UNI CERVANTES
FUNDACIÓN UNIVERSITARIA

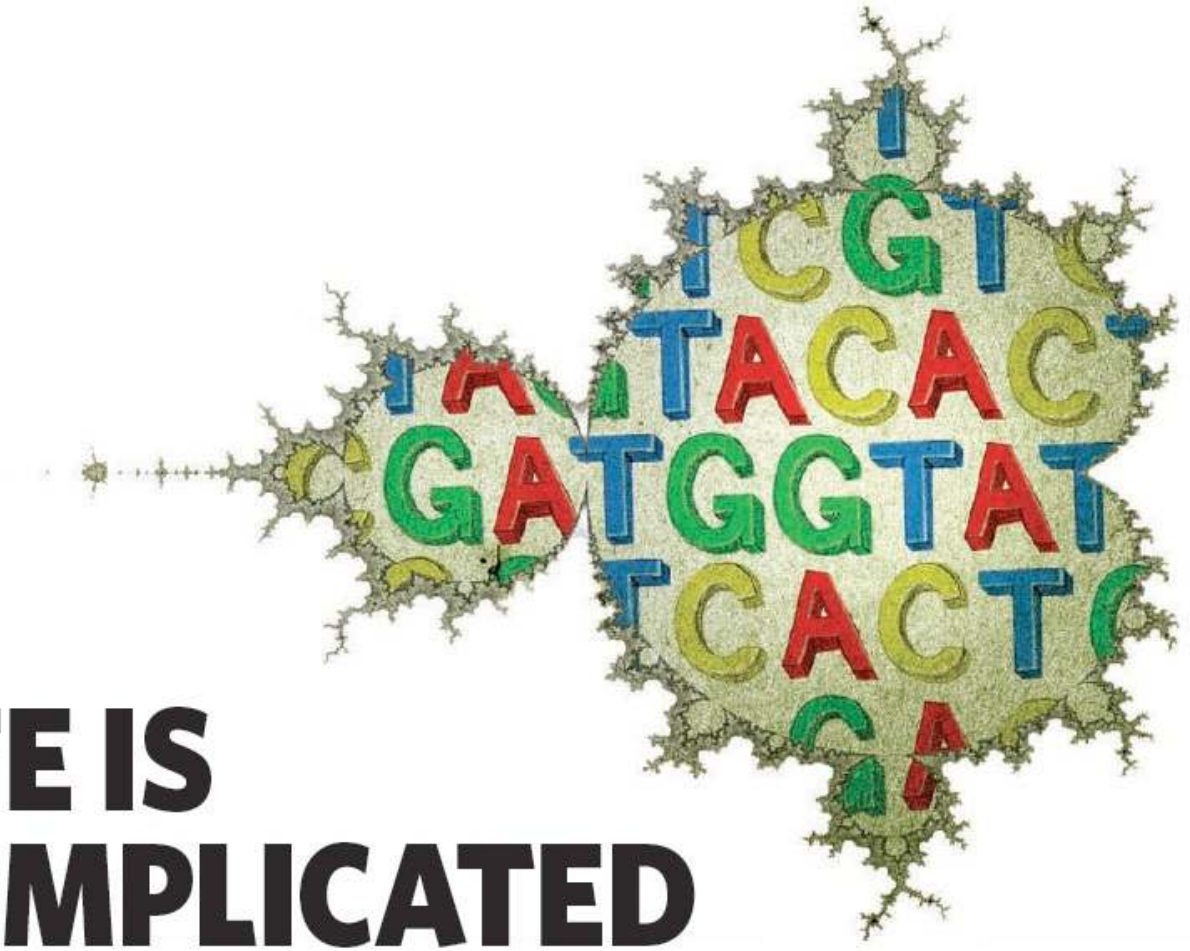
1ER CONGRESO INTERNACIONAL sobre CIENCIA Y FE

What is a 'Gene'? Not a Particle but a Process

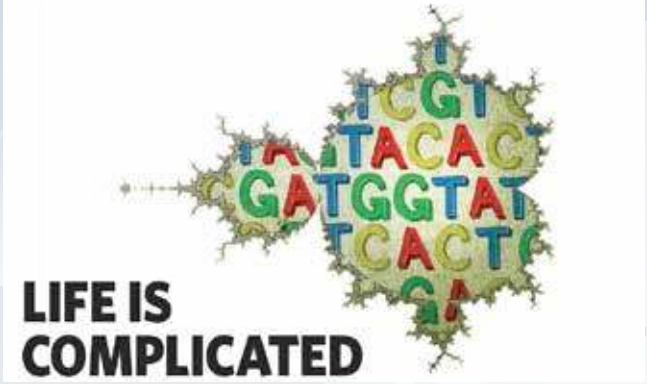
How the 'gene' is a multilevel mediator of 'information' that lacks a material description

Richard v. Sternberg

Biologic Institute



LIFE IS COMPLICATED



LIFE IS COMPLICATED

Few predicted, for example, that sequencing the genome would undermine the primacy of genes by unveiling whole new classes of elements — sequences that make RNA or have a regulatory role without coding for proteins. Non-coding DNA is crucial to biology, yet knowing that it is there hasn't made it any easier to understand what it does. "We fooled ourselves into thinking the genome was going to be a transparent blueprint, but it's not," says Mel Greaves, a cell biologist at the Institute of Cancer Research in Sutton, UK.

Instead, as sequencing and other new technologies spew forth data, the complexity of biology has seemed to grow by orders of magnitude. Delving into it has been like zooming into a Mandelbrot set — a space that is determined by a simple equation, but that reveals ever more intricate patterns as one peers closer at its boundary.

Just one decade of post-genome biology has exploded that view. Biology's new glimpse at a universe of non-coding DNA — what used to be called 'junk' DNA — has been fascinating and befuddling. Researchers from an international collaborative project called the Encyclopedia of DNA Elements (ENCODE) showed that in a selected portion of the genome containing just a few per cent of protein-coding sequence, between 74% and 93% of DNA was transcribed into RNA[2]. Much non-coding

the University of Toronto in Ontario. "Now, we appreciate that the signalling information in cells is organized through networks of information rather than simple discrete pathways. It's infinitely more complex."

A brief history of inheritance as particulate in nature

The earliest concepts of particulate "trait transmission":

Buffon, Diderot, Maupertius; *atomisme*:

"Could one not explain by that means [mutation] how from two individuals alone the multiplication of the most dissimilar species could have followed? They could have owed their first origination only to certain fortuitous productions, in which the **elementary particles** failed to maintain the order they possessed in the father and mother animals; each degree of error would have produced a new species; and by reason of repeated deviations would have arrived the infinite diversity of animals we today." Maupertius, *Système de la Nature* 1751

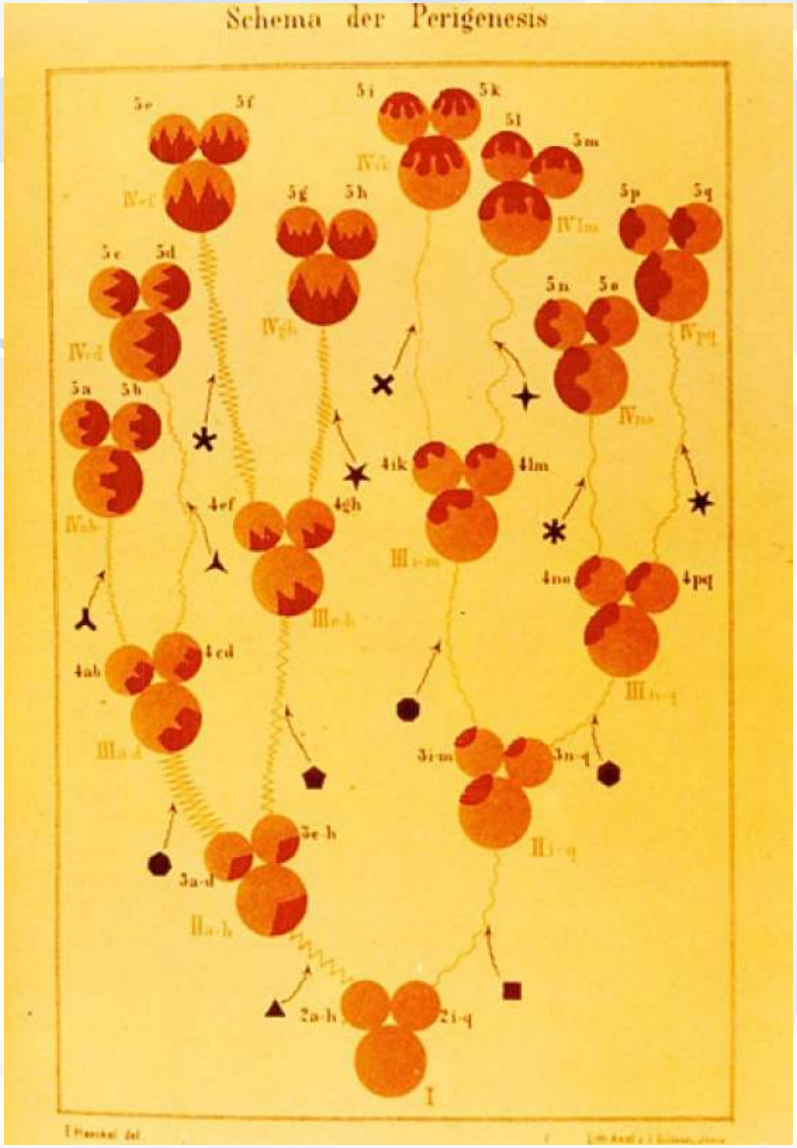
The earliest concepts of particulate "trait transmission":

Physiological units: "The germ cells are essentially nothing more than vehicles in which are contained small groups of the physiological units in a fit state for obeying their proclivity towards the structural arrangement of the species they belong to." Herbert Spencer, *Principles of Biology*, 1864

Gemmules: Cells "throw out minute granules which are dispersed throughout the whole system; that these supplied with proper nutriment, **multiply by self-division** and are ultimately developed into units like those from which they were orginally derived. These granules may be called gemmules." Charles Darwin, *Animals and Plants under Domestication*, 1867

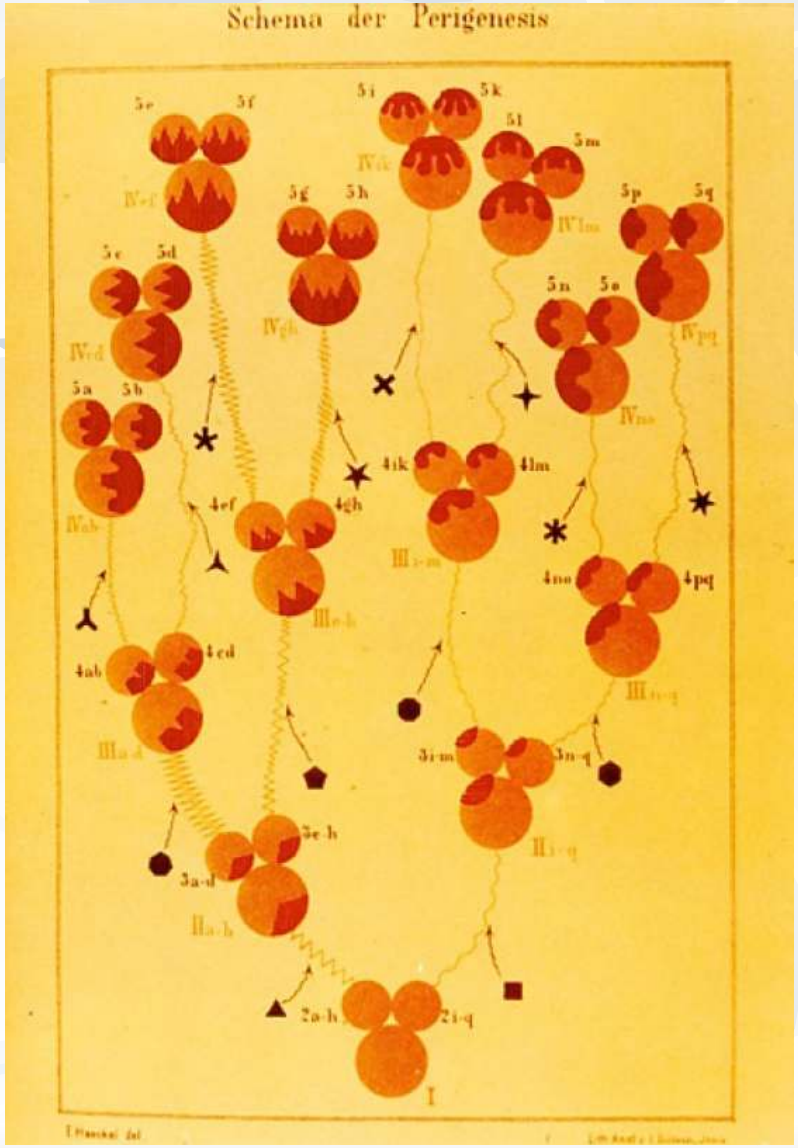
The earliest concepts of particulate “trait transmission”:

Ernst Haeckel, *The Perigenesis of the Plastidule, or the Wave Generation of the Small Vital Particle: An Attempt at a* ***Mechanical Explanation*** *of the Elementary Processes of Development*, 1876; Nucleus as the source of traits, 1866



The earliest concepts of particulate “trait transmission”:

The “plastidule” could **‘self-develop’**, **‘self-propagate’**, and **‘self-transform’** because it was endowed with an *atom-soul* said Haeckel; a materialistic-monistic soul



The earliest concepts of particulate "trait transmission":

Carl v. Nägeli, *Mechanischphysiologie Theorie der Abstammungslehre* (*Mechanical-Physiological Theory of the Doctrine of Descent*), 1884:

- "Idioplasm" as inherited, material substance

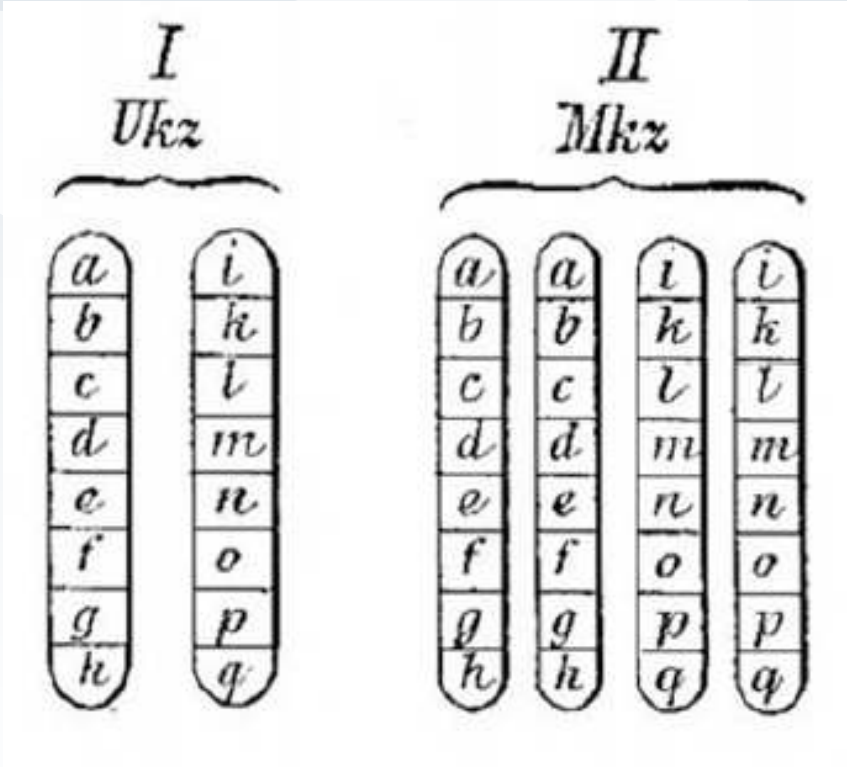
Oskar Hertwig, 1884: Chromatin of a nucleus = "idioplasm"

Hugo de Vries, *Intracelluläre Pangenesis*, 1889:

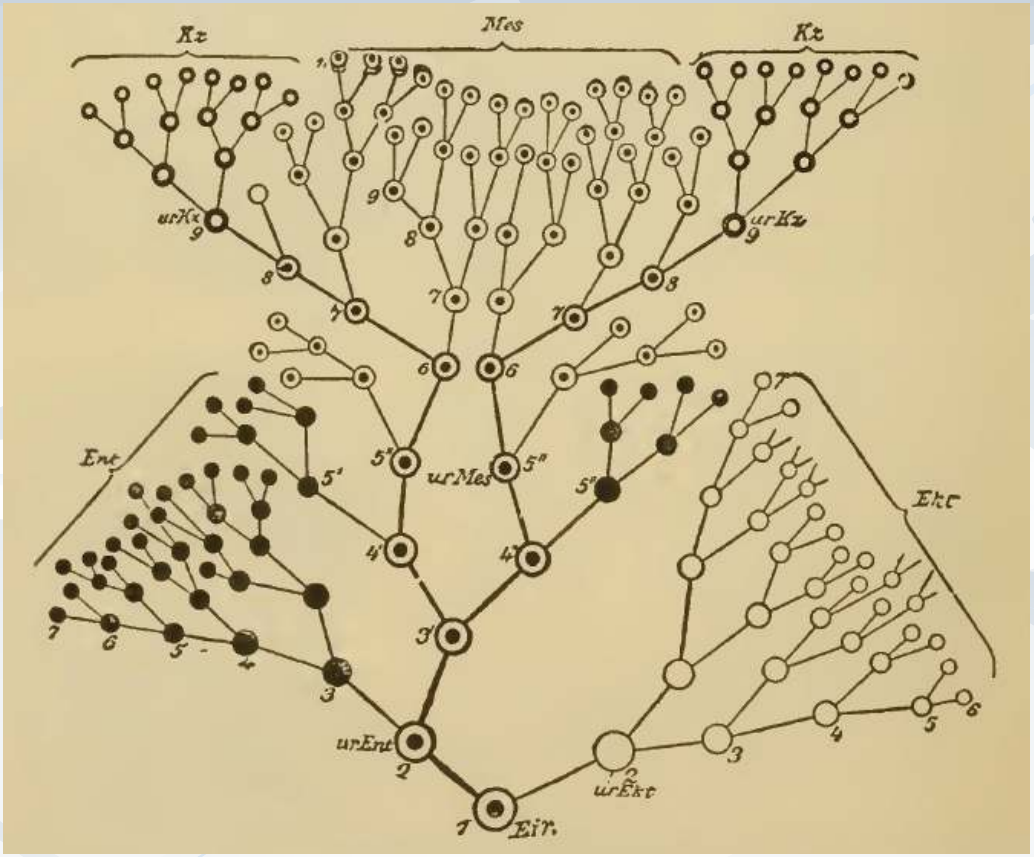
- Basic discreteness of the hereditary particles
- "Independence and miscibility, these are the essential properties of the hereditary dispositions [*erbliche Anlagen*] of all organisms"
- "Pangenes" are endowed with the basic characteristics of life, that is the faculties of assimilation and multiplication and can undergo potentially unlimited variation

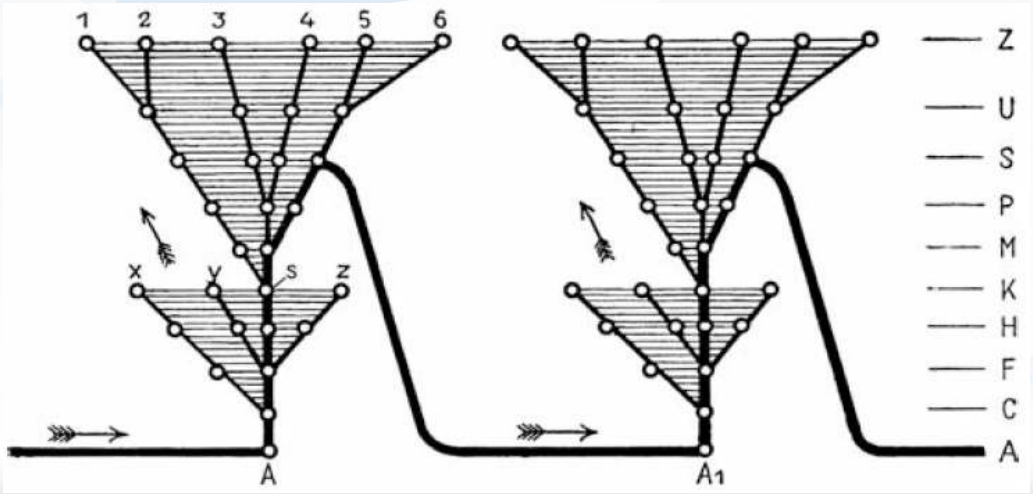
The earliest concepts of particulate "trait transmission":

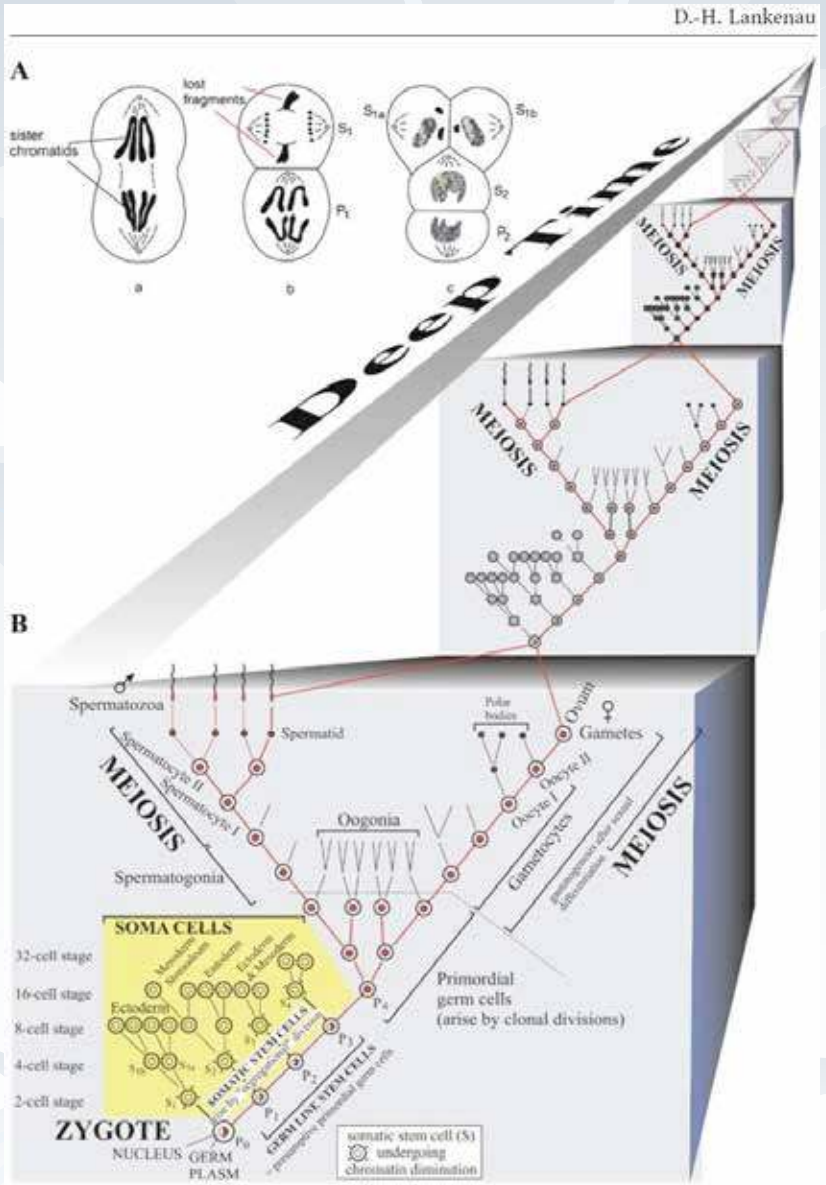
August Weismann, *Das Keinplasm* (*The Germ Plasm*), 1892: To explain the process of heredity and development, Weismann speculated that all the cells of the body, both germ and somatic, contained a hierarchical series of particulate elements, starting with what he called "idants," the visible chromosomes in the nucleus.



The Legacy of the Germ Line







"The most important concept in genetics is the view that the organization and activities of living matter rest on a system of *self-replicating living* units. Living units ("physiological units," "gemmules," "micellae," "pangenes") appeared in biological thinking long before Johannsen introduced the term "gene" in 1909." (L. C. Dunn, 1991)

"These various approaches were all a reflection of an underlying commitment to mechanistic materialism that, by 1900, characterized biology in general, and the study of heredity in particular." (G. E. Allen 2014)

A brief history of inheritance as being “more than matter” in nature

Characters and their inheritance

Gregor Mendel, *Versuche über Pflanzen-Hybriden* (*Experiments in plant hybridization*), 1866: A formal law:

- *Merkmal*(*e*): 157x
- *Anlage*: 1x
- *Elementen*: 10x
- *Beschaffenheit*: 10x

Enter William Bateson (the man who coined the terms *genetics*, *allelomorph*, *heterozygote*, *homozygote*, …)

After an empirical and exhaustive survey of "trait transmission, he presented his findings in *Materials for the Study of Variation*, 1894, wherein he wrote:

"I have as far as possible avoided any use of the terms Heredity and Inheritance. These terms which have taken so firm a hold on science and on the popular fancy, have had a mischievous influence on…biological thought. They are of course metaphors from the descent of property, and were applied to organic descent in a time when the nature of the process of reproduction was wholly misunderstood. The metaphor from the descent of property is inadequate chiefly for two reasons.

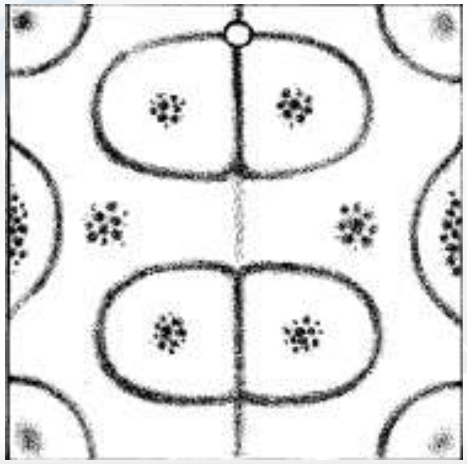
Enter William Bateson (the man who coined the terms *genetics*, *allelomorph*, *heterozygote*, *homozygote*, …)

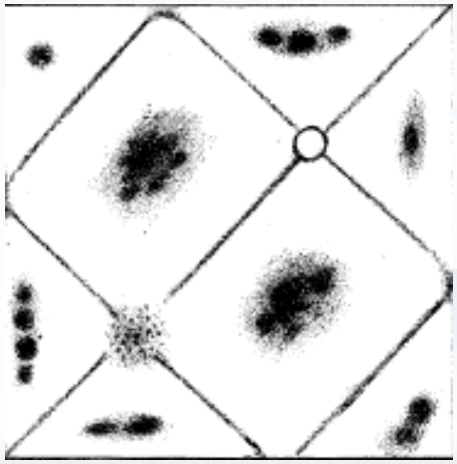
"First, by emphasizing the fact that the organization of the offspring depends on material transmitted to it by its parents, the metaphor of Heredity, through an almost inevitable confusion of thought, suggests the idea that the actual body and constitution of the parent are thus in some way handed on. No one perhaps would now state the facts this way, but something like this material view of Descent was indeed actually developed into Darwin's Theory of Pangenesis. From this suggestion that the body of the parent is in some sort remodeled into that of the offspring, a whole series of errors is derived. …

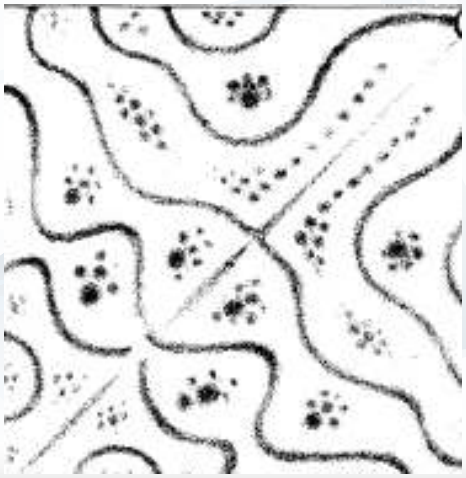
"Secondly, the metaphor of Heredity misrepresents the essential phenomenon of reproduction. In the light of modern investigations, and especially those of Weismann on the continuity of the germ-cells, it is likely that the relation of parent to offspring, if it has any analogy with the succession of property, is rather that of trustee than of testator."

The paper the introduced Mendel's Principles to the Anglophone world

Bateson, W & Saunders E, *The facts of heredity in the light of Mendel's discovery*, 1902: "*We have no warrant for regarding any hereditary character as depending on a material substance for its transmission*." (Cf. Chladni figures)







Why say such? Well, we read in 1896(!) in Edmund Wilson's *The Cell in Development and Heredity*

"Now, chromatin is known to be closely similar to, if not identical with, a substance known as nuclein…composed of **nucleic acid** (a complex organic acid rich in phosphorus) and albumin [protein]. And thus we may reach the remarkable conclusion that inheritance may, perhaps, be effected by the physical transmission of a particular chemical compound from parent to offspring."

THEN in 1909/1991 the terms ‘gene’ and ‘genotype’ were given by Wilhelm Johannsen

J Physiol 592.11 (2014) pp 2431–2438

TOPICAL REVIEW

The holist tradition in twentieth century genetics. Wilhelm Johannsen’s genotype concept

Nils Roll-Hansen

Professor emeritus Nils Diderik Roll-Hansen University of Oslo Department of philosophy, history of art and Nydalen Alle 5, apt. 111 Oslo 0484 Norway

Abstract The terms ‘genotype’, ‘phenotype’ and ‘gene’ originally had a different meaning from that in the Modern Synthesis. These terms were coined in the first decade of the twentieth century by the Danish plant physiologist Wilhelm Johannsen. His bean selection experiment and his theoretical analysis of the difference between genotype and phenotype were important inputs to the formation of genetics as a well-defined special discipline. This paper shows how Johannsen’s holistic genotype theory provided a platform for criticism of narrowly genocentric versions of the chromosome theory of heredity that came to dominate genetics in the middle decades of the twentieth century. Johannsen came to recognize the epoch-making importance of the work done by the Drosophila group, but he continued to insist on the incompleteness of the chromosome theory. Genes of the kind that they mapped on the chromosomes could only give a partial explanation of biological heredity and evolution.

(Received 1 February 2014; accepted after revision 5 April 2014;

THEN in 1909/1991 the terms ‘gene’ and ‘genotype’ were given by Wilhelm Johannsen

Elemente der exakten Erblichkeitslehre (*Introduction to the Elements of an Exact Theory of Heredity*), 1909; *The genotype conception of heredity*, 1911: An accurate concept of heredity could not have as either its axioms, or as it corollaries, the hypotheses “that ‘elements’ in the zygote correspond to *special organs*,” and that “*discrete particles of the chromosomes* are ‘bearers’ of special parts of the whole inheritance in question.” Indeed, for him as for William Bateson the idea that “discrete particles of the chromosomes are ‘bearers’ of special organizing functions in…ontogenesis, a chromatin-particle in the nucleus of a gamete being in some way the representative of an organ or a group of tissues,” was only speculation.

Why ‘gene’ and ‘genotype’ and ‘phenotype’?

“Of all the Weismannian armory of notions and categories it [the concept of inheritance that is least speculative] may use nothing. It is a well-established fact that language is not only our servant, when we wish to express—or even to conceal—our thoughts, but that it may also be our master, overpowering us by means of the notions attached to certain words. This fact is the reason why it is desirable to create a new terminology in all cases where new or revised conceptions are being developed.”

Why ‘gene’ and ‘genotype’ and ‘phenotype’?

A ‘gene’ is a hypothesis-less placeholder:

“By no means have we the right to define the gene as a morphological structure in the sense of Darwin’s gemmules or biophores or determinants or other speculative morphological concepts of that kind. Nor have we any right to conceive that each special gene (or a special kind of genes) corresponds to a particular phenotypic unit-character…of the developed organism.”

“The conception of a gene as an organoid, a little body with independent life and similar attributes, is no longer to be considered. Assumptions which would make such a conception necessary, fail utterly. Putting a horse in a locomotive as a cause of its motion—to use Lange’s classical example—is just as “scientific” an hypothesis as the organoid “explanation” of heredity.”

Why ‘gene’ and ‘genotype’ and ‘phenotype’?

“The question of *chromosomes* as the presumed ‘bearers of hereditary qualities’ seems to be an idle one. I am not able to see any reason for localizing “the factors of heredity” (*i.e.*, the genotypical constitution) to the nuclei. The organism is in its totality penetrated and stamped by its genotype-constitution.”

“The genotype-conception is…an ‘ahistoric’ view of the reactions of living beings—of course only as far as true heredity is concerned. This view is an analog to the chemical view; chemical compounds have no compromising ante-act, H_2O is always H_2O, and reacts always in the same manner, whatsoever may be the “history” of its formation or the earlier states of its elements. I suggest that it useful to emphasize this “radical” ahistoric genotype-conception of heredity in its strict antagonism to the transmission- or phenotype-view.”

So how did we end up with a Haeckelian-Weismannian 'gene'?

Short answer: Thomas Morgan, Alfred Sturtevant, Hans Muller, and Calvin Bridges, *The Mechanism of Mendelian Heredity*, 1915 and thereafter:

"Exception may perhaps be taken to the emphasis we have laid on the chromosomes as the material basis of heredity. … But it should not pass unnoticed that even if the chromosome theory be denied, *there is no result dealt with in the following pages that may not treated independently of the chromosomes*; for *we have made no assumption concerning heredity that cannot also be made abstractly without the chromosomes as bearers of the postulated hereditary factors*."

"The factorial [Mendelian] theory *as such* deals with the behavior of its factors *in an abstract way, quite apart from any material basis on which they may happen to be composed*. In this way it may measure their constancy, segregation, linkage, etc."

So how did we end up with a Haeckelian-Weismannian 'gene'?

The Mechanism of Mendelian Heredity, 1915:

"Why then, we are often asked, do you drag in the chromosomes? Our answer is that since the chromosomes furnish exactly the kind of mechanism that the Mendelian laws call for; and since there is an ever-increasing body of information that points clearly to the chromosomes as the bearers of the Mendelian factors, it would be folly to close one's eyes to so patent a relation. Moreover, as biologists, we are interested in heredity *not primarily as a mathematical formulation* but rather as a problem concerning the cell, the egg, and the sperm. ."

"Weismann's ideas of *heredity* concerning the segregation in the reduction divisions of the egg and sperm of inherited materials present in the chromosomes, furnish the basis of our present attempt to explain heredity in terms of the cell."

The ‘gene’ problem *per* mathematicians:

- Erwin Schrödinger (1944): A ‘gene’ has to combine all four Aristotelian causes (material, efficient, *formal*, and *final*).
- John von Neumann (1950s-1950s): If you consider a ‘gene’ as being a program, then it is teleological.
- A. C. R. Dean, Sir Cyril Hinshelwood (1940s-1950s): The ‘gene’ of heredity is imbued with “mystical” properties.
- Nicholas Rashevsky (1959): “[A] DNA molecule first splits in two parts which stand to each other in the relation of a pattern to its mold. Then each part rebuilds a full new DNA molecule by reconstructing the “opposite” part. In other words, before the DNA molecule multiplies, *it simply ceases to exist as such*. If a gene is a DNA molecule or a combination of DNA molecules, then although the identity and individuality of each gene may be preserved through generations, this preservation is not of a static but of a dynamic nature. The “mother gene” has actually ceased to exist before the two “daughter genes” are formed.[1959; emphasis his] (Cf. Max Delbrück in 1949.)
- Robert Rosen (1959-1998): ‘Gene’ ‘replication is dependent on a formal cause.

The ‘gene’ problem *per* mathematicians:

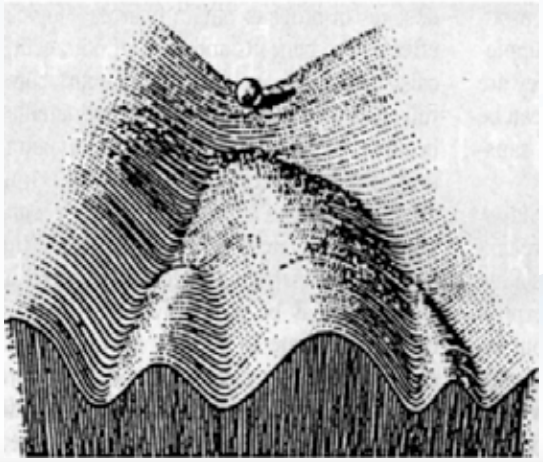
- Robert Rosen (1958, 1962): From the standpoint of classical set theory, ‘gene’ ‘self-replication’ entails a logical paradox.
- Eugene Wigner (1961): The probability of evolving a replicator *de novo* is (nearly) zero; this is now known to be impossible; see the *no-cloning* and *no-deletion theorems* of quantum information. (Strict conservation of information.)

The 'gene' problem *per* mathematicians:

- Lars Löfgren (w.r.t. *Rosen's paradox*) (1968): One can find a logically consistent scheme of a "gene-replicator" provided that one steps out of classical logic/set theory (von Neumann-Bernays-Gödel theory)—the very theory on which population genetics is based: "There is, however, one area of biology where self-reproduction, in the complete sense of this paper, is of interest. This is the theory of biological evolution. If we ask not merely how a cell can reproduce in a suitable surrounding, but how this property has evolved, then we are faced with an explanation of reproduction in a complete sense. Our results show that such a theory of evolution cannot be derived with an ordinary logical-mathematical-biological reasoning, but that it instead will have to contain new and independent axioms. The question of how these should be formulated offers interesting problems. (…) Finally we want to emphasize that it is not *a priori* clear that a formal theory of biological evolution should exist. If such a theory could be formulated, however, it still remains to investigate whether it is recursively axiomatizable (that is, whether there are effective methods for the decision if a given well- formed expression is an axiom or not)."[1968]
- Ion Baianu (1970s): Extended the theorems of Löfgren to *category theory* (where they fit quite nicely).

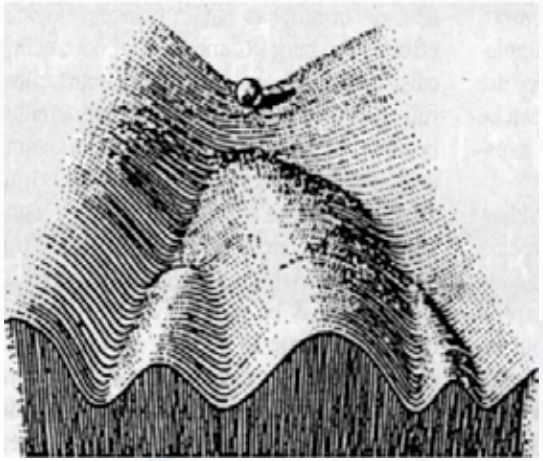
The ‘gene’ problem *per* mathematicians:

- René Thom (1970s-1990s): An investigation of the “epigenetic landscape” of Conrad Waddington by way of differential geometry/topology: “[The] set of living local states is parameterized by a space U (a finite or infinite-dimensional function space); the figure F is embedded in U, but not canonically, encapsulated in some sense in a tubular neighborhood with reflecting walls R_j, and it is the collection of reflecting walls (which describe the mechanisms of epigenesis, regulation, homeostasis, and even morphogenetic regeneration) that forms, strictly speaking, the total geometric structure associated with the species. In the end the spatial form F is only a kind of necessary realization, a compulsory pathway…which develops in a space U of many more dimensions than space-time $\mathbf{R}^4$.” Structural Stability and Morphogenesis, 1975, Pg. 154; emphasis mine



The ‘gene’ problem *per* mathematicians:

- René Thom: The whatness and whereness of ontogenetic information is counter-intuitive: “It may seem difficult to accept the idea that a sequence of stable transformations of our space-time could be directed or programmed by an organizing center consisting of an algebraic structure outside space-time itself.”Structural Stability and Morphogenesis, 1975, Pg. 119



For a hereditary sequence is "directed or programmed by an organizing center consisting of an algebraic structure outside space-time itself" *per* Thom. We are back to the *εντελεχηεια* of Aristotle, Goethe, and Driesch.

Now consider the “writing” of ‘gene’ sequences using inputs that are “junk”

Trypanosoma brucei cytochrome c oxidase III "written" by RNA editing:

AGGAGAGGGGAGGCUUUCGGACCAAGGGAAGGAAGGGAGGUUAAGAAAAAGGAAAAACAAUUUGUGAGGGAGAAGGGUUUUUGGAGGG
GUUUUGGGAAGAGAGGGGUUUUGGGGAAACCAGAUGAGAGUUUUUGCAGAAACAAAGGGGUUUUUGGGCAAAGGGAAUACAAUUUGCU
GAGGGGGGAGAGCGGAAAGAGGGGGGAGAGCGGAAGGAGGAACACGGGAGGGAAGACAGGAUUUAGGAAGCGAGAGAGAGGAGAGGGG
AAAGGGUUUAGUUGGAAUGAAGAGGUAGUUUGUAGGAAGAAAAAAAAAAA

Snapshots of the RNA editing machine in trypanosomes captured at different assembly stages *in vivo*

Monika M Golas[1,2], Cordula Böhm[3], Bjoern Sander[1,4], Kerstin Effenberger[3], Michael Brecht[3], Holger Stark[1,5,*] and H Ulrich Göringer[3,*]

The EMBO Journal (2009), 1–13 | © 2009 European Molecular Biology Organization | www.embojournal.org

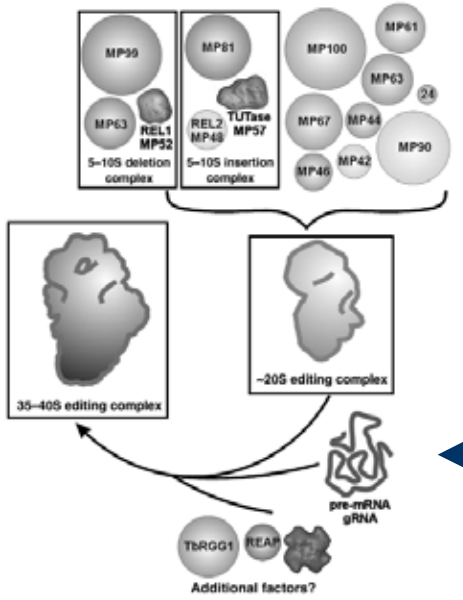


Figure 10 Model for the assembly pathway of the RNA editing machine. The ~20S complexes represent common core modules providing the majority of editing proteins. Through the interaction with gRNA, its target pre-mRNA and possibly additional proteins, the assembly is converted to the ~35–40S complex, which is suggested to perform the RNA editing reaction.

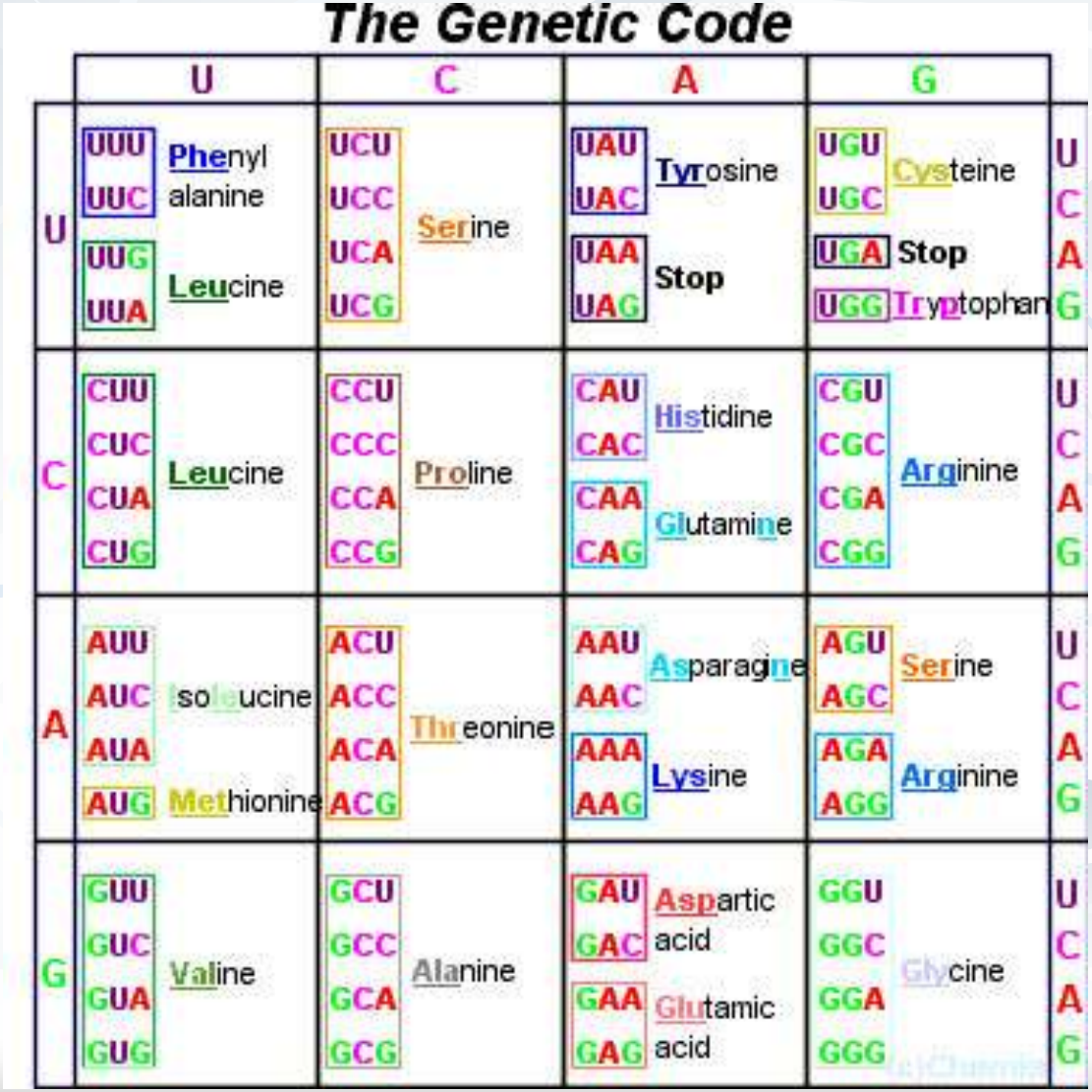
uuAuuuGGuuuAuGuuuAuuuuGuGuuGuGAGuuuGCUUUCGuuuuuuGuuuACCuuAuAuGuuuuGuuGuuuAuuAuGuGAuuAuGGuuuuGuuuuuuAuuGGUA
uuuuuuAGAuuuAuuuAuuuAAuuuGuuGAuAAAuACAuuuuAUUUGuuUGuuAGuGGuuuAuuuGuuAAuuuuuuuGuuuGuGUUUUUGGuuuAGGuuuuuuuG
uuGUUGuuGuuuuGuAuuAuGAuuGAGuuuGuuGuuuGGuuuuuuGuuuuuGuGAAACCAGuuAUGAGAGUUUGCAuuGuuAuuuAuuACAuuAAGuuGuGG
UGuuuuuGGuuCuAuuuuAuuuuuAuuGGAuuuAuUACAuuuuAUGCAuGuuuuuuuAGGuGuuuuGuuGuuGuuuAuuuGuuuuAuGCGuuuGuuuAAuuuAuGuuu
uuuuAGGuGuuuuGuuGuuGuuuAuuuGuuuuAuGCGuuuGuuuAAuuuuuuuuGuGuAuGGAuACACGuuuuGuuuuuuuGuAuuGuGuuuGuuuAuAuuGACAuuuu
GuuGAUUUAGuuuGAuuuuuuuuuAuuGCGAuuuGuuuAuuuuGAuGuuuuAuGuGuuAuGuAuuuGuGuGuGuAAuuuuAuuGGuGuuuuUUUAGUUGuuGAuuA
GuuAAuuuGuAuuGGUAGUUUGUAGGAAGuuuuuuuuuAuAAAAAAAAAA

Trypanosome brucei cytochrome c oxidase III mRNA "written" by RNA editing:

58% of the mRNA written by RNA editing

347 Uridines added to 121 sites; 16 Uridines deleted from 7 sites

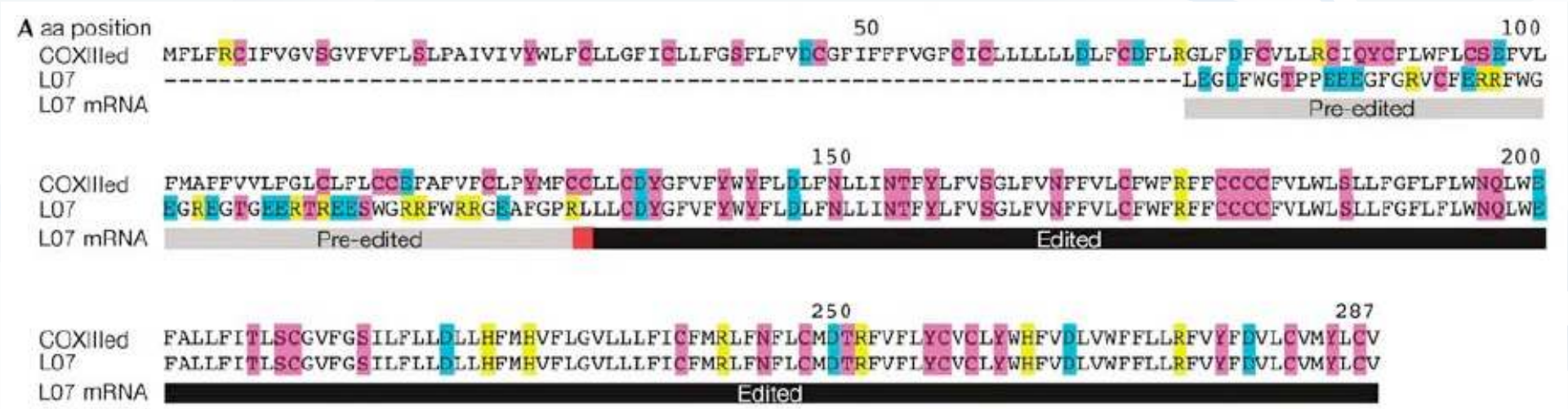
The uridines have to be added or subtracted precisely, for 37/64 codons (including each of the stop codons) contain this nucleotide:



http://library.thinkquest.org/C0123260/basic%20knowledge/images/basic%20knowledge/RNA/genetic%20code.jpg

Different mRNAs and therefore proteins can be generated from the same pseudogene transcript:

AGGAGAGGGGAGGCUUUCGGACCAAGGGAAGGAAGGGAGGUUAAGAAAAAGGAAAAACAAUUUGUGAGGGAGAAGGGUUUUUGGAGGG
GUUUUGGGAAGAGAGGGGUUUUGGGGAAACCAGAUGAGAGUUUUUGCAGAAACAAAGGGGUUUUUGGGCAAAGGGAAUACAAUUUGCU
GAGGGGGGAGAGCGGAAAGAGGGGGGAGAGCGGAAGGAGGAACACGGGAGGGAAGACAGGAUUUAGGAAGCGAGAGAGAGGAGAGGGG
AAAGGGUUUAGUUGGAAUGAAGAGGUAGUUUGUAGGAAGAAAAAAAAAAA



Alternative editing of cytochrome *c* oxidase III mRNA in trypanosome mitochondria generates protein diversity

Torsten Ochsenreiter & Stephen L. Hajduk+

EMBO *reports* VOL 7 | NO 11 | 2006

MOLECULAR AND CELLULAR BIOLOGY, Sept. 2008, p. 5595–5604 Vol. 28, No. 18
0270-7306/08/$08.00+0 doi:10.1128/MCB.00637-08

Alternative RNA Editing Produces a Novel Protein Involved in Mitochondrial DNA Maintenance in Trypanosomes▽†

Torsten Ochsenreiter,‡ Sedrick Anderson,‡ Zachary A. Wood, and Stephen L. Hajduk*

Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602

Received 19 April 2008/Returned for modification 11 June 2008/Accepted 27 June 2008

The mitochondrial genome of trypanosomes is composed of thousands of topologically interlocked circular DNA molecules that form the kinetoplast DNA (kDNA). Most genes encoded by the kDNA require a posttranscriptional modification process called RNA editing to form functional mRNAs. Here, we show that alternative editing of the mitochondrial cytochrome *c* oxidase III (COXIII) mRNA in *Trypanosoma brucei* produces a novel DNA binding protein, alternatively edited protein 1 (AEP-1). AEP-1 localizes to the region of the cell between the kDNA and the flagellum and purifies with the tripartite attachment complex, a structure believed to physically link the kDNA and flagellar basal bodies. Expression of the DNA binding domain of AEP-1 results in aberrant kDNA structure and reduced cell growth, indicating that AEP-1 is involved in the maintenance of the kDNA. Perhaps most important, our studies show a gain of function through an alternatively edited mRNA and, for the first time, provide a link between the unusual structure of the kDNA and RNA editing in trypanosome mitochondria.

A set of RNA/protein products is thus generated from "pseudogene" transcripts

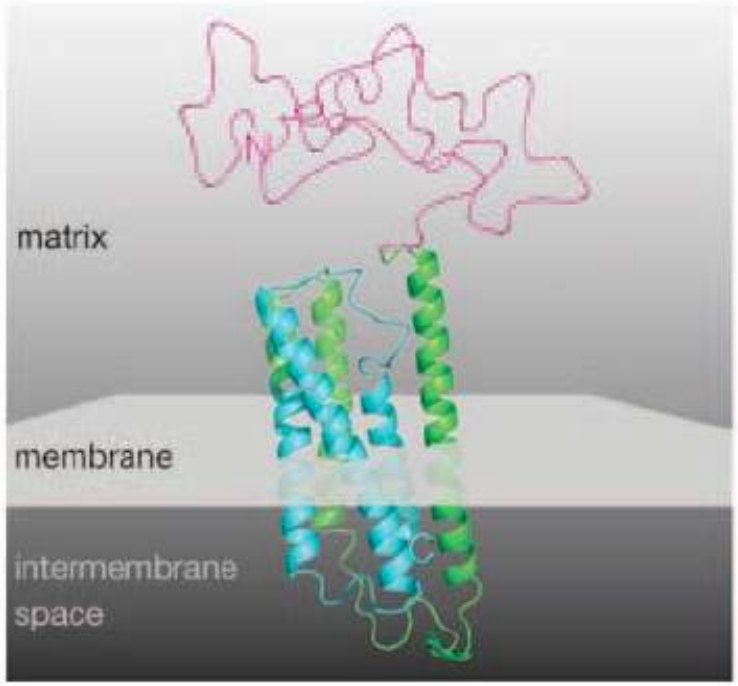


FIG. 1. Orientation of AEP-1 in the mitochondrial membrane. (A) Pri

What can be observed, then, is a dramatic increase in "genetic information" in "real time" that does not entail the destruction of components.

No Codons ⟹ Start, Stop, and Amino Acid Codons

No ORF ⟹ ORF

"Junk" RNA ⟹ Cytochrome c oxidase III mRNA

Ribosomal and transfer RNAs must be highly edited in order to become functional in all known taxa

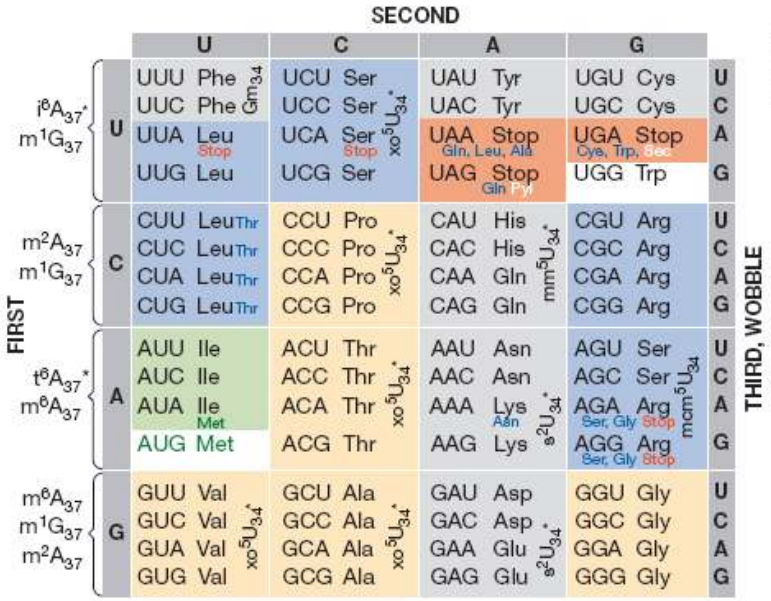
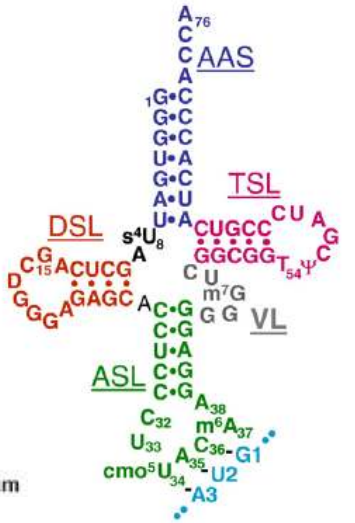


Fig 1 | Universal genetic code. The 64 codes are associated with the transfer RNA (tRNA) modifications that are important for decoding and/or translocation. Twofold degenerate amino-acid codes are highlighted in grey and fourfold degenerate codes are highlighted in tan. Amino acids with six codons are highlighted in blue. The threefold degenerate codons of Ile are highlighted in green, whereas the single codons of Met and Trp are highlighted in white. The three stop codons are highlighted in orange. Non-canonical codon use by some organisms and the mitochondrion is shown by using a small font for the amino acids (blue) or translational stop codons (red). The modified nucleoside abbreviations are defined in the text. Selenocysteine (Sec) and pyrrolysine (Pyl) codons are denoted in white. In the mitochondrion, $tRNA^{Met}$ responds to AUG and AUA, which is not used as an Ile codon (Agris *et al*, 2007; Szymański & Barciszewski, 2007; Björk *et al*, 1987).

Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications

Paul F. Agris

EMBO *reports* VOL 9 | NO 7 | 2008



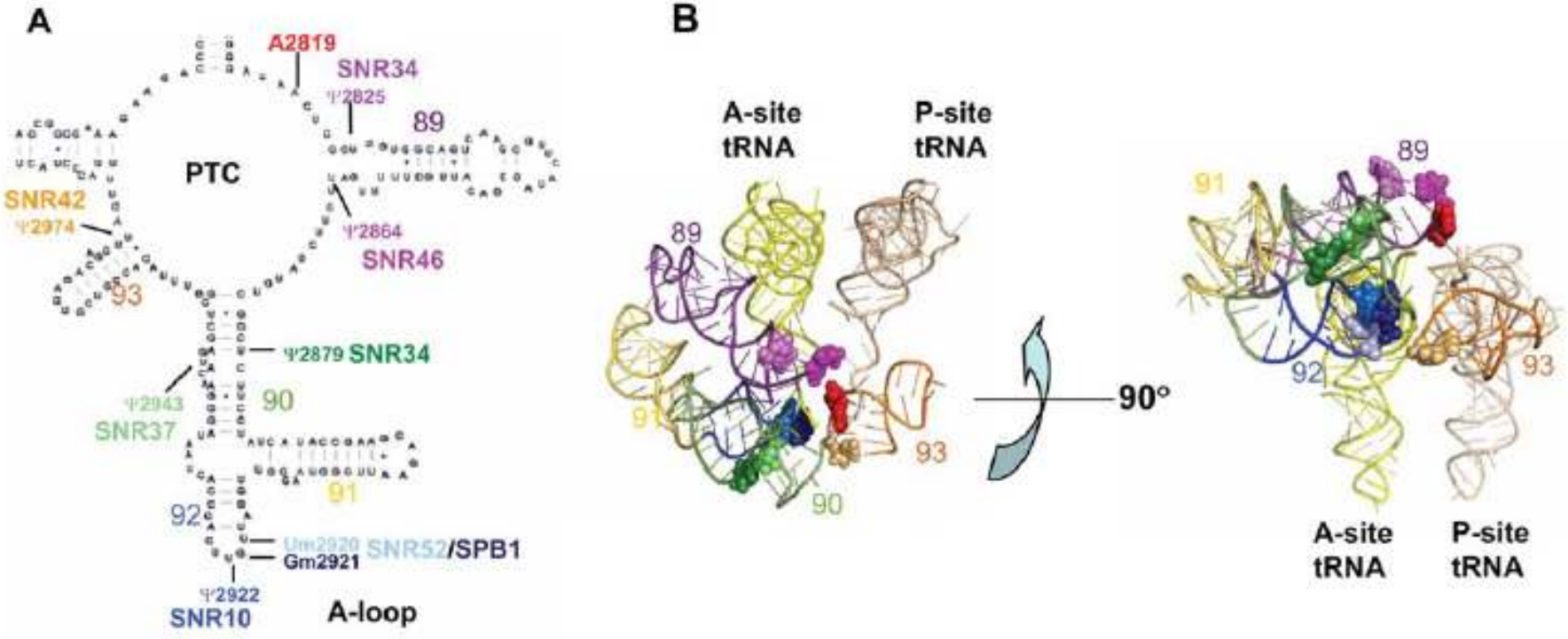
tRNA's Wobble Decoding of the Genome: 40 Years of Modification

Paul F. Agris*, Franck A. P. Vendeix and William D. Graham

J. Mol. Biol. (2007) 366, 1–13

Figure 2. tRNA primary sequence, secondary structure, and codon binding. The sequence and secondary structure of *E. coli* $tRNA^{Val}$. The physical and functional domains of the *E. coli* $tRNA^{Val}_{UAC}$ sequence and secondary structure are the amino acid-accepting stem, AAS (dark blue), the dihydrouridine stem and loop, DSL (red), the anticodon stem and loop, ASL (green), the variable loop, VL (gray), and the thymidine stem and loop, TSL (purple). The modified nucleosides in this tRNA are: s^4U, 4-thiouridine; D, dihydrouridine; cmo^5U, uridine-5-oxyacetic acid; m^6A, N6-methyladenosine; m^7G, 7-methylguanosine; ribothymidine, T; and pseudouridine, Ψ. Because of the wobble nucleoside modification, cmo^5U_{34}, *E. coli* $tRNA^{Val}_{UAC}$ is capable of decoding all of the fourfold degenerate valine codons.[32–34] The tRNA is shown binding the cognate codon for valine, GUA, in light blue.

Ribosomal and transfer RNAs also must be highly edited in order to become operational in all known taxa

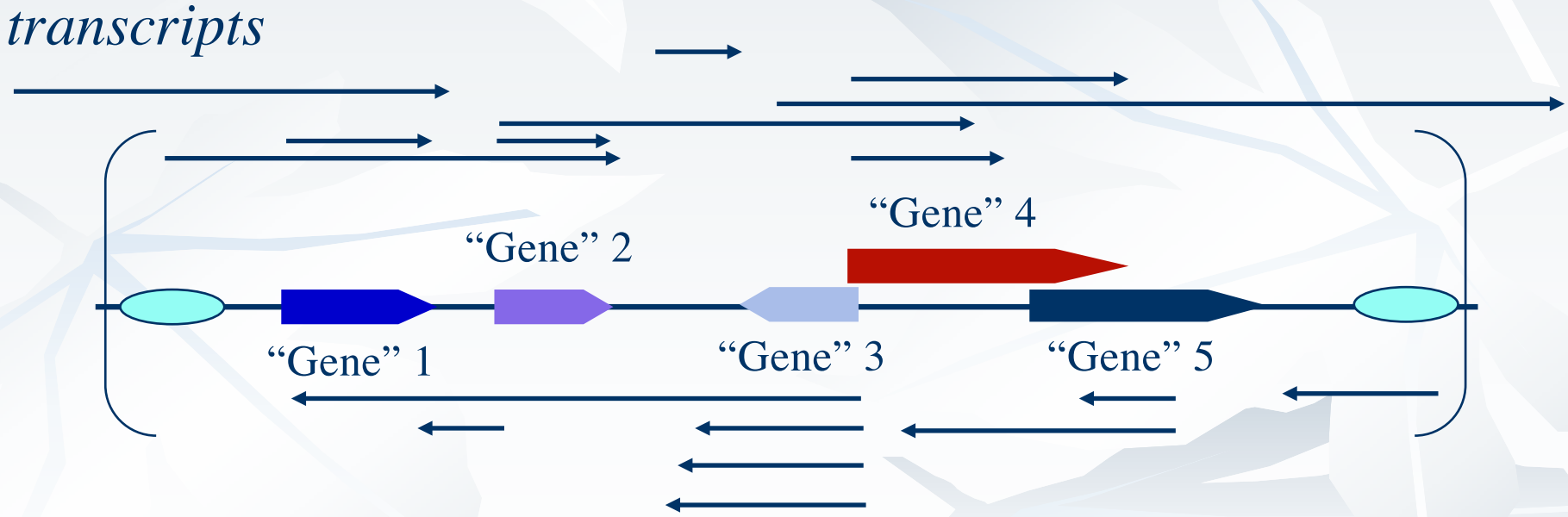


Optimization of Ribosome Structure and Function by rRNA Base Modification

Jennifer L. Baxter-Roshek, Alexey N. Petrov, Jonathan D. Dinman*

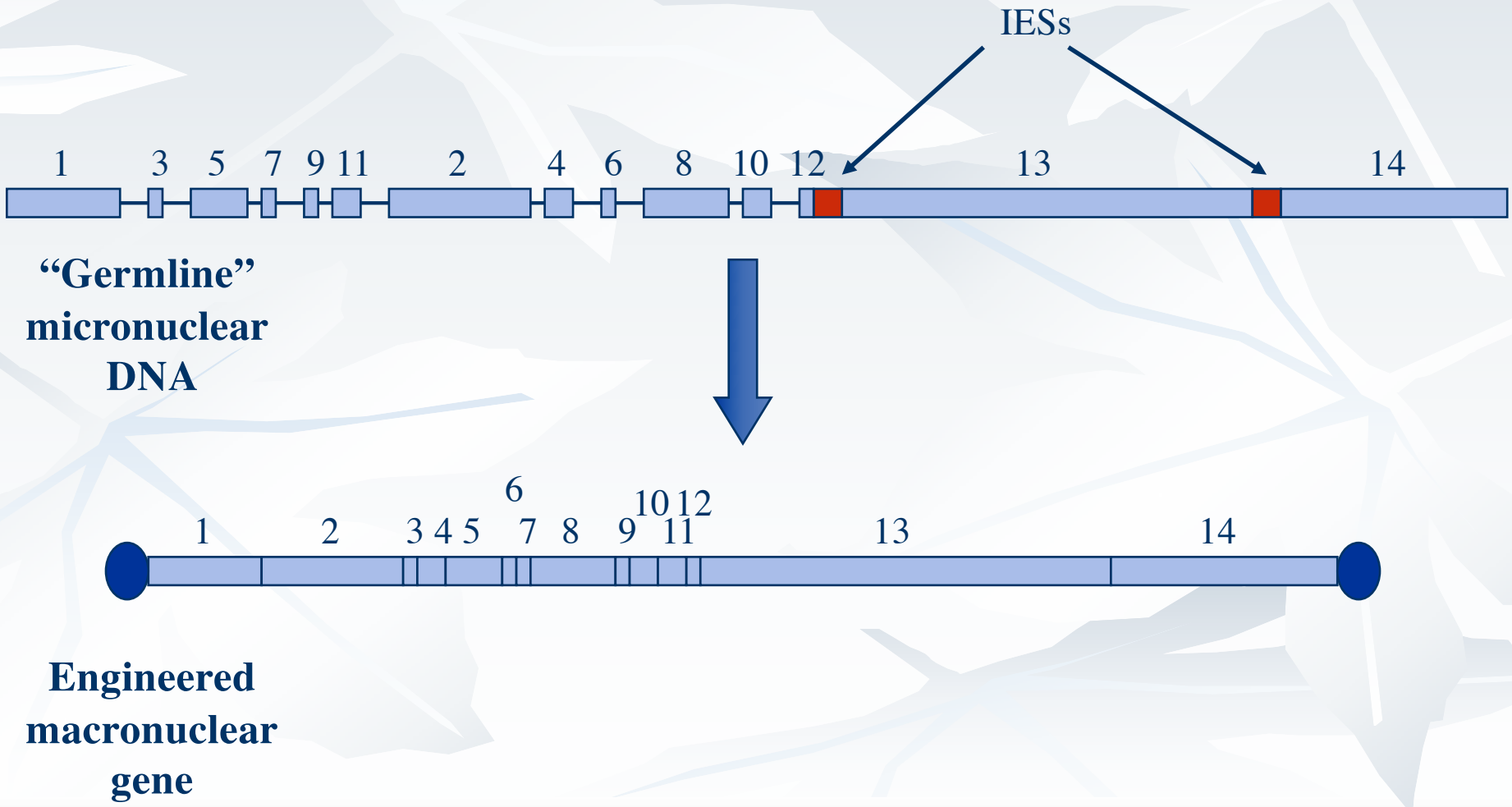
In a sense, each rRNA and tRNA locus is a pseudogene! After all, none encode a fully operational product—the transcripts must be rewritten

Even more startling was discovery in 2004-2007 that making RNA copies often begins in one "gene" and ends in another, and occurs on both DNA strands and in "non-gene" regions



Beyond this, it is now clear that DNA can be extensively rewritten at the genotypic level

Unscrambling the α-subunit gene of the telomere-binding protein in *Oxytricha nova*



Based on such evidence, some of us are asking: What exactly is a 'gene'?

What is a gene, post-ENCODE? History and updated definition

Mark B. Gerstein,[1,2,3,9] Can Bruce,[2,4] Joel S. Rozowsky,[2] Deyou Zheng,[2] Jiang Du,[3] Jan O. Korbel,[2,5] Olof Emanuelsson,[6] Zhengdong D. Zhang,[2] Sherman Weissman,[7] and Michael Snyder[2,8]

17:669–681 ©2007
Genome Research

Theory Biosci.
DOI 10.1007/s12064-008-0025-0

ORIGINAL PAPER

"Genes"

Sonja J. Prohaska · Peter F. Stadler

Theory Biosci.
DOI 10.1007/s12064-009-0067-y

SHORT COMMUNICATION

Defining genes: a computational framework

Peter F. Stadler · Sonja J. Prohaska · Christian V. Forst · David C. Krakauer

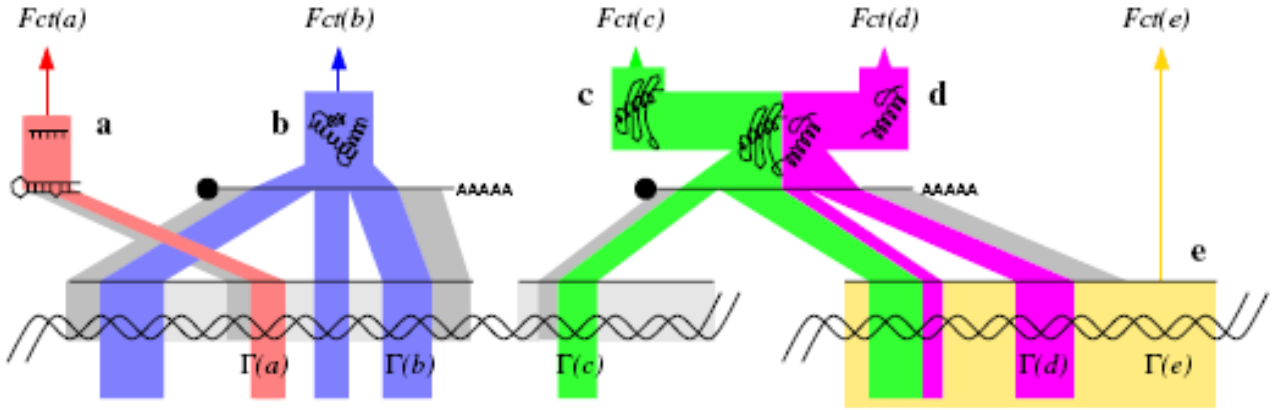


Fig. 1 Functional objects *a* to *e* and relationships with their genomic footprints $\Gamma(a)$ to $\Gamma(e)$. A functional RNA molecule (e.g., a miRNA) with function $\mathrm{Fct}(a)$ is processed in two steps from an intronic sequence. Its image on the DNA is the genomic footprint $\Gamma(a)$. The genomic footprint $\Gamma(b)$ of the functional protein *b* is a discontinuous stretch of DNA corresponding to the coding sequence (CDS) including the start codon but excluding the stop codon. The mRNA includes UTRs that also map back to the DNA as well as parts without footprints on the DNA (the 5′-cap and the poly-A tail). The functional proteins *c* and *d* are obtained by cleavage of the (non-functional) precursor *cd*. The later is encoded by a trans-spliced mRNA. The footprint $\Gamma(c)$ is distributed over two DNA molecules. The primary transcript *e* has an additional function $\mathrm{Fct}(e)$ that is independent of its role as precursor of the mRNA of *cd*. As a consequence, $\Gamma(e)$ overlaps with both, $\Gamma(c)$ and $\Gamma(d)$. In all cases, the gene is the pair $(\Gamma(x),x)$ composed of the genomic footprint $\Gamma(x)$ and the resulting functional molecule *x*

A new ‘gene’ definition and what it implies…

genes previously assumed to be distinct genetic loci. This supports and is consistent with earlier observations of a highly interleaved transcribed genome[1,2], but more importantly, prompts the reconsideration of the definition of a gene. As this is a consistent characteristic of annotated genomes, we would propose that the transcript be considered as the basic atomic unit of inheritance. Concomitantly, the term gene would then denote a higher-order concept intended to capture all those transcripts (eventually divorced from their genomic locations) that contribute to a given phenotypic trait. Co-published

Landscape of transcription in human cells

Sarah Djebali[1*], Carrie A. Davis[2*], Angelika Merkel[1], Alex Dobin[2], Timo Lassmann[3], Ali Mortazavi[4,5], Andrea Tanzer[1], Julien Lagarde[1], Wei Lin[2], Felix Schlesinger[2], Chenghai Xue[2], Georgi K. Marinov[4], Jainab Khatun[6], Brian A. Williams[4], Chris Zaleski[2], Joel Rozowsky[7,8], Maik Röder[1], Felix Kokocinski[9], Rehab F. Abdelhamid[3], Tyler Alioto[1,10], Igor Antoshechkin[4], Michael T. Baer[2], Nadav S. Bar[11], Philippe Batut[2], Kimberly Bell[2], Ian Bell[12], Sudipto Chakraborty[2], Xian Chen[13], Jacqueline Chrast[14], Joao Curado[3], Thomas Derrien[1], Jorg Drenkow[2], Erica Dumais[12], Jacqueline Dumais[12], Radha Duttagupta[12], Emilie Falconnet[15], Meagan Fastuca[2], Kata Fejes-Toth[2], Pedro Ferreira[1], Sylvain Foissac[12], Melissa J. Fullwood[16], Hui Gao[17], David Gonzalez[1], Assaf Gordon[2], Harsha Gunawardena[13], Cedric Howald[14], Sonali Jha[2], Rory Johnson[1], Philipp Kapranov[12,17], Brandon King[4], Colin Kingswood[1,10], Oscar J. Luo[16], Eddie Park[5], Kimberly Persaud[2], Jonathan B. Preall[2], Paolo Ribeca[1,10], Brian Risk[6], Daniel Robyr[15], Michael Sammeth[1,10], Lorian Schaffer[4], Lei-Hoon See[2], Atif Shahab[16], Jorgen Skancke[1,11], Ana Maria Suzuki[3], Hazuki Takahashi[3], Hagen Tilgner[1†], Diane Trout[4], Nathalie Walters[14], Huaien Wang[2], John Wrobel[6], Yanbao Yu[13], Xiaoan Ruan[16], Yoshihide Hayashizaki[3], Jennifer Harrow[9], Mark Gerstein[7,8,18], Tim Hubbard[9], Alexandre Reymond[14], Stylianos E. Antonarakis[15], Gregory Hannon[2], Morgan C. Giddings[6,13], Yijun Ruan[16], Barbara Wold[4], Piero Carninci[3], Roderic Guigó[1,19] & Thomas R. Gingeras[2,12]

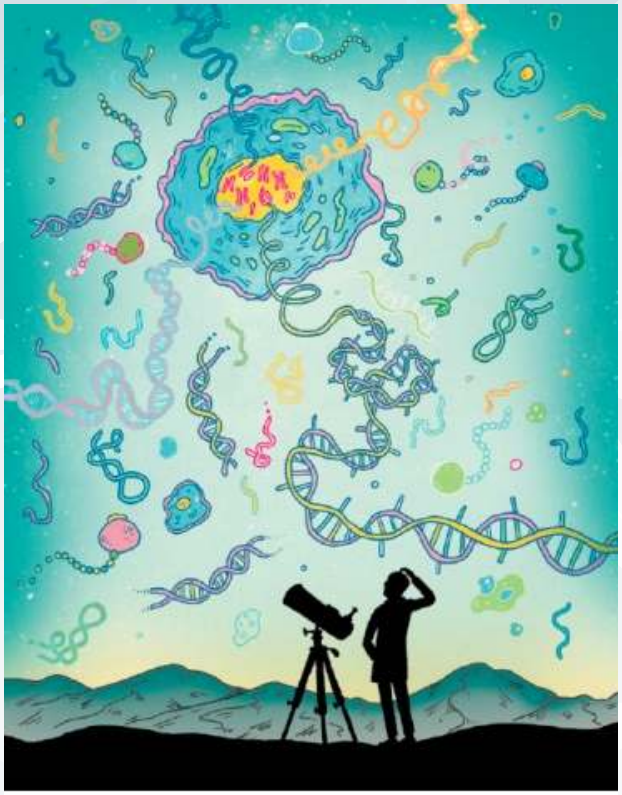
Although the gene has conventionally been viewed as the fundamental unit of genomic organization, on the basis of ENCODE data it is now compellingly argued that this unit is not the gene but rather the transcript (Washietl et al. 2007; Djebali et al. 2012a). On this view, genes represent a higher-order framework around which individual transcripts coalesce, creating a polyfunctional entity that assumes different forms under different cellular states, guided by differential utilization of regulatory DNA.

What does our genome encode?

John A. Stamatoyannopoulos

Genome Res. 2012 22: 1602-1611

Thesis. **In order for a ‘gene’ to be a ‘gene’, to be a “higher-order poly-functional entity” that takes on different forms at different times, it has to be more than an invariant particle.**

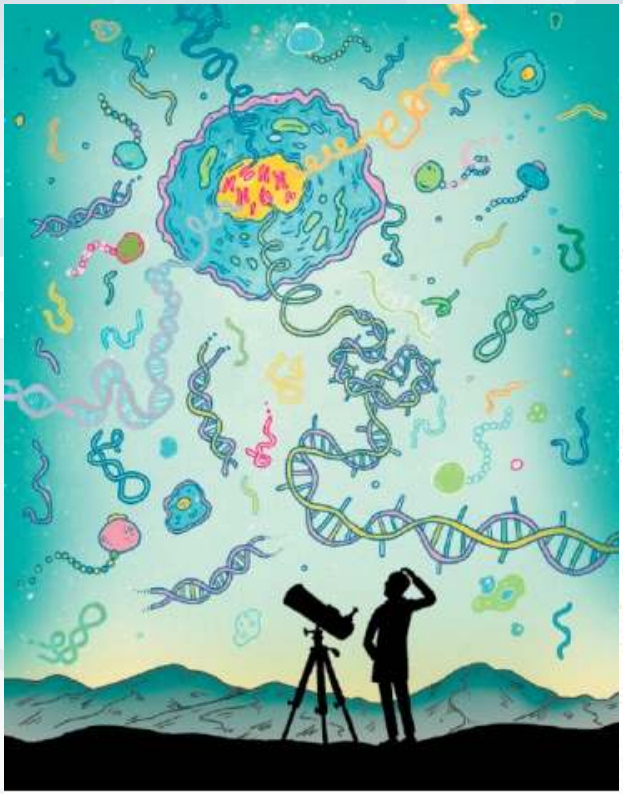


Celebrate the unknowns

On the 60th anniversary of the double helix, we should admit that we don't fully understand how evolution works at the molecular level, suggests **Philip Ball**.

Sixty years on, the very definition of 'gene' is hotly debated. We do not know what most of our DNA does, nor how, or to what extent it governs traits. In other words, we do not fully understand how evolution works at the molecular level.

That sounds to me like an extraordinarily exciting state of affairs, comparable perhaps to the disruptive discovery in cosmology in 1998 that the expansion of the Universe is accelerating rather than decelerating, as astronomers had believed since the late 1920s. Yet, while specialists debate what the latest findings mean, the rhetoric of popular discussions of DNA, genomics and evolution remains largely unchanged, and the public continues to be fed assurances that DNA is as solipsistic a blueprint as ever.



Celebrate the unknowns

On the 60th anniversary of the double helix, we should admit that we don't fully understand how evolution works at the molecular level, suggests **Philip Ball**.

In short, the current picture of how and where evolution operates, and how this shapes genomes, is something of a mess. That should not be a criticism, but rather a vote of confidence in the healthy, dynamic state of molecular and evolutionary biology.

A PROBLEM SHARED

Barely a whisper of this vibrant debate reaches the public