

Systembiologin ger möjlighet att förstå komplex sjukdom i detalj

Åderförkalkning ett exempel



JOHAN BJÖRKEGREN, läkare, docent i molekylär medicin, Karolinska Universitetssjukhuset Solna; forskargruppen för beräkningsmedicin, institutionen för medicin, Karolinska institutet, Stockholm
johan.bjorkegren@ki.se

JESPER TEGNÉR, professor i beräkningsbiologi, beräkningsbiologi, institutionen för fysik, kemi och biologi, Linköpings universitet
jespert@ifm.liv.se

Medicinsk forskning har under 1900-talet till stor del bestått i att bereda hypoteser om sjukdomars uppkomst vad gäller betydelsen av enskilda eller ett begränsat antal biologiska substanser (gener, proteiner eller metaboliter). Generellt sett har denna s k kandidatgenansats försett oss med den samlade medicinska vetenskapen, inklusive en rad betydelsefulla behandlingar. Kandidatgenansatsen har varit framgångsrik framför allt när det gällt att utröna mekanismer för sjukdomstillstånd som uppstår på grund av nedärvalda mutationer i enstaka gener (monogena sjukdomar [single gene disorders] Figur 1 överst).

Emellertid är kandidatgenansatsen förknippad med ett antal begränsningar, som gör den mindre lämpad för att studera mer komplexa problem, där sjukdomsorsaken inte kan hänföras till enstaka eller ett fåtal gener (Figur 1 underst). För dessa »komplexa sjukdomstillstånd«, dit majoriteten av dagens folksjukdomar som hjärt-kärlsjukdom, cancer och inflammatoriska tillstånd som neurodegenerativa sjukdomar räknas, förefaller det i stället som om orsaken ligger i små aktivitetsförändringar i ett stort antal gener som induceras av genetiska varianter i samklang med miljöfaktorer (Figur 1 underst).

För att kartlägga mekanismer för komplexa sjukdomsförlopp måste en mer komplett och detaljerad bild av de biologiska förlopp som orsakar dessa sjukdomar erhållas. Detta har hittills inte varit möjligt.

I och med Watsons och Cricks klassiska beskrivning av DNA-spiralen i början av 1950-talet [1] och vår samtida kartläggning av den kompletta mänskliga DNA-koden (genomet) [2, 3] har förutsättningarna emellertid förändrats.

I kölvattnet av dessa historiska upptäckter har tekniker utvecklats, som gör det möjligt att mäta stora delar av genomets aktivitet samtidigt. Den idag mest utvecklade tekniken är den för att mäta genomets genexpression (s k global genexpression).

Parallellt utvecklas andra tekniker som på samma sätt gör det möjligt att samtidigt mäta koncentrationer av 10 000-tals

En mer strikt och enligt vår mening korrekt definition av systembiologi är läran om hur biologiska komponenter integrerar i nätverk.

proteiner, metaboliter, fettsyror och proteinmodifieringar, t ex fosforyleringar.

Gemensamt skapar dessa tekniker förutsättningarna för att vi nu på allvar ska kunna ta oss an komplexa biologiska fenomen som de stora folksjukdomarna.

Problemet med komplexa sjukdomar är emellertid inte löst enbart med möjligheten att mäta en allt större del av de biologiska aktiviteter som ligger till grund för hur de uppstår. En avgörande utmaning består i hur de massiva datamängder som dessa tekniker genererar ska analyseras och tolkas. Till detta har systembiologi visat sig kunna bidra. Vi har sedan fem år använt en systembiologisk ansats för att studera åderförkalkning (ateroskleros), huvudorsaken till hjärtinfarkt och stroke, som vi i korthet redogör för här.

Åderförkalkning – multifaktoriell ålderssjukdom

Åderförkalkning är en multifaktoriell ålderssjukdom som hos de flesta föreligger i de stora artärerna redan i tonåren. Sjukdomen orsakas primärt av att cirkulationens kolesterol, som huvudsakligen transporteras i s k low density lipoproteins (LDL)-partiklar, fastnar i det subendoteliala rummet i kärlväggen, primärt där turbulent blodflöde föreligger, som i kärlbifurkationer. Subendotelialt modifieras LDL-partiklarna, bl a genom oxidering, vilket initierar en inflammatorisk reaktion [4].

En första väsentlig del av denna reaktion består i rekrytering av monocyter från blodet. De modifierade LDL-partiklarna påverkar närliggande endotelceller att uttrycka receptorer som medverkar till att monocyter först kan fastna på endotelet och sedan migrera in i kärlväggen. I kärlväggen differentieras monocyterna till makrofager, som uttrycker receptorer som binder till de modifierade LDL-partiklarna, varpå dessa fagocyteras. I makrofagerna bryts LDL-partiklarna ner till huvudsakligen kolesterolestrar, som ackumuleras och ger upphov till skumceller, det första histologiska beviset för åderförkalkning. Skumcellerna ansamlas som gula strimlor (fatty streaks) i kärlväggen. Dessa förändringar i kärlväggen kan påvisas hos de flesta av oss redan i tonåren.

Efter den initiala processen (och delvis parallellt) aktiveras en rad inflammatoriska, immunologiska och andra biologiska processer (Figur 2). I senare skeden av aterosklerosförloppet sker även fibrotisering och nekrotisering i det plack som successivt bildas. Beroende på vilka av dessa processer som får överhanden bildas ett stabilt eller ett instabilt plack. Ju stabilare plack, desto mindre benägenhet har placket att rupturera. Om ett plack rupturerar orsakar det en trombos som i värsta

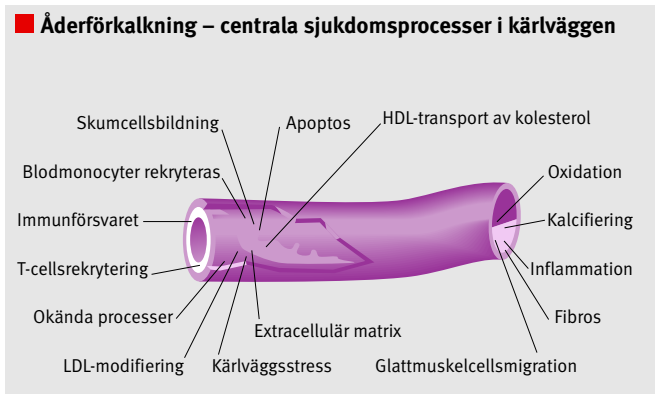
SAMMANFATTAT

Komplexiteten i hur majoriteten av de stora folksjukdomarna uppstår kräver nya ansatser vid sidan av kandidatgenansatsen.

Den idag största utmaningen för folkhälsan, hjärt-kärlsjukdom, som uppstår till följd av en accelererad åderförkalkningsutveckling, lämpar sig väl för en systembiologisk ansats.

Tekniker utvecklade i kölvattnet av kartläggningen av det mänskliga genomet kan tillsammans med systembiologi öka möjligheten att kartlägga de biologiska nätverk som ligger till grund för hur komplexa sjukdomar uppstår.

I sin förlängning kan denna utveckling bereda vägen för individualiserad preventiv hälsovård till gagn för individen och samhället.

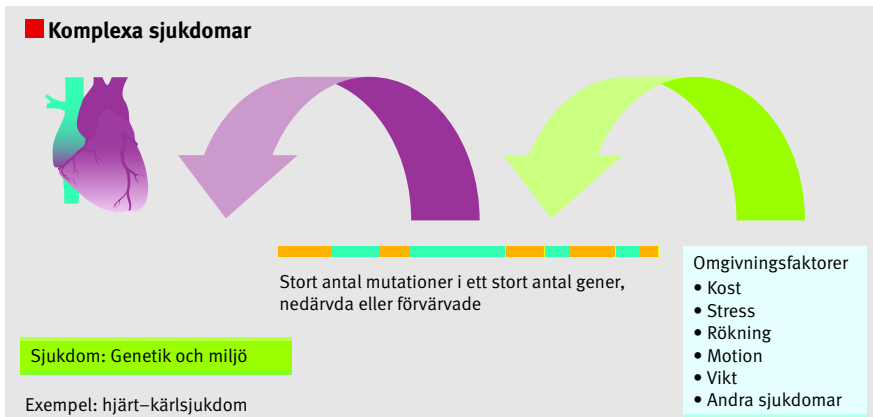
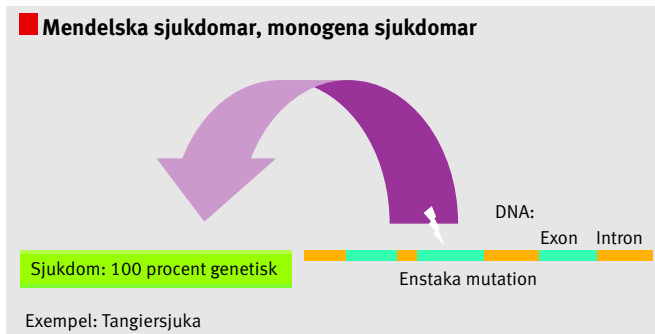


Figur 2. Exempel på biologiska processer som orsakar åderförkalkning. Alla processer är sannolikt inte kända, och hur olika processer integrerar och hur den övergripande regleringen ser ut är i stor utsträckning okänt.

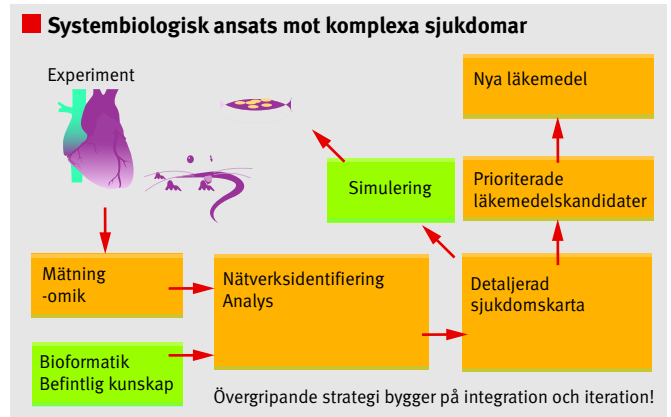
fall kan förhindra blodflödet i den aktuella artären, vilket i sin tur medför ischemi distalt och möjligen en infarkt.

Utöver de biologiska processer som sker i själva kärlväggen påverkas aterosklerosförloppet också av en rad riskfaktorer. I Figur 1 (undre figuren) anges några av de faktorer som påverkar aterosklerosutvecklingen. Den mest avgörande riskfaktorn är ärftlighet för aterosklerosrelaterade sjukdomar som hjärtinfarkt och stroke. Vad den ärftligheten består i är i de flesta fall oklart och kan omfatta en eller flera genetiska varianter med påverkan på en eller flera aterosklerosprocesser eller uppkomsten av riskfaktorer för ateroskleros, t ex höga blodfetter eller högt blodtryck.

Heterogeniteten i aterosklerosprocessen vad gäller dess



Figur 1. Etiologi till mendelska sjukdomar (överst) och komplexa sjukdomar (underst). Den enskildes omgivningsfaktorer avgör vilka genetiska mutationer som bidrar till utveckling av komplexa sjukdomar.



Figur 3. En systembiologisk ansats mot komplexa sjukdomar (system) bygger på att integrera systemexperiment (manipulation), -mätning (measurements), -analys (mining) och -simulering (modeling). Genom sådan integration kan en detaljerad bild av ett systems delar och deras samspel i ett nätverk erhållas. Även komplexa biologiska problem som åderförkalkning tros regleras i nätverk. Nätverksstrukturen är sannolikt sjukdomsspecifik, vilket gör att fokus på den faktiska sjukdomen som den uppstår hos patienter sannolikt är viktigt. Systembiologi bygger på att mäta systemets alla komponenter, vilket -omiktekniker nu erbjuder. Etablerad kunskap integreras. Till skillnad från kandidatgenansatsen är det med en systembiologisk ansats möjligt att prioritera de i sjukdomsprocessen delaktiga substanserna utifrån lämplighet som läkemedelskandidat, vilket förhoppningsvis leder till effektivare läkemedel med färre biverkningar.

multicellulära engagemang och inblandning i ett stort antal biologiska processer och omgivande riskfaktorer gör den speciellt intressant i ett systembiologiskt perspektiv.

Systembiologi – biologiska komponenter integreras i nätverk

Systembiologi är ingen ny företeelse [5], utan har funnits som begrepp sedan 1960-talet [6]. Termen systembiologi är sedan ett antal år emellertid alltmer förekommande. Anledningen är sannolikt utvecklingen av tekniker för att mäta biologisk aktivitet för hela biologiska system. När en term på detta sätt ökar i popularitet finns det anledning att vara extra noggrann med vad man avser. Systembiologi definieras ibland som läran om biologiska system i motsats till systemets enskilda delar. En sådan definition är förvisso inte felaktig men knappast tillräcklig, eftersom en sådan definition i princip omfattas av vad vi sedan länge känt som fysiologi eller patologi.

En mer strikt och enligt vår mening korrekt definition av systembiologi är läran om hur biologiska komponenter integreras i nätverk. För detta är »fyra M« (manipulation [experiment], measurement [mätning], mining [analys] och modeling [simulering]) essentiella ingredienser (Figur 3).

Manipulation avser själva experimentet och kan bestå i kliniska eller experimentella studier i djurmodeller eller i cellkultur.

Mätning (measurements) skiljer sig från den traditionella forskningen genom att man inte enbart fokuserar på någon del av systemet, utan att all aktivitet i systemet mäts (eller åtminstone en så stor del som möjligt). Parallell mätning av all aktivitet i

systemet är essentiell för att en systembiologisk ansats ska vara framgångsrik.

Dataanalys (mining) avser användandet av datorstödda algoritmer för nätverksidentifiering (se om nätverksidentifiering nedan).

Simulering är nästa steg där de nätverk som identifierats i analysen av mätdata används för att skapa modeller *in silico* (i dator) för sjukdomsförloppet. I ett nästa steg testas dessa modeller och sluter därmed cirkeln, eftersom testet består i experiment för att manipulera systemet. Detta belyser betydelsen av att iterera förloppet när man använder sig av en systembiologisk ansats (Figur 3). En annan viktig aspekt är integration av olika datatyper inklusive redan etablerad kunskap i den iterativa processen (Figur 3).

Olika sätt att identifiera nätverk

Ett nätverk ses i ett matematiskt perspektiv som en graf som definieras av noder (kopplingspunkter i nätverket) som sammanfogas av kanter (linjer mellan kopplingspunkterna) [6]. Ett proteinnätverk är ett exempel på ett nätverk med kanter utan riktning, eftersom en kant i detta fall endast anger att två proteiner binder. Ett regulatoriskt nätverk består emellertid av kanter med riktning (informationen mellan två kopplingspunkter överförs i en riktning). Riktningen bestäms av en kausalitet eller ett sannolikhetsberoende. Som exempel kan nämnas en transkriptionsfaktor som påverkar uttrycket av en viss gen till vilken den binder. En kant i ett gennätverk som har härlett från genexpressionsdata kan alltså representera antingen transkriptionsfaktoraktivitet eller indirekt en protein-proteininteraktion, som i sin tur påverkar transkriptionen.

De senaste fem åren har datorsimuleringar visat att det går att med hjälp av ordinära differentialekvationer (s k ODE) identifiera ett nätverk med riktade kanter från upprepade mätningar av simulerad aktivitet under olika störningsförhållanden. I biologiska försök har riktade nätverk på liknande sätt isolerats med hjälp av störningar, såsom knockoutexperiment eller siRNA (silencing RNA). Gennätverk har också härlett från en serie av experiment över tid, där global genexpression

analyserats vid varje tidpunkt. Dessa nätverksstudier har huvudsakligen utförts i bakterier (framför allt i *E coli*) och i jäst (*S cerevisiae*). Utöver ODE har även sannolikhetsmodeller (bayesianska) framgångsrikt använts för att identifiera nätverk från global genexpressionsdata. Den bayesianska modellen skapar kanter genom att beräkna sannolikheter för samband mellan gener.

Ett återkommande problem för nätverksidentifiering är antalet experiment i relation till antalet noder i nätverket. Det är inte möjligt att identifiera ett nätverk med 1 000-tals noder från endast 100 experiment. Att drastiskt utöka antalet experiment är naturligtvis en teoretisk möjlighet, men i praktiken en alltför kostsam sådan för de flesta laboratorier. I stället kan detta problem delvis lösas genom att komplexiteten i nätverket reduceras. Detta kan åstadkommas genom att inkorporera befintlig kunskap. Till exempel kan kända transkriptionsfaktorgen- och proteininteraktioner användas. Algoritmer för textanalyser som automatiskt identifierar möjliga gen-geninteraktioner från artiklar på Internet är ett annat alternativ.

En systembiologisk ansats mot åderförkalkning

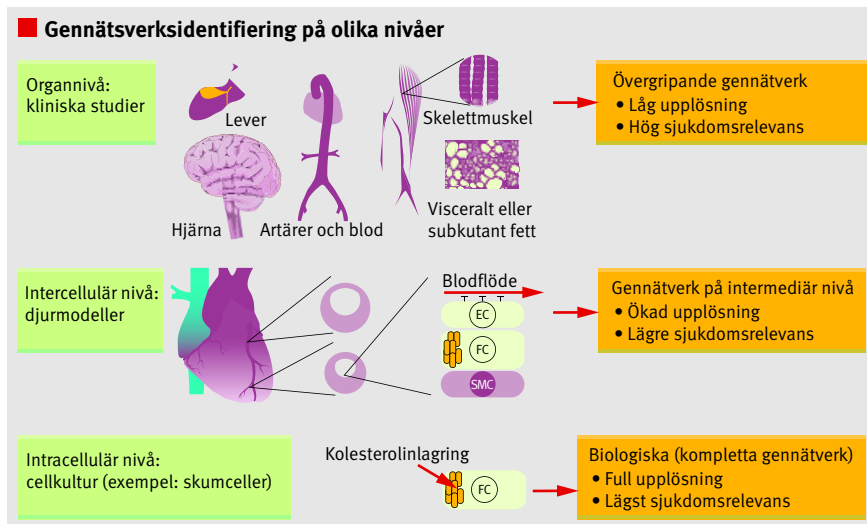
En viktig lärdom från nätverksidentifiering i jäst och bakterier är att nätverkskanter förefaller att inte vara permanenta utan skiftar beroende på omgivningsfaktorer [7]. Det förefaller som om en och samma transkriptionsfaktor aktiverar olika gener, beroende på celltyp, vävnadstyp eller omgivningsfaktorer.

Detta är ett potentiellt stort problem, eftersom vi idag i stor utsträckning lever med föreställningen att regleringen av exempelvis inflammation sker på liknande sätt oavsett var inflammationen sker eller vilken sjukdom det gäller. Så behöver alltså inte vara fallet. Det utesluter inte att delar av exempelvis en inflammatorisk reaktion sker på liknande sätt oavsett betingelser, men den genetiska regleringen av densamma kan alltså variera från sjukdom till sjukdom.

Vävnadsspecifik reglering och dynamisk reglering som beror på omgivningsfaktorer skapar frågetecken kring värdet av forskning som baseras enbart på modellsystem för att påvisa sjukdomsmekanismer (framför allt användandet av cellkultur-

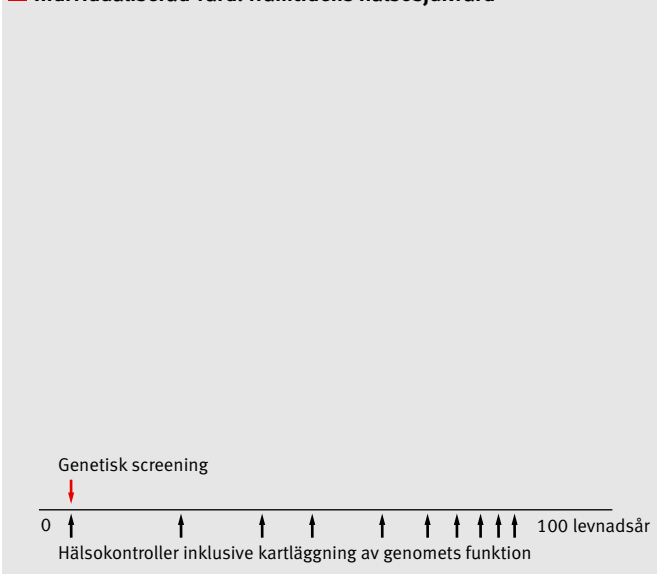
system, men i viss utsträckning även djurmodeller där omgivningsfaktorerna är atypiska för den humana situationen). För att undvika studier av biologiska processer i sjukdomsmodeller som har liten eller ingen relevans för hur en sjukdom faktiskt uppstår, är det sannolikt centralt att en systembiologisk ansats mot komplexa sjukdomar, som åderförkalkning, utgår från humana studier.

Vi har sedan fem år använt en systembiologisk ansats (Figur 3) för att identifiera gennätverk som ligger till grund för åderförkalkning. En viktig princip i vårt arbete har varit att först identifiera övergripande gennätverk utgående från globala genexpressionsprofiler av plack isolerade från patienter. I ett övergripande nätverk representerar kanter ett samband mellan två noder (gener), men det kan finnas en eller flera intermediära noder som inte identifierats. Dessa nätverk har hög sjukdomsrelevans men låg resolution (Figur 4). Med hjälp av övergripande nätverk kan centrala mekanismer för uppkomst av åderförkalkning först identifieras hos patienter, för att sedan valideras och utvidgas i djur- och cellmodel-



Figur 4. Övergripande gennätverk kan identifieras i den sjukdomsdrabbade vävnaden (den åderförkalkade kärlväggen). För att studera omgivningsfaktorer kan andra organ med indirekt påverkan på sjukdomsförloppet studeras (i fallet åderförkalkning t ex lever, skelettmuskel och fett). Det övergripande nätverket kan studeras mer i detalj över tid på intercellulär nivå i djurmodeller. Det kompletta intracellulära nätverket i full upplösning kan slutligen identifieras i modellsystem i cellkultur (t ex skumcellsmodell). (EC = endotelceller, FC = skumceller, SMC = glattmuskelceller.)

■ Individualiserad vård: framtidens hälsosjukvård



Figur 5. Med ökad förståelse kommer ett allt större antal genetiska varianter (mutationer) att kopplas till individuell risk för att utveckla komplexa sjukdomar. Med en säkrare beräkning av den individuella risken kommer ett tydligare hälsfokus. Regelbundna hälsokontroller som omfattar screening av ett stort antal sjukdomsmarkörer, sannolikt i blod, kommer att öka chanserna till tidig sjukdomsupptäckt och preventiv behandling.

ler. I cellmodellerna kan nätverk identifieras till fullo med hjälp av störningar kopplat till globala genexpressionsmätningar [6], på samma sätt som tidigare beskrivits för studier in silico och av biologiska nätverk i jäst och bakterier.

Utmaningar vid nätverksidentifiering

Nätverksidentifiering i vävnader som drabbats av komplexa sjukdomar är inte uppenbar. Det finns flera utmaningar. För åderförkalkning är plackets komposition med många olika celltyper en avgörande utmaning. Genexpressionsförändringar som härleds från ett plack kan inte säkert tolkas som faktiska förändringar i genaktivitet, utan de kan lika väl avspegla förändringar i plackets cellulära komposition. Parallella genexpressionsmätningar också från enskilda celler isolerade från placket kan till en del lösa detta problem.

Ett annat problem är hur man skapar störningar för nätverksidentifiering i humana vävnader. I cellinjer och i viss utsträckning i modelldjur kan systemmanipulering ske på basis av genmanipulering. Det låter sig inte göras i human vävnad. Vi utvecklar för närvarande en idé om att använda de kliniska fenotyperna relaterade till åderförkalkning (diabetes och hyperlipidemi med flera) som möjliga störningar av åderförkalkningsförloppet [7]. Det återstår att se om dessa störningar är specifika nog för att vara behjälpliga vid nätverksidentifiering.

Möjliga konsekvenser för framtidens hälso- och sjukvård

Enligt vår mening råder det inget tvivel om att den systembiologiska ansatsen kommer att medföra stora förändringar för den medicinska forskningen, i kombination med nya tekniker för parallella och heltäckande mätningar av den biologiska aktivitet som ligger till grund för framför allt de komplexa sjukdomarnas uppkomst.

Det är uppenbart att alla de patologiska processer som redan har identifierats vid åderförkalkning (Figur 2) måste sättas in i

en gemensam regulatorisk kontext för att en komplett bild av sjukdomsförloppet ska erhållas. Först när vi har detta klart för oss kan vi identifiera vilka nätverkskomponenter som bäst lämpar sig som behandlingsmål. Men valet av behandlingsmål kommer inte att vara trivialt eller generellt. Det är mycket möjligt att den individuella genetiska sammansättningen och hur den utsätts för omgivningsfaktorer kommer att leda till olika behandlingar för olika individer mot vad som utåt sett förefaller vara en och samma sjukdom, t ex åderförkalkning.

Systembiologi i egenskap av algoritmer för hantering av stora datamängder kan, tillsammans med den utveckling vi redan bevittnat vad gäller informationshantering, i längden påverka hur sjukvård bedrivs. Ett framtida scenario kan mycket väl vara att varje individ i ett tidigt skede i livet kartlägger sin genetiska sammansättning (redan idag finns tekniker för samtidig kartläggning av miljontals singelnukleotidpolymorfismer [SNP] som avspeglar den totala variationen i arvsmassan); denna kommer successivt (parallellt med att ny kunskap om olika genetiska varianter ökar) att användas för att beräkna den individuella risken att drabbas av sjukdom (Figur 5).

När beräkningar av risken att drabbas av sjukdom går från statistiska sannolikheter för populationer (som idag) till att bli individuella lär intresset för att förebygga sjukdom öka, och därmed skärps individens hälsfokus. Enklare hälsostatus (vikt, puls och blodtryck) kommer att kunna mätas i hemmet med apparatur, som kopplas direkt till datorn där ny information kontinuerligt lagras på en personlig hälsowebbsplats. Utvecklingen av biomarkörer, i första hand sannolikt proteinmarkörer i blod, kommer att öka antalet parametrar för att tidigt förutsäga sjukdom (Figur 5). Idag kartläggs 10-tals markörer i blod för att förutsäga hälsostatus. Utveckling av känsliga mätmetoder kommer antagligen att medföra att denna siffra markant ökar till 100-tals, som i kombination med individens genetiska profil ger förutsättningar för att generera ett mer komplett hälsostatus med individuell risk att utveckla sjukdom och eventuella tecken på att sjukdom redan föreligger.

Djupare insikter om genomets funktion i hälsa såväl som i sjukdom parallellt med ökad möjlighet att kartlägga enskilda individers arvs massa (Figur 5) kommer också att medföra utmanande etiska frågeställningar. Svaret på dessa utmaningar måste redan nu förberedas så att vinsterna av tidig upptäckt och möjligheten att förebygga sjukdom kan realiseras – för samhället och för den enskilde individen.

■ *Potentiella bindningar eller jävsförhållanden: Inga uppgivna.*

Kommentera denna artikel på www.lakartidningen.se

REFERENSER

1. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*. 1953;171(4356):737-8.
2. 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(6822):860-921.
3. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science*. 2001;291(5507):1304-51.
4. Libby P. Inflammation in atherosclerosis. *Nature*. 2002;420(6917):868-74.
5. Kitano H. Computational systems biology. *Nature*. 2002;420(6912):206-10.
6. Tegnér J, Björkegren J. Perturbations to uncover gene networks. *Trends Genet*. 2007;23(1):34-41.
7. Tegnér J, Skogsberg J, Björkegren J. Thematic review series: systems biology approaches to metabolic and cardiovascular disorders. Multi-organ whole-genome measurements and reverse engineering to uncover gene networks underlying complex traits. *J Lipid Res*. 2007; 48(2):267-77.