

# Ihmisten erilaisuuden geneettinen perusta

Petter Portin

Ihmisen genomien tutkimus on astunut uuteen vaiheeseen kun on alettu tutkia ihmisen geneettisen monimuotoisuuden määrää ja laatua. Useita tätä tutkimushanketta koskevia artikkeleita ilmestyi *Nature*-lehdessä viime lokakuussa [1]. Myös *Annual Review of Genomics and Human Genetics* -aikakauskirjassa on äskettäin ilmestynyt aiheesta laaja kokooma-artikkeli (Crawford ym. 2005). Useista henkilöistä tehtyjen genomien analyysien perusteella on käynyt mahdolliseksi myös luonnonvalinnan osuuden tutkiminen ihmisen evoluutiossa (Bustamante ym. 2005).

Geneettinen polymorfismi eli monimuotoisuus tarkoittaa sitä, että populaatiossa esiintyy säännöllisesti ja samanaikaisesti kaksi tai useampia toisistaan selvästi erottuvia periytyviä variantteja tai genotyyppisiä eli morfeja niin runsaasti, että sitä ei voida harvinaisimmankaan morfin tapauksessa selittää yksin toistuvista mutaatioista eli mutaatiopaineesta johtuvaksi. Käytännössä polymorfismista puhutaan kun harvinaisimmankin morfin suhteellinen runsaus on yli yksi prosentti.

Luonnossa kaikki populaatiot ovat polymorfisia ja polymorfismin määrä on suuri. Kun on tutkittu monisoluiden eläinten polymorfismia proteiinien tasolla, on havaittu, että yleensä kaikilla eläimillä, ihminen mukaan luettuna, noin yksi kolmasosa proteiineja koodaavista geneista on polymorfisia. DNA-tasolla polymorfismi on vieläkin laajempaa. Eri eläimillä DNA:n polymorfismin määrä vaihtelee suunnilleen sellaisissa rajoissa, että noin joka sadas tai joka tuhannes nukleotidipari DNA:ssa on monimuotoinen. Tällöin puhutaan yhden nukleotidin polymorfismeista eli ns. snipeistä (engl. *SNP* = *single nucleotide polymorphism*). Yhden nukleotidin polymorfismit ovat aina kaksialleelisiä, koska DNA:n nukleotidipareilla on vain kaksi mahdol-

lista tilaa. Pari on joko adeniinin (A) ja tymiinin (T) tai sitten guaniinin (G) ja sytosiinin (C) muodostama.

Yhden nukleotidin polymorfismien analysointi antaa tarkimman mahdollisen geneettisen polymorfismin määrän mitan. Tällä hetkellä meillä on jo varsin hyvä käsitys näiden snippien lukumäärästä ihmisen genomissa. Ihmisen genomi koostuu 3,2 miljardista nukleotidiparista.

## Afrikkalaiset populaatiot muuntelevampia kuin eurooppalaiset

*Naturessa* julkaistussa tutkimuksessa selvitetiin neljästä Maapallon eri puolilla asustavasta ihmispopulaatiosta kaikkiaan 269 yksilön genotyyppit siinä laajuudessa, että pystyttiin havaitsemaan vähintään yksi snippi jokaista 5 000 nukleotidiparin jaksoa kohti. Lisäksi kaikista yksilöistä analysoitiin kymmenen 500 000 nukleotidiparia käsittävää DNA-jaksoa kokonaan (Cheung ym. 2005).

Ihmisen yhden nukleotidin polymorfismin kokonaismääräksi havaittiin yli 10 miljoonaa niin tässä kuin 137 eri etnistä alkuperää olevaa yhdysvaltalaisista ihmistä koskevassa tutkimuksessa (Crawford ym. 2005). Tämä merkitsee sitä, että verrattaessa ketä tahansa kahta ihmistä, he poikkeavat toisistaan miljoonien nukleotidiparien osalta.

Tutkimuksissa (The International HapMap Consortium 2005) on havaittu, että ihmisen genomissa keskimäärin joka 279. nukleotidipari on polymorfinen. Tämän perusteella laskemalla saataisiin ihmisen snippien kokonaismääräksi 11,47 miljoonaa. Tästä taas voidaan laskea, että Mendelin lakien mukaan ihmisen kaikkien mahdollisten genotyyppien teoreettinen kokonaismäärä olisi huikeat kolme potenssiin 11,47 miljoonaa. Kymmenen potensseina tämä on kymmenen potens-

siin 5,47 miljoonaa. Tämä on niin suuri luku, että jos yhden nollan kirjoittamiseen menee tilaa yksi millimetri, seuraa ykköstä 5,47 kilometriä pitkä jono nollia.

Suppeammassa tutkimuksessa (Crawford ym. 2005) tultiin vieläkin suurempaan polymorfismin määrään. Arvioitiin, että suunnilleen joka 180. nukleotidipari ihmisellä olisi polymorfinen. Jos näin on, olisi ihmisen kaikkien mahdollisten genotyyppien määrä kymmenen potenssiin 8,48 miljoonaa. On siis sula mahdollisuus, että nyt olisi tai olisi joskus ollut olemassa kahtakaan genotyyppiltään identtistä yksilöä (luonnollisesti lukuun ottamatta identtisten kaksosten erikois-tapausta).

Tässä tutkimuksessa havaittiin myös, että verrattaessa mitä tahansa kahta kromosomia, keskimäärin joka 1110. nukleotidipari afroamerikkalaisilla ja joka 1 435. eurooppalaista alkupe-  
 rää olevilla amerikkalaisilla oli erilainen. Tästä voidaan laskea, että ensin mainitussa ryhmässä mitkä hyvänsä kaksi genomia poikkeavat toisistaan 2,89 miljoonan ja jälkimmäisessä ryhmässä 2,23 miljoonan nukleotidiparin osalta.

Kuten muutkin tutkimukset ovat aikaisemmin osoittaneet, ovat afrikkalaiset populaatiot siis muuntelevampia kuin eurooppalaiset, mikä osoittaa että ihmisen alkukoti on Afrikassa. [2]

### *Poikkeamme toisistamme vain muutamana promillen verran*

Yhden nukleotidin polymorfismit ovat siis erilaisuutemme geneettinen perusta. Tämä monimuotoisuus tekee meistä ulkonäöltämme erilaisia, säätelee alttiuksiamme sairastua tiettyihin tauteihin sekä ohjaa osittain myös käyttäytymisemme eroja. Jokainen meistä poikkeaa kaikista muista ihmisistä miljoonien nukleotidiparien verran. Samalla kuitenkin snippien moinen muihin eläinlajeihin verrattuna varsin pieni suhteellinen määrä on myös samanlaisuutemme perusta. Ihmiskunta nimittäin polveutuu hyvin suppeasta, vain noin 2 000 henkeä käsittävästä Itä-Afrikassa asuneesta populaatiosta, mikä on rajoittanut lajimme geneettisen muuntelun määrää (ks. Valste 2004).

Vaikka siis ihmisen mahdollisten genotyyppien määrä on tähtitieteellistäkin suurempi, poikkeamme kuitenkin toisistamme kaikkein perustavimmalla geneettisellä tasolla vain 3,6 – 5,6 promillen verran ([1 000 : 279] – [1 000 : 180]). Tässä on ihmisten tasa-arvoisuuden geneettinen perusta ja oikeutus, ja kaikenlaisen et-

nisen ynnä muun syrjinnän biologiset perustelut voidaan heittää romukoppaan. Itse asiassa Maapallolla 10 000 kilometrin päässä toisistaan elävät ihmiset poikkeavat geneettisesti vähemmän kuin Afrikassa toistensa naapureina elävät gorillapopulaatiot.

### *Hyötyä yleisten sairauksien syiden etsinnässä*

Yhden geenin mutaatioiden aiheuttamien, Mendelin sääntöjen mukaan periytyvien sairauksien geenivirheiden etsiminen on jo pitkään ollut ihmisgeneetikkojen rutiinia. Jo ennen kuin ihmisen genomien molekulaarinen tutkimus aloitettiin, etsittiin tällaisten monogeenisten sairauksien ehdokasgeenejä klassillisen genetiikan kartoitusmenetelmiä soveltaen.

Genomimme kartan valmistuttua vuonna 2001 on periaatteessa kaikkien geneettisten sairauksien ehdokasgeenejä voitu etsiä myös suoraan genomitiedostosta. Useiden tautien osalta tämä on ollut mahdollista jo paljon aikaisemmin kun on ollut käytössä osan genomiamme kattavia molekyyllitason karttoja.

Monitekijäisten sairauksien taustalla olevien geenivirheiden etsiminen sen sijaan on paljon hankalampaa. Niiden löytäminen on kuitenkin tärkeää, koska monet yleiset, kansanterveyden kannalta tärkeät taudit, kuten verenpainetauti, diabetes ja sydänverisuonisairaudet ovat tällaisia. Tällaisten tautien puhkeamiseen ja kehitykseen vaikuttavat geenien ohella myös elintavat ja muut ympäristötekijät.

Yhden nukleotidin polymorfismien avulla voidaan etsiä tällaisten monitekijäisten sairauksien taustalla olevia geenimuotoja. Yleisenä metodisena periaatteena on se, että yksilöillä, jotka tutkittavaa tautia sairastavat, on usein tietty snippi kromosomistossa lähellä sitä geneettistä varianttia, joka sairaudelle altistaa. Tätä korrelaatiota kutsutaan kytkentäepätasapainoksi (engl. *linkage disequilibrium*, LD). Tiettyjen geneettisten varianttien ryhmää kromosomissa kutsutaan puolestaan haplotyypiksi. Kahden tai useamman geneettisen markkerin, kuten esimerkiksi tietyn snipin ja jonkin sairaudelle altistavan variantin esiintymistä yhdessä taas kutsutaan geneettiseksi assosiaatioksi.

Naturen lokakuun niteen artikkelissa (The International HapMap Consortium: The International HapMap Consortium: "A haplotype map of the human genome", 2005) esitellään tietokantana ihmisen haplotyyppien kartta ja sen laadinnan

perusta. Vastaavasti kyseistä projektia kutsutaan HapMap- projektiksi (Hap = haplotyyppi, Map = kartta). Tässä julkaistu tietokanta on HapMap-projektin ensimmäisen vaiheen tulos, joka perustuu 1 007 329 tietyn kriteerin täyttävän snipin analyysiin. Projektin toisessa vaiheessa on tarkoitus kartoittaa noin 4,6 miljoonaa snippiä lisää jokaisesta neljästä tutkimuksen kohteeksi valitusta populaatiosta.

Projekti aloitettiin lokakuussa 2002, ja sen tarkoituksena on laatia julkinen, koko genomien kattava tietokanta ihmisen yleisistä DNA-sekvenssin variaatioista. Tietokanta tarjoaa informaation, joka on tarpeen ihmisen kliinisten fenotyyppien geneettisessä tutkimuksessa ja se muodostaa ihmisen genomiprojektin luontevan jatkon. Tietokanta on internetissä vapaasti käytävissä osoitteessa [www.hapmap.org](http://www.hapmap.org)

### *Geneettisten assosiaatiotutkimusten strategia*

HapMap-projektin aloittamisen tärkein motiivi oli siis etsiä ihmisen genomien snipit jotta niitä voitaisiin käyttää geneettisten sairauksiemme ja sairausalttiuksiemme assosiaatiotutkimuksissa. Tietyn sairauden assosiaatio-tutkimuksen yhteydessä löydettyä snippien ryhmää kutsutaan osoitelapuksi (engl. *tag*). Kutakin osoitelappusarjaa voidaan analysoida assosiaatioiden löytämiseksi erilaisin tilastollisin menetelmin. Osoitelapun avulla voidaan sairaudelle altistava geenimuoto löytää keneltä hyvänsä ihmiseltä, jolla sellainen on.

Haplotyyppit purkautuvat useimmilla kromosomialueilla vähitellen geneettisen rekombinaation johdosta vaikkakin monia haplotyyppijä rajaavat myös spesifiset rekombinaatiokohdat. Toisin sanoen haplotyyppijä reunustavat ns. rekombinaation kuumat pisteet, joissa tapahtuu geneettistä rekombinaatiota paljon runsaammin kuin haplotyyppin sisällä.

Geneettisen taudin aiheuttaman mutaation iästä riippuu kuinka voimakkaasti se on edelleen kytkeytynyt alkuperäisiin snippeihin. Mitä nuorempi mutaatio, sen todennäköisemmin alkuperäinen haplotyyppi on pysynyt koossa. Jos assosiaatiotutkimuksesta halutaan todella kattava, tutkimuksen tulee käsittää kaikki kysymyksen tulevat geenimuodot, jotka voivat altistaa tutkimuksen kohteena olevalle taudille.

### *Suurin osa eroista DNA:n ei-koodaavalla alueella*

Snippejä esiintyy yleisesti kautta koko ihmisen genomien. Kuitenkin suuri osa (64 %) iistä on niin harvinaisia, että niiden harvinaisemman muodon taajuus on vähemmän kuin viisi prosenttia. Vain neljä prosenttia snipeistä sijaitsee geenien koodaavalla alueella ja 96 prosenttia ei-koodaavalla alueella (Crawford ym. 2005). Tämä ei kuitenkaan ole odottamatonta, sillä genomistamme vain noin viisi prosenttia on proteiineja koodaavaa DNA:ta.

Keskimääräinen geeni sisältää 126 snippiä, joista kuitenkin vain 46:n yleisyys on suurempi kuin viisi prosenttia. Näistä tutkituista 126:sta vain keskimäärin viisi on geenin koodaavalla alueella (eksoneissa) ja loput koodaamattomissa eksonien väleihin sijoituvissa introneissa. Tutkittiin aineistoa, joka koski sekä farmakogenetiikkaa että sydänverisuonisairauksien riskitekijöitä. Tässä aineistossa kaikkiaan 69 geenin koodaavalla alueella olevan snipin 541 tutkitusta havaittiin olevan haitallisia ko. geenin koodaaman proteiinin toiminnan kannalta. Kuitenkin kaikki mainitut 541 snippiä olivat sellaisia, jotka saavat aikaan aminohapon vaihdoksen proteiinissa (Crawford ym. 2005)

### *Ihmisen geeneihin on kohdistunut sekä positiivista että negatiivista valintaa*

Vertaamalla koodaavien alueiden polymorfismia useammassa kuin 11 000 geenissä 39 ihmisen joukossa ihmisen ja simpanssin eroihin saatiin vahvoja todisteita siitä, että luonnonvalinta, eikä yksin mutaatiopaine, on muokannut lajimme viimeaikaista evoluutiota (Bustamante ym. 2005).

Geneettisessä koodissa esiintyy synonyymisiä koodisanoja. Synonyymiset koodisanat koodaavat samaa aminohappoa proteiinin primäärirakenteessa, ei-synonyymiset puolestaan koodaavat eri aminohappoja. Tutkimalla yhtäältä synonyymisiin ja toisaalta ei-synonyymisiin koodisanoihin johtavia DNA:n vaihteluita, voidaan tehdä päätelmiä luonnonvalinnan voimakkuudesta. Tutkimuksen kohteena oli kaikkiaan 11 624 sellaista ihmisen geeniä, joilla on vastine simpanssin genomissa. Havaittiin, että 10 767 geenissä näistä (92,6 %) oli koodaavan alueen muuntelua joko ihmislajin sisällä tai ihmisen ja simpanssin välillä. Ihmisen ja simpanssin välillä havaittiin 34 099 fiksoitunutta synonyymistä eroa (siis sellaisia muunnoksia, jotka esiintyivät vain

joko ihmisessä tai simpanssissa). Tämä merkitsee 1,02 prosentin suuruista eroa. (Sekä ihmisen että simpanssin genomissa on kaikkiaan 3,2 miljardia nukleotidiparia.) Vastaavasti löydettiin 20 467 fiksoitunutta ei-synonyymistä eroa 11,81 miljoonan nukleotidiparin joukossa, eli 0,242 prosentin suuruinen ero.

Ihmislajin sisällä havaittiin vastaavasti 15 750 synonyymistä ja 14 311 ei-synonyymistä snippiä. Kun otettiin huomioon kaikkien mahdollisten synonyymisten ja ei-synonyymisten nukleotidivaihdosten määrä koodissa, havaittiin että ei-synonyymisten snippien suhteellisen määrän suhde synonyymisten snippien suhteelliseen määrään oli pienempi kuin synonyymisten snippien suhteellisen määrän suhde ei-synonyymisten suhteelliseen määrään. Tämä osoittaa, että aminohappovaihdoksia on enemmän kuin pelkän mutaatiopaineen perusteella voitaisiin odottaa syntyvän. Tämä puolestaan osoittaa luonnonvalinnan toimineen ihmisen evoluution aikana. [4]

### *Fenotyyppien erot johtuvat myös geenitoiminnan säätelyn eroista*

Kun analysoitiin ihmisen geenitoiminnan luontaista muuntelua, tehtiin assosiaatiotutkimuksia tiheällä snippien sarjalla HapMap-projektin tuottamassa aineistossa geenien toiminnan säätelykohtien löytämiseksi (Cheung ym. 2005). Tutkittiin 27 geeniä, joiden toiminnan tason tiedettiin aikaisempien perheaineistoihin perustuvien selvitysten perusteella vaihtelevan geneettisistä syistä johtuen.

Alustava perheaineistoihin perustuva tutkimus oli sisältänyt kaikkiaan 3554:n geenien toiminnan eroihin pohjautuvan fenotyypin kytkentäanalyysin 14 utahilaisessa perheessä. Näistä noin tuhannessa löydettiin viitteitä siitä, että kulloinkin tutkimuksen kohteena olleen geenin lähellä olisi sitä säätelevä kohta DNA:ssa. HapMap-projektiin pohjautuva tarkempi tutkimus näistä jakautui kahteen osaan.

Ensinnäkin tehtiin assosiaatiotutkimus 374 fenotyypin osalta. Näistä 65 (17 %) osoitti erittäin merkitsevää assosiaatiota vähintään yhteen snippiin. Lisäksi löydettiin 136 (36 %) sellaista fenotyyppiä, joissa assosiaatio oli vähemmän merkitsevä, mutta kuitenkin selvä. Osoittautui myös, että perheaineistossa havaitun kytkennän voimakkuus ennusti selvästi assosiaation löytymistä. Seuraavaksi tehtiin tarkempi, 27 fenotyypin rajoitettu tutkimus. Näillä fenotyypeillä oli

osoitettu voimakas kytkentä kyseisen geenin lähellä olevaan, sitä säätelevään DNA-jaksoon.

Kun kokonaisiin genomeihin kohdistuvan tutkimuksen mahdollisuudet ja resurssit ovat kasvaneet, on alettu kiinnittää entistä enemmän huomiota assosiaatiotutkimuksiin. Jo nyt saavutetut tulokset osoittavat, että myös geenien toiminnan säätelyyn kohdistuvat assosiaatiotutkimukset auttavat suuresti kun etsitään monitekijäisten ominaisuuksien ja sairauksien taustalla olevia genejä. Jatkossa, kun snippi-kartat vielä tihtentyvät ja genotyyppien määrittämisen menetelmät paranevat, tulee suuren mittakaavan assosiaatiotutkimuksista käytännöllinen keino tällaisten geenien etsinnässä.

### *Olemme kaikki hyvin yksilöllisiä*

HapMap-projekti ja sen tulevat sovellukset, joiden tarkoituksena on tulkita ihmisen yhden nukleotidin polymorfismien merkitys terveytemme kannalta, on monessa suhteessa käänteentekevä työ. Sen tulokset tulevat edesauttamaan paitsi monitekijäisten sairauksien taustalla olevien geenimuotojen ymmärtämystä myös monien muiden piirteidemme genetiikan selvittelyssä.

Ihmiskuvamme kannalta kuitenkin projektin mielenkiintoisin tulos on se, että verrattaessa ketä tahansa kahta ihmistä, on heidän välillään aina miljoonien DNA:n nukleotidiparien ero. Olemme siis kaikki hyvin yksilöllisiä myös geeniemme puolesta, ja voidaan sanoa, että geneettisessä mielessä ei ole olemassa "normaalia" ihmistyyppiä.

*Kirjoittaja on Turun yliopiston perinnöllisyystieteen emeritusprofessori.*

### VIIITTEET

- [1] Goldstein ym. (2005): "Understanding human diversity." / The International HapMap Consortium (2005): "A haplotype map of the human genome." / Cheung ym. (2005): "Mapping determinants of human gene expression by regional and genome-wide association." *Nature* 437.
- [2] Ks. Bustamante ym. (2005): "Natural selection on protein-coding genes in the human genome." *Nature* 437.
- [3] Itse asiassa havaittiin, että keskimääräinen nukleotididiversiteti geeniä kohti laskettuna afroamerikkalaisilla oli  $9,01 \times 10^{-4}$  ja eurooppalaisista alkuperää olevilla amerikkalaisilla  $6,97 \times 10^{-4}$  (Crawford ym. 2005)
- [4] Kun tutkitaan sitä mihin geeneihin on kohdistunut negatiivinen ja mihin positiivinen valinta, ver-

rataan polymorfismin määrää (P) yhtäältä synonyymisissä (S) kohdissa ja toisaalta ei-synonyymisissä (N) kohdissa näiden kohtien divergenssiin (D). Jos suhdeluku ( $P_S/D_S$ ):( $P_N/D_N$ ) on pienempi kuin yksi, on geeniin kohdistunut negatiivinen valinta. Jos taas suhdeluku on suurempi kuin yksi, on geeniin kohdistunut positiivinen valinta. Jos sen sijaan suhdeluku on yksi, on geeni valinnan kannalta neutraali.

Sen selvittämiseksi kohdistuuko geeniin negatiivinen vai positiivinen valinta tai ovatko ne valinnan kannalta neutraaleja tutkittiin kaikkiaan 4 916 ihmisen geeniä. Useimmat niistä (3 799; 77,3 %) olivat valinnan kannalta neutraaleja. Kuitenkin 813 geenin (16,5 %) havaittiin olleen tilastollisesti merkitsevän negatiivisen ja 304 (6,2 %) positiivisen valinnan kohteena. 45 geenin havaittiin olevan nopean evoluution kohteena. Mielenkiintoista on, että erityisesti toisten geenien toimintaa sääteleviä transkriptiofaktoreita koodaavien geenien havaittiin olevan nopeasti kehittyviä.

#### KIRJALLISUUTTA

- Bustamante, Carlos D., Fledel-Alon, Adi, Williamson Scott, Nielsen, Rasmus, Todd Hubitsz, Melissa, Glanowski, Stephen, Tanenbaum, David M., White, Thomas J., Sninsky, John J., Hernandez, Ryan D., Civello, Daniel, Adams, Mark D., Cargill, Michelle & Clark, Andrew G. (2005): "Natural selection on protein-coding genes in the human genome." *Nature* 437: 1153-1157.
- Cheung, Vivian G., Spielman, Richard S., Ewens, Kathryn G., Weber, Teresa M., Morley, Michael & Burdick, Joshua T. (2005): "Mapping determinants of human gene expression by regional and genome-wide association." *Nature* 437: 1365-1369.
- Crawford, Diana C., Akey, Dayana T. & Nickerson, Deborah A. (2005): "The patterns of natural variation in human genes." *Ann. Rev. Genomics Hum. Genet.* 6: 287-312.
- Goldstein, David B. & Cavalleri, Gianpiero L. (2005): "Understanding human diversity." *Nature* 437: 1241-1242
- The International HapMap Consortium (2005): "A haplotype map of the human genome." *Nature* 437: 1299-1320
- Valste, Juha (2004): *Apinasta ihmiseksi*. WSOY, Helsinki, 343 s.