
Nya metoder inom genetisk diagnostik och forskning – är vi på väg mot personlig genomik?

KATARINA PELIN

Metoderna för sekvensbestämning av DNA har utvecklats snabbt under de senaste åren. Vi har gått från Sanger-dideoxy-metoden till pyrosekvensering mot nanosekvensering. Det har blivit allt snabbare och billigare att sekvensbestämma hela genom. Samtidigt har mikromatristmetoderna utvecklats för snabb analys av hundratusentals enbaspolymorfier och strukturella förändringar, såsom deletioner, duplikationer och amplifikationer i genomet. Metoderna har t.ex. möjliggjort stora associationsstudier för identifiering av riskgener för allmänna sjukdomar, och nya resultat publiceras kontinuerligt. De första personliga genomsekvenserna har nyligen publicerats och det första steget mot personlig genomik har tagits. Diagnostik och skraddarsydd medicinering baserad på personlig genomik har förutspåtts, men steget dit är ännu långt.

Inledning

Våra kunskaper om människans gener och genvarianter ökar med accelererande fart i och med utvecklingen av allt snabbare och specifika metoder inom genomforskningen. Människans genomprojekt som avslutades år 2003 tog 13 år i anspråk och kostade nästan fyra miljarder dollar. Sedan dess har metoderna för sekvensbestämning av DNA blivit betydligt snabbare och billigare. I dag är det möjligt att sekvensera människans hela genom på två månader till ett pris på mindre än en miljon dollar (1). De allra nyaste sekvenseringsmetoderna, vilka ännu är under utveckling, baserar sig på nanoteknik och kommer i framtiden sannolikt att möjliggöra sekvensering av människans hela genom på ungefär en timme till priset av några hundra

dollar (2, 3). Det här öppnar helt nya möjligheter för genetisk forskning och diagnostik.

För att kartlägga genetiska likheter och skillnader mellan olika populationer inleddes det internationella HapMap-projektet år 2002. Resultaten från den andra fasen av projektet publicerades i fjol och omfattar genotypdata för över 3,1 miljoner enbaspolymorfier hos 270 individer från fyra olika populationer (4). Dessa resultat, som är fritt tillgängliga via internet, utgör grunden för bl.a. genomomfattande associationsstudier för identifiering av riskgener för allmänna sjukdomar. Både HapMap-projektet och genomomfattande associationsstudier utnyttjar mikromatriser (genchips, microarrays) för bestämningen av enbaspolymorfiernas alleler. Med mikromatriser kan man snabbt och relativt billigt bestämma allelerna för hundratusentals enbaspolymorfier i en hybridisering. Detta har lett till en enorm ökning av stora associationsstudier med syftet att hitta riskgener för multifaktoriella, eller komplexa sjukdomar så som hjärt-kärlsjukdomar, allergier, cancer m.m.

I början av detta år lanserades ”1000 Genomes Project” vars mål är att sekvensera minst tusen personers genom. Projektet väntas komma med detaljerade uppgifter om

FÖRFATTAREN

FD **Katarina Pelin** är docent i genetik och innehar en universitetslektorstjänst i humangenetik och modern genomforskning vid Helsingfors universitet.

all genetisk variation i människans genom. Resultaten kommer inte att omfatta enbart enbaspolymorfier utan även stora strukturella variationer så som deletioner, duplikationer och translokationer, vilka inte är ovanliga i vårt genom (<http://www.1000genomes.org>). Projektets resultat väntas ha stor betydelse för biomedicinsk forskning och är det första steget mot medicinsk diagnostik baserad på personliga genomsekvenser.

Sekvensbestämning av DNA

Människans genomprojekt och största delen av den rutinmässiga sekvensbestämningen av DNA baserar sig på metoden med så kallad kedjeterminering som också kallas Sangerdideoxy-metoden och är utvecklad av Frederick Sanger för över 30 år sedan (5). I början användes radioaktivt märkta nukleotider i sekvenseringsreaktionerna, och sekvensbestämningen gjordes med hjälp av polyakrylamidgelelektrofores och autoradiografi. Ersättande av de radioaktiva nukleotiderna med fluorescerande nukleotider (6) utgör grunden för den långt automatiserade sekvenseringen av DNA som rutinmässigt används i dag.

Automatiserad sekvensering med Sangers dideoxymetod lämpar sig mycket bra för sekvensbestämning av PCR-produkter (polymerskedjereaktion) och plasmider eftersom ett stort antal DNA-kopior behövs för ett lyckat resultat. Det här innebar också att sekvensbestämningen av människans genom gjordes genom att man först spjälkte DNA i fragment, som klonades i plasmider och amplifierades i bakterier, varefter plasmid-DNA isolerades och sekvenserades (7). Arbetet var tidskrävande och dyrt.

Gemensamt för de nyaste sekvenseringsmetoderna är att det arbetskrävande kloningssteget inte längre behövs. Den så kallade 454-sekvenseringsmetoden (8) är en massiv parallell DNA-sekvenseringsmetod baserad på pyrosekvensering. Den innebär detektering av pyrofosfat under pågående DNA-syntes, metoden kallas därför också sekvensering via syntes (sequencing by synthesis) (9).

Sekvensering av ett genom med 454-metoden sker i korthet på följande sätt: Det genomiska DNA spjälks i 300–800 basers fragment, små adaptermolekyler kopplas till fragmenten, de enkelsträngade fragmenten fästs vid 28 µm-pärlor (ett fragment per pärla), pärlorna fångas i oljedroppar som innehåller reagenser för PCR-reaktioner och DNA-fragmenten amplifieras inom oljedrop-

pen, som fungerar som en mikroreaktor. Efter amplifieringen kommer varje pärla att bära på flera miljoner kopior av den ursprungliga DNA-sekvensen. Pärlorna flyttas till en picotiterplatta för sekvenseringsreaktionen, och sekvenseringsreagenser (bl.a. DNA-polymeras och nukleotider) tillförs plattan. Alla DNA-kopior på en pärla sekvenseras parallellt genom en DNA-syntesreaktion utgående från adaptermolekylen som finns i ändan på varje fragment. Varje gång DNA-polymeraset inkorporerar en viss nukleotid som är komplementär till DNA-fragmentet på pärlan, detekteras händelsen som en ljussignal som avläses av en CCD-kamera och på så sätt fås sekvensen (8).

Med 454-metoden är det i dag möjligt att producera DNA-sekvenser på över 100 miljoner baser på 7,5 timmar, vilket betyder att människans diploida genom (6 600 miljoner baser) kan sekvenseras på drygt två månader. Detta har också nyligen gjorts. James D. Watsons genom sekvenserades med 454-metoden, arbetet tog två månader i anspråk och kostade under en miljon dollar (1).

Framtidens sekvensering kommer emellertid sannolikt att basera sig på nanoteknik. Nanosekvensering innebär avläsning av en enkelsträngad DNA-molekyl som passerar genom en nanopor. Avläsningen kan ske endera genom mätning av DNA-basernas elektriska laddning eller med optisk avläsning av fluorescerande DNA-baser, av dessa två strategier har den senare visat sig ge mer tillförlitliga resultat (3). Nanosekvenseringsmetoderna är ännu under utveckling, och de närmaste åren kommer att utvisa hur användbar tekniken i praktiken är.

Mikromatriser

Mikromatriser baserar sig på hybridisering av DNA eller RNA till DNA-sonder fästa på en glas- eller silikonyta. En mikromatris kan bestå av hundratusentals DNA-sonder. DNA-sondernas storlek varierar från ett tiotal baser till flera hundra kilobaser, beroende på mikromatrisens användningsändamål. Med hjälp av mikromatriser är det bl.a. möjligt att analysera genernas uttryck i olika vävnader, bestämma allelerna för enbaspolymorfier och detektera deletioner, duplikationer och amplifikationer i genomet. Mikromatriser kan indelas i två typer: sådana vars sonder täcker hela genomet och sådana vars sonder representerar vissa delar av genomet, t.ex. exoner.

Mikromatrisbaserad komparativ genomisk hybridisering (ArrayCGH)

Genomomfattande mikromatriser, som används för att detektera stora strukturella förändringar, såsom deletioner och duplikationer, består av 100-kilobasers sonder som delvis överlappar varandra och täcker hela genomet. En nyare version av genomomfattande mikromatriser består av 60-basers oligonukleotidsonder jämnt utspridda över genomet (10). Dessa typer av mikromatriser används i rutindiagnostiken för att hitta specifika förändringar i cancervävnad eller för att diagnostisera deletionssyndrom. Metoden kallas ArrayCGH eller aCGH för Array Comparative genomic Hybridization, dvs. mikromatrisbaserad komparativ genomisk hybridisering.

Inom cancerdiagnostiken används ArrayCGH-metoden för att jämföra DNA isolerat från cancervävnad med DNA isolerat från normal vävnad. DNA från cancervävnad märks med en fluorescerande färg, ofta röd, och DNA från normal vävnad märks med en annan fluorescerande färg, ofta grön. Lika mycket av de båda DNA-proverna hybridiseras samtidigt till mikromatrisen, vilket betyder att DNA-proverna tävlar om hybridisering till de komplementära DNA-sekvenserna på matrisen. Vid deletioner i DNA-provet från cancervävnad kommer endast det normala DNA-provet att hybridisera till de sonder som motsvarar det deleterade området, vilket syns som en grön signal. Vid duplikationer eller amplifikationer i DNA från cancervävnad kommer signalen från motsvarande sonder att vara röd. De områden som är lika i DNA från cancervävnad och från normal vävnad ger gula signaler efter hybridisering till sönerna på mikromatrisen (Figur 1). Signalerna från varje sond mäts med hjälp av en laserskanner, och intensiteternas proportioner för de båda fluorescerande färgerna räknas ut. På så sätt fås kvantitativa skillnader i mängden DNA som hybridiseras till en viss sond (11).

När ArrayCGH-metoden används för att diagnostisera mikrodeletionssyndrom jämförs patientens DNA med normalt referens-DNA och principen är densamma som ovan.

Mikromatriser för detektering av enbaspolymorfier

Människans genom uppskattas innehålla ungefär tio miljoner enbaspolymorfier (4). De kallas allmänt för SNP (uttalas "snipp") från engelskans single nucleotide polymorphism och representerar normal variation av

enskilda nukleotider i genomet. För att kunna klassificeras som SNP måste den ovanligare allelen finnas hos mer än en procent av populationen.

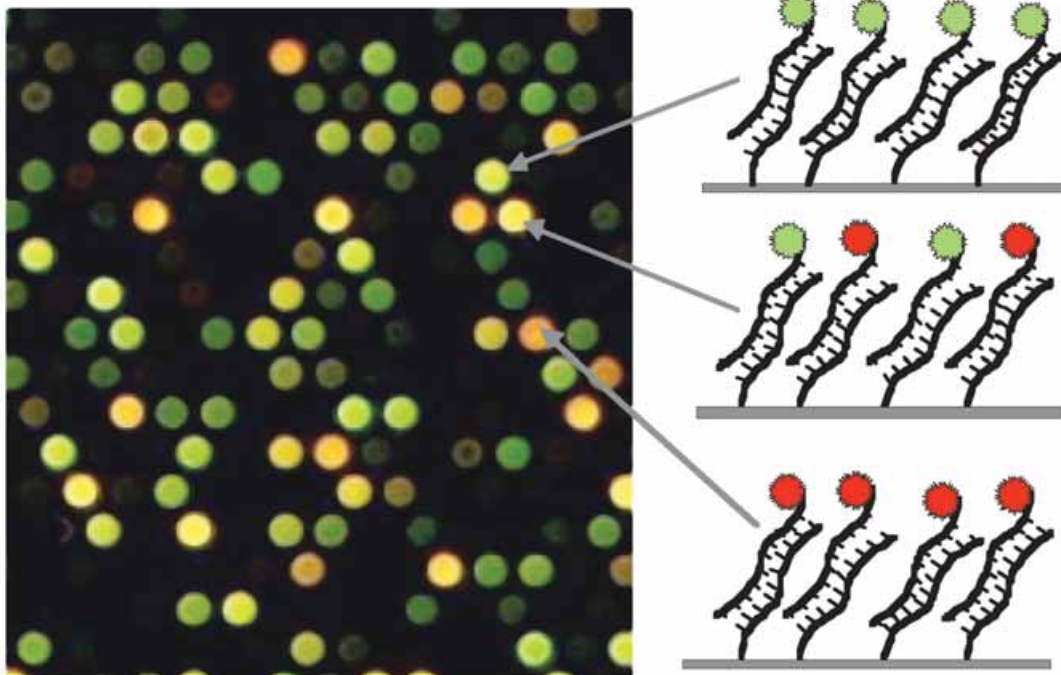
Varje SNP har två alleler och de flesta finns i icke-kodande områden av genomet, dvs. mellan gener och i introner. Dessa antas inte påverka genernas funktioner. Däremot kan enbaspolymorfier som finns i genernas styr-områden (promotorområden) ha en inverkan på genernas uttryck. Enbaspolymorfier inom exoner, dvs. genernas kodande områden, kan orsaka strukturella skillnader mellan proteiner som kodas från olika alleler. Enbaspolymorfier som påverkar genernas eller proteinernas funktioner kan utgöra riskgener för multifaktoriella sjukdomar, eller modifiera läkemedels och andra kroppsfrämmande ämnens metabolism. Det senare alternativet är grunden för farmakogenetiken.

Riskgener för multifaktoriella sjukdomar identifieras genom associationsanalys. Associationsanalyserna utförs ofta som fall-kontrollstudier, där enbaspolymorfiernas allelfrekvenser hos en grupp patienter jämförs med motsvarande frekvenser hos en grupp friska kontrollpersoner. För att en associationsanalys ska ge statistiskt signifikanta resultat måste både antalet enbaspolymorfier och antalet prover som analyseras vara mycket stort (12). I typiska genomomfattande associationsstudier analyseras 100 000–300 000 enbaspolymorfier i DNA-prover från flera tusen patienter och kontrollpersoner (13–15).

Mikromatriser som används i genomomfattande associationsstudier består av 25-basers oligonukleotidsonder för hundratusentals enbaspolymorfier jämnt utspridda i genomet. DNA-proverna som analyseras spjälks först i fragment med hjälp av restriktionsenzymer, adaptermolekyler kopplas till DNA-fragmenten och fragmenten amplifieras med PCR. PCR-produkterna spjälks i mindre delar och märks med en fluorescerande färg innan de hybridiseras till mikromatrisen. De PCR-produkter som perfekt motsvarar oligonukleotidsonden för en viss SNP-allel binds till sonden, vilket syns som en fluorescenssignal, som detekteras av en laserskanner. Resultaten analyseras med hjälp av datorprogram (16).

Personlig genomik och individualiserad medicin

Utvecklingen av allt snabbare och billigare genomomfattande DNA-analysmetoder möjliggör åtminstone i teorin en framtid av



Figur 1. Principen för ArrayCGH. Till vänster visas resultatet av en färdig hybridisering till en mikromatris. Varje fläck innehåller ett stort antal oligonukleotidsonder, som representerar en viss del av genomet. Två olika DNA-prover har hybridiserats till matrisen, det ena provet har märkts med grön fluorescerande färg och det andra med röd fluorescerande färg. DNA-proverna tävlar om hybridisering till de komplementära sönerna på matrisen, vilket schematiskt visas till höger i bilden.

individualiserad medicin baserad på personlig genomik, dvs. analys av individens hela genom. Ett stort antal gentest finns redan nu tillgängliga för ärftliga sjukdomar som orsakas av mutationer i enskilda gener (monogena sjukdomar) och för en del enzymvarianter, som har betydelse för läkemedelsmetabolismen, men analys av ett helt genom ger resultat av en helt annan storleksordning och komplexitet, vilket medför stora utmaningar då resultaten ska tolkas.

Det första steget mot personlig genomik togs när två personers, Craig Venters och James Watsons, hela genomiska DNA sekvenserades, analyserades och nyligen publicerades (1, 17). Den normala variationen i människans genom visade sig vara betydligt större och komplexare än vad man tidigare hade trott. Resultaten visar bl.a. att både Venters och Watsons genom innehåller över tre miljoner enbaspolymorfier och ett stort antal strukturella förändringar, såsom insertioner, deletioner, duplikationer och inversioner. James Watson visade sig vara heterozygot bärare av

åtminstone tio recessiva monogena sjukdomsgener (1). Motsvarande resultat rapporterades inte för Craig Venter (17).

I dag känner man till över 2 000 sjukdomsgener där de allra flesta orsakar monogent ärftliga sjukdomar och kan klassificeras som högriskgener. Ungefär 490 av de kända sjukdomsgenerna har minst en identifierad riskallel eller skyddande allel för komplexa sjukdomar, och de klassificeras som lågriskalleler. Många av lågriskallelerna representerar vanliga polymorfier med allelfrekvenser över en procent (18). Typiskt för lågriskallelerna är att de endast ökar risken för en komplex sjukdom med några procent (13–15). Genotypresultat för lågriskalleler är därför kliniskt irrelevanta så länge vi inte känner till nätverket av och samspelet mellan alla genetiska och icke-genetiska faktorer som ligger bakom komplexa sjukdomar (19, 20).

Trots detta har biotekniska företag inte varit sena med att kommersiellt utnyttja resultaten från associationsanalyser och mikromatrismetoderna för snabb och billig

analys av enbaspolymorfier. Ett exempel är det isländska företaget deCODE Genetics, som till priset av 985 dollar, via internet erbjuder privata gentest för lågriskalleler associerade med ett tiotal vanliga sjukdomar (<http://www.decodeme.com/>). Det här kan verka intressant för gemene man, men det är uppenbart att resultaten kan ge upphov till onödig oro, i synnerhet som lågriskalleler för bl.a. Alzheimer, bröstcancer, prostatacancer, kolorektalcancer och hjärtinfarkt ingår i testet. I praktiken har alltså varken positiva eller negativa testresultat i dessa fall någon medicinsk betydelse, vilket inte framgår av företagets hemsida, snarare tvärtom. Privata gentest på internet har väckt berättigad kritik bland forskare (21).

Den största optimismen med personlig genomik ligger kanske ändå inom farmakogenetik och farmakogenomik. Med farmakogenetik avses skillnader i läkemedelsmetabolismen som beror på enskilda enzympolymorfier, medan farmakogenomikens mål är att identifiera nätverket av alla genetiska faktorer som påverkar läkemedelsmetabolismen. Resultaten av farmakogenomisk forskning väntas leda till skraddarsydd personlig medicinering baserad på patientens genom. Trots att enskilda farmakogenetiska gentest redan erbjuds, är steget ännu långt till en skraddarsydd medicinering baserad på personlig genomik (20).

Analys av personliga genomsekvenser medför även etiska problem, som kräver klara och enhetliga regler. De viktigaste frågorna gäller hur och i vilken form resultaten tillhandahålls den som undersökts, huruvida det finns förpliktelser mot den undersöktes släktingar och hur resultaten får användas i framtiden (22).

Tekniskt är det i dag möjligt att få fram enorma mängder genomdata på en relativt kort tid. De största utmaningarna ligger nu i analys och tolkning av resultaten, vilket är betydligt mera komplicerat och tidskrävande. Även intensiv forskning på RNA- och proteinnivå krävs för att identifiera människans alla ca 30 000 gener och för att förstå deras funktioner. Först efter det kan vi i sann bemärkelse tala om personlig genomik och individualiserad medicin.

Docent Katarina Pelin
Institutionen för Bio- och miljövetenskaper
Genetik
PB 56 (Viksågen 5)
00014 Helsingfors universitet
katarina.pelin@helsinki.fi

Referenser

- 1 Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 2008; 452:872-876.
- 2 Ryan D, Rahimi M, Lund J, Mehta R, Parviz BA. Toward nanoscale genome sequencing. *Trends Biotechnol.* 2007;25:385-389.
- 3 Soni GV, Meller A. Progress toward ultrafast DNA sequencing using solid-state nanopores. *Clin.Chem.* 2007; 53:1996-2001.
- 4 International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;449:851-861.
- 5 Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc.Natl.Acad.Sci.U.S.A.* 1977;74:5463-67.
- 6 Prober JM, Trainor GL, Dam RJ, Hobbs FW, Robertson CW, Zagursky RJ, et al. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* 1987;238:336-341.
- 7 Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860-921.
- 8 Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;437:376-380.
- 9 Ronaghi M, Uhlen M, Nyren P. A sequencing method based on real-time pyrophosphate. *Science* 17;281:363, 365.
- 10 Barrett MT, Scheffer A, Ben-Dor A, Sampas N, Lipson D, Kincaid R, et al. Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. *Proc. Natl.Acad.Sci.U.S.A.* 2004; 101:17765-70.
- 11 Pinkel D, Seagraves R, Sudar D, Clark S, Poole I, Kowbel D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.* 1998;20:207-211.
- 12 Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat.Rev.Genet.* 2005;6:95-108.
- 13 Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007;447:1087-93.
- 14 Helgadóttir A, Thorleifsson G, Manolescu A, Gretarsdóttir S, Blondal T, Jonasdóttir A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* 2007;316:1491-93.
- 15 McPherson R, Pertsemidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, et al. A common allele on chromosome 9 associated with coronary heart disease. *Science* 2007;316:1488-91.
- 16 Matsuzaki H, Dong S, Loi H, Di X, Liu G, Hubbell E, et al. Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat.Methods* 2004;1:109-111.
- 17 Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, et al. The diploid genome sequence of an individual human. *PLoS Biol.* 2007;5:e254.
- 18 McKusick VA. Mendelian Inheritance in Man and its online version, OMIM. *Am.J.Hum.Genet.* 2007; 80:588-604.
- 19 Couzin J, Kaiser J. Genome-wide association. Closing the net on common disease genes. *Science* 2007; 316:820-822.
- 20 Nebert DW, Zhang G, Vesell ES. From human genetics and genomics to pharmacogenetics and pharmacogenomics: past lessons, future directions. *Drug Metab.Rev.* 2008;40:187-224.
- 21 Kornfeldt T. Privata gentester sågas av forskare. DN 13.07.2008. <http://www.dn.se/DNet/jsp/polopoly.jsp?d=597&a=804144>
- 22 McGuire AL, Caulfield T, Cho MK. Research ethics and the challenge of whole-genome sequencing. *Nat.Rev.Genet.* 2008;9:152-156.