

9

Nederlandse samenvatting en discussie



Variaties in het humane genoom: geschikte markers voor weefsel identificatie en het testen van allelische instabiliteit

9.1 Samenvatting en algemene discussie

Sinds de ontdekking van "Nuclein" in de winter van 1868/1869 door Friedrich Miescher, wat later DNA werd genoemd, is er een enorme wetenschappelijke vooruitgang geboekt op het gebied van biotechnologie en genetica. De daaropvolgende ontwikkeling van DNA sequencing en de polymerase chain reaction (PCR) hebben deze vooruitgang versneld wat heeft geleid tot een nieuw tijdperk in de wetenschap en geneeskunde. Met behulp van deze technieken werden wetenschappers in staat gesteld specifieke delen van het DNA te bestuderen. Hierdoor werden variaties in oncogenen en tumor suppressor genen (TSG) geïdentificeerd. De sequenties van verschillende humane genomen en de onderlinge overeenkomsten en verschillen werden door de voltooiing van het Humane Genoom Project (HGP) onthuld. De meest frequent gevonden variaties, ongeveer eens in de 100-1.000 baseparen, waren Single Nucleotide Polymorphisms (SNPs). Deze worden gedefinieerd als variaties van één enkele nucleotide op een bepaalde positie van het genoom. Daarnaast wordt de term SNP regelmatig gebruikt met de restricties dat deze variatie bij ten minste 1% van de individuen binnen een bepaalde populatie voorkomt en dat deze niet ziekmakend is. Indien een variatie niet aan deze criteria voldoet, wordt deze beschouwd als een mutatie in plaats van een polymorfisme. Beide beperkingen zorgen echter voor een dilemma. Ten eerste zou een SNP in de ene populatie een mutatie zijn in de andere, aangezien het voorkomen van de SNP tussen verschillende populaties kan variëren (van bv. >1% tot <1%). Ten tweede kan een mutatie van een enkele nucleotide theoretisch gezien in het ene individu leiden tot ziekte terwijl deze in het andere individu geen ziekmakend effect heeft. In het laatste geval zou de enkele nucleotide variatie geen mutatie zijn, maar een SNP. Om dergelijke onduidelijkheden te voorkomen is de term SNP in dit proefschrift gebruikt als een enkele base variatie op een specifieke positie in het genoom die overgeërfd wordt binnen een populatie, ongeacht de frequentie van het voorkomen en de eventuele potentie om ziekte te veroorzaken.

Ieder individu heeft een unieke combinatie van deze SNPs. Een dergelijk SNP genotype kan dan ook als genetische vingerafdruk beschouwd worden. SNPs komen minder frequent voor in coderende gebieden van het genoom in vergelijking tot gebieden die niet coderend zijn. Wanneer een SNP zich in een coderende regio bevindt dan kan de SNP een direct effect hebben op het betreffende eiwit. Een SNP in een niet coderende regio kan door het teweegbrengen van veranderingen in de eiwit expressie echter ook een functioneel effect hebben. De enorme hoeveelheid SNPs verspreid over het genoom en het mogelijke effect van deze enkele nucleotide variaties op de gecodeerde eiwitten en de expressie daarvan zijn

grotendeels verantwoordelijk voor de fenotypische diversiteit. Doordat SNPs zeer frequent en verspreid over het gehele genoom voorkomen, zijn deze enkele nucleotide variaties in potentie een zeer waardevolle (en bijna onuitputtelijke) bron van biomarkers voor o.a. weefsel identificatie en het testen van instabiliteit van tumor gerelateerde allelen.

9.1.1. Technologische vooruitgang

Verskillende technologische ontwikkelingen gedurende de afgelopen decennia hebben er voor gezorgd dat het routinematig gebruik van moleculaire methodes enorm is toegenomen. Momenteel is o.a. de (real time) PCR techniek breed geïmplementeerd voor de detectie van micro-organismen, aan kanker gerelateerde mutaties en afwijkende gen expressie op mRNA niveau. De (real time) PCR techniek is daardoor gemakkelijk toegankelijk voor diagnostische laboratoria en kent vele ervaren gebruikers. Daarnaast worden momenteel meer gespecialiseerde technieken gebruikt voor de eerder benoemde toepassingen van weefsel identificatie en het testen van instabiliteit van tumor gerelateerde allelen. Voor deze toepassingen is real time PCR mogelijk een goed alternatief, aangezien deze techniek gemakkelijk te implementeren is en uitermate geschikt is voor de detectie van SNPs.

Hoofdstuk 2 | Optimalisatie van DNA extractie

Gedurende verschillende studies beschreven in dit proefschrift werd gebruikt gemaakt van formaline gefixeerde weefsels die ingebed waren in paraffine (FFPE weefsel). De kwaliteit van de DNA isolatie uit deze FFPE weefsels was van grote invloed op de analyses die op het DNA werden uitgevoerd. Het isoleren van DNA uit deze weefsels is echter lastig. Ten eerste resulteert de formaline fixatie in de vorming van crosslinks die de amplificatie van het DNA kunnen blokkeren. Ten tweede kan fragmentatie van het DNA plaatsvinden door de zuurgraad van het fixatief en veroudering van het betreffende patiëntenmonster. In hoofdstuk 2 wordt een studie beschreven waarin een vergelijking van DNA extractie methoden plaatsvond voor gebruik in combinatie met de FFPE weefsels. De geteste methodes waren: verhitting, QIAamp, EasyMAG en Gentra. Deze 4 methodes werden getest met en zonder proteïnase K digestie als voorbehandeling. De DNA extracten werden vervolgens getest op i) het cumulatieve effect van extractie efficiëntie en aanwezigheid van remming van de PCR middels de detectie van een interne controle (DNA virus), ii) geschiktheid voor gebruik in real time SNP PCR en iii) geschiktheid voor gebruik in een conventionele multiplex PCR welke amplicons genereerde van 200, 400 en 600 bp. Gentra DNA extractie werd ongeschikt bevonden voor het zuiveren van DNA uit FFPE weefsels. De voorbehandeling middels proteïnase K digestie werd noodzakelijk bevonden voor een zo optimaal mogelijke DNA isolatie uit FFPE weefsels. Wanneer het geëxtraheerde DNA gebruikt werd voor real time SNP PCR bleken de QIAamp en EasMAG methodes het meest geschikt. De amplificatie van de

400 en 600 bp producten was het meest succesvol na QIAamp extractie gevolgd door de EasyMAG methode.

We concludeerden dat DNA extractie methode een grote invloed had op de daaropvolgende analyse, wat overeenkomt met eerdere bevindingen. Het is daarom belangrijk om de DNA isolatie methode af te stemmen op de beoogde analyse die gebruikt gaat worden. Voor de experimenten beschreven in dit proefschrift zal voor het extraheren van DNA uit FFPE weefsels ten behoeve van real time SNP PCR gebruik gemaakt worden van de EasyMAG methode.

Hoofdstuk 3 en 4 | Identificatie van humane weefsels

In klinische laboratoria heeft het voorkomen van verwisselingen van patiënten materialen hoge prioriteit. In de praktijk vinden dergelijke verwisselingen echter plaats, wat kan leiden tot ernstige gevolgen voor de betrokken patiënten. Het is daarom belangrijk potentiële verwisselingen op te lossen door identificatie van de betrokken weefsels. Het in de literatuur gerapporteerde verwisseling percentage van FFPE weefsels en serum ligt rond respectievelijk 0.08-1.2% en 0.1%. Dit benadrukt het belang van stringente monsteridentificatie en het oplossen van potentiële verwisselingen.

Er bestaan momenteel verschillende technieken, veelal gebaseerd op de karakterisatie van korte, repeterende DNA sequenties, het humane leukocyten antigeen systeem (HLA) en andere polymorfe regionen van het genoom, om humaan materiaal te identificeren. Implementatie van deze technieken is echter lastig vanwege het gebruik van kostbare apparatuur, het contaminatie-risico in verband met post-PCR handelingen, slechte DNA kwaliteit (FFPE weefsel) en lage hoeveelheden DNA (serum). Om de implementatie problemen van deze technieken te omzeilen werd een real time PCR gebaseerde test voor het opmaken van SNP profielen ontwikkeld en beschreven in hoofdstuk 3 en 4. De toepasbaarheid van deze test werd gedemonstreerd middels het oplossen van een aantal mogelijke verwisselingen uit de praktijk.

Er werd een panel van 17 SNPs verdeeld over 16 verschillende chromosomen en met allel frequenties van $\approx 0,5$ samengesteld. Zeven van deze 17 SNPs werden uiteindelijk geëxcludeerd op basis van i) een afwijkende allel frequentie binnen de geteste populatie (chi-square test; P waarde $<0,05$, $n=3$), ii) op basis van een *in silico* onderzoek naar de LOH (loss of heterozygosity; verlies van heterozygotie) frequentie van de DNA regio waarin de SNP zich bevindt ($n=4$, waarvan 2 reeds geëxcludeerd op basis van afwijkende allel frequentie) en iii) op basis van technische tekortkomingen ($n=2$). Bij het gebruik van een dergelijk panel kan gesteld worden dat de kans dat 2 willekeurig gekozen individuen hetzelfde SNP profiel hebben 1 op 18.000 is. Hoewel deze kans bij de huidige (commercieel beschikbare) technieken veel kleiner is, bleek de SNP test geschikt om de beschreven casussen van mogelijke monsterverwisselingen (totaal $n=7$; FFPE weefsels $n=6$; serum monsters $n=1$) op te lossen. Onze bevindingen werden bevestigd door aanvullend histologisch of serologisch onderzoek.

Een voordeel van de beschreven SNP test is dat er zeer kleine stukjes DNA van ongeveer 80 bp geamplificeerd worden, terwijl bij andere beschikbare technieken een amplicon grootte van wel 360 bp benodigd is. Daarnaast kan de gepresenteerde SNP test verricht worden middels iedere standaard real time PCR cyclus, in plaats van dure sequencing apparatuur. Andere significante voordelen van de SNP test zijn de snelle doorlooptijd en het gebrek aan post-PCR handelingen waardoor de kans op amplicon contaminatie erg klein is.

Hoofdstuk 5 | Bepalen van de *JAK2V617F* mutatie status

Het *Janus Kinase 2 (JAK2)* gen codeert voor het JAK2 tyrosine kinase. JAK2 speelt een belangrijke rol in de door cytokines gemedieerde signaaltransductie en heeft daardoor een grote invloed op de hematopoïese. De activerende *JAK2V617F* mutatie resulteert in onafgebroken activatie van de JAK-STAT signaaltransductie en wordt frequent gevonden in myeloproliferatieve ziekten (MPZ) zoals polycythemie vera (PV), essentiële thrombocythemie (ET) en chronische idiopathische myelofibroze (IMF). Naast de detectie van de *JAK2V617F* mutatie, is ook de hoeveelheid die van dit mutant allel aanwezig is mogelijk van klinisch belang. Hoofdstuk 5 beschrijft de ontwikkeling van een real time PCR voor de detectie en kwantificering van *JAK2V617F* in patiënten met een mogelijke MPZ. Om de gevoeligheid van de beschreven real time PCR te maximaliseren werd amplificatie van het wildtype allel geblokkeerd door gebruik van een locked nucleic acid (LNA) probe in vergelijking tot een peptide nucleic acid (PNA) probe. De gevoeligheid en lineariteit werden vervolgens bepaald middels een verdunningsreeks (0,006% tot 97% *JAK2V617F* DNA verdunt in wildtype DNA). Zowel de gevoeligheid als de lineariteit van de test in combinatie met het gebruik van de PNA probe (0,05-97%) lagen hoger in vergelijking tot het gebruik van de LNA probe (0,2-97%). De ontwikkelde real time PCR in combinatie met het gebruik van de PNA probe liet daarmee een gevoeligheid zien vergelijkbaar met die van verschillende reeds beschreven testen.

In 9 van de 100 geteste bloedmonsters van gezonde donoren (<30 jaar) werden *JAK2V617F* achtergrondsignalen gezien. Eén van deze donoren liet in duplo een dergelijk achtergrond-sig-naal zien. Deze observatie kan veroorzaakt zijn door een technisch artefact of mogelijk was het *JAK2V617F* allel daadwerkelijk aanwezig bij deze patiënten. Deze bevindingen komen overeen met die in de literatuur, al blijft de klinische relevantie van dit fenomeen onduidelijk. Om te voorkomen dat de test in een diagnostische setting resulteert in vals positieven werd er een Ct-waarde van 37,9 aangehouden als grens.

Hoofdstuk 6 | LOH van het *JAK2* locus

LOH is een genetisch incident waarbij heterozygotie van een bepaald locus verloren gaat. LOH kan daarmee bijdragen aan het vergaren van de mutaties die benodigd zijn voor het

ontstaan van kanker. In combinatie met *JAK2V617F* kan LOH resulteren in homozygotie van dit mutant allel waardoor het effect ervan mogelijk versterkt wordt. Het detecteren van LOH van het *JAK2* locus is daarom mogelijk klinisch relevant. In hoofdstuk 6 wordt de ontwikkeling van een *JAK2* LOH test met behulp van real time PCR gebaseerde SNP detectie beschreven. Daarnaast werd de SNP gebaseerde *JAK2* LOH test vergeleken met een in de literatuur beschreven STR (short tandem repeat) gebaseerde test. Twee patiënten groepen werden geïncludeerd: een groep bestaande uit 12 bekend *JAK2V617F* positieve patiënten (waarvan 6 met 25-50% mutant allel en 6 >50% mutant allel) en een groep van 81 patiënten met een mogelijke MPZ. In de 6 patiënten met 20-25% mutant allel en in de 6 patiënten met >50% mutant allel werd op basis van de SNP test *JAK2* LOH vastgesteld in respectievelijk 2 en 6 patiënten. In de groep patiënten met mogelijke MPZ werd bij 6 van hen *JAK2* LOH vastgesteld. Bij het beschouwen van de STR test als gouden standaard, had de SNP test een gevoeligheid en negatief voorspellende waarde van 100% in beide patiënten groepen. De positief voorspellende waarde van de SNP test in de groep met bekend *JAK2V617F* positieven en in de groep met mogelijke MPZ patiënten waren respectievelijk 85,7% en 100%. De specificiteit van de SNP test was 80% en 100% in deze groepen. De SNP test bleek een hoger slagingspercentage te hebben dan de STR test (3 versus 9 niet informatieve profielen). Daarnaast genereerde de SNP test meer gedetailleerde informatie met betrekking tot het *JAK2* locus.

Hoofdstuk 7 | Allelische instabiliteit van het *HER2/TOP2A* locus

Borstkanker is een heterogene aandoening die een verscheidenheid aan biologische aspecten kent. Op basis van o.a. een verhoogde expressie van eiwitten zoals de oestrogeen, progesteron en humane epidermale groeifactor receptoren (ER, PR en *HER2*), wordt borstkanker verdeeld in verschillende categorieën. Deze categorisatie wordt gebruikt om een voorspelling te doen van het klinisch verloop van de ziekte en om bijpassende therapieën te selecteren. Borsttumoren negatief voor zowel ER, PR als *HER2* worden “triple negative” (TN) genoemd. Deze TN tumoren worden geassocieerd met een slechte prognose en een gebrek aan gerichte therapie.

Amplificatie van het *HER2* gen leidt tot overexpressie van het *HER2* eiwit. Dit eiwit speelt een belangrijke rol in de *HER* signaaltransductie. *HER2* amplificatie komt ongeveer in 25-30% van de borstkanker gevallen voor. Daarnaast vindt in 30-45% van de *HER2* geamplificeerde gevallen coamplificatie plaats van het naastgelegen topoisomerase II alpha (*TOP2A*) gen. *TOP2A* is belangrijk bij de topologische veranderingen van het DNA benodigd gedurende de DNA replicatie. *TOP2A* deletie vindt plaats in ongeveer 15-40% van de *HER2* geamplificeerde borsttumoren. Patiënten met een dergelijke *TOP2A* amplificatie lijken meer baat te hebben bij behandeling met anthracycline (AC), terwijl deletie van *TOP2A* juist een negatieve uitwerking op de effectiviteit van deze behandeling lijkt te hebben.

Instabiliteit van het *HER2/TOP2A* allel wordt geassocieerd met een verhoogde chromosomale instabiliteit wat mogelijk een verhoogde effectiviteit van AC therapie tot gevolg heeft. In hoofdstuk 7 wordt de ontwikkeling van een op real time PCR gebaseerde SNP test om deze allelische instabiliteit vast te kunnen stellen beschreven. Er werden 44 patiënten met 3 verschillende borstkanker typen geïnccludeerd: borstkanker met HER2 overexpressie (n=15), TN borstkanker (n=16) en ER en PR positieve borstkanker (n=13). Daarnaast werd in deze studie een negatieve controle groep van 10 patiënten zonder borstkanker opgenomen. Alle borsttumoren met HER2 overexpressie lieten allelische instabiliteit van het *HER2/TOP2A* locus zien. Daarnaast was dit bij 12/16 (75%) van de TN borsttumoren en bij 5/13 (38%) ER en PR positieve borsttumoren het geval. Alhoewel het aantal onderzochte borsttumoren in deze studie gering was, kwamen de percentages allelische instabiliteit in de TN groep en de ER en PR positieve groep van respectievelijk 75% en 38% ruwweg overeen met de AC responspercentages van 85% en 47% in deze groepen. Een (5-jaars) klinische vervolgstudie is echter benodigd om vast te stellen of allelische instabiliteit van het *HER2/TOP2A* locus daadwerkelijk klinisch relevant is.

9.2 Toepassing buiten de scope van dit proefschrift: bacteriële SNP genotypering

Naast de focus op diverse humane toepassingen, werd de inzetbaarheid van real time PCR gebaseerde SNP detectie op het gebied van bacteriële genotypering onderzocht. Gedurende een periode van 4 jaar (2007-2010) werd Nederland geconfronteerd met de grootste Q-koorts (*Coxiella burnetii*) epidemie ooit. Op het moment van deze uitbraak was er geen epidemiologische data beschikbaar, noch bestond een directe typeringsmethode. Als eerste werd een panel van 10 onderscheidende SNPs gebruikt om een internationale verzameling van 28 referentie stammen met beschikbare Multi Locus Sequence Typing (MLST) data en 5 commerciële stammen te genotypen. Ten tweede werden 40 monsters gerelateerd aan de uitbraak (14 van humane en 26 van dierlijke oorsprong) getypeerd middels het SNP panel. Als derde werd de SNP genotypering gebruikt om vast te stellen of de verplichte vaccinatie van melkgeiten ertoe kon leiden dat *C. burnetii* DNA afkomstig van het vaccin in de bulkmelk van deze geiten terecht zou kunnen komen. Dit zou namelijk kunnen leiden tot het onterecht ruimen van de betrokken dieren.

De verzameling referentie stammen waarin zich 14 verschillende MLST types bevonden bevatte 9 SNP genotypen. Onder de uitbraak gerelateerde monsters werden 5 verschillende SNP genotypen geïdentificeerd (3 in de humane monsters en 4 in de dierlijke monsters). Dit suggereert dat de Nederlandse uitbraak waarschijnlijk niet te wijten is aan 1 hypervirulent *C. burnetii* type, maar dat omgevingsfactoren (zoals een hoge bevolkingsdichtheid in combinatie met de aanwezigheid van veel geitenboerderijen en droge perioden gedurende de lente) de

uitbraak bespoedigd hebben. In de bulkmelkmonsters werd vastgesteld dat het aanwezige *C. burnetii* DNA zeer waarschijnlijk afkomstig was van het toegediende vaccin. Dit had tot gevolg dat het Nederlandse Ministerie van landbouw, Natuur en Voedselkwaliteit het vaccinatiebeleid herzag en een 2 weken durende periode tussen vaccinatie en het testen van de bulkmelk introduceerde.

9.3 Slotbeschouwing

Zoals gedemonstreerd middels de studies beschreven in dit proefschrift, heeft real time gebaseerde SNP detectie momenteel een breed spectrum aan klinisch relevante toepassingen. De techniek kan daarnaast gemakkelijk geïmplementeerd worden in ieder moleculair diagnostisch laboratorium. Dit maakt real time gebaseerde SNP detectie een aantrekkelijk diagnostisch middel. We voorzien daarom dat de beschreven testen een startpunt zijn voor de ontwikkeling van andere, diagnostisch relevante analyses op het gebied van de oncologie en bacteriële SNP genotyping. Daarnaast zou de SNP identificatie van alle humane weefsels in behandeling binnen een klinisch laboratorium bewerkstelligd kunnen worden door een brede implementatie van de daarvoor ontwikkelde SNP test. Dit zou resulteren in een verbeterde monsteridentificatie en het aantal verwisselingen kunnen reduceren. Om dit te realiseren zal het aantal te detecteren SNPs uitgebreid moeten worden zodat het onderscheidend vermogen van de test toereikend is. Daarnaast zullen de kosten drastisch gereduceerd moeten worden door de real time PCR volumina sterk te verlagen en gebruik te maken van apparatuur met 384- of zelfs 1536-wells formaat.

Uiteindelijk zal (next generation) sequencing van het gehele genoom de toegevoegde waarde van andere moleculaire technieken zoals real time SNP PCR mogelijk teniet doen. Momenteel is de benodigde apparatuur hiervoor echter nog zeer kostbaar en niet toegankelijk voor de meeste moleculair diagnostische laboratoria. Daarnaast is de analyse van de enorme hoeveelheid gegenereerde data een grote uitdaging. Als diagnostisch middel zal real time PCR in de nabije toekomst daarom waarschijnlijk de meest voor de hand liggende keuze blijven voor moleculair diagnostische laboratoria.