

---

# Individualiserad prognosbedömning vid cancer

JOHAN LUNDIN, MIKAEL LUNDIN, JORMA ISOLA OCH HEIKKI JOENSUU

---

*Metodiken för analys av gen- och proteinuttryck i tumörceller och -vävnad har genomgått en utveckling från enskilda experiment till högupplösande, parallella och miniatyriserade analyser. Överbryggande forskning för att studera den kliniska betydelsen av grundforskningsrön har samtidigt effektiviserats. För att testa uppställda hypoteser och dra klinisk nytta av nya biomarkörer och molekylära "fingeravtryck" krävs avancerade bioinformatiska metoder till vilka hör databashantering, sammanlänkning med kliniska uppgifter och en omfattande repertoar av datorstödda analyser. I denna artikel presenteras ett system för webb-baserad, individuellt anpassad prognosbedömning för patienter med bröstcancer*

## BAKGRUND

De flesta cancersjukdomar uppvisar betydande variationer då det gäller tumörens aggressivitet och sjukdomens förlopp, även då graden av utbredning vid tidpunkten för diagnos tagits i beaktande. Tumörer kan t.o.m. histologiskt vara av samma typ, men trots detta förknippade med avsevärda skillnader i behandlingssvar och risk för återfall. Om man kunde utveckla effektivare metoder för att identifiera de patienter som har sämre eller bättre utsikter att dra nytta av en viss behandling, kunde rutinemässig, likriktad och kostsam behandling – "för säkerhets skull" – undvikas. Som exempel kan nämnas att man på basis av nyligen publicerade, randomiserade studier föreslagit att nästan samtliga patienter som insjuknat i bröstcancer borde få tilläggsbehandling utöver den kirurgiska (1). Under rådande incidensprognoser innebär detta att minst var tionde kvinna någon gång under sitt liv kommer att få bröstcancerbehandling i form av behandling med antingen cytostatika eller antiöstrogen i 2–5 år eller båda behandlingarna. Samtidigt vet man att mer än 70 procent av patienter som behandlats med enbart operation är recidivfria ännu efter tio års uppföljning; följdaktligen överbehandlas en avsevärd del av patienterna. Problemet är att man med de prognosbedöm-

ningsmetoder som för närvarande används, endast kan placera en mycket liten del av patienterna i en kategori med så låg risk för återfall att ett utelämnande av tilläggsbehandling är motiverat.

Forskningsaktiviteten när det gäller biologisk karakterisering av cancersjukdomar är mycket hög, och nya faktorer som har samband med behandlingssvaret och prognosen rapporteras i en ständig ström. Analyser av gen- och proteinuttryck i tumörceller och -vävnad har genomgått en utveckling från enskilda experiment till högupplösande, paral-

## FÖRFATTARNA

*MD Johan Lundin är docent i biomedicinsk informatik vid Helsingfors universitet och forskare vid Folkhälsans forskningscentrum, samt vid Cancerkliniken, HUCS*

*MK Mikael Lundin är doktorand vid Kirurgiska kliniken, HUCS, med specialintresse inom webb- och databasprogrammering*

*Jorma Isola är professor i cancerbiologi vid Tammerfors universitet*

---

lella och miniaturiserade metoder. Den överbryggande forskningen för att studera den kliniska betydelsen av grundforskningsrön har samtidigt blivit effektivare. Nya tekniker som exempelvis mikromatriser (cDNA och protein microarrays), ger möjlighet till studium av uttrycksprofiler för tusentals gener eller proteiner vid ett enda experiment (2, 3). Associationer mellan ett visst gen- eller proteinuttryck och en tumörfenotyp eller -undergrupp, kan därefter studeras i hundratals- eller tusentals tumörpreparat med hjälp av vävnadsmikromatriser (tissue microarrays) (4).

För att testa uppställda hypoteser och dra klinisk nytta av tumörbiologiska och molekylära "fingeravtryck" krävs avancerade bioinformatiska metoder till vilka hör databashantering, sammanlänkning med kliniska uppgifter och en omfattande repertoar av statistiska (biostatistics) och datatekniska analyser (data mining). De bioinformatiska analyserna i sig kommer i framtiden att vara av största betydelse, och en stor del av framtidens hypoteser kommer att testas på gen- och proteindata, där själva laboratorieexperimentet gjorts i ett tidigare skede eller av en annan forskargrupp. De forskare som ursprungligen utfört experimentet har inte haft kapacitet att analysera de enorma mängder data som alstras t.ex. vid mikromatris hybridiseringar, men har gjort resultaten tillgängliga i databaser för ytterligare experiment som görs av andra forskare. För dessa "omedelbara experiment" (instant experiments) har även myntats uttrycket *in silico* för att beskriva analyser där nya biostatistiska analyser görs på redan existerande databaser.

För att testa hypoteser, klassificera tumörer och förutspå sjukdomsförlopp *in silico* används konventionell statistik, som korrelations- och regressionsanalyser, men även metoder som härstammar från forskningen kring artificiell intelligens, som artificiella neurala nätverk. Dessa informationstekniska hjälpmedel har framgångsrikt använts för igenkännandet av mönster och prognosbedömning i komplexa uppsättningar av data. Man har nyligen med hjälp av högupplösande molekylära analyser i kombination med bioinformatiska metoder lyckats identifiera prognostiska undergrupper i flera olika cancerformer (5-9).

Omsorgsfullt sammanställda kliniska databaser som innehåller noggranna uppgifter om patienten, diagnosen, de patologiska och radiologiska fynden, behandlingen och behandlingsresponsen, samt uppföljningen är ytterst

värdefulla. Dessa är en förutsättning för att man skall kunna testa de kliniskt relaterade hypoteser som uppställs utifrån molekylärbiologiska analyser av tumörvävnadsprover. Nya datatekniska framsteg gör det även möjligt att foga t.ex. digitala bilder av histopatologiska snitt och radiologiskt bildmaterial till databaserna för att ytterligare utöka den framtida nyttan av dessa.

Inom en riksomfattande bröstcancerstudie har vår forskningsgrupp utvecklat modeller för en mer individualiserad prognosbedömning. Dessa modeller baserar sig på patientdatabaser med noggranna kliniska, patologiska, terapi- och uppföljningsrelaterade uppgifter sammanlänkade med molekylärbiologiska data. Nedan beskrivs en metod kallad "case-match" som kombinerar information från ett flertal enskilda prognosfaktorer och beräknar genomsnittlig överlevnad på basis av tidigare patientfall med en viss prognosfaktorprofil. Denna metod kan användas av forskare som vill studera prognosen för olika undergrupper enligt tumörbiologisk profil. Den lämpar sig även som stöd för kliniker som genom att mata in uppgifter om en ny patient kan erhålla ett prognosestimat baserat på tidigare liknande patientfall. Målet är att med hjälp av ovannämnda tumörbiologiska data och bioinformatiska metoder uppnå en mer individuell bedömning av sjukdomens prognos, risken för återfall och den mest lämpliga behandlingen.

#### DET WEBB-BASERADE SYSTEMET FÖR UPPSKATTNING AV PROGNOSEN

Det har skapats en webbplats för uppskattning av prognos på adressen <http://www.finprog.org>. Man får tillgång till prognosmodellerna efter att ha intygat att man har behörig utbildning och kunskaper för att förstå det medicinska innehållet på webbplatsen. Ett urval prognostiska faktorer finns tillgängliga för s.k. casematch ("patientfallsmatch") uppskattning av prognosen (Figur 1). Det förhandsinställda valet för varje faktor är "All" vilket betyder att denna variabel inte har specificerats. Användaren kan sedan mata in prognosfaktorer för en ny bröstcancerpatient genom att på rullgardinsmenyerna välja kategorier som motsvarar den aktuella patientens sjukdomsprofil och därefter klicka "calculate". En fråga (query) körs då automatiskt för att i databasen plocka ut alla de tidigare patientfall som har en likadan prognosfaktorprofil och för vilka utgången är känd. En Kap-

---

lan-Meier överlevnadskurva kalkyleras omedelbart för dessa matchande patienter och framställs grafiskt samt numeriskt i tabellform med procentuell överlevnad vid bestämda tidpunkter. Endast en exakt matchning enligt de på förhand definierade prognosfaktorkategorierna i rullgardinsmenyerna genomförs, således används inte någon nearest neighbor-teknik eller liknande.

På webbplatsen kan även två prognosfaktorprofiler jämföras visuellt och statistiskt. I tvåprofilsläget jämförs två överlevnadskurvor statistiskt med ett logrank test och resultatet presenteras i form av ett chi-square värde, samt motsvarande p-värde. Fördelningen av de matchande patienterna enligt överlevnadsstatus, erhållen terapi eller någon av de andra klinisk-patologiska faktorer som finns registrerade i databasen kan även visas i form av en tabell eller diagram (Figur 1). Utöver prognosuppskattningssystemet innehåller webbplatsen basinformation om överlevnadsanalys, prognostiska faktorer, riktlinjer för val av faktorer samt förklaring av hur resultaten i modellen skall tolkas.

#### BRÖSTCANCERDATABASEN FINPROG

Den ovan nämnda casematch-metoden för uppskattning av prognos kan appliceras på vilken klinisk databas som helst som innehåller uppföljningsuppgifter. På webbplatsen <http://www.finprog.org> har systemet kopplats till en bröstcancerdatabas som har sammanställts inom den s.k. FinProg-studien. Samtliga universitetssjukhus och ett större antal andra sjukhus i Finland deltar i studien, som omfattar mer än hälften av de ca 6 000 patienter som fick diagnosen bröstcancer under åren 1991 och 1992. Om dessa patienter har insamlats data som sammanlagt utgör ett femtiotal faktorer med potentiell betydelse för patientens prognos, bl.a. ålder, tumörstorlek, tumörtyp, differentieringsgrad, genomförd behandling, samt ett flertal biomarkörer. Den registrerade uppföljningstiden är för tillfället 9,5 år. Vävnadsprover tagna vid diagnostillfället har även insamlats och tumörvävnadsmikromatriser tillverkats som omfattar ca 2 000 av FinProg-patienterna. Mikromatriserna har analyserats med ett flertal tillgängliga markörtest med potentiell prognostisk betydelse. Resultaten för följande tumörbiologiska faktorer har redan inregistrerats i databasen: östrogen och progesteronreceptorer, immunhistokemisk bestämning av HER-2 (10), Herceptest (kommersiellt HER-2 test), chro-

mogen *in situ*-hybridisering av HER-2, immunhistokemisk bestämning av Cox-2 (cyclooxygenas 2) (11), ezrin, p53, Ki-67, ER- $\beta$ , xantinoxidas och SnoN (nukleärt protoonkogen protein)(12). Alla personuppgifter och data som kunde göra det möjligt att identifiera enskilda patienter har avlägsnats och ersatts med interna koder.

Den recidivfria överlevnaden räknas från diagnostillfället till upptäckten av distant metastas och bröstcancerspecifik överlevnad från diagnos till död orsakad av bröstcancer.

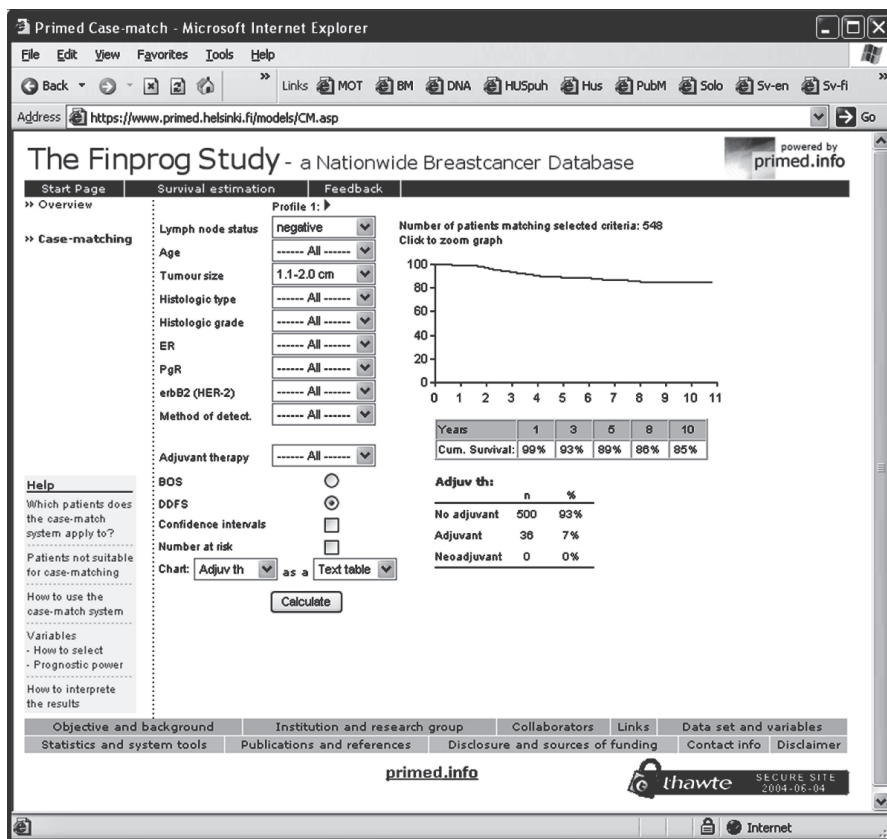
#### SYNPUNKTER PÅ CASEMATCH-METODEN

Den presenterade casematch-systemet innebär att forskare och kliniker via den ovan nämnda webbplatsen får tillgång till överlevnadsestimat och tumörbiologiska karakteristika för ett stort antal tidigare patientfall. Dessa estimat och data kan användas för forskningsändamål men även som stöd t.ex. vid val av terapi. Systemet kan ses som ett slags artificiellt minne, där användaren definierar en prognosfaktorprofil och "återkallar" alla de patientfall som överensstämmer med den valda profilen. Det kan även ses som ett svar på önskemål om ökad öppenhet när det gäller kliniska databaser. Forskare från andra institutioner och länder ges möjlighet att själva gå in och utforska databasen, dock utan att själva grunddata överläts.

Till fördelarna jämfört med traditionella prognostiska modeller så som index eller nomogram i tryckt form (t.ex. TNM-indexet för cancerpatienter) hör att användaren erhåller en överlevnadskurva för hela den uppföljningstid som finns registrerad i databasen och inte enbart estimat för en viss tidpunkt (t.ex. 10-års överlevnad). Estimaten baserar sig inte heller på någon approximation eller regressionsformel, utan direkt på genomsnittlig överlevnad för de patienter som matchar den valda profilen. Även det att systemet är åtkomligt med en vanlig webbläsare (browser) och inte kräver installation av något specifikt program kan ses som en fördel.

Systemet kan länkas till andra databaser och på detta sätt kunde antalet patientfall som har en mer sällsynt tumörbiologisk profil eller som fått viss terapi utökas. Det är givetvis även möjligt att tillämpa metoden vid andra cancerformer och andra sjukdomar där överlevnadsestimat används.

Ett problem vid uppskattning av överlevnad med den presenterade casematch-metoden är att Kaplan-Meier-estimatens tillförlit-



Figur 1. Bildskärmsavbildning av casematch-systemet för uppskattning av prognos i bröstcancer

lighet givetvis är fullständigt beroende av kvaliteten på den databas som systemet är kopplat till. Endast välbeskrivna och kvalitetsgranskade patientmaterial kan användas, och även estimat baserade på stora, oselektade databaser kräver en sund kritisk inställning av användaren. För att sammanställa databasen med överlevnadsuppgifter har en kortare eller längre uppföljningstid varit nödvändig, vilket i sin tur gör att den dagsaktuella situationen för prognosen inte nödvändigtvis är densamma. Förändringar i diagnostik och terapi måste beaktas av användaren, och bl.a. därför är systemet i sin nuvarande form inte lämpat att användas av personer utan sakkunskap på området, t.ex. patienterna själva.

Det är också givet att ju mer specifik profil av prognosfaktorer som anges av användaren, desto färre patienter kommer att matcha denna profil i databasen. I detta fall kunde en prognosmodell baserad på en regressionsformel tänkas ge ett bättre estimat, speciellt för sällsynta profiler. Å andra sedan måste man minnas att regressionsmodeller även konstruerats på basen av patientdata, och då

har sällsynta profiler sannolikt haft en mycket liten inverkan på parametrarna i regressionen. Sålunda kan kanske en prognostisk modell baserad på regressionsekvation inte ge ett mer exakt estimat än det man kan få genom att exempelvis med casematch-metoden granska tidigare patientfall i en stor databas. Då gäller det förvisso att analysera resultaten noggrant och hålla ett öga på antalet matchande patienter och konfidensintervallen för överlevnadsestimaten.

#### VIDAREUTVECKLING AV METODERNA

Den webbplats som är länkad till bröstcancerdatabasen FinProg öppnades i januari 2003 har och har för närvarande ca 5 000 användare per månad från olika delar av världen (13). Vi planerar som bäst en enkät som i första hand skall riktas till användare i Finland, för att utreda hur och i vilka situationer forskare och kliniker använder sig av systemet.

I övrigt pågår vidareutveckling på flera plan. Då det gäller bröstcancerdatabasen studeras nya biomarkörers inverkan på behandlings-



---

ressvar och -prognos. Resultaten fogas till databasen, som för närvarande innehåller data på mer än 50 olika prognosfaktorer. Vidare studeras användningen av mer avancerade statistiska modeller och artificiella neurala nätverk för att förbättra träffsäkerhet i klassificering och prognosbedömning.

En utvidgning av databasen till att omfatta allt fler fall av bröstcancer – men även övriga cancerformer – är en av huvudstrategierna för tillfället. Detta gäller inte enbart en utökning av patientdata från Finland utan även från andra länder. På så vis kunde allt fler sällsynta fall inkluderas och öka värdet av metoden. Vi

har tagit initiativ till ett europeiskt samarbetsprojekt och har även samarbetspartner i USA.

Sammanfattningsvis kan sägas att webb-baserad tillgång till välkarakteriserade och omfattande kliniska databaser utgör ett nytt verktyg för forskare och kliniker. I denna artikel har vi beskrivit en metod som gör det möjligt för användaren att definiera en sjukdomsprofil och omedelbart erhålla en uppskattad prognos för de patientfall i en databas som överensstämmer med den valda profilen. Detta verktyg kunde potentiellt fungera som ett stöd vid val av behandling och medföra en mer individuellt anpassad vård.

DOCENT JOHAN LUNDIN  
FORSKNINGSGRUPPEN FÖR BIOMEDICINSK  
INFORMATIK  
AVDELNINGEN FÖR CANCERSJUKDOMAR  
INSTITUTIONEN FÖR KLINISK MEDICIN  
HELSINGFORS UNIVERSITET  
HUCS INSTITUT,  
PB 105, 00290 HELSINGFORS  
johan.lundin@helsinki.fi

MK MIKAEL LUNDIN  
FORSKNINGSGRUPPEN FÖR BIOMEDICINSK  
INFORMATIK  
AVDELNINGEN FÖR CANCERSJUKDOMAR  
INSTITUTIONEN FÖR KLINISK MEDICIN  
HELSINGFORS UNIVERSITET  
HUCS INSTITUT, PB 105,  
00290 HELSINGFORS  
mikael.lundin@helsinki.fi

PROFESSOR JORMA ISOLA  
INSTITUTET FÖR MEDICINSK TEKNOLOGI  
33014 TAMMERFORS UNIVERSITET  
jorma.isola@uta.fi

PROFESSOR HEIKKI JOENSUU  
AVDELNINGEN FÖR CANCERSJUKDOMAR  
INSTITUTIONEN FÖR KLINISK MEDICIN  
HELSINGFORS UNIVERSITET  
PB 180, 00290 HELSINGFORS  
heikki.joensuu@hus.fi

## REFERENSER

1. Goldhirsch A, Wood WC, Gelber RD, Coates AS, Thurlimann B, Senn HJ. Meeting highlights: updated international expert consensus on the primary therapy of early breast cancer. *J Clin Oncol* 2003;21(17):3557-65.
2. Quackenbush J. Computational analysis of microarray data. *Nat Rev Genet* 2001;2(6):418-27.
3. Lakhani SR, Ashworth A. Microarray and histopathological analysis of tumours: the future and the past? *Nat Rev Cancer* 2001;1(2):151-7.
4. Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, et al. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* 1998;4(7):844-7.
5. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286(5439):531-7.
6. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;98(19):10869-74.
7. Dhanasekaran SM, Barrette TR, Ghosh D, Shah R, Varambally S, Kurachi K, et al. Delineation of prognostic biomarkers in prostate cancer. *Nature* 2001;412(6849):822-6.
8. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415(6871):530-6.
9. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003;100(14):8418-25.
10. Joensuu H, Isola J, Lundin M, Salminen T, Holli K, Kataja V, et al. Amplification of erbB2 and erbB2 Expression Are Superior to Estrogen Receptor Status As Risk Factors for Distant Recurrence in pT1N0M0 Breast Cancer: A Nationwide Population-based Study. *Clin Cancer Res* 2003;9(3):923-30.
11. Ristimäki A, Sivula A, Lundin J, Lundin M, Salminen T, Haglund C, et al. Prognostic significance of elevated cyclooxygenase-2 expression in breast cancer. *Cancer Res* 2002;62(3):632-5.
12. Zhang F, Lundin M, Ristimäki A, Heikkilä P, Lundin J, Isola J, et al. Ski-related novel protein N (SnoN), a negative controller of transforming growth factor-beta signaling, is a prognostic marker in estrogen receptor-positive breast carcinomas. *Cancer Res* 2003;63(16):5005-10.
13. Lundin J, Lundin M, Isola J, Joensuu H. A web-based system for individualized survival estimation in breast cancer. *BMJ* 2003;326:29.