

Gebruik van microroosters in de klinische oncologie

F. DE SMET ^{1,3}, K. MARCHAL ¹, D. TIMMERMAN ², I. VERGOTE ², B. DE MOOR ¹, Y. MOREAU ¹

¹ ESAT-SISTA/COSIC/DocArch, K.U.Leuven

² Dienst Gynaecologie-Verloskunde, Universitaire Ziekenhuizen, Leuven

³ Correspondentieadres: ir. dr. F. De Smet, ESAT-SISTA/COSIC/DocArch, K.U.Leuven, Kasteelpark Arenberg 10, 3001 Leuven-Heverlee; e-mail: frank.desmet@esat.kuleuven.ac.be

Samenvatting

Microroosters of microarrays laten toe het expressieniveau van duizenden genen in een cel tegelijkertijd te bepalen. Een van de toepassingsdomeinen van deze technologie is de oncologie.

Omdat het ontregelde expressiepatroon van een groot aantal genen aan de basis ligt van het gedrag van kwaadaardige processen, is het mogelijk de kennis van deze patronen te gebruiken om de klinische besluitvorming te helpen sturen. Vermits de analyse van expressieniveaus van tumoren met microroosters voor één enkele patiënt duizenden waarden oplevert, is het nodig om bepaalde procedures of methoden te volgen, die uit deze enorme hoeveelheid gegevens de klinisch relevante informatie halen.

Inleiding

Recente technologische ontwikkelingen in de moleculaire biologie hebben het mogelijk gemaakt het expressieniveau (d.i. een maat voor de transcriptie of de hoeveelheid mRNA dat van een bepaald gen wordt aangemaakt) van duizenden genen in een cel simultaan te meten. Een van deze technologische ontwikkelingen zijn de microroosters (ook wel microarrays genoemd) (1-3).

Gedurende de laatste maanden zijn verschillende publicaties verschenen die aantonen hoe gegevens, bekomen met deze microroosters, zouden kunnen gebruikt worden voor klinische doeleinden - veelal in de oncologie (4-15). Deze gegevens kunnen, eventueel samen met andere klinische parameters, als hulp aangewend worden om het proces van diagnose, bepaling van prognose en therapieplanning van kwaadaardige processen mee te sturen. De data afkomstig van experimenten met microroosters kunnen eventueel ook gebruikt worden om klassen of subtypen te ontdekken in bepaalde categorieën tumoren (bijvoorbeeld voor non-Hodgkin-lymfomen (4) of borstcarcinomen (13)). Deze klassen kunnen overeenkomen met een gekende indeling (die dan gebaseerd is op de meer klassieke parameters, bijvoorbeeld de gegevens verkregen door het histopathologisch onderzoek), maar het is ook mogelijk dat via deze weg een nieuwe classificatie wordt gevonden die toelaat patiënten meer gericht te gaan behandelen of het verloop van de ziekte beter te voorspellen (overeenstemmend met het herschikken van diagnostische categorieën). De hoge kostprijs van deze technologie is voorlopig een van de beperkende factoren die tot op heden het routinematig gebruik voor klinische doeleinden heeft verhinderd. Wanneer deze technologie in de toekomst op grote schaal zal worden gebruikt, valt het te verwachten dat de kostprijs ervan zeker zal dalen, waardoor deze techniek het experimenteel stadium zal ontgroeien en binnen het bereik komt van de meeste artsen (8).

Ook voor het wetenschappelijk onderzoek naar het ontstaan van kanker (carcinogenese) kunnen de gegevens bekomen uit metingen van expressieniveaus van onschatbare waarde zijn. Zo kan er op zoek gegaan worden naar bepaalde genen die differentieel tot expressie komen wanneer bijvoorbeeld gezond weefsel vergeleken wordt met tumorale weefsel (12, 16-23) of wanneer weefsel van tumoren met en zonder invasief of metastatisch fenotype (17, 19, 24, 25) vergeleken wordt. Het is dus mogelijk dat op deze manier

de gebeurtenissen die de transformatie naar een kwaadaardig fenotype veroorzaken, kunnen worden achterhaald. Op deze manier zouden bijvoorbeeld ook genen (of hun proteïnen) kunnen geïdentificeerd worden die goede kandidaten zijn om als doelwit te dienen voor bepaalde geneesmiddelen (26). Meer algemeen is het zelfs zo dat de functie van de meeste genen in de cel nog niet is gekend. Dankzij de gegevens beschikbaar gesteld door het Humaan Genoom Project (27), zullen deze genen, waarvan de plaats in het geheel van de moleculaire interacties nog niet is achterhaald, nog in aantal toenemen. Gegevens afkomstig uit experimenten met microroosters kunnen de onderzoekers helpen om deze moleculaire interacties te ontcijferen (afleiden van genetische netwerken uit expressedata (28)). Dit zal dan op zijn beurt nieuwe perspectieven openen voor het gebruik van deze technologie in de kliniek.

Het bepalen van expressieniveaus van genen was vroeger reeds mogelijk (bijvoorbeeld Northern blot), maar de grote kracht van microroosters bestaat erin dat deze bepaling op redelijk eenvoudige wijze voor *duizenden genen tegelijk* kan gebeuren. Dit wil zeggen dat voor ieder staal (bijvoorbeeld biopsie van een tumor) dat wordt geanalyseerd op een microrooster, duizenden getallen (één per gemeten gen) als resultaat verkregen worden. Het spreekt voor zich dat het niet eenvoudig is om beslissingen te nemen of voorspellingen te doen aan de hand van deze enorme hoeveelheid data. Er zal dus een zekere verwerking van deze gegevens moeten gebeuren, welke tot het vakgebied van de bio-informatica behoort (29).

In deze tekst zullen enkele van de methoden besproken worden die het mogelijk maken om uit metingen met microroosters klinisch relevante gegevens af te leiden. Dit zal o.a. geïllustreerd worden met een eenvoudig voorbeeld uit de hematologie (diagnose van acute leukemie). Er zal ook een korte beschrijving worden gegeven van de microroostertechnologie zelf en een link worden gelegd naar de rol die wijzigingen in expressieniveaus spelen in het ontstaan van kanker. Ten slotte zullen enkele kritische bedenkingen worden gemaakt.

Verband met carcinogenese

Het ontstaan van kanker is een proces dat zich voor een groot deel afspeelt op het niveau van het genoom. Onder invloed van bepaalde factoren (bestraling, virale infecties, ...) kunnen mutaties ontstaan in bepaalde genen (bijvoorbeeld proto-oncogenen en tumorsuppressorgenen) met eventueel ongecontroleerde celgroei en de mogelijkheid tot invasie en metastasering tot gevolg. Door deze mutaties kan er ook bij

andere genen (waarin geen mutatie optreedt, maar waarvan de expressie direct of indirect door het product of de proteïne van een gemuteerd gen wordt geregeld, bijvoorbeeld als het gemuteerd gen codeert voor een transcriptiefactor) verstoring van hun expressie optreden. Het is nu de verzameling van deze ontregelde genexpressies die het fenotype of gedrag van de tumorcel bepaalt (30). Met dit fenotype kunnen o.a. de patholoog-anatomische diagnose, het stadium, de histopathologische differentiatiegraad, de prognose, het antwoord op therapie, ... van de tumor worden bedoeld. Het meten van een groot gedeelte van deze expressieniveaus zou dus van grote waarde kunnen zijn om het werkelijk gedrag van de tumorcellen te kennen, te voorspellen en te begrijpen. Door de expressieniveaus te betrekken in de klinische besluitvorming, zouden de fundamentele processen die aan de basis liggen van de carcinogenese mee in rekening kunnen worden gebracht om het beleid te bepalen. Dit zou een verbetering betekenen in vergelijking met de meer empirische beslissingsschema's die nu gebruikt worden.

Het zijn nu de microroosters die ons precies in staat stellen om expressieniveaus van genen te meten.

Microroostertechnologie

Microroosters bestaan uit een groot aantal sondes samengebracht op een klein oppervlak. Sterk vereenvoudigd kan gesteld worden dat ieder van deze sondes bestaat uit DNA dat complementair is aan één welbepaalde mRNA-streng (ze zijn dus specifiek voor één welbepaald gen). Iedere mRNA-streng zal dus specifiek binden aan (of hybridiseren met) zijn complementaire sonde(s) wanneer het totaal mRNA, afkomstig uit cellen van een welbepaald celtype, in contact wordt gebracht met de sondes op het microrooster. De binding van iedere complementaire sonde met zijn overeenkomstig mRNA kan gemeten worden en is dus een maat voor de hoeveelheid mRNA (expressieniveau) afkomstig van één welbepaald gen. Door het groot aantal sondes aanwezig op het microrooster kan dus het expressieniveau van duizenden genen tegelijk worden gemeten. De keuze van de genen of sondes wordt meestal bepaald door het probleem dat wordt bestudeerd (bijvoorbeeld weefselspecifieke genen).

Een onderscheid kan worden gemaakt tussen twee verschillende soorten microroosters:

cDNA-microroosters (31):

Deze microroosters zijn kleine glazen plaatjes waarop kleine hoeveelheden, vooraf gemaakte, DNA-strengen worden gedrukt of “gestippeld”. Deze “DNA-vlekjes” zullen dan dienst doen als de sondes (zoals hierboven beschreven, één vlekje per gen). De DNA-strengen zijn hier gewoonlijk enkele honderden basenparen lang en afgeleid van delen van het genoom waarvan geweten is dat ze tot expressie komen. Bij deze techniek worden altijd twee stalen tegelijk op het microrooster gehybridiseerd (één referentiestaal en één teststaal). Als uiteindelijk resultaat wordt dan voor ieder gen het relatieve expressieniveau genomen (expressietest / expressiereferentie). Ter verduidelijking wordt in fig. 1 een schematisch overzicht gegeven van de procedure die gevolgd kan worden bij een experiment met een cDNA-microrooster.

Oligonucleotideroosters (ook wel DNA-chips genoemd - GeneChip[®], Affymetrix Inc.) (32):

Hier worden de oligonucleotiden (= sondes) in situ (dus op het microrooster zelf) gesynthetiseerd met behulp van technieken die eveneens gebruikt worden bij de productie van elektronische chips. Hierdoor bereikt men een enorm hoge densiteit aan sondes. Ieder gen wordt hier op een unieke wijze vertegenwoordigd door 15 tot 20 verschillende oligonucleotiden van 25 basenparen lang. Ter controle worden ook enkele oligonucleotiden toegevoegd die op één basenpaar na exact gelijk zijn aan de vorige oligonucleotiden. Deze controles laten toe een schatting te maken van de kruishybridisatie en de lokaal aanwezige ruis. Ieder staal wordt hier op een aparte chip gehybridiseerd (hier worden dus geen relatieve maar absolute expressieniveaus verkregen).

Doelstellingen en methodologie

voor het gebruik in de klinische oncologie

Zoals reeds vermeld, kan men bij de studie van tumorstalen met microroosters niet onmiddellijk besluiten trekken uit de enorme hoeveelheid gegevens (één waarde per gen of sonde) die hieruit resulteren. Dit is bijvoorbeeld te vergelijken met een uitslag in de klinische biologie waarop de resultaten van enkele duizenden tests (als die al zouden bestaan) zouden staan. Het is dus duidelijk dat er, om uit deze gegevens klinisch bruikbare resultaten te halen, bepaalde procedures moeten worden gevolgd (29). Deze procedures vinden meestal hun oorsprong in de statistiek of in een relatief nieuwe wetenschap die “bio-informatica”

wordt genoemd; deze discipline heeft dan weer sterke banden met technieken die worden gebruikt in “data mining”, d.i. de wetenschap die zich bezighoudt met de analyse van grote verzamelingen gegevens.

Tijdens het bespreken van de verschillende doelstellingen en methodologieën zullen we, ter illustratie, gebruikmaken van een verzameling van 72 patiënten met acute lymfatische (ALL, in deze tekst wordt gesteld dat deze patiënten tot klasse 1 behoren) of acute myeloïde (AML, in deze tekst behoren deze patiënten tot klasse 2) leukemie, wier bloed- of beenmergstalen werden geanalyseerd met een oligonucleotiderooster van Affymetrix (waarop sondes aanwezig waren voor een 7000-tal genen). De resultaten van deze metingen, zoals die geproduceerd werden door Golub et al. (9), zijn publiek beschikbaar op het internet (33). De analyse van de gegevens, zoals ze in deze tekst wordt getoond, werd op onze dienst uitgevoerd. De 72 patiënten werden opgesplitst in 2 groepen: één groep die we de trainingsset noemen (met de eerste 38 patiënten van wie 27 met ALL en 11 met AML) en één groep die we de testset noemen (met de laatste 34 patiënten van wie 20 met ALL en 14 met AML). Merk nogmaals op dat voor iedere patiënt ongeveer 7000 verschillende expressieniveaus gemeten werden. Merk ook op dat iedere patiënt een kenteken (ook label genoemd) draagt (d.i. de uiteindelijke diagnose: ALL of AML – klasse 1 of klasse 2).

De gegevens afkomstig van microroosters kunnen voor de volgende doelstellingen worden gebruikt:

Selectie van kenmerken

Een eerste doelstelling is het verminderen van het aantal gegevens (of waarden) per patiënt of per meting op het microrooster. Dit wordt ook het probleem van de afname van de dimensionaliteit genoemd. Voor de dataset bestaande uit de patiënten met acute leukemie is de dimensionaliteit zonder afname ongeveer 7000. Deze afname is meestal noodzakelijk vooraleer gestart kan worden met het maken van voorspellingen of het ontdekken van klassen (zie volgende doelstellingen). Belangrijk is echter dat deze afname gepaard gaat met een minimaal verlies aan essentiële informatie. Enkel de meest essentiële kenmerken, nodig voor het bestuderen van een bepaald probleem, moeten worden geselecteerd. Deze selectie kan het best op een van de twee volgende manieren gebeuren:

Selectie van individuele genen:

De meest eenvoudige manier is de selectie van individuele genen waarvan de expressie het best gecorreleerd is met een bepaald klassenverschil, waarin men op een bepaald moment geïnteresseerd is. Hiervoor moeten de verschillende klassen en de classificatie van ten minste een deel van de patiënten (bijvoorbeeld kentekens van de trainingsset) reeds gekend zijn, wat niet zo is wanneer men de verschillende tumorklassen juist wil gaan bepalen – zie laatste doelstelling. Deze selectie is logisch vermits niet alle genen een expressiepatroon hebben dat informatie bevat over een bepaald klassenverschil zodat deze genen kunnen worden weggelaten. Zo zouden bijvoorbeeld voor de patiënten met acute leukemie de 10 genen kunnen worden geselecteerd waarvan het expressieniveau tussen de 2 klassen (ALL en AML) in de trainingsset het meest beduidend verschilt. De andere genen worden, voor dit probleem, gewoon buiten beschouwing gelaten. Dit komt dan overeen met een dimensionaliteitsreductie van 7000 naar 10. Op deze manier kunnen ook onmiddellijk de genen worden geïdentificeerd die een rol spelen bij een bepaald klassenverschil, wat eventueel zou kunnen helpen om hun rol in de carcinogenese te bepalen. In de literatuur kan men reeds een groot aantal artikels vinden die op een dergelijke manier genen trachten te identificeren. Dit gebeurt aan de hand van metingen van expressieniveaus (soms gemeten met andere technieken dan microarrays) in cellen onder verschillende omstandigheden. Deze verschillende omstandigheden komen overeen met de verschillende klassen (bijvoorbeeld cellen met een metastatisch fenotype versus cellen zonder metastatisch fenotype (17, 19, 24, 25), gezonde cellen versus tumorale cellen (12, 16-23), cellen van tumoren met een verschillende prognose (4), cellen van tumoren met verschillende histopathologische diagnose (9, 13), cellen van tumoren met een verschillende respons op therapie (9, 34, 35), cellen met een bepaalde mutatie versus cellen zonder een bepaalde mutatie (36), ...).

Xu et al. bijvoorbeeld, hebben, onder andere door screening van verschillende weefsels met microarrays, een drietal genen (*P503S*, *P504S* en *P510S*) geïdentificeerd die een verhoogde expressie vertonen in normaal en/of tumorale prostaatweefsel (23). *P504S* zou een gen zijn dat specifiek tot expressie komt in sommige prostaattumoren (vergelijk met de veelgebruikte tumormarker PSA die dikwijls ook in verhoogde hoeveelheden kan voorkomen bij benigne prostaathypertrofie en prostatitis). Dergelijke kanker-specifieke genen zouden eventueel geschikte kandidaten kunnen zijn om te dienen als merkers voor screening, diagnose, evaluatie van therapierespons en opsporing van recidief. Ze zouden zelfs als doelwit

kunnen dienen voor bepaalde therapeutische interventies (bijvoorbeeld gebruik van antilichamen of therapeutische vaccins).

Een gelijkaardig resultaat werd door Wang et al. voor ovariumcarcinomen gemeld (21). Zij vergeleken de genexpressie in normale ovaria met de genexpressie in ovariumtumoren door middel van een cDNA-microrooster, waarop het expressieniveau van 5766 genen werd gemeten. Hun onderzoek resulteerde in de identificatie van een stel genen (waaronder bijvoorbeeld *HE4*, een mogelijke merker voor ovariumcarcinomen – zie ook (18)) waarvan het expressieniveau verschillend was in tumoraal versus normaal ovariaal weefsel.

Selectie van combinaties van genen:

Een tweede methode om het probleem van de afname van dimensionaliteit op te lossen is de selectie van verschillende (lineaire of niet-lineaire) combinaties van genexpressies. Ieder van deze combinaties kan dan beschouwd worden als één enkel kenmerk. Merk op dat in de vorige methode ieder kenmerk overeenkomt met één enkel expressieniveau. Omdat het onderscheid tussen verschillende klassen over het algemeen niet volledig zal bepaald worden door de individuele activiteit van een gen, maar door de interactie van verschillende genen, is het beter om met combinaties van genexpressies te werken.

Een van deze methoden voor het bepalen van geschikte combinaties van genexpressies is de “Principale Component Analyse” (PCA) (37). Kort samengevat kan men zeggen dat PCA lineaire combinaties van genexpressies zoekt zodanig dat deze een maximale spreiding hebben over een verzameling van patiënten waarvoor metingen met een microrooster werden uitgevoerd. PCA zoekt als het ware de combinaties die het meest “informatief” zijn voor een verzameling patiënten. Deze combinaties (of kenmerken) worden dan de principale componenten genoemd van deze verzameling patiënten. Zo kunnen voor de trainingsset, zoals die hierboven werd beschreven (acute leukemie), twee lineaire combinaties van de gemeten genexpressies (= principale componenten) worden bepaald waarvoor de spreiding maximaal is over deze trainingsset. De twee combinaties van genexpressies kunnen dan uitgerekend worden voor iedere patiënt afzonderlijk (zowel voor patiënten van de trainings- en de testset), zodanig dat men voor iedere patiënt twee waarden of kenmerken volgens de principale componenten van de trainingsset verkrijgt. In dit geval is er dus sprake van een afname van de dimensionaliteit van 7000 naar 2 (iedere patiënt komt uiteindelijk overeen met een koppel getallen dat men kan uitzetten in een grafiek – zie fig. 2). Deze

methode waarborgt bovendien dat deze afname gebeurt met een minimaal verlies aan inlichtingen. Merk ook op dat voor deze methode de verschillende tumorklassen en de classificatie van de verschillende patiënten niet op voorhand moet gekend zijn (dit wordt ook een niet-gesuperviseerde methode genoemd, terwijl de selectie van individuele genen in de vorige methode wel gesuperviseerd is, d.w.z. dat de kentekens of labels moeten gekend zijn). Dit is dus een geschikte methode om juist de tumorklassen te bepalen – zie laatste doelstelling.

Maken van voorspellingen

In een klinische omgeving is het belangrijk dat, aan de hand van metingen met microroosters (eventueel aangevuld met andere klinische gegevens), voor individuele patiënten voorspellingen kunnen worden gedaan i.v.m. prognose, antwoord op therapie, stadiumbepaling, histopathologische diagnose, ... Dit gebeurt door middel van modellen die aan de hand van de in het vorige punt geselecteerde kenmerken de tumor trachten te classificeren. De parameters van het model moeten worden bepaald aan de hand van een verzameling patiënten van wie reeds geweten is tot welke klasse ze behoren (m.a.w. patiënten voor wie bijvoorbeeld de stadiumbepaling, histopathologische diagnose, prognose, effect van therapie, ... reeds gekend zijn). Deze verzameling van patiënten wordt de trainingsset genoemd, waarvan gesteld wordt dat ze gebruikt wordt om het model te trainen (= bepaling van de parameters van het model). Het getrainde model kan achteraf aangewend worden om voorspellingen te doen voor patiënten van wie de classificatie nog niet gekend is (deze zijn de patiënten van de testset). Enkele voorbeelden van veel gebruikte modellen zijn: Artificiële Neurale Netwerken (38-40), Bayesiaanse Netwerken (41), Lineaire Discriminant Analyse (38) en Support Vector Machines (42).

Na bepaling van de eerste twee principale componenten van de trainingsset kan bijvoorbeeld, voor de verzameling patiënten met acute leukemie, een rechte bepaald worden die de patiënten van de trainingsset optimaal scheidt in patiënten met ALL en patiënten met AML (fig. 2). In dit geval is de rechte een voorstelling van het model en de coëfficiënten van de vergelijking van de rechte zijn dan de parameters van het model, die bepaald worden aan de hand van de trainingsset. Door middel van deze rechte kan een voorspelling gemaakt worden voor de patiënten die tot de testset behoren. Dit is weergegeven in fig. 3 waar de patiënten van de testset worden voorgesteld door punten die bepaald werden volgens de eerste twee principale componenten van de trainingsset. Het model classificeert de punten die boven de rechte liggen

als ALL en de punten eronder als AML. Zoals kan gezien worden in fig. 3, resulteert dit in 3 misclassificaties. Door een gelijkaardige procedure, waar met 5 principale componenten gewerkt wordt, zou men het aantal misclassificaties kunnen verminderen tot 2. Door het gebruik van niet-lineaire modellen (bijvoorbeeld Artificiële Neurale Netwerken - de rechte wordt dan vervangen door een gekromde curve) kan men dit aantal misclassificaties nog verder terugdringen.

Volgens de hierboven beschreven methode kunnen modellen ontwikkeld worden waarin metingen op microroosters gebruikt worden om de arts te helpen bij de diagnose of waarmee gepoogd wordt een voorspelling te maken van het klinisch verloop van een tumoraal proces. *Enkele voorbeelden:*

- Zoals hierboven reeds vermeld, kunnen in sommige gevallen genen geïdentificeerd worden die differentieel tot expressie komen in tumoren met en zonder invasief of *metastatisch fenotype*. Dit wil zeggen dat in principe modellen kunnen worden gebouwd die, aan de hand van de primaire tumor, een voorspelling maken van het al dan niet gelokaliseerd zijn van het tumoraal proces (stadiumbepaling). Voor tumorale processen zonder macroscopische aantasting van de lymfeklieren of zichtbare metastasen op afstand, kan het model op een dusdanige manier worden getraind dat het de aanwezigheid van subklinische metastasen tracht te voorspellen, zodanig dat de patiënten die het meest baat hebben bij een aanvullende therapie worden geselecteerd. Omgekeerd kunnen patiënten worden geïdentificeerd voor wie de aanvullende therapie waarschijnlijk onnodige toxiciteit of mutilatie zou betekenen.

- Het *klinisch wisselend verloop* van sommige tumoren (bv. adenocarcinomen van de prostaat) kan zich vertalen op moleculair-biologisch niveau (differentieel tot expressie komende genen). Dit maakt het mogelijk om modellen te maken die voorspellen of de tumor een eerder traag dan wel agressief verloop zal kennen. Deze modellen kunnen dus gebruikt worden om een *schatting van de prognose* te maken. Dit kan vergeleken worden met het bepalen van de histopathologische differentiatiegraad van een tumor, wat op dit moment een van de meest gebruikte methoden is om het gedrag van een tumor te beschrijven en te voorspellen (bijvoorbeeld Gleason-score bij prostaatkanker). Soms kan het echter gebeuren dat tumoren die er klinisch en histopathologisch gelijkaardig uitzien, toch een zeer verschillend verloop kennen. In sommige gevallen (zie ook (4)) zullen modellen gebaseerd op metingen met microroosters uitkomst kunnen bieden om dit verschillend verloop te voorspellen.

- Modellen die tot doel hebben het *antwoord op therapie* te voorspellen, zouden kunnen gebruikt worden om, aan de hand van expressieniveaus in de onbehandelde tumor, een schatting te maken van de kans op succes van een behandelingsschema. Er kan bijvoorbeeld gepoogd worden te bepalen welke patiënten tijdens behandeling met chemotherapie een volledige remissie of, integendeel, progressie van het tumorale proces zullen vertonen. Zo ook kan getracht worden te voorspellen welke patiënten die, na het behalen van een volledige remissie, al dan niet zullen hervallen. De gekozen modellen kunnen dan gebruikt worden om een keuze te maken van de meest geschikte therapie.

Ontdekken van klassen

Door gebruik te maken van het uitgebreid arsenaal aan klinische en morfologische parameters kan men een kwaadaardig proces indelen in verschillende categorieën of klassen. De manier van classificeren zal in de meeste gevallen ook het beleid bepalen. Zoals reeds vermeld kunnen patiënten met een gelijkaardige diagnose en therapie (dus patiënten die volgens de huidige kennis tot dezelfde categorie behoren) een variabel verloop kennen. Door het herschikken of opdelen van diagnostische categorieën kan gepoogd worden deze variabiliteit binnen eenzelfde klasse te verminderen en kan het in sommige gevallen mogelijk zijn om de therapie te verfijnen en het verloop van de ziekte beter te voorspellen. Hier is het dus niet de bedoeling om voorspellingen te gaan maken voor individuele patiënten (zoals in de vorige doelstelling), maar om te bepalen welke de verschillende tumorklassen en hun eigenschappen zijn.

Het is mogelijk om, op basis van metingen met microarrays (dus op basis van het expressieniveau van een paar duizend genen), een groep patiënten in te delen in verschillende klassen of clusters waarvan de eigenschappen niet op voorhand gekend zijn. Een klasse of cluster groepeerd over het algemeen patiënten die een zekere graad van overeenkomst vertonen. De indeling wordt over het algemeen bepaald d.m.v. clusteringstechnieken (43). Enkele voorbeelden van veelgebruikte methoden zijn: hiërarchische clustering (44), “Self Organizing Maps” (9, 45) en “K-means clustering” (46). In veel gevallen zal de indeling, gevonden door deze clusteringstechnieken, grotendeels overeenkomen met de gekende indeling. In sommige gevallen zal er echter een totaal nieuwe indeling gevonden worden die eventueel klinisch relevant kan zijn.

Deze technieken kunnen mooi geïllustreerd worden aan de hand van de verzameling patiënten met acute leukemie. Veronderstel dat het onderscheid tussen ALL en AML nog niet zou gekend zijn (de

subklassen voor acute leukemie zijn nog niet gekend). Veronderstel ook dat er geen onderscheid voorkomt tussen trainingsset of testset. In dit geval is er dus een verzameling van 72 patiënten met acute leukemie, zonder meer, waarvoor de expressieniveaus van een 7000-tal genen zijn bepaald. Na bepaling van de eerste vijf principale componenten van deze verzameling, werd er d.m.v. K-means clustering gezocht naar twee clusters. Het resultaat is te zien in fig. 4A (enkel de eerste twee principale componenten worden hier getoond). Het algoritme is erin geslaagd om twee groepen te identificeren. Wanneer men de patiënten van de eerste groep bekijkt, valt het op dat dit allemaal patiënten zijn die, op één na, ALL hadden, en wanneer men de patiënten van de tweede groep bekijkt, valt het op dat die patiënten allemaal AML hadden. Dit algoritme is er dus als het ware in geslaagd om de concepten ALL en AML te herontdekken. In dit geval brengt het ontdekken van de twee groepen niets nieuws bij, maar het toont de mogelijkheden van deze techniek duidelijk aan.

Ter illustratie worden in fig. 4B de resultaten getoond van hiërarchische clustering op dezelfde verzameling patiënten (hiërarchische clustering is een methode die in de literatuur zeer regelmatig wordt gebruikt) (44). Deze techniek neemt achtereenvolgens de meest gelijkaardige elementen in de data samen, zodat een boomachtige structuur ontstaat. Afhankelijk van het niveau waarop men deze “boom” doorsnijdt, ontstaan dan verschillende clusters.

Alizadeh et al. (4) en Perou et al. (13) hebben door middel van deze technieken zeer belangrijke resultaten bekomen; deze auteurs hebben hiërarchische clustering toegepast op patiënten met een diffuus grootcellig B-cellymfoom (non-Hodgkin-lymfoom) waarvoor het expressieniveau van een aantal genen werd bepaald d.m.v. een cDNA-microrooster. Voor de clustering werden expressieniveaus van 4026 genen gebruikt. Patiënten met een diffuus grootcellig B-cellymfoom vertonen een klinisch heterogeen verloop. Sommigen reageren goed op therapie en kunnen curatief behandeld worden terwijl anderen een slecht verloop kennen. Door clustering kunnen twee groepen geïdentificeerd worden waarin de overleving significant verschillend is. Terwijl in de ene groep 76% van de patiënten na 5 jaar nog in leven was, was er in de andere groep nog slechts 16% in leven. Door deze indeling in clusters te combineren met een prognostische factor (gebaseerd op klinische gegevens), is het mogelijk om een nog meer verfijnde opdeling te maken van de prognostische categorieën. Dit wil dus zeggen dat in dit geval de inlichtingen bekomen uit de meting van expressieniveaus deels complementair zijn aan de klinische gegevens.

Perou et al. (13) bestudeerden borsttumoren, afkomstig van 42 patiënten, met cDNA-microroosters (8102 genen). Hiërarchische clustering scheidde de patiënten in twee grote groepen: oestrogenreceptor-negatieve en oestrogenreceptor-positieve patiënten. De oestrogenreceptor-negatieve groep werd op haar beurt opgesplitst in twee subtypen met een verschillend moleculair biologisch profiel, die misschien beschouwd zouden moeten worden als twee apart te behandelen entiteiten.

Kritische bemerkingen

Zoals hierboven besproken, kunnen gegevens bekomen met microroosters potentieel van grote waarde zijn in de aanpak van kwaadaardige processen. Er zijn echter nog een aantal problemen eigen aan deze techniek waarvan hieronder enkele worden vermeld (47).

Zoals uit de bespreking hierboven valt af te leiden, worden de beste resultaten bekomen voor hematologische aandoeningen. Dit is eraan te danken dat voor deze categorie van aandoeningen een zeer zuivere selectie van kwaadaardige cellen kan bekomen worden. Vaste tumoren kunnen echter veel moeilijker gezuiverd worden, zodat men in vele gevallen een mengsel bekomt van kwaadaardige cellen met andere cellen (zoals cellen afkomstig van normaal weefsel, stromale cellen, lymfocyten, ...). De gemeten expressieniveaus zullen daarom gemiddelden zijn van de expressieniveaus in deze verschillende cellen en als het ware gedeeltelijk een weerspiegeling zijn van de samenstelling van de tumor. Door deze uitmiddeling is het mogelijk dat de meest essentiële informatie wordt gemaskeerd. De samenstelling van de tumor is natuurlijk relevante informatie maar dit is in feite equivalent aan een histopathologische diagnose. Een mogelijke oplossing voor dit probleem is het gebruik van modellen die de gegevens bekomen met microroosters zodanig filteren dat enkel de informatie eigen aan de kwaadaardige component wordt gebruikt (12, 15). Een andere mogelijke oplossing is het gebruik van lasermicrodissectie om een zuivere selectie van kwaadaardige cellen te bekomen (48, 49).

Een probleem dat zich bij microdissectie (en meer algemeen bij het gebruik van te kleine biopten) stelt, is het te klein aantal cellen dat hieruit voortkomt. Omdat voor metingen met microroosters een minimum aan mRNA noodzakelijk is, moeten hier methoden gebruikt worden die op een betrouwbare wijze (bij de meest courant gebruikte methoden om mRNA te amplificeren - zoals PCR - bestaat immers

het gevaar voor verstoring van de expressiepatronen) dit mRNA lineair amplificeren (11). Lineaire amplificatie van RNA vormt echter nog een probleem vermits deze techniek nog niet optimaal is.

Een tweede, gelijkaardig probleem, is dat binnen eenzelfde tumor de kwaadaardige cellen zelf een verschillend gedrag kunnen vertonen. Zo kunnen deze cellen een verschillende differentiatiegraad hebben of een verschillend invasief potentieel bezitten. De expressieniveaus zullen dan opnieuw een gemiddelde zijn van de niveaus in de verschillende cellen terwijl er vooral interesse is voor de meest agressieve component van een tumor.

Een ander probleem wordt veroorzaakt doordat expressieniveaus van genen over het algemeen niet enkel afhankelijk zijn van de eigenschappen van een tumor. Ten eerste zijn er de variaties in expressieniveaus die te wijten zijn aan interindividuele verschillen. Verschillende individuen kunnen onderling soms zeer verschillende expressieniveaus vertonen zonder dat die te wijten zijn aan enig pathologisch verschijnsel. Ten tweede zijn er de variaties in expressieniveaus die te wijten zijn aan experimentele omstandigheden. Transcriptie is immers extreem gevoelig voor externe prikkels (bv. stress). Daarenboven is mRNA ook onderhevig aan degradatie. Daarom worden tumorbiopsieën, die zullen gebruikt worden voor analyse op een microrooster, best zo snel mogelijk na wegname bewaard in vloeibare stikstof. Ook de gebruikte RNA-extractiemethode kan een invloed uitoefenen op de uiteindelijke kwalitatieve en kwantitatieve samenstelling van het totaal mRNA.

Een andere belangrijke opmerking heeft ermee te maken dat de relatie tussen expressieniveau en eiwit zeker niet eenduidig is en dat de huidige kennis in verband hiermee nog zeer beperkt is. Verschillende processen op het niveau van het mRNA (bijvoorbeeld leesraamverschuivingen onder invloed van deletiemutaties, subtiele veranderingen in genexpressie, ...), op het niveau van de translatie en op het niveau van de posttranslationale verwerking van het eiwit kunnen het uiteindelijke proteïne en zijn functioneren in de cel aanzienlijk beïnvloeden. Deze fenomenen worden dikwijls niet waargenomen door middel van microroosters. Expressieniveaus gemeten met microroosters vertellen dus als het ware maar een beperkt deel van het verhaal. Daarom kan het aan te raden zijn om de resultaten, bekomen met microroosters, op het niveau van de eiwitten te toetsen aan bijvoorbeeld weefselmicroroosters (tissue-microarrays (50)).

Wanneer de meer technische problemen in verband met microroosters zullen zijn opgelost (en indien ons geen verrassingen meer te wachten staan), zullen de verwerking en de analyse van data geproduceerd door deze apparatuur uitermate belangrijk zijn. De ontwikkeling, de verfijning en de validatie van de gebruikte methoden zullen dan de grootste uitdaging vormen. De in deze tekst besproken technieken trachten hiertoe een eerste aanzet te vormen.

Besluit

Microroosters of microarrays laten toe in tumoren de expressie van een groot aantal genen parallel te meten. Omdat de ontregelde expressie van genen aan de basis ligt van het gedrag van tumoren, kan de meting hiervan zeer waardevol zijn om het gedrag van kwaadaardige processen te voorspellen of te modelleren. Hierdoor worden de fundamentele processen die aan de basis liggen van de carcinogenese mee betrokken in de klinische besluitvorming.

Door de grote hoeveelheid gegevens die resulteren uit deze metingen, zijn er methoden nodig die een clinicus of onderzoeker helpen om hieruit relevante informatie te halen. Deze technieken kunnen over het algemeen gebruikt worden om voorspellingen te maken, om diagnostische categorieën te herschikken of om kenmerken te selecteren die belangrijk zijn om het onderscheid te maken tussen verschillende groepen tumoren.

Wanneer in de toekomst de kostprijs van deze technologie zal dalen, behoort het tot de mogelijkheden dat microroosters deel zullen uitmaken van de dagelijkse klinische praktijk.

Dankbetuiging

Frank De Smet is wetenschappelijk medewerker aan de K.U.Leuven. Bart De Moor is gewoon hoogleraar aan de K.U.Leuven. Yves Moreau is een postdoctoraal onderzoeker bij het F.W.O. Dit werk wordt gedeeltelijk ondersteund door verschillende instituten: 1. De Vlaamse regering: a) Onderzoeksraad K.U.Leuven: G.O.A. Mefisto-666. b) De F.W.O. -projecten G.0240.99, G.0256.97, G.0115.01, en onderzoeksgemeenschappen: ICCoS en ANMMM. c) IWT-project: STWW. 2. De Belgische staat, bureau van de Eerste Minister – OSTC: IUAP P4-02 (1997-2001) en IUAP P4-24 (1997-2001).

De wetenschappelijke verantwoordelijkheid berust bij de auteurs.

Fig. 1

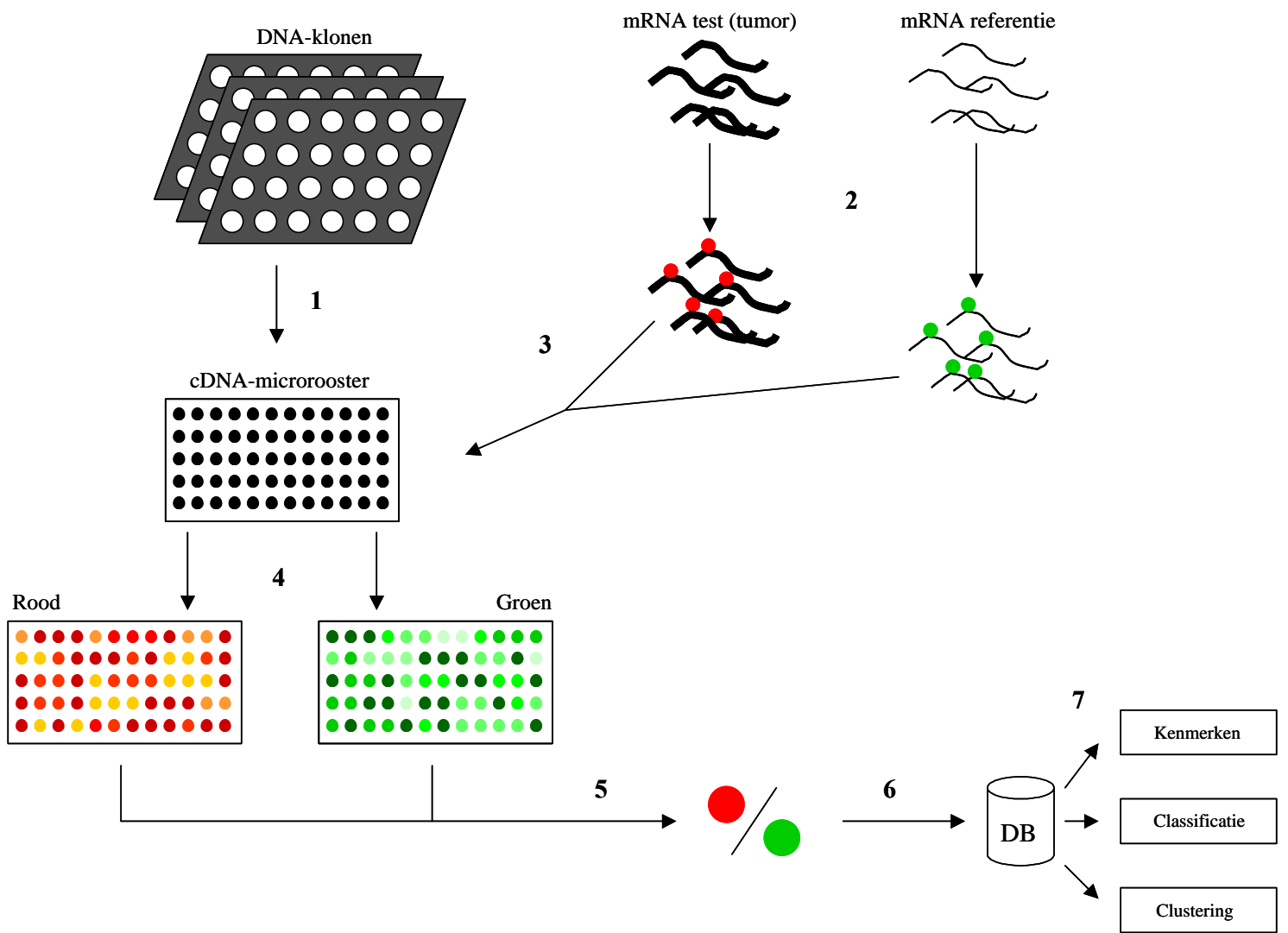


Fig. 2

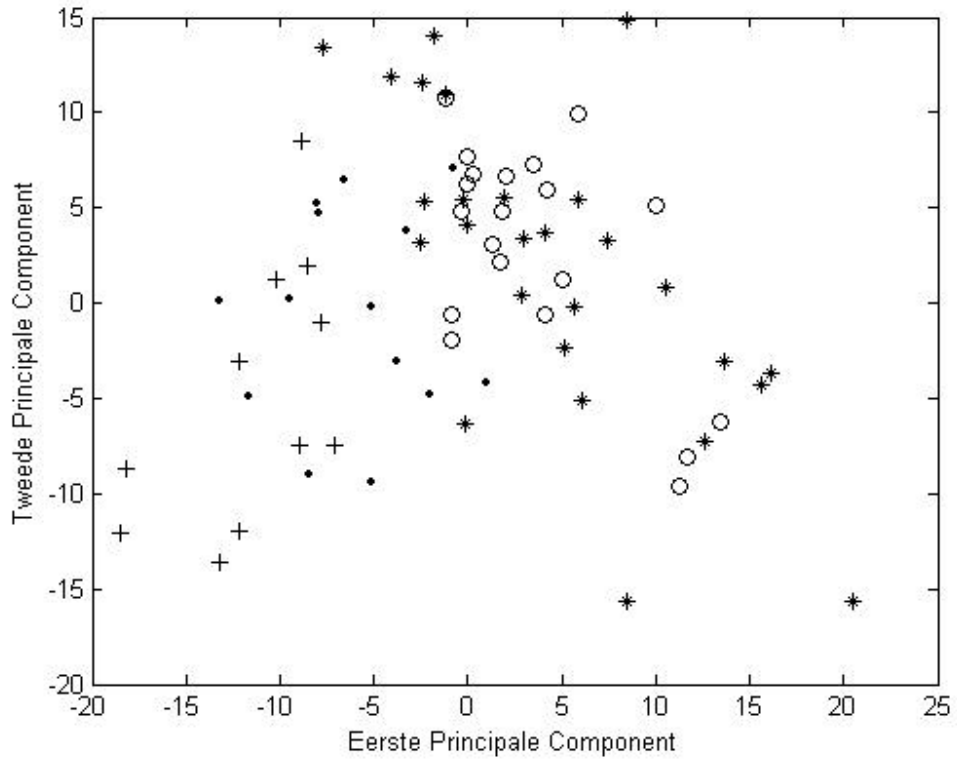


Fig. 3

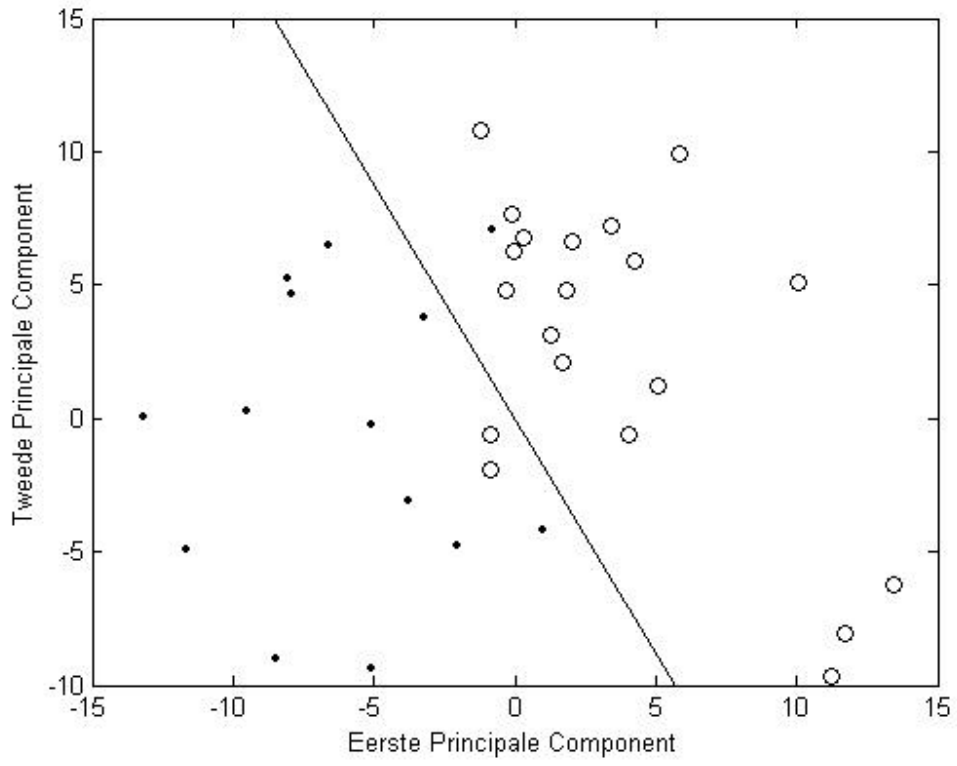
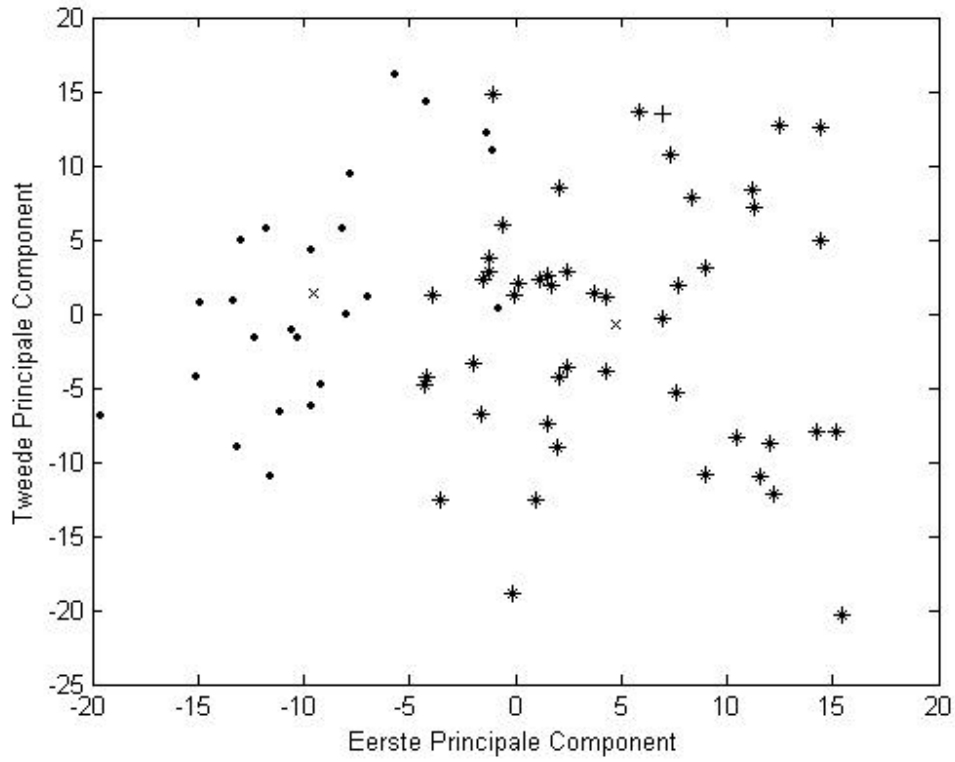
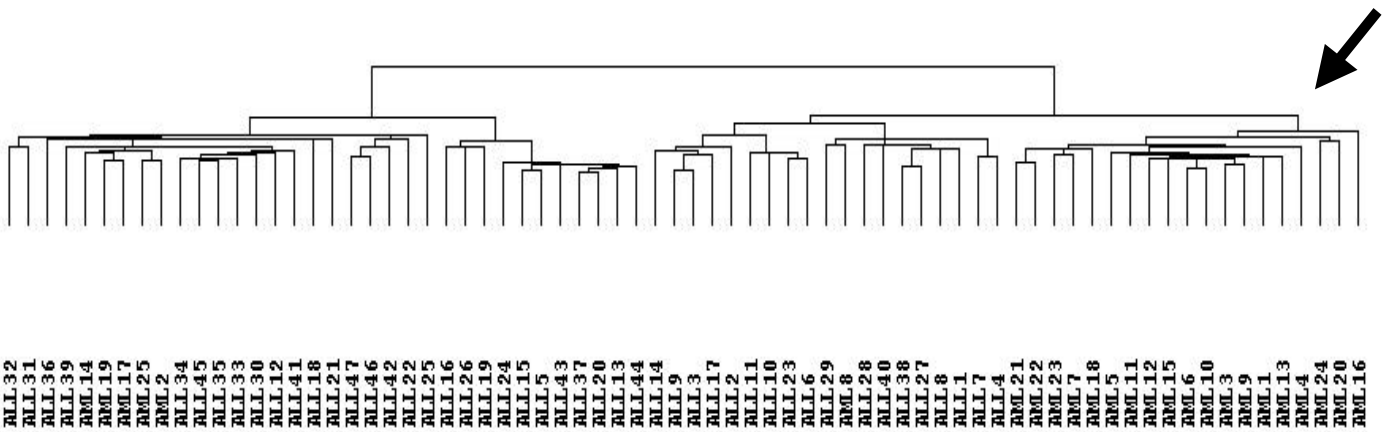


Fig. 4

A.



B.



Figuuronderschriften

Fig. 1: Schematisch overzicht van een experiment met een cDNA-microrooster: (1) Drukken of “stippelen” van de DNA-sondes (afgeleid van de te bestuderen genen) op het microrooster (glazen plaatje). Deze sondes zijn de gezuiverde producten bekomen na PCR-amplificatie van de overeenkomstige DNA-klonen. (2) Merken (eventueel indirect, na “reverse” transcriptie) van totaal mRNA uit het teststaal (tumor) en uit het referentiestaal (rode merker voor mRNA-test, groene merker voor mRNA-referentie). (3) Mengen en in contact brengen van beide stalen met de sondes op het microrooster (hybridisatie). (4) Afzonderlijk aflezen van rode en groene intensiteit (dit is een maat voor de hybridisatie – hier weergegeven door verschillende tinten van rood en groen) ter hoogte van iedere sonde. (5) Berekening van de relatieve expressieniveaus (intensiteit rood / intensiteit groen = hybridisatietest / hybridisatiereferentie) ter hoogte van iedere sonde. Indien dit relatieve expressieniveau ter hoogte van een bepaalde sonde groter is dan 1, is de expressie van het overeenkomstige gen in het teststaal hoger dan in het referentiestaal. Wanneer dit relatieve expressieniveau kleiner is dan 1, is het omgekeerde waar. (6) Opslaan van resultaten in een databank. (7) Verwerking en analyse van gegevens (“data mining”).

Fig. 2: Principale Component Analyse voor de verzameling patiënten met acute leukemie volgens de eerste twee principale componenten van de trainingsset. De waarden op de X- en de Y-as komen overeen met de eerste twee combinaties van genexpressies (bepaald door de eerste twee principale componenten van de trainingsset) uitgerekend voor iedere patiënt (zie ook tekst). Trainingsset: * = ALL, + = AML; Testset: O = ALL, • = AML.

Fig. 3: Model voor de classificatie van patiënten met acute leukemie na principalecomponentanalyse volgens de eerste twee principale componenten van de trainingsset. De parameters van de rechte werden bepaald aan de hand van de patiënten uit de trainingsset (zie patiënten aangeduid met * of + in fig. 2). Enkel patiënten uit de testset worden in deze figuur getoond. De punten boven de rechte worden geclassificeerd als ALL en de punten eronder als AML. Merk op dat deze classificatie voor 3 patiënten verkeerd is. Testset: O = ALL, • = AML.

Fig. 4: Clusteranalyse voor de verzameling patiënten met acute leukemie. (A) “K-means cluster”-analyse na principalecomponentanalyse volgens de eerste vijf principale componenten van de volledige dataset. Enkel de eerste twee principale componenten worden getoond. Er werden 2 groepen of clusters geïdentificeerd. Merk op dat, op één na, alle patiënten in de eerste cluster ALL hadden (er is maar één patiënt die wordt aangeduid met +) en dat alle patiënten in de tweede cluster AML hadden (er is geen patiënt die wordt aangeduid met O). Cluster 1: * = ALL, + = AML; Cluster 2: O = ALL, • = AML; x = clustergemiddelden. (B) Hiërarchische clustering van de volledige dataset. De terminale “takken” van de boomstructuur stellen de individuele patiënten voor (weergegeven door ALL/AML + volgnummer). Merk op dat de meeste patiënten met AML (19 op 25) gegroepeerd zijn in het meest rechtse deel van deze boomstructuur (de “tak” van de “boom” waarnaar de pijl wijst, bevat alleen patiënten met AML).

LITERATUUR

1. DERISI JL, IYER VR, BROWN PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 1997; *278*: 680-686.
2. LANDER ES. Array of hope. *Nat Genet* 1999; *21*: 3-4.
3. SCHENA M, SHALON D, DAVIS RW, BROWN PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995; *270*: 467-470.
4. ALIZADEH AA, EISEN MB, DAVIS RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000; *403*: 503-511.
5. ALON U, BARKAI N, NOTTERMAN DA, et al. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci USA* 1999; *96*: 6745-6750.
6. BERNS A. Gene expression in diagnosis. *Nature* 2000; *403*: 491-492.
7. COLE KA, KRIZMAN DB, EMMERT-BUCK MR. The genetics of cancer--a 3D model. *Nat Genet* 1999; *21*: 38-41.
8. FRIEND SH. How DNA microarrays and expression profiling will affect clinical practice. *BMJ* 1999; *319*: 1306-1307.
9. GOLUB TR, SLONIM DK, TAMAYO P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999; *286*: 531-537.
10. GRAY JW, COLLINS C. Genome changes and gene expression in human solid tumors. *Carcinogenesis* 2000; *21*: 443-452.
11. LOCKHART DJ, WINZELER EA. Genomics, gene expression and DNA arrays. *Nature* 2000; *405*: 827-836.

12. PEROU CM, JEFFREY SS, VAN DE RIJN M, et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc Natl Acad Sci USA* 1999; *96*: 9212-9217.
13. PEROU CM, SØRLIE T, EISEN MB, et al. Molecular portraits of human breast tumors. *Nature* 2000; *406*: 747-752.
14. RIMM DL. Molecular biology in cytopathology: current applications and future directions. *Cancer* 2000; *90*: 1-9.
15. ROSS DT, SCHERF U, EISEN MB, et al. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 2000; *24*: 227-235.
16. MOCH H, SCHRAML P, BUBENDORF L, et al. High-throughput tissue microarray analysis to evaluate genes uncovered by cDNA microarray screening in renal cell carcinoma. *Am J Pathol* 1999; *154*: 981-986.
17. NACHT M, FERGUSON AT, ZHANG W, et al. Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer. *Cancer Res* 1999; *59*: 5464-5470.
18. SCHUMMER M, NG WV, BUMGARNER RE, et al. Comparative hybridization of an array of 21,500 ovarian cDNAs for the discovery of genes overexpressed in ovarian carcinomas. *Gene* 1999; *238*: 375-385.
19. SGROI DC, TENG S, ROBINSON G, LEVANGIE R, HUDSON JRJ, ELKAHLOUN AG. In vivo gene expression profile analysis of human breast cancer progression. *Cancer Res* 1999; *59*: 5656-5661.
20. VILLARET DB, WANG T, DILLON D, et al. Identification of genes overexpressed in head and neck squamous cell carcinoma using a combination of complementary DNA subtraction and microarray analysis. *Laryngoscope* 2000; *110*: 374-381.

21. WANG K, GAN L, JEFFERY E, et al. Monitoring gene expression profile changes in ovarian carcinomas using cDNA microarray. *Gene* 1999; 229: 101-108.
22. WANG T, HOPKINS D, SCHMIDT C, et al. Identification of genes differentially over-expressed in lung squamous cell carcinoma using combination of cDNA subtraction and microarray analysis. *Oncogene* 2000; 19: 1519-1528.
23. XU J, STOLK JA, ZHANG X, et al. Identification of differentially expressed genes in human prostate cancer using subtraction and microarray. *Cancer Res* 2000; 60: 1677-1682.
24. SCHOLL FA, BETTS DR, NIGGLI FK, SCHAFFER BW. Molecular features of a human rhabdomyosarcoma cell line with spontaneous metastatic progression. *Br J Cancer* 2000; 82: 1239-1245.
25. WELFORD SM, GREGG J, CHEN E, et al. Detection of differentially expressed genes in primary tumor tissues using representational differences analysis coupled to microarray hybridization. *Nucleic Acids Res* 1998; 26: 3059-3065.
26. DEBOUCK C, GOODFELLOW PN. DNA microarrays in drug discovery and development. *Nat Genet* 1999; 21: 48-50.
27. DE BAERE E, MESSIAEN L. DNA-onderzoek in het begin van de 21ste eeuw: dromen en werkelijkheid. *Tijdschr Geneesk* 2000; 56: 1696-1703.
28. D'HAESELEER P, LIANG S, SOMOGYI R. Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics* 2000; 16: 707-726.
29. BASSETT DE Jr, EISEN MB, BOGUSKI MS. Gene expression informatics--it's all in your mine. *Nat Genet* 1999; 21: 51-55.
30. SAGER R. Expression genetics in cancer: shifting the focus from DNA to RNA. *Proc Natl Acad Sci USA* 1997; 94: 952-955.

31. DUGGAN DJ, BITTNER M, CHEN Y, MELTZER P, TRENT JM. Expression profiling using cDNA microarrays. *Nat Genet* 1999; *21*: 10-14.
32. LIPSHUTZ RJ, FODOR SP, GINGERAS TR, LOCKHART DJ. High density synthetic oligonucleotide arrays. *Nat Genet* 1999; *21*: 20-24.
33. http://waldo.wi.mit.edu/MPR/data_set_ALL_AML.html
34. HILSENBECK SG, FRIEDRICHS WE, SCHIFF R, et al. Statistical analysis of array expression data as applied to the problem of tamoxifen resistance. *J Natl Cancer Inst* 1999; *91*: 453-459.
35. BUBENDORF L, KOLMER M, KONONEN J, et al. Hormone therapy failure in human prostate cancer: analysis by complementary DNA and tissue microarrays. *J Natl Cancer Inst* 1999; *91*: 1758-1764.
36. KHAN J, BITTNER ML, SAAL LH, et al. cDNA microarrays detect activation of a myogenic transcription program by the PAX3-FKHR fusion oncogene. *Proc Natl Acad Sci USA* 1999; *96*: 13264-13269.
37. BISHOP CM. Pre-processing and feature extraction. In: Bishop CM, ed. *Neural networks for pattern recognition*. New York: Oxford University Press Inc., 1995: 293-331.
38. BISHOP CM. Single-layer networks. In: Bishop CM, ed. *Neural networks for pattern recognition*. New York: Oxford University Press Inc., 1995: 77-115.
39. BISHOP CM. The multi-layer perceptron. In: Bishop CM, ed. *Neural networks for pattern recognition*. New York: Oxford University Press Inc., 1995: 116-63.
40. TIMMERMAN D, VERRELST H, BOURNE TH, et al. Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses. *Ultrasound Obstet Gynecol* 1999; *13*: 17-25.
41. JENSEN FV. *An introduction to Bayesian networks*. New York: Springer-Verlag, 1996.

42. MUKHERJEE S, TAMAYO P, SLONIM D, et al. Support vector machine classification of microarray data. Technical report CBCL182, MIT, 1998:
<http://www.ai.mit.edu/people/sayan/pub.html>
43. SHERLOCK G. Analysis of large-scale gene expression data. *Curr Opin Immunol* 2000; *12*: 201-205.
44. EISEN MB, SPELLMAN PT, BROWN PO, BOTSTEIN D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998; *95*: 14863-14868.
45. KOHONEN T. Self-organizing maps. Berlin: Springer-Verlag, 1997.
46. TOU JT, GONZALEZ RC. Pattern classification by distance functions. In: Tou JT, Gonzalez RC, eds. *Pattern recognition principles*. Reading, MA: Addison-Wesley, 1979: 75-109.
47. MASTERS JR, LAKHANI SR. How diagnosis with microarrays can help cancer patients (letter). *Nature* 2000; *404*: 921.
48. SIMONE NL, BONNER RF, GILLESPIE JW, EMMERT-BUCK MR, LIOTTA LA. Laser-capture microdissection: opening the microscopic frontier to molecular analysis. *Trends Genet* 1998; *14*: 272-276.
49. EMMERT-BUCK MR, BONNER RF, SMITH PD, et al. Laser capture microdissection. *Science* 1996; *274*: 998-1001.
50. MOCH H, KONONEN T, KALLIONIEMI OP, SAUTER G. Tissue microarrays: what will they bring to molecular and anatomic pathology ? *Adv Anat Pathol* 2001; *8*: 14-20.