PERSPECTIVE

Is junk DNA bunk? A critique of ENCODE

W. Ford Doolittle[1]

Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, NS, Canada B3H 4R2

Edited by Michael B. Eisen, Howard Hughes Medical Institute, University of California, Berkeley, CA, and accepted by the Editorial Board February 4, 2013 (received for review December 11, 2012)

Diana Lin[1,2]

MEDG 505
Jan. 24, 2020

[1] Canada's Michael Smith Genome Sciences Centre, BC Cancer, Vancouver, BC, Canada
[2] Bioinformatics Graduate Program, University of British Columbia, Vancouver, BC, Canada

Overview

Background
Main Critiques
Conclusions
Critique
Discussion Questions

Background

The human genome is packed with at least four million gene switches that
of DNA that were
but that t
contr
tis
pla

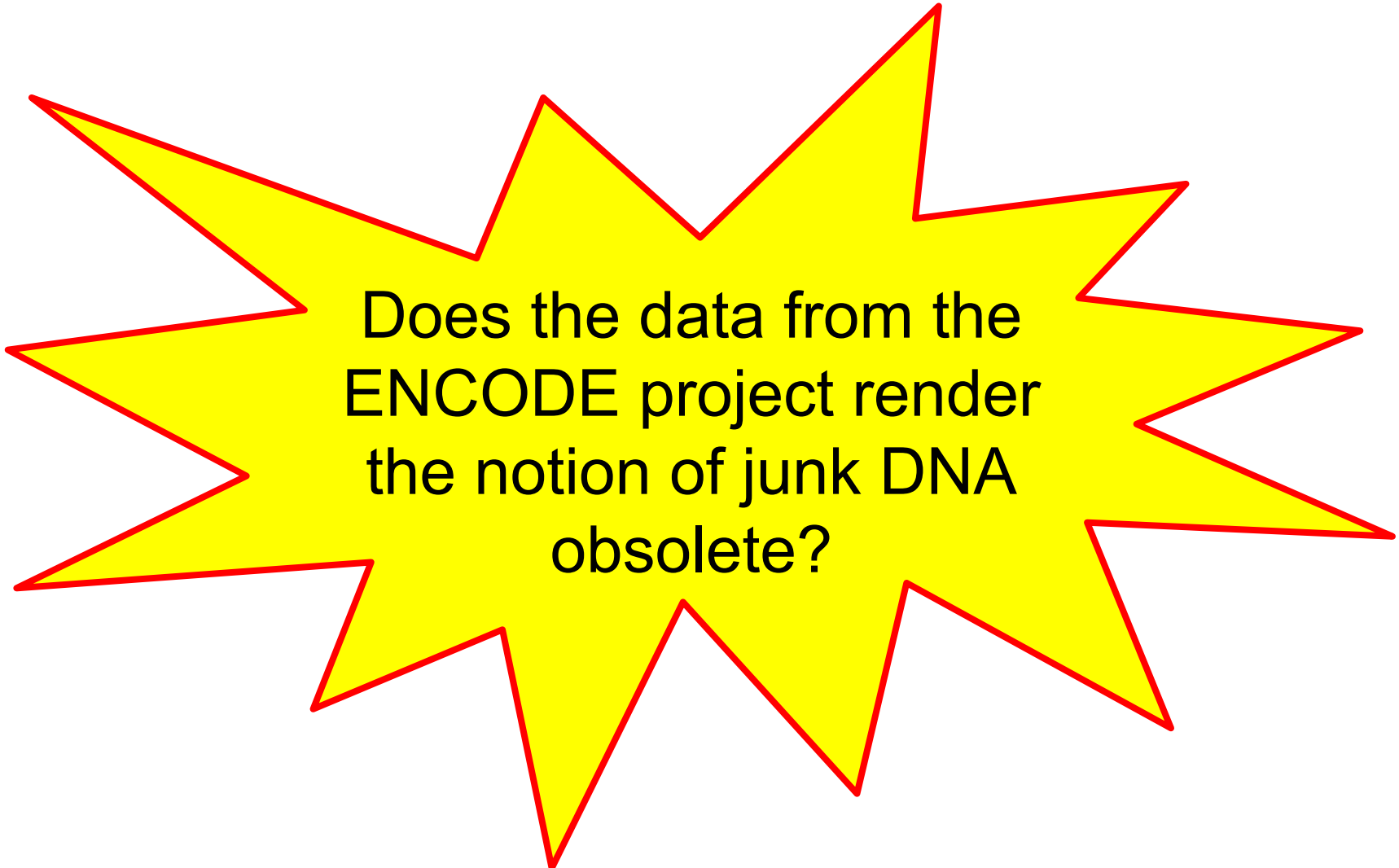
One of the more remarkable findings described
entrée paper is that 80% of
nts linked to
widely
stly

ENCODE project writes

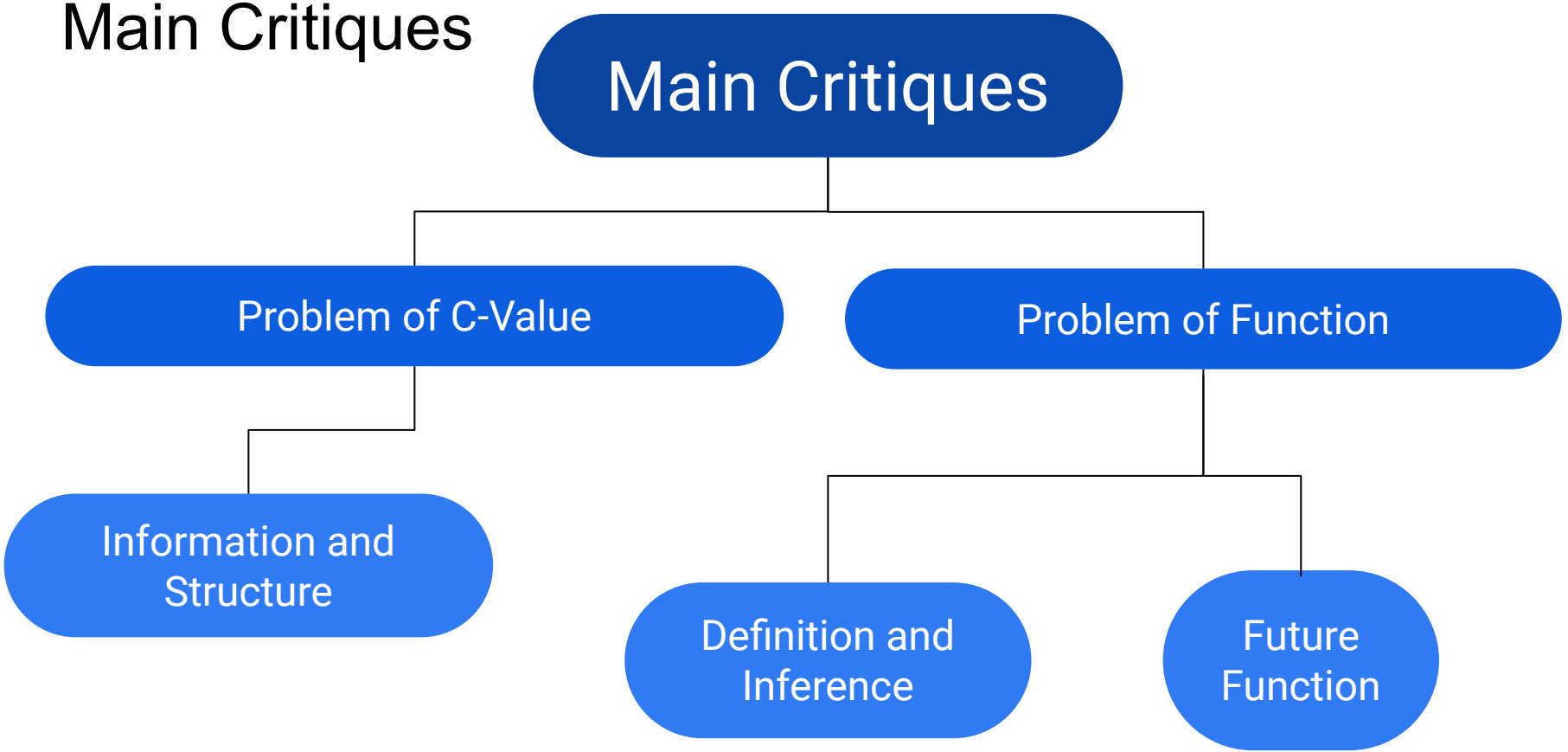
eulogy for junk DNA.

Science

whether by
or by regulating the degree of transcription and consequently the concentrations and function of all proteins.

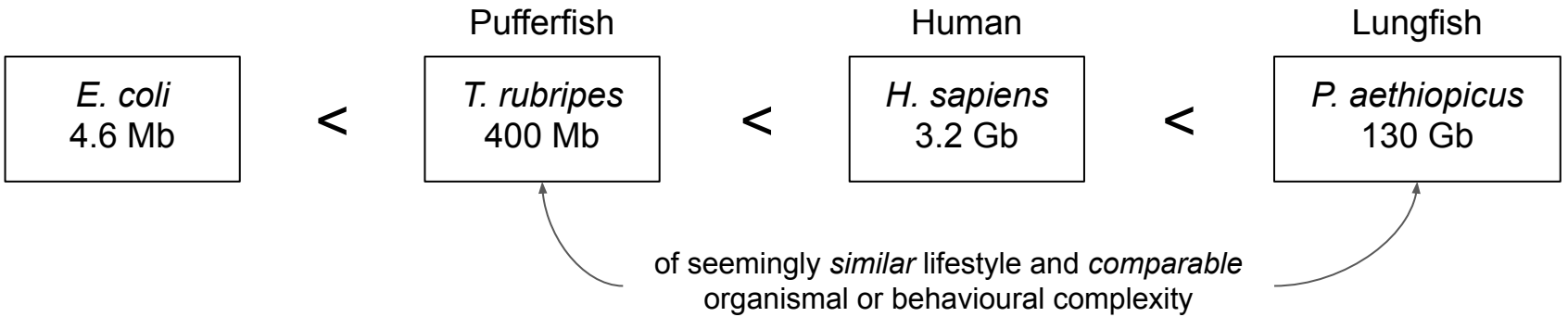

Does the data from the ENCODE project render the notion of junk DNA obsolete?

Main Critiques



Information and Structure

- **"C-Value" Paradox:** the observation that the DNA amounts/complexities correlate very poorly with organismal complexity/evolutionary "advancement"



DEFINITION
C-Value: haploid nuclear DNA content

Information and Structure

The term ‘junk’ DNA is attributed to Ohno, *Brookhaven Symp Biol,* 1972.

What is the current (2012) definition of ‘junk’ DNA?

- ➢ DNA that does not encode information promoting survival and reproduction
- ➢ DNA whose presence cannot be reasonably explained by natural selection at the organism-level for encoded informational roles

Does this include…

- Transposable elements?
- Regulatory regions?
- UTRs?

Information and Structure

- DNA **does** have a basic structural role to play, that is unlinked to
  - specific biochemical activities
  - encoding of information relevant to genes and expression
- Example: Centromeres and telomeres

Are “nucleo-skeletal/nucleotypic” roles considered ‘junk’ DNA? (Cavalier-Smith, *J Cell Sci,* 1978)

- DNA as a macromolecule gives shape to chromosomes, which determines division time, coupled to organismal development
- C-value is considered optimized by organism-level natural selection

DEFINITION
C-Value: haploid nuclear DNA content

Information and Structure

There is room within an expanded, pluralistic, hierarchical theory of C-value for much DNA that makes no contribution whatever to survival and reproduction at the organismal level and thus is junk at that level, although it may be under selection at the sub- or supra-organismal levels.

T. Ryan Gregory, University of Guelph
Principal C-value Theorist

Gregory TR (2001) Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. Biol Rev Camb Philos Soc 76(1):65-101

Gregory TR (2003) Variation across amphibian species in the size of the nuclear genome supports a pluralistic hierarchical approach to the C-value enigma. Biol J Linn Soc Lond 79:329-339

Gregory TR (2004) Macroevolution hierarchical theory and the C-value enigma. Paleobiology 30:179-202

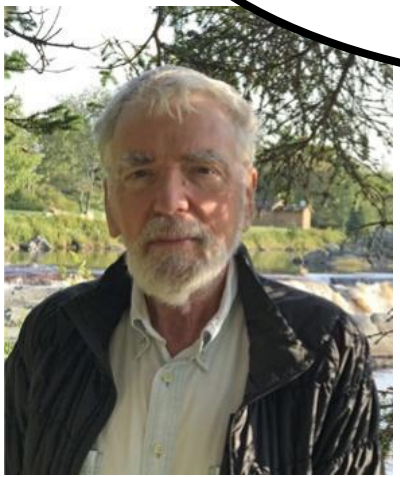
Gregory TR (2005) Synergy between sequence and size in large scale genomics. Nat Rev Genet 6(9):699-708

DEFINITION
C-Value: haploid nuclear DNA content

Information and Structure

DNA fulfilling bulk structural roles remains, in terms of encoded information, is **just junk.**

Junk Advocates



True junk
might be better defined as DNA not currently held to account by selection for any sort of role operating at any level of the biological hierarchy.

W. Ford Doolittle, Dalhousie University

Definition and Inference

How do we define ‘function’, informational or otherwise? (Schwartz, *Philos Sci,* 1999)

➢ The functions of a trait or feature are all and only those effects of its presence for which it was under positive natural selection in the (recent) past and for which it is under (at least) purifying selection now-- known as ‘selected effect’

ENCODE’s Definition

Functional element (FEs): a discrete genome segment that...

- encodes a defined product
- displays a reproducible biochemical signature

Definition and Inference

What about traits that are now under positive selection for one effect but first arose selecting for another trait?

➢ These are called **exaptations.** (Gould and Vrba, *Paleobiology,* 1982)

What about effects that are never selected for, but are side effects? That is, undesirable but apparently unavoidable consequences (e.g. lower back pain in primates walking upright).

➢ These are called **spandrels.** (Gould and Lewontin, *Proc R Soc Land B Biol Sci*, 1979)

What about neutrally selected traits that become positively selected?

Definition and Inference

There are three ways to infer selection, and therefore, infer function:

1. Evolutionary Conservation
2. Experimental Ablation
3. Existence

Definition and Inference

Evolutionary Conservation: if diverse lineages retain a DNA sequence despite the erosive force of mutational divergence, there must be some effect maintained by purifying selection

- The *most* reliable method
- Not universally applicable

Example: Some conserved functions, such as the complementary base-pairings that maintain ribosomal RNA secondary structures, do not require primary sequence conservation.

Definition and Inference

Experimental Ablation: whatever organism-level effect *E* does not occur after deleting or blocking the expression of a region *R* of DNA is taken to be the latter's function (e.g. knock-outs to determine function)

- A causal role (CR) definition of function
- However, causation does not equal function

Example: If you stop a heart, the thumping noise disappears, but the function of the heart is not to create a thumping noise!

Definition and Inference

Existence: the presence of a structure or the occurrence of a process or detectable interaction, is taken as adequate evidence for its being under selection

- The *least* reliable method

Example: Introns and endogenous viruses must be doing us some good since they are present. All point mutations must have a selection-based explanation.

These points are NOT true-- but this method still informs much of molecular and evolutionary genetics.

Definition and Inference

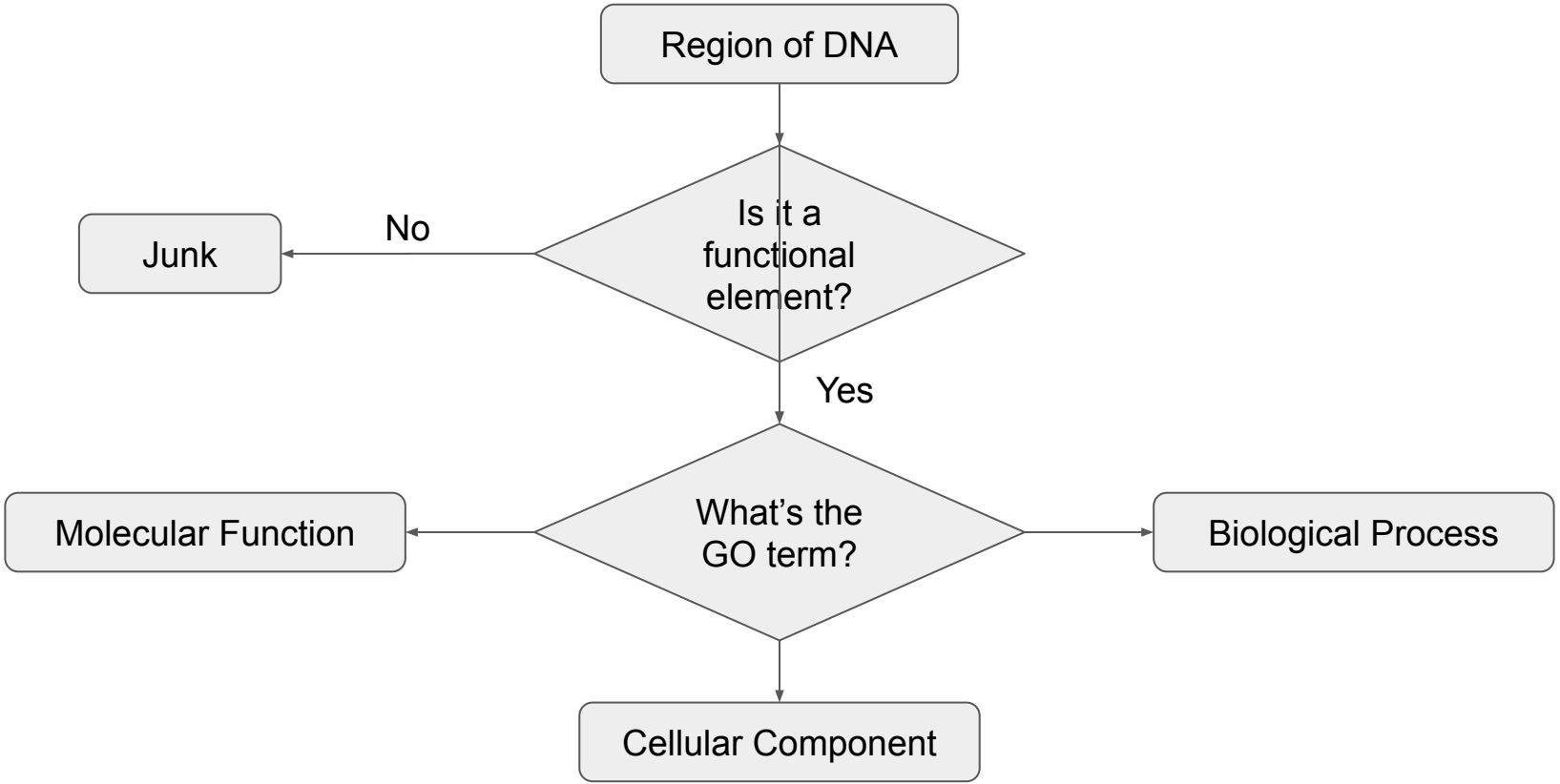
What method of function inference does the ENCODE project use?

➢ ENCODE mainly uses the third method: **Existence**, according to Doolittle.

Although functional elements defined by ENCODE may also be evolutionarily conserved, a lot of them neutrally selected, and exist with unknown purpose.

In the essay *The Frailty of Adaptive Hypothesis for the Origins of Adaptive Complexity* (2007), Lynch argues that **Existence** is no longer a viable means of inferring function as recent observations from genomic sequencing and population genetics has shown otherwise.

Definition and Inference



Definition and Inference

This is wisely worded by Stamatoyannopoulos (*Genome Res* 2012):

- "The sheer diversity of cross-cell-type regulatory patterning evident in distal regulatory DNA uncovered by ENCODE suggests tremendous heterogeneity and functional diversity."

- "ENCODE is thus in a unique position to promote clearer terminology that **separates** the identification of functional elements per se from the ascription of specific functional activities using historical experimentally defined categories."

Future Function

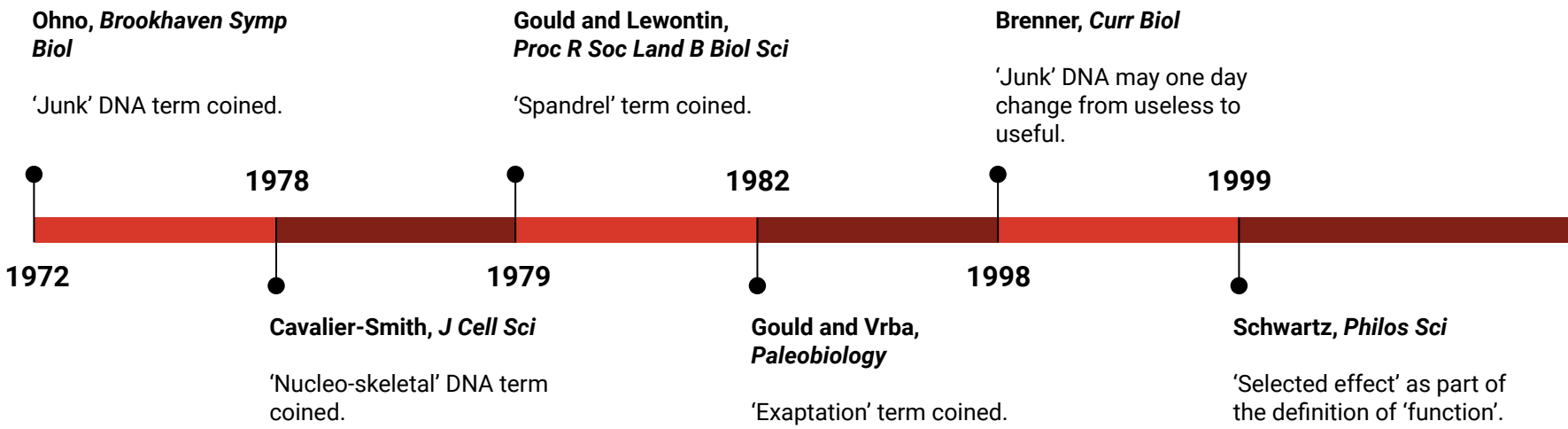
Currently considered (neutrally selected) 'junk' DNA could potentially become regulatory RNAs through years of evolution, as speculated by Barroso (*Nature*, 2012):

- "it seems that widespread transcription from non-coding DNA potentially acts as a reservoir for the creation of new functional molecules, such as regulatory RNAs"

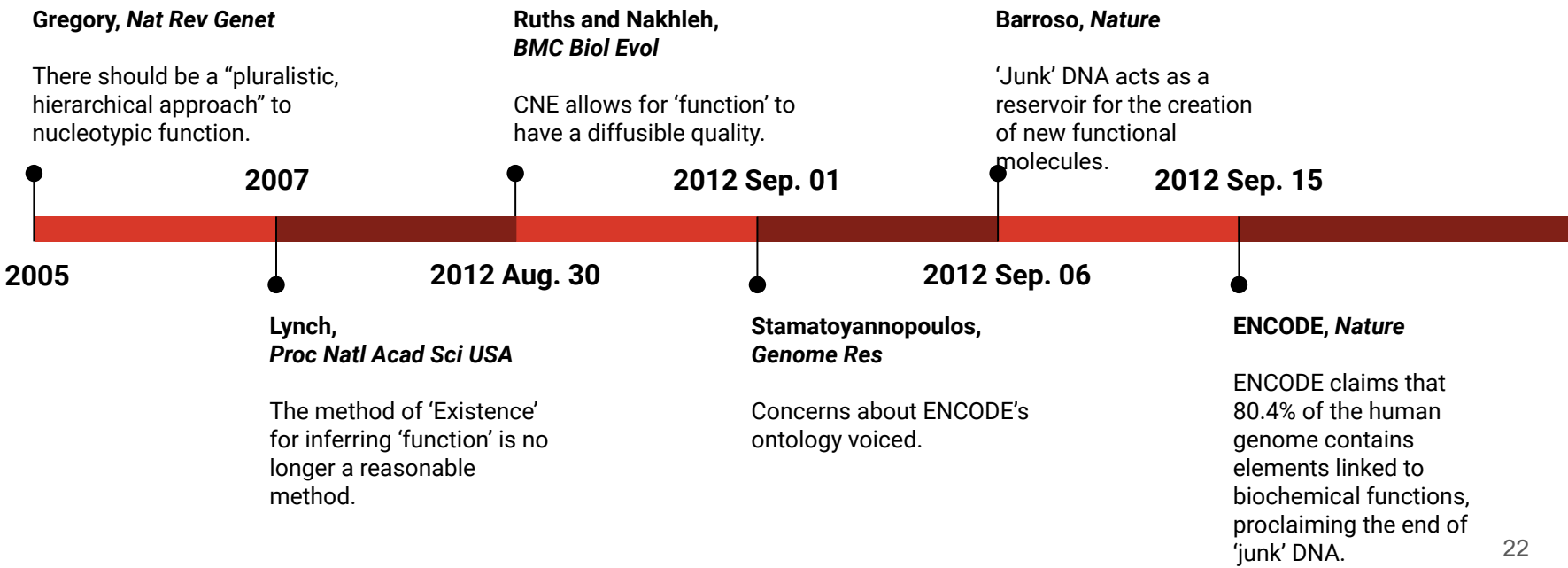
Brenner (*Curr Biol,* 1998) has speculated this years before Barroso:

- "patently useless features of existing organisms are there as an investment for the future"

So Is Junk Bunk?



So Is Junk Bunk?



So Is Junk Bunk?

Answer: Depends on what your definition of ‘junk’ and ‘function’ are!

But generally, there will be DNA whose presence we cannot *currently* explain that may fall under the term ‘junk’ DNA, Doolittle insists.

Conclusion: The field of biology, genomics, genetics, etc. needs to standardize the definition of ‘junk’ and ‘function’ for the time being, and periodically update those definitions as needed, as new research is published.

‘Junk’ DNA needs its own ontology!

With its own ontology, we could avoid the many misconceptions and misrepresentations that are presented with the ENCODE project.

Critiques of the Critique

PERSPECTIVE

Is junk DNA bunk? A critique of ENCODE

W. Ford Doolittle[1]

Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, NS, Ca[illegible]3H 4R2

Edited by Michael B. Eisen, Howard Hughes Medical Institute, University of California, Berkeley, CA, and [illegible] by the Editorial Board February 4, 2013 (received for review December 11, 2012)

- This critique is *not*…
  - a critique of the methodology [illegible]e ENCODE experiments
  - a critique of the results of [illegible] ENCODE experiments
  - a critique of ENCODE

Critiques of the Critique

Main Critique: When ENCODE commentators proclaim the end of 'junk' DNA, it does NOT mean anything significant, because, *what counts as 'junk'?*

Doolittle's Commentary: "Indeed, there would be no excitement in either the press or the scientific literature if all the ENCODE team had done was acknowledge an established theory concerning DNA's structural importance...

Rather, the excitement comes from interpreting ENCODE's data to mean that a much larger fraction of our DNA than until very recently thought contributes to our survival and reproduction as organisms**,** because it encodes information transcribed or expressed phenotypically in one tissue or another, or specifically regulates such expression."

Critiques of the Critique

Doolittle's Conclusion: "In the end, of course, there is no experimentally ascertainable truth of these definitional matters other than the truth that many of the most heated arguments in biology are not about facts at all but rather about the words that we use to describe what we think the facts might be. However, that the debate is in the end about the meaning of words does not mean that there are not crucial differences in our understanding of the evolutionary process hidden beneath the rhetoric."

Discussion Questions

1. What do you consider ‘junk’ DNA? In your opinion, are structural regions of DNA count as ‘junk’ DNA or ‘functional’ DNA? Why?
2. Should we eliminate the term ‘junk’ DNA? If so, should we replace the term? What would you replace the term with?
3. What should ENCODE have done differently as to not bring about a critique like this one?
4. What is the current (2020) definition of ‘junk’ DNA? Has the term evolved since 2012? How has it evolved?
5. How do you propose we create a new ontology for the study of ‘junk’ DNA and ‘functional’ DNA? What would the ontology terms be?