Microarray-based expression signatures: potential application for individualized cancer treatment

Chapter 9

Summary

Summary

One of the major focuses in cancer research is the identification of novel biomarkers and therapeutic targets to improve treatment outcomes. There is a significant need to enhance both the accuracy and precision of risk prediction and therapy selection. Advances in both molecular and clinical cancer research have improved prognosis and reduced side-effects. Nonetheless, available therapy remains ineffective in a large fraction of patients. New biomarkers could help predict which patients would benefit from a treatment.

With the advent of high-throughput technologies it became possible to study enormous numbers of parameters simultaneously, opening up a new era of possibilities to identify novel markers much faster than previously possible.

One of these novel technologies, gene-expression microarrays have been widely used in cancer research for the last decade. Measuring the expression of thousands of genes in parallel has the potential to answer many questions that were previously impossible to resolve. This type of transcriptomic profiling has, amongst others, led to the identification of dysregulated pathways in cancer. Furthermore, multi-gene markers (so-called gene signatures) have been created that show the promise to improve risk prediction and therapy selection. Despite progress in this area, introduction into clinical practice has been slow. Only two breast cancer gene signatures are currently tested in clinical trials. Reports on minimal gene-wise overlap, disappointing validation in external data and lack of consistency have negatively impacted the field.

The first part of this thesis focuses on the creation of a robust, biology-bases gene signature for multiple cancer types. In the second part of this thesis we elaborate on the data analysis of microarray experiments.

Proliferation signature

Recently it has been reported that, although gene-wise overlap is poor, classification results from different signatures show high agreement. It is thought that this phenomenon reflects common underlying biology represented by each classifier in a different way. In breast cancer, more and more data suggests that cellular proliferation plays a major role in the prognosticity of gene signatures.

To clarify the significance of proliferation in prognostic breast cancer signatures we created a gene expression-based marker reflecting cellular proliferation (**chapter 2**).

Two *in vitro* gene expression datasets were combined to extract this signature: genes were selected that showed a cycling expression pattern after synchronization in one dataset and responded to serum stimulation in the other. This 104-gene proliferation signature had high prognostic power in different breast cancer datasets. We further hypothesized that this marker could also be used for prognostic purposes in other cancer sites and validated it across multiple cancer types. A further independent validation of our proliferation signature was performed by relating gene expression to other proliferation measures in both patient and xenograft material (**chapter 3**).

Since the clinical utility of microarrays can be controversial, we sought to evaluate our classifier using an independent technique. In **chapter 4** the number of genes in the signature was computationally-reduced in an information-preserving manner, followed by experimental evaluation using RT-PCR. This reduced marker was successfully tested in two large patient microarray meta-datasets. Subsequently, this refined marker was validated with PCR *in vitro*, *in vivo* and in an independent primary breast cancer cohort.

Challenges in microarray data analysis

Several aspects of high-throughput gene expression profiling involve major technical challenges. The first lies in the dimensionality of a microarray experiment, where thousands of genes are simultaneously measured in a relatively small number of samples. Without the use of proper statistical and computational methods to account for this “curse of dimensionality” there is a major risk of over-fitting, which can lead to the reporting of over-optimistic and non-generalizable results. Increasingly sophisticated methods are being developed to handle this problem when identifying interesting genes and signatures from microarray data. This has clearly led to an improvement in quality of the reported classifiers. Less notice of the multiple testing issue is taken once a signature is created. Nevertheless, despite all efforts to account for multiple testing and over-fitting, it is crucial to independently validate prognostic gene expression-based. At this stage standard survival statistics are generally used to address prognosticity.

In **chapter 5** we clearly shown that also in this phase great care must be taken. By testing sets of randomly-generated signatures in multiple independent datasets we demonstrate that many gene sets reach statistical significance. In some datasets the proportion of significant gene-sets reached dramatic levels, indicating that a single,

dataset-invariant threshold for significance is inappropriate. Based on this permutation study, we suggest a method to test signature performance to control for spurious random findings. In addition, our data demonstrates that evaluating a signature in multiple independent datasets attenuates the multiple testing issue.

Another reason for discrepancies and inconsistencies in reported studies is the enormous diversity of available microarray platforms and pre-processing techniques. Variations in either of these parameters can result in identification of different genes and signatures. Many previously reported failed data replications can be attributed to unavailability or incompleteness of data or analysis details.

We therefore believe it is essential to follow the original data handling pipeline as precisely as possible (**chapter 6**), as failure to do so may in part explain why independent validation of signatures by other research groups often fails. Although in most of the cases the statistical methods are properly followed, far less attention is usually paid to data pre-processing. We sought to clarify the importance of standardizing the pre-processing schedule for gene expression-based classifiers. Two existing non-small-cell lung cancer signatures were tested in an independent dataset pre-processed with 24 different methods. The data in **chapter 7** clearly demonstrates that even small changes in pre-processing could change a successful marker into one indistinguishable from chance.

Our results do not only reveal the sensitivity of marker performance to the applied pre-processing, but also lead to the report of a new validation of two non-small-cell lung cancer markers in a new, large patient cohort. Interestingly, this study revealed that when focusing on those patients for whom different pre-processing schemes agree prognostic capability increased. This feature might be exploited to improve classifier robustness.

In conclusion, we created a signature for cellular proliferation, which demonstrated high prognostic power in multiple cancer types. We successfully translated this marker to a PCR-based classifier, which will increase clinical utility. In the second part of the thesis we identify several key issues that when kept in mind will improve reported classifier validity. Two novel methods for assessing signature robustness were developed. Taken together, these results make concrete and practical steps towards speeding the integration of molecular diagnostics with clinical practice.

Nederlandse samenvatting

Samenvatting

Het identificeren van nieuwe biomarkers en het stellen van therapeutische doelen ter verbetering van de behandeling is een van de speerpunten van kankeronderzoek. Het is duidelijk dat er nauwkeurigere methoden voor het voorspellen van het behandelingsresultaat en het bepalen van de optimale behandeling per individu nodig zijn. Vooruitgang op het gebied van onderzoek en verbetering van de bestaande behandelmethodes hebben zeer zeker een positief effect gehad op de prognose van patiënten. Echter, voor een groot deel van de patiënten blijven beschikbare therapieën inefficiënt of zelfs onbeduidend.

Met de opkomst van 'high-throughput' technieken is het mogelijk geworden om een zeer groot aantal parameters tegelijkertijd te bestuderen. Door deze ontwikkeling kunnen nieuwe markers vele malen sneller worden geïdentificeerd dan voorheen mogelijk was.

Een van deze nieuwe methoden, genaamd gen-expressie microarray technologie, wordt reeds veelvuldig gebruikt binnen het kankeronderzoek. Met deze techniek kan de expressie van duizenden genen tegelijkertijd worden gemeten, waardoor het mogelijk is om vele vraagstukken te ontrafelen die voorheen onmogelijk te beantwoorden waren. Er zijn sinds de introductie van deze methode verscheidene biologische processen gevonden die ontregeld zijn binnen kankercellen. Daarnaast worden gen-expressie microarrays veelvuldig gebruikt om multi-gen markers te genereren. Een zodanig samengestelde set genen heeft de potentie om te voorspellen welke patiënten goed op een behandeling zullen reageren en welke een slechtere prognose hebben. Het idee is dat deze markers uiteindelijk zullen bijdragen aan het bepalen van de meest optimale behandeling voor een patiënt.

Ondanks de voortgang in onderzoek op dit gebied verloopt de introductie van gen-expressie microarrays en geïdentificeerde multi-gen markers in een klinische setting traag. Momenteel zijn er slechts twee grote klinische studies gestart waarin multi-gen markers voor borstkanker worden getest. Rapportage van minimale gen overlap tussen markers, teleurstellende validatie resultaten in externe data en een gebrek aan consistentie dragen hiertoe bij.

Het eerste gedeelte van deze thesis behandelt de ontwikkeling van een robuuste multi-gen marker die mogelijk voor meerdere kankersoorten kan worden toegepast.

In het tweede deel wordt dieper ingegaan op de analyse van gen-expressie microarray data en wat de invloed hiervan is op marker resultaten.

Multi-gen marker voor proliferatie

Uit recent onderzoek blijkt ondanks dat er een gebrek aan overlap tussen multi-gen markers is, dat de patiënt classificaties met deze markers zeer goed overeenstemmen. Er wordt gedacht dat ongeacht of deze multi-gen markers qua inhoud verschillen of niet, ze gelijkwaardige biologie representeren. Verder wijst steeds meer onderzoek naar borstkanker erop dat proliferatie een belangrijke rol speelt ten aanzien van de prognostische waarde van multi-gen markers.

Om deze redenen hebben we in **hoofdstuk 2** een multi-gen marker ontwikkeld voor proliferatie. Onze verwachting is dat deze marker niet alleen van prognostische waarde zal zijn voor borstkanker patiënten, maar ook toepasbaar zal zijn voor andere kankersoorten. Om deze set van genen te creëren is informatie uit twee *in vitro* microarray datasets gecombineerd. In totaal werden 104 genen geselecteerd die een cyclisch expressie patroon volgden na synchronisatie in de ene dataset en een reactie vertoonden op stimulatie met serum in een andere dataset. Deze multi-gen proliferatie marker heeft een hoge prognostische waarde in verschillende patiënten microarray datasets. Een tweede validatie werd uitgevoerd door expressie van de proliferatie-geassocieerde genen te correleren aan andere metingen van proliferatie materiaal van patiënten en xenografts (**hoofdstuk 3**).

Aangezien de introductie van microarrays in een klinische setting nog moeilijk verloopt, hebben we gezocht naar een alternatieve methode om de proliferatie marker te bekijken. In **hoofdstuk 4** hebben we het aantal genen in de multi-gen marker verkleind, zodat het mogelijk werd om deze te evalueren met een onafhankelijke techniek: PCR. De prognostische waarde van deze nieuwe marker werd succesvol getest in twee grote microarray meta-datasets van patiënten. Daaropvolgend werd de aangepaste marker *in vitro*, *in vivo* en in een onafhankelijke patiënten dataset gevalideerd met PCR.

Microarray data analyse

Verschillende aspecten van 'high-throughput' methodes gaan gepaard met aanzienlijke technische vereisten. De dimensionaliteit van microarray data, waarbij expressie van duizenden genen tegelijkertijd wordt gemeten in een relatief klein aantal monsters, brengt een aantal uitdagingen met zich mee wat betreft de data-

analyse. Zonder het gebruik van geschikte statistische methoden en computer algoritmes bestaat er een groot risico van over-fitting, dat kan leiden tot rapportage van over-optimistische resultaten. Er worden steeds meer geavanceerde procedures ontwikkeld om hiervoor te corrigeren tijdens het identificeren van interessante genen en het afleiden van multi-gen markers. Dit heeft zeker effect gehad op de kwaliteit van gegenereerde markers. Er wordt minder aandacht besteed aan dit probleem nadat een nieuwe marker is gecreëerd, het is echter essentieel om prognostische gen-expressie gebaseerde markers te valideren op onafhankelijke data. Om de prognostische waarde van deze markers te bepalen worden meestal standaard methodes voor overlevings statistiek toegepast.
In **hoofdstuk 5** tonen we aan dat ook in dit stadium rekening gehouden moet worden met de dimensionaliteit van de data. Door het genereren en testen van random multi-gen markers, werd duidelijk dat een groot aantal van deze willekeurige markers statistisch significante resultaten opleverden. Afhankelijk van de dataset bereikte dit aantal een dramatisch hoog niveau, waardoor het gebruik van een enkele drempel voor significantie niet geschikt is voor dit soort analyses. Om te controleren voor random bevindingen stellen we een methode voor om multi-gen markers te evalueren gebaseerd op deze permutatie studie. Daarnaast wijst de data in deze studie erop dat bij het testen van een marker in meerdere onafhankelijke datasets het van minder groot belang is om hier rekening mee te houden.

Een andere reden voor gerapporteerde tegenstrijdigheden en instabiliteit van multi-gen markers is het grote scala aan beschikbare microarray types, data voorbewerkingstechnieken en analyse methoden. Veranderingen in een van deze parameters resulteert in de identificatie van andere genen en dus ook in uiteenlopende multi-gen markers. Eerdere beschrijvingen over mislukte data replicatie konden meestal worden toegeschreven aan onvolledige of niet beschikbare ruwe data of analyse details.
Voor externe validatie denken we dat het essentieel is om de data bewerking zo precies mogelijk te volgen als omschreven in de oorspronkelijke studie (**hoofdstuk 6**). Onafhankelijke validatie van multi-gen markers geïdentificeerd door andere onderzoeksgroepen leveren doorgaans negatieve resultaten op. Hoewel in de meeste gevallen de analyse methoden tot in detail worden gevolgd, wordt er veel minder aandacht besteedt aan de data voorbewerking. Om het belang van

standaardisering van data voorbewerking te verduidelijken, werden twee bestaande multi-gen markers voor niet-kleincellig long carcinoma getest in een onafhankelijke dataset die met 24 verschillende schema's werd voorbewerkt. Uit de data in **hoofdstuk 7** blijkt dat zelfs kleine aanpassingen in een voorbewerkings schema een succesvolle marker kunnen veranderen in een marker die niet te onderscheiden is van toeval.
In dit hoofdstuk laten we zien dat de prognostische waarde van een marker zeer gevoelig is voor de gebruikte voorbewerkings techniek. Ook valideren we twee multi-gen markers voor niet-kleincellig long carcinoma in een nieuwe, grote dataset. Een andere interessante bevinding van deze studie is dat wanneer alleen patiënten in acht worden genomen waarvoor de verscheidene voorbewerkingen een gelijke classificatie geven, de prognostische waarde sterk toeneemt. Deze eigenschap zou in de toekomst gebruikt kunnen worden om marker robuustheid te verbeteren.

Concluderend, omschrijven we in deze thesis de ontwikkeling van een muti-gen marker voor proliferatie, die een hoge prognostische waarde heeft in meerdere kankertypes. Deze marker werd succesvol omgezet in een PCR-gebaseerde marker, waardoor de toepasbaarheid in een klinische setting vergemakkelijkt zal worden. In het tweede deel van deze thesis identificeren we een aantal belangrijke punten waar rekening mee gehouden dient te worden bij de analyse van microarray data. Twee nieuwe methoden werden ontwikkeld om de robuustheid van een marker te bepalen. Dit zal bijdragen aan het verbeteren van de kwaliteit van multi-gen markers en leiden tot een snellere vertaling naar de kliniek.