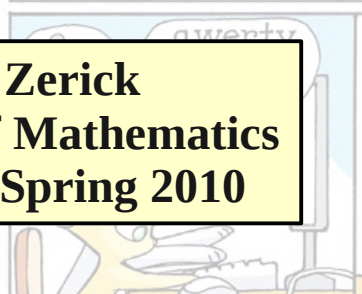
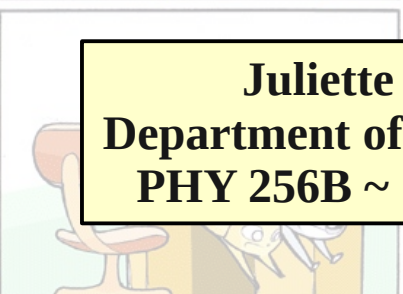
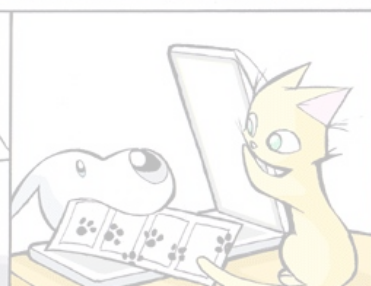
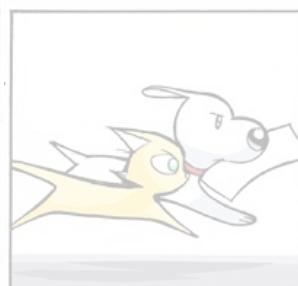
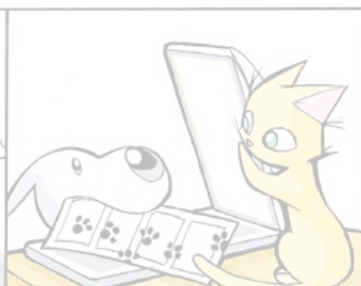
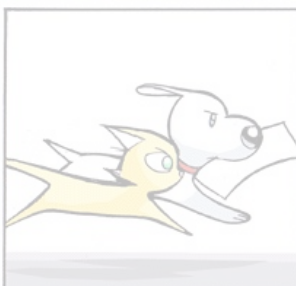
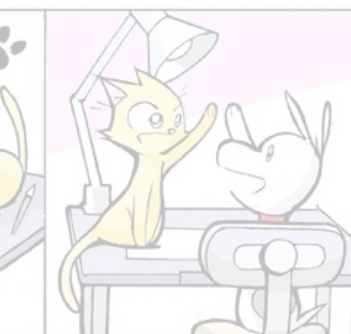
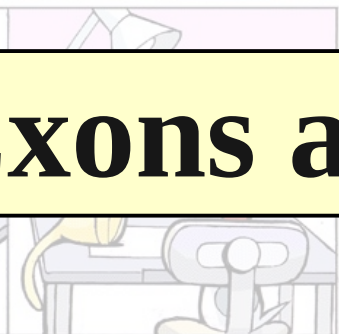
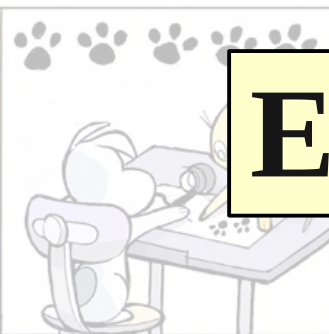
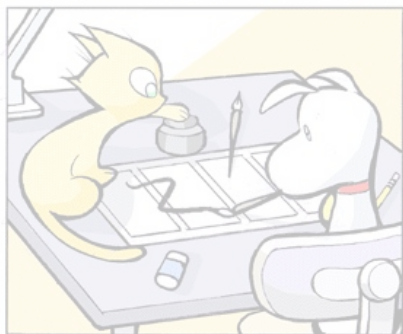


# Exons and You

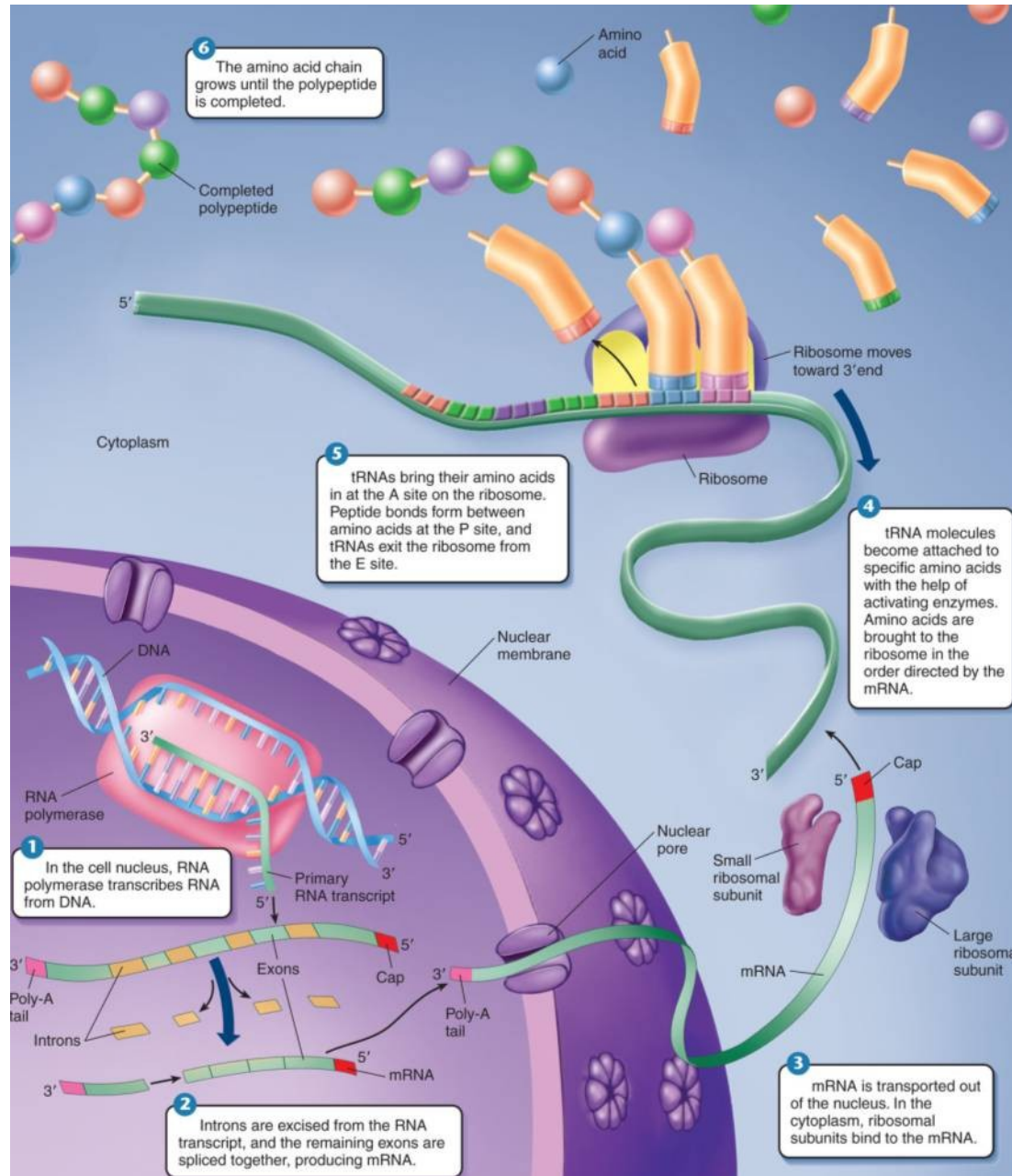


Juliette Zerick  
Department of Mathematics  
PHY 256B ~ Spring 2010

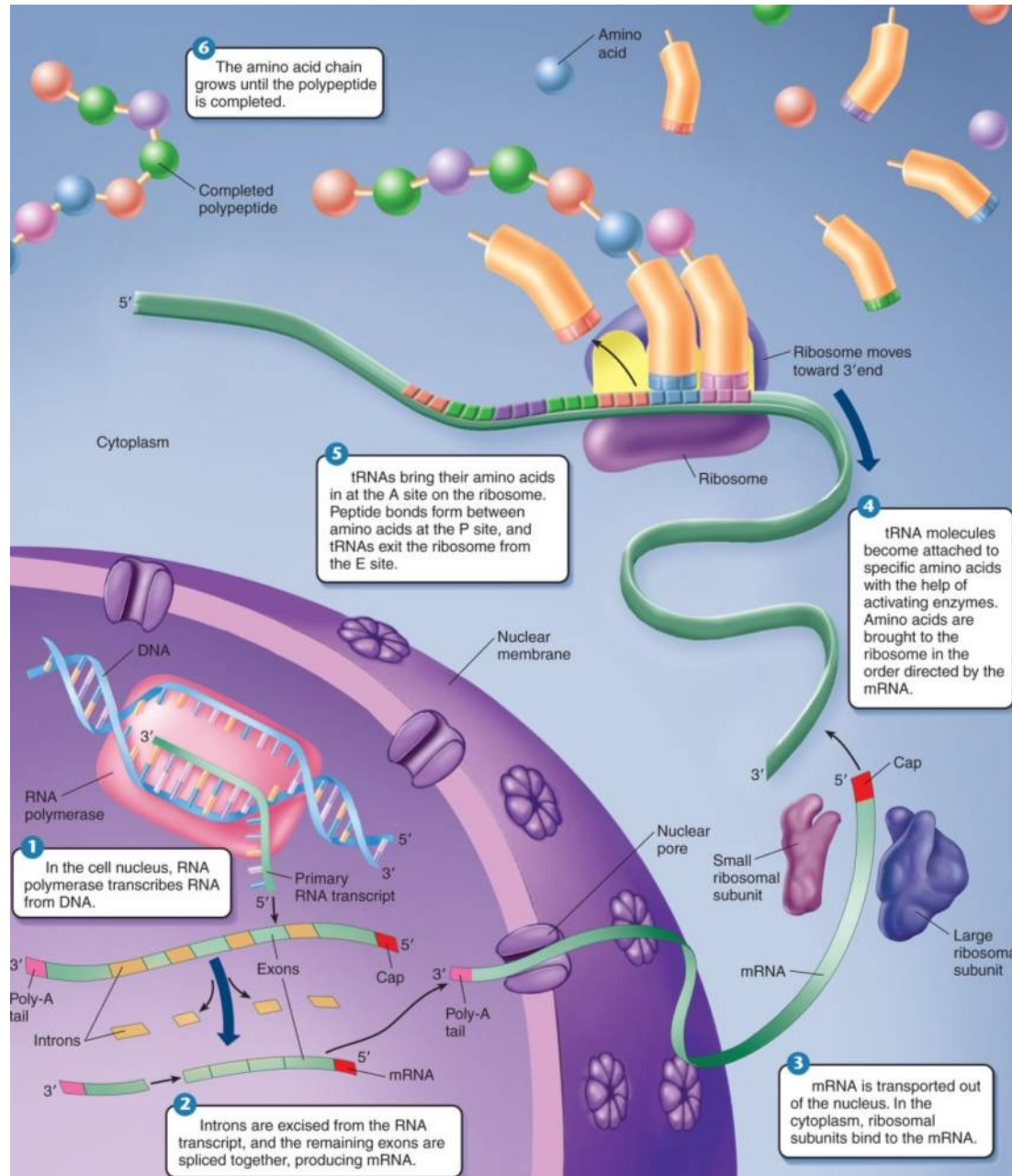
[2]



# Central Dogma of Molecular Biology



# Central Dogma of Molecular Biology (and a sampling of interested parties)

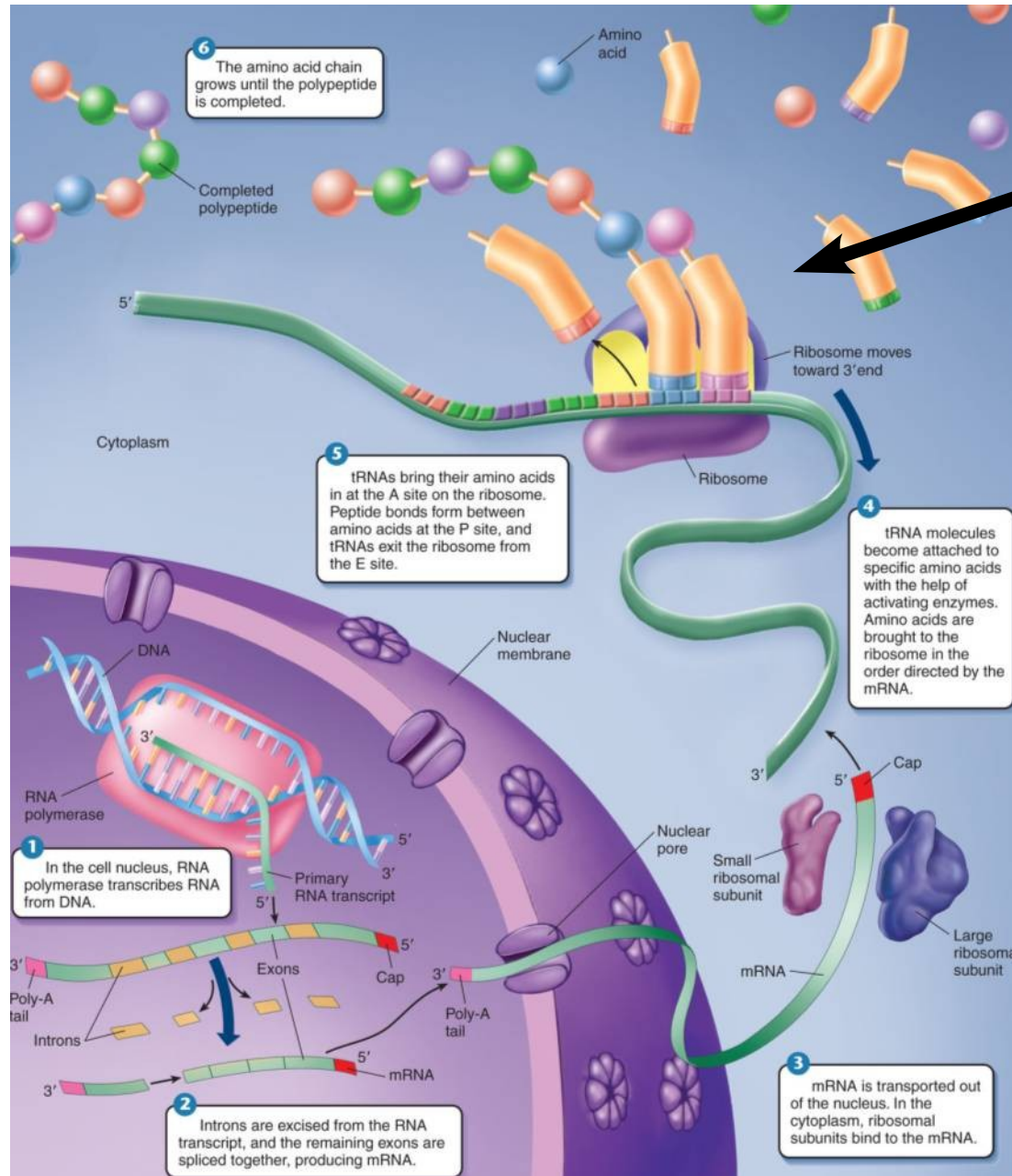


[4]

[3]



# Central Dogma of Molecular Biology (and a sampling of interested parties)



(Coding Theorists)



[5]

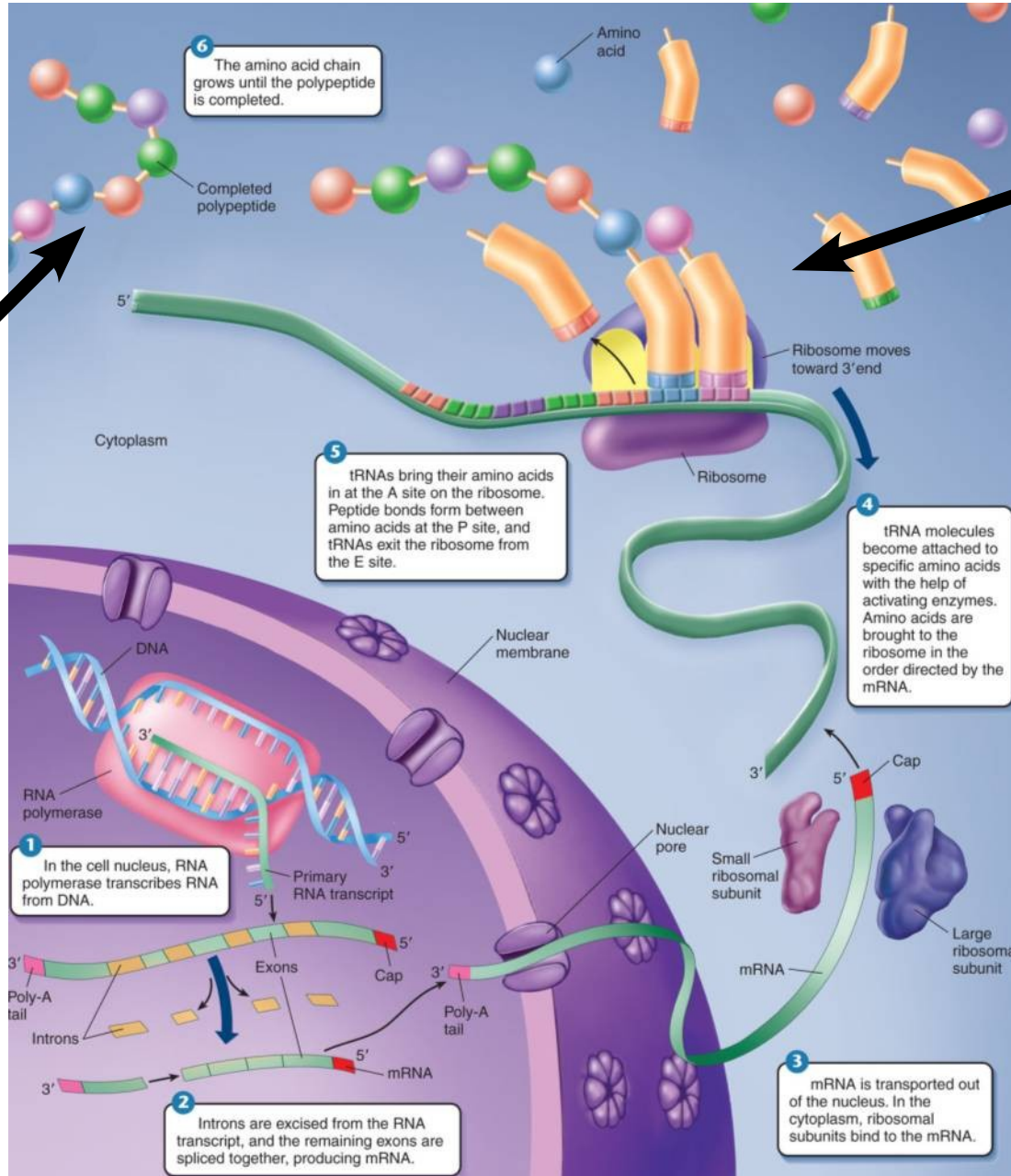


[4]

[3]

# Central Dogma of Molecular Biology (and a sampling of interested parties)

[6] (Biologists of various stripes)



(Coding Theorists)



[5]



[4]

[3]



# Central Dogma of Molecular Biology (and a sampling of interested parties)

[6]

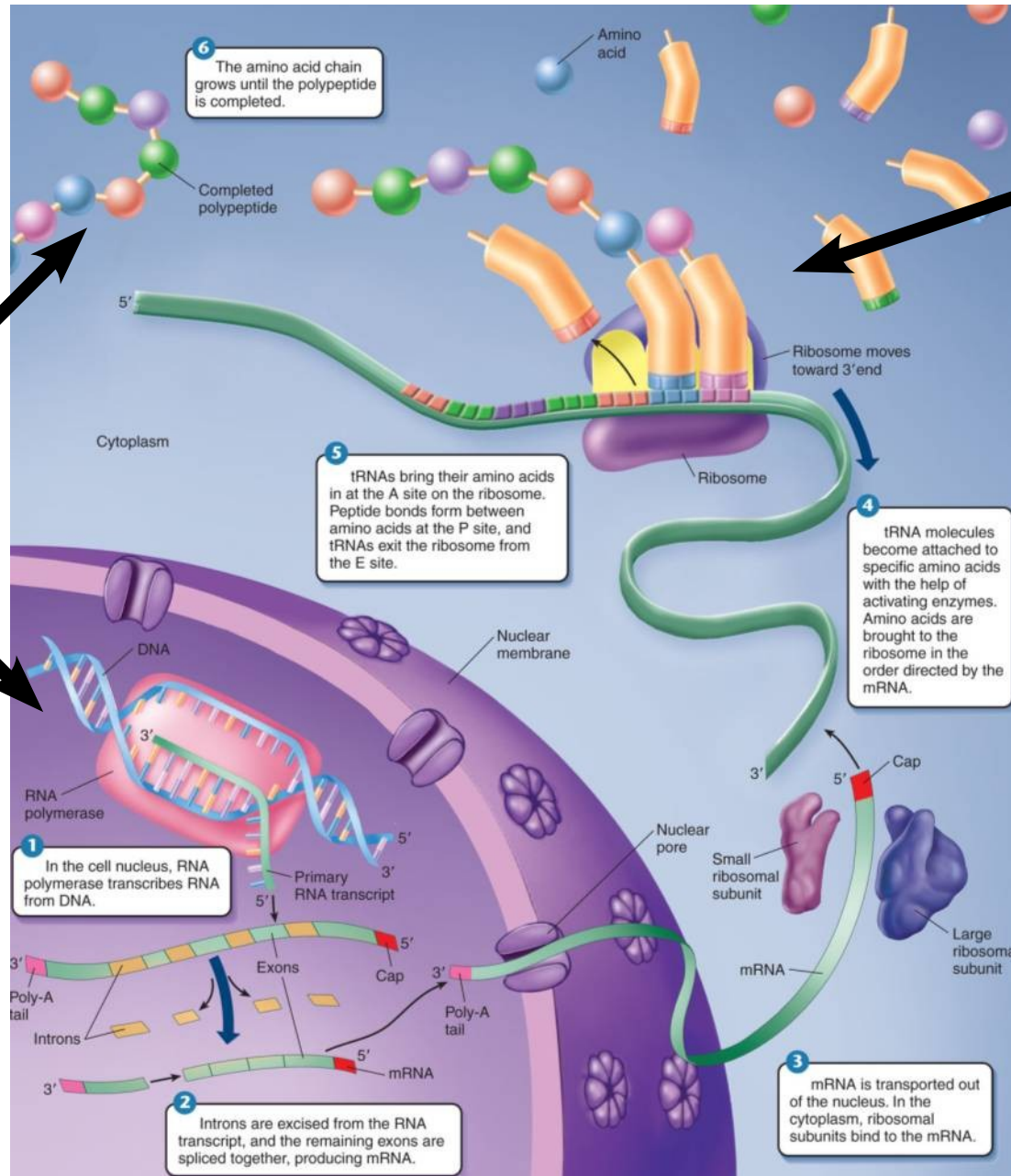
(Biologists of various stripes)



(Physical Biochemists)



[7]



(Coding Theorists)



[5]



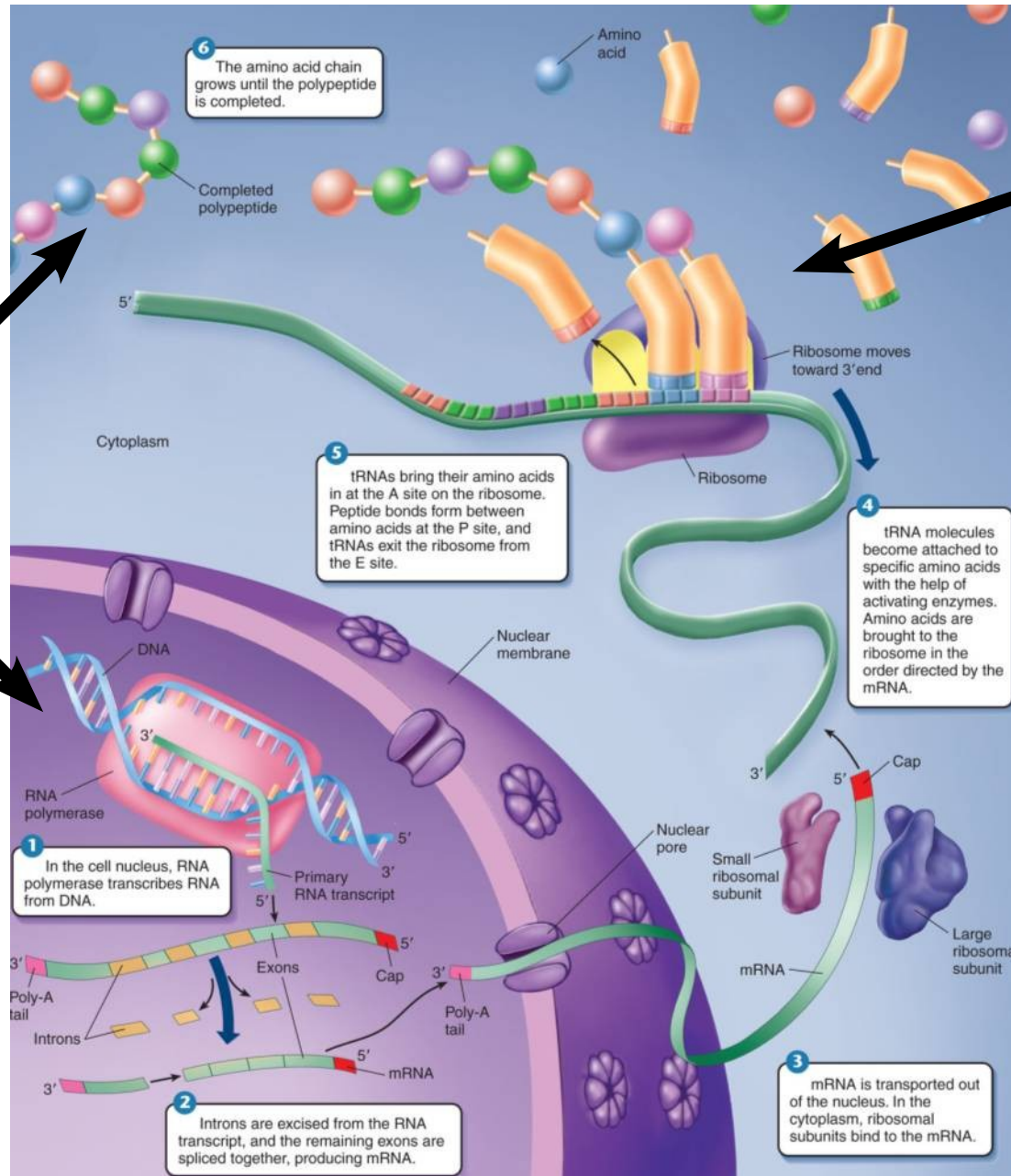
[4]

[3]

# Central Dogma of Molecular Biology (and a sampling of interested parties)

[6]

(Biologists of various stripes)



(Coding Theorists)



[5]

(Physical Biochemists)



[7]

(Exon "Predictors")



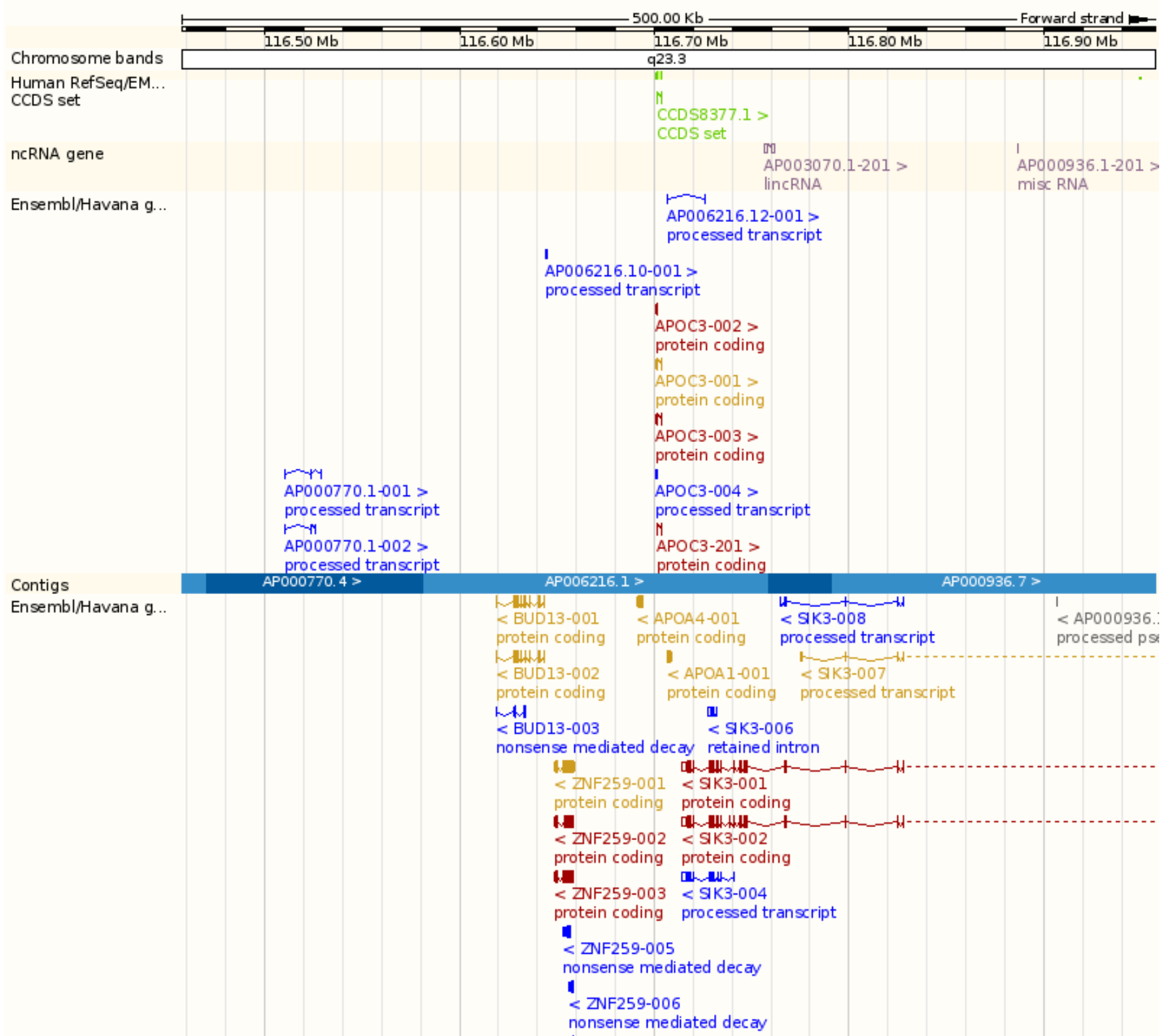
[8]



[4]

[3]







# Exon prediction via

[10]

## Information Theory: mid-1980s - present

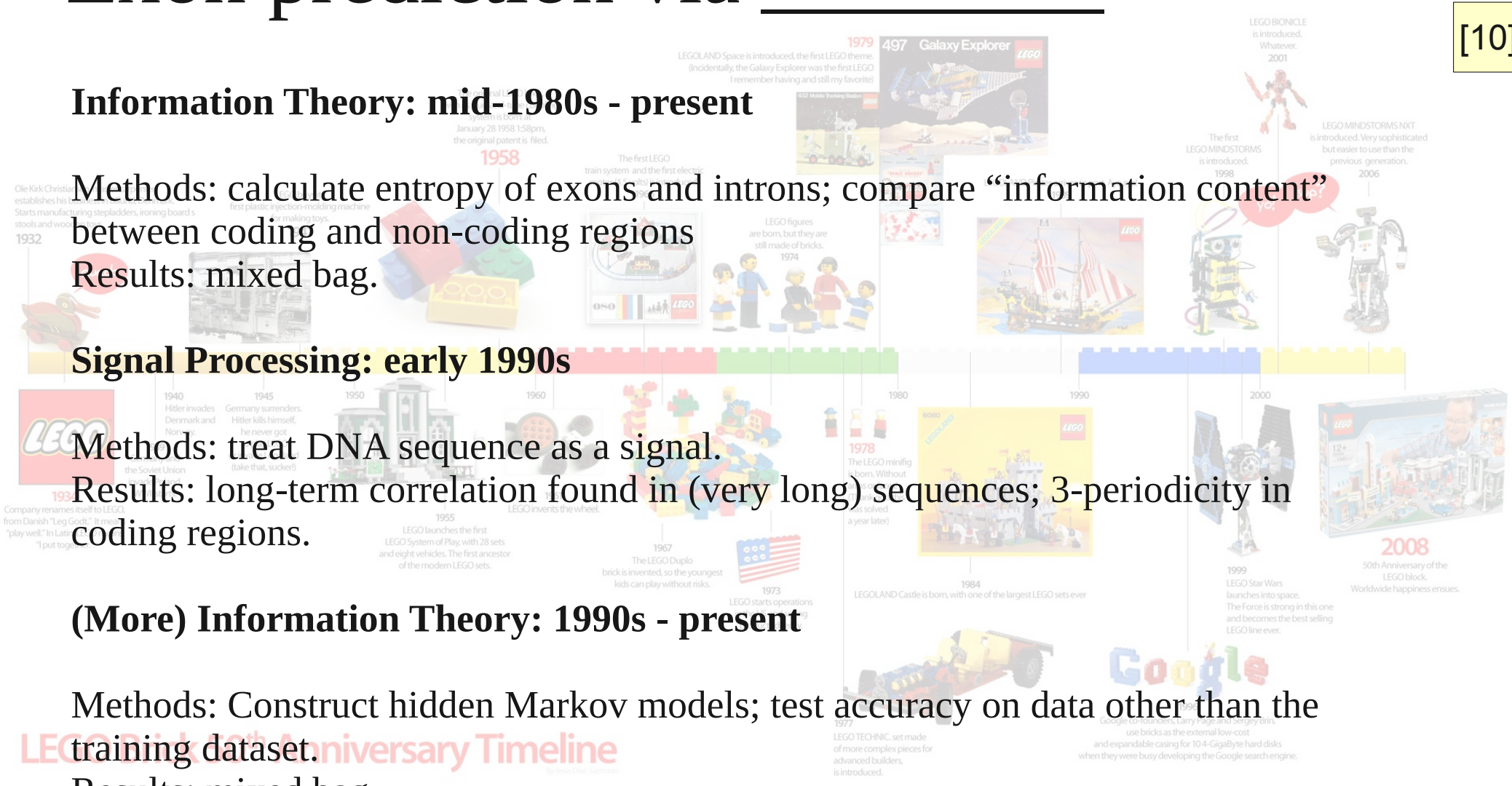
Methods: calculate entropy of exons and introns; compare “information content” between coding and non-coding regions  
Results: mixed bag.

## Signal Processing: early 1990s

Methods: treat DNA sequence as a signal.  
Results: long-term correlation found in (very long) sequences; 3-periodicity in coding regions.

## (More) Information Theory: 1990s - present

Methods: Construct hidden Markov models; test accuracy on data other than the training dataset.  
Results: mixed bag.



# Exon prediction via

[10]

## Information Theory: mid-1980s - present

Methods: calculate entropy of exons and introns; compare “information content” between coding and non-coding regions  
Results: mixed bag.

## Signal Processing: early 1990s

Methods: treat DNA sequence as a signal.  
Results: long-term correlation found in (very long) sequences; 3-periodicity in coding regions.

## (More) Information Theory: 1990s - present

Methods: Construct hidden Markov models; test accuracy on data other than the training dataset.  
Results: mixed bag.

**Current tool of choice: Hidden Markov models.**



[11]



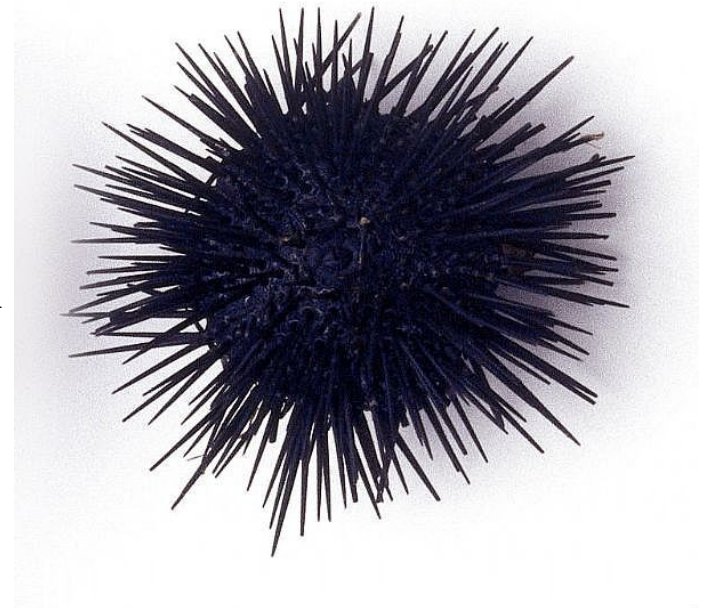
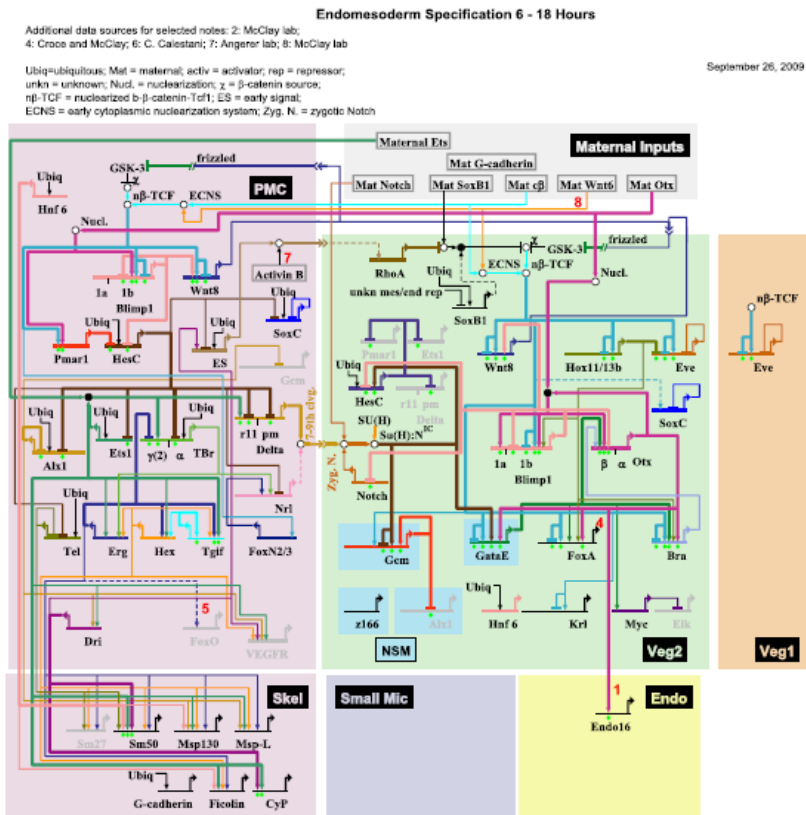
# Maybe . . .

Given real-world constraints, the process of eukaryotic DNA transcription cannot be modeled with the sequence alone. In other words, *de novo* exon prediction is impossible.

# Maybe . . .

Given real-world constraints, the process of eukaryotic DNA transcription cannot be modeled with the sequence alone. In other words, *de novo* exon prediction is impossible.

But consider this:



This model is frequently revised. It is based on the latest laboratory data, some of which is not yet published. Copyright © 2001-2009 Hamid Bokouri and Eric Davidson

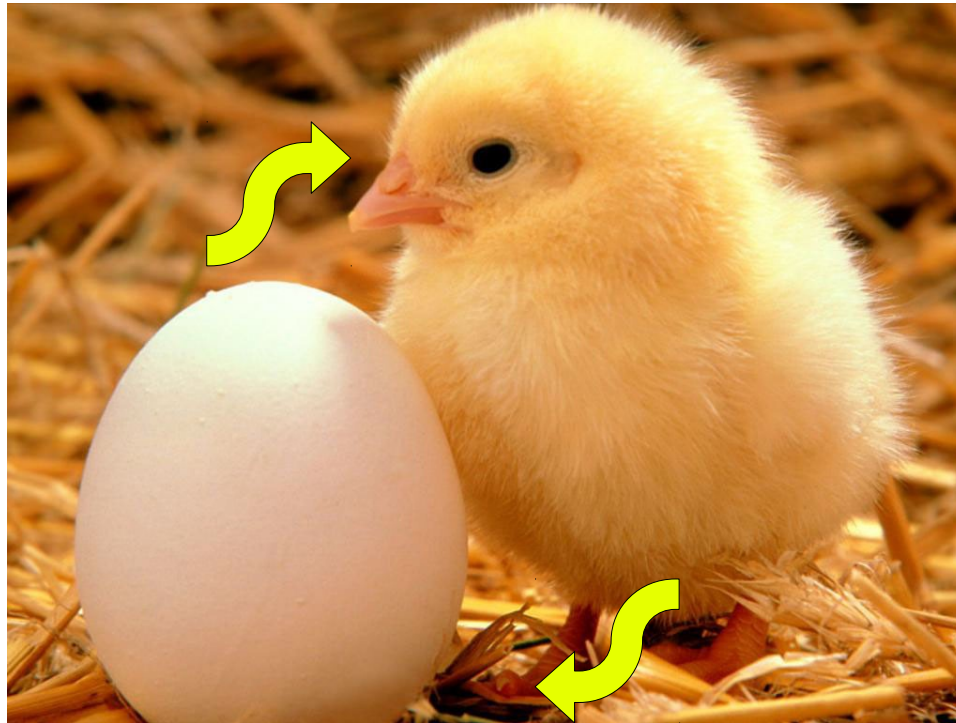
**Fig. 2.** GRN model for the endomesoderm 6-18 h of development. This is a "view from the genome" in which all regulatory interactions occurring through time are portrayed in the various domains (cf Fig. 1). For data and temporal and spatial regulatory views see <http://www.sugp.caltech.edu/endomes/>.



# Maybe . . .

Given real-world constraints, the process of eukaryotic DNA transcription cannot be modeled with the sequence alone. In other words, *de novo* exon prediction is impossible.

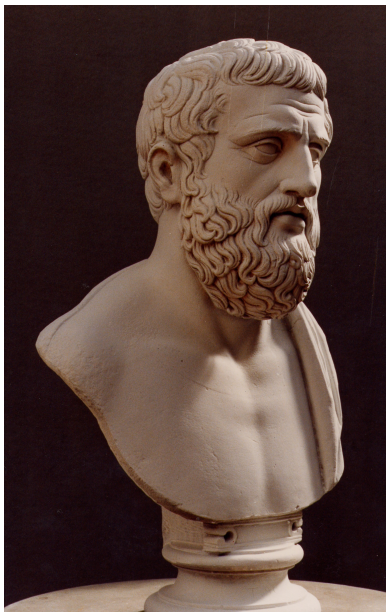
But consider this:



# Maybe . . .

Given real-world constraints, the process of eukaryotic DNA transcription cannot be modeled with the sequence alone. In other words, *de novo* exon prediction is impossible.

But consider this:



[15]

**Aristotle**  
**(Philosopher/Biologist)**

"If there has been a first man he must have been born without father or mother—which is repugnant to nature. For there could not have been a first egg to give a beginning to birds, or there should have been a first bird which gave a beginning to eggs; for a bird comes from an egg."

[16]



# Maybe . . .

Given real-world constraints, the process of eukaryotic DNA transcription cannot be modeled with the sequence alone. In other words, *de novo* exon prediction is impossible.

But consider this:

“[W]e appear to be faced with a paradox: when living organisms develop, the formation of new cell molecules and structures is directed by the base sequence of DNA; but DNA cannot function correctly without the prior presence of many of these molecules and structures arranged in the correct relative positions.

**However, living organisms are not as a result locked in a vicious circle against evolutionary change. During evolution DNA and other cell components pass from generation to generation coupled in a mutual interchange of 'information'.”**

–J.M. Barry (1986)

BEHOLD, THE MAJESTIC WILD BANANA. THIS LONE CUB HAS WANDERED AWAY FROM THE BUNCH TO SIP COOL WATER FROM THE STREAM, LEAVING ITSELF UNPROTECTED.

NEARBY, A PREDATOR MOVES STEALTHILY THROUGH THE GRASS.

# References (1)

BEHOLD, THE MAJESTIC WILD BANANA. THIS LONE CUB HAS WANDERED AWAY FROM THE BUNCH TO SIP COOL WATER FROM THE STREAM, LEAVING ITSELF UNPROTECTED.

NEARBY, A PREDATOR MOVES STEALTHILY THROUGH THE GRASS.

[1] Image: 2008-01-13.gif

Tatsuya Ishida (2008). Distress Signal. *Sinfest*, 2008 January 13. Available online: <[http://www.sinfest.net/archive\\_page.php?comicID=2686](http://www.sinfest.net/archive_page.php?comicID=2686)>.

[2] Image: beaker.jpg

Available online: <<http://lockheart.co.uk/bb/>>.

[3] Image: 3\_10.jpg

Biology 1100: Survey of Biology, College of DuPage. Available online: <[http://bio1100.nicerweb.com/Locked/media/ch13/13\\_10.jpg](http://bio1100.nicerweb.com/Locked/media/ch13/13_10.jpg)>.

[4] Image: bcbobblehead.jpg

Louise Bowman (2009). Buddy Christ. *Christ, Coffee, Chocolate, and the Internet*. Available online: <<http://louisebowmanonline.wordpress.com/2009/12/15/buddy-christ-2/>>.

[5] Image: bunsen-and-beaker.jpg

(2009). Obama Unveils Initiative to Improve Math and Science Education. *Dang That's Cool*.

Available online: <<http://dangthatscool.wordpress.com/2009/11/23/obama-unveils-initiative-to-improve-math-and-science-education/>>.

[6] Image: beaker-ii.jpg

Lower Management Beaker (2009). *HR Girl 4'11" Blog*. Available online: <<http://shortandsweethrgirl.wordpress.com/tag/lower-management/>>.

[7] Image: muppets-beaker.jpg

Dr. Bunsen Honeydew & Beaker – top boffins (2007). *Buffet o' Blog*. Available online: <<http://buffetoblog.wordpress.com/2007/01/18/dr-bunsen-honeydew-beaker-top-boffins/>>.

[8] Image: 4659\_1.jpg

ToyTokyo (2004). Other Plush Muppets Beaker. *ToyTokyo.com*. Available online: <[https://www.toytokyo.com/shopping/index.php/page/product/product\\_id/4659](https://www.toytokyo.com/shopping/index.php/page/product/product_id/4659)>.

AN ALL TOO COMMON DANCE OF DEATH.

AN ALL TOO COMMON DANCE OF DEATH.



BEHOLD, THE MAJESTIC WILD BANANA. THIS LONE CUB HAS WANDERED AWAY FROM THE BUNCH TO SIP COOL WATER FROM THE STREAM, LEAVING ITSELF UNPROTECTED.

NEARBY, A PREDATOR MOVES STEALTHILY THROUGH THE GRASS.

# References (2)

BEHOLD, THE MAJESTIC WILD BANANA. THIS LONE CUB HAS WANDERED AWAY FROM THE BUNCH TO SIP COOL WATER FROM THE STREAM, LEAVING ITSELF UNPROTECTED.

NEARBY, A PREDATOR MOVES STEALTHILY THROUGH THE GRASS.

[9] Image: EVCBOMPACALJQAAAACk.png

Ensembl (2010). Ensembl Genome Browser. Ensembl Project. Available online <  
[http://uswest.ensembl.org/Homo\\_sapiens/Location/View?db=core;r=11:116457105-116957106](http://uswest.ensembl.org/Homo_sapiens/Location/View?db=core;r=11:116457105-116957106)>.

[10] Image: lego-brick4-timeline.jpg

Gizmodo (2008). LEGO Brick Timeline: 50 Years of Building Frenzy and Curiosities. *Gizmodo.com*. Available online: <  
<http://gizmodo.com/349509/lego-brick-timeline-50-years-of-building-frenzy-and-curiousities>>.

[11] Image: Rex\_(Power\_Miners).jpg

(2009). File:Rex (Power Miners).jpg. *Brickipedia*. Available online: <  
[http://lego.wikia.com/wiki/File:Rex\\_%28Power\\_Miners%29.jpg#file](http://lego.wikia.com/wiki/File:Rex_%28Power_Miners%29.jpg#file)>.

[12] Chiou-Hwa Yuh, Hamid Bolouri, and Eric H. Davidson (1998). Genomic Cis-Regulatory Logic: Experimental and Computational Analysis of a Sea Urchin Gene. *Science*, Vol. 279, No. 5358, pp. 1896-1902.

[13] Image: sea\_urchin.jpg

Brady Oshiro (2008). VANA! Hawaiian Sea Urchin. *bradyoshiro.com*. Available online: <  
<http://bradyoshiro.com/blog/2008/06/02/vana-hawaiian-sea-urchin/>>.

[14] Image: Egg\_and\_Chicken.jpg

*Egg and Chicken*. *Wallpapers Free*. Available online: <  
[http://www.wallpapers-free.org/15/-/Egg\\_and\\_Chicken/](http://www.wallpapers-free.org/15/-/Egg_and_Chicken/)>.

[15] Image: aristotle.jpg

Renaissance Bronzes. Aristotle. *Renaissance Bronzes.co.uk*. Available online: <  
[http://www.renaissancebronzes.co.uk/renaissance\\_bronzes\\_plaster\\_aristotle.htm](http://www.renaissancebronzes.co.uk/renaissance_bronzes_plaster_aristotle.htm)>.

[16] H.P. Blavatsky (1877). *Isis Unveiled*, Volume I, Section II. Available online through Google Books: <  
<http://books.google.com/books?id=wNdd7alnRooC&dq=blavatsky%20isis%20unveiled&pg=PA428#v=onepage&q&f=false>>.

[17] J.M. Barry (1986). Informational DNA: a useful concept?. *Trends in Biochemical Sciences*, pp. 317-318.

[18] Image: 20070901.jpg

Tim Buckley (2007). *Bongos*. *Ctrl+Alt+Del*. Available online: <  
<http://www.cad-comic.com/cad/20070901>>.