



Improving Design, Execution & Analysis of Transcriptomics
O. Bruning

Samenvatting

Het Verbeteren van het Ontwerp, de Uitvoering en de Analyse van Transcriptoomexperimenten

De afgelopen 20 jaar is het life sciences onderzoek ingrijpend veranderd door de introductie van innovatieve genoombrede technologieën, waarmee veel monsters met hoge snelheid verwerkt kunnen worden en grote hoeveelheden data kunnen worden gegenereerd. Deze zogenaamde “omics”-technologieën kunnen in potentie onze kennis van de biologie enorm uitbreiden, doordat ze het in principe mogelijk maken om voor het gehele genoom alle moleculaire veranderingen simultaan te kwantificeren. Deze technieken zijn dan ook zeer snel als standaard omarmd door vele onderzoekers binnen de life-sciences, zoals blijkt uit het grote aantal succesvolle studies naar de identificatie van biomarkers en de zogenaamde genome-wide associatie studies (GWAS). Het nettoresultaat is echter dat deze technieken nog niet echt hebben geleid tot een uitgebreid begrip van onderliggende moleculaire mechanismen, maar vooral tot de ontdekking van een veelheid aan nieuwe regulatieniveaus en onderdelen in levende cellen. Ook al kunnen we nu dus vrijwel alle cellulaire onderdelen met hoge efficiëntie meten, het begrijpen van de interacties tussen deze vele onderdelen vereist dat we hun basale eigenschappen, functie(s) en werkingslocatie kennen. Het lijkt er dus op dat de complexiteit van levende organismen ons voorstellingsvermogen vooralsnog te boven gaat. Hierdoor voelt het alsof de belofte van “omics” nog niet is waargemaakt, maar in werkelijkheid hebben we geleerd dat we nog slecht weinig weten van de intrigerende biologische systemen die we bestuderen.

Een gevolg van deze ontdekking van vergrote complexiteit van biologische systemen is dat we onze experimentele strategieën zullen moeten aanpassen. Het lijkt er vaak op dat wetenschappers starten met het gebruik van een “omics”-technologie, zonder dat ze daarbij de wijze waarop ze hun experimenten doen, aanpassen. “Omics”-experimenten verschillen echter zodanig van traditionele moleculair-biologische experimenten dat ze een duidelijk andere aanpak vereisen. Het is mogelijk om een experiment in drie belangrijke fasen op te delen: de ontwerpfasen, de uitvoeringsfase en de data-analysefase, en iedere fase moet in methodologische zin geoptimaliseerd uitgevoerd worden. Tijdens het ontwerpen van een experiment is er gevaar voor onvolkomenheden als: onvoldoende biologische replica's, gebrek aan heldere biologische vraagstellingen, een onbekende ontwerprijmte, etc. In de uitvoeringsfase, kunnen er problemen ontstaan door de steeds kleiner wordende hoeveelheid materiaal van monsters en onvoldoende aandacht voor standaardisatie van protocollen. Als laatste is het belangrijk dat er in de data-analysefase gepaste statistiek en deskundige kennis voor het verwerken van enorme datavolumes beschikbaar is en gebruikt wordt.

In dit proefschrift zijn verscheidene kwesties, die van belang zijn gedurende het proces van het ontwerpen, uitvoeren en analyseren van transcriptoomexperimenten onderzocht aan de hand van voorbeeldstudies. Het doel hierbij was om te identificeren welke onderdelen van het proces van experimenteren, verbeterd zouden kunnen worden om de betrouwbaarheid van experimenten voor transcriptoomonderzoek te verhogen.

Om onderzoek naar de uitvoering van transcriptoomexperimenten mogelijk te maken, hebben we gebruik gemaakt van een interessante biologische test-case: de goed onderzochte rol van het p53-gen in de ontwikkeling van tumoren. Kort samengevat: de transcriptiefactor p53, welke een DNA-schadedetector is, voorkomt de opeenhoping van genetische afwijkingen, onder andere veroorzaakt

door UV-bestraling, in cellen en de daaruit volgende vorming van tumoren. De consensus is dat als gevolg van blootstelling aan UV, het p53-eiwit celdeling blokkeert door in te grijpen in de cellcyclus. Hierdoor is het voor cellen mogelijk om hun door UV-geïnduceerde DNA-schade te herstellen via cellulaire DNA-reparatiemechanismen. Als een bepaalde cel echter te veel, ogenschijnlijk niet meer te repareren DNA-schade heeft, initieert p53 geprogrammeerde celdood, ook wel apoptose genoemd, om zo te voorkomen dat de cel ongecontroleerd gaat delen. In het omgekeerde geval, wanneer deze beschermende cellulaire reacties van p53 verstoord of afwezig zijn, kan de opeenhoping van DNA-mutaties leiden tot genomische instabiliteit en uiteindelijk kankerachtige laesies. Aangezien p53 een transcriptiefactor is, werkt het voornamelijk via transcriptionele activatie van zijn doelgenen.

Vandaar dat we bij dit promotieonderzoek gestart zijn met een transcriptoomanalyse van een grootschalig *in-vitro* onderzoek naar de rol van p53 na DNA-schade (zie Hoofdstuk 2). Hierbij hebben we de genexpressie van de reactie op UV-straling in embryonale fibroblasten van muizen (MEFs) met p53-gemuteerde muismodellen geanalyseerd. Aangezien de fosforylering van het p53-eiwit belangrijk is bij p53-gestuurde DNA-schadereacties en deze fosforylering specifiek plaatsvindt op het muizenresidue Ser389 van het p53-eiwit, hebben we in een tijdsreeks van UV-bestraalde wild-type, p53.S389A mutant en p53^{-/-} knock-out MEFs de onderliggende cellulaire processen onderzocht. De eerste observatie toonde aan dat er zelfs zonder UV-bestraling al duizenden genen een andere expressie lieten zien bij de gemuteerde MEFs dan in de wild-type MEFs, wat elke interpretatie van de UV-blootstelling van deze mutanten bemoeilijkt. Verder bleken er in de wild-type MEFs meer dan 6.000 genen te zijn veranderd in reactie op de UV-straling. Deze verandering vond plaats als een strikte twee-fasen-reactie over de tijd en hoewel deze reactie alleen subtiel veranderd was in de p53.S389A-gemuteerde MEFs, waren er toch vele cellulaire processen aangedaan. Daarom lijkt fosforylering van p53.S389 essentieel voor vele p53-doelgenen en p53-afhankelijke processen, ook al heeft het grote aantal veranderde genen gezorgd voor een grotendeels beschrijvende uitkomst van deze studie.

Hoewel we verscheidene nieuwe en belangrijke inzichten met betrekking tot de rol van p53 in reactie op UV hadden gevonden, waren we in het algemeen niet bijzonder tevreden met de algehele interpretatie van de transcriptoomdata. Daarom hebben we geprobeerd met creatieve bioinformatica-analyses de betrokken cellulaire processen beter te begrijpen (zie Hoofdstuk 3). Op het eerste gezicht leken we behoorlijk succesvol te zijn met onze aanpak en hebben we veel nieuwe inzichten verkregen met de bijbehorende hypothese, dat de waargenomen transcriptoomreactie strikt wordt gereguleerd. Deze regulatie zou het resultaat kunnen zijn van meerdere transcriptiefactoren, die verdeeld over tijd, genen aan en uit zouden schakelen door promotorites met afnemende transcriptiefactor-bindingsaffiniteit. Een van deze transcriptiefactoren zou p53 kunnen zijn. Tijdens het opstellen van een manuscript om onze bevindingen te publiceren, begonnen wij onze aan onze resultaten te twijfelen, ook al waren er in het verleden vergelijkbare bevindingen in andere organismen gepubliceerd. Vooral het hoge percentage differentieel tot expressie gebrachte genen (34%) gecombineerd met hun symmetrische preferentiële genexpressieprofielen over alle tijdstippen en genotypen waren verdacht en deden ons besluiten om de data opnieuw te analyseren. Nauwkeurig onderzoek toonde een aangetaste mRNA/rRNA-ratio aan, die een valide data-analyse verhindert. Verder liet een nieuwe studie naar UV dosis-response zien dat er bij een lage dosis, UV-specifieke en bij een hoge dosis, stress-gerelateerde reacties zijn, wat pleit voor het uitvoeren van studies naar het bereik van UV-dosering bij het ontwerp van experimenten.

De nieuwe “omics”-technieken kunnen beschouwd worden als gevoeliger detectoren om biologische systemen te meten en te observeren. Het gebruik van deze gevoeliger detectoren heeft

echter laten zien dat niet alle gebruikelijke experimentele condities optimaal zijn. Bij de *in-vitro* studies bleken onder andere het opgroeien van de cellen onder atmosferische zuurstofconcentraties behoorlijke specifieke stress op te leveren en *in-vivo* studies met een enkele biopt-afname per muis genereren ook veel ongewenste ruis. Daarom hebben we de experimentele condities geoptimaliseerd voor wat betreft de zuurstofconcentratie niveaus en celcultuur-synchronisatie voor *in-vitro* experimenten (zie Hoofdstuk 3) en RNA-isolatie voor *in-vivo* experimenten (zie Hoofdstuk 4 & 5).

Aangezien de experimentele ontwerpruimte in feite oneindig is, wordt het bereik van elke biologische variabele in de ontwerpruimte meestal gebaseerd op wat gangbaar is en dus vaak op fenotypische eindpunten. Echter, specifieke sub-cellulaire processen zullen vaak slechts ten dele weerspiegeld worden door de fenotypische eindpunten of vallen buiten het hiermee geassocieerde variabelenbereik. Om dit probleem op te lossen hebben we een generiek protocol ontwikkeld, gebaseerd op kleinschalige genexpressie-experimenten, om het relevante variabelenbereik te vinden (zie Hoofdstuk 5). Dit protocol maakt het mogelijk om de juiste locatie in de ontwerpruimte te vinden door het analyseren van de activiteit van al bekende genen van relevante moleculaire processen. Onze pragmatische aanpak is op het volgende gebaseerd: het vaststellen van een specifieke biologische vraag en de bijbehorende sets van genen, het uitvoeren van een experiment met een breed bereik zonder replicatie, het uitsluiten van mogelijk niet-relevante genen, het bepalen van de optimale experimentele locatie door verrijking over de sets van genen te bepalen plus het analyseren van de dosis-respons relatie. Middels *in-vitro* UV-C-blootstelling van MEFs en *in-vivo* UV-B-blootstelling van muizenhuid laten we de toepasbaarheid van deze aanpak zien. In onderzoek naar de vele cellulaire processen die gerelateerd zijn aan de reactie op UV, zoals DNA-reparatie en celcyclus arrest, bleek dat vrijwel ieder cellulair (sub-) proces actief is op een eigen specifieke locatie in de experimentele ontwerpruimte.

Na alle verbeteringen op het gebied van het ontwerp en de uitvoering van transcriptoomexperimenten, hebben we ons gericht op de data-analyse en specifiek op zogenaamde versturende effecten. Bij transcriptoomexperimenten zijn deze versturende effecten regelmatig aanwezig naast de bedoelde experimentele factoren en zullen de resultaten van een analyse beïnvloeden. Zo kunnen onder andere de *in-vivo* experimentele factoren, zoals individu, monstersamenstelling en tijdstip mogelijk geduchte versturende effecten zijn. Om de invloed hiervan te bestuderen hebben we een uitgebreid *in-vivo* transcriptoomexperiment opgezet met UV-B blootstelling van muizenhuid met zes opeenvolgende monsters van elke individuele muis in combinatie met het UV-dosis bereik, dat bepaald was in Hoofdstuk 5. In deze test studie bleek de monstersamenstelling, welke veroorzaakt wordt door muisafhankelijke verschillen in huidsamenstelling, variatie van bemonstering en/of in-/uitstroom van mobiele cellen, het meest prominent versturende effect (zie Hoofdstuk 6). Het bleek dat de versturende effecten: monstersamenstelling, tijdstip, omgangsstress en muis vele genen beïnvloedden met soms tot wel 30 maal veranderde genexpressiewaarden. Hieruit volgde de conclusie dat *in-vivo* transcriptoomexperimenten extreem gevoelig kunnen zijn voor oncontroleerbare en vaak verborgen versturende effecten, die hun resultaten sterk kunnen beïnvloeden en daarmee hun bruikbaarheid beperken.

Gedurende het verloop van dit promotieonderzoek, hebben we meerdere waardevolle lessen geleerd met betrekking tot het doen van transcriptoomexperimenten (zie Hoofdstuk 7). Ook al blijven er nog vele problemen over, toch hebben wij en anderen met ons gedurende de laatste jaren een groot aantal verbeteringen aangebracht met betrekking tot het ontwerp, de uitvoering en de

analyse van transcriptoomexperimenten, welke ongetwijfeld zullen leiden tot betere experimenten, verbeterde resultaten en uiteindelijk diepere inzichten in de complexe biologische systemen.