Machine Learning View of OMICs Integration

ISMB / ECCB 2021

Tutorial 4: A practical introduction to multi-omics integration and network analysis

Nikolay Oskolkov, NBIS SciLifeLab, 22.07.2021

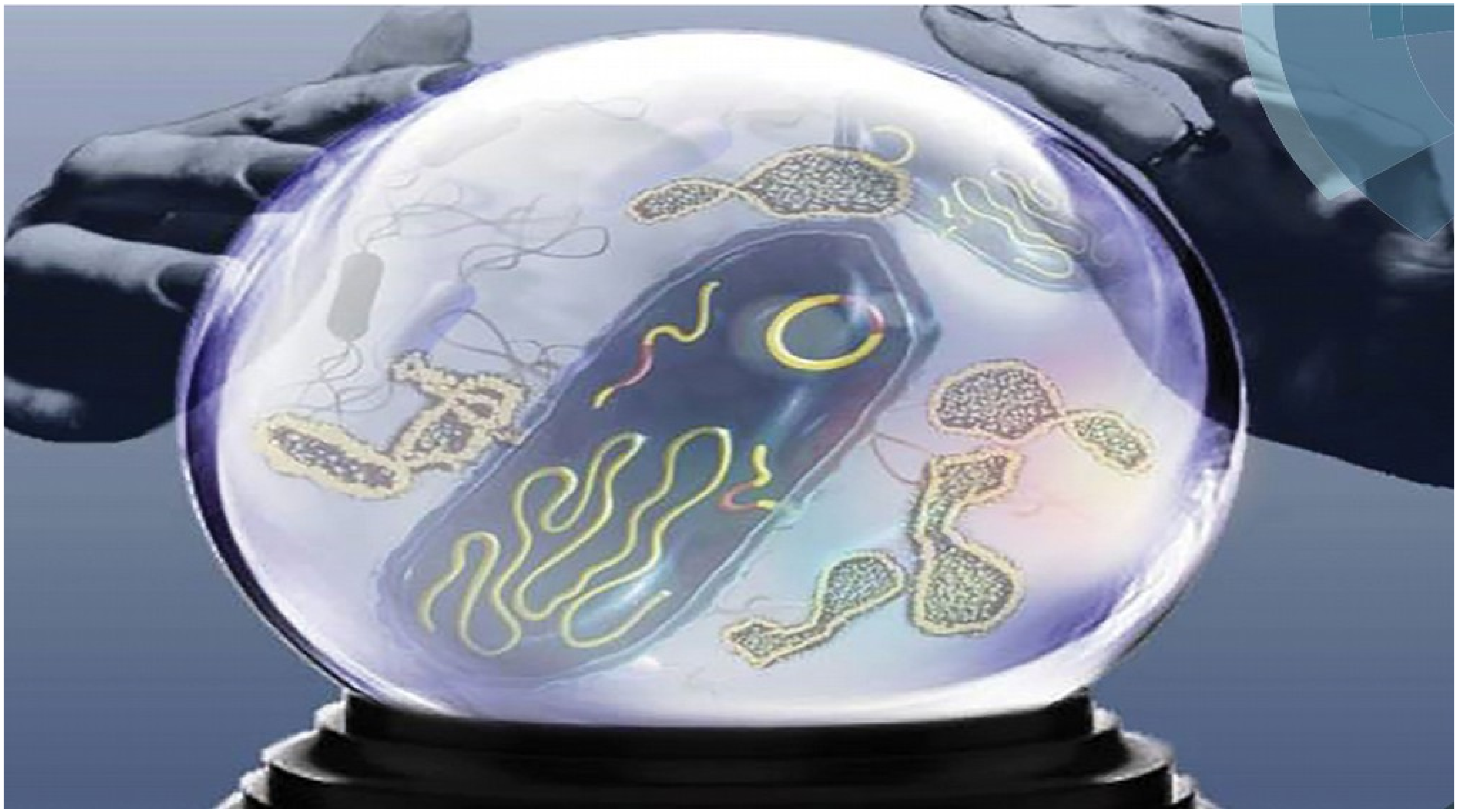
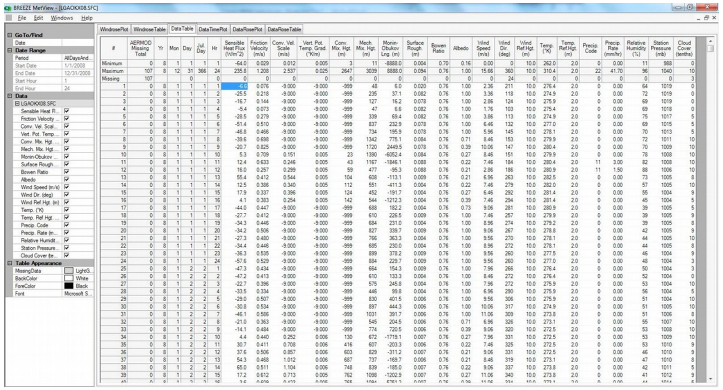


Image adapted from Molecular Omics, Issue 1, 2018

Various Data Distributions

Tabular



BREEZE MetView - [LGAOKX08.SFC]

GoTo/Find
Date

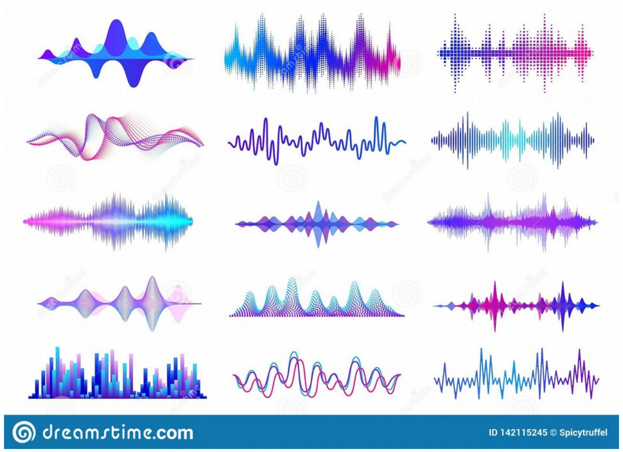
Date Range
Period AllDaysAnd...
Start Date 1/1/2008
End Date 12/31/2008
Start Hour 1
End Hour 24

Data
LGAOKX08.SFC
Sensible Heat Fl... ☑
Friction Velocity ... ☑
Conv. Vel. Scal... ☑
Vert. Pot. Temp.... ☑
Conv. Mix. Hgt. ... ☑
Mech. Mix. Hgt.... ☑
Monin-Obukov ... ☑
Surface Rough.... ☑
Bowen Ratio ☑
Albedo ☑
Wind Speed (m/s) ☑
Wind Dir. (deg) ☑
Wind Ref.Hgt. (m) ☑
Temp. (°K) ☑
Temp. Ref.Hgt. ... ☑
Precip. Code ☑
Precip. Rate (m... ☑
Relative Humidit... ☑
Station Pressure... ☑
Cloud Cover (te... ☑

Table Appearance
MissingData LightG...
BackColor White
ForeColor Black
Font Microsoft S...

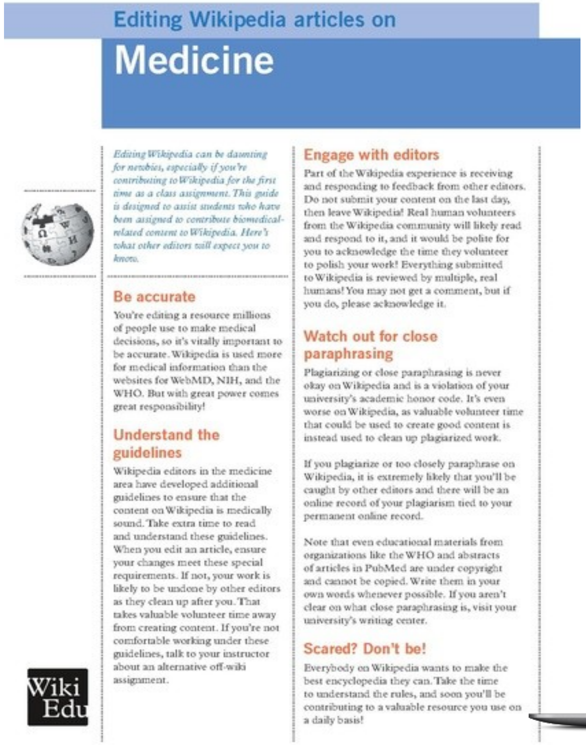
#	AERMOD Missing Total	Yr	Mon	Day	Jul. Day	Hr	Sensible Heat Flux (W/m^2)	Friction Velocity (m/s)	Conv. Vel. Scale (m/s)	Vert. Pot. Temp. Grad. (°K/m)	Conv. Mix. Hgt. (m)	Mech. Mix. Hgt. (m)	Monin-Obukov Lng. (m)	Surface Rough. (m)	Bowen Ratio	Albedo	Wind Speed (m/s)	Wind Dir. (deg)	Wind Ref.Hgt. (m)	Temp. (°K)	Temp. Ref.Hgt. (m)	Precip. Code	Precip. Rate (mm/hr)	Relative Humidity (%)	Station Pressure (mb)	Cloud Cover (tenths)
Minimum	0	8	1	1	1	1	-64.0	0.029	0.012	0.005	3	11	-8888.0	0.004	0.70	0.16	0.00	0	10.0	262.0	2.0	0	0.00	11	988	0
Maximum	107	8	12	31	366	24	235.8	1.208	2.537	0.025	2647	3039	8888.0	0.094	0.76	1.00	15.66	360	10.0	310.4	2.0	22	41.70	96	1040	10
Missing	107			0		0	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0	0	0	0	3
1	0	8	1	1	1	1	-6.6	0.076	-9.000	-9.000	-999	48	6.0	0.020	0.76	1.00	2.36	211	10.0	276.4	2.0	0	0.00	64	1019	0
2	0	8	1	1	1	2	-25.5	0.218	-9.000	-9.000	-999	235	37.1	0.082	0.76	1.00	3.36	118	10.0	274.9	2.0	0	0.00	72	1019	0
3	0	8	1	1	1	3	-16.7	0.144	-9.000	-9.000	-999	127	16.2	0.078	0.76	1.00	2.86	124	10.0	275.9	2.0	0	0.00	69	1019	0
4	0	8	1	1	1	4	-5.4	0.073	-9.000	-9.000	-999	47	6.6	0.082	0.76	1.00	1.76	103	10.0	275.4	2.0	0	0.00	69	1018	0
5	0	8	1	1	1	5	-28.5	0.279	-9.000	-9.000	-999	339	69.4	0.082	0.76	1.00	3.86	113	10.0	274.9	2.0	0	0.00	75	1017	5
6	0	8	1	1	1	6	-51.4	0.510	-9.000	-9.000	-999	837	232.9	0.078	0.76	1.00	6.46	132	10.0	277.0	2.0	0	0.00	69	1015	5
7	0	8	1	1	1	7	-46.8	0.466	-9.000	-9.000	-999	734	195.9	0.078	0.76	1.00	5.96	145	10.0	278.1	2.0	0	0.00	70	1013	5
8	0	8	1	1	1	8	-39.6	0.698	-9.000	-9.000	-999	1342	775.1	0.084	0.76	0.71	8.46	153	10.0	279.9	2.0	0	0.00	72	1011	10
9	0	8	1	1	1	9	-20.7	0.825	-9.000	-9.000	-999	1720	2449.5	0.078	0.76	0.39	10.06	147	10.0	280.4	2.0	0	0.00	70	1009	10
10	0	8	1	1	1	10	5.3	0.709	0.151	0.005	23	1390	-6052.4	0.084	0.76	0.27	8.46	151	10.0	279.9	2.0	0	0.00	78	1008	10
11	0	8	1	1	1	11	12.4	0.633	0.246	0.005	43	1167	-1846.1	0.088	0.76	0.22	7.46	184	10.0	280.4	2.0	11	3.00	82	1008	10
12	0	8	1	1	1	12	16.0	0.257	0.299	0.005	59	477	-95.3	0.088	0.76	0.21	2.86	186	10.0	280.9	2.0	11	1.50	88	1006	10
13	0	8	1	1	1	13	55.4	0.412	0.544	0.005	104	608	-113.1	0.009	0.76	0.21	6.96	263	10.0	282.5	2.0	0	0.00	73	1005	8
14	0	8	1	1	1	14	12.5	0.386	0.340	0.005	112	551	-411.3	0.004	0.76	0.22	7.46	279	10.0	282.0	2.0	0	0.00	57	1005	10
15	0	8	1	1	1	15	17.9	0.337	0.396	0.005	124	452	-191.7	0.004	0.76	0.27	6.46	292	10.0	281.4	2.0	0	0.00	55	1004	9
16	0	8	1	1	1	16	4.1	0.383	0.254	0.005	142	544	-1212.3	0.004	0.76	0.39	7.46	294	10.0	281.4	2.0	0	0.00	45	1004	5
17	0	8	1	1	1	17	-44.0	0.447	-9.000	-9.000	-999	688	182.2	0.004	0.76	0.73	9.06	281	10.0	280.9	2.0	0	0.00	39	1005	5
18	0	8	1	1	1	18	-27.7	0.412	-9.000	-9.000	-999	610	226.5	0.009	0.76	1.00	7.46	257	10.0	279.9	2.0	0	0.00	39	1005	9
19	0	8	1	1	1	19	-34.3	0.446	-9.000	-9.000	-999	684	231.0	0.004	0.76	1.00	8.96	274	10.0	279.2	2.0	0	0.00	39	1005	8
20	0	8	1	1	1	20	-34.2	0.506	-9.000	-9.000	-999	827	339.7	0.009	0.76	1.00	9.06	267	10.0	278.8	2.0	0	0.00	42	1005	9
21	0	8	1	1	1	21	-27.3	0.480	-9.000	-9.000	-999	766	363.3	0.004	0.76	1.00	9.56	270	10.0	278.1	2.0	0	0.00	44	1005	10
22	0	8	1	1	1	22	-34.4	0.446	-9.000	-9.000	-999	685	230.0	0.004	0.76	1.00	8.96	272	10.0	278.1	2.0	0	0.00	44	1005	8
23	0	8	1	1	1	23	-36.3	0.535	-9.000	-9.000	-999	899	378.2	0.009	0.76	1.00	9.56	260	10.0	277.5	2.0	0	0.00	46	1004	9
24	0	8	1	1	1	24	-57.6	0.529	-9.000	-9.000	-999	884	229.7	0.009	0.76	1.00	9.56	260	10.0	277.0	2.0	0	0.00	48	1004	3
25	0	8	1	2	2	1	-47.3	0.434	-9.000	-9.000	-999	664	154.3	0.009	0.76	1.00	7.96	266	10.0	276.4	2.0	0	0.00	50	1004	3
26	0	8	1	2	2	2	-47.2	0.413	-9.000	-9.000	-999	610	133.3	0.004	0.76	1.00	8.46	272	10.0	276.4	2.0	0	0.00	52	1004	0
27	0	8	1	2	2	3	-22.7	0.396	-9.000	-9.000	-999	575	245.8	0.004	0.76	1.00	7.96	272	10.0	275.9	2.0	0	0.00	53	1004	10
28	0	8	1	2	2	4	-33.5	0.334	-9.000	-9.000	-999	446	99.8	0.004	0.76	1.00	6.96	290	10.0	275.9	2.0	0	0.00	56	1004	5
29	0	8	1	2	2	5	-29.0	0.507	-9.000	-9.000	-999	830	401.5	0.006	0.76	1.00	9.56	306	10.0	275.9	2.0	0	0.00	51	1004	10
30	0	8	1	2	2	6	-30.8	0.534	-9.000	-9.000	-999	897	444.3	0.006	0.76	1.00	10.06	317	10.0	274.9	2.0	0	0.00	51	1005	10
31	0	8	1	2	2	7	-46.1	0.586	-9.000	-9.000	-999	1031	391.7	0.006	0.76	1.00	11.06	309	10.0	273.8	2.0	0	0.00	51	1006	8
32	0	8	1	2	2	8	-21.0	0.363	-9.000	-9.000	-999	545	204.5	0.006	0.76	0.71	6.96	326	10.0	273.1	2.0	0	0.00	55	1007	10
33	0	8	1	2	2	9	-14.1	0.484	-9.000	-9.000	-999	774	720.5	0.006	0.76	0.39	9.06	320	10.0	272.5	2.0	0	0.00	53	1008	10
34	0	8	1	2	2	10	4.4	0.440	0.252	0.006	130	672	-1719.1	0.007	0.76	0.27	7.96	331	10.0	272.5	2.0	0	0.00	53	1009	10
35	0	8	1	2	2	11	30.7	0.411	0.708	0.006	416	607	-203.3	0.006	0.76	0.22	7.46	325	10.0	272.5	2.0	0	0.00	53	1010	9
36	0	8	1	2	2	12	37.6	0.506	0.857	0.006	603	829	-311.2	0.007	0.76	0.21	9.06	331	10.0	272.5	2.0	0	0.00	46	1010	9
37	0	8	1	2	2	13	54.3	0.468	1.012	0.006	687	737	-169.7	0.006	0.76	0.21	8.46	319	10.0	273.1	2.0	0	0.00	47	1010	8
38	0	8	1	2	2	14	65.0	0.511	1.104	0.006	748	839	-185.0	0.007	0.76	0.22	9.06	337	10.0	273.8	2.0	0	0.00	42	1011	5
39	0	8	1	2	2	15	17.2	0.612	0.713	0.005	762	1098	-1202.9	0.007	0.76	0.27	11.06	340	10.0	273.8	2.0	0	0.00	41	1012	9

Sound



Image

Text



Editing Wikipedia articles on

Medicine

Editing Wikipedia can be daunting for newbies, especially if you're contributing to Wikipedia for the first time as a class assignment. This guide is designed to assist students who have been assigned to contribute biomedical-related content to Wikipedia. Here's what other editors will expect you to know.

Be accurate

You're editing a resource millions of people use to make medical decisions, so it's vitally important to be accurate. Wikipedia is used more for medical information than the websites for WebMD, NIH, and the WHO. But with great power comes great responsibility!

Understand the guidelines

Wikipedia editors in the medicine area have developed additional guidelines to ensure that the content on Wikipedia is medically sound. Take extra time to read and understand these guidelines. When you edit an article, ensure your changes meet these special requirements. If not, your work is likely to be undone by other editors as they clean up after you. That takes valuable volunteer time away from creating content. If you're not comfortable working under these guidelines, talk to your instructor about an alternative off-wiki assignment.

Engage with editors

Part of the Wikipedia experience is receiving and responding to feedback from other editors. Do not submit your content on the last day, then leave Wikipedia! Real human volunteers from the Wikipedia community will likely read and respond to it, and it would be polite for you to acknowledge the time they volunteer to polish your work! Everything submitted to Wikipedia is reviewed by multiple, real humans! You may not get a comment, but if you do, please acknowledge it.

Watch out for close paraphrasing

Plagiarizing or close paraphrasing is never okay on Wikipedia and is a violation of your university's academic honor code. It's even worse on Wikipedia, as valuable volunteer time that could be used to create good content is instead used to clean up plagiarized work.

If you plagiarize or too closely paraphrase on Wikipedia, it is extremely likely that you'll be caught by other editors and there will be an online record of your plagiarism tied to your permanent online record.

Note that even educational materials from organizations like the WHO and abstracts of articles in PubMed are under copyright and cannot be copied. Write them in your own words whenever possible. If you aren't clear on what close paraphrasing is, visit your university's writing center.

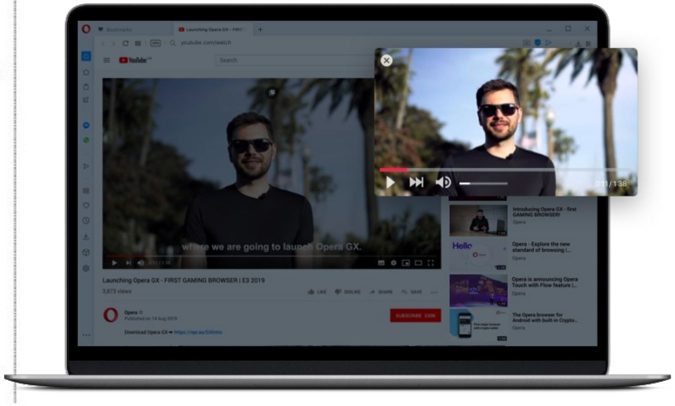
Scared? Don't be!

Everybody on Wikipedia wants to make the best encyclopedia they can. Take the time to understand the rules, and soon you'll be contributing to a valuable resource you use on a daily basis!

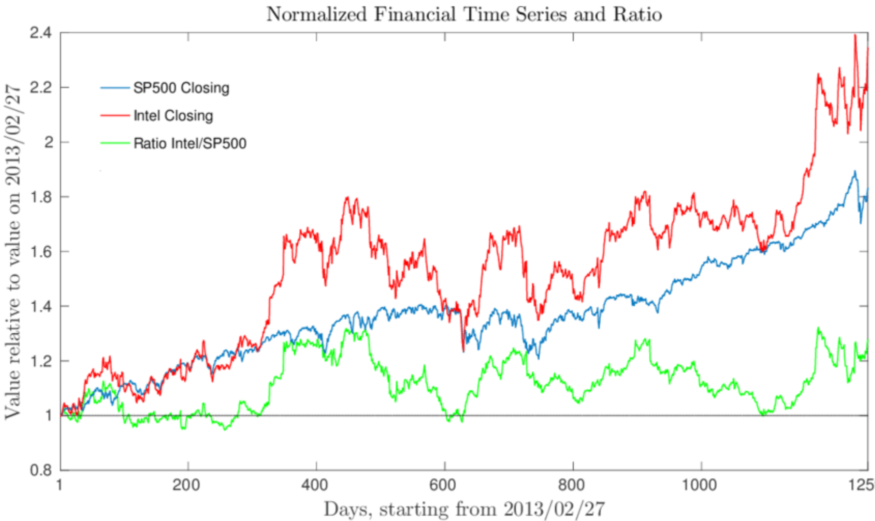
Wiki Edu

DATA

Video



Time Series



High-Dimensional Data

$$\begin{array}{cc} & P_1 \\ N & \begin{pmatrix} 0 & 3 & 1 & 0 & 2 & 3 & 8 & 1 & 1 & 3 \\ 1 & 1 & 0 & 0 & 7 & 1 & 2 & 2 & 3 & 3 \\ 1 & 2 & 2 & 0 & 0 & 6 & 7 & 1 & 2 & 2 \\ 1 & 2 & 3 & 10 & 0 & 4 & 6 & 1 & 0 & 5 \\ 3 & 2 & 2 & 1 & 4 & 3 & 2 & 1 & 6 & 0 \\ 7 & 4 & 4 & 5 & 3 & 9 & 6 & 1 & 6 & 1 \\ 7 & 1 & 1 & 5 & 2 & 8 & 9 & 1 & 3 & 6 \\ 5 & 0 & 1 & 6 & 2 & 0 & 0 & 0 & 1 & 5 \\ 1 & 6 & 3 & 3 & 4 & 6 & 2 & 0 & 1 & 1 \\ 1 & 2 & 2 & 4 & 1 & 1 & 3 & 0 & 8 & 2 \end{pmatrix} \end{array}$$

OMIC1

$$\begin{array}{cc} & P_2 \\ N & \begin{pmatrix} 0 & 3 & 1 & 0 & 2 & 3 & 8 & 1 & 1 & 3 \\ 1 & 1 & 0 & 0 & 7 & 1 & 2 & 2 & 3 & 3 \\ 1 & 2 & 2 & 0 & 0 & 6 & 7 & 1 & 2 & 2 \\ 1 & 2 & 3 & 10 & 0 & 4 & 6 & 1 & 0 & 5 \\ 3 & 2 & 2 & 1 & 4 & 3 & 2 & 1 & 6 & 0 \\ 7 & 4 & 4 & 5 & 3 & 9 & 6 & 1 & 6 & 1 \\ 7 & 1 & 1 & 5 & 2 & 8 & 9 & 1 & 3 & 6 \\ 5 & 0 & 1 & 6 & 2 & 0 & 0 & 0 & 1 & 5 \\ 1 & 6 & 3 & 3 & 4 & 6 & 2 & 0 & 1 & 1 \\ 1 & 2 & 2 & 4 & 1 & 1 & 3 & 0 & 8 & 2 \end{pmatrix} \end{array}$$

OMIC2

$$\begin{array}{cc} & P_3 \\ N & \begin{pmatrix} 0 & 3 & 1 & 0 & 2 & 3 & 8 & 1 & 1 & 3 \\ 1 & 1 & 0 & 0 & 7 & 1 & 2 & 2 & 3 & 3 \\ 1 & 2 & 2 & 0 & 0 & 6 & 7 & 1 & 2 & 2 \\ 1 & 2 & 3 & 10 & 0 & 4 & 6 & 1 & 0 & 5 \\ 3 & 2 & 2 & 1 & 4 & 3 & 2 & 1 & 6 & 0 \\ 7 & 4 & 4 & 5 & 3 & 9 & 6 & 1 & 6 & 1 \\ 7 & 1 & 1 & 5 & 2 & 8 & 9 & 1 & 3 & 6 \\ 5 & 0 & 1 & 6 & 2 & 0 & 0 & 0 & 1 & 5 \\ 1 & 6 & 3 & 3 & 4 & 6 & 2 & 0 & 1 & 1 \\ 1 & 2 & 2 & 4 & 1 & 1 & 3 & 0 & 8 & 2 \end{pmatrix} \end{array}$$

OMIC3

$$\begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

Metabolomics
N ≈ P

$$\begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

Proteomics
N ≈ P

← **manageable**

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 \end{pmatrix}$$

Transcriptomics
N << P
(Single cell: N <= P)

impossible ↓

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 \end{pmatrix}$$

Genomics
N <<<< P

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 \end{pmatrix}$$

Methylomics
N <<<< P

The Curse of Dimensionality complicates OMICs Integration

The Curse of Dimensionality

P is the number of features (genes, proteins, genetic variants etc.)
N is the number of observations (samples, cells, nucleotides etc.)

Biomedicine

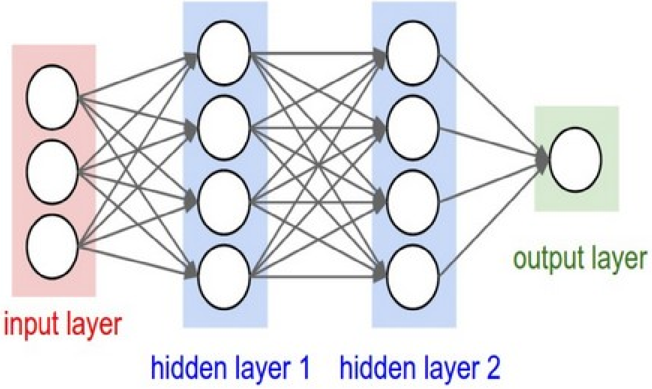
Bayesianism

P >> N

Frequentism

P ~ N

Deep Learning



P << N

Amount of Data

Ex.1

$$Y = \alpha + \beta X$$

$$\beta = \left(X^T X\right)^{-1} X^T Y$$

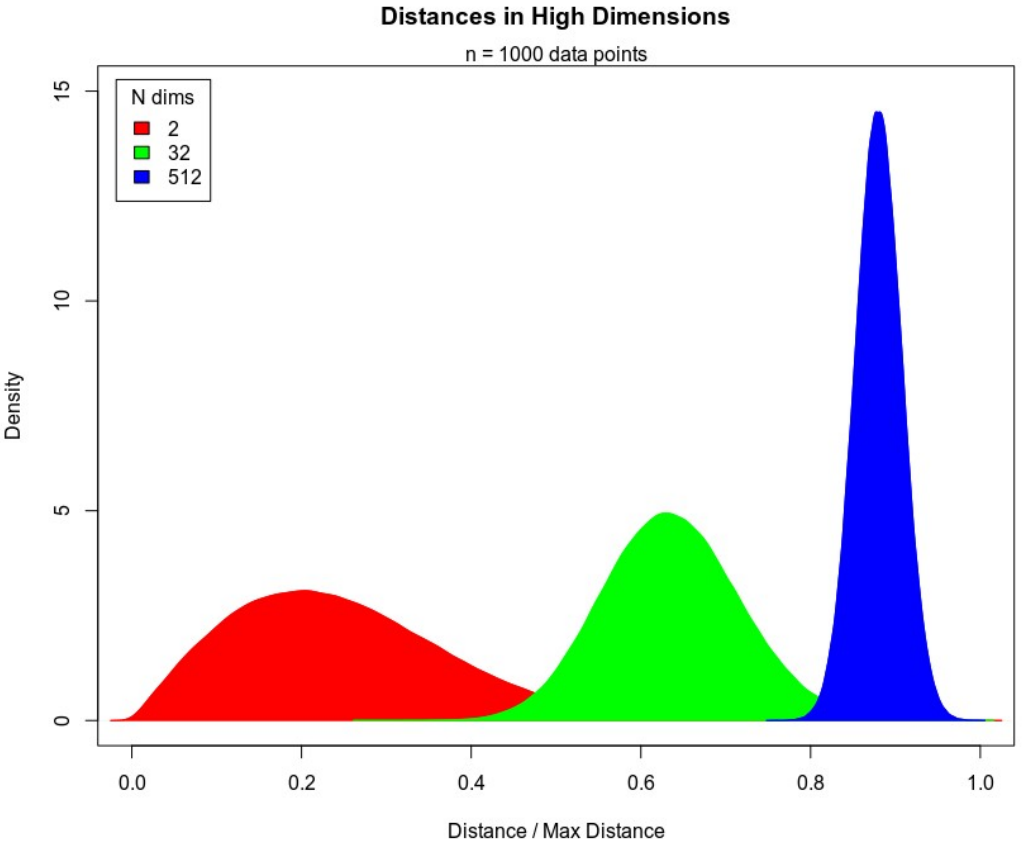
$$\left(X^T X\right)^{-1} \sim \frac{1}{\det\left(X^T X\right)} \cdots \to \infty, \quad n << p$$

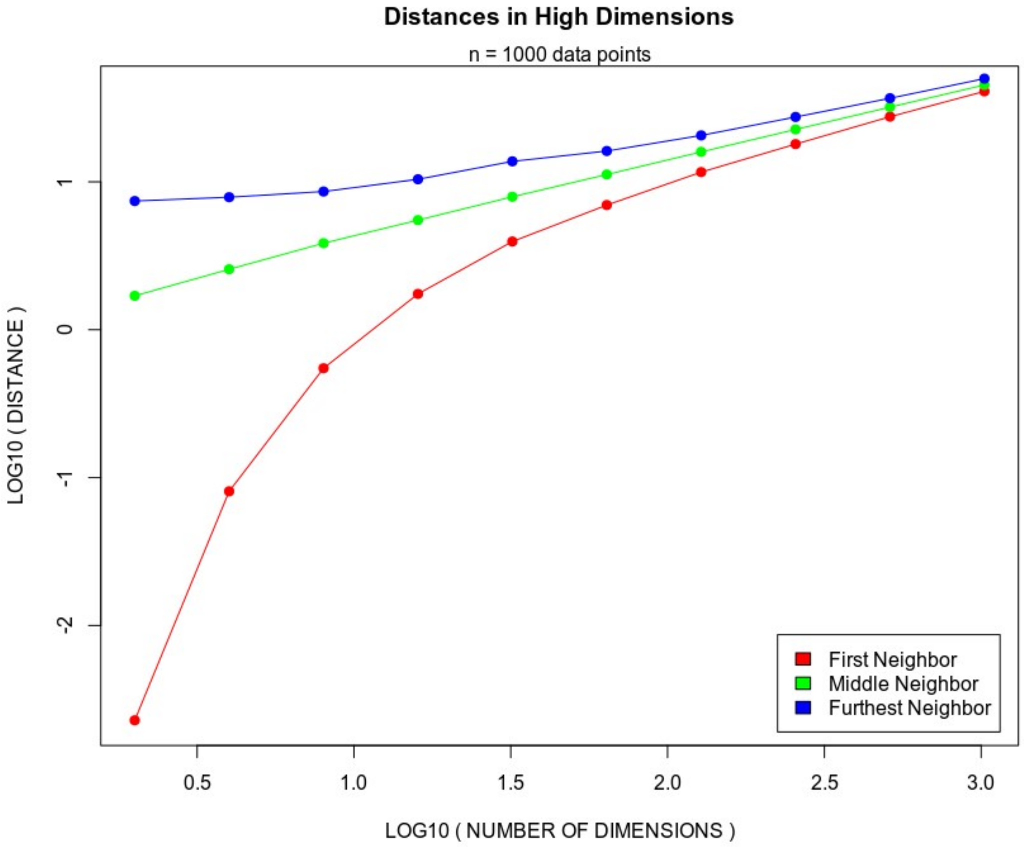
Ex.2

$$E[\hat{\sigma}^2] = \frac{n-p}{n}\sigma^2$$

Biased ML variance estimator in HD-space

Equidistant Points in High Dimensions



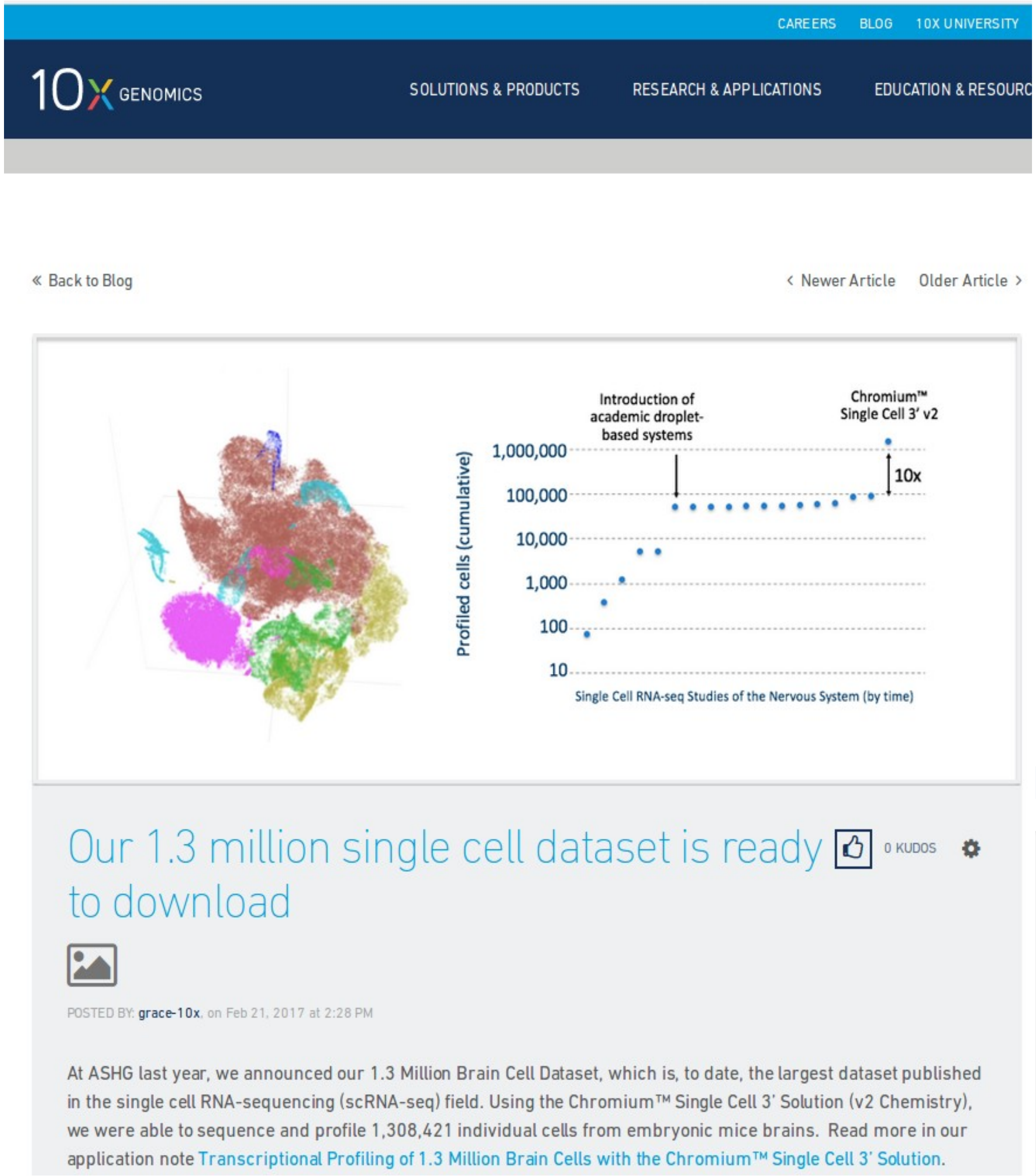


Data points become far from each other and equidistant from each other in high dimensions

The differences between closest and furthest data point neighbours disappears in high-dimensional spaces – can’t cluster

In high-dimensional space we can not separate cases and controls any more

Big Data in Single Cell



Our 1.3 million single cell dataset is ready to download

POSTED BY: grace-10x, on Feb 21, 2017 at 2:28 PM

At ASHG last year, we announced our 1.3 Million Brain Cell Dataset, which is, to date, the largest dataset published in the single cell RNA-sequencing (scRNA-seq) field. Using the Chromium™ Single Cell 3' Solution (v2 Chemistry), we were able to sequence and profile 1,308,421 individual cells from embryonic mice brains. Read more in our application note Transcriptional Profiling of 1.3 Million Brain Cells with the Chromium™ Single Cell 3' Solution.

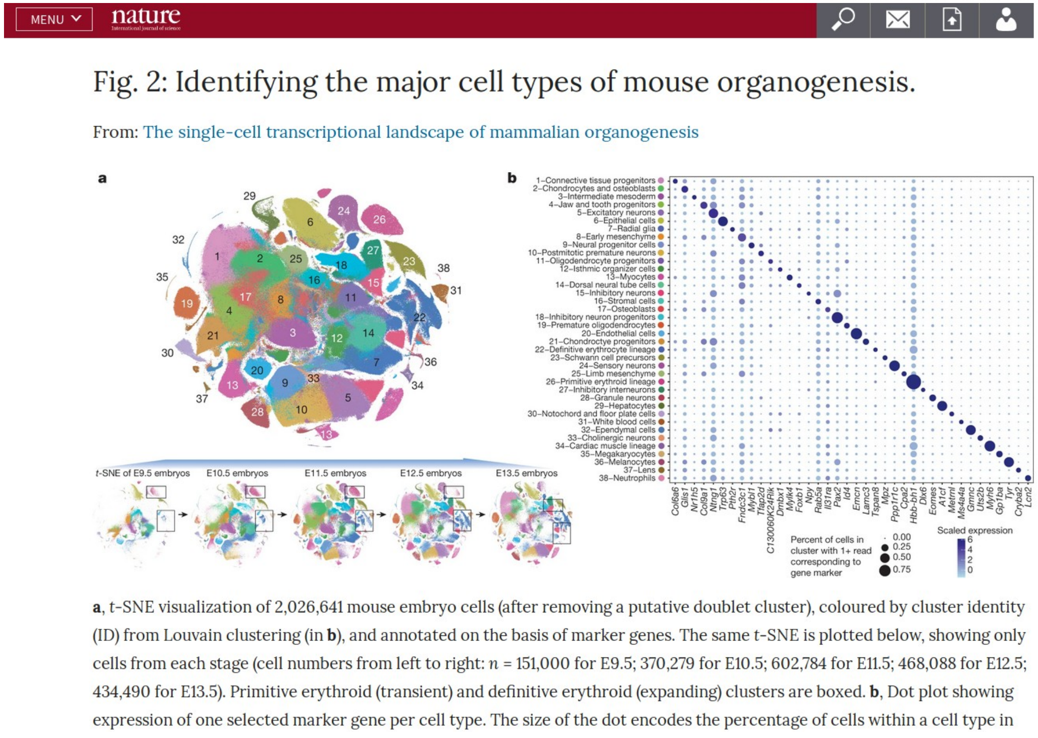
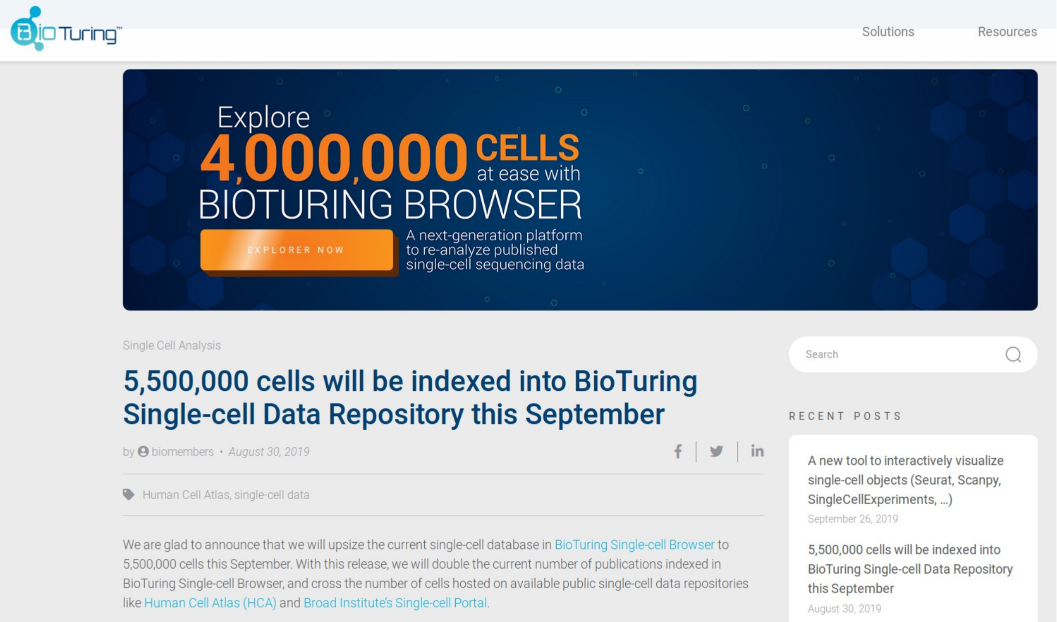


Fig. 2: Identifying the major cell types of mouse organogenesis.

From: The single-cell transcriptional landscape of mammalian organogenesis

a, *t*-SNE visualization of 2,026,641 mouse embryo cells (after removing a putative doublet cluster), coloured by cluster identity (ID) from Louvain clustering (in **b**), and annotated on the basis of marker genes. The same *t*-SNE is plotted below, showing only cells from each stage (cell numbers from left to right: *n* = 151,000 for E9.5; 370,279 for E10.5; 602,784 for E11.5; 468,088 for E12.5; 434,490 for E13.5). Primitive erythroid (transient) and definitive erythroid (expanding) clusters are boxed. **b**, Dot plot showing expression of one selected marker gene per cell type. The size of the dot encodes the percentage of cells within a cell type in



Single Cell Analysis

5,500,000 cells will be indexed into BioTuring Single-cell Data Repository this September

by biomembers • *August 30, 2019*

Human Cell Atlas, single-cell data

We are glad to announce that we will upsize the current single-cell database in BioTuring Single-cell Browser to 5,500,000 cells this September. With this release, we will double the current number of publications indexed in BioTuring Single-cell Browser, and cross the number of cells hosted on available public single-cell data repositories like Human Cell Atlas (HCA) and Broad Institute's Single-cell Portal.

A new tool to interactively visualize single-cell objects (Seurat, Scanpy, SingleCellExperiments, ...)
September 26, 2019

5,500,000 cells will be indexed into BioTuring Single-cell Data Repository this September
August 30, 2019

Watch out Underfitting!
Paradise for Deep Learning!

How to define and evaluate OMICs Integration?

What I Mean by OMICs Integration

mixOmics

Exploration and Integration of Omics datasets

OnPLS

JIVE

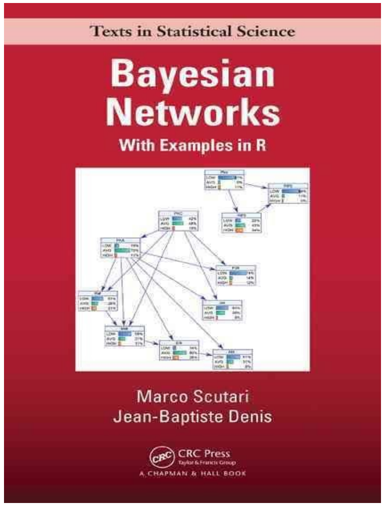
DISCO

Clustering of Clusters

SOFTWARE UPDATES

SIMILARITY NETWORK FUSION (SNF)

As of Jan 26, 2014 our paper is online at Nature Methods



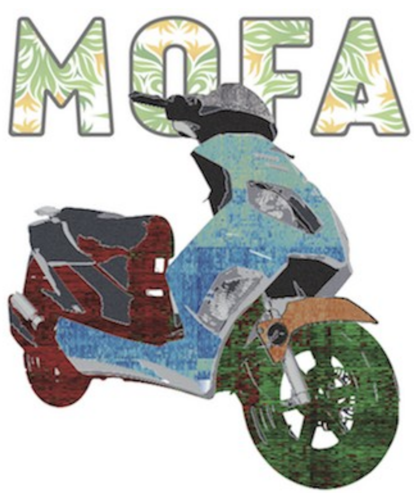
Texts in Statistical Science

Bayesian Networks

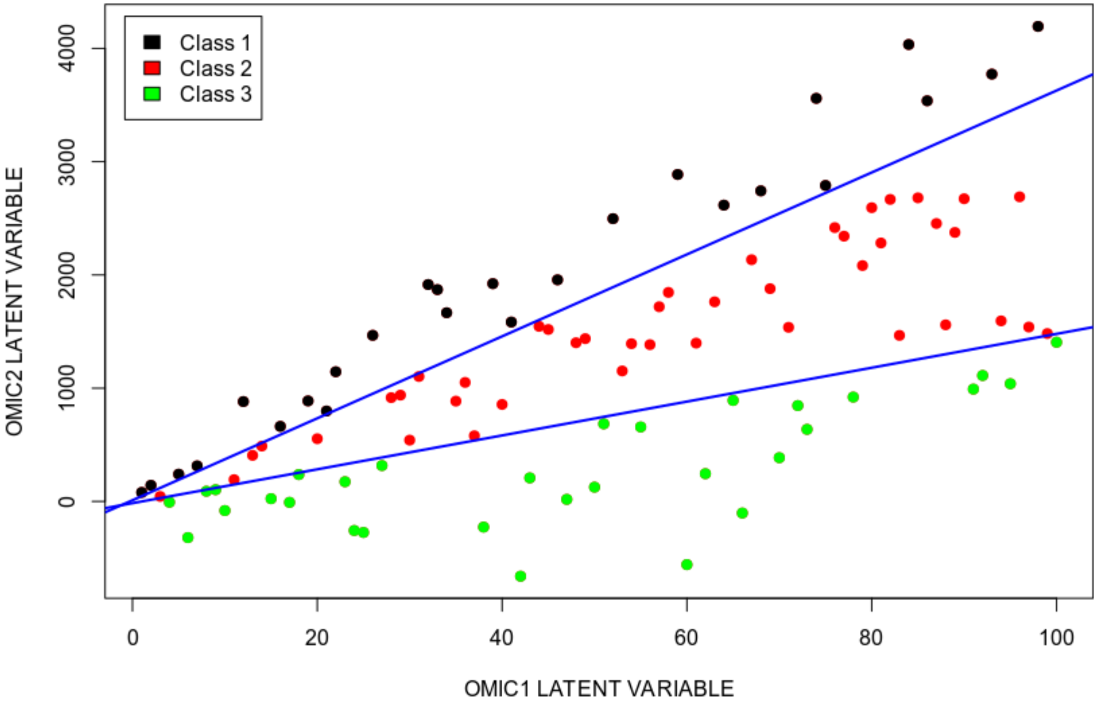
With Examples in R

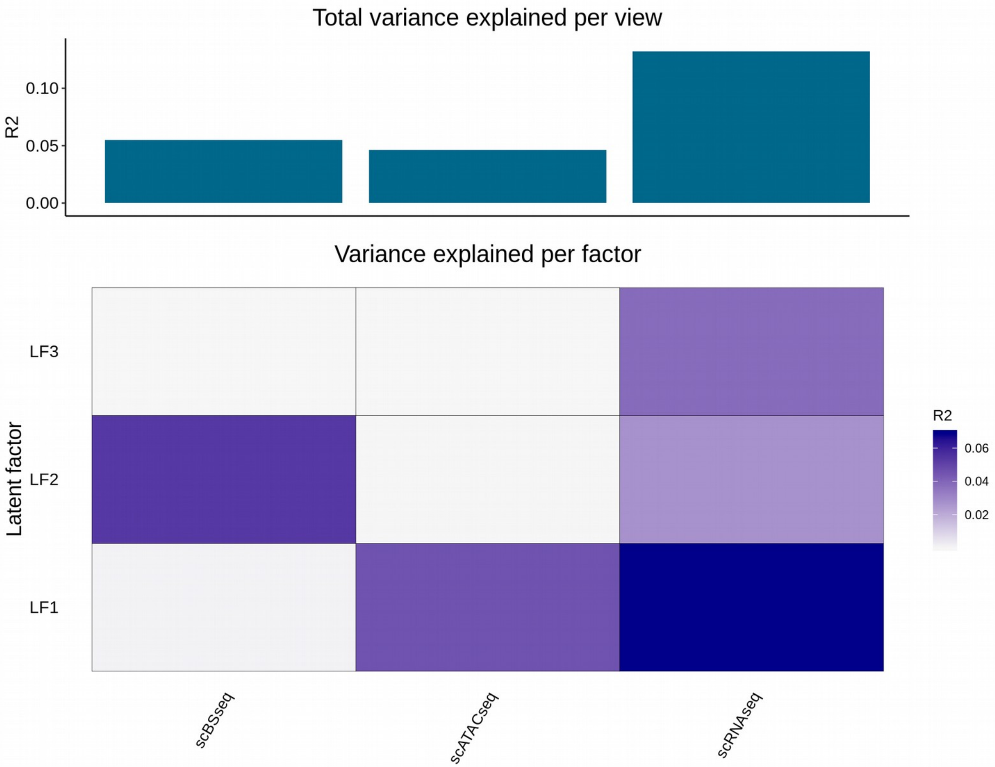
Marco Scutari
Jean-Baptiste Denis

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK



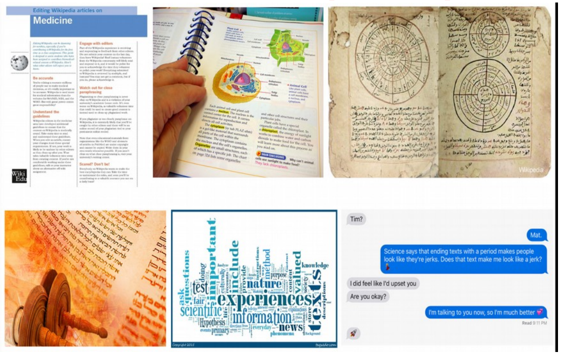
Idea Behind OMICs Integration: See Patterns Hidden in Individual OMICS





How I Evaluate OMICs Integration, Data Science: Boost in Prediction

TEXT (78%)



Tim?

Mat.

Science says that ending texts with a period makes people look like they're jerks. Does that text make me look like a jerk?

I did feel like I'd upset you

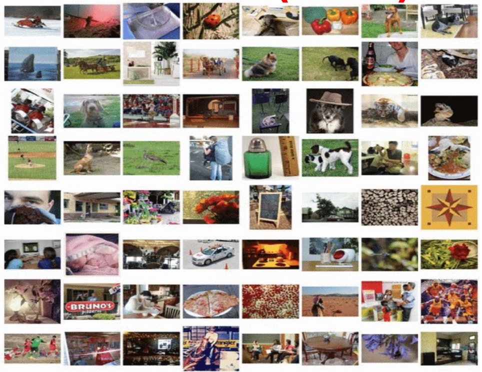
Are you okay?

I'm talking to you now, so I'm much better 💕

Read 9:11 PM

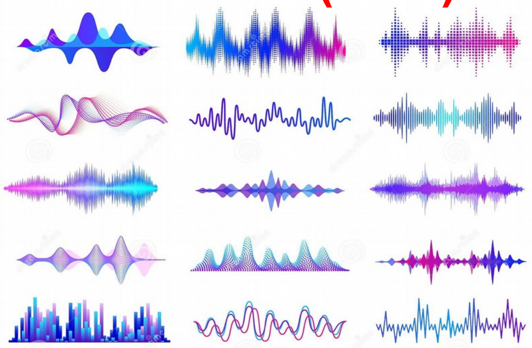
	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-3	-2	0
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

IMAGE (83%)



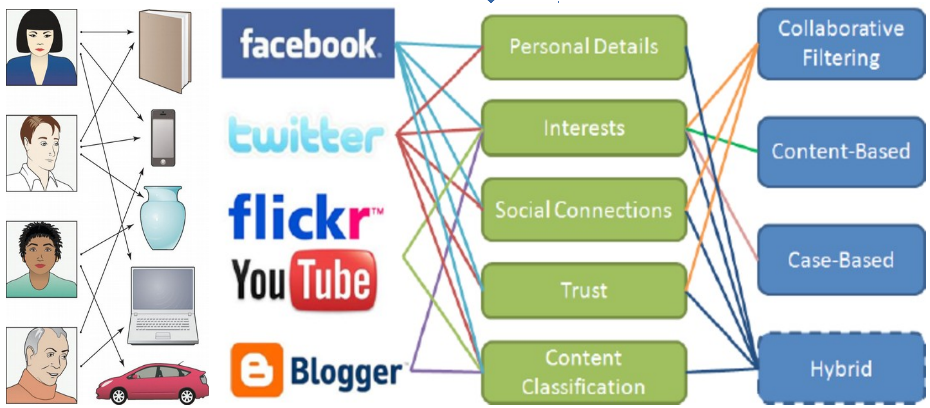
	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-3	-2	0
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

SOUND (75%)



	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-3	-2	0
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

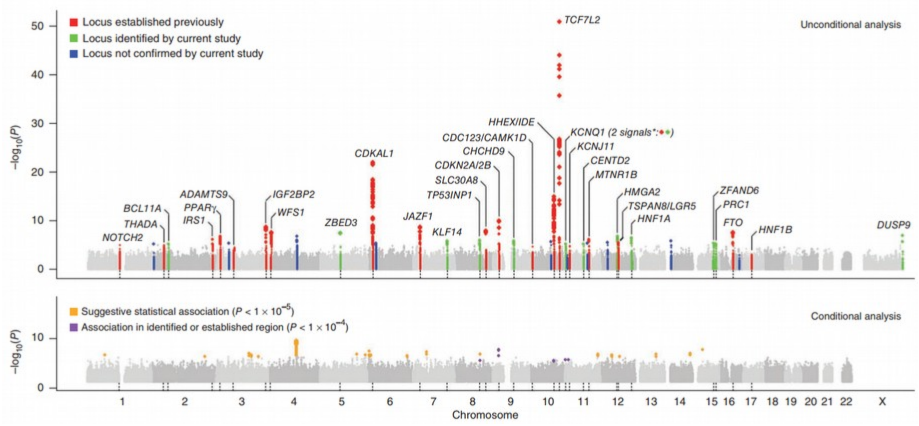
Predict Facebook user interests



Data Integration Accuracy: 96%

Prediction is an Ultimate Criterion of Successful OMICS Integration

Statistics searches for candidates

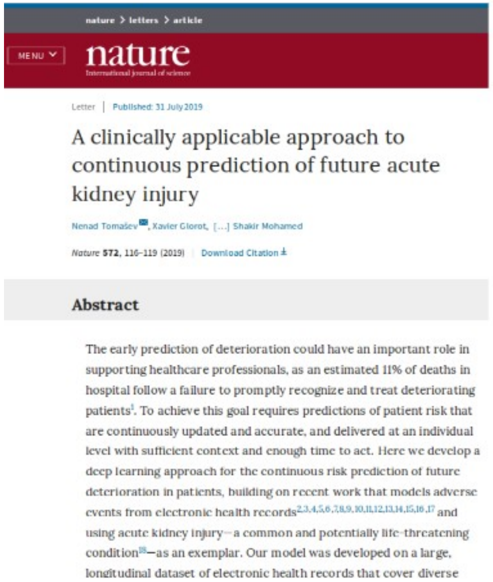


Consequence

The case of the missing heritability

B. Maher, Nature 456, 18-21 (2008)

Machine Learning optimizes prediction



Letter | Published: 31 July 2019

A clinically applicable approach to continuous prediction of future acute kidney injury

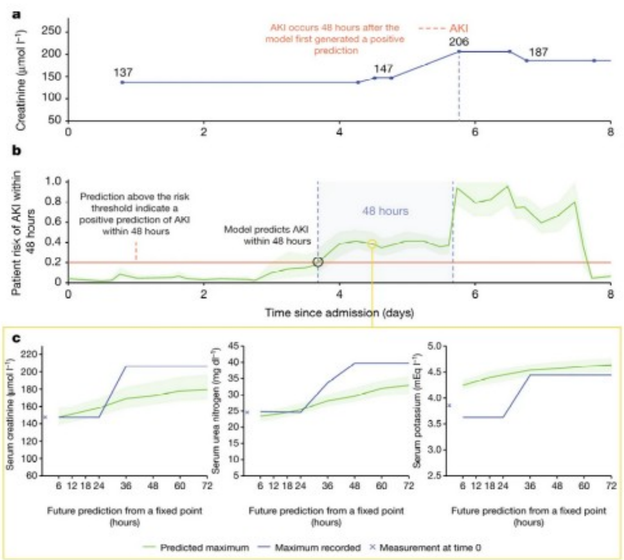
Nenad Tomašev, Xavier Glorot, […] Shakir Mohamed

Nature **572**, 116–119 (2019) | Download Citation

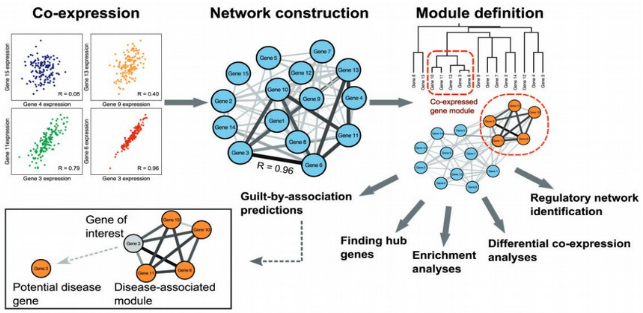
Abstract

The early prediction of deterioration could have an important role in supporting healthcare professionals, as an estimated 11% of deaths in hospital follow a failure to promptly recognize and treat deteriorating patients[1]. To achieve this goal requires predictions of patient risk that are continuously updated and accurate, and delivered at an individual level with sufficient context and enough time to act. Here we develop a deep learning approach for the continuous risk prediction of future deterioration in patients, building on recent work that models adverse events from electronic health records[2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17] and using acute kidney injury—a common and potentially life-threatening condition[18]—as an exemplar. Our model was developed on a large, longitudinal dataset of electronic health records that cover diverse

From: A clinically applicable approach to continuous prediction of future acute kidney injury

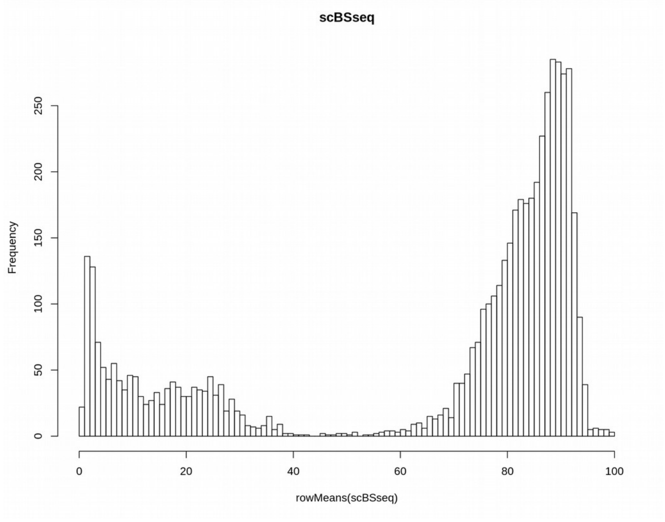


Consequence

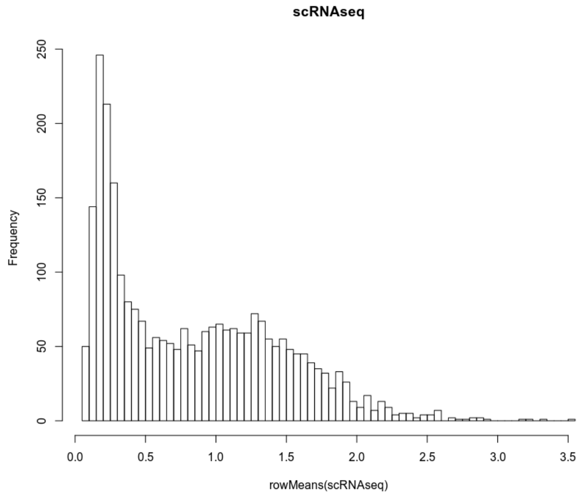


How I Evaluate OMICs Integration, Life Sciences: Boost in Prediction

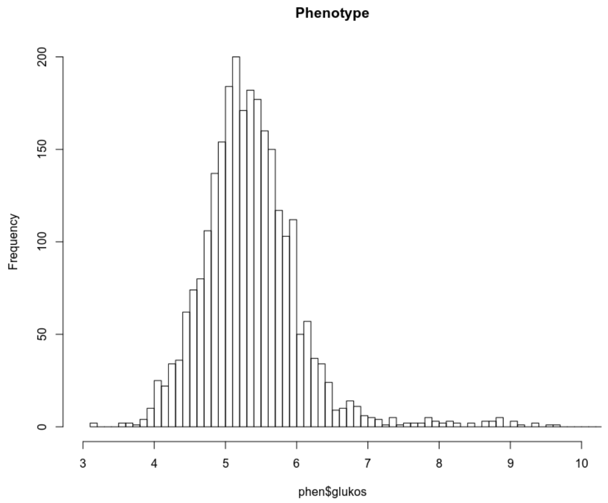
Methylation (78%)

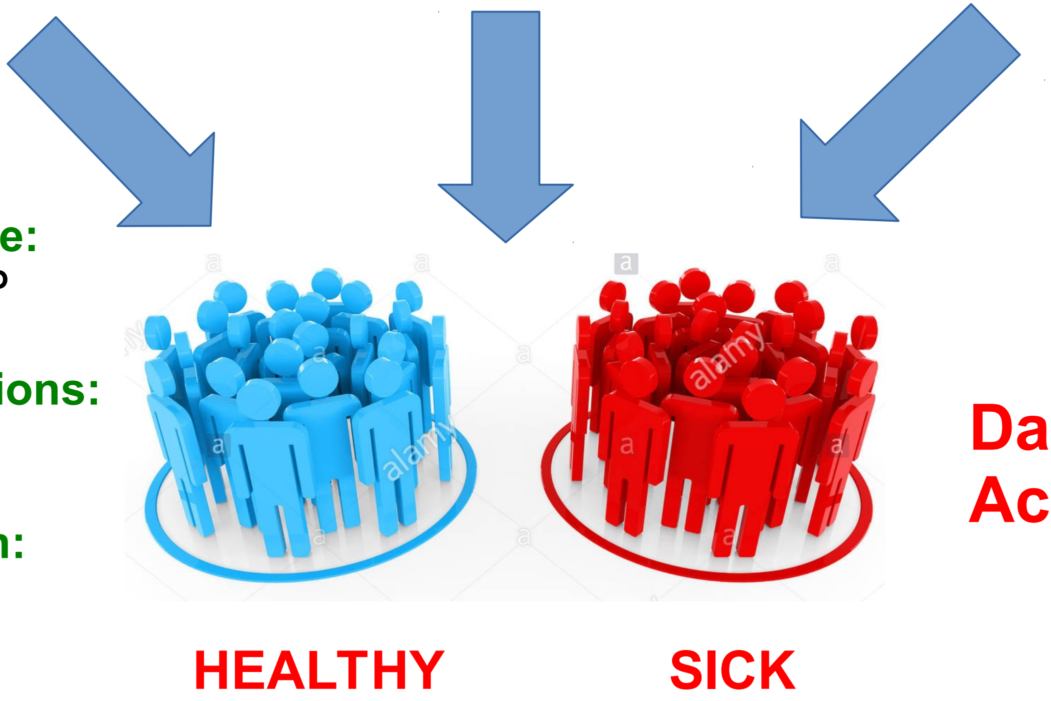


Gene Expression (83%)



Phenotype (75%)





1) Convert to common space:
Neural Networks, SNF, UMAP

2) Explicitly model distributions:
MOFA, Bayesian Networks

3) Extract common variation:
PLS, CCA, Factor Analysis

Data Integration Accuracy: 96%

NBIS

National Bioinformatics Infrastructure Sweden (NBIS)

SciLifeLab

Knut och Alice Wallenbergs Stiftelse

Vetenskapsrådet

LUNDS UNIVERSITET