Deep Aging Clocks

Dmytro Mishchenko

Leading causes of death
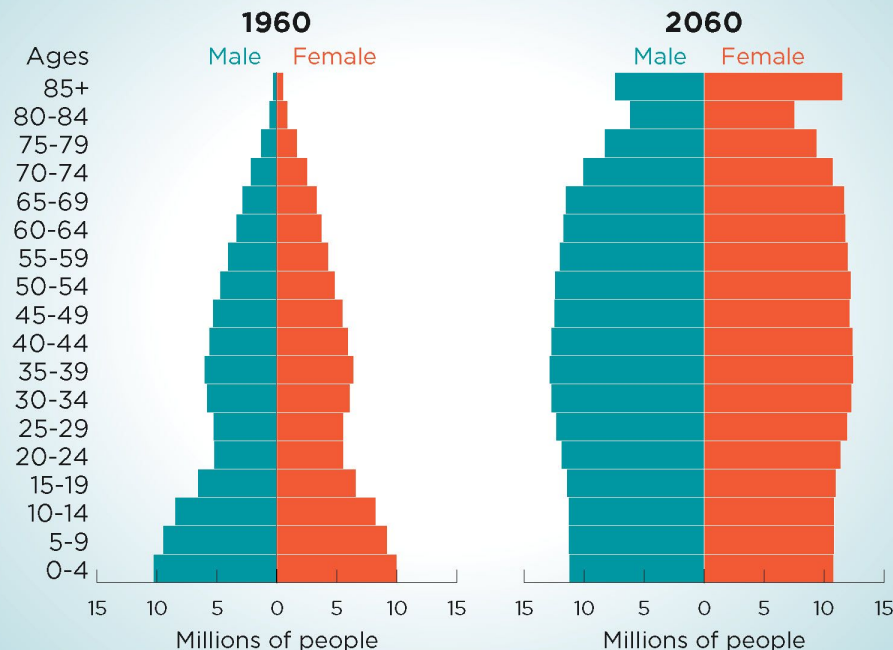World, 2016

Cause	Percentage
Cardiovascular diseases	32.3%
Cancers	16.3%
Respiratory diseases	6.5%
Diabetes	5.8%
Lower respiratory infections	4.4%
Dementia	4.4%
Neonatal deaths	3.2%
Diarrhoeal diseases	3%
Road injuries	2.5%
Liver disease	2.3%

BBC
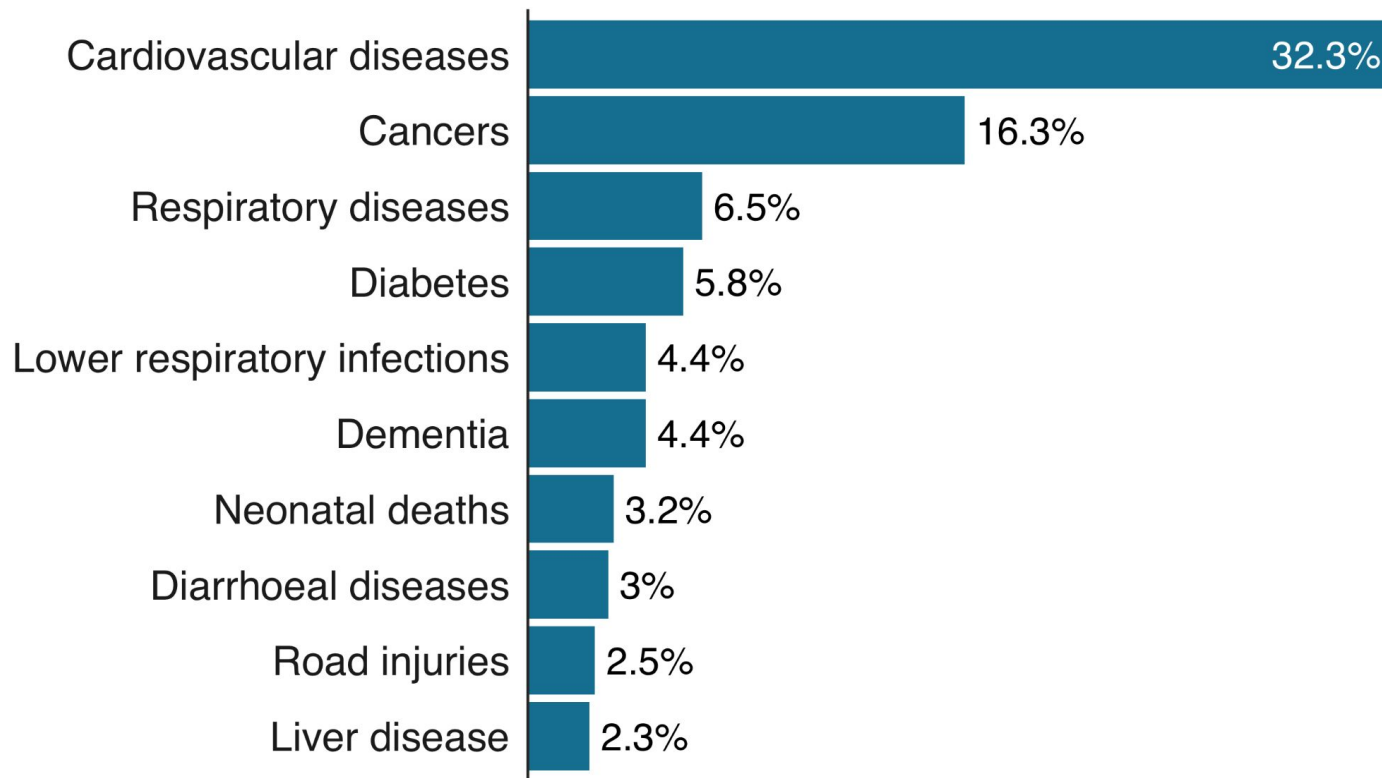
"**Preventive**, **geroscience-based measures** to treat aging at the organismal level would provide a more substantial benefit than reactive therapeutic approaches targeted at a single disease or organ, since those **do not contribute to a significant improvement in healthspan**."

From Pyramid to Pillar: A Century of Change

Population of the United States

1960
Male   Female

2060
Male   Female

Ages
85+
80-84
75-79
70-74
65-69
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
20-24
15-19
10-14
5-9
0-4

15 10 5 0 5 10 15
Millions of people

15 10 5 0 5 10 15
Millions of people

United States Census Bureau
census.gov

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU

Source: National Population Projections, 2017
www.census.gov/programs-surveys/popproj.html

Sources: "Trends in Causes of Death among Older Persons in the US"
https://www.cdc.gov/nchs/data/ahcd/agingtrends/06olderpersons.pdf
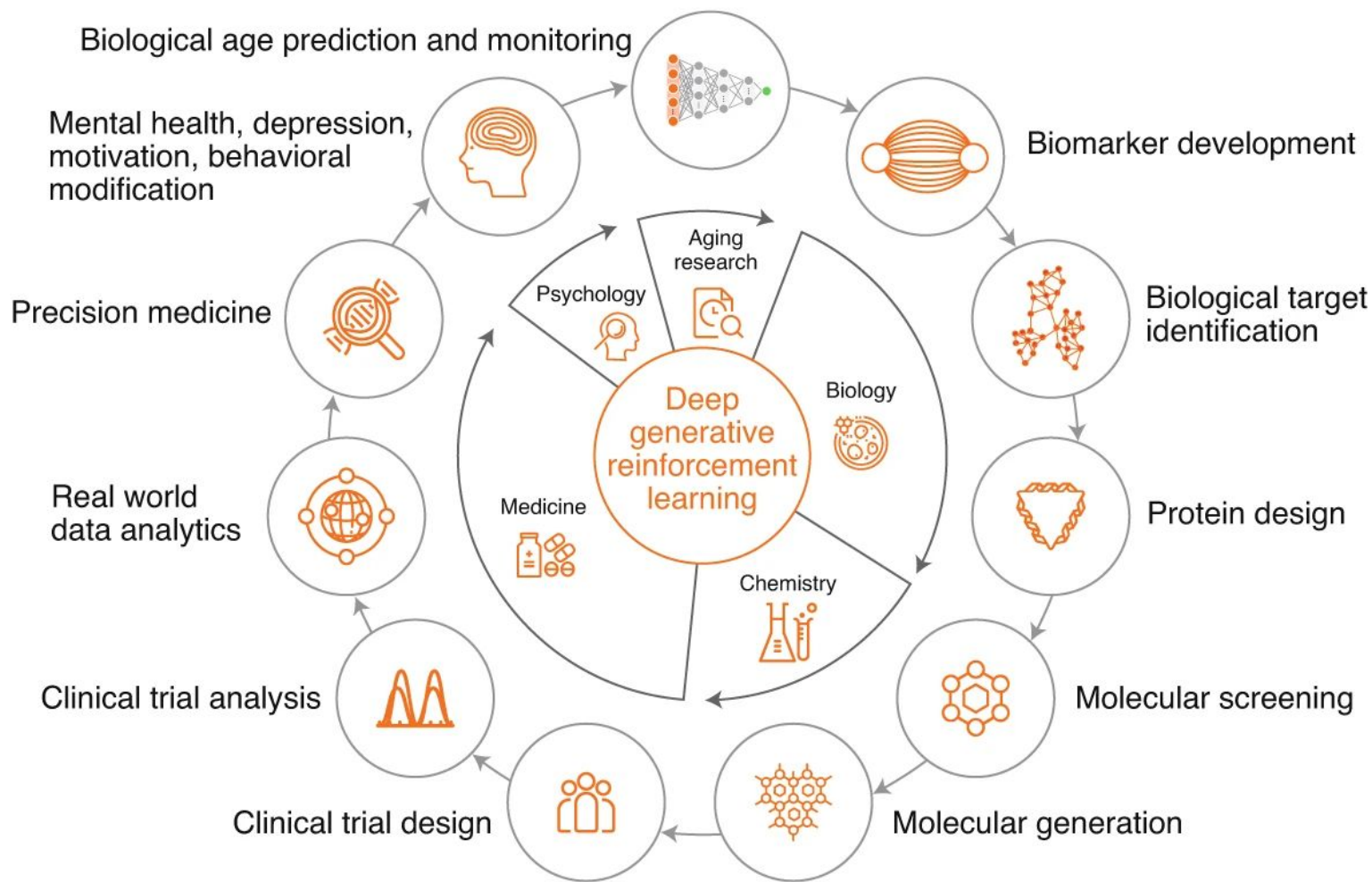Zhavoronkov, A., Bischof, E. & Lee, KF. Artificial intelligence in longevity medicine. *Nat Aging* 1, 5–7 (2021).
https://doi.org/10.1038/s43587-020-00020-4
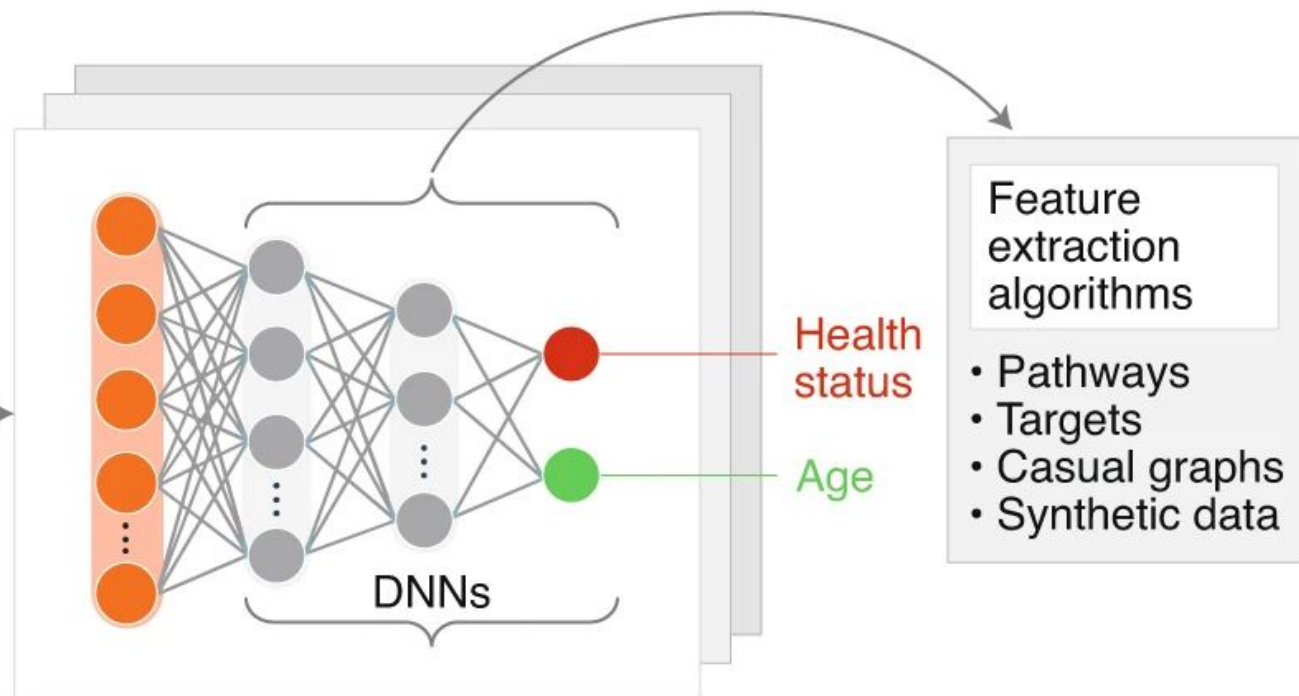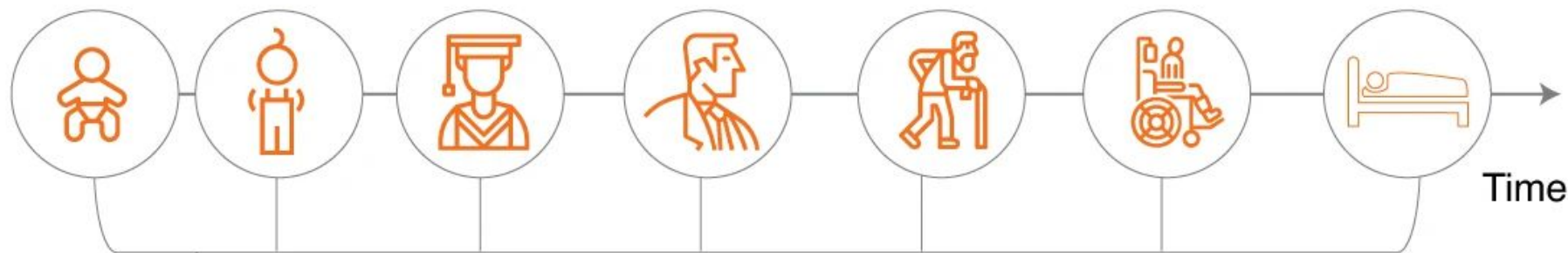
"It has been estimated that the complete elimination of a single fatal disease such as cancer in the USA would merely **lead to a 2.3-year population increase in life expectancy** [...] since the majority of overall mortality is due to age-related diseases."

Source: Zhavoronkov, A., Bischof, E. & Lee, KF. Artificial intelligence in longevity medicine. *Nat Aging* 1, 5–7 (2021). https://doi.org/10.1038/s43587-020-00020-4

Biological age prediction and monitoring

Mental health, depression, motivation, behavioral modification

Precision medicine

Real world data analytics

Clinical trial analysis

Clinical trial design

Molecular generation

Molecular screening

Protein design

Biological target identification

Biomarker development

Aging research

Psychology

Biology

Medicine

Chemistry

Deep generative reinforcement learning

Time

Genome
Transcriptome
Proteome
Methylome
Imaging
Pictures and videos
EEG/ECG
Activity
Blood tests
Behavior

Health status

Age

DNNs

Feature extraction algorithms

- Pathways
- Targets
- Casual graphs
- Synthetic data

Epigenomics

Chromosome

DNA Inaccessible
Gene Turned Off

Histone Tail

Histone

Chemical Tag

DNA

DNA Accessible
Gene Turned On

NIH National Human Genome
Research Institute
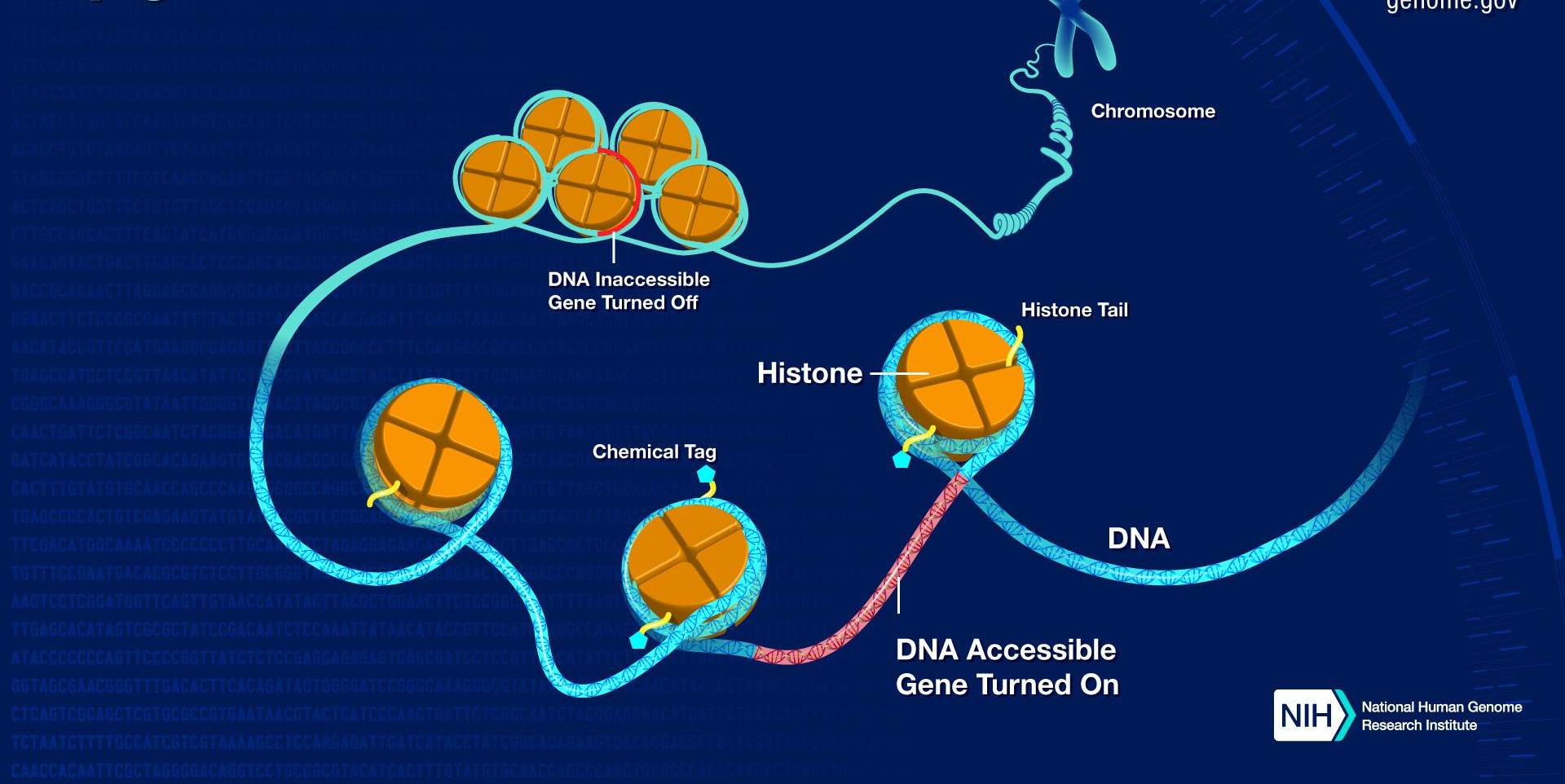
Table 1. Sixteen healthy DNA-methylation datasets.

| DNA Origin	Platform	No.	Age Range	Author and Publication Year	Availability
Whole Blood	27K	93	(49, 74)	Rakyan (2010)	GSE20236
Blood CD4+CD14	27K	50	(16, 69)	Rakyan (2010)	GSE20242
Blood PBMC 1	27K	398	(3.6, 18)	Alisch (2012)	GSE27097
Blood Cord	27K	168	(0, 0)	Adkins (2011)	GSE27317
Blood PBMC	450K	40	(0, 103)	Heyn (2012)	GSE30870
Blood PBMC	450K	71	(3.5, 76)	Harretal (2012)	GSE32149
Blood Cord	27K	84	(0, 0)	Khulan (2012)	GSE34257
Blood Cord	27K	24	(0, 0)	Mallon (2012)	GSE34869
Blood PBMC	450K	78	(1, 16)	Alisch (2012)	GSE36064
Blood Cord	27K	123	(0, 0)	Gordon (2012)	GSE36642
Blood Cord	27K	48	(0, 0)	Turan (2012)	GSE36812
Blood PBMC	27K	91	(24, 45)	Lam (2012)	GSE37008
Whole Blood	450K	500	(26, 101)	Hannum (2012)	GSE40279
Whole Blood	450K	95	(18, 65)	Horvath (2012)	GSE41169
Whole blood	450K	43	(47, 59)	Bell (2013)	GSE53128
Blood	450K	16	(21, 32)	Xu (2015)	GSE65638

[1] Peripheral blood mononuclear cell.

**Table 3.** Information of 6 selected age-related CpG sites.

CpG ID	Gene ID	Chromosome Location [1]	Gene Region [2]	Relation to GpG Island [3]	Correlation Status	Reference
cg09809672	EDARADD	1:236557682	TSS1500	N_Shore	Negative	[1,17,33]
cg22736354	NHLRC1	6:18122719	1stExon	Island	Positive	[2,7,18,19]
cg02228185	ASPA	17:3379567	1stExon	--	Negative	[7,26,33]
cg01820374	LAG3	12:6882083	Body	N_Shore	Negative	[1]
cg06493994	SCGN	6:25652602	1stExon	Island	Positive	[2,7,18,19]
cg19761273	CSNK1D	17:80232096	TSS1500	S_Shore	Negative	[2]

[1] Chromosome location is referred to the Human genome reference GRCh37 version. [2] TSS: transcription start site. TSS1500: 1500 bp flanking region from the TSS. [3] CpGs island table were downloaded from University of California Santa Cruz (UCSC) browser. Distance of 2kb to CpG islands were defined as CpG island shores (N_Shore: downstream of CpG island and S_Shore: up-stream of the CpG island).

	...1	cg09809672	cg22736354	cg02228185	cg01820374	cg06493994	cg19761273	sex	age
1									
2	GSM507152	0.392464194	0.281630294	0.352590915	0.315924458	0.111603714	0.211683062	1	71.06
3	GSM507153	0.37791864	0.238899924	0.520396483	0.323641214	0.127964701	0.184306617	1	69.65
4	GSM507154	0.458521656	0.204077992	0.519273008	0.363714467	0.119891309	0.215066962	1	65.79
5	GSM507155	0.479107642	0.154893332	0.618859411	0.493972405	0.088941341	0.240057105	1	49.39
6	GSM507156	0.418762861	0.217831525	0.561005099	0.398579533	0.085992475	0.207462482	1	60.68
7	GSM507157	0.389851147	0.237511488	0.33308652	0.371843478	0.193483015	0.211772313	1	68.2
8	GSM507158	0.472550037	0.191580972	0.623615165	0.353173498	0.102075645	0.242547404	1	52.13
9	GSM507159	0.389721269	0.174475213	0.554374291	0.290877514	0.102095865	0.196526801	1	71.82
10	GSM507160	0.509076429	0.137656154	0.540985557	0.391201823	0.055698215	0.19982747	1	59.74
11	GSM507161	0.415323323	0.220878433	0.512730113	0.368857906	0.122601447	0.197427766	1	66.74
12	GSM507162	0.458253361	0.205556408	0.516113988	0.425091452	0.069954327	0.180355455	1	64.45
13	GSM507163	0.378139053	0.221221663	0.438043288	0.416106889	0.075606976	0.165156105	1	71.69
14	GSM507164	0.450666073	0.218492844	0.56306017	0.394534391	0.097795844	0.227327603	1	56.12
15	GSM507165	0.361932444	0.180163296	0.568846789	0.364857313	0.088021561	0.224786447	1	66.52
16	GSM507166	0.415051338	0.196526801	0.571111441	0.516113988	0.086964606	0.226468761	1	49.39
17	GSM507167	0.41797811	0.204920542	0.510767292	0.432834777	0.089674574	0.2220192	1	63.09
18	GSM507168	0.481087159	0.203858331	0.574735703	0.373937273	0.110983503	0.262777449	1	52.9
19	GSM507169	0.388601332	0.22178402	0.598611078	0.408912284	0.137848401	0.209589263	1	67.37
20	GSM507170	0.331986813	0.229113554	0.508684599	0.300439871	0.108993605	0.227225861	1	72.58
21	GSM507171	0.318446799	0.183101539	0.478976025	0.387441782	0.09408223	0.205556408	1	62.51
22	GSM507172	0.31173124	0.206780208	0.528639914	0.331649357	0.101966354	0.19035596	1	65.53
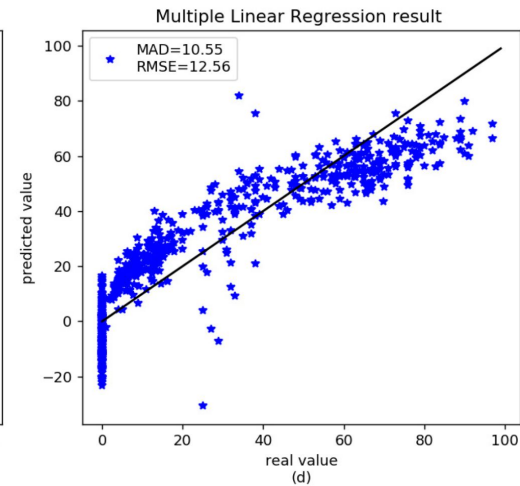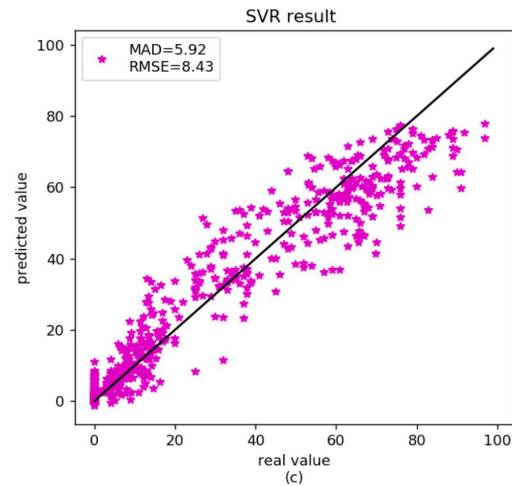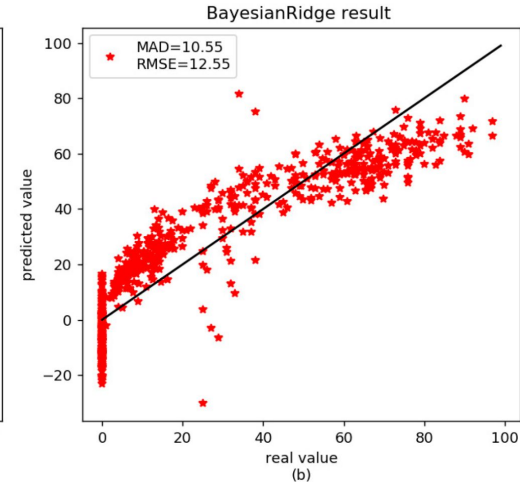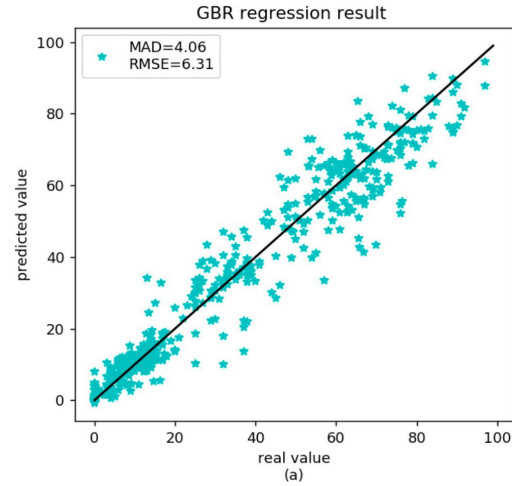
Source: Li, X.; Li, W.; Xu, Y. Human Age Prediction Based on DNA Methylation Using a Gradient Boosting Regressor. *Genes* **2018**, *9*, 424. https://doi.org/10.3390/genes9090424
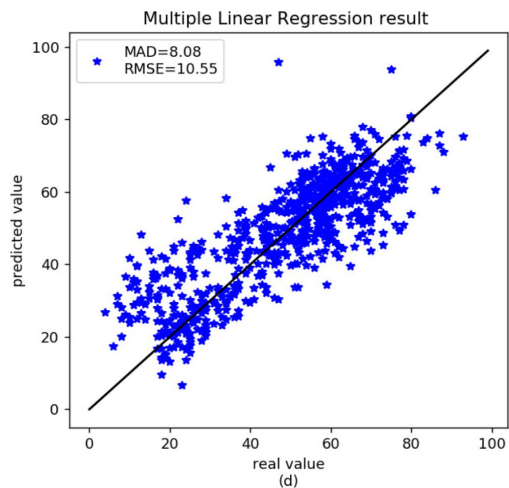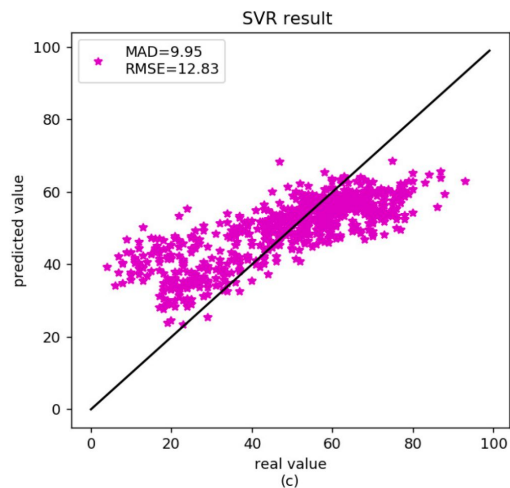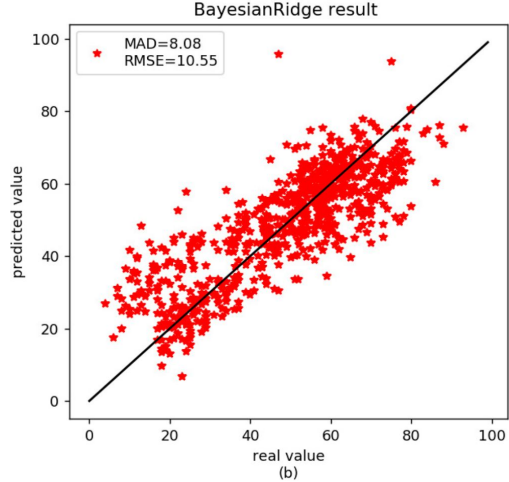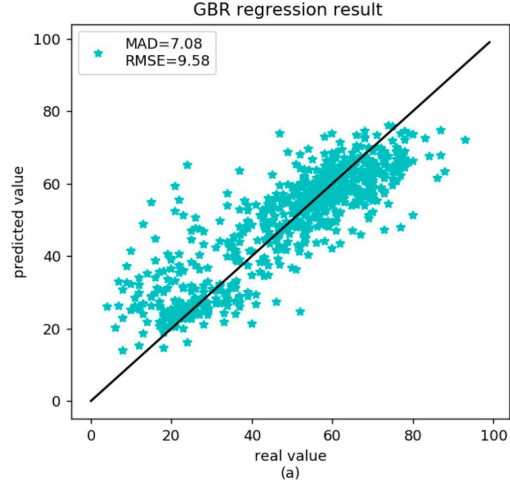
Source: Li, X.; Li, W.; Xu, Y. Human Age Prediction Based on DNA Methylation Using a Gradient Boosting Regressor. *Genes* **2018**, *9*, 424. https://doi.org/10.3390/genes9090424
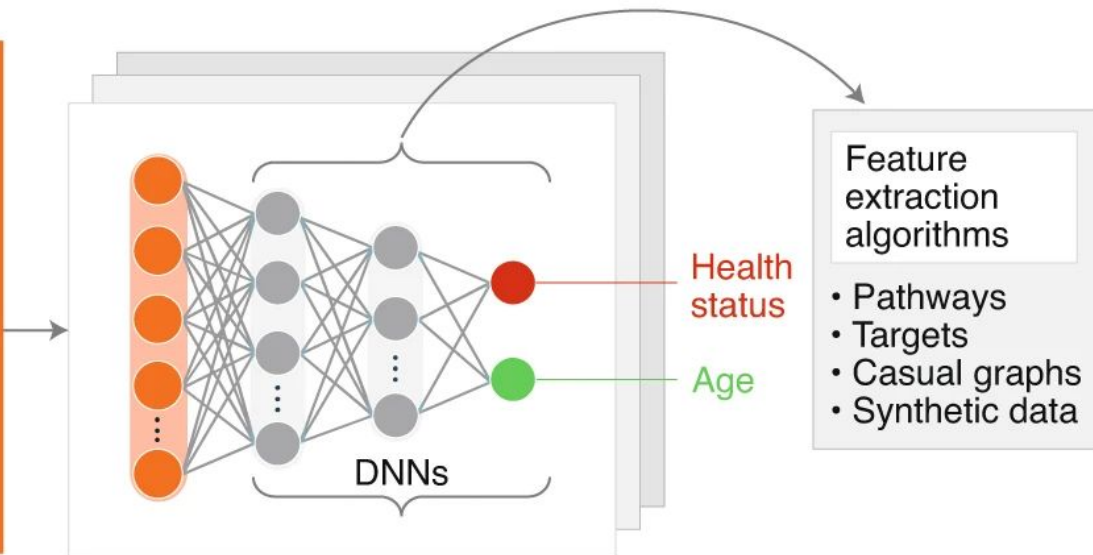
Table 4. Comparison of gradient booster regressor (GBR) with the other three methods on healthy datasets.

	R^2	MAD	MSE	RMSE
Training				
Gradient Boosting Regressor	0.9747	2.7171	20.7243	4.5524
BayesianRidge	0.8055	10.2561	158.3044	12.5819
Support Vector Regression	0.9267	5.1338	60.0420	7.7487
Multiple Linear Regression	0.8055	10.2448	158.2800	12.5809
Testing				
Gradient Boosting Regressor	0.9523	4.0593	39.8269	6.3109
BayesianRidge	0.8101	10.5654	157.8721	12.5647
Support Vector Regression	0.9151	5.9267	71.2060	8.4384
Multiple Linear Regression	0.8104	10.5510	157.6726	12.5568
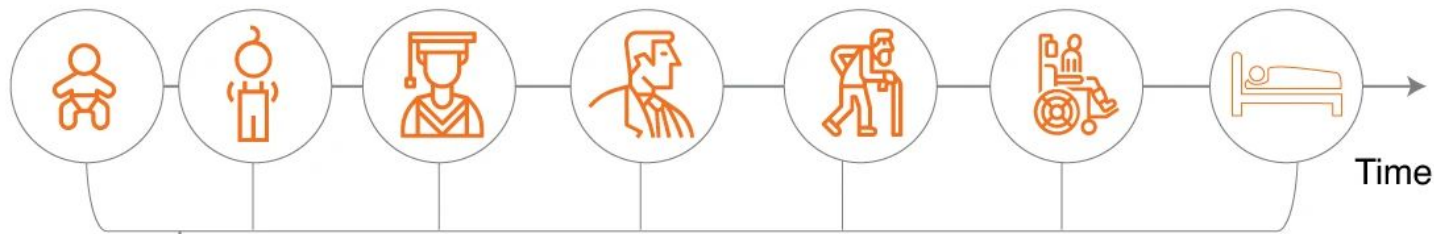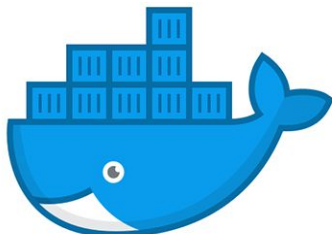
Time

Genome
Transcriptome
Proteome
Methylome
Imaging
Pictures and videos
EEG/ECG
Activity
Blood tests
Behavior

Health status

Age

DNNs

Feature extraction algorithms

• Pathways
• Targets
• Casual graphs
• Synthetic data

Thank you for your attention

Discord nickname: Dmitry
Mail: mdmittriy@gmail.com