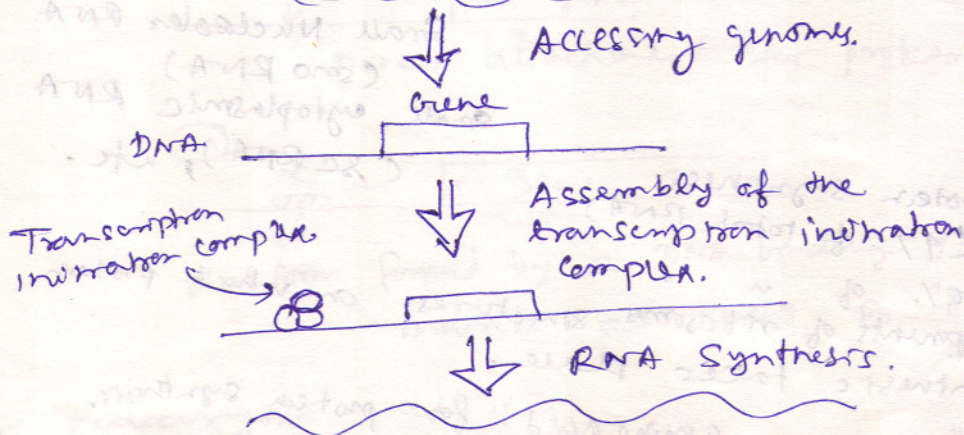
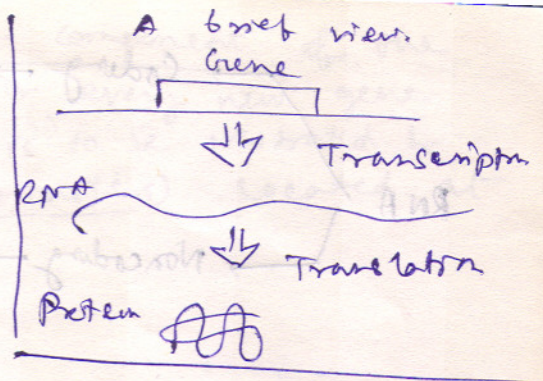
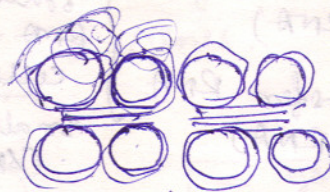
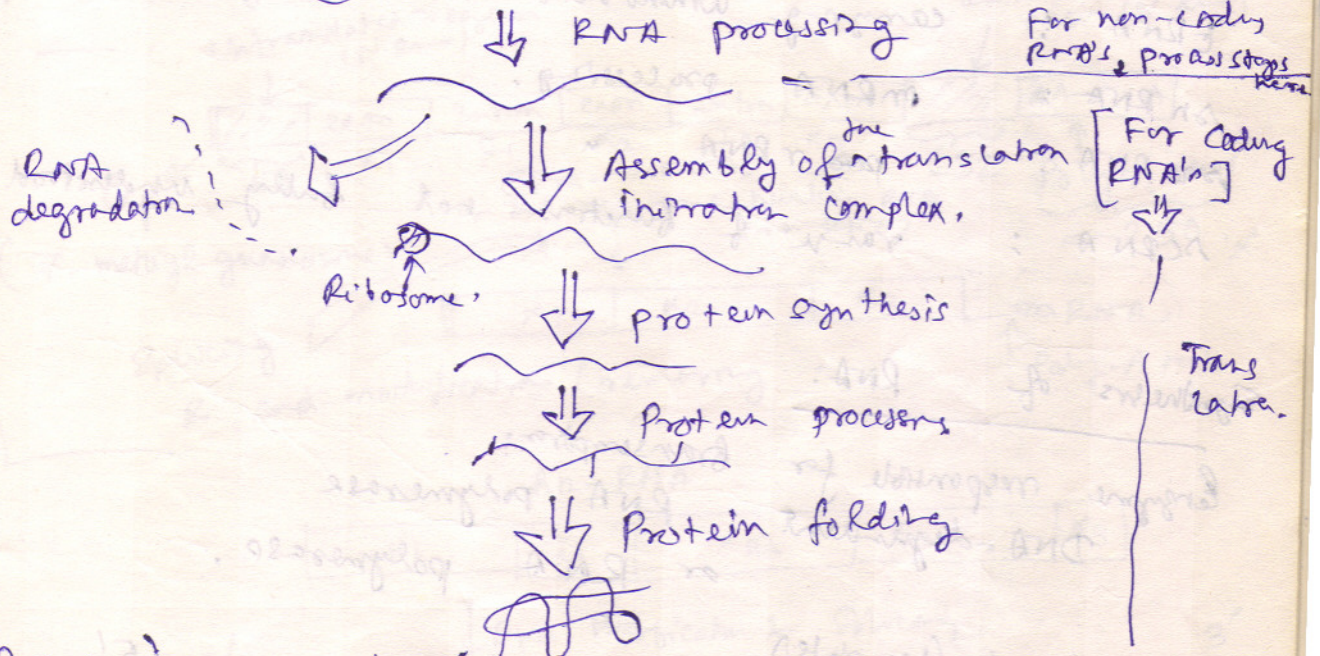


Gene Expression & Genome Anatomy.



Transcription



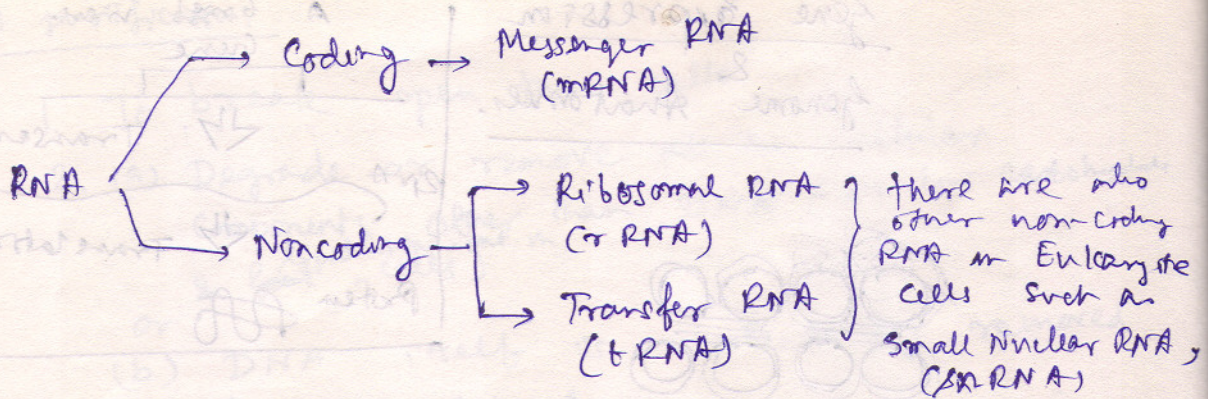
For non-coding RNAs, processing stops here.

For coding RNAs

Translation

Protein degradation

These schemes



Roles:

mRNA: Protein synthesis.
(1% of total RNA)

rRNA: 80% of " "
components of ribosomes structures on which protein synthesis takes place.

tRNA: carrying amino acid for protein synthesis.

snRNA: mRNA processing.

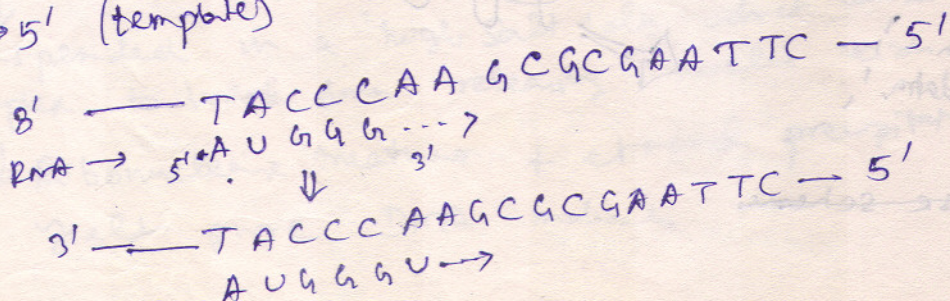
snoRNA: ~~pro~~ rRNA "

scRNA: various functions not fully understood

Synthesis of RNA:

Enzyme responsible for DNA-dependent transcription:
RNA polymerase
or RNA polymerase.

DNA: 3' → 5' (template)

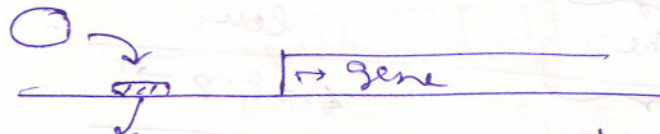


RNA polymerase is the central component of the transcription initiation complex. For every new gene transcription, a new complex is to be initiated & bound at target site (or promoters) located at upstream of genes.

RNA polymerase



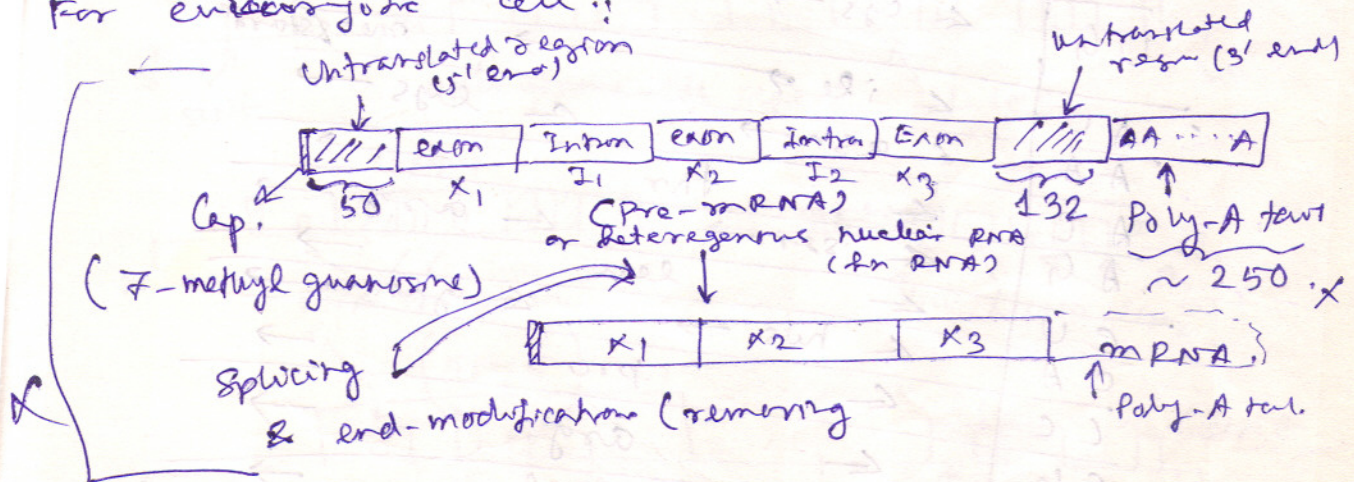
Direct attachment for prokaryotic cell



platform formed by a DNA binding protein

indirect attachment for Eukaryotic cell.

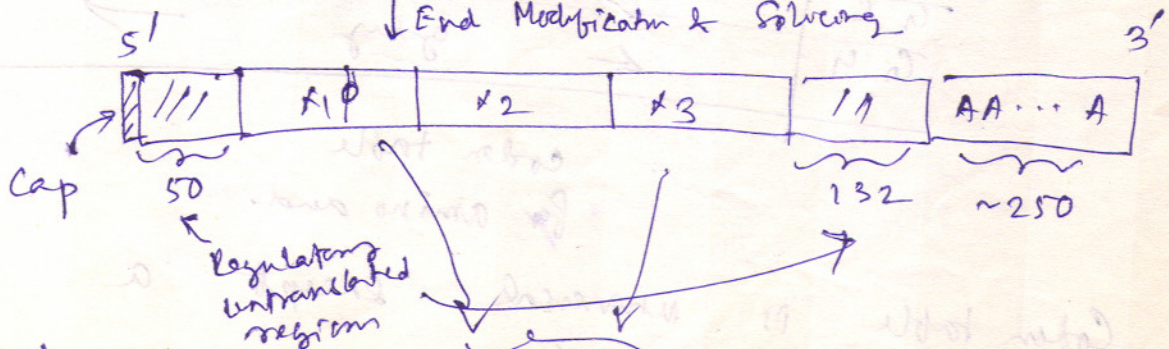
For eukaryotic cell:



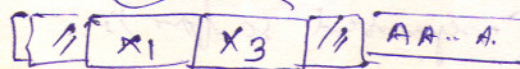
hn-RNA



End Modification & Splicing



Alternative splicing:



Translation:

5' - AUG | CUU | CAU | - - - | UAG | UAA | 3'

start
Codon
& Methionine

tyr | stop
Codon

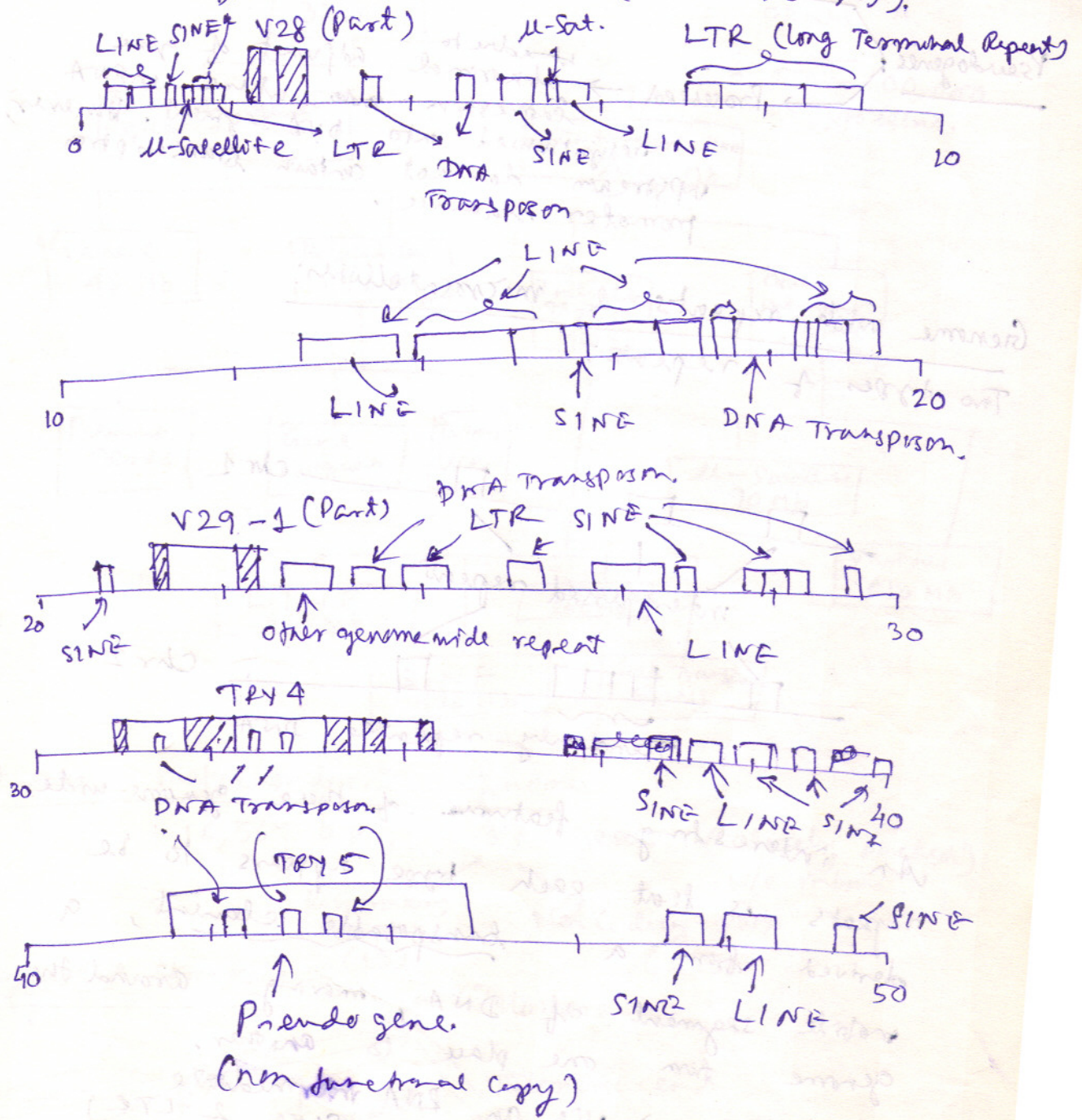
Genetic Code:

$X_1, X_2 \backslash X_3$	U	AC	CA	G
UU	phe		leu	
UA	tyr		stop	
UC		ser		
UG	cys		stop	stop
AU	ile			met & start
AA	asn		lys	
AC		thr		
AG	ser		arg	
CU		leu		
CA	his		gln	
CC		pro		
CG		arg		
GU		val		
GA	asp	ala	glu	
GC		ala		
GG		gly		

codon table
for amino acids

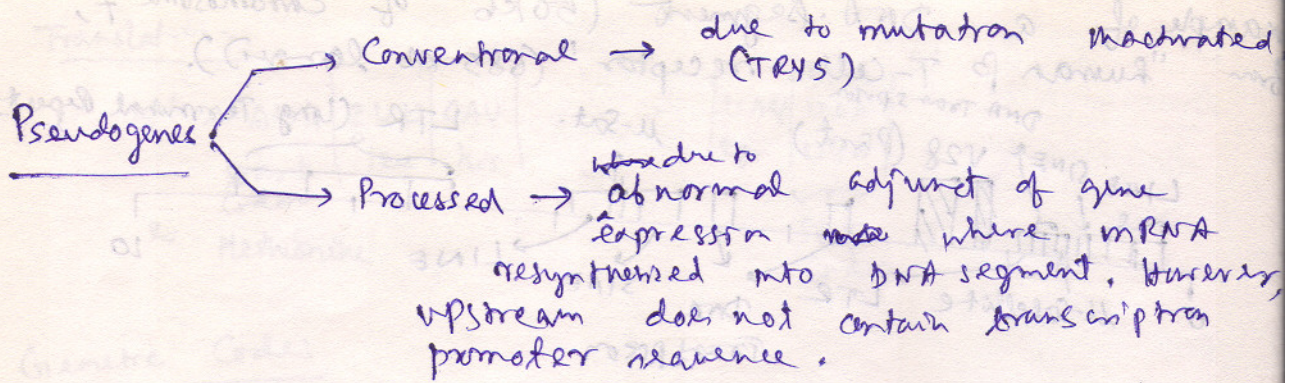
Codon table is universal except a few cases. (reported so far),

Example of a DNA Segment (50 kb of chromosome 7, from "human β T-cell receptor" (685 kb long)).



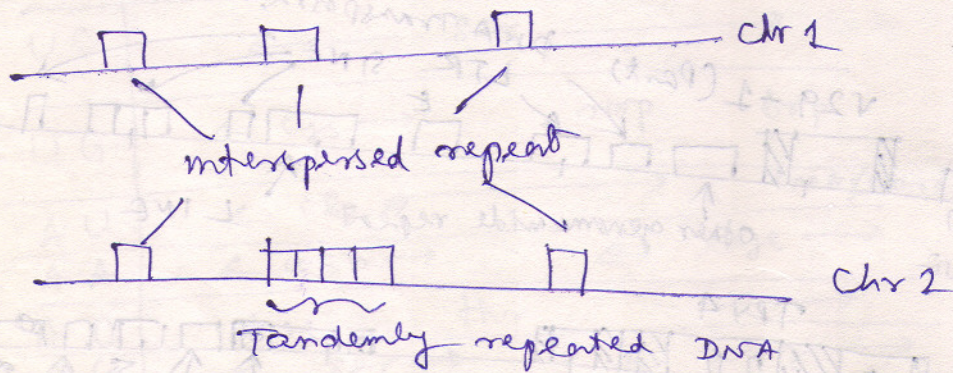
(9d 61 >)

...

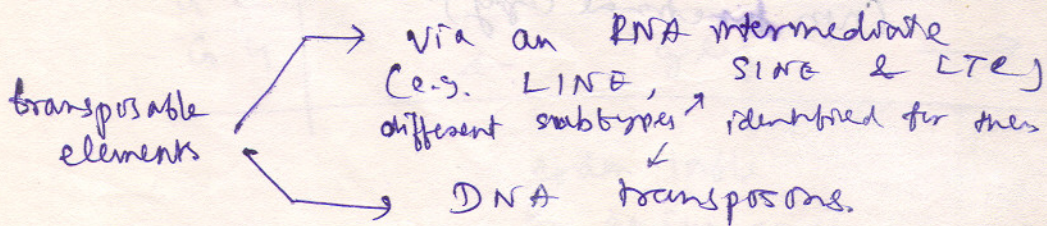


Genome wide repeats & microsatellites:

Two types of repeats:

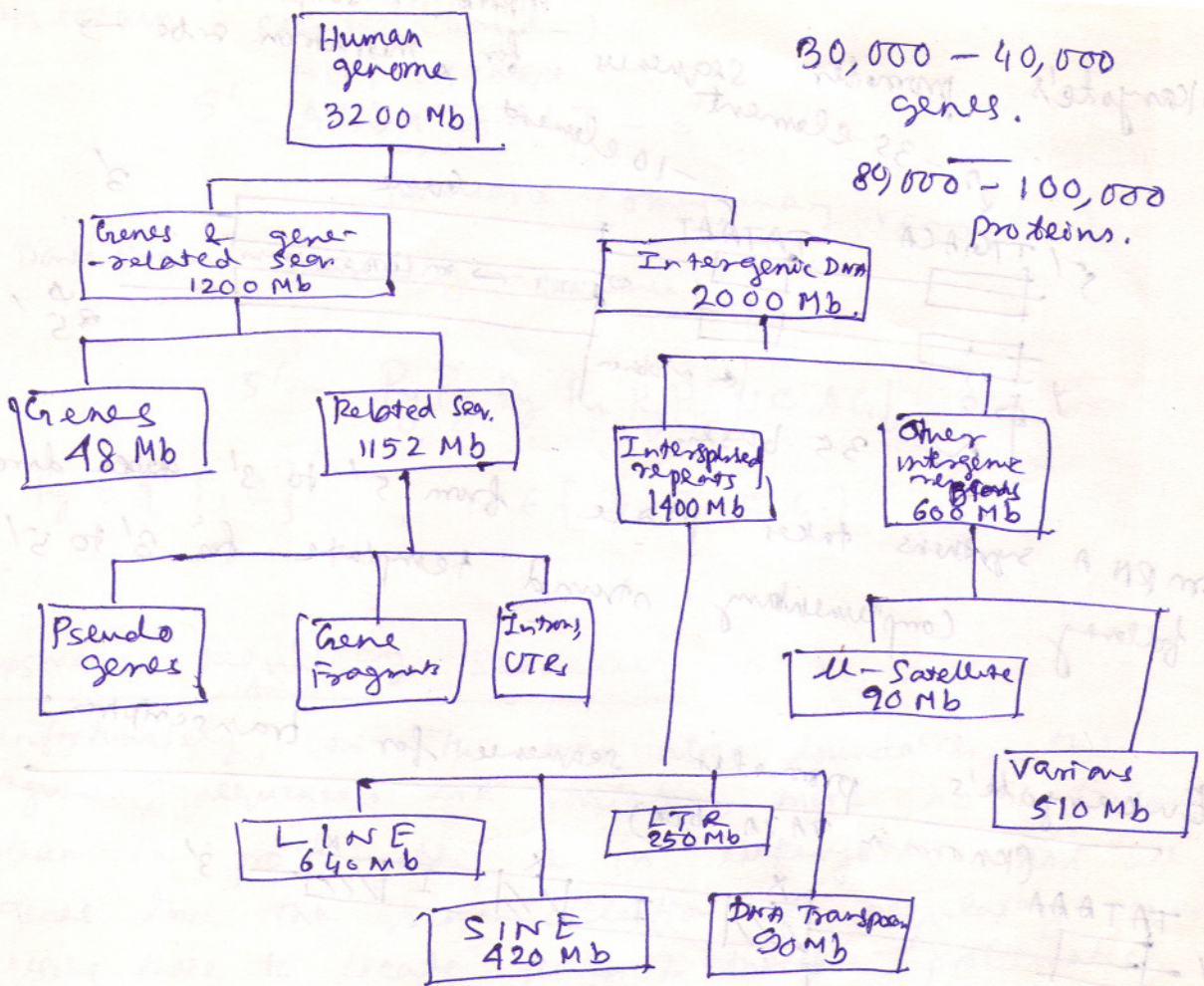


An interesting feature of these genome-wide repeats is that each type appears to be derived from a transposable element, a mobile segment of DNA, moving around the genome from one place to another.



Microsatellite: repeat unit is short (< 13 bp)

"A" repeat (120k in humans)
 "CA" repeat (140,000 copies)

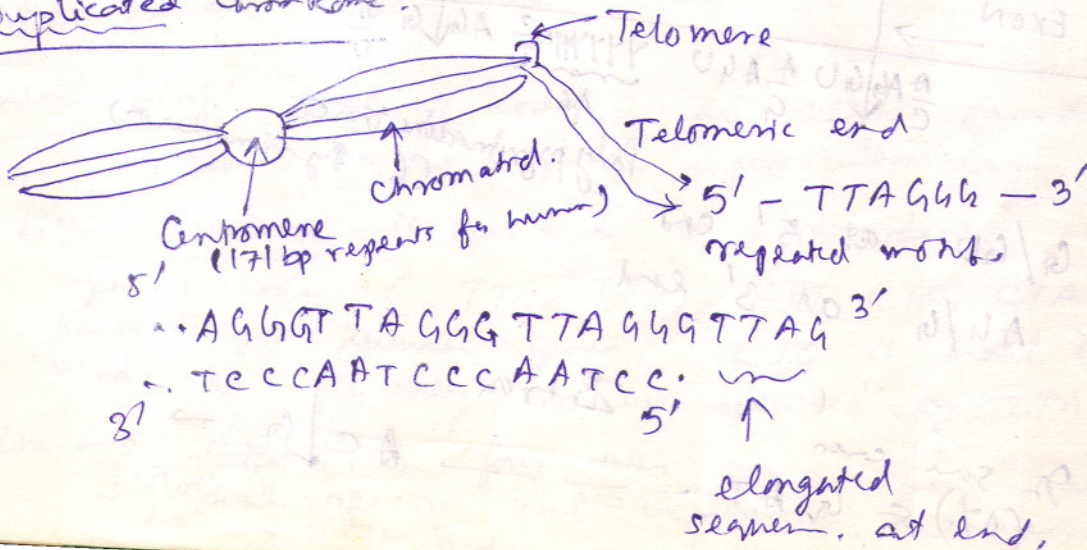


Human mitochondrial genome:

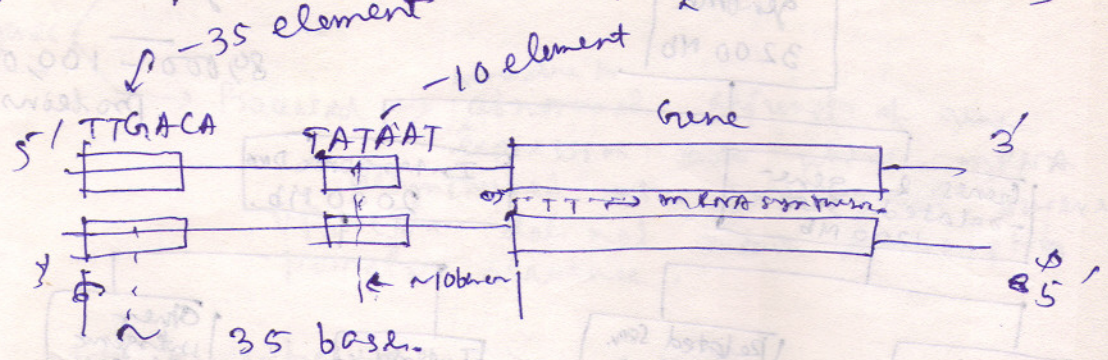
16,569 bp, 37 genes (closely packed) w/o introns

Respiratory Complex (13) Non Coding RNA (24)

Duplicated chromosome:

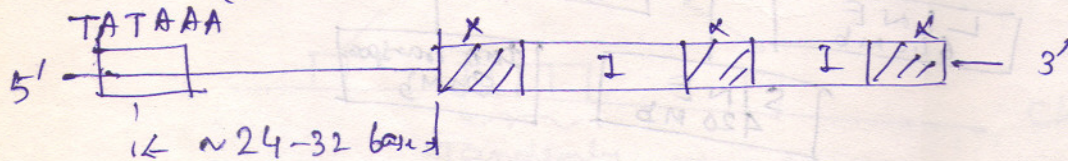


⑦ Prokaryote's promoter sequences for mRNA transcription initiation site →

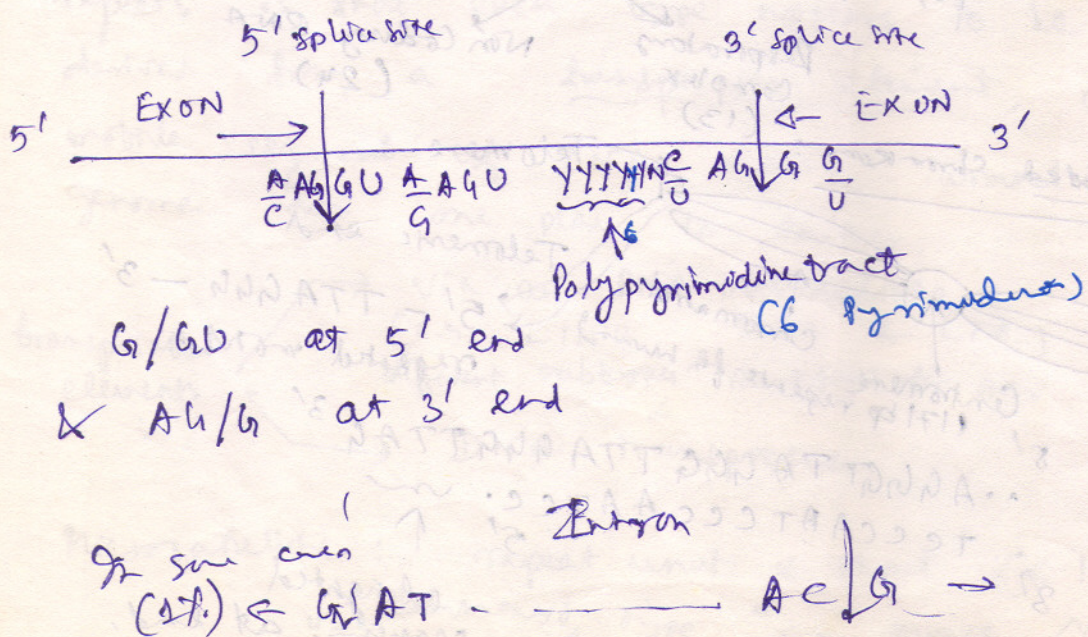


mRNA synthesis takes place from 5' to 3' direction following complementary strand template from 3' to 5'.

Eukaryote's promoter sequence for transcription:
(known as TATA box)



✱ EXON - INTRON boundaries:



upstream sequence: (border)

EXON; INTRON
5' - AG↓GTAAT - 3'

invariable (common many runs)

Downstream boundary sequence:

5' - PyPyPyPyPyPyNCAG↓ - 3'

Py $\in \{T, C\}$, N $\in \{A, T, C, G\}$

Upstream Regulatory Sequence:

unfortunately, as with exon-intron boundaries, the regulatory sequences are variable, more so in Eukaryotes than in prokaryotes, & in eukaryotes not all genes have the same collection of regulatory sequences. Using these to locate genes is therefore problematic.

Vertebrate genomes contain CpG islands upstream of many genes. These sequences of approximately 1 kb in which the GC content is greater than the avg. for the genome as a whole.

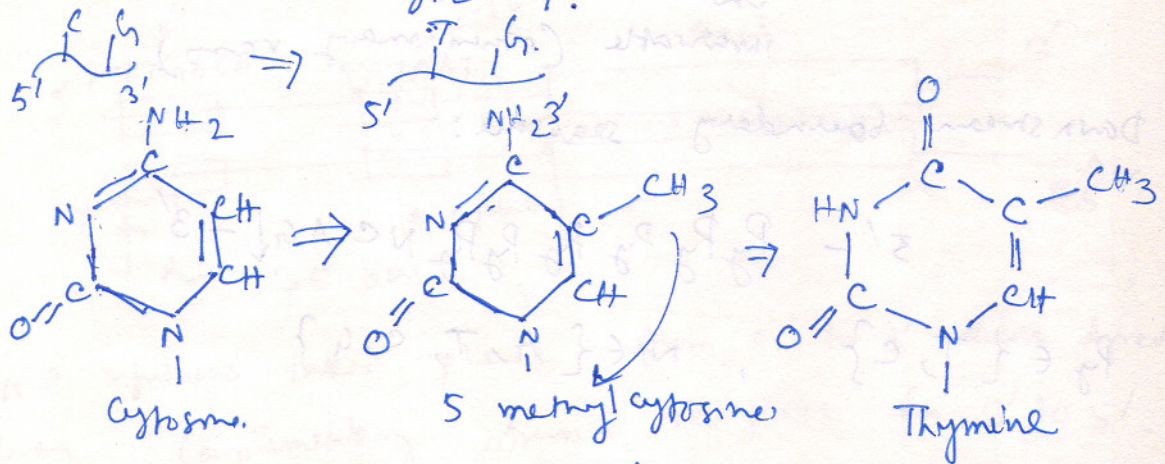
For human: 40-50% gene have upstream CpG islands.

Codon bias: Not all codons are used equally frequently in the genes of a particular org.

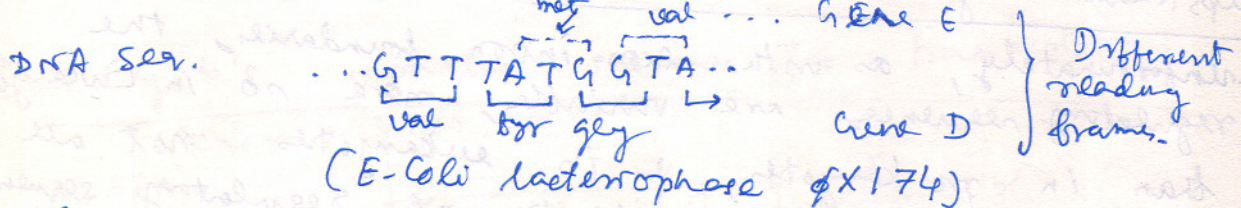
e.g. leucine is specified by six codons in the genetic code (TTA, TTG, CTT, CTC, CTA & CTG); however for human genes CTG is most frequently used & is only rarely specified by TTA & CTA.

Valine $\rightarrow$ ~~CTG~~ GTG four times more freq. than GTA. Biological reason not understood. But even org. has a codon bias.

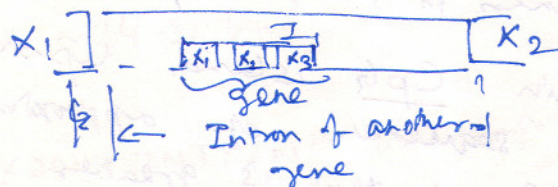
- ⑧ The sequence 5'-CG-3' is rare in human DNA because of methylation of the C, followed by deamination to give T.



- ⑨ Overlapping genes: (small genome met an virus)



- ⑩ Genes-within-genes (relatively common in Eukaryotes)



Recently it has been discovered many snRNA are encoded within Introns.

- ⑪ multi gene families: Groups of genes of identical or similar sequences.

- multiple copies of rRNA. (e.g. 200 5S rRNA of human genome)
- 28S, 5.8S, 18S rRNA repeat units grouped into 5 clusters. $\Rightarrow$ sedimentation factor.
- families arise due to gene duplication
- individual members of multigene families of higher organisms may perform quite different functions (diverging)
- $\Rightarrow$ human α & β -globin cluster (genes in this cluster are expressed from embryo to fetus development)
- multigene family may be observed as cluster or dispersed throughout the genome $\rightarrow$ 5 human genes for alcohol dehydrogenase (in chr. 3, 9, 10, 16 & 17).