

A GENOM

Venetianer Pál

A genom fogalma

A huszadik század első felében, a klasszikus örökléstan (mendeli-morgani genetika) úgy tudta, hogy az egyes örökletes tulajdonságok meghatározói, a mintegy az átöröklés atomjainak tekinthető gének (amelyek természetéről, anyagáról akkor még semmit sem lehetett tudni), a minden sejt magjában megtalálható fonalszerű kromoszómákban, egymás mellett gyöngysorszerűen helyezkednek el és többé-kevésbé változatlanok. Az örökléstan, a genetika tudománya elsősorban arra törekedett, hogy az egyes gének egymáshoz viszonyított elhelyezkedését (térképezés), a következő generációkban történő együttes, vagy külön-külön megjelenését, hatásuk érvényesülését jellemezze. A genom fogalma (Winkler, 1920) ekkor az adott élőlény génjeinek összességét jelentette. A gén természetének megismerésében jelentős lépés volt annak tisztázása, hogy a gének azáltal képesek tulajdonságokat determinálni, hogy egyes enzimek (azaz fehérjék, mert minden enzim: fehérje) szerkezetét határozzák meg. A genetika döntő fordulatát az jelentette, mikor kiderült, hogy az átöröklés egyetlen jól ismert anyaghoz a dezoxiribonukleinsavhoz (DNS) köthető (Avery, 1945) és ennek az anyagnak a szerkezete (Watson-Crick, 1953) magyarázatot kínál arra, hogyan kettőződnek meg a gének a sejtosztódás során, és hogyan határozhatják meg fehérjék szerkezetét. A DNS hosszú láncmolekulái ugyanis négy különböző alapegység (nukleotidok) egymáshoz kapcsolódásával épülnek föl és e négy nukleotidféleség sorrendje az az információ, amely egy – az egész élővilágban egységes rejtjelkulcs, a genetikai kód – segítségével meghatározza minden egyes fehérjében a fehérjeláncot alkotó aminosavak sorrendjét (20 különböző aminosavból épülnek fel az átlagosan 300-400 aminosav hosszúságú fehérjeláncok). Mióta ezt tudjuk, azóta a genom fogalma az egyes élőlények, illetve egyes sejtek összes DNS-ét jelenti, a gének a DNS-lánc meghatározott, hosszabb-rövidebb szakaszait, amelyek egy-egy fehérjelánc szerkezetét határozzák meg (kódolják).

A genom megismerése

Már régen tudjuk, hogy a a genom mérete nagyjából arányos az élőlények fejlettségével, komplexitásával. A legkisebb vírusoknál néhány ezer nukleotid, a baktériumoknál néhány millió, élesztőnél tíz-, muslicánál százmilliós nagyságrendű, a gerinces állatok többségénél, így az embernél is 3 milliárd nukleotid (a fenti adatok az úgynevezett *haploid* genomra vonatkoznak, azaz a magasabbrendű élőlényeknél *egy* kromoszómaszerelvényre. Mint tudjuk, ilyet csak az ivarsejtekben találunk, a testi sejtek *két* kromoszómaszerelvényt tartalmaznak azaz *diploid* genomot. Mivel a magasabbrendű élőlények

mitochondriumi és a növények kloroplasztiszi is tartalmaznak DNS-t, az eddig említett nukleáris genom mellett szokás beszélni külön mitochondriális, illetve kloroplasztisz-genomról is). Az is régen ismert, hogy az élővilágban találkozunk a sorból kilógó kivételesen nagy, illetve kivételesen kis genomokkal is.

	Legkisebb		Legnagyobb	
	Név	Méret	Név	Méret
Vírus	MS2 fág	3.569	Mimivírus	1,181.404
Baktérium	Carsonella ruddii	159.662	Solibacter usitatus	9,970.000
Állat	Protylenchus coffeae (féreg)	20,000.000	Polychaos dubium (amőba)	670.000,000.000
Gerinces	Tetraodon nigroviridis (fuguhal)	385,000.000	Pratipterus aethiopicus (tüdőshal)	130.000,000.000

1. ábra. A genomméret extrémításai (a méretadatok DNS nukleotidpárt jelentenek, illetve az MS2 fágna RNS nukleotidot)

A huszadik század hetvenes éveiben dolgozták ki azokat a módszereket, amelyekkel a DNS nukleotidsorrendje meghatározható (Sanger, 1977) így 1990-ben lehetővé vált a Humán Genomprogram megindítása, amely célul tűzte ki az ember mintegy 3 milliárd nukleotidnyi teljes haploid genomjának megismerését. A program 2004-ben fejeződött be, azóta viszont a DNS szerkezet megismerése (szekvenálás) újszerű technológiák és az azokat alkalmazó automatizált készülékek segítségével nagyságrendekkel gyorsabbá és olcsóbbá vált, így ma (2012 március) már közel tizenhatezerkülönböző faj teljes genomszerkezetét ismerjük és ez a szám napról napra nő. Elmondhatjuk, hogy a teljes genomok vizsgálatával, összehasonlításával foglalkozó tudomány a *genomika* lassan elfoglalja a *genetika* helyét. Érdeemes megjegyezni, hogy az utóbbi évtizedben a *genom* és a *genomika* fogalmainak analógiájára számos új „-omika” terjedt el a tudományban, amelyek a sejt, illetve organizmus egyes szerkezeti illetve funkcionális elemeinek teljességét kívánják leírni és tanulmányozni. Így beszélünk *transzkriptómáról* (valamennyi DNS-átírási termék), *glikómáról* (valamennyi szénhidrát), *lipidómáról* (valamennyi lipid), *ribonómáról* (valamennyi RNS), *metabolómáról* (valamennyi kismólsúlyú anyagcseretermék), *interaktómáról* (valamennyi fehérjék közötti kölcsönhatás, stb).

Az emberi genom

A klasszikus értelemben vett gének (fehérjét kódoló DNS-szakaszok) a genomnak kevesebb mint 2%-át teszik ki és számuk kb. 20-25000 (régében sokkal nagyobb génszámot feltételeztek). A genom körülbelül felét olyan ismétlődő DNS-szakaszok adják, amelyek eredetileg mintegy parazitaként szállták meg a genomot, helyüket változtatni képes („ugráló”) gének voltak. Ezeknek döntő többsége ma már stabil, és evolúciós multunk maradványainak tekinthetők. Vannak olyan rövid, ismétlődő DNS-szakaszok,

amelyeknek nincs semmiféle ismert funkciójuk, de az egyes helyeken történő ismétlődéseik száma nagy egyedi változatosságot mutat és ezek mintázata szolgál az egyedi DNS-azonosítás (DNS-ujjlenyomat készítés) alapjául. Vannak ezen kívül olyan egészen rövid, de milliószámra ismétlődő szakaszok, amelyek minden bizonnyal csak szerkezeti elemei a kromoszámának. A genom jelentős része (egyeb becslések szerint 70%-a) átíródik (azaz RNS-másolat készülhet róla), noha ezek az RNS-ek nem kódolnak fehérjét. Egy (a közelmúltban felfedezett) részük igen fontos (jelenleg a kutatások középpontjában álló) szabályozó szerepet játszik a génműködésben ezek az ún. kisRNS-ek, más részüknek egyelőre nem ismerjük a szerepét. Az mindenesetre biztos, hogy a fehérjét nem-kódoló DNS, amelyet korábban (helytelenül) „szemét-DNS”-nek neveztek igen fontos és részben még nem ismert szerepet játszik a sejtek működésében.

Fehérjét kódoló szekvenciák	2 %
Intronok	26 %
Egyéb egyedi szekvenciák	12 %
Egyszerű szekvenciaismétlődések	3 %
Nagyobb régiók megkettőződése	5 %
Heterokromatin	8 %
Mobilis ismétlődő elemek	44 %

2. ábra. Az emberi genom szerkezeti elemeinek közelítő megoszlása

Az emberi genom egyedi különbségei

Mivel ma már több tucatnyi emberi egyed teljes genomját ismerjük (nem teljes szekvenciainformációk már többzezernyi emberről állnak rendelkezésre), megállapítható, hogy genomi szinten háromféle különbség adja az ember egyediségét. A nukleotidsorrendben (szekvenciában) mutatkozó különbség 0.1 %, azaz átlagosan 1000 nukleotidonként egy különbözik minden egyénben, e különbségeket a szakmai zsargon SNP-nek (= single-nucleotide-polymorphism = egyes nukleotidok többalakúsága) nevezi. Ennél nagyobb (1-3 %) különbség mutatkozhat abban, hogy hosszabb DNS-szakaszok ismétlődhetnek és az ismétlődések száma és helye egyénenként különböző. Ezek neve CNV (= copy-number-variant = példányszámkülönbség). Végül a harmadik különbség nem a nukleotidsorrendben mutatkozik, hanem a DNS másodlagos kémiai módosulásaiban (például egy-egy metilcsoport jelenléte vagy távolléte valamelyik nukleotidon), illetve a DNS-hez kapcsolódó fehérjék, a kromatin bizonyos módosulásaiban. Ezek neve: epigenetikus módosulások. Régebben ezeket kizárólag az egyedi élet során szerzett megváltozásnak tekintették, de a mai felfogás szerint ezeknek egy része is átörökíthető. A jelenlegi orvosi-genetikai kutatások központi kérdése, hogy az itt felsorolt tényezőknek milyen szerepe van a legfontosabb emberi betegségek kialakulásában. Ebből a szempontból igen fontosak azok a

vizsgálatok (GWAS = genome-wide association studies = genomi-szintű asszociációs kutatások) amelyek igen nagyszámú beteg, illetve egészséges egyed genomját, illetve egyes genomszakaszait vizsgálják, és azt kérdezik, hogy, hogy milyen SNP-k, illetve CNV-k fordulnak elő gyakrabban a betegek, mint az egészséges kontrollok között, mivel ezekről joggal feltételezhető, hogy az illető betegségek kialakulására hajlamosító genetikai tényezők.

Az emberi egyedek közötti DNS-szekvencia különbségről még feltétlenül megjegyzendő, hogy ennek értéke (0.1 %) rendkívül alacsony (a csimpánzegyedek közti különbség négyszer ekkora), ami cáfolhatatlanul bizonyítja, hogy az egész emberiség egyetlen faj és mindannyian aránylag közeli rokonai vagyunk egymásnak.