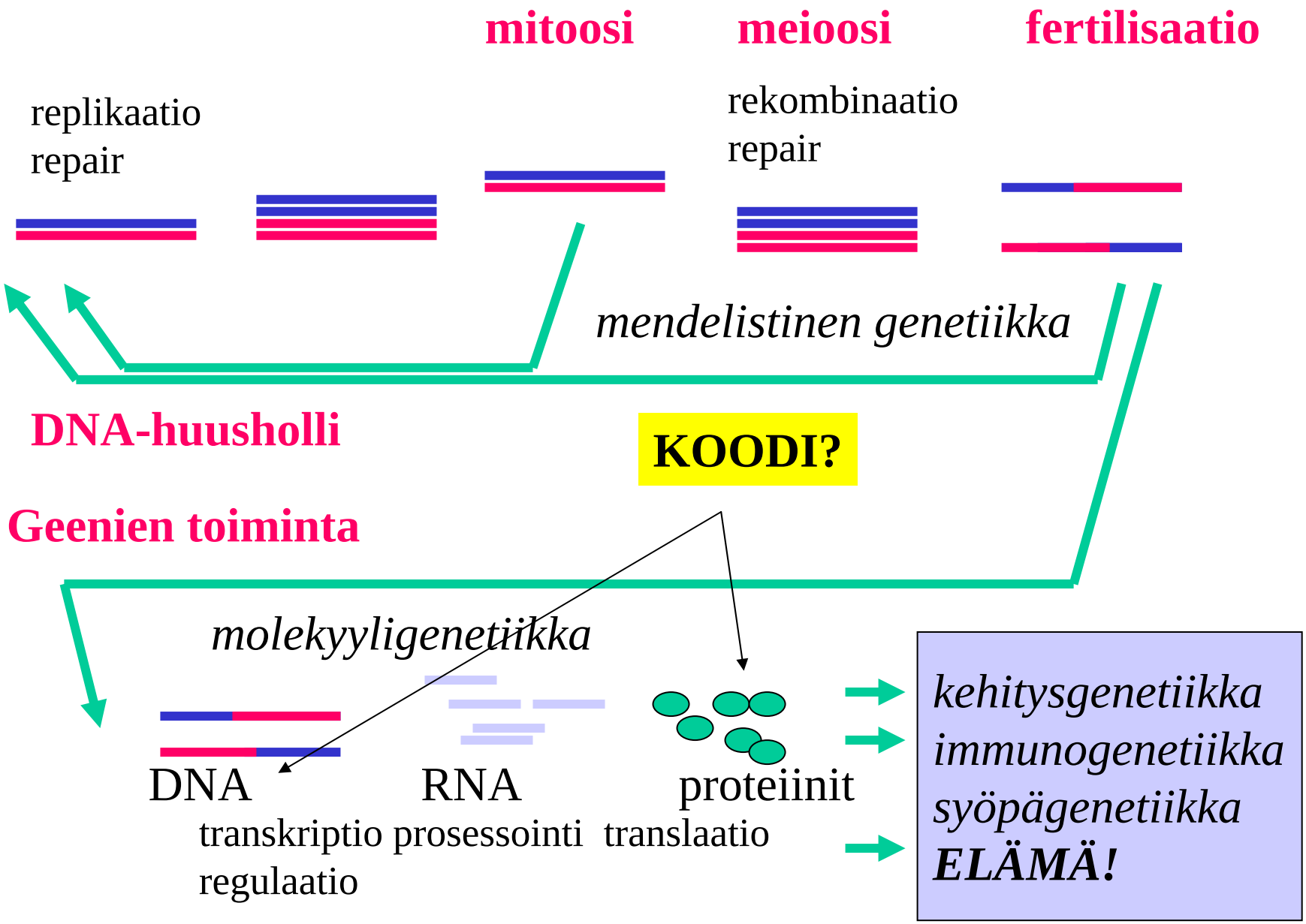


Genetiikan perusteiden miellekartta: EST ja geenien ekspressio

Geeni oli alunperin abstrakti yksikkö, jonka (alleelien) periytymistä saatettiin seurata mendelistisellä mallilla

Geeni rupesi saamaan konkreettista sisältöä, kun kartoitus eteni oikein tarkaksi. Ruvettiin puhumaan lokuksesta (locus, loci), jossakin vaiheessa kistronista (cistron: esimerkiksi lac-geeni on "polykistroninen"). Operoni jo opeteltiin: geeni sekin, jossa erotellaan erilaisia funktioita

DNA:n sekvensoinnin jälkeen geenillä on varsin tarkka sisältö, mutta *määritteleminen* on tullut vaikeammaksi

Mitä määritteleminen on? Biologiassa erityisesti?

Jonkun asian rajaamista yhdellä kauniilla lauseella siten että se sisältää kaikki joukon objektit ja sulkee pois kaikki ne jotka eivät kuulu joukkoon

Geeni?

Useiden geenien lopputuote on polypeptidiketju, joista solu rakentelee proteiineja rakenteisiin ja reaktioihin

Sekundaarirakenne - tertiaarirakenne - polymerisaatio etc

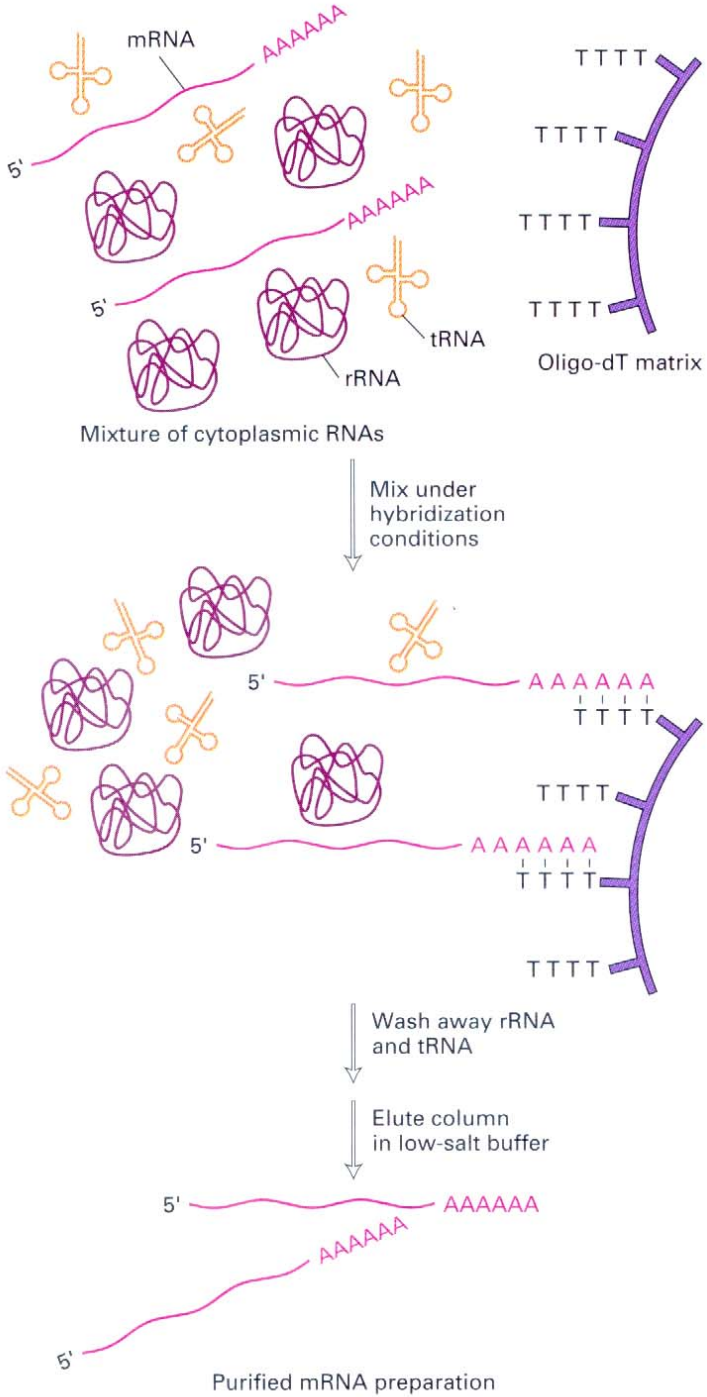
Hävityskin kuuluu proteiinin olemukseen

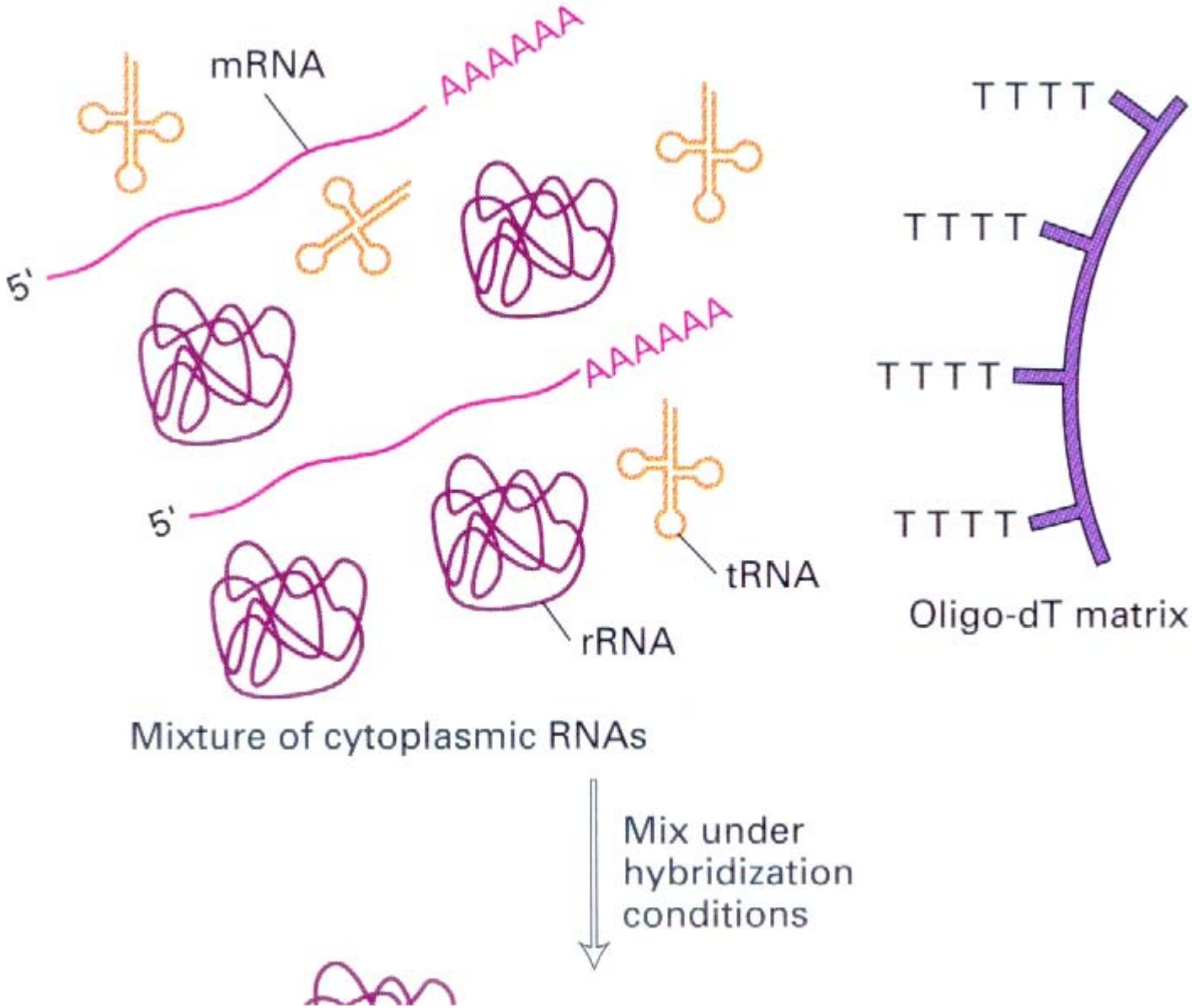
Tällaisen "geenin" ominaisuuksiin kuuluvat:
- säätelyyn liittyvät DNA-jaksot ylä- ja alavirrassa
  promoottori, operaattori, enhancerit yms
- "turhat" välikkeet (intronit), joilla kuitenkin on merkitystä
- RNA:han kirjoitettavat tekstit:
  varsinainen *koodi*, aloitus, lopetus, poly-A -signaali
  Untranslated regions (UTR), IRES etc)

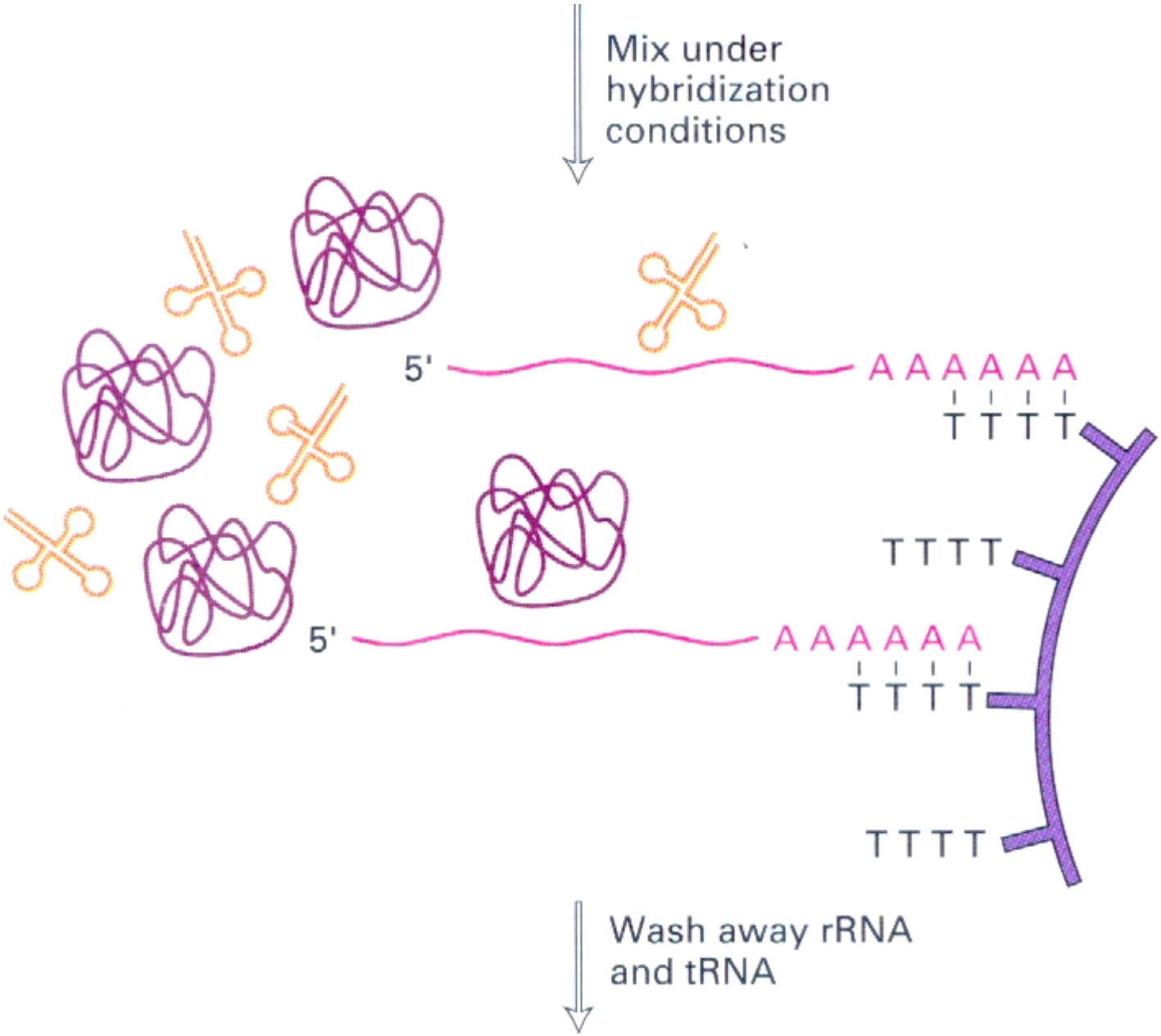
Tärkeä keino “geenien” löytämiseksi on eristää ja sekvensoida mRNA-molekyylejä

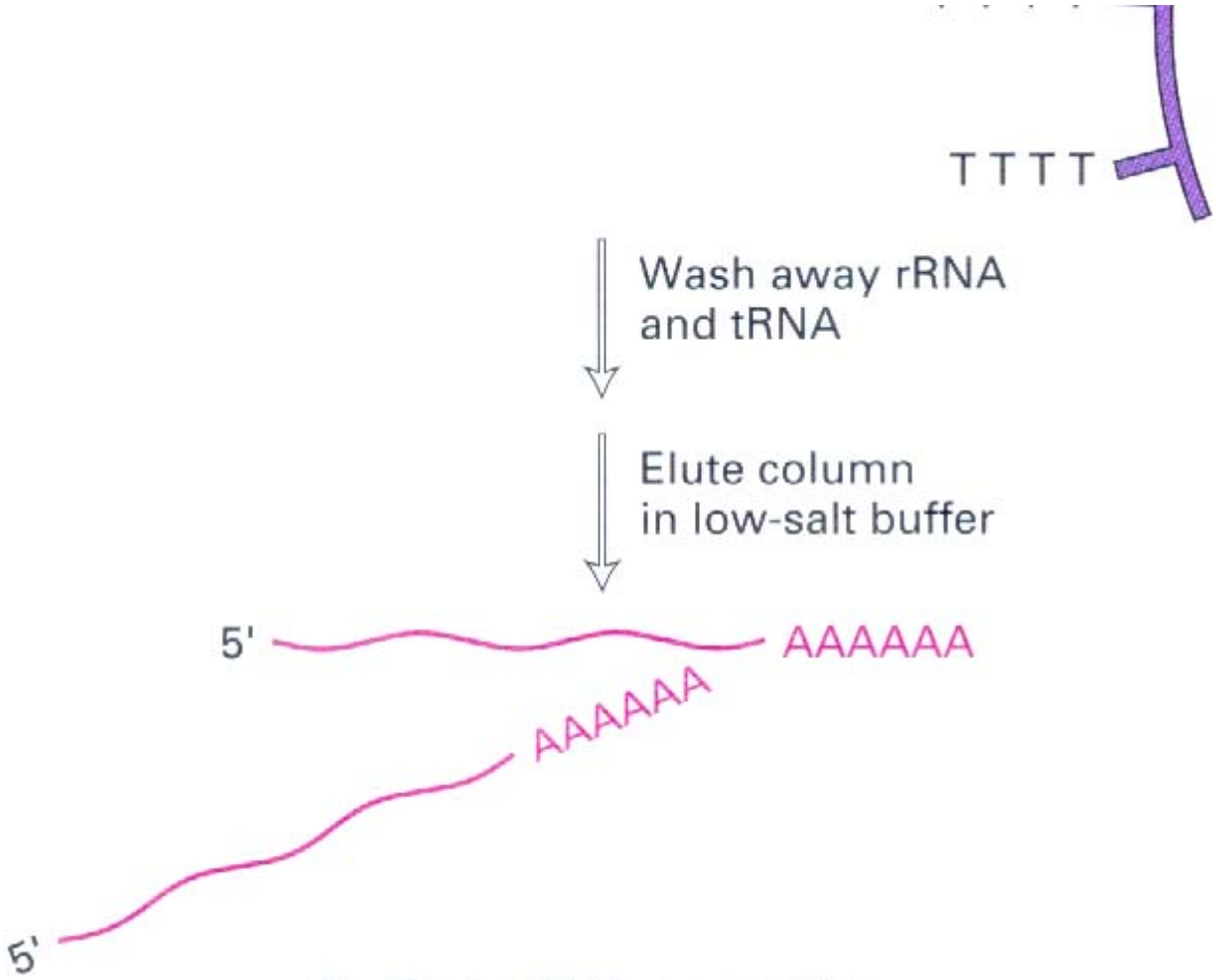
1. askel: **eristetään mRNA**

No miten? No hännästä nostamalla, tietenkin, sitä vartenhan se poly(A)-häntä on


mRNA
AAAAAA
5'
AAAAAA
5'
tRNA
rRNA
Mixture of cytoplasmic RNAs
TTTT
TTTT
TTTT
TTTT
Oligo-dT matrix
Mix under hybridization conditions
5'
AAAAAA
TTTT
TTTT
5'
AAAAAA
TTTT
TTTT
Wash away rRNA and tRNA
Elute column in low-salt buffer
5'
AAAAAA
AAAAAA
5'
Purified mRNA preparation


mRNA
AAAAAA
5'
AAAAAA
5'
tRNA
rRNA
TTTT
TTTT
TTTT
TTTT
Oligo-dT matrix
Mixture of cytoplasmic RNAs
Mix under
hybridization
conditions


Mix under
hybridization
conditions
5'
A A A A A A
T T T T
T T T T
5'
A A A A A A
T T T T
T T T T
Wash away rRNA
and tRNA



Purified mRNA preparation

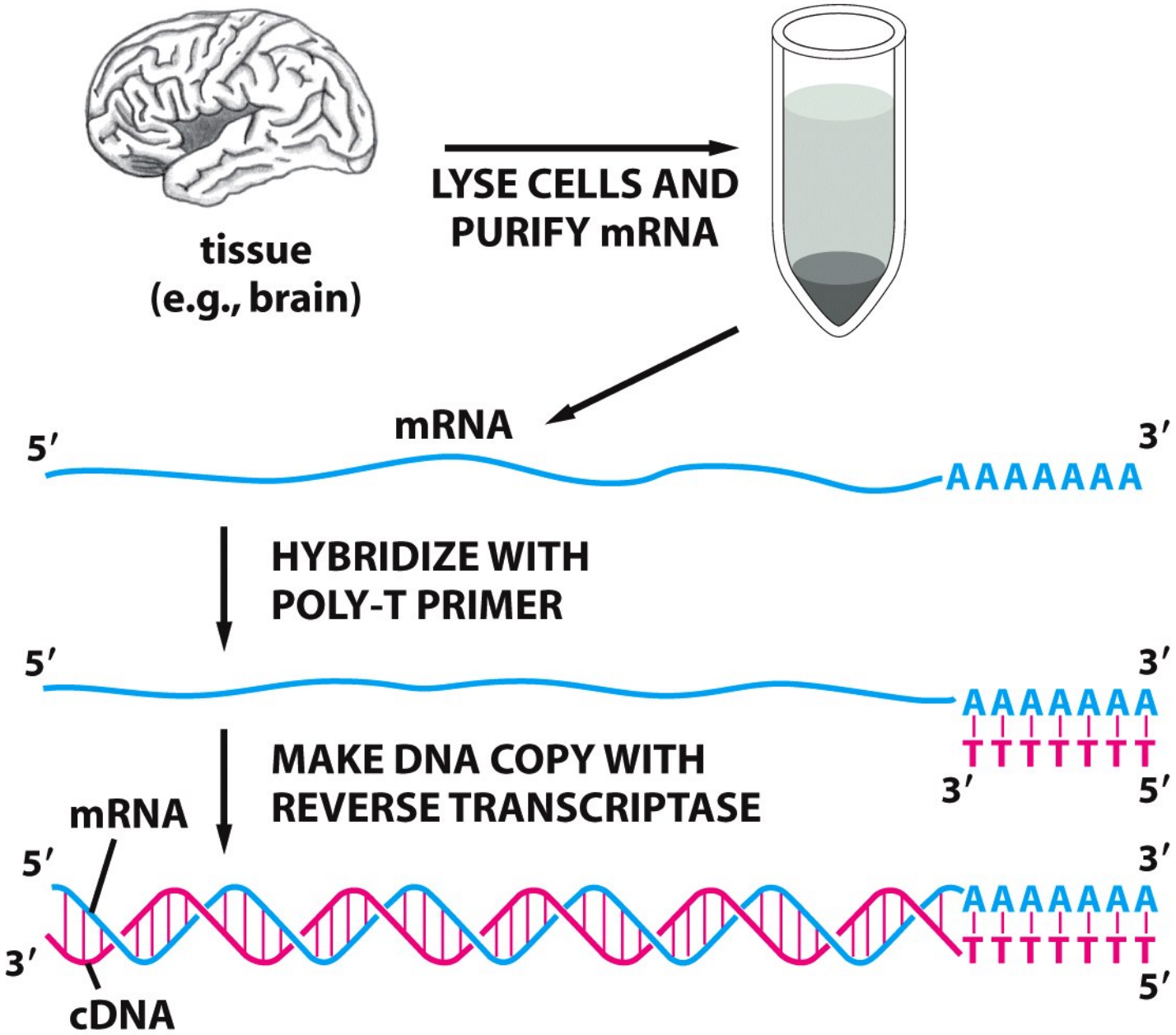


Figure 8-43 part 1 of 2 Molecular Biology of the Cell 5/e (© Garland Science 2008)

CELL 543

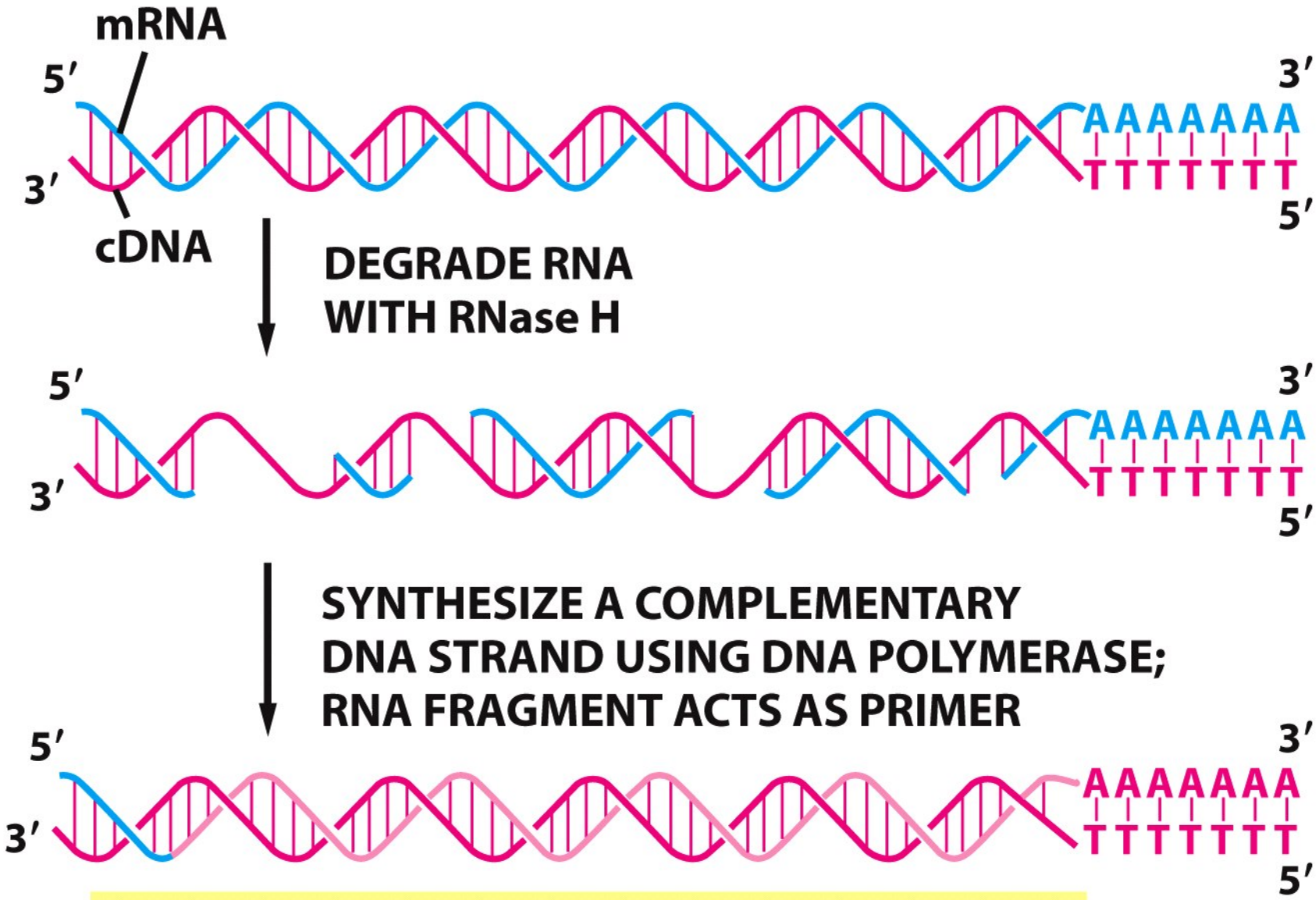


Figure 8-43 part 2 of 2 Molecular Biology of the Cell 5/e (© Garland Science 2008)

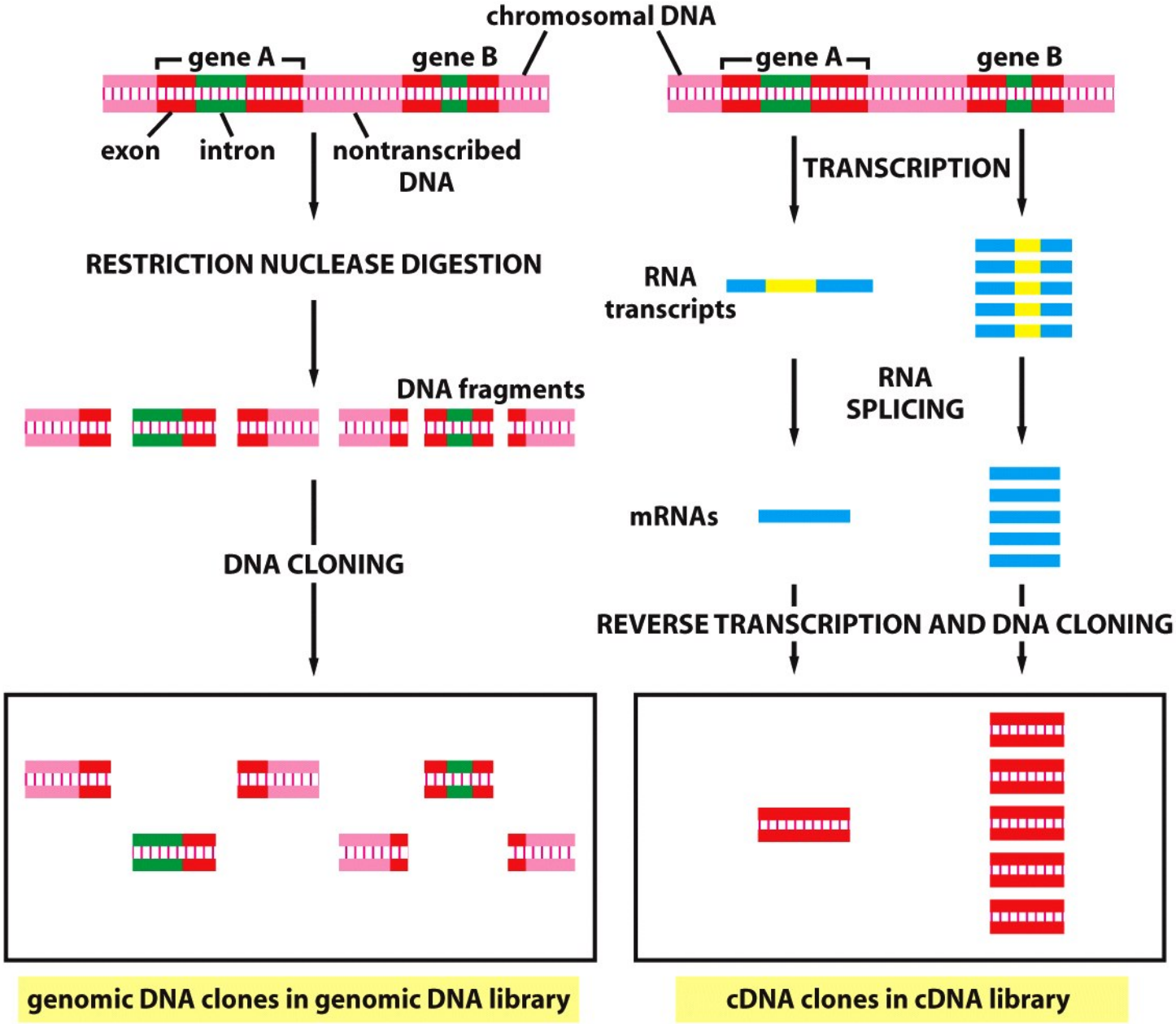


Figure 8-44 Molecular Biology of the Cell 5/e (© Garland Science 2008)

Geeni?

Joidenkin geenien ainoa tuote on RNA

Defining Genes in the Genomics Era

Michael Snyder and Mark Gerstein

A genome is defined as the entire collection of genes encoded by a particular organism. But what is a gene? Historically, the term gene, attributed to Johansson, first appeared in the early 1900s as an abstract concept to explain the hereditary basis of traits (*1, 2*). Phenotypic traits were ascribed to hereditary factors even though the physical basis of those factors was not known. Subsequently, early genetic studies by Morgan and others associated heritable traits with specific chromosomal regions. In the 1930s, Beadle introduced the concept of "one gene, one enzyme," which later became "one gene, one polypeptide."

With the advent of recombinant DNA and gene cloning, it became possible to combine the assignment of a gene to a specific segment of DNA and the production of a gene product. Although it was originally presumed that the final product was a protein, the discovery that RNA has structural, catalytic, and even regulatory properties made it evident that the end product could be a nucleic acid (*3*). Thus, we now define a gene in molecular terms as "a complete chromosomal segment responsible for making a functional product." This definition has several logical components: the expression of a gene product, the requirement that it be functional, and the inclusion of both coding and regulatory regions. According to this definition, it should be possible to use straightforward criteria to identify genes in the DNA sequence of a genome. Five such criteria are in common use, but their application is not straightforward.

Open reading frames (ORFs). An ORF is a string of codons bounded by start and stop signals, where codons are nucleotide triplets encoding amino acids. An obvious way to find protein-coding genes is through identifying large ORFs in the genome. This is particularly applicable to prokaryotes and other organisms with few introns (the regions spliced out of RNA) in their genes. Even so, many genes are short and difficult to identify in this way. Moreover, organisms with genes that undergo an appreciable amount of RNA splicing often have small exons sandwiched between large introns, making ORFs especially difficult to find.

Sequence features. Once an ORF is identified, codon bias often is used to determine whether the ORF is a gene (*4*). The value of this measure stems from the fact that genes, particularly highly expressed genes, exhibit biased nonrandom use of codons. However, for many genes, the bias is weak, and small ORFs (or exons) contain too few codons to exhibit statistically significant bias. Beyond overall bias, one can also look for specific patterns in the DNA sequence such as splice sites to help locate genes (*5*). Computer programs that use DNA sequence features alone predict fewer than 50% of exons and 20% of complete genes (*5*). Moreover, while both the existence of an ORF and favorable sequence features may imply the presence of a gene product, they say nothing about that product's function.

Sequence conservation. In contrast to focusing on an individual DNA sequence, genes can be identified by comparing multiple sequences among organisms (*4, 5*). DNA sequence conservation among species is an excellent method to gauge the importance of the gene product. However, conserved sequences could be nontranscribed regulatory elements. Another problem with using conservation to find genes is that it requires sequences of related organisms that are separated by appropriate evolutionary distances. A current estimate of the number of genes in an organism can never be an absolute, unchanging number, because it is contingent on the specific related organisms used for comparison.

Evidence for transcription. A non–sequence-based approach for identifying genes is to search for RNA or protein expression, the hallmark of a gene product. This is commonly accomplished using microarray hybridization, serial analysis of gene expression (SAGE), cDNA mapping, or sequencing of expressed sequence tags (*6–8*). Large-scale tagging of genes with transposons reveals many new regions in the yeast genome that are capable of producing proteins (*9*) (see the figure). Likewise for humans, hybridization of labeled cDNAs to

ORF
Codon bias
Seq. conservation
Transcriptio
Gene inactivation

j.n.e.

Science 11 April 2003