

Genomet og genekspression

STATUSARTIKEL

Finn Cilius Nielsen

Det humane genom er sammensat af 3 mia. basepar fordelt på 2×23 par kromosomer. Den færdige genomsekvens tydede i første omgang på, at vores genom indeholdt ca. 35.000 gener, men sammenligning med musegenomet har reduceret antallet til ca. 26.000, da mange gener har vist sig at være pseudogener. Det relativt beskedne antal gener hos mennesker sammenlignet med antallet hos orme og fluer, hvis genomer rummer hhv. 19.099 og 13.061 gener, er en af de store overraskelser ved den humane genomsekvens. Men en nærmere gennemgang har vist, at gener hos mennesker udviser en større grad af f.eks. alternativ splejsning end hos de øvrige sekventerede organismer. Alt tyder således på, at det humane proteom – trods alt – er mere komplekst end de øvrige organismers. De fleste gener i det humane genom er evolutionært set meget gamle. Faktisk tyder alt på, at mindre end 10% af proteinfamilierne i vores genom er enestående for hvirveldyr. De mest elementære cellulære funktioner er således direkte overtaget fra gærceller og bakterier, og man har blot bygget et nyt hus med gamle sten (1, 2). Når alt kommer til alt er størrelsen måske ikke altafgørende, og i denne artikel vil vi fokusere på generne og deres ekspression. Selv om genekspression ofte udtrykkes som en lineær proces, er der langtfra proportionalitet mellem

transkriptionen af et gen og mængden af det færdige bioaktive protein (**Fig. 1**). Hvert enkelt led i ekspressionskaskaden er reguleret, og mange af mekanismerne er centrale for at forstå cellens funktion og konsekvensen af sygdomsfremkaldende mutationer.

Generne

Et proteinkodende gen defineres som den nukleinsyresekvens, der er nødvendig for syntesen af et givent protein. Et gen omfatter derfor mere information end den, der specificerer aminosyresekvensen. Den resterende gensekvens regulerer syntesen og processeringen af RNA. Genet er opbygget af en promotorregion, der regulerer omsætningen af DNA-sekvensen til RNA (transkriptionen), samt sekvensen af exoner og introner, der tilsammen udgør det primære RNA-transkript. Exoner defineres som de områder af genet, der findes i det modne mRNA. Exoner indeholder ikke kun proteinkodende sekvenser, da de såkaldte 5' og 3' untranslatede regioner, som er involveret i kontrollen af mRNA'ets translation, stabilitet og transport, også er en del af det færdige mRNA (**Fig. 2**). Promotorregionen ligger lige opstrøms for transkriptionsinitieringsstedet og er ca. 200-500 basepar lang. Området indeholder bindingssteder for specifikke transkriptionsfaktorer og det basale transkriptionsapparat, som kontrollerer bindingen af RNA-polymerasen, der oversætter DNA til RNA. Endvidere medvirker locuskontrolregioner (LCR) og såkaldte enhancere til at regulere

Fig. 1. Det centrale dogme angiver, at informationsstrømmen går fra DNA via mRNA til protein og ikke omvendt. Der er ikke altid proportionalitet mellem et gens transkription og mængden af bioaktivt protein. Genekspressionen reguleres på alle trin – fra processeringen af det primære transkript og eksporten af mRNA til dets translation og lokalisering. Endelig har den posttranslationelle processing og kovalente modifikation af de nysyntetiserede proteiner stor betydning for deres bioaktivitet. Fra (2).

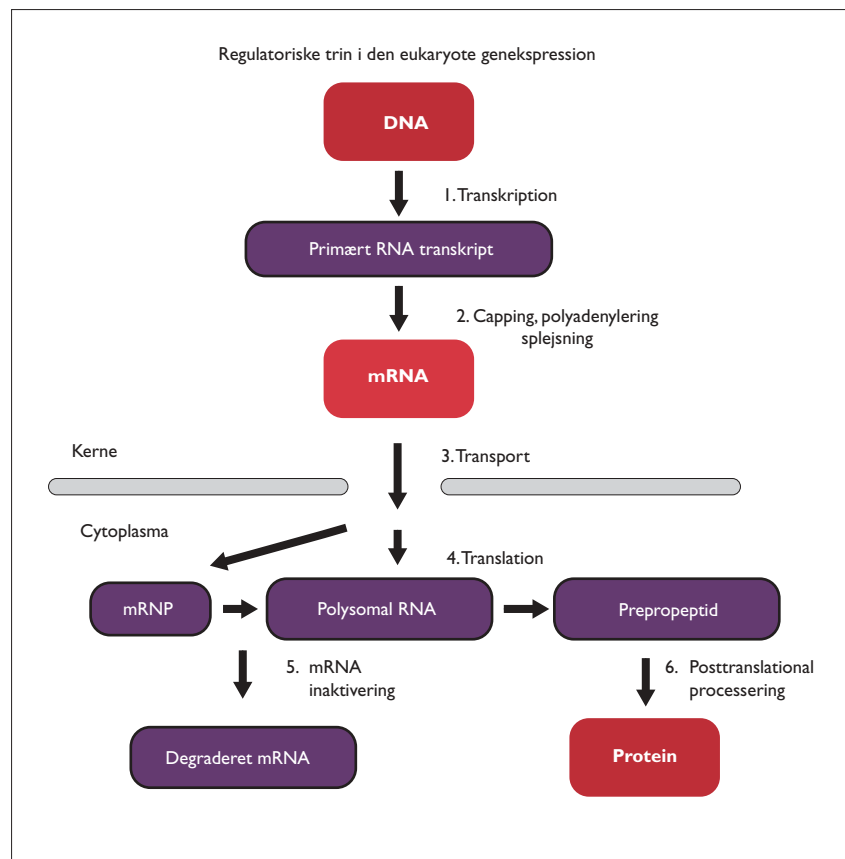
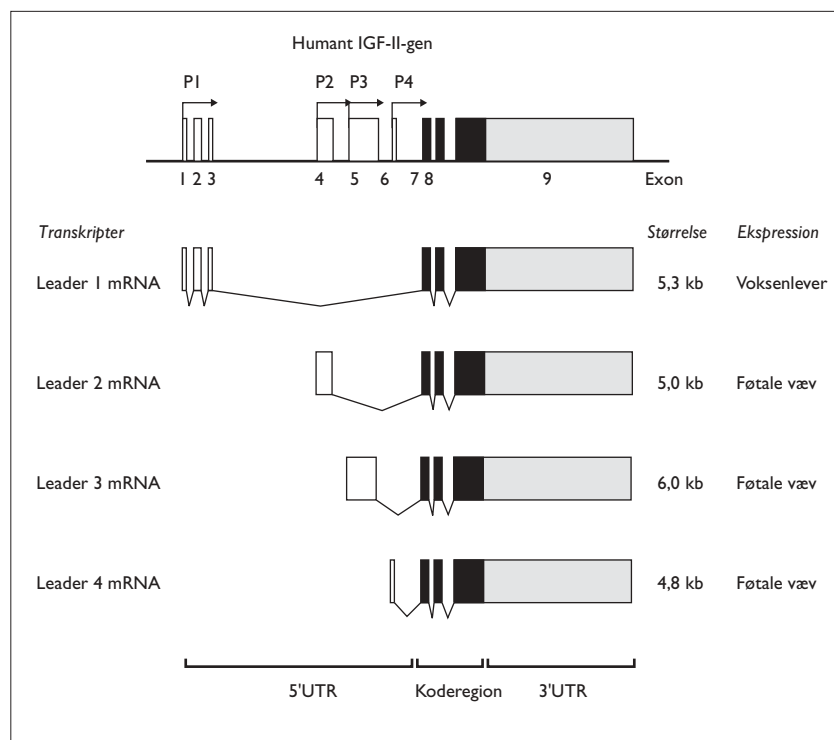


Fig. 2. Strukturen af gener hos mennesker. Eukaryote gener eller transkriptionsenheder er sammensat af et antal exoner og introner samt en eller flere promotorer, der er beliggende lige opstrøms for det første exon. Desuden findes ofte en enhancer-region længere væk fra genet. Det humane IGF-II-gen består f.eks. af ni exoner, og ved hjælp af fire forskellige promotorer genereres fire mRNA med forskellige 5' utranslaterede regioner (UTR), men samme koderegion (KR) og 3' UTR. Promotorerne er vævsspecifikke. Bemærk at de største exoner ikke er proteinkodende, men udgør de utranslaterede regioner af mRNA. Fra (2).



promotoraktiviteten. Lokuskontrolregioner og enhancere er mindre veldefinerede end promotorregionerne og kan ligge flere tusinde basepar væk fra selve genet.

Den genetiske kode

DNA makes RNA makes Protein udtrykker strømmen af genetisk information i levende organismer. Den genetiske information, der er lagret i DNA overføres via mRNA til proteinet under proteinsyntesen. Mekanismen for overførslen af den genetiske information blev afdækket i 1961 af *Brenner & Crick*, der foreslog, at rækkefølgen af aminosyrer i et protein var bestemt af tre på hinanden følgende nukleotider i DNA – de såkaldte tripletkodoner. Med fire forskellige nukleotider G, C, A og T fik man 64 (4^3) forskellige kodoner. Dette antal er tilstrækkeligt til at definere de 20 forskellige aminosyrer, og hvor proteinsyntesen (translationen) skal starte og slutte. Oversættelsen af den genetiske kode (translationsstarten) starter altid ved AUG-kodonet, der markerer, hvor samlingen af de to ribosomale subunits skal finde sted. AUG specificerer aminosyren methionin, og stort set alle proteiner starter derfor med denne aminosyre. Translationsstop defineres af kodonerne UGA, UAA eller UAG, der betegnes *nonsense*-kodoner. Den korrekte rækkefølge af triplet kodoner er proteinets læseramme. Hvis læserammen forskydes ved insertion eller deletion af et eller flere nukleotider, aflæses en forkert proteinsekvens. Koblingen mellem tripletkodonet i mRNA og elongeringen af polypeptidkæden medieres af tRNA. tRNA binder til mRNA via sit tripletantikodon og medbringer en bestemt aminosyre.

Transkription

DNA oversættes til RNA af RNA-polymeraser. Der findes tre typer: RNA-polymerase I transkriberer de store ribosomale

gener, og RNA-polymerase III transkriberer tRNA-generne, hvorimod RNA-polymerase II transkriberer alle de proteinkodende gener til mRNA. Transkriptionen starter ved det basepar, der korresponderer til 5'-enden på mRNA. RNA-polymerasen rekrutteres til genets promotor af en række generelle transkriptionsfaktorer, der binder til promotorens TATA-box (navngivet efter TATAAA-sekvensen) eller initiatorsekvens (**Fig. 3**).

Mens nogle gener er konstant aktive, er andre underlagt en temporal og spatial kontrol af deres ekspression. Vævsspecifik transkription er fortrinsvis medieret via bindingen af specifikke transkriptionsfaktorer og kofaktorer til den proksimale promotorregion og via epigenetiske mekanismer (se nedenfor). Transkriptionsfaktorer binder sig til 6-8 basepar lange sekvensmotiver og kan via interaktion med kofaktorer eller det basale transkriptionsapparat både stimulere eller hæmme aktiviteten af promotoren. Transkriptionsfaktorer inddeles som hovedregel enten efter deres bindingsspecificitet eller strukturen af deres DNA-bindende domæne. Faktorerne danner et samlet proteinkompleks med det basale transkriptionsapparat, og kombinationen af faktorer sikrer promotorens korrekte aktivitet.

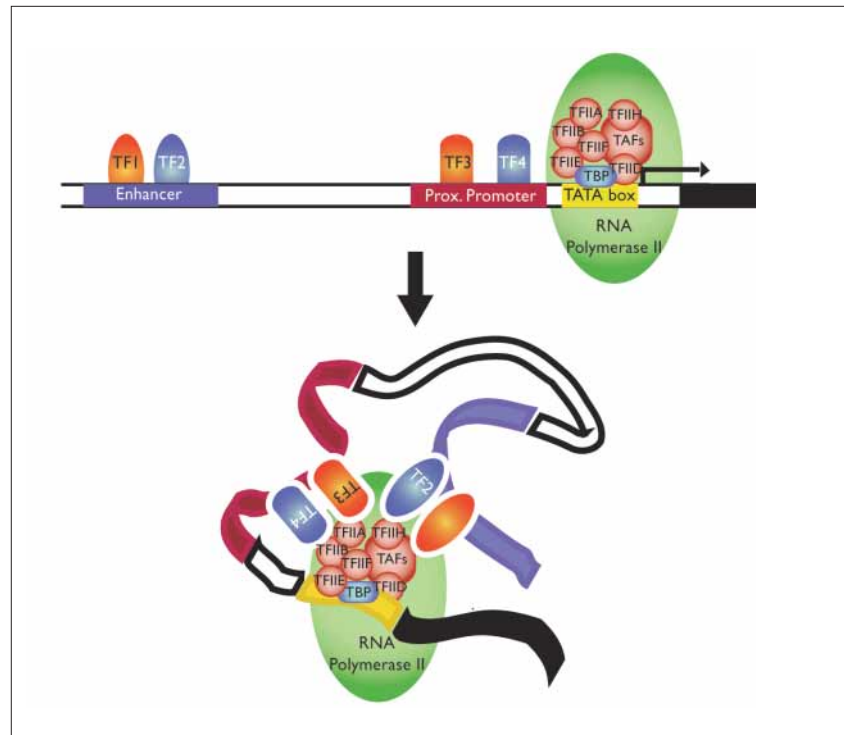
Posttranskriptionel regulering

RNA medierer overførslen af genetisk information fra DNA til protein. En lang række biologiske processer, som går under fællesbetegnelsen posttranskriptionel regulering, kontrollerer mRNA's skæbne.

mRNA-splejsning

Under og efter transkriptionen gennemgår det primære RNA-transkript en række modifikationer, før det modne mRNA er klart. Cappingenzymet påsætter en 7-methylgua-

Fig. 3. Transkriptionsinitieringen bestemmes af promotor- og enhancer-regionerne. Den proksimale promotorregion er normalt 100-400 bp lang. Den er placeret lige opstrøms for TATA-boksen eller et initiatorelement, som binder det basale transkriptionsapparat bestående af det TATA-boksbindende protein (TBP), RNA-polymerase II og generelle transkriptionsfaktorer (TAFs). TBP associerer med transkriptionsfaktoren TFIIB, og sammen binder de RNA-polymerase II og TFIIF. Efter binding af TFIIE og TFIIH starter elongeringen af RNA-kæden. I både promotorregionen og enhanceren findes bindingselementer for både generelle og vævsspecifikke transkriptionsfaktorer (TF1-4). De associerede faktorer interagerer i et tredimensionalt kompleks med det basale transkriptionsapparat og stimulerer transkriptionen af genet. Adapteret fra (2).



nosine på 5'-enden af RNA-strengen, og poly(A)-polymerasen påhæfter ca. 200 adenylater på enden af RNA'et. Herefter splejses introner ud, og exoner sættes sammen i spliceosomet. Splejsningen foregår i 98% af alle tilfælde efter den såkaldte GU-AG-regel, hvor GU er de første og AG er de sidste nukleotider i intronet. Mange mutationer rammer netop de konserverede nukleotider i *splice sites*, og dette medfører som regel, at hele exoner udelades af mRNA. Efter splejsning transporteres det modne mRNA ud af kernen.

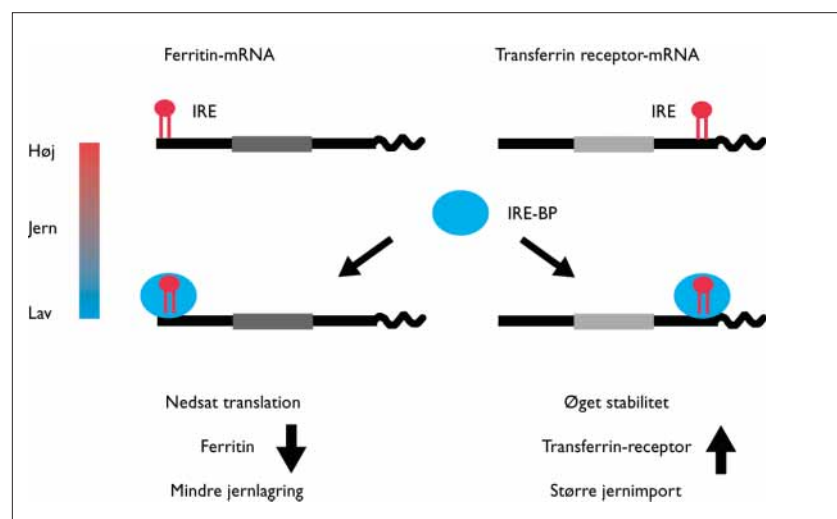
mRNA translation

I cytoplasma er mRNA forbundet med translationsapparatet, eller det opbevares i kortere eller længere tid i kompleks med forskellige RNA-bindende proteiner. Translationen af mRNA bestemmes af både RNA'ets struktur og af forskel-

lige translationsfaktorer. Translationsinitieringen finder almindeligvis sted ved mRNAs 5'-ende. Initieringskomplekset sætter den lille ribosomale 40S-subunit ned på mRNAets 5' utranslaterede region, som scannes ned til AUG-translationsstart-kodonet. Her bindes den store ribosomale 60S-subunit sammen med 40S-subuniten, og aflæsningen af læserammen og peptidsyntesen (elongeringen) påbegyndes. mRNA-halveringstiden i cytoplasma varierer fra få minutter til dage. Nedbrydningen af mRNA medieres af nukleaser, der ledes til mRNA'et via sekvensmotiver i mRNA'ets 3' UTR. Mange cytokin- og protoonkogen-mRNA indeholder en AUUUA-sekvens, som destabiliserer transkripterne og sikrer, at proteinerne kun er kortvarigt til stede i cellen.

Leverens jernmetabolisme er paradigmet på selektiv mRNA-translation og stabilitet (Fig. 4) (3). Når jern-

Fig. 4. Posttranskriptionel regulering af leverens optagelse og lagring af jern. Den transferrinreceptor- og ferritinmedierede transport og lagring af jern i hepatocytter er paradigmet på posttranskriptionel kontrol. Levercellerne indeholder aconitase iron responsive element binding protein (IRE-BP), som er i stand til at binde jern og fungerer som en sensor af jernindholdet. Når jernniveauet er lavt, binder IREBP til et element (IRE) i ferritin mRNA's 5' UTR og forhindrer translationen af transkriptet samt til samme element i 3' UTR af transferrinreceptor-mRNA, således at det stabiliseres, og der transporteres mere jern ind i cellen. Adapteret fra (2).



niveauet er højt, øges syntesen af ferritin i leveren, og samtidigt nedsættes leverens evne til at optage jern fra serum via en nedregulering af transferrinreceptoren, der transporterer jern over levercellernes plasmamembran. Omvendt reduceres ferritinindholdet og transferrinreceptorantallet øges, når jernniveauet er lavt. Processen styres af *iron responsive element binding protein* (IREBP). Når jernniveauet er lavt, binder IREBP til en RNA-struktur i 5' UTR af ferritin-mRNA og i 3' UTR af transferrinreceptor-mRNA, hvorved translationen af ferritin-mRNA hæmmes, og transferrinreceptor-mRNA'et stabiliseres. På den måde sikres en øget optagelse af jern og nedsat lagringskapacitet ved lave koncentrationer og en reduceret optagelse samt øget lagringskapacitet ved høje jernkoncentrationer.

mRNA-editering

En sammenligning mellem mRNA og genomsekvenser har vist, at de to sekvenser ikke altid stemmer overens (4). Dette tilsyneladende brud med det centrale dogme betegnes RNA-editering. Hos mennesket er der fundet flere tilfælde af mRNA-editering. Apolipoprotein-B (Apo-B) forekommer i en 100 kDa-form, der produceres i leveren og i en 48 kDa-form, der primært udtrykkes i tarmen. Begge Apo-B-molekylerne transporterer serumlipider, men det er kun 100 kDa-formen, der bindes til *low density lipid* (LDL)-receptoren på overfladen af levercellerne. Den vævsspecifikke ekspresion af Apo-B skyldes, at mRNA'et i leveren editeres af en deaminase, som ændrer cytosin 6666 til uracil, således at et CAA-kodon bliver til et UAA-stopkodon.

Nonsense medieret mRNA-decay

Nonsense-medieret mRNA-decay (NMD) er et særlig aspekt af posttranskriptionel regulering med stor klinisk betydning (5). Det antages, at ca. en tredjedel af alle arvelige sygdomme og mange sporadiske cancere skyldes *frameshift*- eller *nonsense*-mutationer, der begge medfører en for tidlig afslutning af translationen. Langt de fleste af de afkortede proteiner kommer aldrig til udtryk på grund af *nonsense*-medieret mRNA-decay, der genkender og nedbryder det muterede mRNA. NMD kræver tilstedeværelsen af et intron nedenstrøms for mutationen og en aktiv translation for at blive aktiveret. Betydningen af NMD er tydelig ved β -thalassæmi, idet mutationer i β -globingenet medfører translationsstopkodoner i det første eller andet exon. NMD giver en kraftig nedregulering af det muterede allel, og bærere af mutationerne er som regel helt asymptomatiske. Patienter med mutationer i det sidste exon får derimod β -thalassaemia intermedia, da ekspresionen af det muterede allel ikke nedreguleres (2).

Epigenetisk regulering

En række af de mest spektakulære genetiske fænomener forsøges af mekanismer, der ikke relaterer sig til genernes sekvens, men derimod til en epigenetisk regulering af histonbindingen og metyleringen af generne (6). Promotor-metylering modificerer bindingen af transkriptionsfaktorer og er ofte forbundet med repression af genet. En anden vigtig epigenetisk regulatorisk mekanisme er imprinting, som

medfører, at ekspresionen af et allel vil være afhængig af, om det kommer fra faderen eller moderen. Imprinting er essentiel for den normale udvikling og særlig for transporten af næringsstoffer over placenta. Fænomenet har desuden betydning for en række sygdomme, herunder for mange cancerformer og forstyrrelser af fostervæksten.

Summary

Finn Cilius Nielsen:
The human genome and gene expression.

Ugeskr Læger 2003;165:773-6.

Recent data estimate that the human genome contains only about twice as many genes as the *c. elegans* and *drosophila* genomes. However, this may not reflect the true difference in the complexity of the organisms, since splicing and other postranscriptional mechanisms may occur more frequently in humans. This review highlights the current knowledge of human gene structure and gene expression, with focus on mechanisms that play a role in human diseases.

Reprints: *Finn Cilius Nielsen*, Klinisk Biokemisk Afdeling, H:S Rigshospitalet, Blegdamsvej 9, DK-2100 København Ø.

Antaget den 24. januar 2003.

H:S Rigshospitalet, Klinisk Biokemisk Afdeling.

Litteratur

1. Lodish H, Baltimore D, Berk A et al. *Molecular cell biology*. New York: W.H. Freeman and Company, 1995.
2. Nielsen F. *Klinisk molekylærbiologi*. København: Munksgaard, 2002.
3. Rouault T, Klausner R. Regulation of iron metabolism in eukaryotes. *Curr Top Cell Regul* 1997;35:1-19.
4. Gott JM, Emeson RB. Functions and mechanisms of RNA editing. *Annu Rev Genet* 2000;34:499-531.
5. Lykke-Andersen J. mRNA quality control: marking the message for life or death. *Curr Biol* 2001;11:R88-91.
6. Reik W, Walter J. Genomic imprinting: parental influence on the genome. *Nat Rev Genet* 2001;2:21-32.