment-level analysis across thousands of unique markers for 39 cell types sorted from 205 healthy tissue samples.

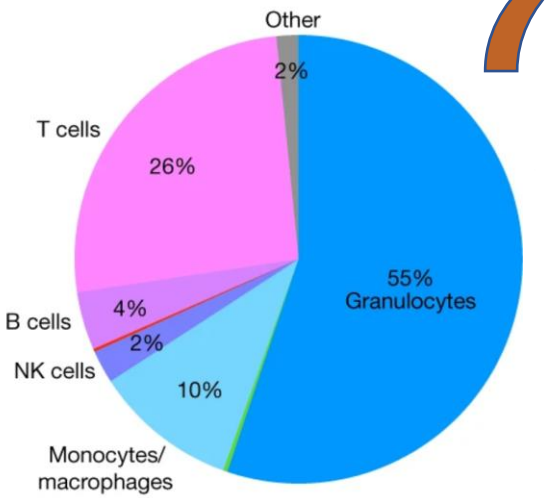
DNA methylation creates a cell-specific epigenetic signature

Found blocks of homogeneously methylated CpG sites

2,783,421 methylation blocks of at least three CpGs with an average length of 544 bp

Let's talk about the cell-free nucleotide methylation data

eurofins | BioPharma Services



Cell-types % in healthy blood cell-free DNA

nature communications

Explore content | About the journal | Publish with us

nature > nature communications > articles > article

Article | Open Access | Published: 13 December 2022

Tumor fractions deciphered from circulating cell-free DNA methylation for cancer early diagnosis

Xiao Zhou, Zhen Cheng, Mingyu Dong, Qi Liu, Weiyang Yang, Min Liu, Junzhang Tian & Weibin Cheng

Nature Communications **13**, Article number: 7694 (2022) | Cite this article

5788 Accesses | **19** Altmetric | Metrics

A Publisher Correction to this article was published on 19 January 2023

nature biotechnology

Explore content | About the journal | Publish with us

nature > nature biotechnology > brief communications > article

Brief Communication | Open Access | Published: 07 February 2022

Cell types of origin of the cell-free transcriptome

Sevahn K. Vorperian, Mira N. Moufarrej, Tabula Sapiens Consortium & Stephen R. Quake

Nature Biotechnology **40**, 855–861 (2022) | Cite this article

24k Accesses | **14** Citations | **156** Altmetric | Metrics

A Publisher Correction to this article was published on 28 March 2022

Explore content | About the journal | Publish with us | Subscribe

nature > nature reviews nephrology > review articles > article

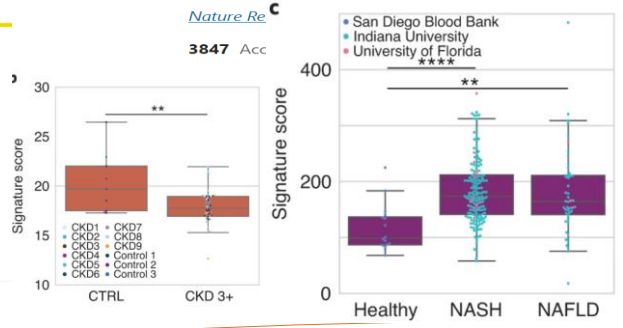
Review Article | Published: 24 May 2021

Liquid biopsies: donor-derived cell-free DNA for the detection of kidney allograft injury

Michael Oellerich, Karen Sherwood, Paul Keown, Ekkehard Schütz, Julia Beck, Johannes Stegbauer, Lars Christian Rump & Philip D. Walson

Nature Re

3847 Acc



Solving biological mixture models

eurofins | BioPharma Services

Blood cell-free DNA is a biological-mix

Microbiome is another biological-mix

Blood transcriptome is a biological-mix

We assume knowing all the components of these biological-mixtures.

It helps us using following constraints:

1- Sum of proportion of all known components would be 1.0 (or 100%).

2- We deal with non-negative numbers, since proportions are positive numbers.

Solving biological mixture models
(Quadratic Programming)

eurofins | BioPharma Services

We solve this equation: E = S x C

S (signature matrix) :
rows = component-id, columns = feature-ids
matrix-values = feature-frequency of a component

E (bulk matrix) :
rows = sample-id, columns = feature-ids
matrix-values = frequency of a feature for the given sample

C (proportion matrix) :
rows = sample-id, columns = component-id
matrix-values = proportion of a component in a sample

We eventually try to minimize the difference between the observed matrix (E) and computed matrix (S x C) *i.e.,* E – (S x C) = 0. Which yields optimal values for the matrix C (proportion of each component in a biological-mixture).

Solving biological mixture models
(Quadratic Programming)

eurofins | BioPharma Services

We implemented Quadratic-programming to compute the fraction of multiple tissues in the blood cell-free DNA, using cell-specific methylation patterns.

HOW TO VALIDATE OUR PROTOCOL

Human Methylome Atlas

eurofins | BioPharma Services

Chromosomal location

Cell Types

Probability of methylation

		start	end	bcell_1	bcell_2	bcell_3	bcell_4	bcell_5	bcell_6	bcell_7	bcell_8	bcell_9	colon_1	colon_2	colon_3	colon_4	colon_5	colon_6	colon_7	colon_8	colon_9	colon_10	colon_11	colon_12	colon_13	colon_14	colon_15	colon_16	colon_17
2	chr1	836950	837233	0.842222	0.857222	0.942	0.7975	0.832778	0.731111	0.875	0.764333	0.768667	0.419039	0.255106	0.477394	0.4713	0.588611	0.359039	0.583167	0.307333	0.363502	0.555335	0.397236	0.261111	0.328889	0.303333	0.361111	0.430372	0.342222
3	chr1	850702	851077	0.639444	0.664167	0.824286	0.447917	0.523333	0.5025	0.77625	0.601706	0.632647	0.315106	0.367725	0.111319	0.440853	0.351097	0.183447	0.439769	0.380978	0.203888	0.455693	0.397159	0.155	0.351111	0.352778	0.437222	0.447467	0.463333
4	chr1	864780	865152	0.808571	0.899688	0.878529	0.863636	0.805909	0.816944	0.932941	0.845235	0.855294	0.1545	0.372608	0.339042	0.573236	0.536642	0.528861	0.392539	0.467261	0.561499	0.600314	0.45218	0.200556	0.413333	0.375	0.457222	0.5465	0.493333
5	chr1	868110	868741	0.939688	0.874118	0.892647	0.767857	0.752778	0.876087	0.970789	0.832261	0.848348	0.238196	0.119233	0.243783	0.402622	0.444587	0.193602	0.356065	0.307896	0.430329	0.480779	0.294862	0.158696	0.133913	0.08913	0.132609	0.253622	0.137826
6	chr1	920168	920354	0.69125	0.835	0.786875	0.84	0.89125	0.820625	0.826875	0.767875	0.715625	0.185613	0.119213	0.171169	0.277556	0.475231	0.181294	0.311706	0.207519	0.164282	0.327283	0.173214	0.097143	0.142857	0.087143	0.092857	0.145088	0.108571
7	chr1	986214	986847	0.9666	0.662656	0.928226	0.680556	0.805682	0.926711	0.931351	0.663561	0.674881	0.408202	0.713205	0.429708	0.685894	0.755956	0.274271	0.737506	0.396559	0.499692	0.673115	0.513039	0.157558	0.599535	0.749186	0.762093	0.856528	0.75686
8	chr1	995950	996317	0.622609	0.692895	0.613864	0.828125	0.67675	0.790652	0.85175	0.640652	0.627478	0.230417	0.245543	0.132789	0.515107	0.555326	0.181913	0.353346	0.29428	0.428922	0.462753	0.280737	0.131087	0.230652	0.226087	0.240652	0.380276	0.256957
9	chr1	1045135	1045787	0.928158	0.938421	0.967105	0.974375	0.895	0.912778	0.959444	0.955105	0.952526	0.168897	0.064382	0.063671	0.501521	0.779697	0.1798	0.248705	0.227584	0.772482	0.835483	0.207016	0.036316	0.061316	0.025263	0.042632	0.154866	0.019412
10	chr1	1058075	1058994	0.758378	0.760781	0.5816	0.70625	0.6785	0.709865	0.850735	0.771647	0.711351	0.238964	0.067834	0.113764	0.448839	0.654955	0.190281	0.24372	0.221988	0.639448	0.642565	0.171531	0.091111	0.076081	0.056528	0.055143	0.128334	0.047361
11	chr1	1061092	1061660	0.98375	0.86125	0.922778	0.9835	0.8925	0.7742	0.9398	0.936231	0.9375	0.159631	0.170269	0.279185	0.758629	0.868752	0.30415	0.352404	0.239475	0.721272	0.747068	0.212132	0.196923	0.136538	0.262692	0.081923	0.437217	0.100192
12	chr1	1062143	1062366	0.935455	0.644167	0.926923	0.923214	0.604167	0.771	0.888	0.724636	0.7424	0.453903	0.225547	0.084593	0.748627	0.752907	0.15962	0.303133	0.1394	0.688944	0.643062	0.240998	0.253	0.47	0.618667	0.462	0.679933	0.524333
13	chr1	1062973	1063885	0.925	0.888068	0.873214	0.905789	0.891585	0.904271	0.907653	0.91134	0.933653	0.291846	0.067002	0.026553	0.790367	0.497405	0.27113	0.282361	0.179864	0.234542	0.450056	0.175376	0.023571	0.037872	0.009149	0.035444	0.127936	0.022245
14	chr1	1064792	1065434	0.947308	0.9415	0.825714	0.9112	0.813704	0.933846	0.971346	0.908444	0.908741	0.186611	0.051396	0.077141	0.874224	0.814713	0.249843	0.2619	0.227487	0.230876	0.492252	0.159698	0.043519	0.036667	0.021111	0.033148	0.168522	0.036667
15	chr1	1065665	1066181	0.870833	0.779348	0.771333	0.818148	0.806053	0.826552	0.926207	0.82975	0.87663	0.283105	0.082502	0.074078	0.763357	0.628884	0.137238	0.259174	0.219745	0.207932	0.361339	0.193325	0.01431	0.024138	0.031379	0.014655	0.395093	0.01
16	chr1	1066236	1066586	0.905667	0.87	0.981667	0.797	0.952857	0.818333	0.883667	0.816867	0.860333	0.28044	0.038607	0.1462	0.78972	0.637787	0.094213	0.338107	0.227747	0.269934	0.541576	0.147665	0.062	0.004	0.021333	0.029333	0.353023	0.059333
17	chr1	1067378	1067547	0.9495	0.928333	0.9375	0.916667	0.935909	0.815909	0.967	0.90075	0.904167	0.255188	0.505208	0.357738	0.951983	0.9625	0.397721	0.203758	0.352413	0.372809	0.582676	0.190778	0.1025	0.089167	0.080833	0.0625	0.082508	0.100833
18	chr1	1090810	1091043	0.961667	0.956111	0.9375	0.961111	0.937778	0.933889	0.892222	0.907	0.932	0.940094	0.97985	0.955583	0.921822	0.92065	0.934694	0.917983	0.936511	0.947112	0.950938	0.956895	0.937778	0.94	0.946667	0.933333	0.8913	0.955556
19	chr1	1098515	1099650	0.841455	0.848462	0.833507	0.810392	0.867424	0.90223	0.937324	0.90947	0.822122	0.204607	0.167783	0.089262	0.117331	0.253946	0.169263	0.268848	0.369507	0.249086	0.319376	0.166676	0.036849	0.044795	0.0525	0.017933	0.218236	0.028767
20	chr1	1108460	1108723	0.8675	0.77	0.9	0.955385	0.753214	0.768571	0.739643	0.779615	0.756286	0.6818	0.696514	0.786864	0.638886	0.691943	0.824479	0.7466	0.752618	0.838209	0.776602	0.71934	0.757143	0.807143	0.835714	0.862143	0.813018	0.828571
21	chr1	1161352	1161624	0.979375	0.961364	0.95	0.924545	0.9465	0.91	0.931818	0.967273	0.970273	0.918014	0.952341	0.894486	0.837432	0.909091	0.880559	0.907532	0.906014	0.888918	0.897149	0.921996	0.879091	0.932727	0.911818	0.937273	0.937441	0.910909
22	chr1	1164456	1164780	0.895	0.919545	0.872308	0.959	0.865	0.924667	0.941	0.8726	0.921667	0.92921	0.697803	0.899787	0.92095	0.81679	0.93916	0.89509	0.880863	0.912776	0.897925	0.915613	0.857333	0.862667	0.860667	0.863333	0.889657	0.830667

An algorithm of simulate biological-mixtures using DNA Methylation patterns

eurofins | BioPharma Services

In order to test our method, we create a bulk data which shows the percentage of methylation in each position in a specific sample.

Step 1 : Simulate mixtures: define the proportions of each cell types in the mix (*i.e.,* c1:c2:c3 = 10:20:70)

Step 2 :For each position in the Methylome table:

* Let's say we have (**N**) methylation records (based on the NGS data): let's make N = 10
* Let's say the methylation-probabilities for each cell types are: c1 = 70%, c2 = 30%, c3 = 80%
* Run a uniform random-number generator **N** times
* based on the simulated mix, the methylation records (MR) would ideally have following distribution: c1 = 1 MR, c2 = 2 MR, c3 = 7MR
* Run a uniform random-number generator 10 times-- by looking at the original number which shows the probability of the methylation in that specific position, if the random number is greater than the original methylation probabilities, convert the number to 0(unmethylated), otherwise convert the number to 1 (methylated)

Step 3: Finally, sum all the 1s and divide by **N** (10), which is the observed frequency of methylation on a position.

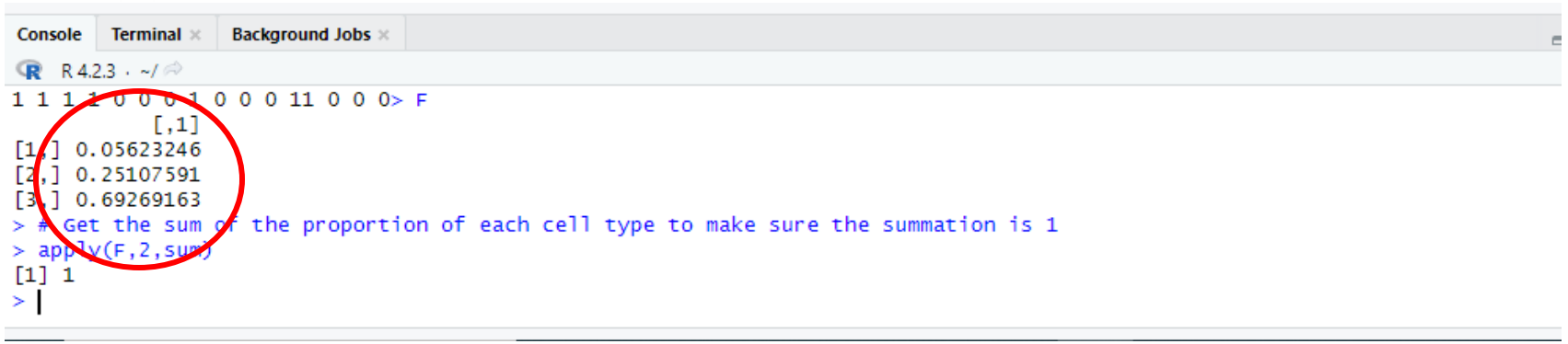
Now the fun begins… let's run a few simulations

eurofins | BioPharma Services

Simulation mix:

3 cell types

cell proportions: 2, 8 ,20 MRs (2/30 = 0.06; 8/30 = 0.26; 20/30 = 0.66)



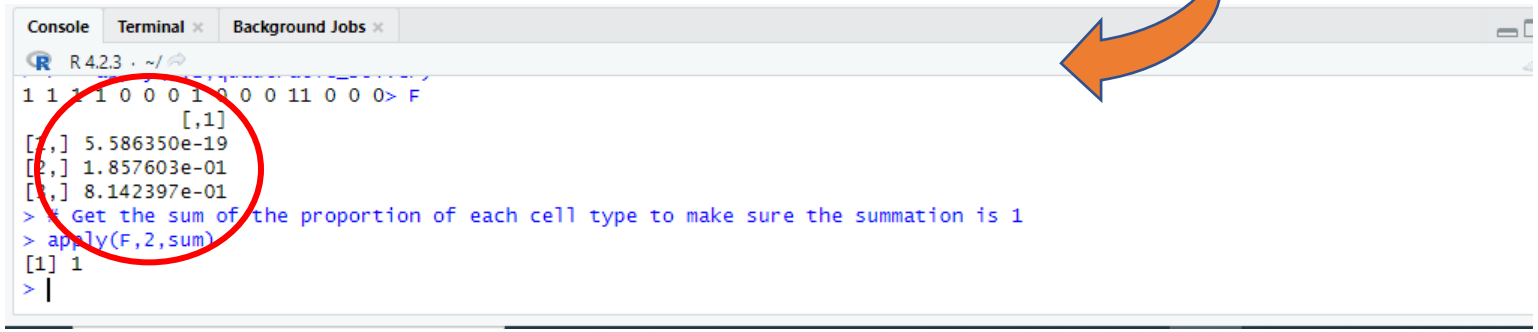
Now the fun begins… let's run a few simulations

eurofins | BioPharma Services

Simulation mix:

3 cell types

cell proportions: 1, 19 ,80 MRs (1/100 = 0.01; 19/100 = 0.19; 80/100 = 0.80)



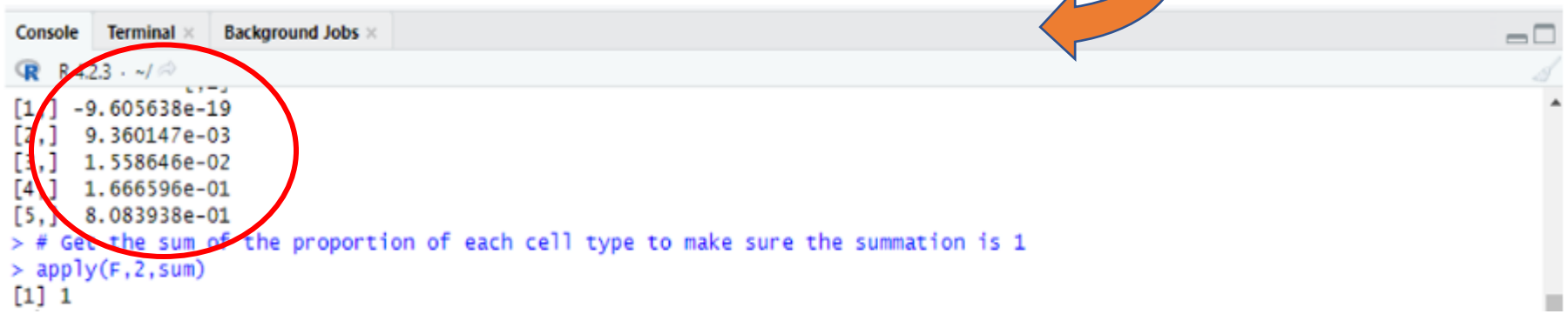
Now the fun begins… let's run a few simulations

eurofins | BioPharma Services

Simulation mix:

5 cell types

cell proportions: 5, 10 ,15, 170, 800 MRs (5/1000 = 0.005; 10/1000 = 0.01; 15/1000 = 0.015, 170/1000 = 0.17, 800/1000 = 0.80)



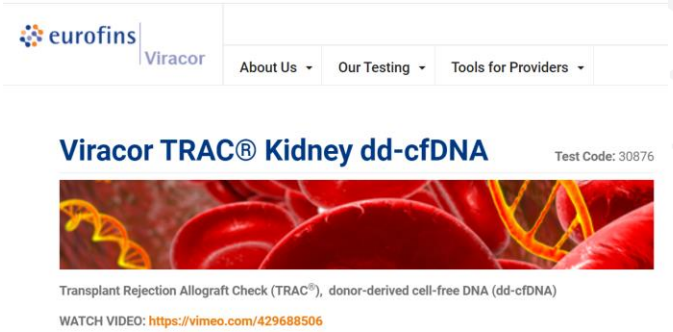
NOW LET'S TRY IT ON THE REAL DATA

eurofins | BioPharma Services

We extracted cell-free DNA from multiple Kidney-transplant patients

Generated cell-free DNA methylation data

Processed cf-DNA methylation data using our protocol



	Kidney 1	Kidney 3	Kidney 9	Kidney 6	AVG
Sample 1	1.6	2.3			1.95
Sample 2	3.6			5.7	4.65
Sample 3				5	5
Sample 4					0
Sample 5	2.2				2.2
Sample 6	1		2.2		0.61
Sample 7	2.1				2.1
Sample 8			1		1
Sample 9	1		0.44		0.72
Sample 10			4.9		4.9
Sample 11	1.7				1.7
Sample 12	2.2				2.2

Summary:

We discussed:

- How a centralized Bioinfo & AI unit helps our plans.
- Why the current A.I/M.L are more data hungry.
- What skill sets we may need for the optimal utilization of the big-data.
- Usage of the public-data and associated challenges.
- Strategies to account for the heterogeneity of the biological data.
- Epigenetic data and solving mixture-models.
- Our pursuit to use the Methylation-data to better understand the allograft injuries.

Thanks for joining us & listening!

eurofins | BioPharma Services