



Genoma Umano IV

Gabriella De Vita

gdevita@unina.it

13 ottobre 2022

Eterocromatina e DNA
altamente ripetitivo

Classi di DNA ripetuto del genoma umano

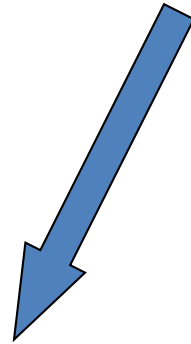
Alto numero di copie:

- Trasposoni (oltre 40% del genoma)
- Eterocromatina (satellite e minisatellite, 6.5%)
- Microsatellite

Basso numero di copie:

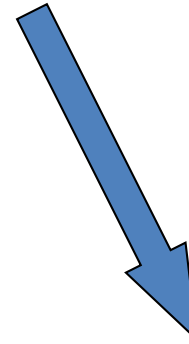
- Lunghe sequenze ripetute spesso contenenti geni (duplicazioni segmentali)
- Famiglie geniche

Il DNA altamente ripetitivo non-codificante



clusterizzato

(→ → → →)



disperso

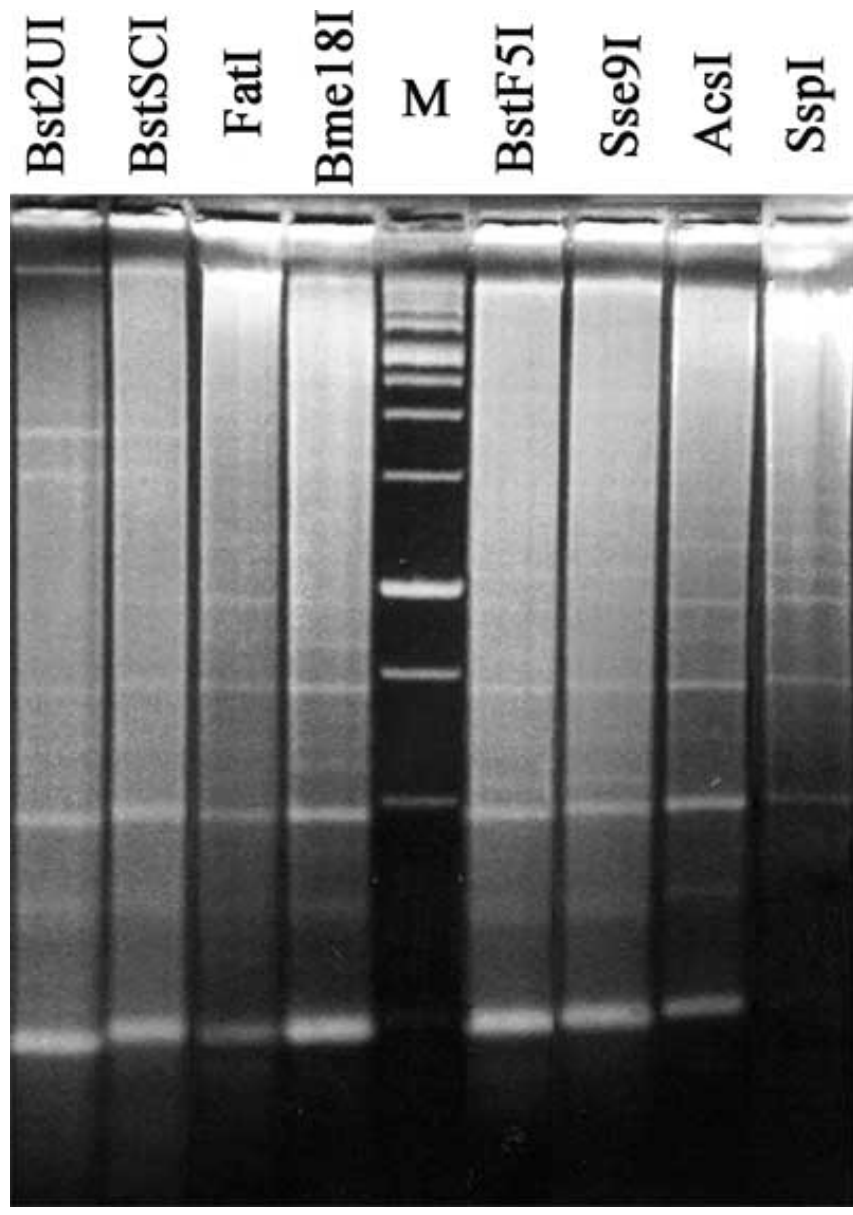
DNA ripetitivo non-codificante clusterizzato

Blocchi di ripetizioni in tandem

Classificato in:

- **DNA satellite**
- **DNA minisatellite**
- **DNA microsatellite**

in base alla taglia dell'array (blocco) di ripetizioni



Il DNA satellite in gradiente di densità

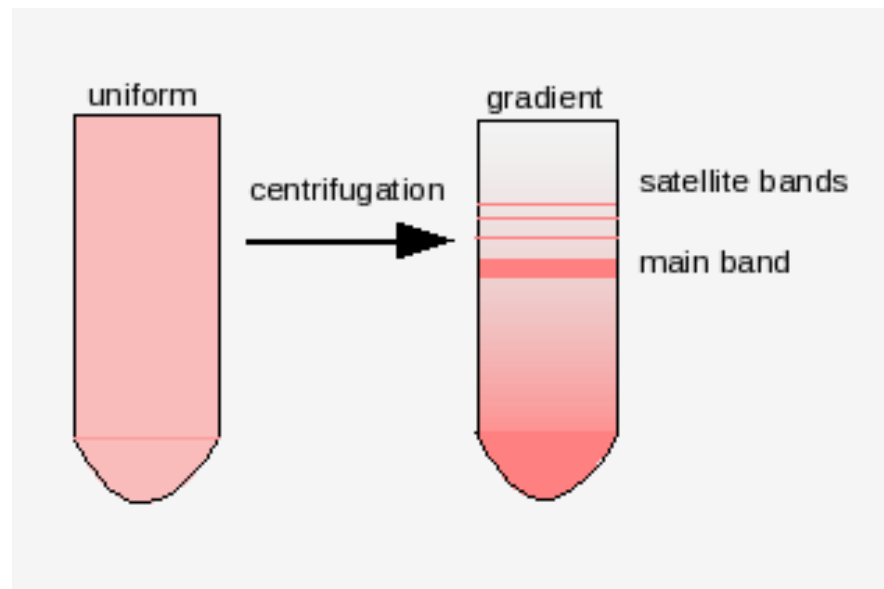


TABLE 9.13 MAJOR CLASSES OF HIGH-COPY-NUMBER TANDEMLY REPEATED HUMAN DNA

Class ^a	Total array size unit	Size or sequence of repeat unit	Major chromosomal location(s)
Satellite DNA^b	often hundreds of kilobases		associated with heterochromatin
α (alphoid DNA)		171 bp	centromeric heterochromatin of all chromosomes
β (<i>Sau3A</i> family)		68 bp	notably the centromeric heterochromatin of 1, 9, 13, 14, 15, 21, 22, and Y
Satellite 1		25–48 bp (AT-rich)	centromeric heterochromatin of most chromosomes and other heterochromatic regions
Satellite 2		diverged forms of ATTCC/GGAAT	most, possibly all, chromosomes
Satellite 3		ATTCC/GGAAT	13p, 14p, 15p, 21p, 22p, and heterochromatin on 1q, 9q, and Yq12
DYZ19		125 bp	~400 kb at Yq11
DYZ2		AT-rich	Yq12; higher periodicity of ~2470 bp
Minisatellite DNA	0.1–20 kb		at or close to telomeres of all chromosomes
Telomeric minisatellite		TTAGGG	all telomeres
Hypervariable minisatellites		9–64 bp	all chromosomes, associated with euchromatin, notably in sub-telomeric regions
Microsatellite DNA	< 100 bp	often 1–4 bp	widely dispersed throughout all chromosomes

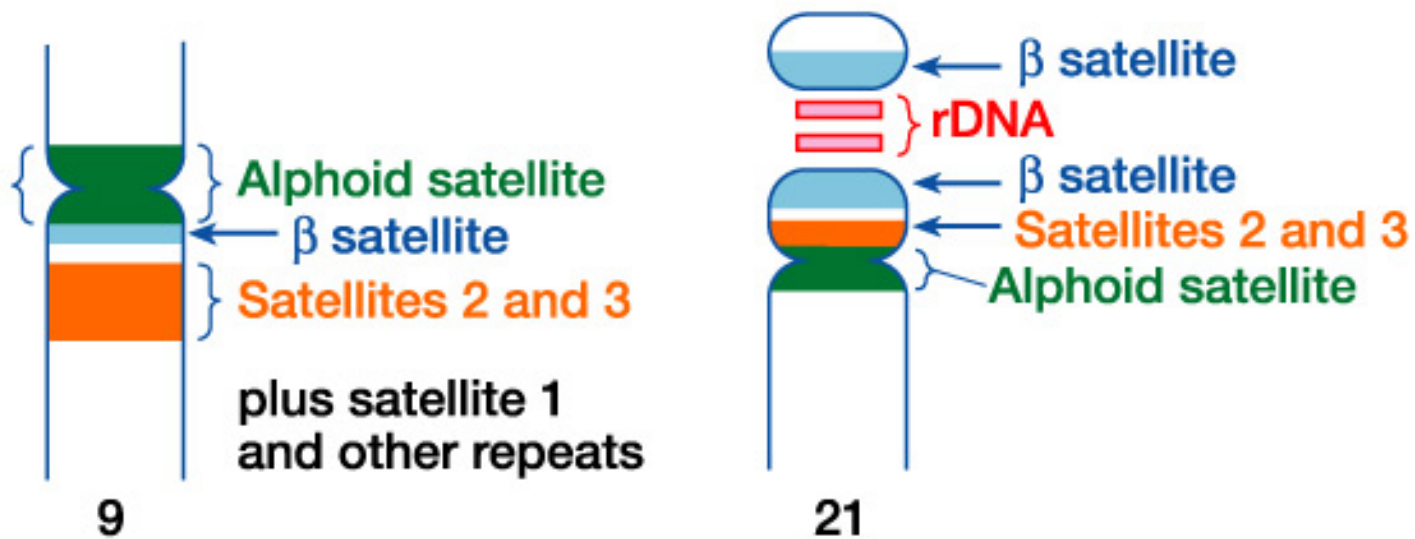
^aThe distinction between satellite, minisatellite, and microsatellite is made on the basis of the total array length, not the size of the repeat unit.

^bSatellite DNA arrays that consist of simple repeat units often have base compositions that are radically different from the average 41% G+C (and so could be isolated by buoyant density gradient centrifugation, when they would be differentiated from the main DNA and appear as *satellite bands*—hence the name).

Satellite α

- Tandem repeats di un'unità di 171 bp che lega la proteina centromerica CENP-B
- Costituisce la gran parte dell'eterocromatina centromerica di tutti i cromosomi
- Elevata divergenza di sequenza tra cromosomi diversi
- Clonato in cellule umane genera nuovi centromeri

Organizzazione del Dna satellite ai centromeri



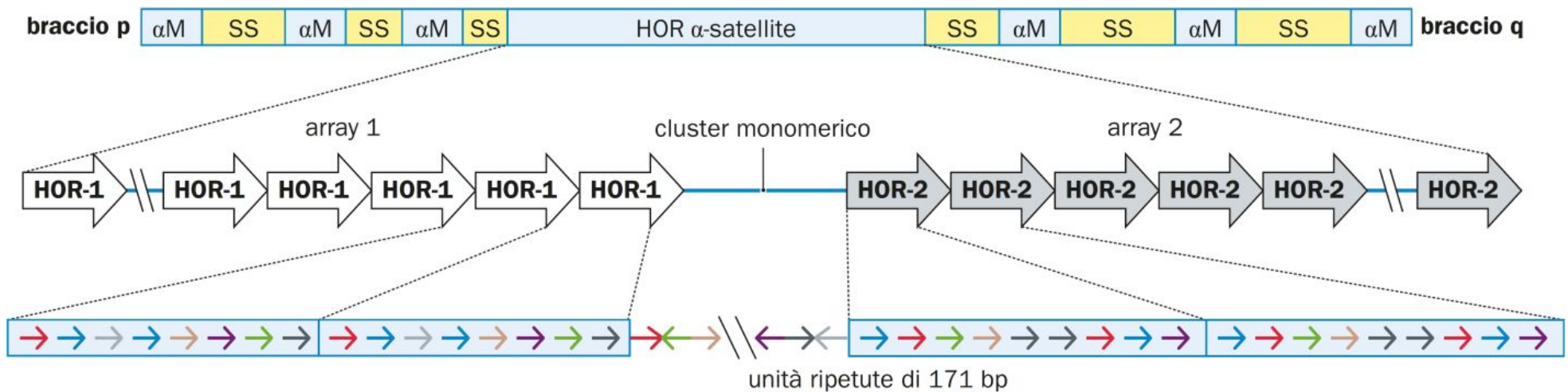
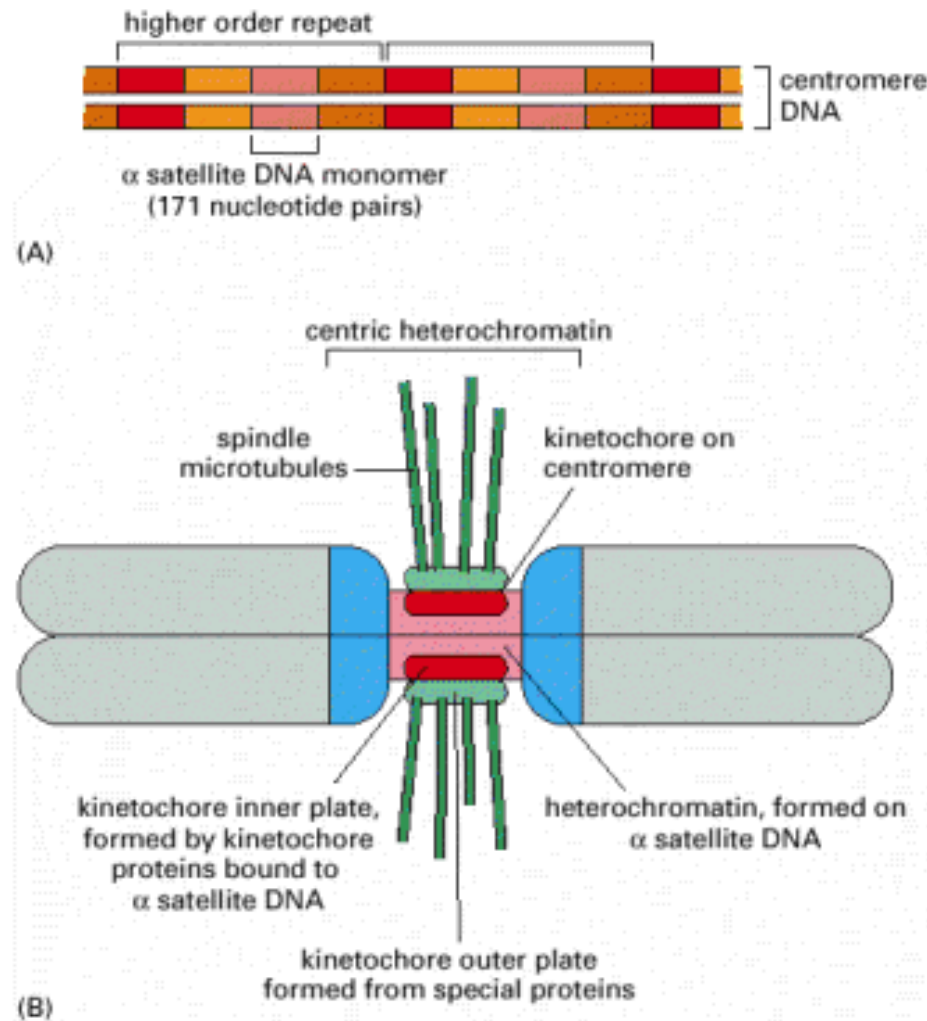
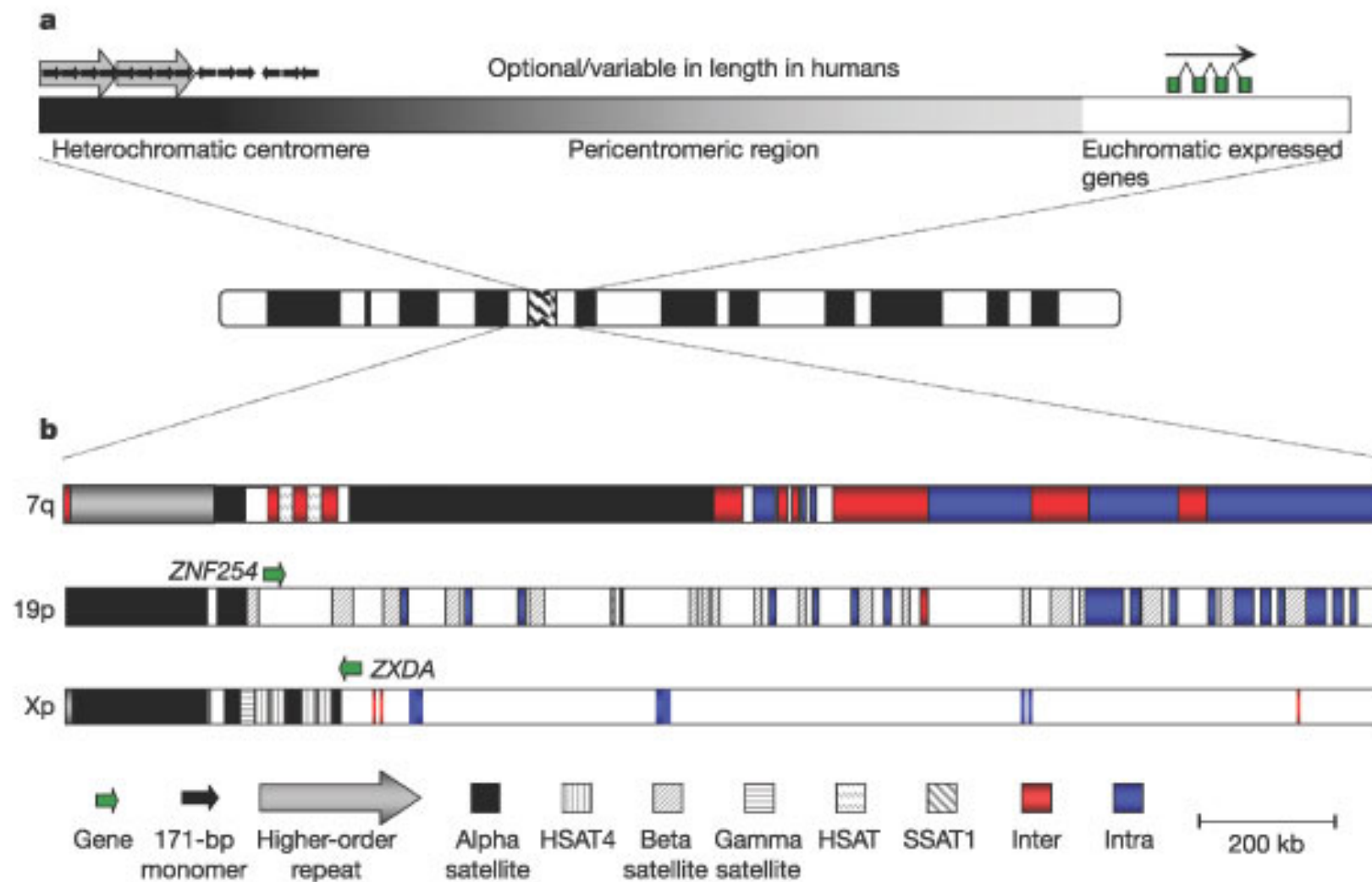


Figura 9.11 Il centromero umano e l'organizzazione del DNA α -satellite. Il DNA α -satellite (DNA alfoide) è una parte fondamentale del centromero di tutti i cromosomi umani. È composto di ripetizioni di una sequenza di 171 bp, ma tra le diverse ripetizioni di 171 bp, qui rappresentate dalle frecce diverse colorate in basso, si riscontrano differenze significative nella sequenza; i monomeri possono variare fino al 30-40% delle posizioni dei nucleotidi. Le ripetizioni di ordine superiore (HOR, *Higher-Order Repeats*) contrassegnano le regioni in cui una serie di ripetizioni di monomeri orientati in modo simile è ripetuta in tandem per formare lunghe serie multimeriche con livelli molto elevati d'identità di sequenza (97-100%) tra le diverse HOR

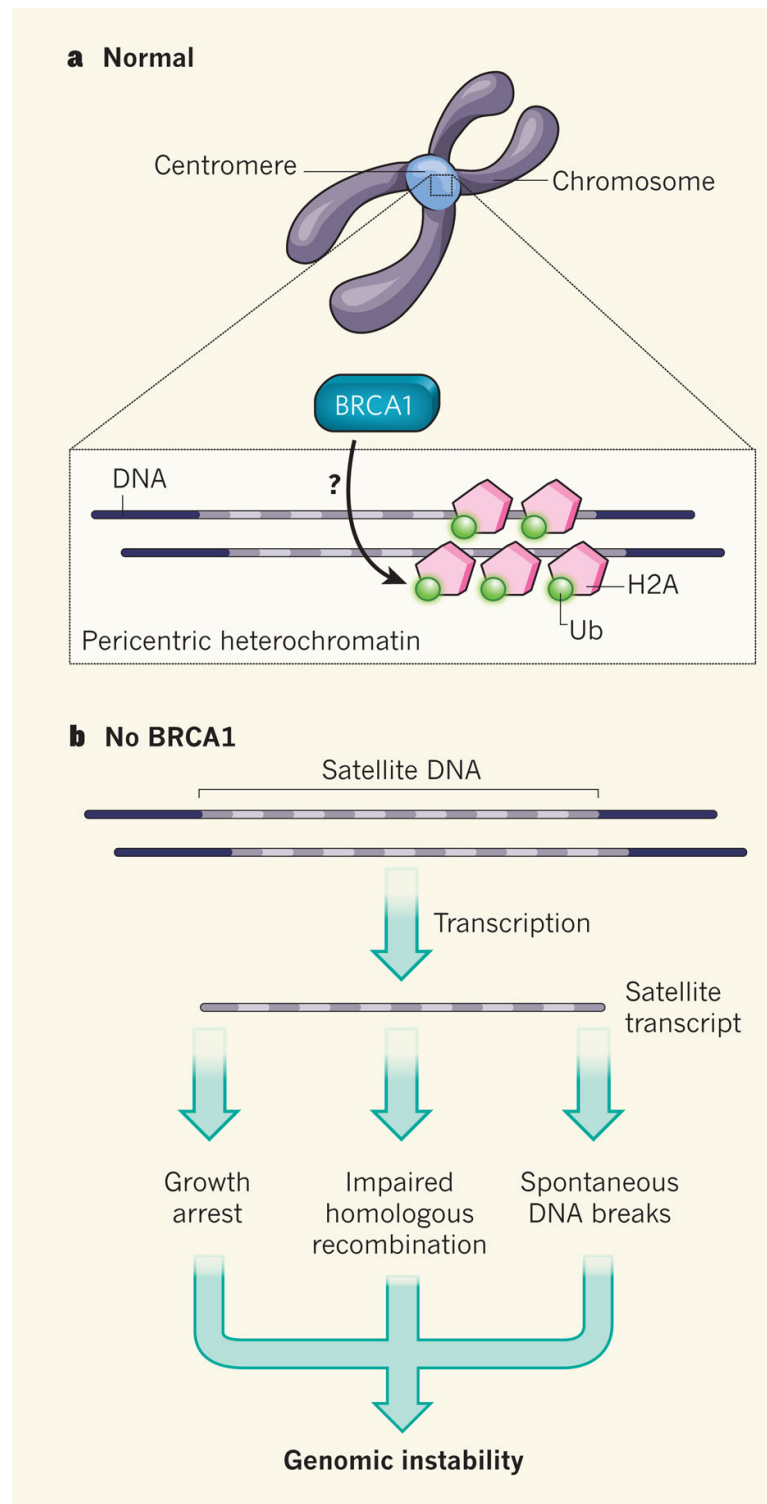
in una serie. In questo ipotetico esempio, immaginiamo due matrici composte rispettivamente da ripetizioni HOR-1, ciascuna composta da una sequenza di otto monomeri, e ripetizioni HOR-2, con una sequenza di nove monomeri. Gruppi di semplici monomeri che possono avere orientamenti diversi si trovano in genere nell'intervallo tra le matrici multimeriche. Al di fuori delle HOR α -satelliti, le sequenze dei centromeri spesso includono gruppi di monomeri α -satellite (αM) e DNA satellite a semplice sequenza (SS). Per un esempio specifico, il centromero del cromosoma umano 10, vedi l'articolo di Aldrup-Macdonald & Sullivan (2014) (PMID 24683489) riportato nei *Suggerimenti di lettura*.

DNA satellite α centromerico



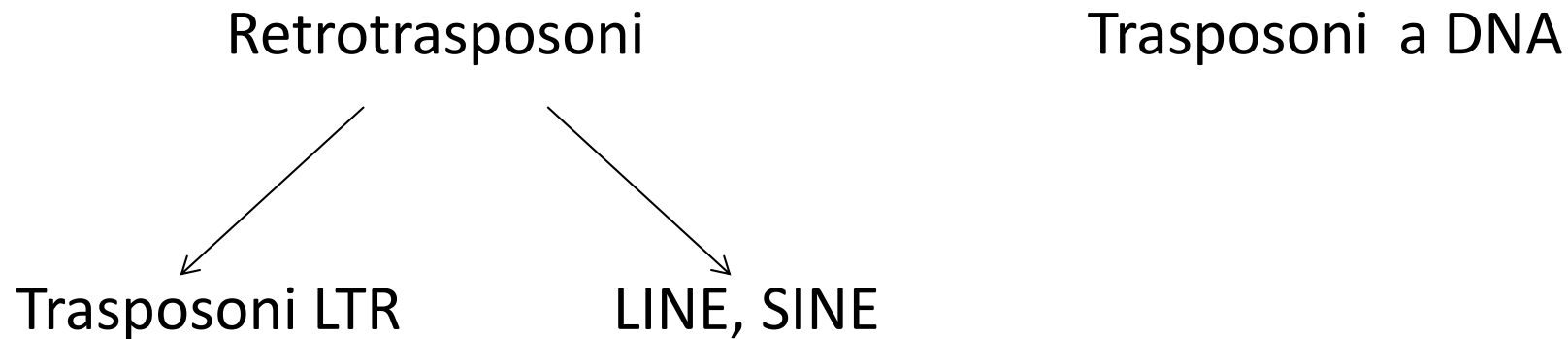


BRCA1 è una ubiquitina
ligasi dell'istone H2A.
La sua perdita di
funzione induce la
trascrizione
dell'eterocromatina
costitutiva centromerica

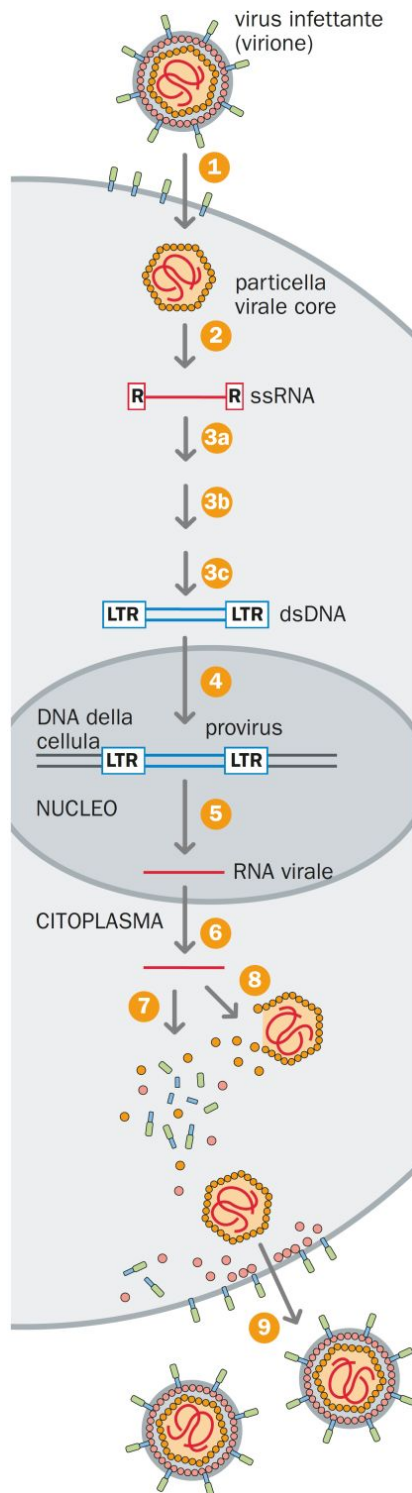


Ripetizioni derivate da trasposoni

Costituiscono quasi tutto il DNA ripetitivo non codificante non clusterizzato del genoma umano (45%)

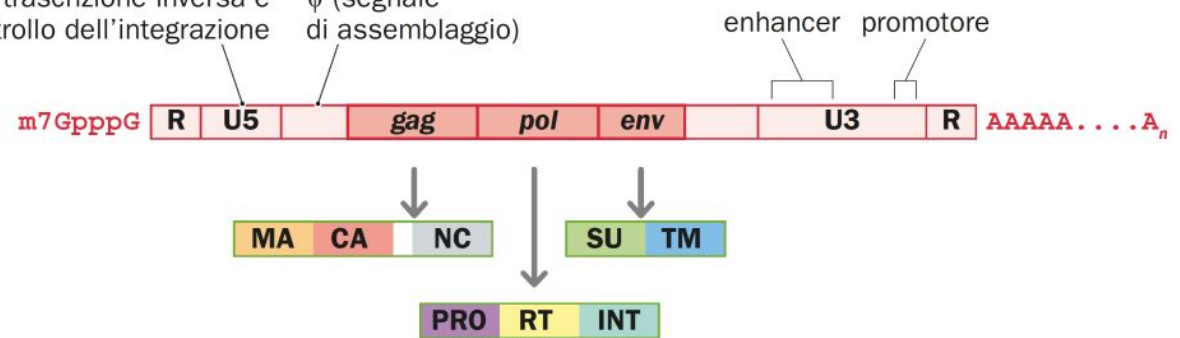


I retrovirus e il loro genoma



(A) GENOMA RETROVIRALE (ssRNA)

trascrizione inversa e controllo dell'integrazione ψ (segnale di assemblaggio)

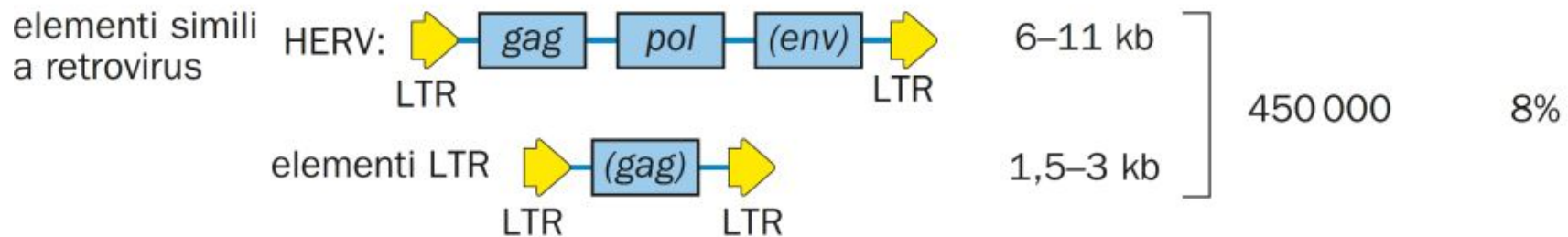


(B) GENOMA DEL PROVIRUS INTEGRATO (dsDNA)



Retrotrasposoni LTR

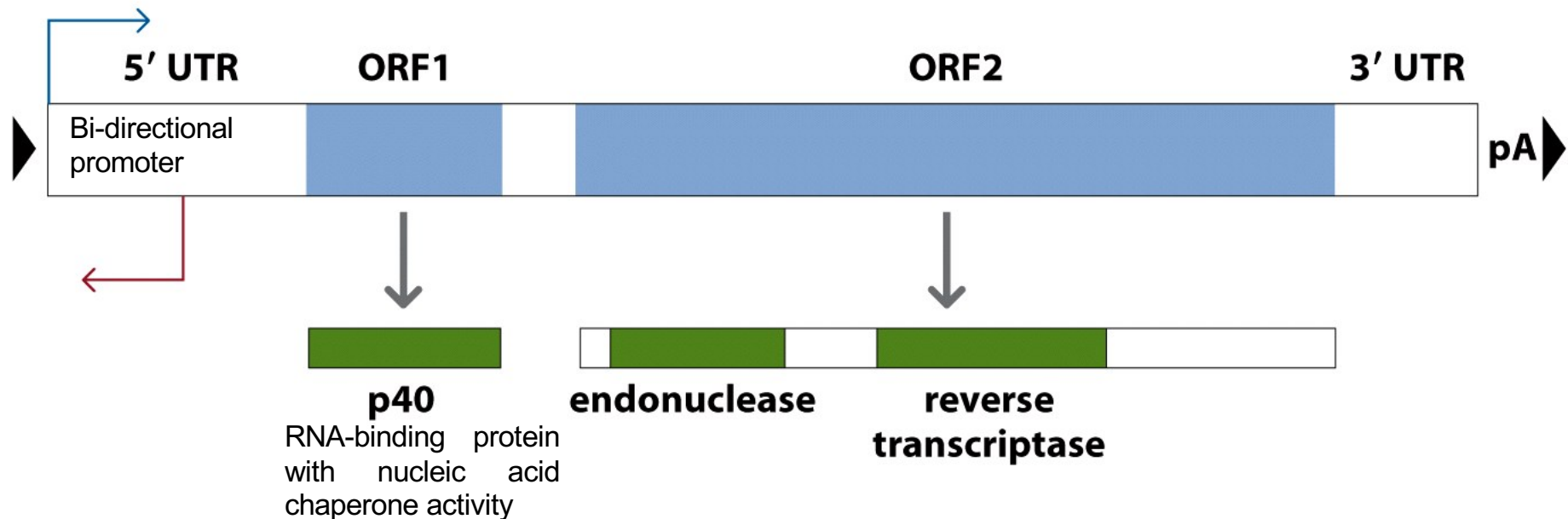
- Trasposoni retrovirus-like (Human Endogenous Retro Virus, HERV) circa 4.6% del genoma
- Hanno tipiche regioni LTR ripetute in tandem alle estremità
- Traspongono in maniera simile agli pseudogeni
- Autonomi e non-autonomi
- Trasposizione molto rara negli ultimi milioni di anni



Retrotrasposoni LINE

- Long Interspersed Nuclear Elements
- Sono i trasposoni più attivi e sono molto antichi (topo)
- Codificano tutti i prodotti necessari alla trasposizione
- 3 famiglie (L1, L2, L3) comprendenti il 20% del genoma umano
- L1 è la famiglia predominante (17%) ed è l'unica che traspone attivamente (circa 6.1 kb)
- L1 è responsabile di quasi tutta la retrotrascrizione che avviene nel genoma
- Localizzati in regioni AT-rich

(A) LINE-1 repeat element



- I trascritti LINE-1 sono tradotti nel citoplasma, si assemblano ai loro prodotti e tornano nel nucleo
- La nucleasi taglia un filamento del DNA in sequenze TTTT↓A lasciando un 3'OH che serve come innesco della RT
- La RT spesso non arriva al 5' dell'elemento
- Solo 1/100 copie di LINE è completa
- Di queste 6000, solo 100 sono in grado di trasporre

Trasposoni SINE

- Short Interspersed Nuclear Elements
- Sono lunghi 100-400bp e hanno colonizzato il genoma umano con successo
- Non codificano proteine
- Condividono le sequenze terminali con le LINE
- Traspongono parassitando il macchinario delle LINE
- Si sono originati per copia di geni ad RNA (7SL, tRNA...)
- La più abbondante è la famiglia Alu
- Localizzati in regioni GC-rich

transcriptase

(B) Alu repeat element

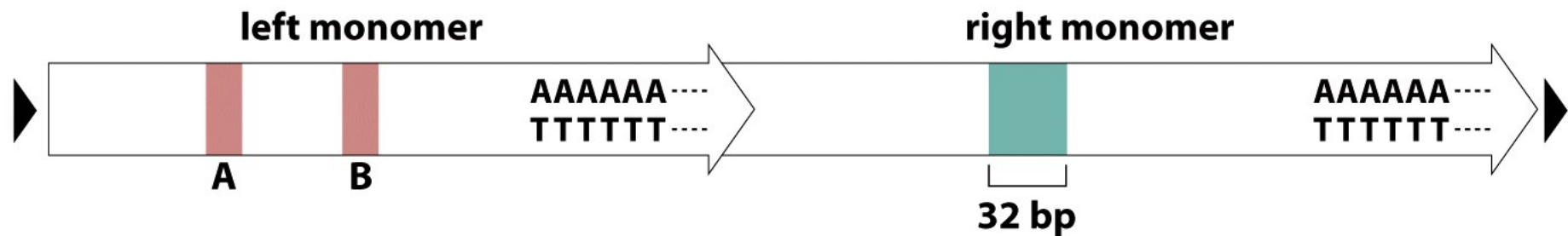


Figure 9.21 Human Molecular Genetics, 4ed. (© Garland Science)

- Gli elementi Alu sono le sequenze a maggior numero di copie del genoma umano (circa 1,5 milioni)
- L'unità completa è di circa 280 bp, costituita da ripetizioni in tandem di 120 bp
- Sono presenti anche come monomeri
- Sono relativamente ricche in GC e prevalentemente posizionate in regioni attive della cromatina

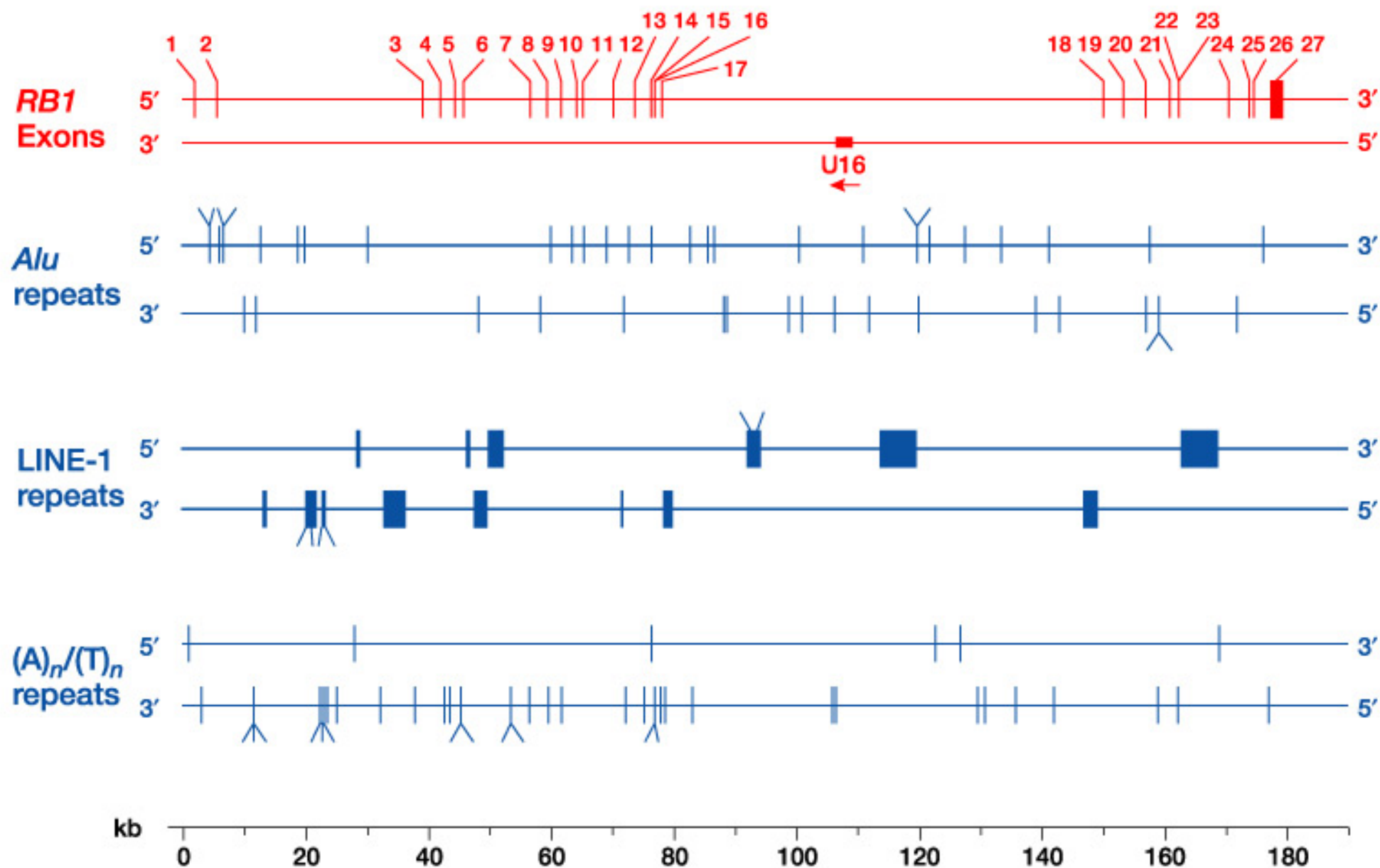
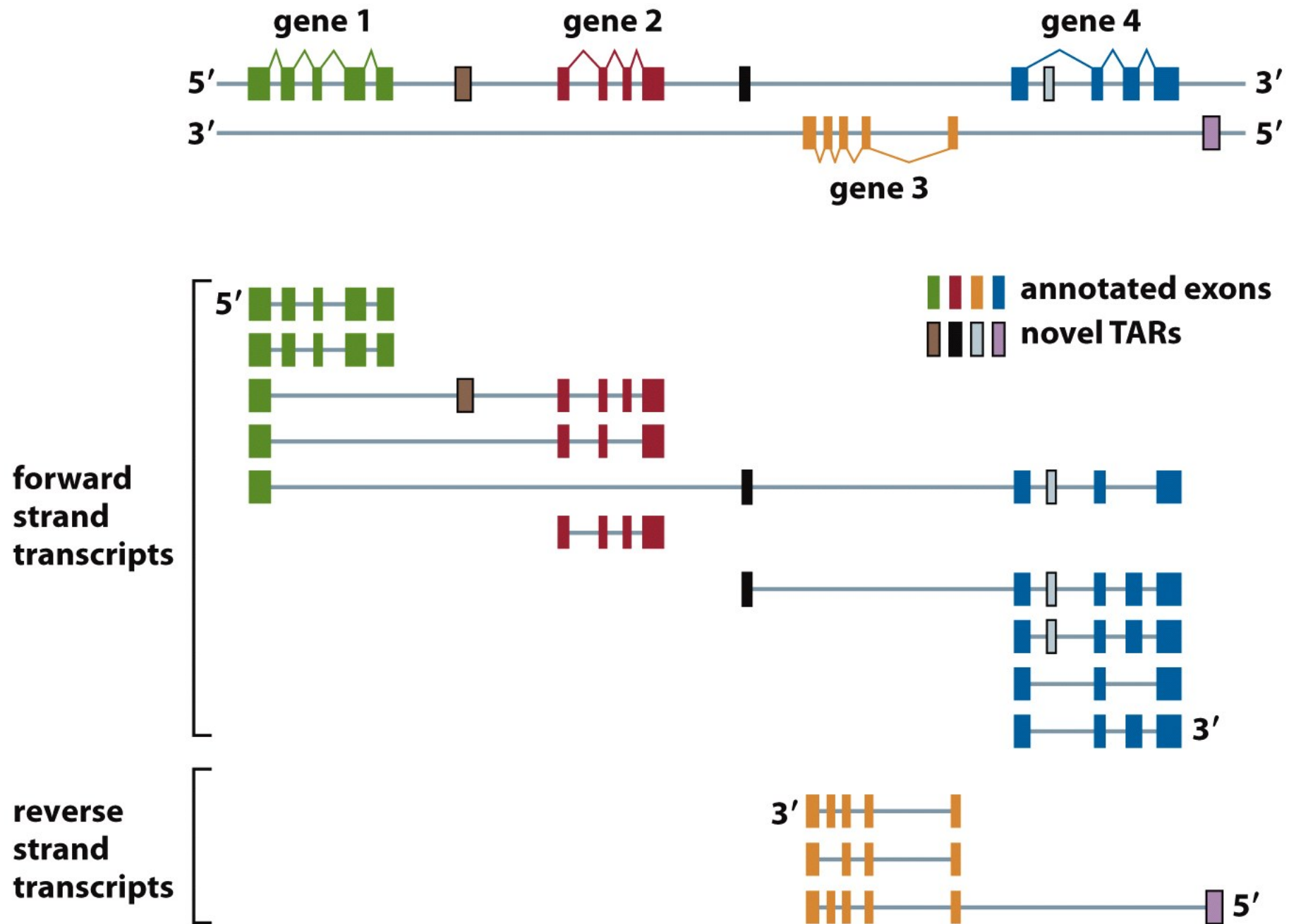


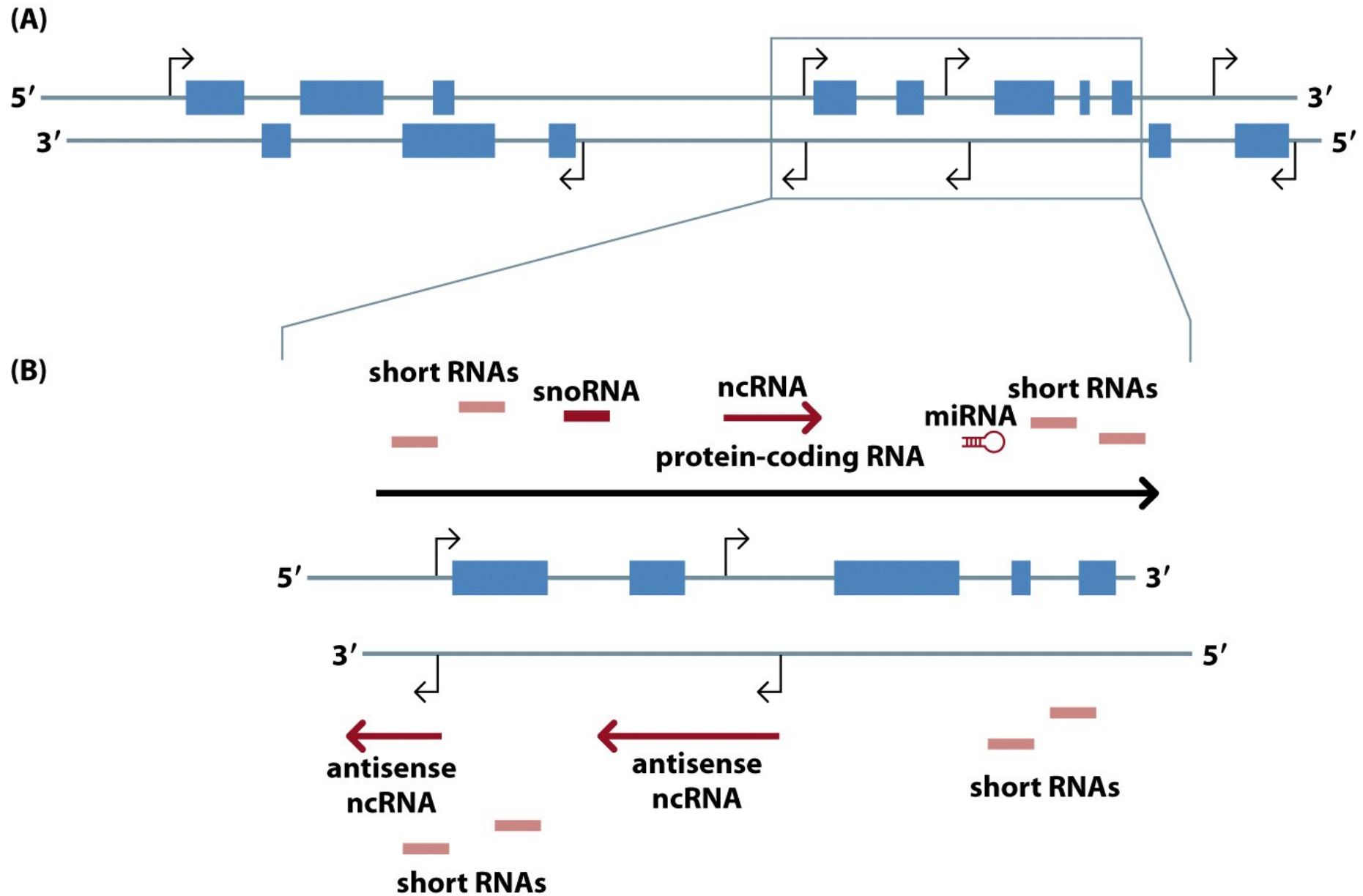
Figure 9-19 Human Molecular Genetics, 3/e. (© Garland Science 2004)

Ruolo dei trasposoni nell'evoluzione del genoma umano

- Fonte di nuove sequenze funzionali (regolatorie/esoni o geni)
- Hanno favorito la duplicazione di geni o esoni
- Possono alterare l'espressione dei geni vicini o ospiti (sequenze regolative, nuovi siti di splicing)
- I retrotrasposoni traspongono attivamente durante la neurogenesi, generando un alto livello di variabilità genetica che può promuovere la diversità dei neuroni e la plasticità delle reti neuronali

I geni delle proteine coprono 1.1% del genoma, ma le analisi genomiche hanno rivelato che almeno l'85% del genoma umano viene trascritto





I geni degli RNA

- Numero in continua crescita
- Principalmente coinvolti nei processi di espressione genica (rRNA e tRNA)
- Altre classi di RNA con ruoli regolatori

Un mondo a RNA (1960)

l'RNA avrebbe avuto un ruolo duplice nelle fasi più precoci della vita, fungendo da materiale genetico (con capacità di autoreplicazione) e anche da molecola effettrice.

- Materiale genetico (retrovirus)
- Molecole effettrici (Ribozimi, RNAsi P, autosplicing)

Un'altra evidenza che gli RNA siano venuti prima del DNA è che nella cellula i deossiribonucleotidi sono sintetizzati da ribonucleotidi.

Oltre a portare informazione genetica, gli RNA in seguito sarebbero stati usati per sintetizzare le proteine da aminoacidi: diversi RNA sono essenziali nella sintesi proteica, è possibile eliminare molte proteine ribosomali senza distruggere la funzione dei ribosomi e la peptidil trasferasi è un ribozima

Infine sarebbe stato rimpiazzato dal DNA che è molto più stabile, anche per la convenienza di separare funzione di conservazione e trasmissione dalla sintesi proteica. Naturalmente ciò sarebbe stato possibile grazie ad una trascrittasi inversa.

I geni degli RNA non codificanti

Funzioni degli RNA

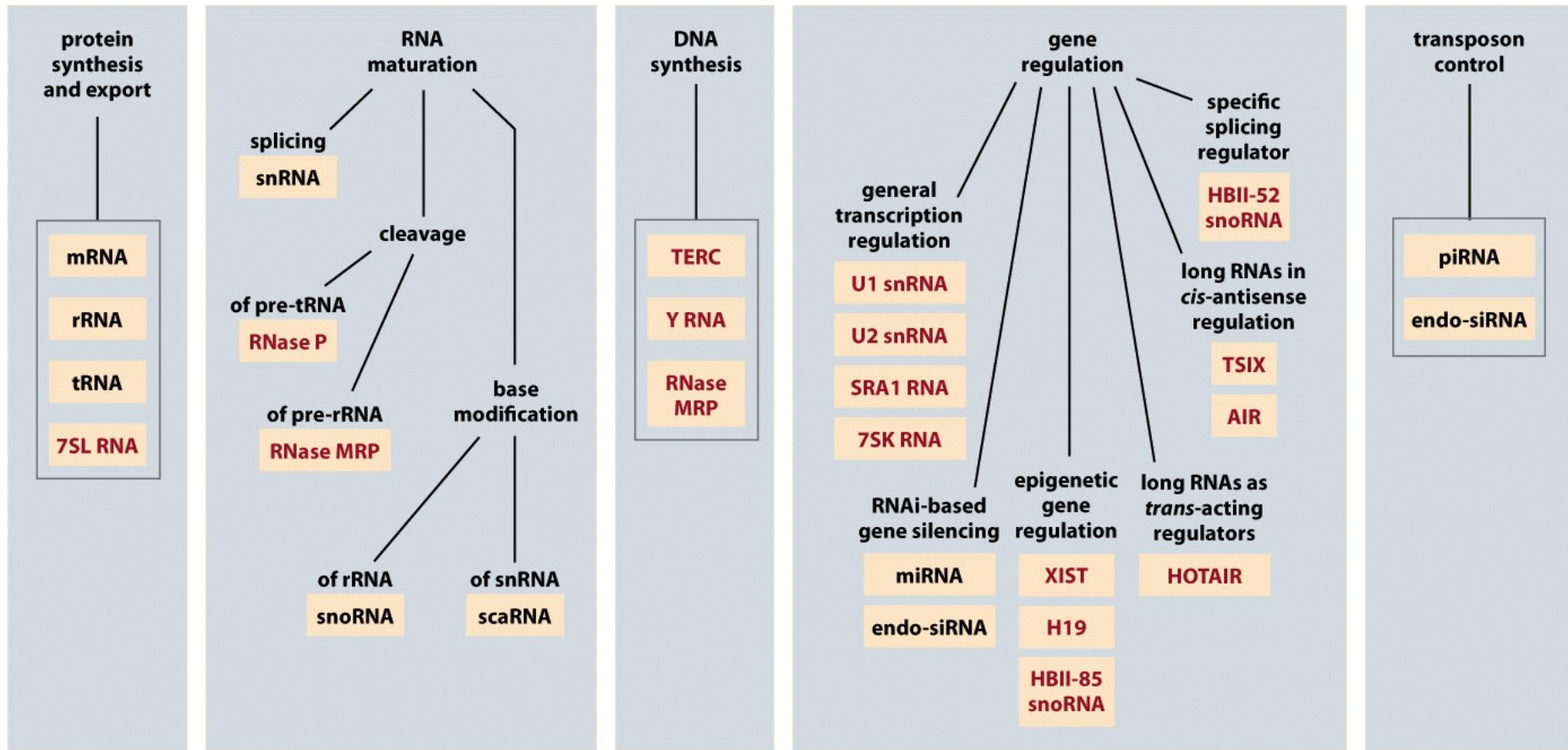


Figure 9.13 Human Molecular Genetics, 4ed. (© Garland Science)

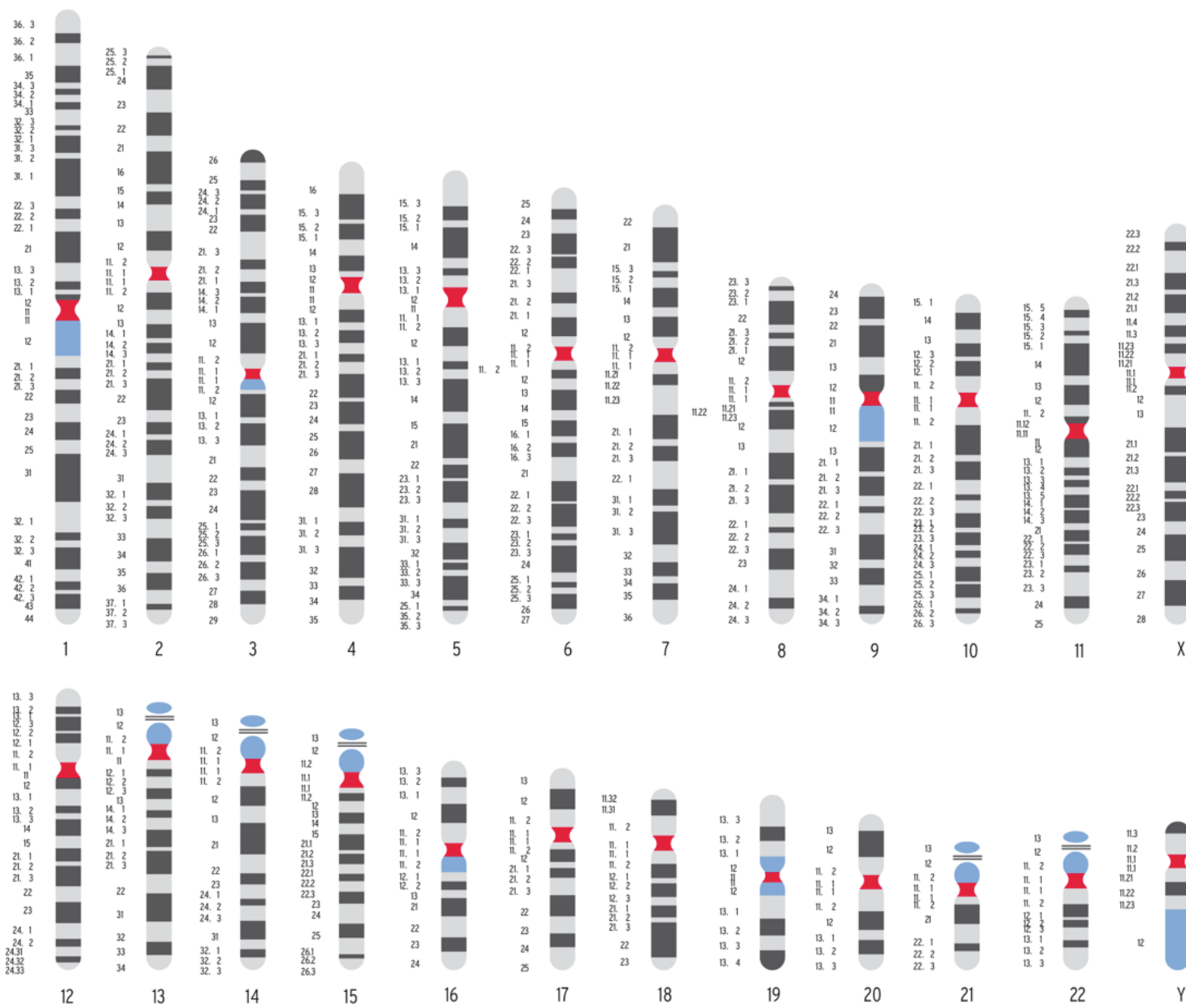
RNA families

Individual RNA

rRNA e tRNA sono codificati da oltre 1000 geni, principalmente clusterizzati

rRNA

- 4 tipi: 28S, 5.8S e 5S associati alla subunità maggiore e 18S alla minore
- Il 5S è prodotto da piccoli cluster genici sul cromosoma 1q42 (molti pseudogeni)
- 28S, 18S e 5.8S sono prodotti da una sola unità trascrizionale multigenica
- Tandem repeat di 30-40 copie nel braccio p dei cromosomi acrocentrici 13, 14, 15, 21 e 22
- Esclusi dal progetto genoma



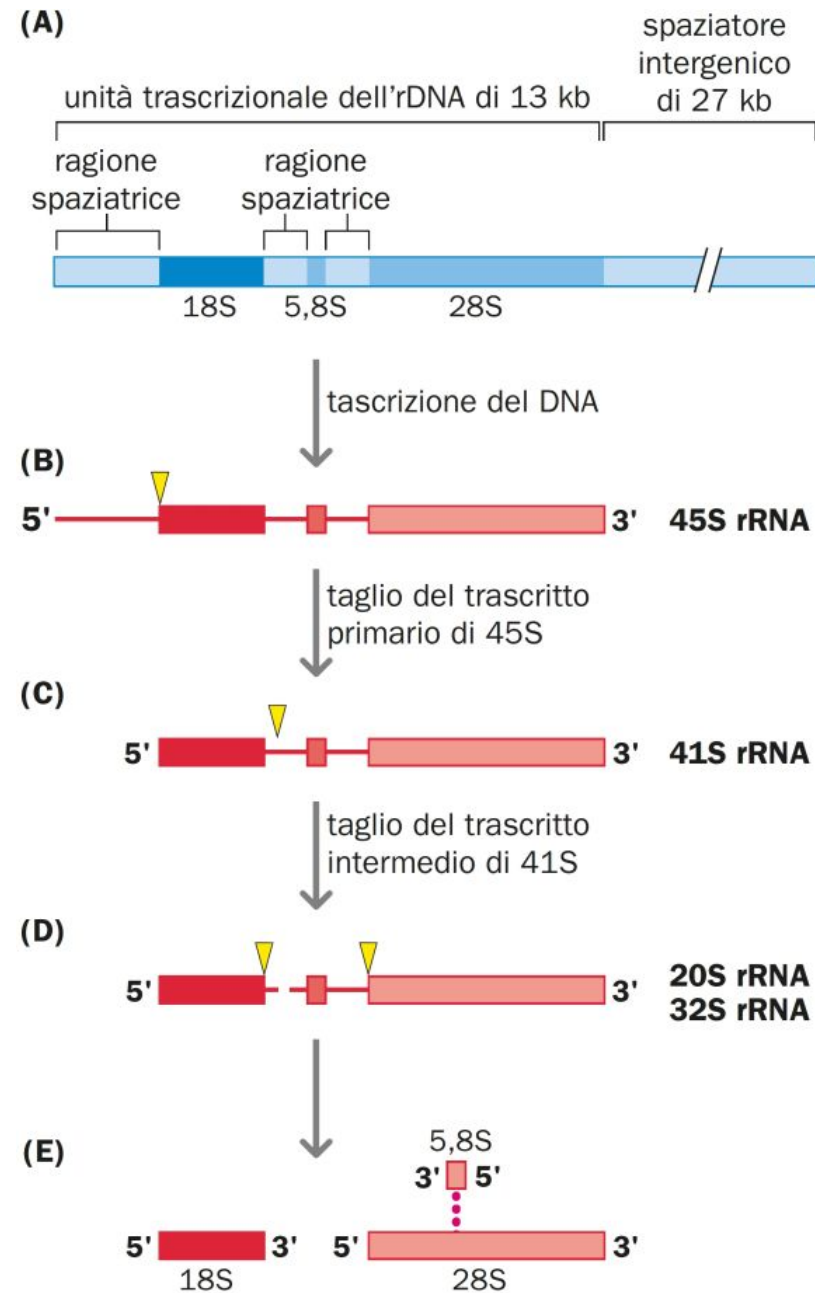
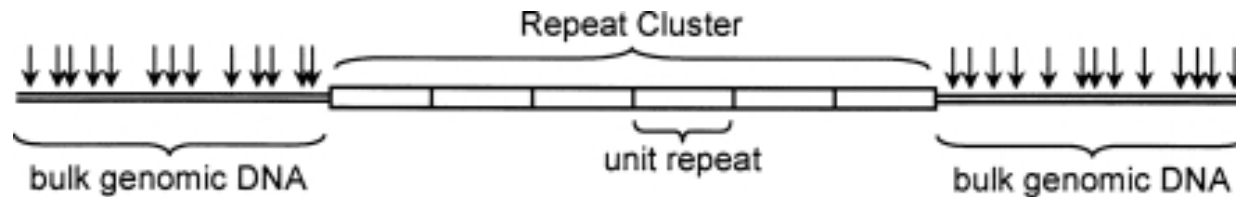


Figura 1.24 I tre principali tipi di rRNA sono prodotti dal taglio di un trascritto primario condiviso. **(A)** Nelle

I cluster dei geni per rRNA sono molto variabili

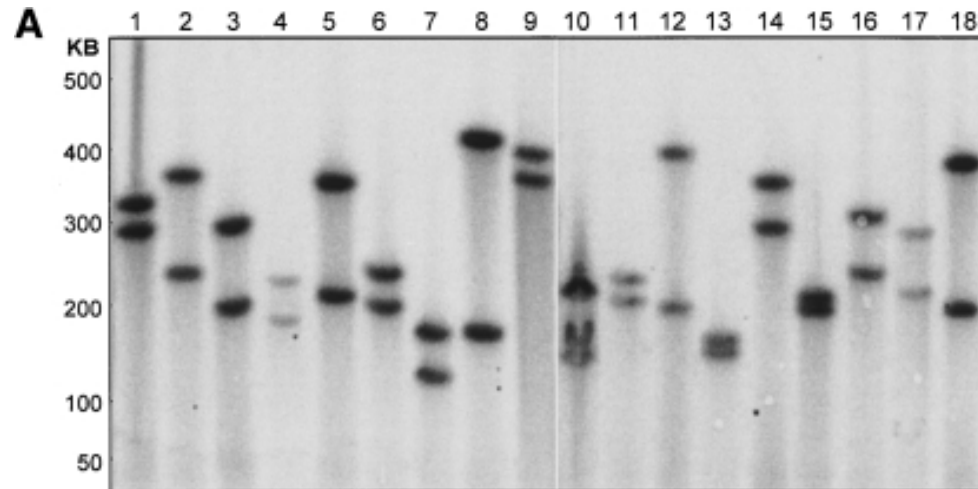


Digestion by a restriction enzyme lacking a recognition site in a single rRNA gene to leave intact the clusters

Pulse field gel electrophoresis to separate the entire clusters

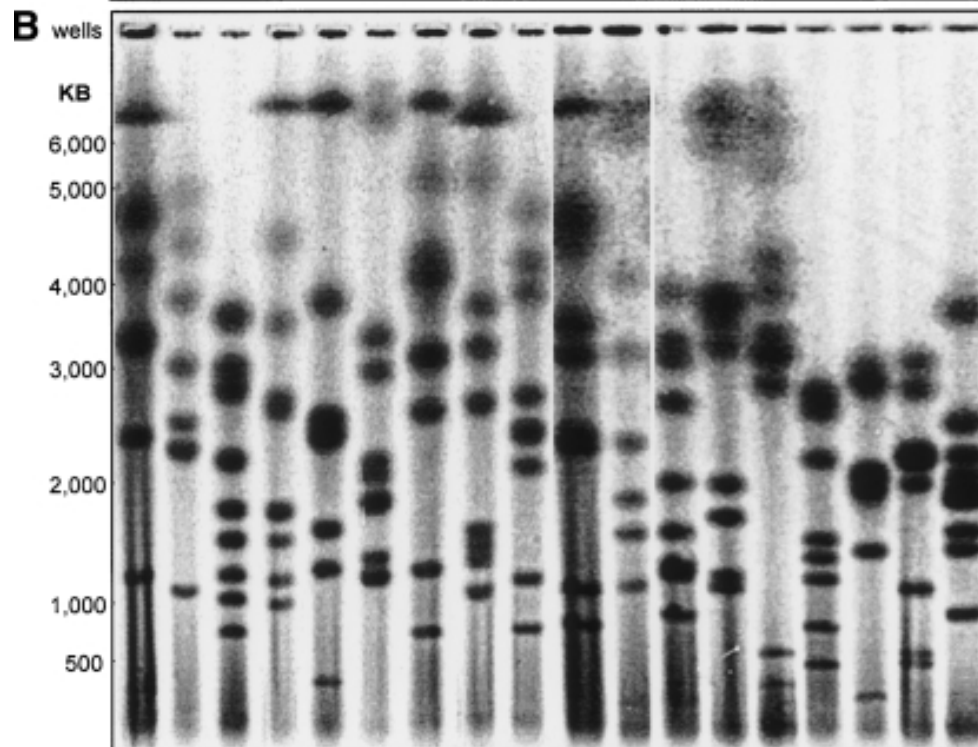
Hybridization with a probe for 5S or for 45S

Variabilità della lunghezza dei cluster degli rRNA



18 healthy individuals
were analyzed

A: 5S probe



B: 45S probe

tRNA

- Oltre 500 geni appartenenti a 49 famiglie (anticodoni)
- Poca correlazione tra numero di geni e frequenza dell'aa
- Dispersi nel genoma, ma più della metà raggruppati sui cromosomi 6 e 1

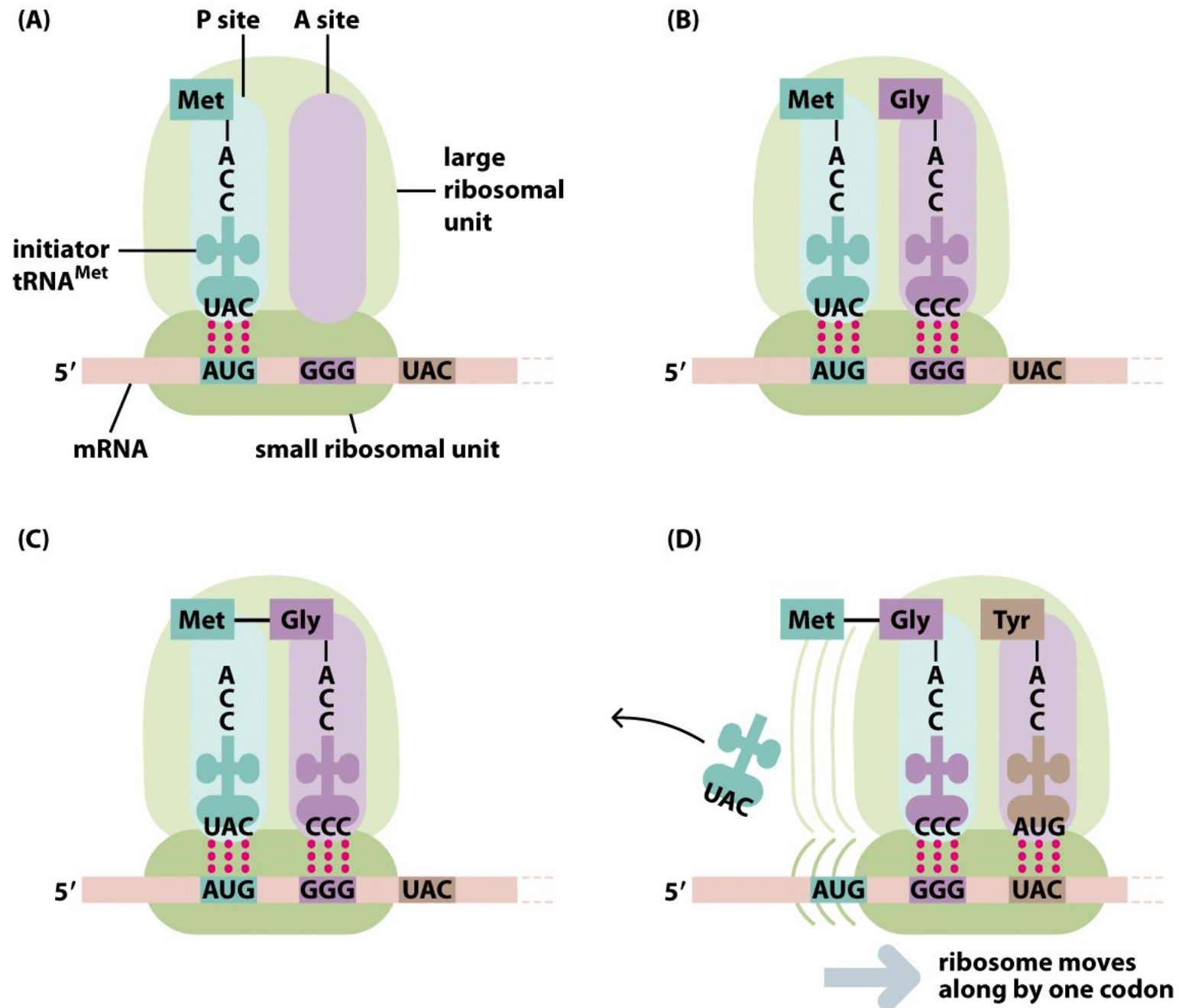


Figure 1.24a-d Human Molecular Genetics, 4ed. (© Garland Science)

Human tRNA gene families

Phe	<table><tr><td>UUU</td><td>AAA</td><td>–</td></tr><tr><td>UUC</td><td>GAA</td><td>12</td></tr></table>	UUU	AAA	–	UUC	GAA	12
UUU	AAA	–					
UUC	GAA	12					
Leu	<table><tr><td>UUA</td><td>UAA</td><td>7</td></tr><tr><td>UUG</td><td>CAA</td><td>7</td></tr></table>	UUA	UAA	7	UUG	CAA	7
UUA	UAA	7					
UUG	CAA	7					

Ser	<table><tr><td>UCU</td><td>AGA</td><td>11</td></tr><tr><td>UCC</td><td>GGA</td><td>–</td></tr><tr><td>UCA</td><td>UGA</td><td>5</td></tr><tr><td>UCG</td><td>CGA</td><td>4</td></tr></table>	UCU	AGA	11	UCC	GGA	–	UCA	UGA	5	UCG	CGA	4
UCU	AGA	11											
UCC	GGA	–											
UCA	UGA	5											
UCG	CGA	4											

Tyr	<table><tr><td>UAU</td><td>AUA</td><td>1</td></tr><tr><td>UAC</td><td>GUA</td><td>14</td></tr><tr><td>UAA</td><td>UUA</td><td>–</td></tr><tr><td>UAG</td><td>CUA</td><td>–</td></tr></table>	UAU	AUA	1	UAC	GUA	14	UAA	UUA	–	UAG	CUA	–
UAU	AUA	1											
UAC	GUA	14											
UAA	UUA	–											
UAG	CUA	–											

Cys	<table><tr><td>UGU</td><td>ACA</td><td>–</td></tr><tr><td>UGC</td><td>GCA</td><td>30</td></tr><tr><td>UGA</td><td>UCA</td><td>– (3) *</td></tr><tr><td>UGG</td><td>CCA</td><td>9</td></tr></table>	UGU	ACA	–	UGC	GCA	30	UGA	UCA	– (3) *	UGG	CCA	9
UGU	ACA	–											
UGC	GCA	30											
UGA	UCA	– (3) *											
UGG	CCA	9											

Leu	<table><tr><td>CUU</td><td>AAG</td><td>12</td></tr><tr><td>CUC</td><td>GAG</td><td>–</td></tr><tr><td>CUA</td><td>UAG</td><td>3</td></tr><tr><td>CUG</td><td>CAG</td><td>10</td></tr></table>	CUU	AAG	12	CUC	GAG	–	CUA	UAG	3	CUG	CAG	10
CUU	AAG	12											
CUC	GAG	–											
CUA	UAG	3											
CUG	CAG	10											

Pro	<table><tr><td>CCU</td><td>AGG</td><td>10</td></tr><tr><td>CCC</td><td>GGG</td><td>–</td></tr><tr><td>CCA</td><td>UGG</td><td>7</td></tr><tr><td>CCG</td><td>CGG</td><td>4</td></tr></table>	CCU	AGG	10	CCC	GGG	–	CCA	UGG	7	CCG	CGG	4
CCU	AGG	10											
CCC	GGG	–											
CCA	UGG	7											
CCG	CGG	4											

His	<table><tr><td>CAU</td><td>AUG</td><td>–</td></tr><tr><td>CAC</td><td>GUG</td><td>11</td></tr></table>	CAU	AUG	–	CAC	GUG	11
CAU	AUG	–					
CAC	GUG	11					
Gln	<table><tr><td>CAA</td><td>UUG</td><td>11</td></tr><tr><td>CAG</td><td>CUG</td><td>20</td></tr></table>	CAA	UUG	11	CAG	CUG	20
CAA	UUG	11					
CAG	CUG	20					

Arg	<table><tr><td>CGU</td><td>ACG</td><td>7</td></tr><tr><td>CGC</td><td>GCG</td><td>–</td></tr><tr><td>CGA</td><td>UCG</td><td>6</td></tr><tr><td>CGG</td><td>CCG</td><td>4</td></tr></table>	CGU	ACG	7	CGC	GCG	–	CGA	UCG	6	CGG	CCG	4
CGU	ACG	7											
CGC	GCG	–											
CGA	UCG	6											
CGG	CCG	4											

Ile	<table><tr><td>AUU</td><td>AAU</td><td>14</td></tr><tr><td>AUC</td><td>GAU</td><td>3</td></tr><tr><td>AUA</td><td>UAU</td><td>5</td></tr></table>	AUU	AAU	14	AUC	GAU	3	AUA	UAU	5
AUU	AAU	14								
AUC	GAU	3								
AUA	UAU	5								
Met	<table><tr><td>AUG</td><td>CAU</td><td>20</td></tr></table>	AUG	CAU	20						
AUG	CAU	20								

Thr	<table><tr><td>ACU</td><td>AGU</td><td>10</td></tr><tr><td>ACC</td><td>GGU</td><td>–</td></tr><tr><td>ACA</td><td>UGU</td><td>6</td></tr><tr><td>ACG</td><td>CGU</td><td>6</td></tr></table>	ACU	AGU	10	ACC	GGU	–	ACA	UGU	6	ACG	CGU	6
ACU	AGU	10											
ACC	GGU	–											
ACA	UGU	6											
ACG	CGU	6											

Asn	<table><tr><td>AAU</td><td>AUU</td><td>2</td></tr><tr><td>AAC</td><td>GUU</td><td>32</td></tr><tr><td>AAA</td><td>UUU</td><td>16</td></tr><tr><td>AAG</td><td>CUU</td><td>17</td></tr></table>	AAU	AUU	2	AAC	GUU	32	AAA	UUU	16	AAG	CUU	17
AAU	AUU	2											
AAC	GUU	32											
AAA	UUU	16											
AAG	CUU	17											

Lys	<table><tr><td>AAA</td><td>UUU</td><td>16</td></tr><tr><td>AAG</td><td>CUU</td><td>17</td></tr></table>	AAA	UUU	16	AAG	CUU	17
AAA	UUU	16					
AAG	CUU	17					

Ser	<table><tr><td>AGU</td><td>ACU</td><td>–</td></tr><tr><td>AGC</td><td>GCU</td><td>8</td></tr></table>	AGU	ACU	–	AGC	GCU	8
AGU	ACU	–					
AGC	GCU	8					
Arg	<table><tr><td>AGA</td><td>UCU</td><td>6</td></tr><tr><td>AGG</td><td>CCU</td><td>5</td></tr></table>	AGA	UCU	6	AGG	CCU	5
AGA	UCU	6					
AGG	CCU	5					

Val	<table><tr><td>GUU</td><td>AAC</td><td>11</td></tr><tr><td>GUC</td><td>GAC</td><td>–</td></tr><tr><td>GUA</td><td>UAC</td><td>5</td></tr><tr><td>GUG</td><td>CAC</td><td>16</td></tr></table>	GUU	AAC	11	GUC	GAC	–	GUA	UAC	5	GUG	CAC	16
GUU	AAC	11											
GUC	GAC	–											
GUA	UAC	5											
GUG	CAC	16											

Ala	<table><tr><td>GCU</td><td>AGC</td><td>29</td></tr><tr><td>GCC</td><td>GGC</td><td>–</td></tr><tr><td>GCA</td><td>UGC</td><td>9</td></tr><tr><td>GCG</td><td>CGC</td><td>5</td></tr></table>	GCU	AGC	29	GCC	GGC	–	GCA	UGC	9	GCG	CGC	5
GCU	AGC	29											
GCC	GGC	–											
GCA	UGC	9											
GCG	CGC	5											

Asp	<table><tr><td>GAU</td><td>AUC</td><td>–</td></tr><tr><td>GAC</td><td>GUC</td><td>19</td></tr></table>	GAU	AUC	–	GAC	GUC	19
GAU	AUC	–					
GAC	GUC	19					
Glu	<table><tr><td>GAA</td><td>UUC</td><td>13</td></tr><tr><td>GAG</td><td>CUC</td><td>13</td></tr></table>	GAA	UUC	13	GAG	CUC	13
GAA	UUC	13					
GAG	CUC	13					

Gly	<table><tr><td>GGU</td><td>ACC</td><td>–</td></tr><tr><td>GGC</td><td>GCC</td><td>15</td></tr><tr><td>GGA</td><td>UCC</td><td>9</td></tr><tr><td>GGG</td><td>CCC</td><td>7</td></tr></table>	GGU	ACC	–	GGC	GCC	15	GGA	UCC	9	GGG	CCC	7
GGU	ACC	–											
GGC	GCC	15											
GGA	UCC	9											
GGG	CCC	7											

Il codice genetico è universale ma....

- Ogni specie utilizza preferenzialmente alcuni codoni sinonimi (**CODON USAGE**)
- Ogni specie ha un set di tRNA (anticodoni) diverso
- L'utilizzo dei codoni sinonimi è correlato con l'abbondanza dei tRNA corrispondenti

Piccoli RNA nucleari (60-360 bp)

- Varie famiglie che assistono l'espressione genica generale nel nucleo
- Funzionano in RNPs
- Alcuni sono molto ricchi in uridina (U1, U2, ecc.)
- 3 categorie: snRNA dello spliceosoma o con altre funzioni, snoRNAs localizzati nel nucleolo, scaRNA
- Codificati da centinaia di geni per lo più dispersi, oltre a centinaia di pseudogeni
- snoRNA e scaRNA sono localizzati in introni di altri geni