
”Det är väldigt svårt att förklara de här sakernas komplexitet för dem som inte alls förstår sig på databaser och informationsteknik”

HENRIK BRUUN

Bioinformatikens intåg i biologin och biomedicinen¹ går inte att stoppa. Genomiken, proteomiken, systembiologin och många av de övriga nya forskningsplattformarna bygger på användningen av bioinformatiska verktyg för datalagring, -manipulering, -analys och modellering. Det är därför dags att fundera på vad det här har för praktiska konsekvenser för forskarna och deras institutioner. I denna kolumn diskuterar jag utvecklingens kompetensrelaterade följder: hur integrera kunskap över disciplingränser? Utformandet av DNA-mikromatrisexperiment analyseras som ett paradigmatiskt exempel på de utmaningar den biologiska forskningen står inför.

Influenser från andra discipliner är givetvis inget nytt i biologins historia (1-4). Flera av 1900-talets viktigaste upptäckter var ett resultat av växelverkan över disciplingränser. Jim Watsons och Francis Cricks upptäckt av DNA-molekylens dubbelspiralstruktur är kanske det mest spektakulära exemplet. Watson var biolog till utbildningen, medan Crick hade sin bakgrund i fysiken. De var inte unika i sitt tvärvetenskapliga samarbete. Särskilt Crick var ett tidstypiskt fenomen: en fysiker som valt att söka nya utmaningar inom biologin. Fysiker och tekniker från fysikens värld kom att spela en viktig roll för uppkomsten av mole-

kylärbiologin. Detta inflytande var dock inte direkt i den bemärkelsen att biologin skulle ha reducerats till fysik eller kemi, eller att den skulle ha blivit en matematisk vetenskap. Nej, inflytandet verkar ha rört sig på ett annat plan. Enligt vetenskapshistorikern Morange (3) låg fysikernas insats i att göra biologin till en *operationell* vetenskap. De bidrog med andra ord till införandet av en kultur där de vetenskapliga begreppen förutsattes ha en motsvarighet i enkla, experimentella operationaliseringar. En annan och kanske ännu viktigare insats var att de övertygade biologerna om att ”livets gåta inte var ett evigt mysterium, utan inom räckhåll” (3).

FÖRFATTAREN

FD Henrik Bruun är forskare vid Tekniska högskolan. Han bedriver forskning inom området vetenskaps- och teknikstudier

1. För enkelhetens skull kommer jag att använda ordet ”biologi” som en förkortning för den biologiska och biomedicinska forskning som grundar sig på en förståelse av cellers och organismers biokemiska processer. Biologin som disciplin är naturligtvis mycket mer än detta, men jag fann inget annat bättre ord. De mera specialiserade beteckningarna, såsom ”molekylärbiologi”, kändes alltför specifika i detta sammanhang.

Idag befinner sig biologin i en ny period av epistemisk transformation. Bakgrunden är att man genom årtionden av forskning om DNA, gener och proteiner tillägnat sig en omfattande kunskap om deras egenskaper. De internationella genomprojekten producerar ständigt ny information om olika organismers DNA-struktur. Däremot vet man mindre om genernas funktion samt om mönstren för växelverkan mellan gener och gener, gener och proteiner, samt proteiner och proteiner. Nästa steg är därför, enligt ett växande antal forskare, att koncentrera sig på de biokemiska processernas dynamiska aspekter: signaleringsvägar, regleringsnätverk, molekylära maskiner, organeller och så småningom cellen som helhet. Från metodologisk synvinkel innebär detta ett skifte från ett "lokalt" till ett "globalt" perspektiv: i stället för att studera enskilda gener eller proteiner, studerar man hur de fungerar tillsammans (5, 6). En förutsättning för sådan forskning är att man kan hantera en stor mängd data, vilket möjliggörs av olika bioinformatiska redskap. Just dessa två aspekter – det globala perspektivet och hantering av stora mängder information – utgör kärnan i dagens "nya" biologi (7). Situationen är jämförbar med perioden av växelverkan mellan biologi och fysik under början och mitten av 1900-talet, men nu är det data- och informationsvetenskapen som utgör den disciplinära motparten.

Bioinformatikens centrala roll för biologin blev uppenbar i samband med 1990-talets stora genomprojekt. Dels var elektroniska databaser nödvändiga för att lagra den enorma mängd information som sekvenseringsprojekten gav upphov till. En bok med exempelvis bananflugans, *Drosophila Melanogaster*, hela DNA-sekvens skulle omfatta 27 000 sidor (bokstavsstorlek 8, inga marginaler). Databasernas vetenskapliga betydelse stod klar senast år 1994, då Bert Vogelstein, en av USA:s vid denna tid mest ansedda cancerforskare, upptäckte cancerförklarande genmutationer (grovtarmscancer) genom att analysera informationen i en gensignaturdatabas, dvs. en EST-databas. Den nya biologin skulle inte vara möjlig utan den uppsjö av databaser som existerar idag: DNA-databaser, SNP-databaser, EST-databaser, proteindatabaser, bibliografiska databaser och så vidare. Bioinformatiken är dessutom viktig som ett redskap för datamanipulering och -analys; i denna egenskap hade den en helt avgörande roll i tävlan om att sekvensera det mänskliga genomet. Craig Venters och hans företags (Celera Ge-

nomics) s.k. shot gun-sekvenseringsteknik byggde på att man spjälkade upp det mänskliga genomet i små, slumpmässigt ordnade DNA-fragment, för att sedan låta datorerna koppla samman dem till en löpande sekvens. Detta var inget litet arbete, vilket man kan förstå om man betänker att Celera under vissa perioder av 1999 avläste ca 40 miljoner baser per dag (8).

För forskningsinstitutionerna innebär den nya biologin en mängd utmaningar. Införandet av tekniker med s.k. high-throughput i forskningens explorativa fas innebär att grundforskningen blir ännu mer kapitalintensiv. Laboratoriernas konkurrenskraft blir i allt högre grad beroende av mekanisk automation, instrumentering och system för behandling av laboratoriedata (5). Förändringen väcker många frågor om finansiering och organisation. I det här sammanhanget vill jag ta fasta på de senare. De framväxande plattformarna ställer nya krav på kompetens. Forskningsansatserna tenderar att bli bredare, vilket innebär att man behöver en mer heterogen vetenskaplig bas. Det räcker inte med att man behärskar en enda forskningsplattform, såsom studiet av genexpression med hjälp av DNA-mikromatriser (DNA arrays, biochips). Det globala perspektivet innebär att man integrerar dessa resultat med forskning där man använder andra plattformar. Expressionsstudierna kombineras således med forskning på andra områden, t.ex. proteomik, metabolomik, cellbiologi och klinisk forskning med modellorganismer. Den långsiktiga målsättningen kan vara att skapa en modell för cellernas funktion och sjukdomarnas uppkomstmekanismer (7, 9). Bioinformatikens olika vektyg – allt från databaser till analys- och modelleringsprogram – intar en central roll i denna typ av forskning. Laboratorierna måste därför noggrant överväga vilken bioinformatisk kompetens som behövs och hur de på bästa sätt kan anamma den.

Min egen, aktuella forskning tyder på att laboratoriernas kompetensutveckling inte går helt smärtfritt. Studien jag hänvisar till bygger på intervjuer med finländska forskare som använder sig av DNA-mikromatriser. Denna teknik för analys av genexpression utvecklades under 1990-talet vid Stanfords universitet och har sedermera kommit att bli en av den nya biologins viktigaste forskningsplattformar (10). DNA-mikromatrisernas stora förtjänst är att man med hjälp av dem kan analysera förändringar i genexpressionen hos

tusentals gener i ett enda experiment, medan man tidigare fick avancera gen för gen (11). Mikromatrisexperimenten genererar en enorm mängd information, och forskarna är därför helt beroende av bioinformatiska verktyg för datalagring, -manipulering och -analys (12). Bioinformatiken utgör dock inte en del av utbildningen hos dagens biologiska och biomedicinska forskare, och intervjuerna visar tydligt att många uppfattar den bioinformatiska biten som svår och problematisk.

En möjlig lösning på kompetensproblemet vore att genomföra en arbetsfördelning där datamanipulerandet och -analysen helt enkelt skulle lämnas till utomstående bioinformatiker. Mycket av den biologiska forskningen bygger ju på en arbetsfördelning där var och en har en avgränsad uppgift i enlighet med sitt expertområde. Mina intervjuer och litteraturen på området tyder dock på att det är svårt att organisera DNA-mikromatrisexperimenten modulärt, dvs. så att det biologiska elementet skulle avskiljas från det bioinformatiska. Insikter om dataanalysens begränsningar och möjligheter behövs redan vid planeringen av experimenten (13). På motsvarande sätt är den biologiska kunskapen viktig vid dataanalysen efter experimentet: den behövs för att särskilja de biologiskt relevanta resultaten och de resultat som har statistisk men inte biologisk relevans. (14). Frågan är hur laboratorerna och forskningsgrupperna bör reagera på detta dilemma. Borde de sätta biologerna på skolbänken och i framtiden förutsätta bioinformatisk kompetens vid anställning av nya forskare? Eller borde de kanske i högre grad anta människor med matematisk eller datavetenskaplig grad till forskarutbildningen? Eller kan man trots allt lösa problemet med utgångspunkt i arbetsfördelningsfilosofin, dvs. genom att anställa flera bioinformatikexperter eller samarbeta med bioinformatiker vid andra institutioner? Eller kan man räkna med att den nya biologins forskningsinstrument inom kort kommer att automatiseras till den grad att bioinformatisk kompetens inte behövs?

Det finns inga givna svar på dessa frågor och praxis varierar från laboratorium till laboratorium. Uppfattningarna om framtiden varierar också. Vissa forskare förutspår att långt över hälften av laboratoriernas personal kommer att utgöras av matematiker, dataexperter, bioinformatiker etc. (intervju med forskare vid Harvard Medical School). Men även det motsatta argumentet har presenterats: en långt driven automatisering kommer

att leda till ett ganska litet behov av den typen av kompetens (intervju med finländsk "toppforskare"). Med tanke på att bioinformatikens integration med biologin är en så central del av de nya forskningsplattformarna, verkar det finnas förvånansvärt lite diskussion om hur det hela skall gå till i praktiken. I stort sett alla jag intervjuat har ansett att integrationen verkligen är ett problem, och de har haft personliga erfarenheter av det. Detta gäller inte minst de bioinformatiker jag intervjuat. En av dem sammanfattar (översättning från finskan) : "Det är väldigt svårt att förklara de här sakernas komplexitet för dem som inte alls förstår sig på databaser och informationsteknik".

Det vore kanske dags att ta tjuren vid hornen.

FD HENRIK BRUUN
TEKNISKA HÖGSKOLAN
MILJÖSKYDDSLABORATORIET
PB 2300
02015 TH
henrik.bruun@hut.fi

REFERENSER

1. Kay LE. Who Wrote the Book of Life? A History of the Genetic Code. Stanford University Press; 2000.
2. Bechtel W. Integrating Sciences by Creating New Disciplines. The Case of Cell Biology. *Biology and Philosophy* 1993;8(3):277-299.
3. Morange M. A History of Molecular Biology. Cambridge, MA, and London, UK: Harvard University Press; 2000 [1998].
4. Fujimura J. Crafting Science. A Sociohistory of the Quest for the Genetics of Cancer. Cambridge, MA, and London, UK: Harvard University Press; 1996.
5. Duyk GM. Sharper Tools and Simpler Methods. *Nature Genetics* 2002;32:465-469.
6. Lander ES. Array of Hope. *Nature Genetics Supplement* 1999;21(1):3.
7. Kallioniemi O. Geenisiruista biosiruihin. *Uuden bioteknikan haasteet ja mahdollisuudet. Duodecim* 2002;118:1149-56.
8. Davies K. Cracking the Genome. Inside the Race to Unlock Human DNA. Baltimore and London: The John Hopkins University Press; 2002 [2001].
9. Mäkelä T, Porkka K. "Omiikat" tulevat - yksi geeni ei enää riitä [The "omics" are coming - one gene is not enough any more]. *Duodecim* 2002;118:1146-8.
10. Keating P, Cambrosio A. Biomedical Platforms. Realigning the Normal and the Pathological in Late-Twentieth-Century Medicine. Cambridge, MA, and London, UK: The MIT Press; 2003.
11. Brown PO, Botstein D. Exploring the New World of the Genome with DNA Microarrays. *Nature Genetics* 1999;21:33-37.
12. Bassett DE, Eisen MB, Boguski MS. Gene Expression Informatics. It's All in Your Mine. *Nature Genetics* 1999;21:51-55.
13. Churchill GA. Fundamentals of Experimental Design for cDNA Microarrays. *Nature Genetics* 2002;32:490-495.
14. Kohane IS, Kho AT, Butte AJ. Microarrays for an Integrative Genomics. Cambridge, MA, and London, UK: The MIT Press; 2003.