



Från genomikrevolutionens årtionden till enstaka cellers funktioner

JUHA KERE

De tre senaste årtiondena har ofta ansetts vara genomikrevolutionens era. Den kunskap och de tekniker som utvecklades när det mänskliga genomets DNA sekvenserades för första gången har fått spridning överallt i medicinsk forskning. Genomtäckande associationsanalyser (GWAS) har identifierat hundratals gener som påverkar sjukdomsrisker. Individuell DNA-sekvensering har kunnat identifiera sjukdomsgener också hos enstaka patienter. Även om utvecklingen har varit oförutsett snabb, fortsätter den i accelererande takt.

De nya metoderna kan öppna helt nya frågeställningar, som alldeles nyligen var omöjliga att besvara. De allra senaste åren har man uppnått den känslighet i sekvenseringsmetoderna som gör det möjligt att sekvensera mRNA-innehållet i enstaka celler och mäta genuttryck oerhört exakt. Vi har tillämpat RNA-sekvenseringsmetoder för att förstå de första stegen i människans utveckling: hur fostrets genom aktiveras 2–3 dygn efter befruktningen. Resultaten tyder på att aktiveringen av genomet börjar med 32 gener som uttrycks vid 4-cellsstadiet och fortsätter med ytterligare 129 gener vid 8-cellsstadiet. Bland dessa gener finns flera sådana som forskarna aldrig tidigare har hittat i vävnader, dvs. gener som uttrycks unikt i det tidiga fostret för att starta genomet. Resultaten tyder också på att den tidiga utvecklingen skiljer sig avsevärt mellan olika arter: motsvarigheter för flera av dessa tidiga gener finns till exempel inte i musgenomet. Viktiga problem och obesvarade frågor återstår alltså även långt efter att människans genom sekvenserades.

SKRIBENTEN

Juha Kere är professor i molekylär genetik vid Karolinska Institutet, Stockholm, sedan 2001, och professor i genetik och molekylär medicin vid King's College, London, sedan 2016. Hans forskning täcker områden från positionell kloning av människans sjukdomsgener och riskgener för komplexa sjukdomar till studier av mekanismer för dessa geneffekter. På senare tid har han fokuserat på att förstå hur de första stegen av livet reglerar dagarna efter befruktningen.

Genomikrevolutionen

Det är inte lätt att lägga fast en tydlig startpunkt för genomikrevolutionen. Precis som vid andra revolutioner mognade de revolutionerande idéerna så småningom och den stora

förändringen började synas först när revolutionen fick fotfäste inom många olika fält, i detta fall inom medicinsk forskning (Tabell I). Det går att räkna upp ett antal kritiska uppfinningar som var nödvändiga för att möjliggöra den senare utvecklingen. Naturligtvis var det viktigt att förstå ärftlighetens molekylära bakgrund och DNA-dubbelspiralens struktur (1). Men lika viktigt var det att rengöra enzymer från mikrober som kunde klyva, klistra och kopiera DNA-molekyler på ett noga definierat sätt – därmed fick vi sådana verktyg som restriktionsenzymer, ligaser, polymeraser m.fl. (2). Vidare lärde man sig isolera och kopiera delar av stora genom, och senare multiplicera definierade DNA-molekyler helt enzymatiskt med polymeraskedjereaktion (PCR) (3). Det behövdes dessutom metoder för att sekvensbestämma DNA-molekylernas struktur (4), och allt var klart för den dåtida megalomaniska tanken att definiera innehållet i människans hela arvs massa. Efter ett

antal möten och kritiska värderingar tog man i USA beslutet att satsa på ett femtonårigt projekt, The Human Genome Project, ofta även kallat HUGO-projektet enligt Human Genome Organisation, som är en internationell akademi (5).

Projektet startade officiellt 1990 – samma år som författaren till denna artikel åkte till Washington-universitetet i St. Louis, Missouri, som postdoktoral forskare och inledde sitt arbete inom projektet med målet att kartlägga människans X-kromosom som det förberedande steget till de senare DNA-sekvenseringsprojekten.

Låt oss dock hoppa över ett årtionde av kompetitivt och högst intressant arbete för att få ihop de olika delarna av projektet – en genetisk karta över genomet, en fysikaliskt klonad DNA-karta över genomet, motsvarande kartor för andra genom såsom mus och masken *Caenorhabditis elegans*, och slutligen en ordnad DNA-sekvens över alla kromosomer. Låt oss i stället gå direkt till de breda tillämpningar och konsekvenser för medicinsk forskning som projektet har genererat. Människans genom förklarades vara sekvenserat första gången 2001 med en första systematisk inblick i genomets struktur (6, 7) och mer färdigt sekvenserat 2003, även om stora luckor fanns kvar. Också idag uppdateras fortfarande nya kontinuerligt förbättrade genomsekvensversioner.

Från genetisk koppling till genomets blockstruktur

Redan de tidigaste genetiska kopplingskartorna gjorde det möjligt att kartlägga eller placera sjukdomsgener i olika kromosomer i genomet (8). I början kunde denna metodik tillämpas enbart på gener med tydliga nedärvningsmönster, alltså sådana sjukdomsgener som kunde klassifieras som dominant, recessiva eller X-kromosomala enligt Mendels lagar. Metoderna ledde till kartläggningen och senare till identifieringen av gener för alla de sällsynta ärftliga sjukdomar som hör till det så kallade finska sjukdomsarvet (9, 10).

Men aptiten på nya rön ökade. Så småningom ville man även lösa gåtor bakom de så kallade multifaktoriella, eller komplexa, sjukdomar som inkluderar många av de kliniskt viktigaste och vanligaste sjukdomarna: astma, diabetes, kardiovaskulära sjukdomar, schitsofreni, reumatiska sjukdomar m.fl. Efter de första kartläggningsförsöken med samma metoder som användes för klart ärftliga

mendelistiska sjukdomar blev besvikelsen stor. Med kopplingsanalyser fick man resultat som oftast inte gick att replikera i fortsatta studier. Orsaken var sannolikt att gen-effekterna är så svaga att det skulle behövas mycket större forskningsmaterial, men det var inte praktiskt möjligt eftersom genetiska kopplingsstudier förutsätter familjer i stället för enstaka patienter och kontroller. Bättre metodik skulle behövas när enstaka gen-effekter är svagare.

Ytterligare två saker behövdes innan man kunde gå vidare, nämligen förståelse för hur det mänskliga genomet hade formats över årtusenden och en viss teknisk utveckling. Man började storskaligt lista punkter på olika kromosomer, där DNA-baspar varierade mellan individer, så kallade SNP:er (Single Nucleotide Polymorphisms). I takt med detta arbete påbörjades ett nytt internationellt projekt, HapMap-projektet, som hade sitt ursprung i identifieringen av så kallade blockstruktur i det humana genomet (11, 12).

Genomets blockstruktur var ett viktigt fynd med stora konsekvenser för medicinsk forskning. Med blockstruktur avses att genomet är organiserat i DNA-segment bestående av i medeltal några tusen baspar, där det aldrig har funnits rekombinationer i hela mänsklighetens historia (11, 13). Resultaten ledde till en viktig slutsats, nämligen att man inte behöver analysera alla varierande (polymorfiska) punkter i genomet, utan man kan välja ett mycket mindre urval av SNP:er som representerar hela den vanliga variationen i så gott som alla människor. Upptäckten av genomets blockstruktur var övertygande även på ett annat sätt. Det kunde nämligen konstateras att den genetiska variationen var organiserad på ett sätt som tydligt bekräftade att alla människor har samma ursprung i Afrika (13). Målet med HapMap-projektet blev senare att systematiskt identifiera all vanlig genetisk variation hos människor från de tre stora kontinenterna, Afrika, Asien och Europa. Man förstod att det var viktigt för att kunna tillämpa kunskapen på forskning kring sjukdomsgener.

Genomtäckande associationsanalyser, GWAS

För att kunna analysera tillräckligt många SNP-varianter på en gång måste tekniken utvecklas. Den viktiga tekniska lösningen var mikrochipmetoden baserad på DNA-hybridisering. Med mikrochip kunde man

Tabell I. Genomikrevolutionens årtionden och medicinska tillämpningar.

Projekt på 1990-talet	Forskningsmål	Viktiga nya metoder och resurser
Start för genomprojektet	Sekvensera människans hela arvmassa	Automatiserad Sanger-sekvensering
Kartläggning av sjukdomsgener	Kartlägga monogena sjukdomsgener i genomet	Genetisk kopplingsanalys med mikrosatellit-PCR
Identifiering av sjukdomsgener	Positionell kloning	Kloning av stora DNA-segment, t.ex. konstgjorda jästkromosomer
Projekt på 2000-talet		
Genomprojektet färdigt	Människans genomsekvens	Genomsekvens tillgänglig i databaser
HapMap-projektet	Kartlägga variationen i det mänskliga genomet	Grundinformationen för SNP-matrisdesign
Genomtäckande associationsanalyser i komplexa sjukdomar	Identifiera gener som påverkar risken för vanliga folksjukdomar	SNP-matriser för genomtäckande associationsanalyser
Projekt på 2010-talet		
Individuell DNA-sekvensering	Identifiera sjukdomsgener, ta fram en helhetsbild av människans genetiska variation	Den nya generationens DNA-sekvenseringsmetoder, bioinformatik och databaser
Epigenetik	Förstå påverkan mellan miljö och genetik	Mikromatrismetoder för DNA-metyleringsstudier
RNA-sekvensering	Förstå genuttryck även på en resolution av enstaka celler	RNA-sekvenseringsmetoder baserade på ny DNA-sekvensering-smetodik

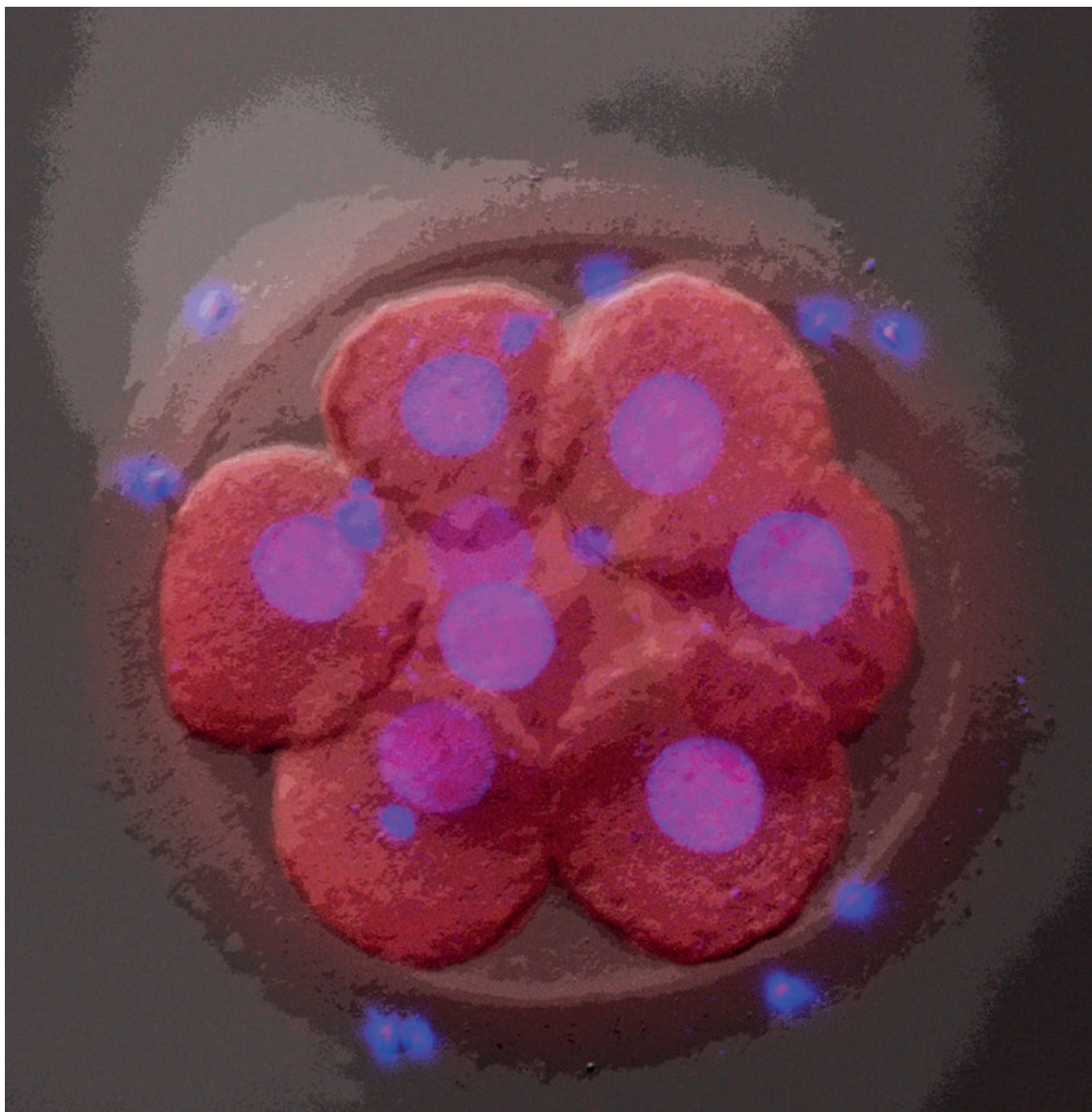
samtidigt analysera först ett tiotusental, senare hundratusentals och nu ett miljontal definierade punkter i genomet. Mikrochip designade enligt resultaten från HapMap-projektet kom att bli viktiga verktyg i mer än ett decennium av genetisk forskning. Nyligen har man tagit fram mikrochip som fokuserar exempelvis på variationen i gener relaterade till immunsystemet. Det finns dock en hel del mer sällsynta varianter som inte kan täckas in med dessa verktyg.

Med hjälp av mikrochipmetodik kunde tusentals individer snabbt och relativt billigt analyseras för hundratusentals i populationer varierande punkter. Dessa studier kallas numera GWAS efter engelskans Genome-Wide Association Study. En av de första stora studierna, som använde strategin organiserades som ett konsortium med stöd från The Wellcome Trust, och den kallas Wellcome Trust Case Control Consortium (WTCCC) (14). Studien publicerades sommaren 2007 och inkluderade sju sjukdomar med 2 000 patienter var och 3 000 kontroller. Efter studien har ett par tusen liknande arbeten publicerats och de har identifierat hundratusentals genomvarianter, som associeras med ett antal

vanliga komplexa sjukdomar och fenotyper. Antalet studier ökade snabbt, och mot slutet av 2008 fanns det i litteraturen cirka 250 GWAS-studier, i slutet av 2010 närmare 800 och i slutet av 2012 långt över 1 500 studier (15). Listan fortsätter att växa snabbt.

Genreglering påverkar sjukdomsrisiker

Resultaten av GWAS bjöd på ett par överraskningar. Den första överraskningen var att de flesta sjukdomsassocierande SNP-varianter lokaliserades utanför gener i funktionellt okänd mark. Först nyligen har forskarna börjat få ihop bevis för att många av dem påverkar regulatoriska genomelement, som ofta kan öka genexpression ("enhancer" på engelska). Sådana regulatoriska element kunde kartläggas systematiskt först när data för genexpressionen insamlades från hundratusentals vävnads- och cellprover i stora internationella samarbetsprojekt (16). När genexpressionen analyserades för att identifiera alla startpunkter för gener (promotorer), sågs även låga nivåer av transkription utanför gener, som kunde kopplas ihop med genreglering. Senare



Figur 1. Människans embryo vid 8-cellsstadiet, som visar proteinuttrycket av den viktiga startnyckelgenen LEUTX i cellkärnor (lila färg, överlappande rött för LEUTX och blått för cellkärnor).

har det visat sig att sjukdomsassocierade SNP:er ofta befinner sig i dessa områden av genomet.

Den försvunna nedärvningen

Den andra stora överraskningen från GWAS-studierna var att en del av den förväntade genetiska effekten inte kunde kartläggas, ett fenomen som man började kalla för "försvunnen nedärvning" (eng. missing inheritance) (17). Ett exempel är gener för tillväxt. Det är välkänt att människans längd till största delen, över 80 procent, är under genetisk kontroll. Ett barns vuxenlängd kan uppskattas ganska

exakt på grund av föräldrarnas längd. Men när de tjugo första generna för tillväxt hade identifierats, kunde de förklara bara 3 procent av variationen i längd mellan individer (18). Senare har ytterligare hundratals gener med väldigt små effekter kartlagts och tillsammans förklarar de nu långt över hälften av längdvariationen, men inte på långt när allt.

Ett antal möjliga mekanismer bakom denna observation har förslagits. Det kan delvis bero på att gener kan påverka varandra, alltså ha så kallade epistatiska effekter. Det är dock svårt att mäta effekterna och för det krävs väldigt många deltagare i studier, i storleksordningen en halv miljon. Det har påpekats

att de statistiska modeller som nu används för att beräkna genetiska risker inte alls kan beakta epistatiska effekter. En annan möjlighet är interaktioner mellan miljö och gener. Individer med vissa genvarianter kan reagera på olika sätt på olika miljöfaktorer. Också sådana studier är svåra att utföra på grund av svårigheter att definiera miljöfaktorer på samma sätt i olika stora kohorter i olika länder. En tredje möjlighet är epigenetisk reglering, alltså förändringar i genregleringen som beror på långvarig miljöpåverkan, till exempel barndomseffekter senare i livet. Det har också framförts misstankar om att epigenetiska mekanismer kan ha effekter över flera generationer, vilket är ett koncept som inte stämmer särskilt väl överens med våra genetiska modeller om nedärvning.

Individuell DNA-sekvensering

När människans genom hade blivit sekvenserat och genomprojektet hade förklarats slutfört, undrade många vad man skulle göra med den stora DNA-sekvenseringskapacitet som hade byggts för projektet. Andra i sin tur undrade hur man kunde öka sekvenseringskapaciteten och göra sekvenseringen ännu mycket billigare. Det gick inte många år innan en helt ny teknik utvecklades. Den var baserad på upprepade DNA-syntessteg på mikroskopiska molekylkluster på en glasyta och fotografering av ytan med miljontals sådana molekylkluster för att mäta deras fluorescens. Färgförändringar över upprepade syntesomgångar signalerar ordningsföljden i baserna i DNA-sekvensen.

Metoden revolutionerade DNA-sekvenseringen. Idag kostar det billigaste alternativet för att sekvensera hela människans genom högkvalitativt mindre än 1 000 euro. Jämfört med kostnaden för den tidigare tekniken har den nya tekniken minskat priset på det individuella genomet med en faktor på 10 000 till 100 000. Den nya tekniken resulterade i att kostnaden för att sekvensera en individs genom hade sjunkit till 10 000 amerikanska dollar 2011, och bara fyra år senare var kostnaden nere i 1 000 amerikanska dollar. Denna utveckling har varit mycket snabbare än den berömda Moores lag för datorer och datorminnen.

Möjligheten att sekvensera patienternas individuella genom har lett till en renässans av forskningen kring sällsynta sjukdomar med misstänkt mendelistisk nedärvning. Individuell DNA-sekvensering också av

enstaka patienter kan identifiera mutationer som aldrig tidigare har listats bland tusentals sekvenserade genom. Mutationernas effekter kan sedan testas och definieras på cellmodeller, som kan vara till hjälp för att formulera ökad mekanistisk förståelse. Många exempel av detta slag finns redan publicerade. Individuell DNA-sekvensering används nu ofta som första laboratorietest för nyfödda barn med misstänkt nedärvd sjukdom eller okänt syndrom. Genomikmetodik har vunnit insteg inom det vardagliga kliniska arbetet.

Stora projekt har startat för att ta fram genomsekvenser av tusentals, rentav hundratusentals individer. Ett av de nyaste initiativen inleddes i England i slutet av 2012 med målet att mot slutet av 2017 ha sekvenserat 100 000 DNA-prov från patienter med sällsynta sjukdomar för att identifiera nya sjukdomsgener. Patienter ska väljas ut för genomsekvenseringen på grund av en okänd diagnos som rimligen kan tänkas vara av genetisk ursprung, alltså orsakad av ett sällsynt syndrom alternativt sällsynt cancer (19). I projektet ingår en bioinformatikdel som syftar till att utveckla tolkningen av de identifierade sekvensförändringarna. De individuella genomsekvenserna ska förvaras i ett nytt datacenter, som även tar hand om analysprogram så att inga individuella sekvenser släpps ut. Konfidentialiteten måste vara på högsta nivå när det gäller individuella genomsekvenser.

Genuttrycksanalys av enstaka celler

Utvecklingen av den nya sekvenseringstekniken har inte bara gynnat DNA-sekvensering och mutationsanalyser från hela genomet. Samma teknik kan nu även tillämpas på funktionell analys av gener med hjälp av cDNA-sekvensering. Man kan konvertera cellulärt mRNA, alltså de uttryckta kopiorna av gener, till cDNA som sedan kan sekvenseras med hjälp av de nya metoderna för massiv parallell sekvensering. Den känslighet har nåtts som tillåter analyser av mRNA-innehållet i enstaka celler (20). För det första har denna känslighet gjort det möjligt att forska på den tidigaste fasen av livet, alltså de första dagarna efter befruktningen när fostret genomgår de första cellfördelningarna och fostrets gener aktiveras så småningom för senare differentieringsprocesser. Man kan även börja dissekera och identifiera de olika celltyperna i komplexa vävnader som exempelvis hjärnan, man kan analysera samtalet mellan celler i vävnader som består av olika celltyper och man kan

bättre förstå komplexiteten i cancervävnader, som oftast består av heterogena tumörceller efter upprepade omgångar av nya mutationer. Detta bara för att nämna några exempel.

Hur börjar livet? Genuttrycket i det tidiga fostret

Genregleringen de första dagarna har varit otillgänglig för analys eftersom mängden molekylärt material är så litet i enstaka celler i det tidiga fostret. Efter befruktningen i äggledarna börjar fostrets celledelningar i en takt av ungefär en celledelning per dygn, alltså till två celler efter första dygnet, till fyra efter två dygn, till åtta efter tre, ända upp till ungefär hundra celler vid implantationen som sker ungefär en vecka efter befruktningen. Under hela preimplantationsperioden ska celler syntetisera hundra gånger mängden DNA i den befruktade äggcellen när celler samtidigt blir mindre och mindre – fostrets diameter förändras inte innan kläckningen av fostret ur dess zona pellucida-skäl och implantationen i livmoderns slemhinna.

Den känslighet i metodiken som de nya DNA-sekvenseringsmetoderna har tillåtit har nyligen gjort det möjligt att plocka fram enstaka celler från tidiga embryon och analysera genuttrycket i dem (21). I vår studie samlade vi över 300 äggceller, befruktade ägg samt enstaka celler från 4- och 8-cellsfoster och sekvenserade deras mRNA-innehåll efter konversion till cDNA. Våra resultat bekräftade en stark mRNA-degradation vid 4-cellsstadiet, när en stor del av äggcellens mRNA-innehåll bryts ned. Samtidigt kunde vi uppmäta en signifikant ökning av bara 32 nyuttryckta gener när vi jämförde 4-cellstadiet med äggceller. Bland dessa fanns sju nya gener som ännu inte var kända i de tusentals vävnadsprov som nyligen hade analyserats med de känsligaste metoderna (16) eller annars i databaser. Dessa gener tillhörde genfamiljen homeobox, särskilt de så kallade paired-liknande (eng. PRD-like) homeoboxgenerna. Vid analysen av genaktivering ett dygn senare, alltså jämförelsen mellan 4- och 8-cellsstadierna, kunde vi identifiera ytterligare 129 nyaktiverade gener, nästan alla redan tidigare kända.

I senare studier har vi kunnat bekräfta att alla sju nya homeoboxgenerna kodar för transkriptionsfaktorer, alltså proteiner som reglerar de andra genernas uttryck. Bland generna finns en, LEUTX, som ökar genuttrycket för tusentals andra gener i cellmodeller, och andra, som DPRX, som kraftigt kan bromsa

genuttrycket (22). Båda genprodukterna kan binda sig till en 35–36 baspar lång sekvens, som är högt signifikant överrepresenterad i de regulatoriska områdena av de tidigt aktiverade generna. Dessa regulatoriska genprodukter fungerar unikt som startnyckel i början av det nya fostrets liv och aktiveras aldrig på nytt i de differentierade vävnaderna. Figur 1 visar proteinuttrycket av LEUTX i ett embryo vid 8-cellsstadiet.

Mot nya landskap

För mer än 25 år sedan, 1990, började projektet beskriva människans genomsekvens i detalj. Även om DNA-sekvenseringsteknikerna inte alls var tillräcklig utvecklade på den tiden och arbetet inleddes med kartläggning, tog det mindre än femton år att få fram sekvensen på allmän nivå. Den första sekvensen var en kombination av många individers sekvenser och kallades för referenssekvens. De senaste tio åren har man tack vare utvecklingen i DNA-sekvenseringsteknikerna kunnat fokusera på individuella genomsekvenser och deras tolkning för klinisk nytta.

Utvecklingen har varit oväntat snabb. Genomforskningen har emellertid även andra intresseområden än bara genomets allmänna och individuella struktur och dess variation mellan individer. Nu används DNA-sekvensering bland annat för att karaktärisera den mikrobflora som finns på huden eller i tarmen. En annan viktig dimension är att kartlägga vilka gener uttrycks var och när och hur genuttrycket är reglerat. De första stegen i denna riktning har tagits i form av stora internationella projekt som ENCODE och FANTOM5 (16).

Trots alla avancerade framsteg har projektet sina begränsningar. Genuttryck har traditionellt analyserats på vävnadsprovsnivå. RNA har isolerats från biopsier av vävnader, men egentligen består vävnader av enstaka celler, som kan vara väldigt olika till sina funktioner och fysiologiska roller. Exemplevis hjärnan består inte bara av neuroner (och dessutom av många olika typer av neuroner), utan också av gliaceller och blodkärl. Huden innehåller keratinocyter, fibroblaster, melanocyter, langerhanska cellerna, lymfocyter och mycket mer. De nyaste metoderna för att mäta genuttryck är tillräckligt känsliga för att genuttrycksprofiler för enstaka celler ska kunna tas fram. Denna utveckling gjorde det möjligt för oss att karaktärisera den tidigaste fasen av utvecklingen, det vill säga hur det mänskliga

genomet aktiveras vid 4- och 8-cellsstadierna i foster (21, 22).

Den nya tekniken och de ökade kunskaperna kommer ha stor betydelse även för cancerforskningen. Cancer börjar som en klonal expansion av en malign cell, men större tumörer är komplexa redan i det stadiet och består av olika cellinjer, som i de flesta fall också har olika antal kromosomer som ett tecken på sin genetiska instabilitet. Om man sekvenserar genomet eller transkriptomet av cancerbiopsier, får man fram ett genomsnitt av miljontals olika genom- och transkriptionsprofiler av heterogena cancerceller, men därutöver även blodkärl och bindvävsceller som inte är maligna.

För att förstå den klonala evolutionen av cancertumörer, deras benägenhet för metastasering och deras potential för resistens mot cancerbehandlingar måste man profilera ett större antal enstaka cancerceller med olika biologiska egenskaper. En framtidsvision som tycks självklar redan nu är att nästa steg i cancerforskningen kommer att använda och behöva metoder för profilering av enstaka cancerceller isolerade direkt från tumörer. Odlade celler duger inte, eftersom deras egenskaper förändras snabbt i cellodling och analysresultaten inte längre är representativa för den ursprungliga tumörens egenskaper.

En annan typisk egenskap för tumörer är deras snabba tillväxt, men också hög grad av celldöd som delvis helt enkelt beror på cellernas genetiska instabilitet, dock även på andra faktorer som syrebrist i tumören. Ökad celldöd leder till att ökade mängder intracellulära proteiner, men också de döda cellernas DNA, sköljs ut i blodet. Det pågår intensivt arbete för att utveckla en cancerdiagnostik, som är baserad på analys av cirkulerande DNA och identifiering av avvikande DNA-sekvenser typiska för cancer.

Även vid andra sjukdomar än cancer pågår aktiv forskning med fokus på DNA- och RNA-molekyler, som finns cirkulerande i blodet antingen i celler (för fosterdiagnostik), så kallade exosomer (små membranpartiklar som frigörs från så gott som alla celler), eller som fria nukleinsyror.

Detta är bara några exempel på de nya landskap som den fortsatta revolutionen av genomforskningen har möjliggjort och fortsätter att möjliggöra. Det känns nästan trivialt att säga att utvecklingen de närmaste trettio åren kommer att bjuda på ännu större överraskningar än vi har vittnat om de senaste trettio åren.

Juha Kere
juha.kere@ki.se
juha.kere@helsinki.fi

Referenser

1. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 1953;171:737–738.
2. Smith HO, Wilcox KW. A restriction enzyme from *Hemophilus influenzae*. I. Purification and general properties. *J. Mol. Biol.* 1970;51:379–391.
3. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 1977;74:5463–67.
4. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* 1986;51:263–273.
5. Watson JD. The human genome project: past, present, and future. *Science* 1990;248:44–49.
6. Lander ES, Linton LM, Birren B, Nussbaum C, Zody MC, Baldwin J, Devon K, et al.; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.
7. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, et al. The sequence of the human genome. *Science* 2001;291:1304–51.
8. NIH/CEPH Collaborative Mapping Group. A comprehensive genetic linkage map of the human genome. *Science* 1992;258:148–162.
9. Norio R, Nevanlinna HR, Perheentupa J. Hereditary diseases in Finland; rare flora in rare soil. *Ann. Clin. Res.* 1973;5:109–141.
10. Kere J. Human population genetics: lessons from Finland. *Annu. Rev. Genomics Hum. Genet.* 2001;2:103–128.
11. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nature Genet.* 2001;29:229–232.
12. International HapMap Consortium. The International HapMap Project. *Nature* 2003;426:789–796.
13. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225–29.
14. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. *Nature* 2007;447:661–678.
15. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorf L, and Parkinson H. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucl. Acids Res.* 2014;42:D1001–D1006.
16. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, Nüni E, Arner E, Valen E, Li K, Schwarzfischer L, Glatz D, Raithe J, Lilje B, Rapin N, Bagger FO, Jørgensen M, Andersen PR, Bertin N, Rackham O, Burroughs AM, Baillie JK, Ishizu Y, Shimizu Y, Furuhashi E, Maeda S, Negishi Y, Mungall CJ, Meehan TF, Lassmann T, Itoh M, Kawaji H, Kondo N, Kawai J, Lennartsson A, Daub CO, Heutink P, Hume DA, Jensen TH, Suzuki H, Hayashizaki Y, Müller F; FANTOM Consortium, Forrest AR, Carninci P, Rehli M, Sandelin A. An atlas of active enhancers across human cell types and tissues. *Nature* 2014;507:455–461.
17. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, et al. Finding the missing heritability of complex diseases. *Nature* 2009;461:747–753.

-
18. Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, Mangino M, Freathy RM, Perry JR, Stevens S, Hall AS, et al., Genome-wide association analysis identifies 20 loci that influence adult height, *Nature Genet.* 2008;40:575–583.
 19. www.genomicsengland.co.uk
 20. Islam S, Kjällquist U, Moliner A, Zajac P, Fan JB, Lönnerberg P, Linnarsson S. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* 2011;21:1160–67.
 21. Töhönen V, Katayama S, Vesterlund L, Jouhilahti EM, Sheikhi M, Madisson E, Filippini-Cattaneo G, Jaconi M, Johnsson A, Bürglin TR, Linnarsson S, Hovatta O, Kere J. Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. *Nature Commun.* 2015;6:8207.
 22. Jouhilahti E-M, Madisson E, Vesterlund L, Töhönen V, Krjutshkov K, Plaza Reyes A, Petropoulos S, Månsson R, Linnarsson S, Bürglin T, Lanner F, Hovatta O, Katayama S, Kere J. The human PRD-like homeobox gene LEUTX has a central role in embryo genome activation. *Development* 2016;143:3459-69.

Summary

From the decades of genetic revolution to single-cell functions

The past three decades are those of genetic revolution. Knowledge and technologies that were developed to study and sequence the human genome have spread to all areas of medical research. Most recently, genome-wide association studies have identified hundreds of genetic variants associated with complex disease risks, and full-genome sequencing with massively parallel DNA sequencing technologies have made it possible to resolve causes of rare diseases even in individual patients. Recent advances have allowed the analysis of gene expression at single-cell resolution, exemplified here by the delineation of the steps of embryo genome activation two to three days after conception.