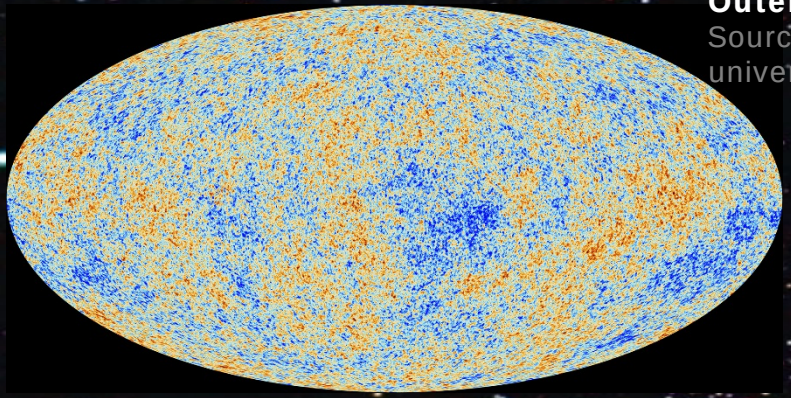


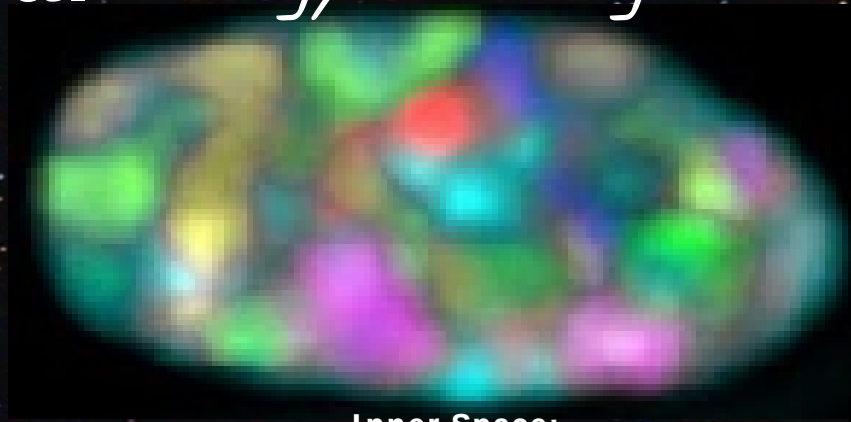
Outer Space:
Source -- Cosmic Microwave Background (CMB): the oldest light in our universe, Planck spacecraft, 2013 (Copyright: ESA and the Planck Corporation)

Cold Data Archival

Biology as Storage?

Sanjay Joshi
EMC² Isilon Storage Division

SDC STORAGE DEVELOPER CONFERENCE SNIA SANTA CLARA, 2014



Inner Space:
Source -- Bolzer A, et al., "Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes.", (2005), PLoS Biol 3(5)

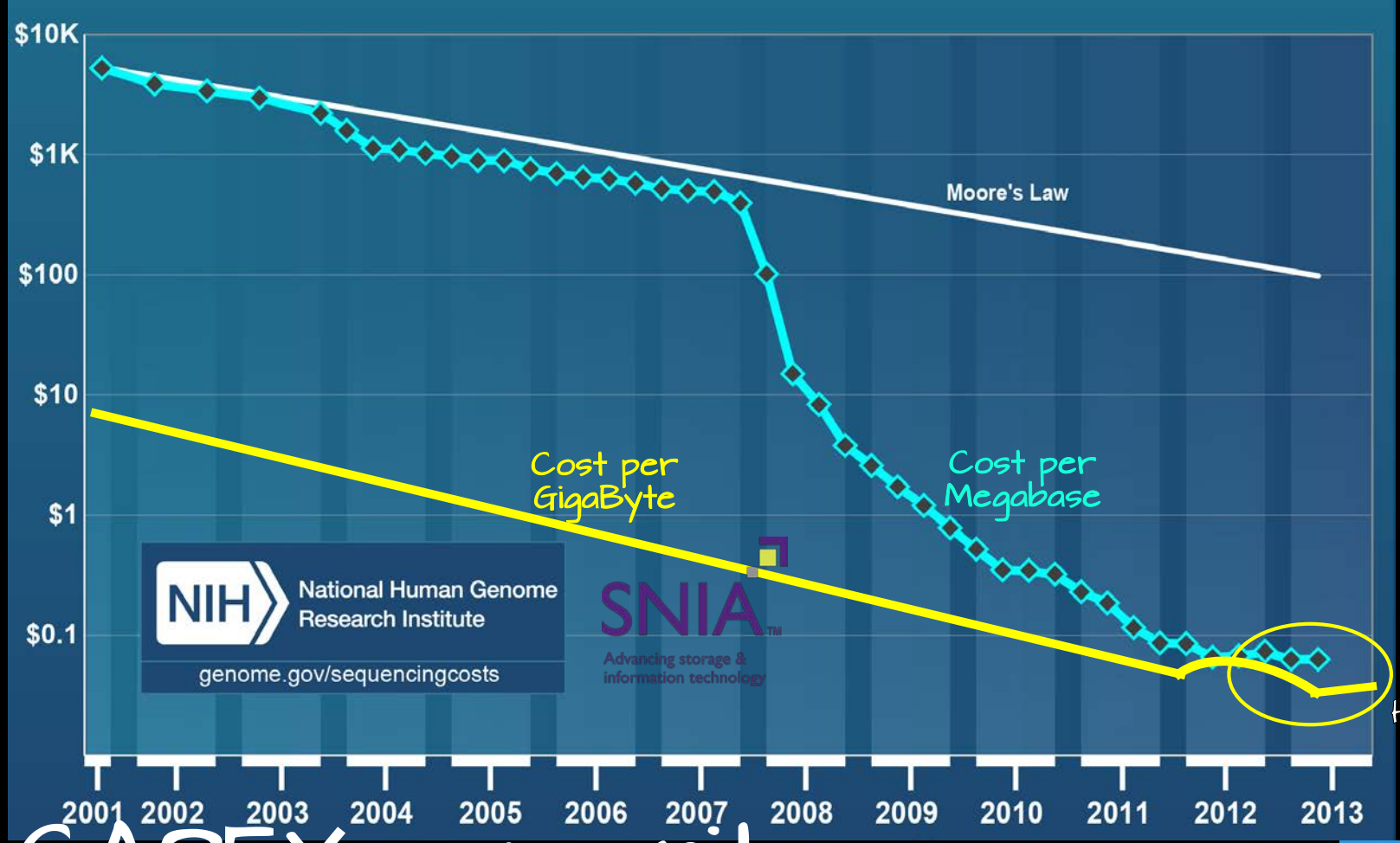
"...because as we know, there are known knowns;
there are things that we know that we know.

We also know there are known unknowns;
that is to say we know there are some things we don't know.

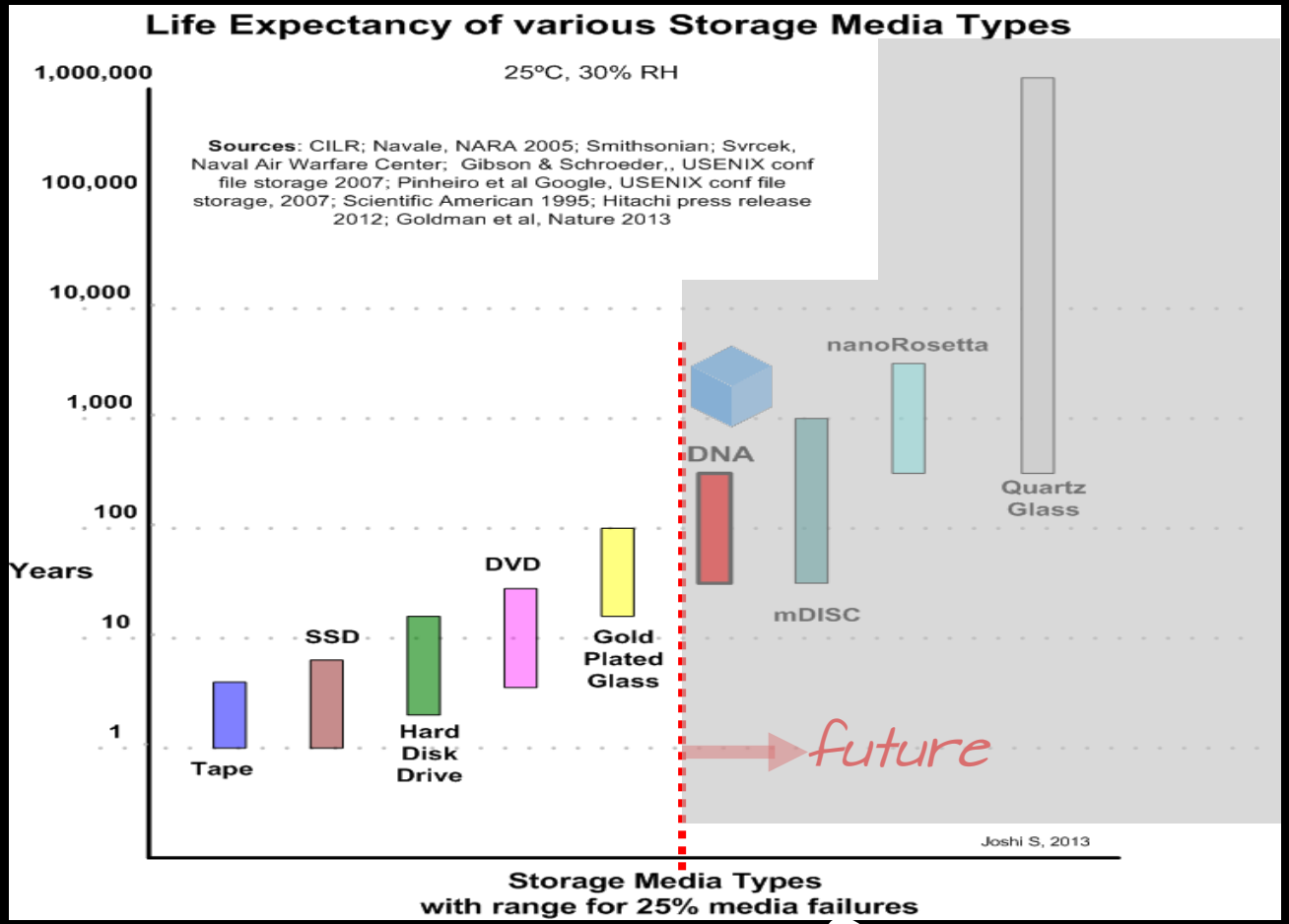
But there are also unknown unknowns,
the ones we don't know we don't know."

Donald Rumsfeld, United States Secretary of Defense
1975-1977, 2001-2007

{'un'|''}known



CAPEX per unit

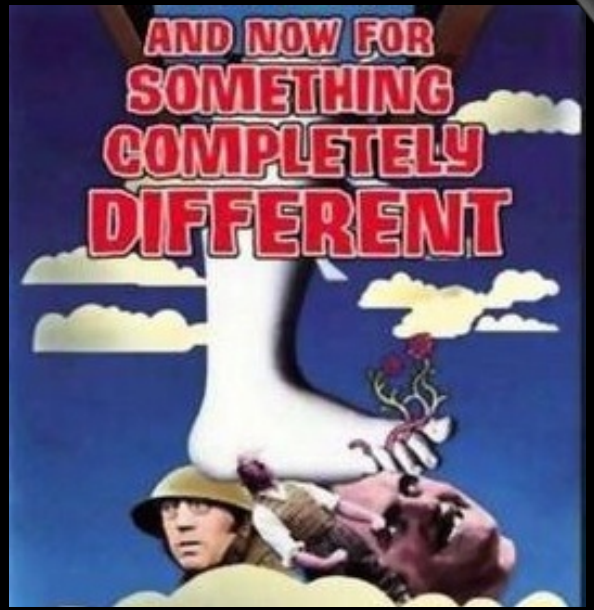


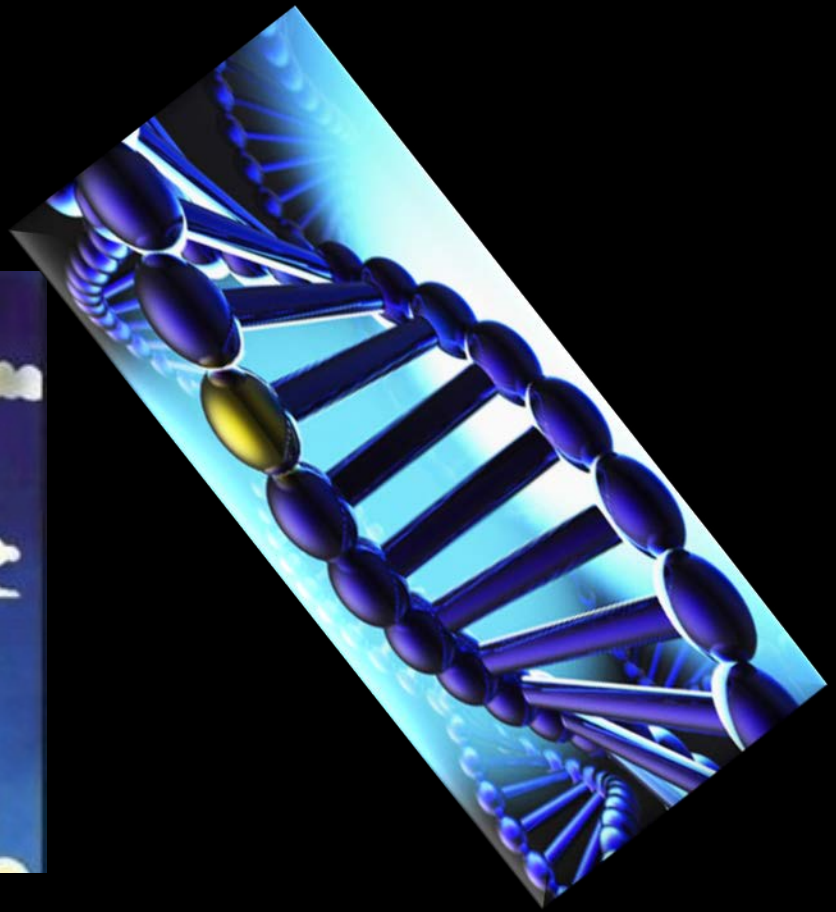
The Reality of MTBF

40 ZB by 2020

Source: EMC Digital Universe Study, 2014
Research and Analysis by IDC

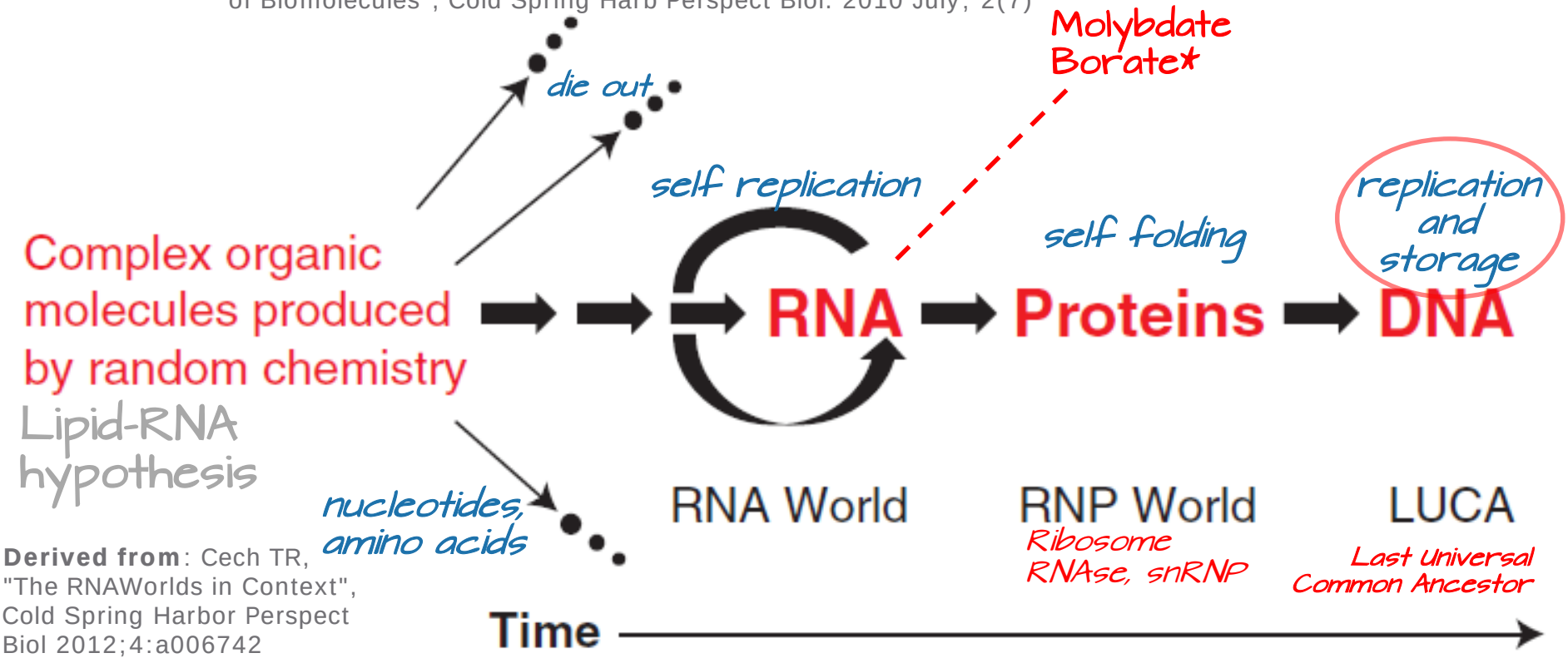
The Digital Universe





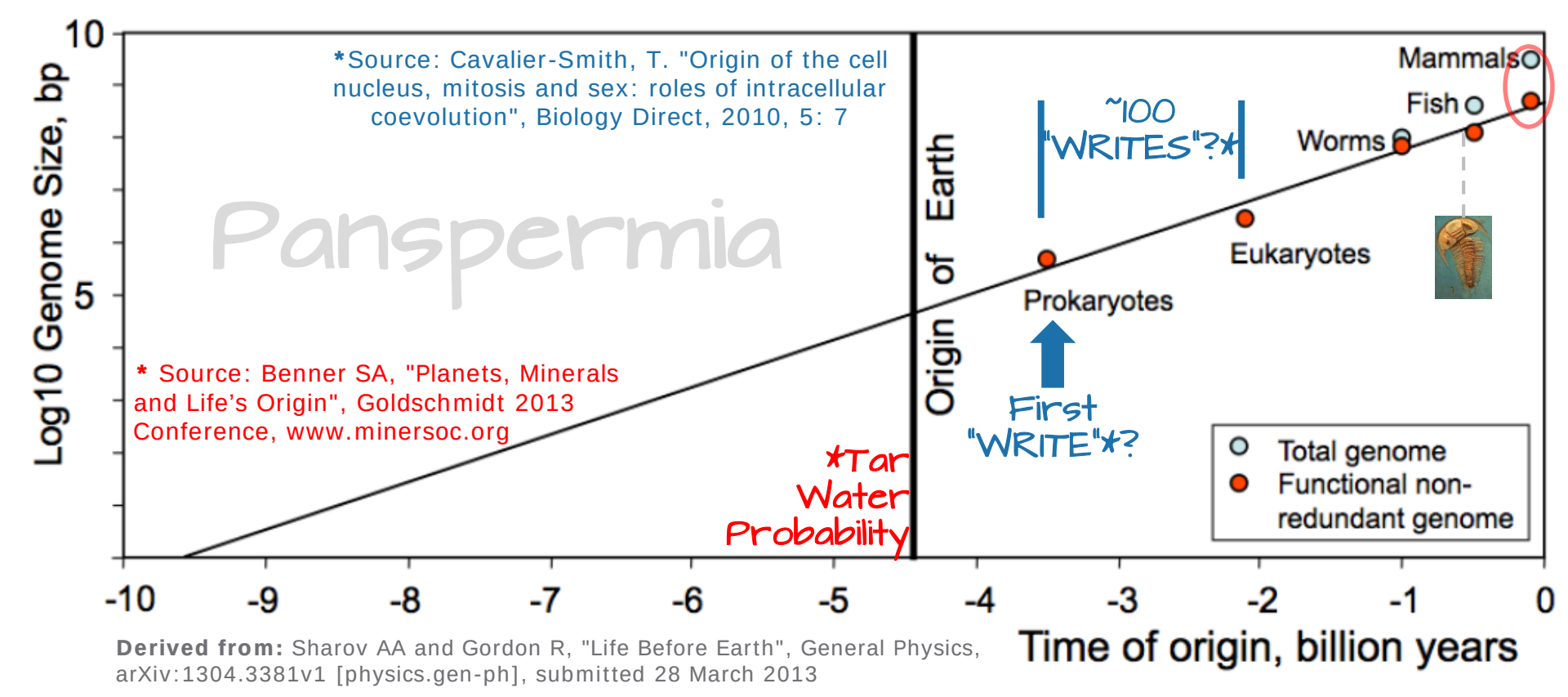
Monty Python, 1972. © BBC

* Source: Benner SA, et al, "Planetary Organic Chemistry and the Origins of Biomolecules", Cold Spring Harb Perspect Biol. 2010 July; 2(7)



Derived from: Cech TR, "The RNAWorlds in Context", Cold Spring Harbor Perspect Biol 2012;4:a006742

Life: RNA came first?



Derived from: Sharov AA and Gordon R, "Life Before Earth", General Physics, arXiv:1304.3381v1 [physics.gen-ph], submitted 28 March 2013

The first "write" process in the cell?

char(3*10^9) human_genome

3 gigabases [(3*10^9)*2]/8 = ~750MB

with overlaps, ~1 GB per cell

DNA base: A, G, T, C

RNA base: A, G, U, C

DNA Sizing as Storage

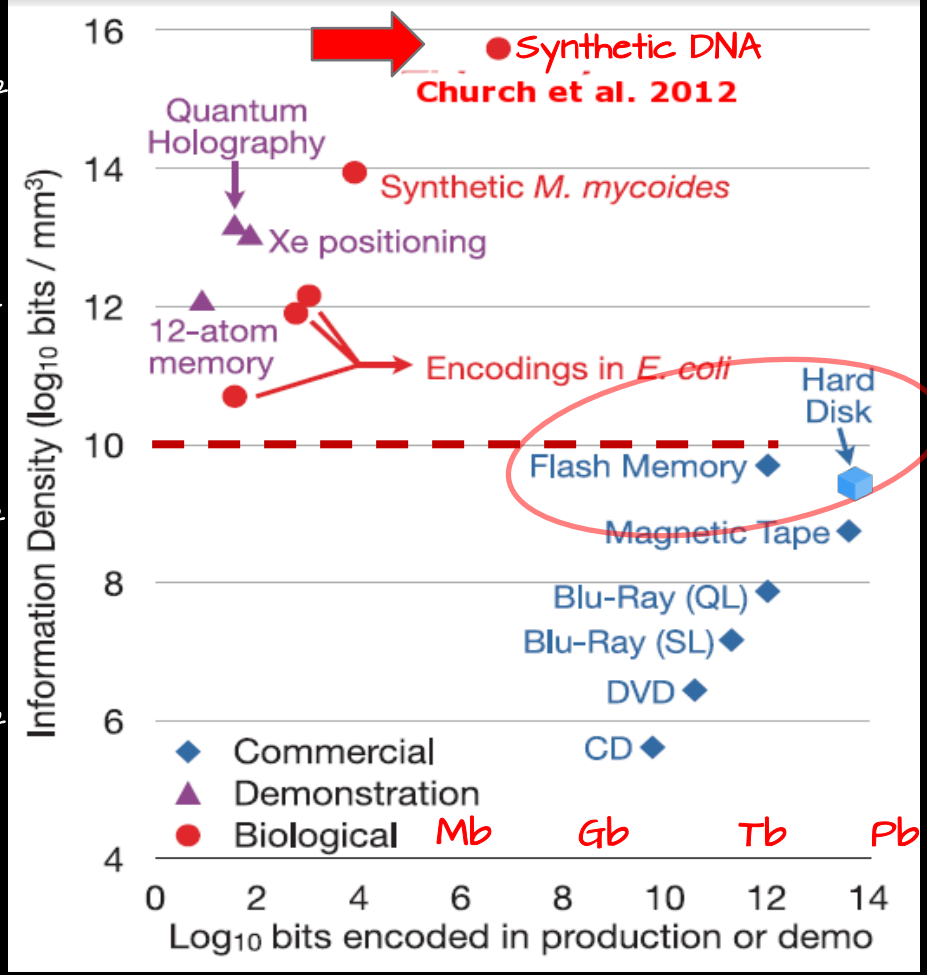
Storage Density

Pb/mm²

Tb/mm²

Gb/mm²

Mb/mm²

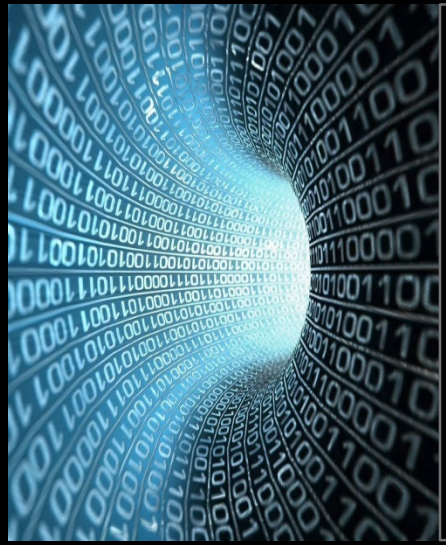


Storage Size

Source: Church et al., Science, Vol 337, p1628, Sep 2012

EMC²

WRITE



bits to trits (base 3 encoding)

Oligo library

ink-jet printer

Parallel Flow Cell Reactor
1 error/500 bases

Purify, mag beads
PCR amplify

15 ng/µL

Derived from: Goldman et al., Nature (Jan 2013) doi:10.1038/nature11875

Note: Images from Agilent, GE and Life Technologies for example purposes only.

2 PB/gram

cheaper than tape

by ~2025?

DNA as Storage: Process

EMC²

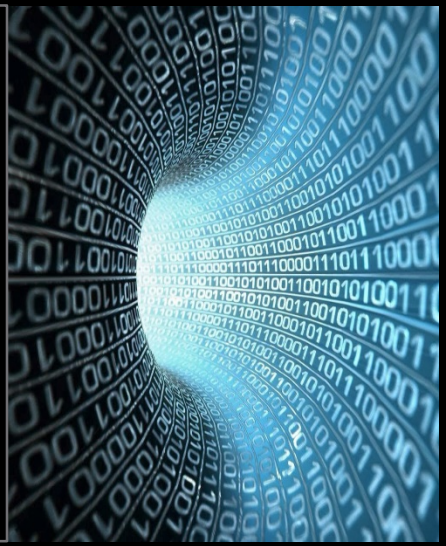
READ

DNA Sequencer

Derived from: Goldman et al., Nature (Jan 2013) doi:10.1038/nature11875

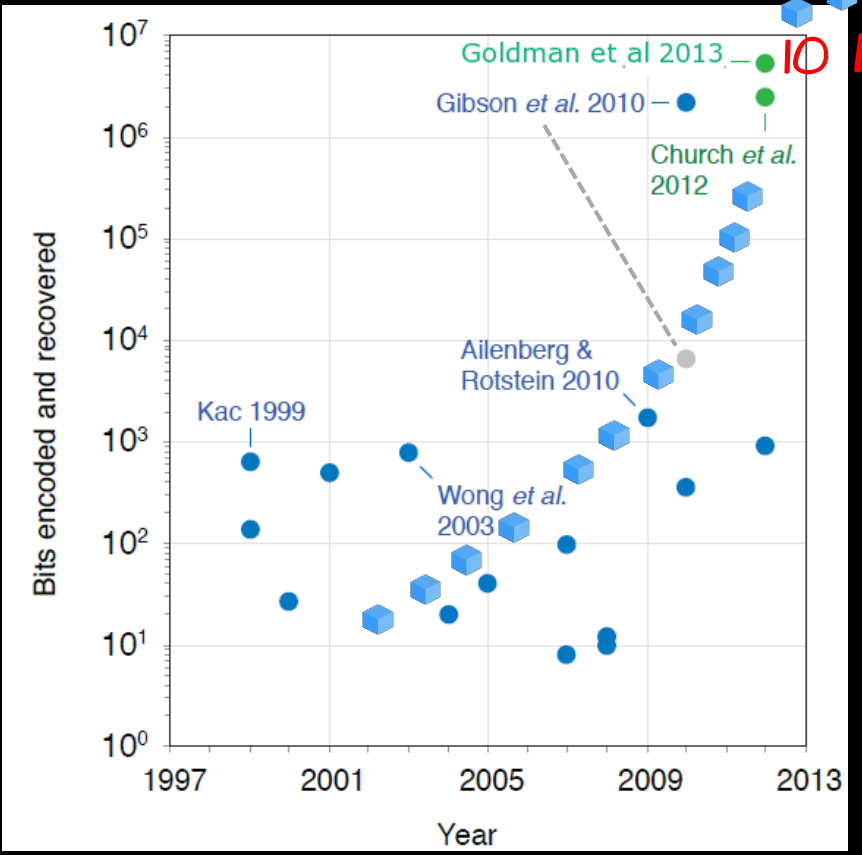
Note: Image from Oxford Nanopore sequencer for example purposes only.

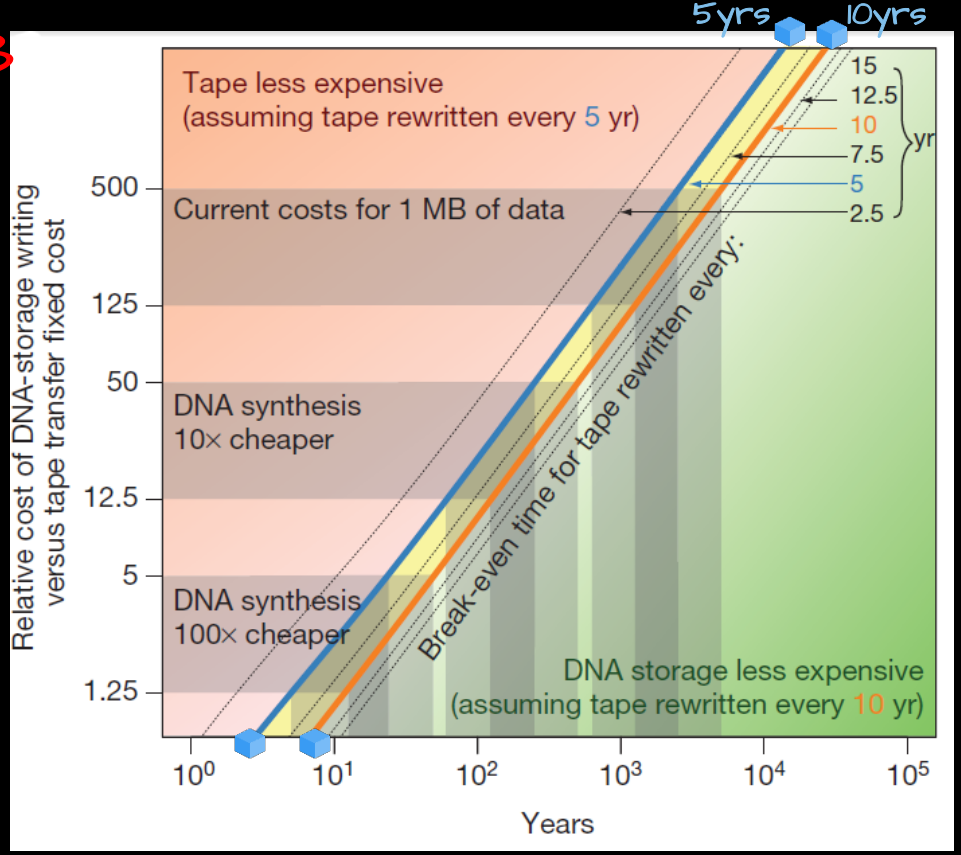
trits to bits



DNA as Storage: Process

Derived from: Goldman et al., Nature (Jan 2013) doi:10.1038/nature11875





DNA Storage Scale & Cost

Brewer's Theorem

Consistency

Acknowledgement

ACID

Atomicity, Consistency, Isolation, Durability

BASE

Basically Available Soft-state Eventual consistency

Partition Tolerance

Source: Brewer E, "CAP Twelve Years Later: How the "Rules" Have Changed,", http://www.infoq.com/articles/cap-twelve-years-later-how-the-rules-have-changed

Distributing it...

Trust

"Information moves at the speed of trust."

Dr. Douglas Fridsma, CSO

The Office of the National Coordinator for Health Information Technology

"Non-reproducible single occurrences are of no significance to science." Karl Popper, 1959

Big Data = V⁶

Volume, Velocity, Variability, Visualization, Viscosity, Veracity

Control + Visibility + Transparency

From: Security guidance for critical areas of focus in Cloud Computing v3.0:
http://www.cloudsecurityalliance.org/guidance/csaguide.v30.pdf

	Infrastructure Managed By[1]	Infrastructure Owned By[2]	Infrastructure Located[3]	Accessible and Consumed By[4]
Public	Third Party Provider	Third Party Provider	Off-Premise	Untrusted
Private/ Community	Organization Or Third Party Provider	Organization Third Party Provider	On-Premise Off-Premise	Trusted
Hybrid	Both Organization & Third Party Provider	Both Organization & Third Party Provider	Both On-Premise & Off-Premise	Trusted & Untrusted

41

of 91 HC respondents are concerned about Public Cloud Security*

*Ponemon Institute Mar 2014

5.3

Days: length of security breach, worst for Healthcare/Pharma**

** BitSight Industry Security Report, May 2014

30M

data breaches since 2009 (avg. $2M fine per breach, 2 yrs)

US Govt. Office of Civil Rights, HHS

£1.79B fines in UK for NHS 2012 breaches

343 US incidents through Jul 2013
321 in 2012 *ITRC

4

position points (GPS) determine identity . .

. . Scientific Reports, 3, Mar 2013

PGP Encrypted De-ID reverse engineered +

+ Data Privacy Lab, CMU, Apr 2013

genome Y-STR

Surnames from ancestry data

Science, Jan 2013:
339:6117 pp. 321-324

4M

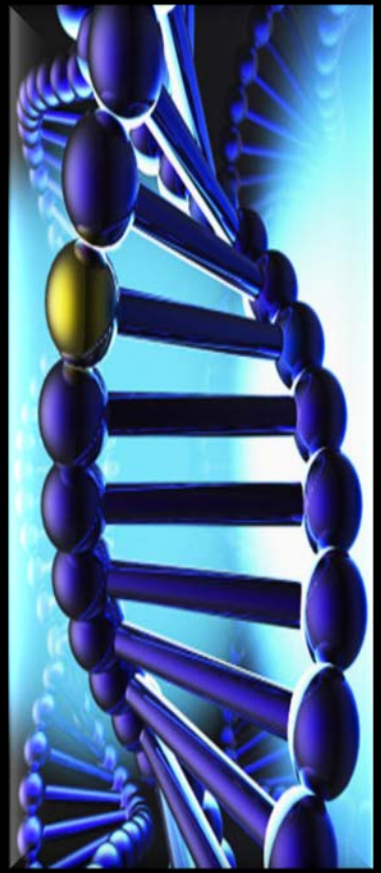
Common Variants in trace DNA

Craig DW, et al, (2008), PLoS Genet 4(8)

Private & Community Cloud

EMC²

sanjay.joshi@emc.com



+

=

(2*23)

~ 5 trillion cells...

50 ZB@ 10^{21}

100W/day

One Word: "Polymers"

Inspired by Mr. McGuire
in movie 'The Graduate',
© Embassy Pictures, 1967

Thank You!