The Gene Concept

Sonja Prohaska

Computational EvoDevo
Universitaet Leipzig

June 19, 2014

What is a gene?

"I can't tell but I recognize a gene when I see one."
a biologist

"Something is a gene when a biologist says it is one."
a bioinformatician

"A gene is a database entry with an Ensembl gene ID."
a computer scientist

"A gene is what Wikipedia says it is."
a student

"A gene is a locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions and/or other functional sequence regions." Wikipedia

Historical view – really short

In the beginning...

- a *phenotype* has characteristics
- some characteristics are independent
- some characteristics are heritable
- all heritable characteristics need to go through a single cell (gamete)

How to put (all) characteristics of a phenotype into a gamete?

- miniature organism within gamete?
- gemmule, shed by the organs accumulated in gametes? (Darwin 1868)
- distinct, discrete entities that specify characteristics (Mendel 1866)

"special conditions, foundations and determiners which are present [in the gametes] in unique, separate and thereby independent ways [by which] many characteristics of the organism are specified" by Johannsen (1909)

... the **gene** is a (unknown) substance **representing a characteristic**.

Historical view – really short

linkage of genes

- Morgan (1915)
- segregation experiments and crossbreeding
- the observed linkage of genes best fitted a model of a linear arrangement
- size of genes and distance between genes could be inferred
- the model had predictive power in breeding

How did this change the understanding of a gene?

- genes are continuous
- genes are nonoverlapping
- distinct genes have distinct dimensions
- genes are linked to verying degrees

A gene is an abstract entity whose existance is reflected in the way a phenotype is transmitted between generations.

Historical view – really short

- **1941** Beadle and Tatum: *"one gene, one enzyme"*
  The gene is the information behind the individual molecule.
- **1955** Hershey and Chase: the substance for genes is DNA
- **1955** Benzer: a cistron (gene) is a region of DNA defined by mutations that in *trans* could not genetically complement each other.
- **1953** Watson and Crick: how DNA could function as a molecule of heredity
- **1958** Crick: flow of information from DNA → RNA → protein
- **1970 – 1980** Fiers: RNA and DNA sequencing
- understanding of how genes are expressed, discovery of splicing
- development of computational tools
- **the "nominal gene"** is defined by its **predicted sequence** rather than a genetic locus
- **1986** the gene effectively became identified as an annotated open freading frame (ORF)

pre-ENCODE: the birth of the structural gene

a gene is...

"... a DNA segment that contributes to phenotype/function. In the absence of demonstrated function a gene may be characterized by sequence, transcription or homology." Human Genome Nomenclature Organization

"... the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding seqments (exons)." textbook *Genes V* by Lewin (1994)

" ... the entire nucleic acid sequence that is necessary for the synthesis of a functional polypeptide (or RNA)" by Lodish (2000)

Servey – “Representing Genes” by Griffiths and Stotz

Problematic issues with the gene concept

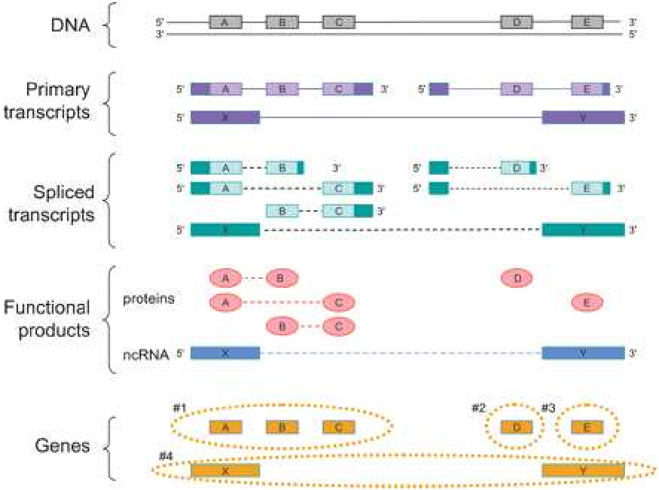
- **regulatory sequence**: part of a gene or associated with a gene?
- **overlapping genes**: same strand different reading frame or readingframes on opposite strands
- **splicing**: open reading frame is segmented
- **alternative splicing**: multiple different transcripts with different function
- **trans-splicing**: distinct transcripts can be joint
  the gene as a single locus no longer applies
- **run-through transcripts and fusion proteins**
- **parasitic and mobile elements**
- **pseudogenes**: retrotranscposed "dead" genes

A gene is a set of connected transcripts where "connected" means sharing of exons.

How ENCODE ruined/challenged the gene concept

- functional non-coding RNAs
- unannotated transcription: only 50% of spliced transcripts are annotated
- transcription from (distal) alternative transcription start sites (TSS)
- alternative 3'UTRs
- transcription at regulatory elements
- dispersed regulation and elements (upstream, downstream, within the first exon, within the first intron, anywhere else)
- blurring of the destinction between genic and intergenic, exonic and intronic
- act of transcription of functional importance, transcript irrelevant
- pseudogenes
- highly conserved elements, only 20% in annotated regions

The Gerstein-Snyder gene definition



"The gene is a union of genomic sequences encoding a coherent set of potentially overlapping functional products." Gerstein et al. 2007

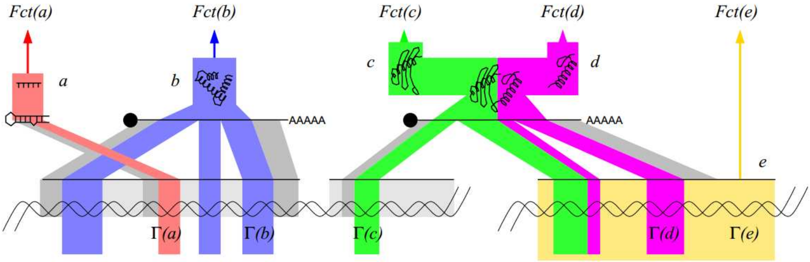
What any good gene concept needs

A **structural** and a **functional** component.

- a well-defined structural representation on the genetic material
- a well-defined notion of function (by measurement)
- **heritability** (conveyed by the genetic material)
- compartibility with the concept of **homology** on structural and/or functional level
- processes of **gene expression** connecting the structural and functional components
- relation between mutations and changes in a gene product

A gene is a heritable elementary functional unit.

The Stadler-Prohaska gene definition



Functional objects (a to e) and relationships with their genomic footprints (Γ(*a*) to Γ(*e*)).

Btw. **Genes** are irrelevant for **genome annotation**! Annotation should focus on the observable intermediates (transcripts, translation products, precursors, etc.) of the expression cascade.

Literature

Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, Emanuelsson O, Zhang ZD, Weissman S and Snyder M (2007). *What is a gene, post-ENCODE? History and updated definition.* Genome Res. 17:669-681

Griffiths PE (2002). *Lost: One Gene Concept. Reward to Finder.* Biology and Philosophy 17:271-283

Prohaska SJ and Stadler PF (2008) *"Genes"* Theory Biosci. 127: 215-221

Stadler PF, Prohaska SJ, Forst CV and Krakauer DC (2009). *Defining genes: a computational framework.* Theory in Biosciences 128:165-170

Engelhardt, Kirsten T, Stadler PF and Prohaska S (2010). *Genome Annotation without Genes.* Technical report.

Teaser

- What is “function” or a “functional gene”?
- Does function answer the question:

  “What is it (good) for”?
  “What does it do”?
  “What has it been doing in the past”?
- How are the relations between “unit of function”, “unit of selection” and “unit of heredity”?

- Can a small peptide be encoded in a non-coding RNA?
- Is a functional pseudogene a “gene zombie”?