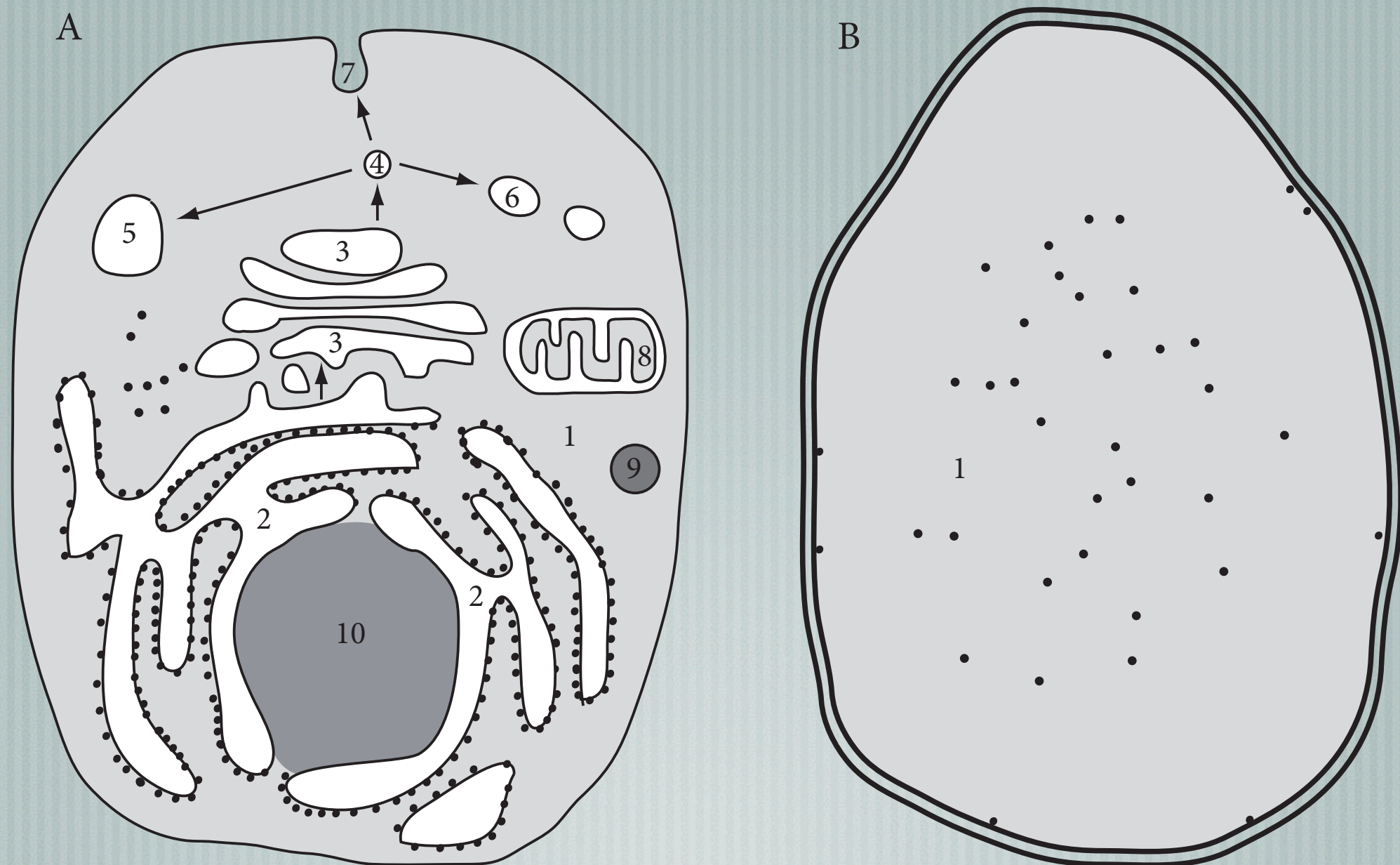


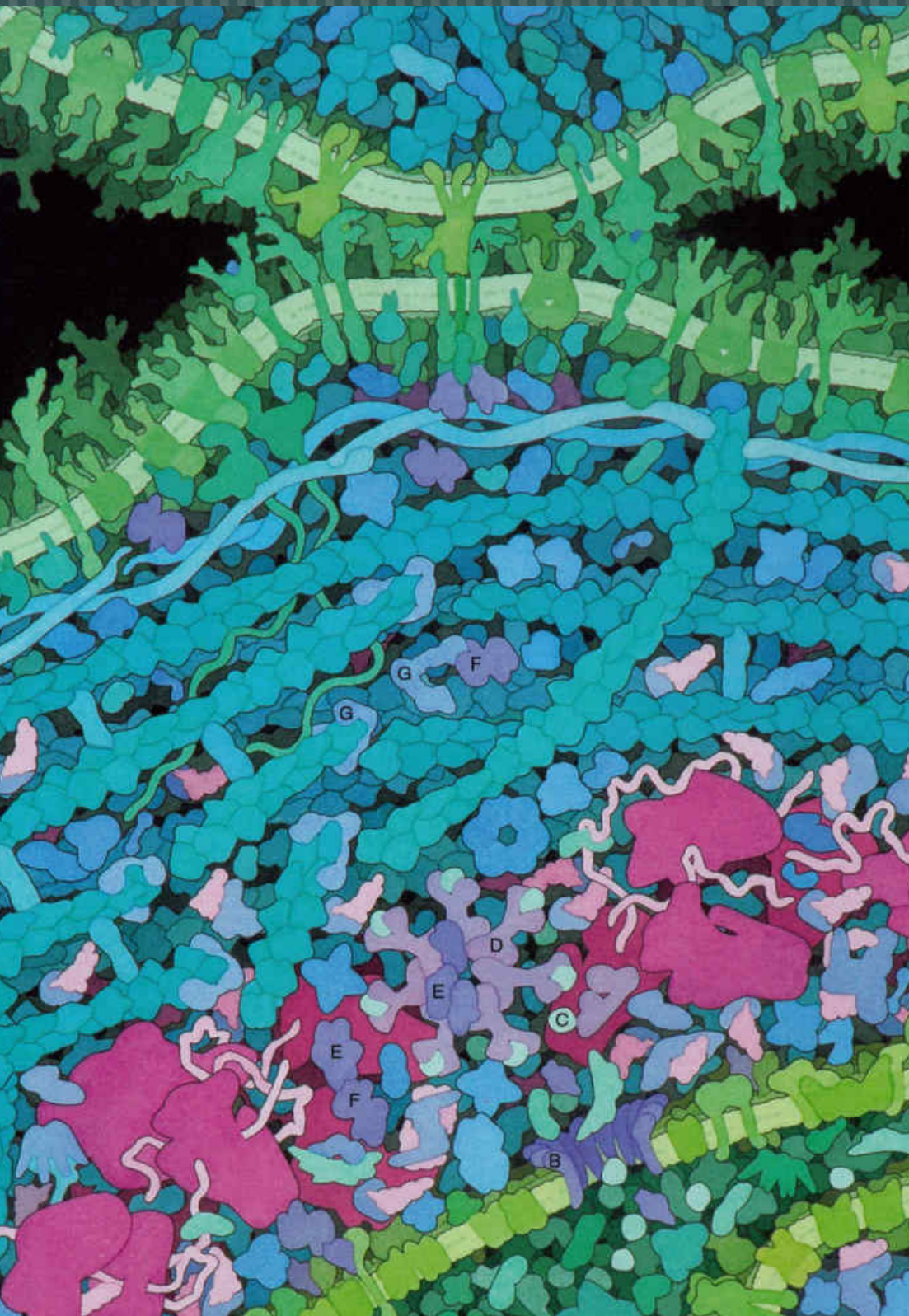
ANALYSIS OF DATA FROM HIGH-
THROUGHPUT MOLECULAR BIOLOGY
EXPERIMENTS

LARS ARVESTAD
OLOF EMANUELSSON
LUKAS KÄLL

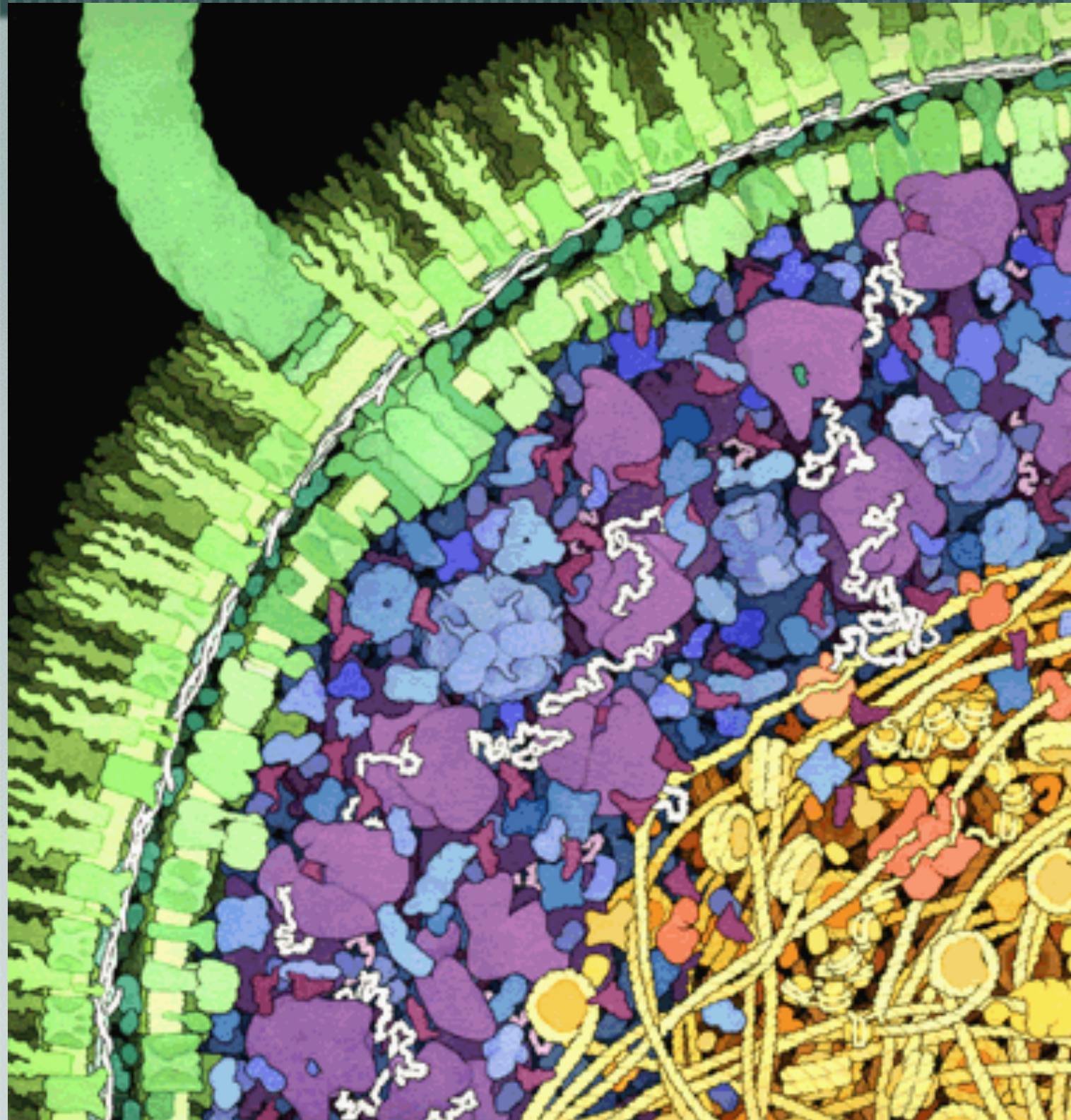
I used to have a naïve view of Cells



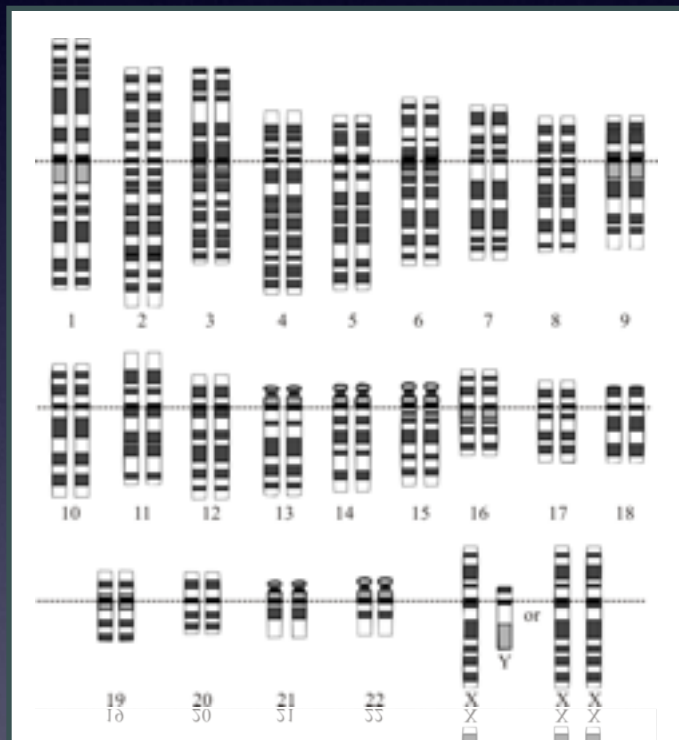
More realistic sketches of Cells



[David S. Goodsell, The machinery of Life]



ConcepTest 1: How many kinds of protein-species are there in the human body?



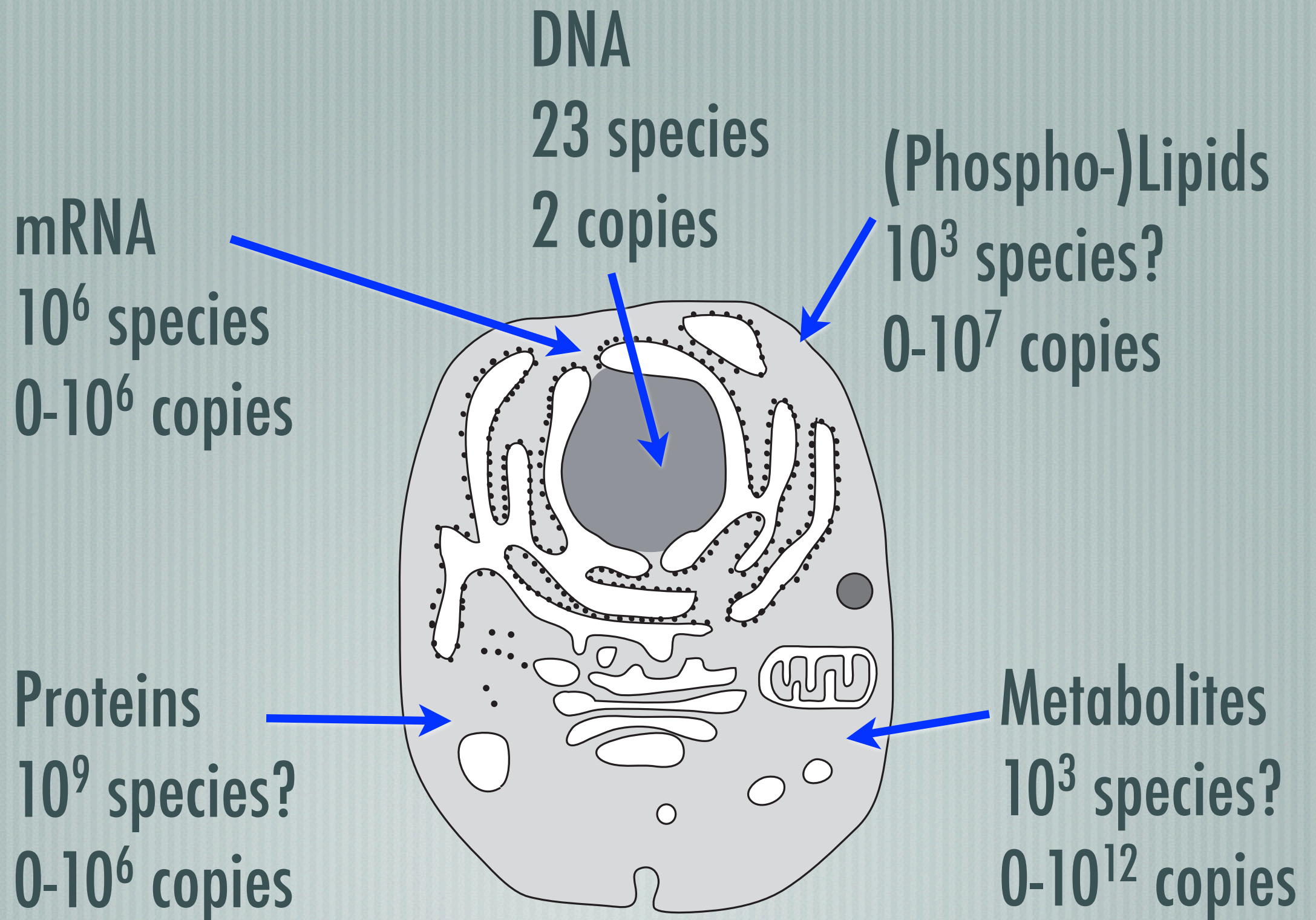
A) $<40,000$

B) 40,000-200,000

C) $>200,000$



A human cell - a system



Two ways to investigate a system

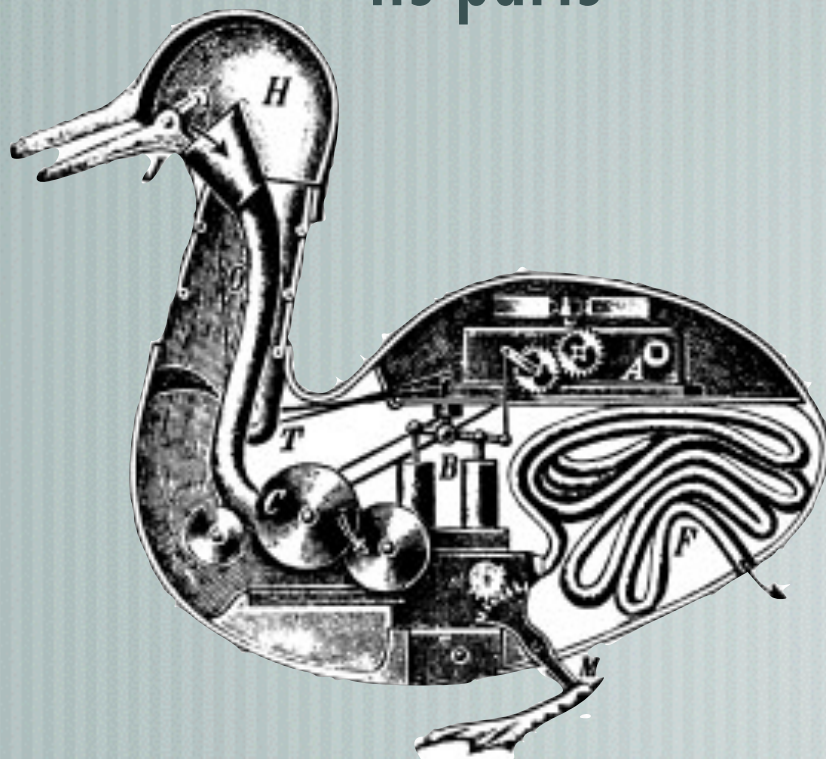
Reductionism

- Isolate and examine parts of the system
- Reducing the system to the interactions of its parts

Holistic (System)

- Measure everything at once
- Not necessarily investigating all causal relationships

Canard Digérateur
(Vaucanson)



$$\frac{dS}{dt} \geq 0$$

Omic

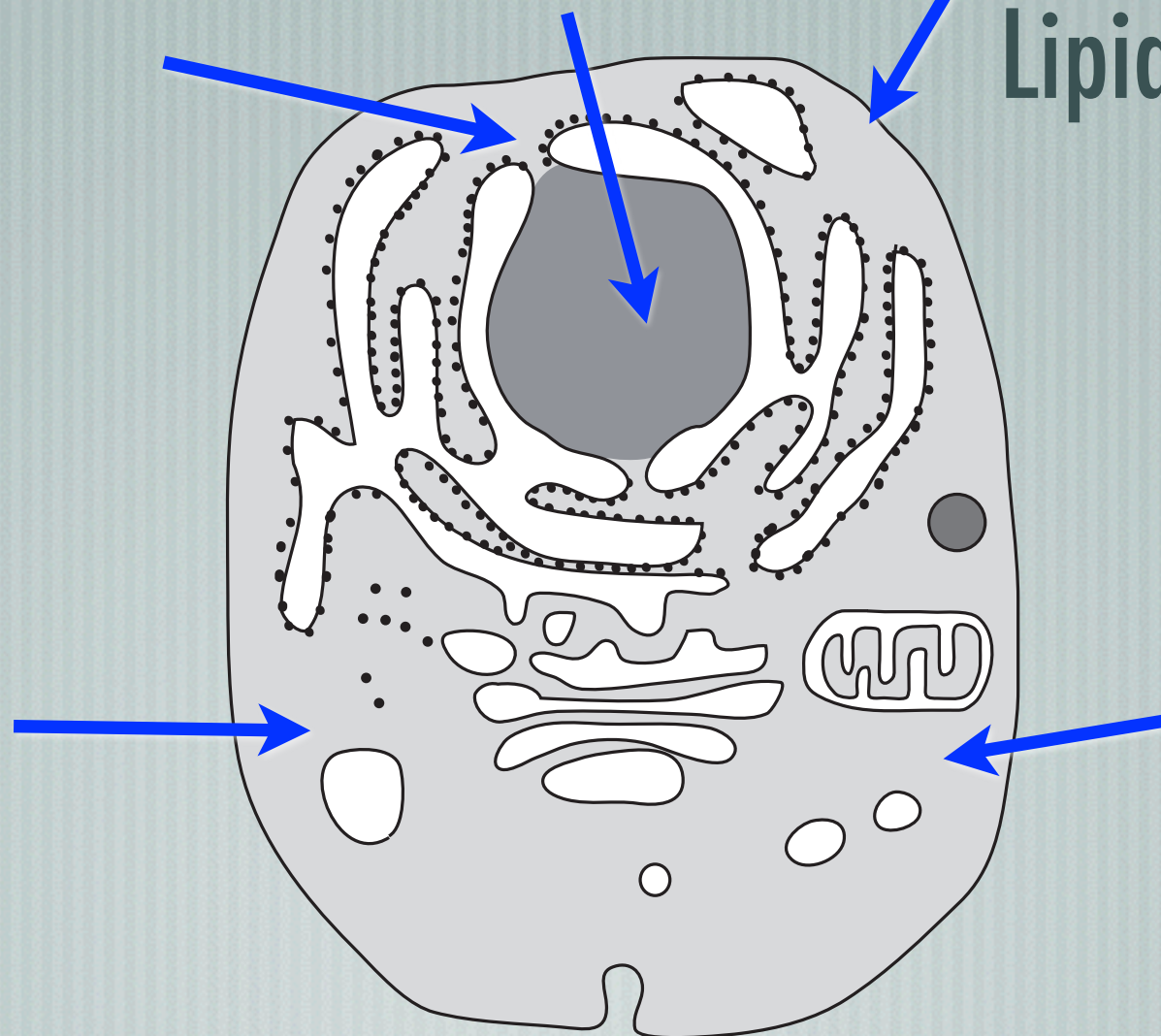
mRNA
Transcriptomics

DNA
Genomics

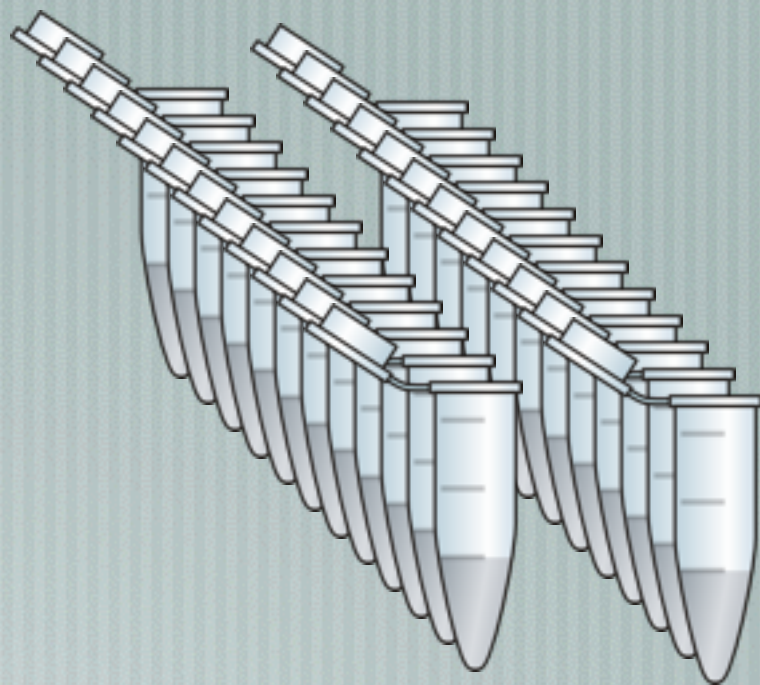
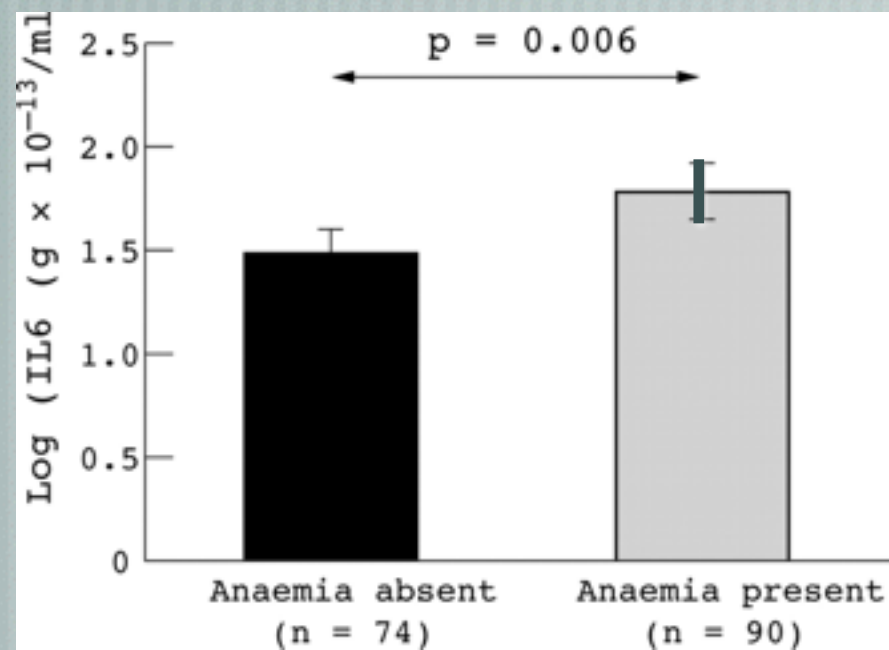
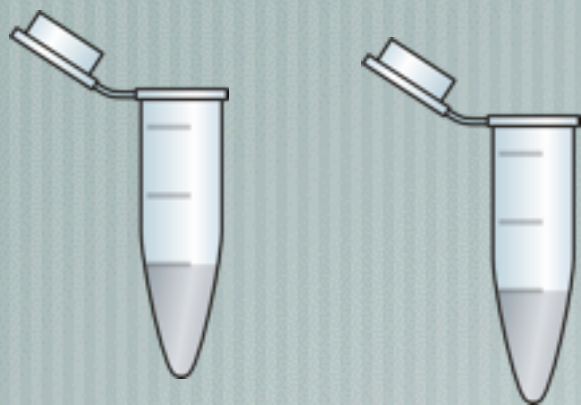
Lipids
Lipidomics

Proteins
Proteomics

Metabolites
Metabolomics

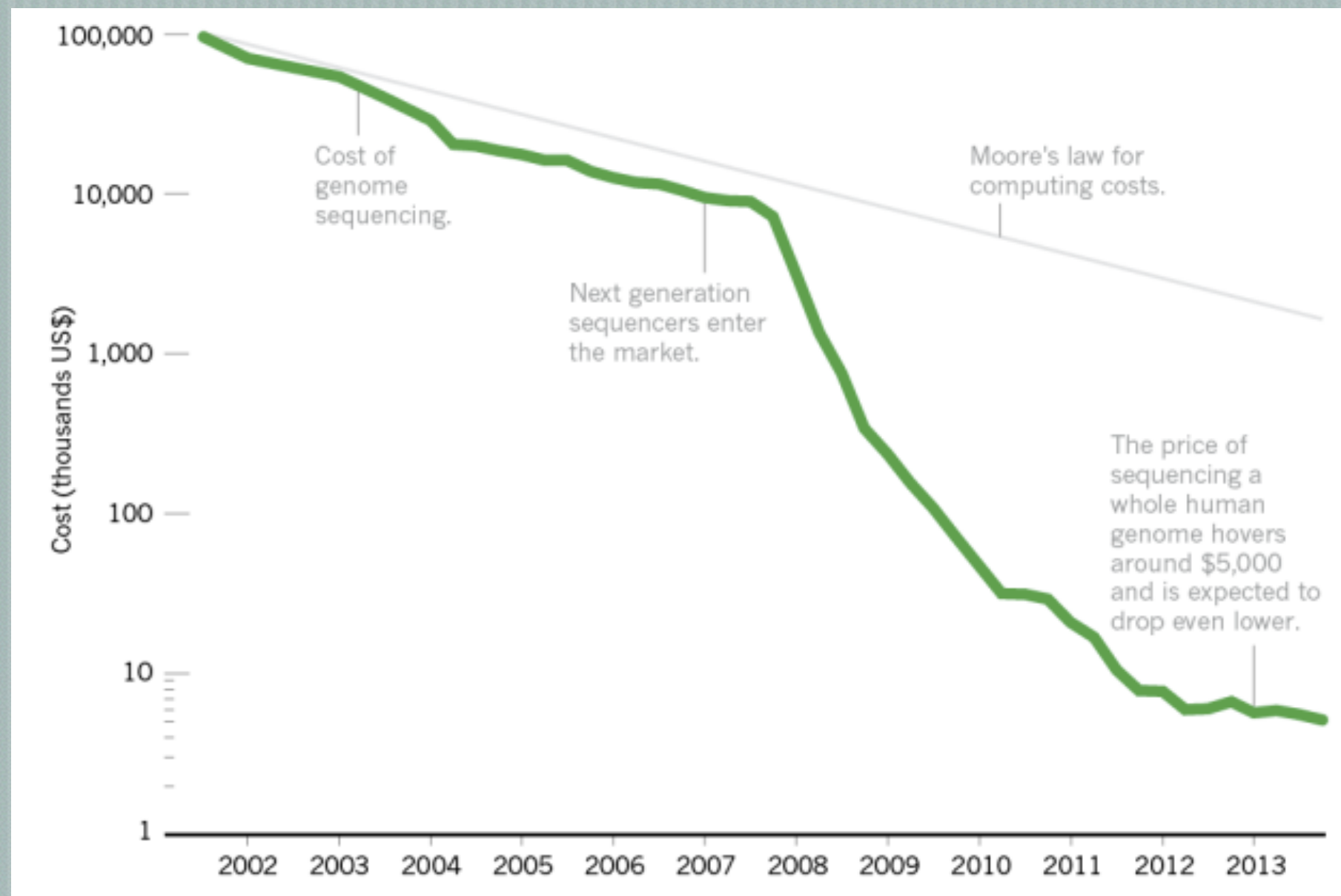


New role of evidence



t_12_2_1	1.09445	0	1	K.AIVLTLSGIFLLGNIPAILGDNLWK.N	71074392
t_8_3_1	1.07078	0	1	R.IAAGDEDHVSIQPIAAGSVAQNGEAGAMFMK.P	71074392
t_9_2_1	0.828281	0.224336	1	R.YWVFTAFGDTSGDYM#K.N	71074392
t_15_3_1	0.773491	0.224336	1	R.TLFSENGQNVSFYLKEINQTLDR.I	71074392
t_19_2_1	0.483015	0.859955	1	K.DKC*SWLYC*ETAC*ASAAMDVCR.G	71074392
t_5_2_1	0.434554	0.859955	1	R.PVDIESAVKKIHEMGC*LAMADC*SNLEEGLYC*K.A	71074392
t_21_2_1	0.247071	0.859955	1	K.EKMIM#ALQR.L	71074392
t_11_2_1	0.222999	0.859955	1	R.MNPPLYVVITFFITGR.A	71074392
t_21_3_0	0.159285	0.859955	1	K.HHINIGDVFOIVPSR.R	71074392
t_13_3_1	0	0.859955	1	K.QFQDLSLFRWANNHGAFM#YHEKIYFIVIYVTTITYLAAVER.K	71074392
t_11_3_1	-0.0840831	0.859955	1	K.LPGAAVEQLAQRPKFIESVPGVM#EAK.N	71074392
t_20_2_1	-0.0840831	0.859955	1	K.KLTALFNLPELKNDELHNMVLDNR.K	71074392
t_19_3_1	-0.0840831	0.859955	1	K.YLTYIIGM#SFGVIANM#VMC*WLLTDFSPNYLGAK.N	71074392
t_4_3_1	-0.0840831	0.859955	1	K.PHGEAQALEIPINSFDLEDMPYELMKNLDEIACVIGLSR.N	71074392
t_16_3_1	-0.0840831	0.859955	1	K.QQRAAENDAILNNIWSPMVIGMQTVDSK.A	71074392
t_20_3_1	-0.198275	0.859955	1	R.HLFVQLLKEYGAEQDFSFIEQNGMFSFSGLTGEQVDR.L	71074392
t_7_2_1	-0.284387	0.859955	1	R.DIEQDAKASKNTLAVR.L	71074392
t_12_3_1	-0.284387	0.859955	1	K.IARNM#KDYLALAWADLEDLTALTISAGATFGFNWK.L	71074392
t_3_2_1	-0.515813	0.862831	1	K.NORPYYFHTAEP.R	71074392
t_3_3_1	-0.515813	0.859955	1	K.FSLEHTPSAQKALAMSCRDLC*R.E	71074392
t_4_2_1	-0.515813	0.859955	1	K.EACILGM#LYLVANLEIGQSHTIMKSK.A	71074392
t_16_2_1	-0.515813	0.859955	1	R.LHLIDHVKDADVITGEGR.I	71074392
t_15_2_1	-0.515813	0.862831	1	R.FGLFDPEATGLGADNER.A	71074392
t_7_3_1	-0.515813	0.859955	1	K.DC*LPHGKAGNIMISICQDAPVRLK.K	71074392
t_9_3_1	-0.515813	0.859955	1	K.IQITRNSGMATFENLHYSHWC*YR.K	71074392
t_8_2_1	-0.515813	0.859955	1	K.VMALKIC*ANPEKQDTPC*R.K	71074392

Historical trends in DNA sequencing costs



2000 - HUGO 30G\$

2007 - Venter 10M\$

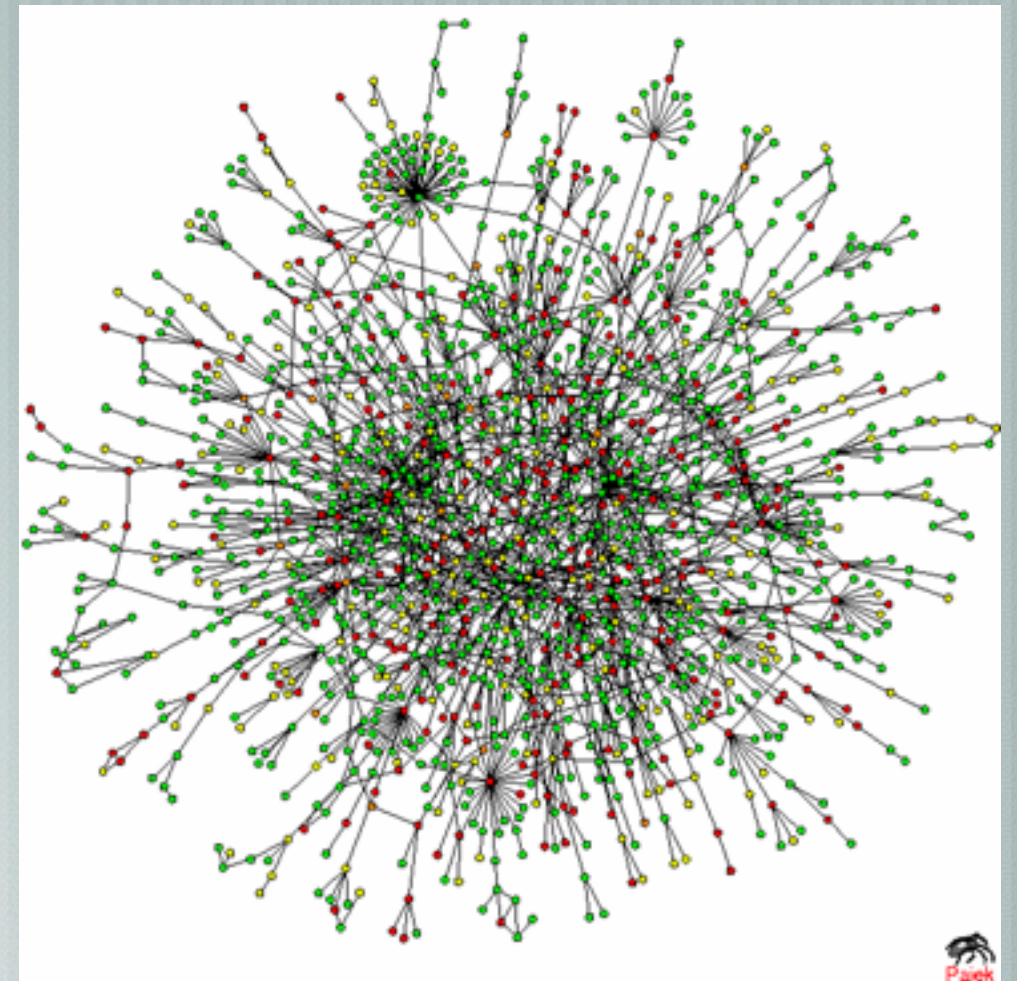
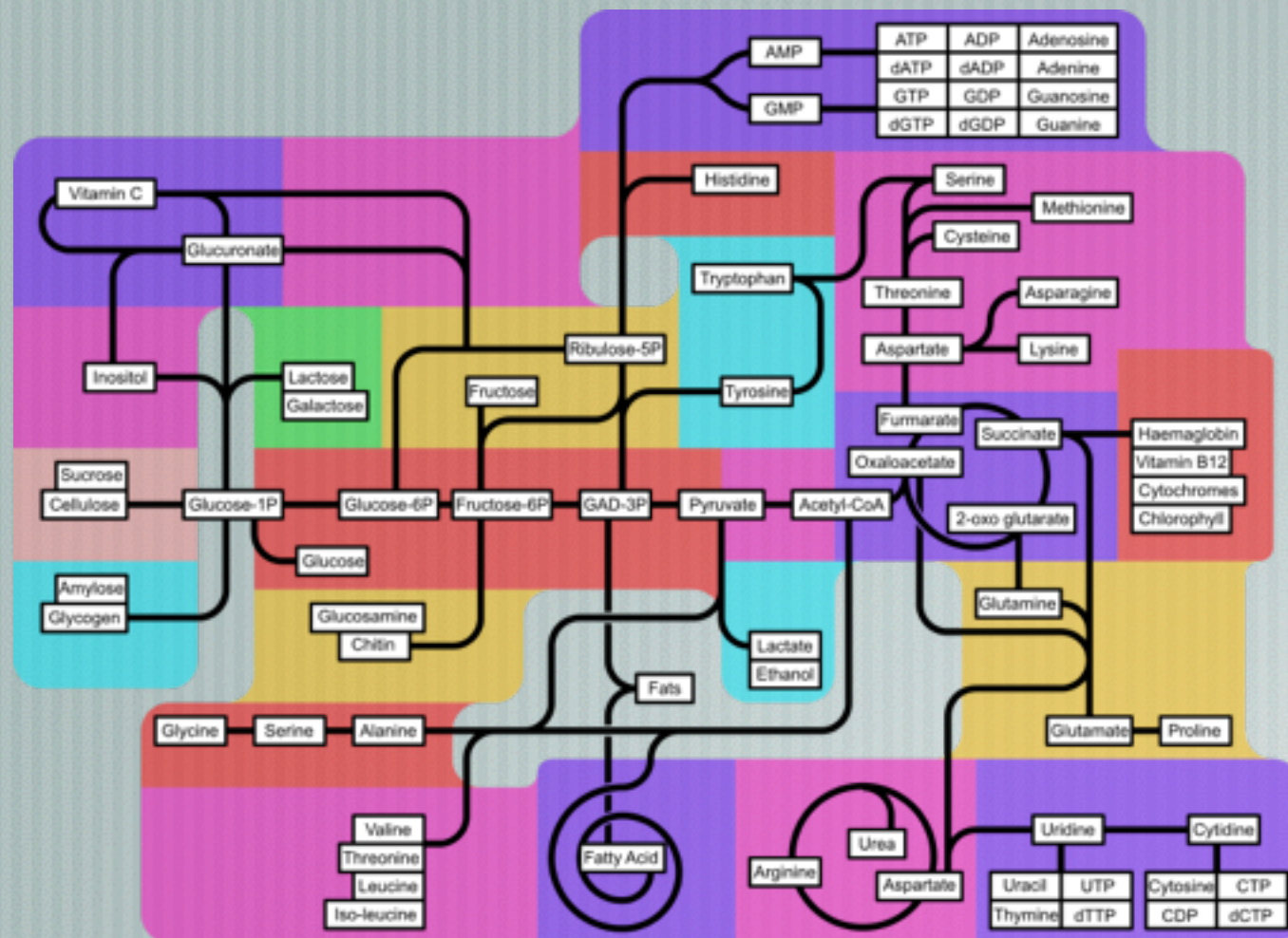
2008 - Watson 1M\$

Today - 5k\$

[Hayden, Nature 2014]

New role of bioinformatics

Collection of data is no longer the bottleneck,
analysis is!



Central Dogma of Molecular Biology

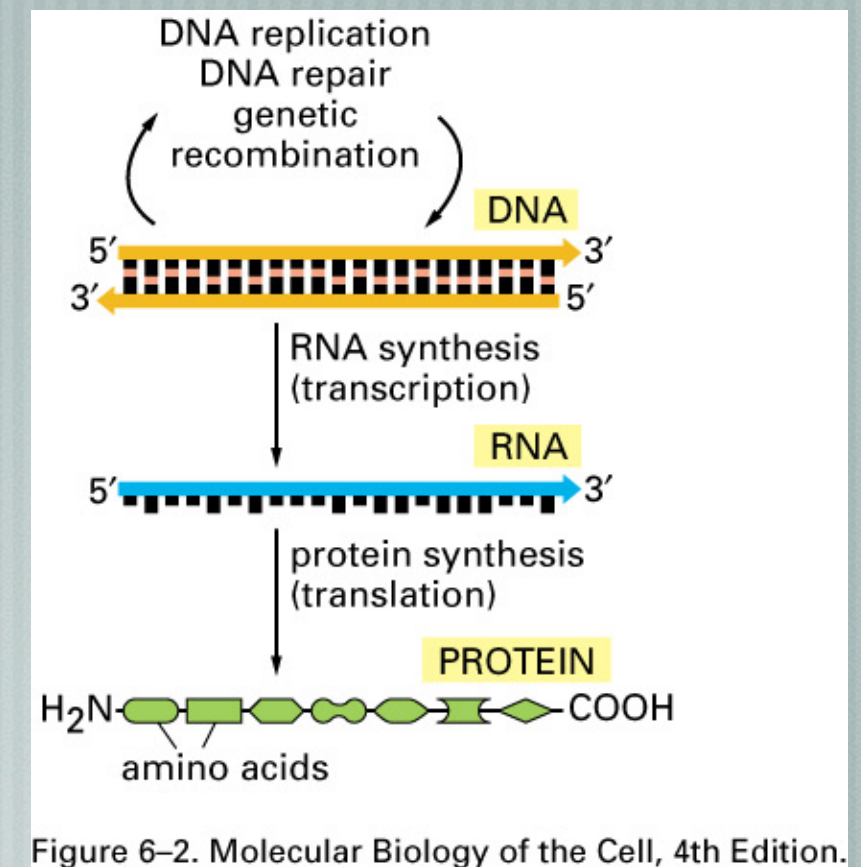
DNA: Genome



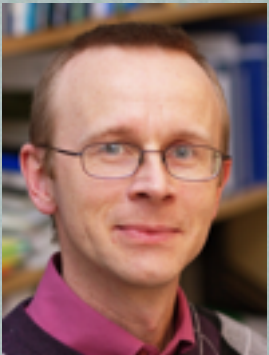
mRNA: Transcriptome



Proteins: Proteome



Central Dogma of Molecular Biology



DNA: Genome



mRNA: Transcriptome



Proteins: Proteome

