

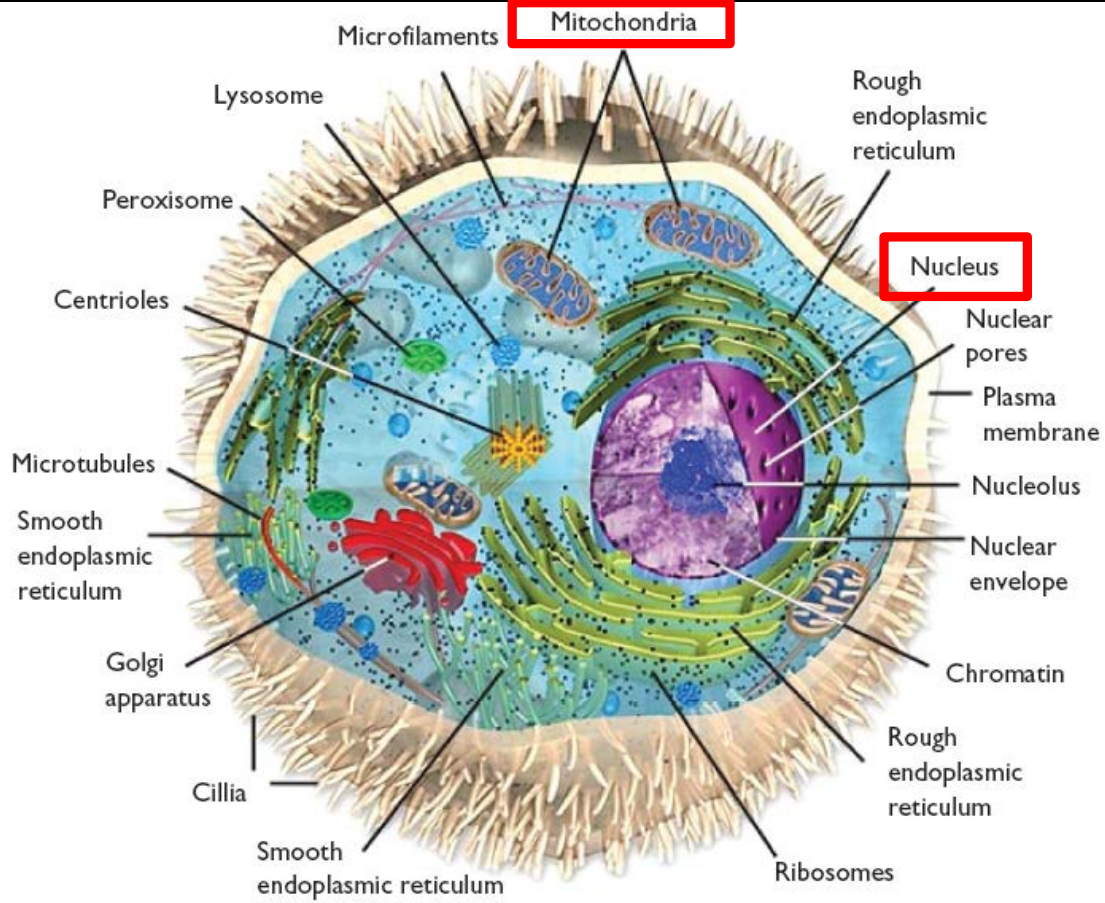
Sanjay Joshi
EMC² Emerging Technologies Division

/etc/sysctl/human/cell
non-reductionist systems thinking in biology



Unit





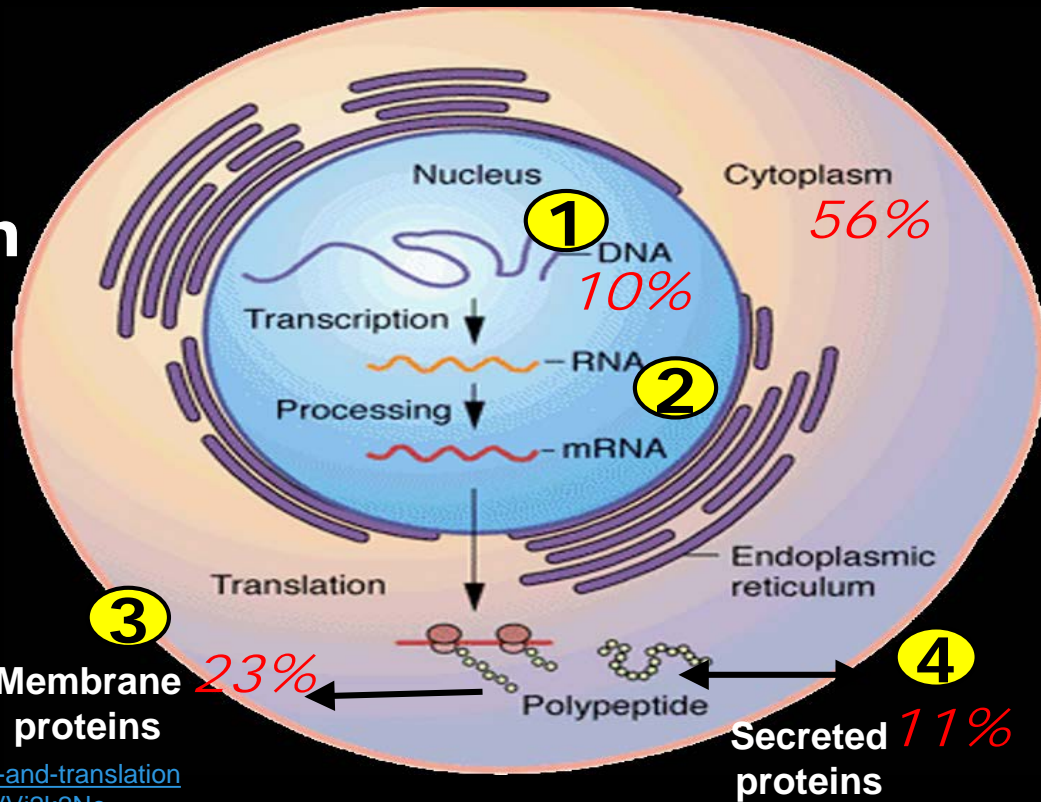
The Human Cell



15 - 70 trillion cells*



DNA_Transcription_Translation.webm

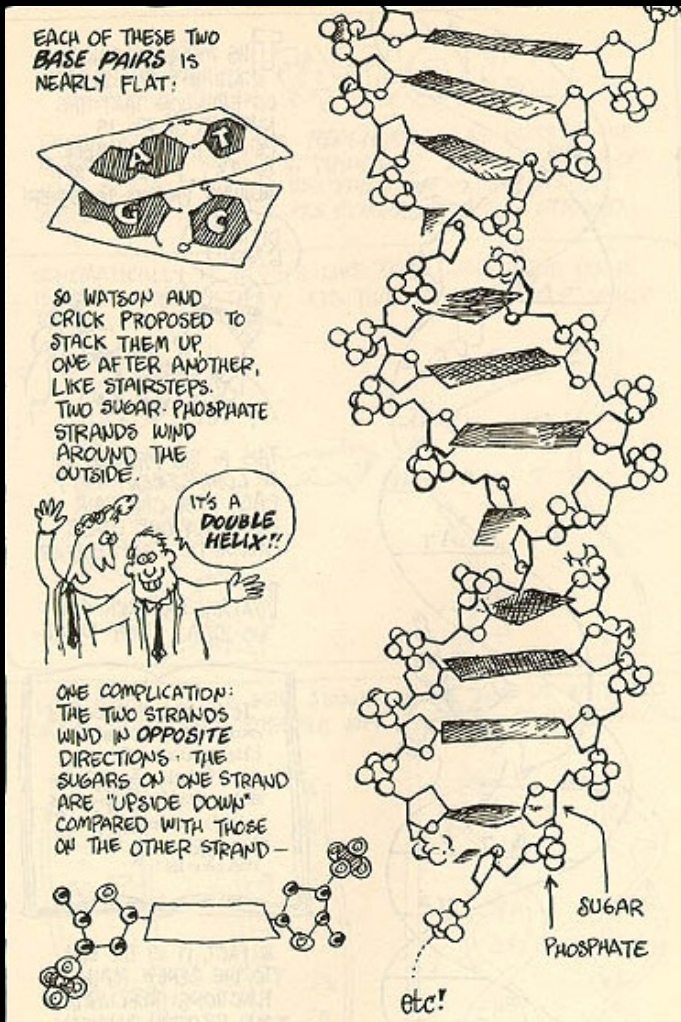


<http://www.moleclues.org/videos/transcription-and-translation>
<http://www.youtube.com/watch?v=J3HVVi2k2No>

Nuclear DNA: Transcription and Translation

EMC²

DNA Structure



Base Pairs

A - T

G - C

purines

U

pyrimidines

Double Helix

Source: Larry Gonick, Mark Wheelick;
Cartoon Guide to Genetics

EMC²

`[char(3 * 10^9) human_genome]`_{strand}

3 gigabases $[(3 * 10^9) * 2] / 8 = \sim 750\text{MB}$

with overlaps, ~ 1 GB per cell



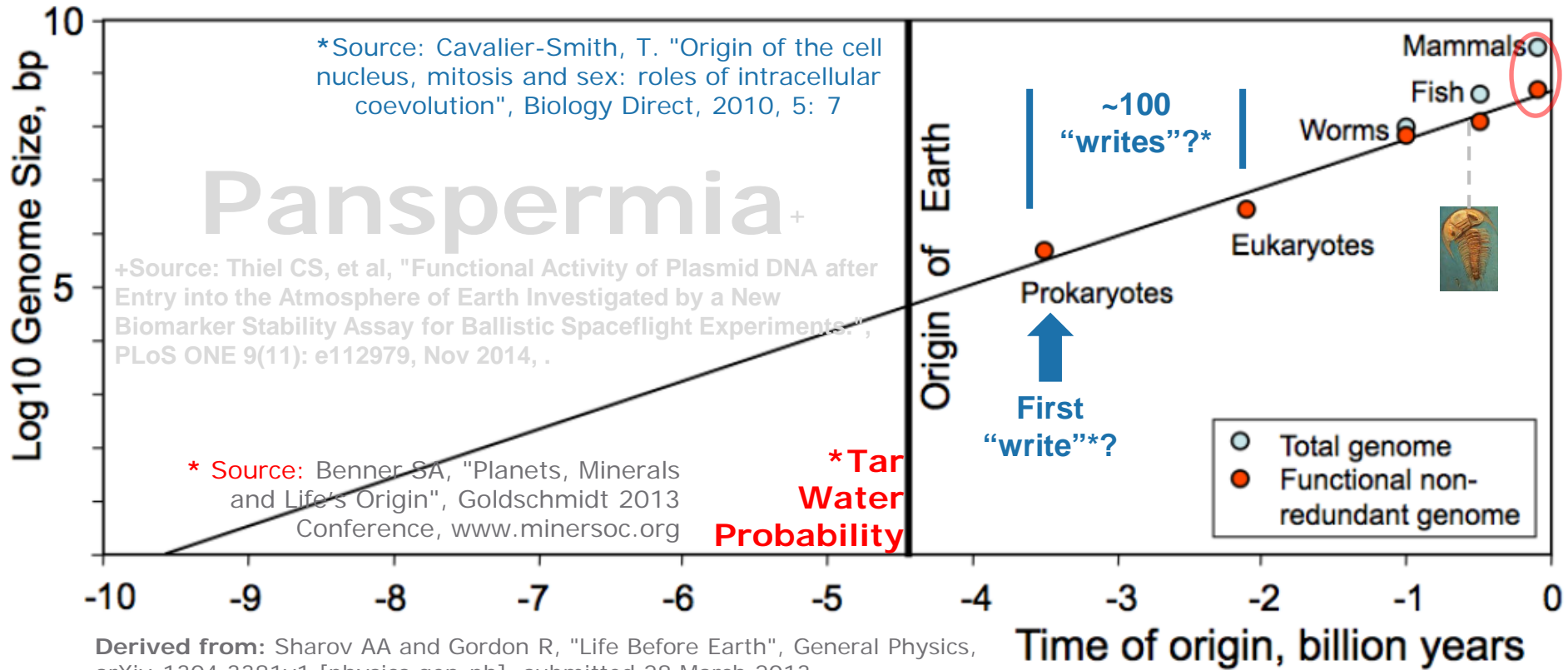
DNA base: A, G, T, C

RNA base: A, G, U, C



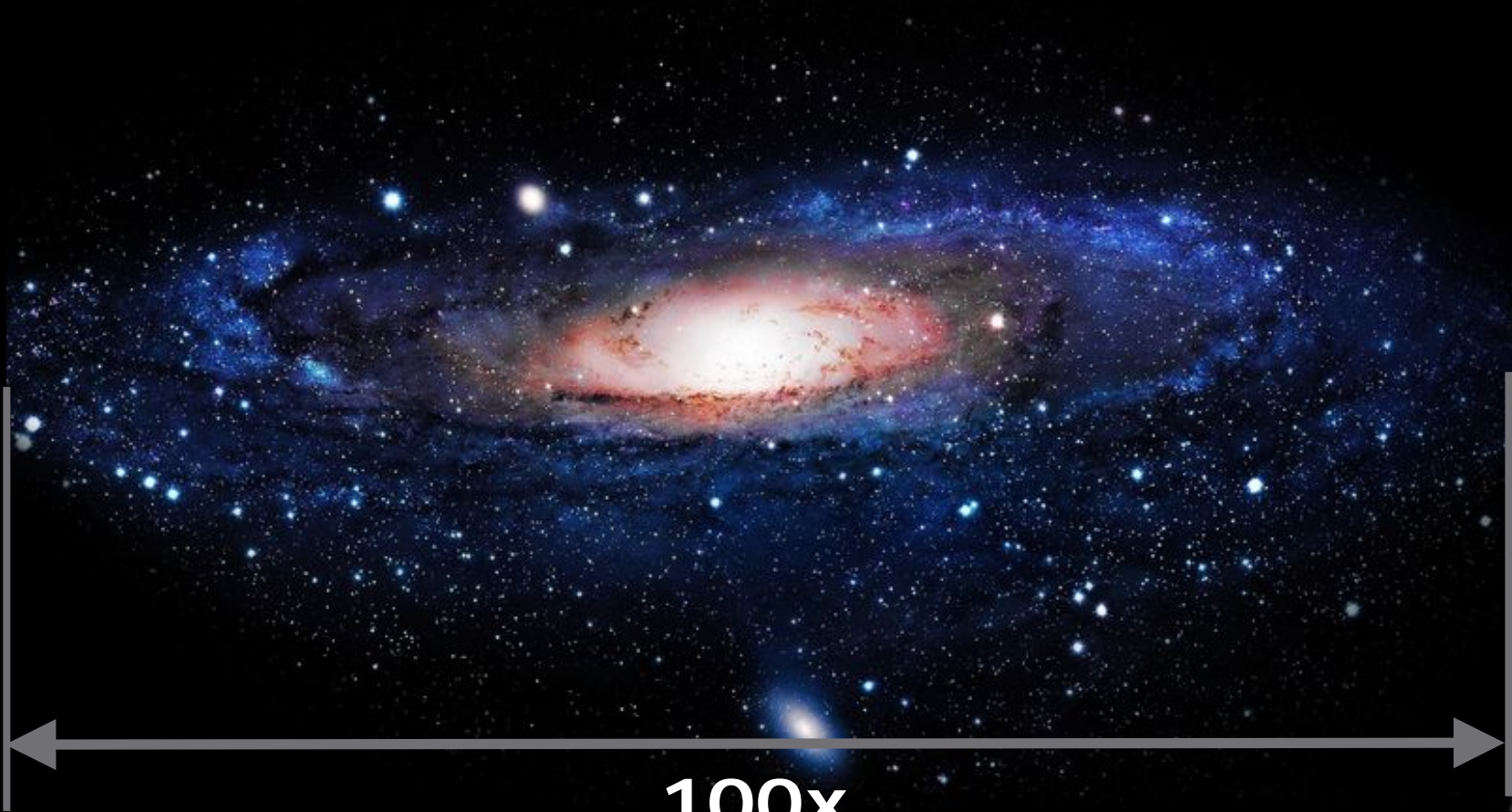
Genome Sizing as Storage

EMC²



The Gold Master?

EMC²

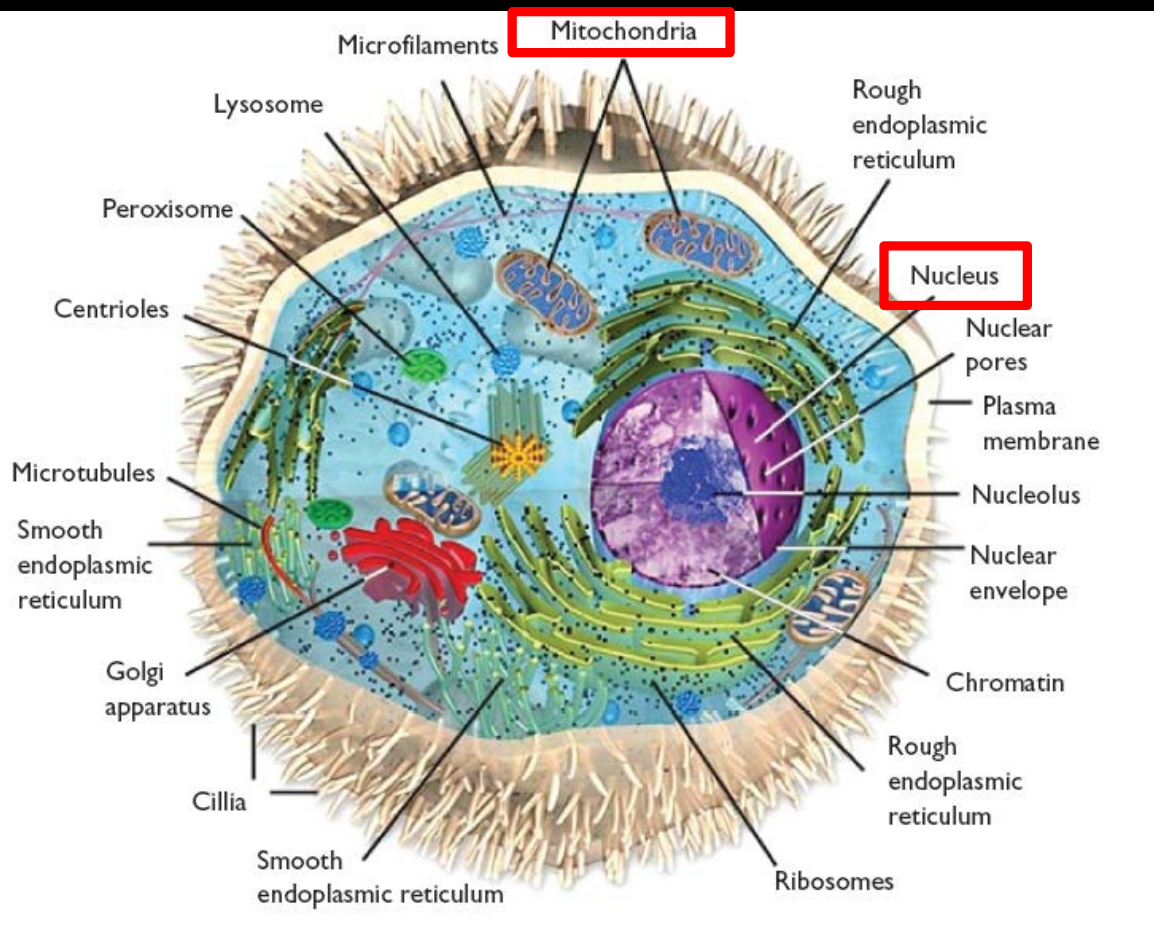


100x

(all human cells)

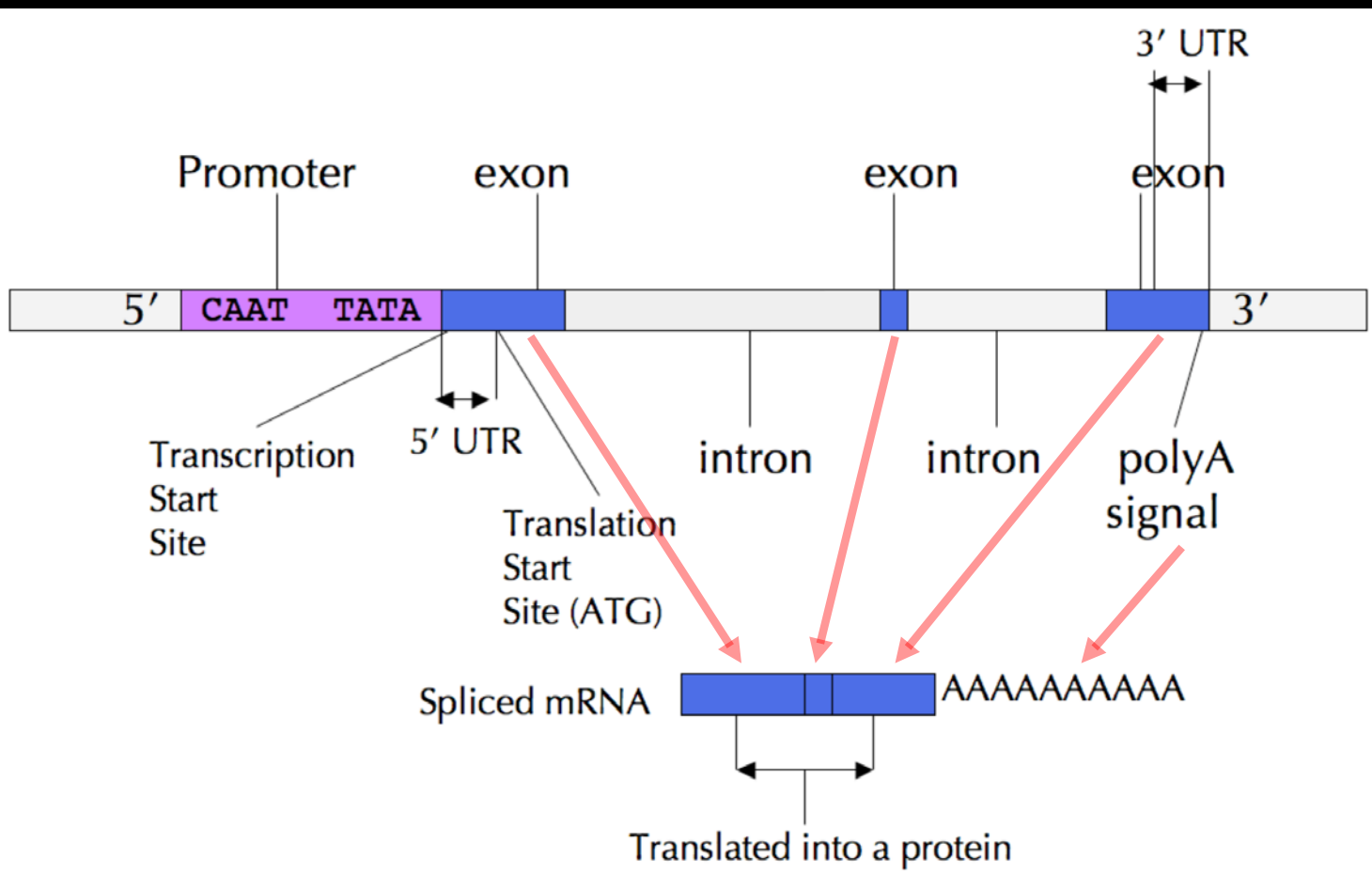


EMC²



Information \equiv DNA

EMC²



Search Header

Interpret

Compile

Execute



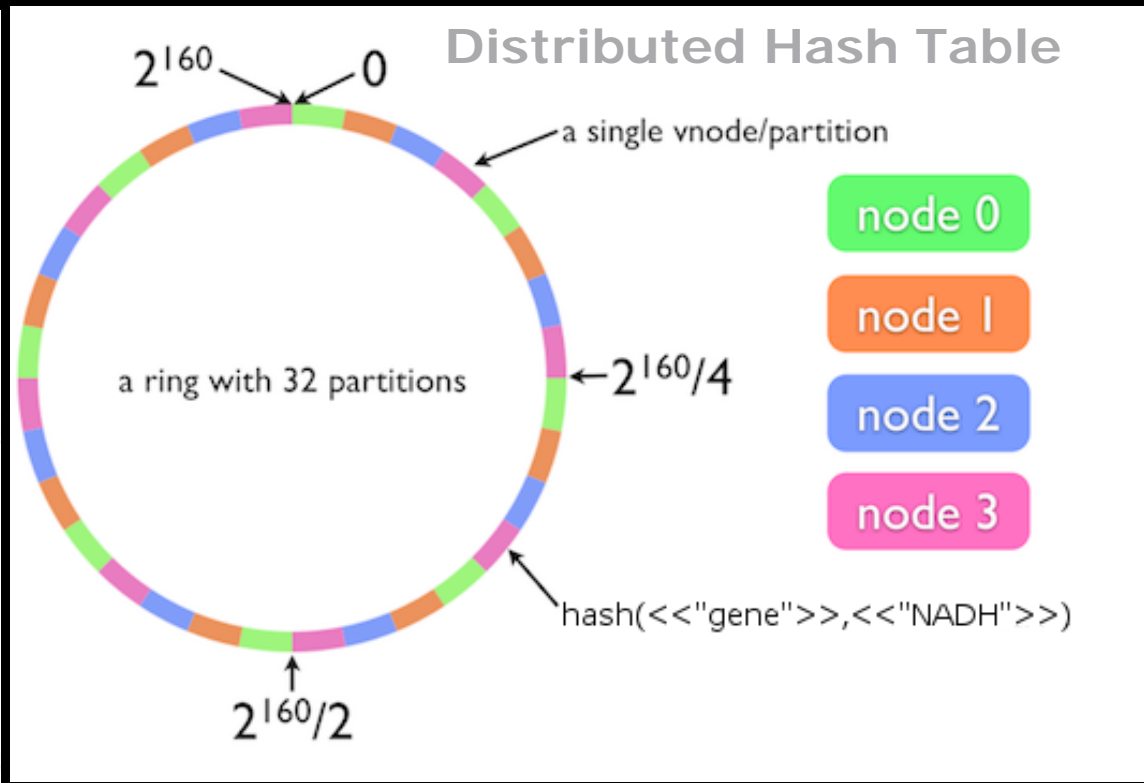
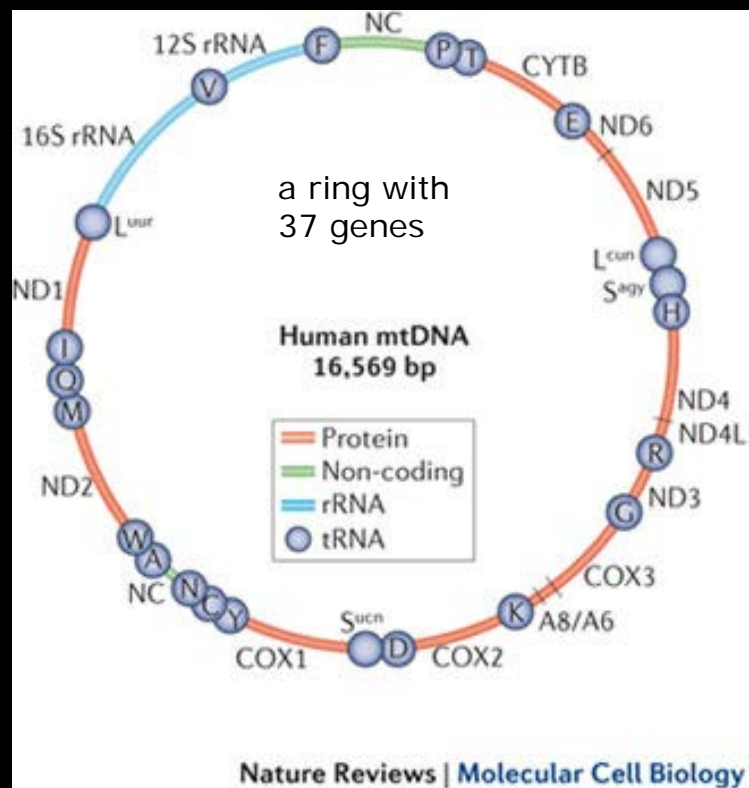
Nuclear DNA: File?



- ❑ **Large:** ~20,000 genes per cell
- ❑ **Dynamic:** Append, Overwrite, Truncate...
- ❑ **Semantics:** Strict (protects behavior at scale)
- ❑ **Consistent:** No (unique, diploid, large numbers)
- ❑ **Metadata:** Fixed (semantics, replication)
- ❑ **View:** One-to-Many (complex interactions)



Nuclear DNA Properties



Mitochondrial DNA: Object?

- ❑ **Small:** ~40 genes per cell (500x smaller)
- ❑ **Static: Constancy, Energy functions**
- ❑ **Semantics: single origin**
- ❑ **Consistent: Yes** (static, exp. fewer interfaces)
- ❑ **Metadata: System based** (X chromosome)
- ❑ **View: One-to-One** (minimal functions)



Mitochondrial DNA Properties

Function	File	Object
Behavior	Dynamic append, overwrite, truncate, etc.	Static* Practical for CAS and Cloud
ID	Path + Name Uniquely identified	URI Acct:DB, Obj:File, MD:file attr
Hierarchy	Strict Semantics protects behavior	None* by name, not enforced by system
Consistency	No* Not Eventually Consistent	Yes Static content behavior (above)
Metadata (MD)	Fixed Semantics based	Arbitrary System based
View to Consumer	One-Many "read-after-write" consistency	One-One* Two consumers may have diff ver




File versus Object

* challenges
Source: EMC Isilon
Office of CTO



System



7,500 named parts

206 regularly occurring bones
(newborns close to 300)

About 640 skeletal muscles (320 pairs)

60+ organs **Distributed Cluster**
(depending on how they are classified)

37 trillion cells (15-70 range) *

Source: Federative Committee on Anatomical Terminology (1998), Terminologia Anatomica. Theime, Stuttgart, Germany

*Source: Bianconi E, et al, "An estimation of the number of cells in the human body.", Annals of Human Biology, 2013 Nov-Dec;40(6):463-71



The Human Body

EMC²

max 10^{45} states

at $4 \cdot 10^{53}$ state-changes/sec*

{send , receive} event
{start , cancel} timer
decision()

* Source::Burks AW, "Logic, Computers, and Men", Proceedings and Addresses of the American Philosophical Association, Vol. 46 (1972 - 1973), pp. 39-57



cellular **Finite State Machine**

EMC²

- **Einstein:** transfer < speed of light c
- **Landauer:** erasure \equiv heat $[kT \cdot \ln 2 / \text{erased bit}] @$ temp T
- **Margolus & Levitin:** process rate < $[4E/h] @$ Energy E

Speed of light $c = 2.98 \times 10^8$ m/s

Boltzmann's constant $k = 1.4 \times 10^{-23}$ Joule/Kelvin

Planck's constant $h = 6.6 \times 10^{-34}$ Joule/Hertz

*Source: Beenakker C, "Hempel's dilemma and the physics of computation", published in: Knowledge in Ferment: Dilemmas in Science, Scholarship and Society (Leiden University Press, 2007)



Information Constraints

EMC²

- ❑ **SAFETY** under all conditions: Apoptosis
- ❑ **AVAILABILITY**: billions of replicate copies
- ❑ **NOT TIMING DEPENDENT**: Asynchronous
- ❑ **COMMAND COMPLETION**: 10 base errors in every 10,000 protein translations (10 AAs/sec)

Derived from: Lamport L, "Paxos made simple", Nov 2001

Ongaro D and Osterhaut J, "In Search of an Understandable Consensus Algorithm", Proc ATC

'14., USENIX Annual Technical Conference, May 20 2014



Consensus in Cell Biology

EMC²

- ❑ **Object: Maternal, Static, Haploid**
- ❑ **Object: Small, Simple, Energy, *Early***

- ❑ **File: Maternal and Paternal, Diploid**
- ❑ **File: Scalable, Dynamic, Complex**

- ❑ **All cells are female first**

Inferences & Thoughts



- ❑ **Implementation and vendor incompatibilities**
- ❑ **Location transparency**
- ❑ **Design and Process deficiencies**
- ❑ **Implementation issues, Geo management**
- ❑ **Firewalls and Security**
- ❑ **Code and Data migration, validation**
- ❑ **Performance**

Source: Wikipedia, Joshi S notes from 1998

*Common Object Request Broker Architecture

EMC²®