



Universiteit  
Leiden  
The Netherlands

## **Mutagenic mechanisms in normal and neoplastic B cells: from AID-induced diversification to genome-wide patterns**

Sepúlveda Yáñez, J.H.

### **Citation**

Sepúlveda Yáñez, J. H. (2024, November 12). *Mutagenic mechanisms in normal and neoplastic B cells: from AID-induced diversification to genome-wide patterns*. Retrieved from <https://hdl.handle.net/1887/4108983>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

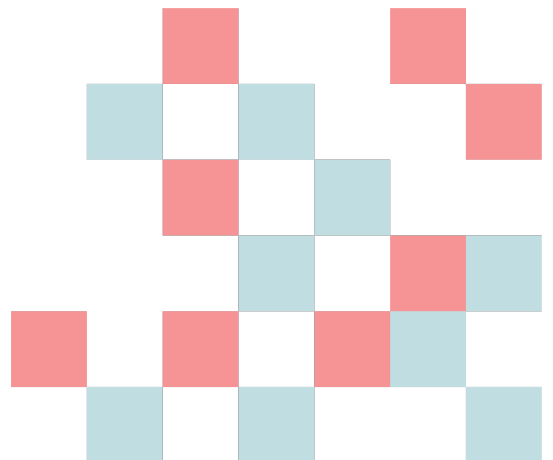
Downloaded from: <https://hdl.handle.net/1887/4108983>

**Note:** To cite this publication please use the final published version (if applicable).



*Estrecho de Magallanes, Chile*

# APPENDIX





## NEDERLANDSE SAMENVATTING

### ANTILICHAAM DIVERSITEIT

Het adaptieve immuunsysteem is een essentieel onderdeel van onze afweer en is in staat om heel gericht te beschermen tegen onder andere schadelijke indringers van buitenaf. Immunoglobulinen, ofwel antilichamen, spelen hierin een cruciale rol door te binden aan structuren (antigenen) van ziekteverwekkers en ze zo onschadelijk te maken. Immunoglobulinen zijn gespecialiseerde eiwitten die worden geproduceerd door B-cellen en tot expressie worden gebracht op het oppervlak als een B-cel receptor (BCR). De B-cel kan uitgrijpen tot plasmacel die de antilichamen produceert en uitscheidt en zo zorg draagt voor het elimineren van de ziekteverwekker. Daarnaast kan de B-cel een memory cel worden die in geval van een her-infectie direct overgaat tot snelle productie van effectieve antilichamen. De specificiteit van immunoglobulinen om antigenen te herkennen wordt bepaald door twee verschillende processen. Tijdens de ontwikkeling van elke B-cel vormen combinaties van verschillende gen segmenten de zware en lichte keten van de BCR. Dit proces wordt VDJ recombinatie genoemd en plaatst een variabel (V), diversiteit (D) en joining (J) segment na elkaar zodat een aaneengesloten VDJ gen ontstaat. Tijdens deze recombinatie dragen zowel combinaties van willekeurige gen segmenten als variaties die ontstaan bij het aan elkaar koppelen van de gen segmenten bij aan de diversiteit van immunoglobulinen. Vanuit deze diversiteit kan een veelheid van antigenen tot activatie van de B-cel kan leiden. Ten tweede kan de bindingscapaciteit van immunoglobulines verbeteren, ofwel rijpen, door het verkrijgen van mutaties. Dit proces vindt plaats onder zeer strenge regulatie in kiemcentra van reactieve lymfeklieren en heet somatische hypermutatie (SHM). SHM wordt geïnitieerd door het enzym activation-induced cytidine deaminase (AID). AID deamineert de nucleobase cytidine (C) en heeft een voorkeur voor specifieke motieven in het DNA waaronder het canonieke WRCY motief en de niet-canonieke WA, en RCG motieven. De letters W, R en Y staan voor combinaties van meerdere nucleobasen.

Het fysiologische mutagene AID-proces splitst een amine-groep af van cytosine waardoor de nucleobase uracil ontstaat. Dit enzymatische proces is uiterst efficiënt omdat de hoge affiniteit van AID voor enkelstrengs DNA een bindingsduur van meer dan 5 minuten waarborgt. Daardoor kan AID meerdere deaminaties per keer uitvoeren, wat de incidentie van mutaties verhoogt en uiteindelijk een snelle rijping van de antilichamen veroorzaakt. De aanwezigheid van de nieuwe uracil nucleobase tegenover guanine is een DNA-mismatch en moet daarom hersteld worden. De B cel heeft hiervoor twee mechanismen tot haar beschikking.

### DNA-HERSTELMECHANISMEN

Base Excision Repair (BER) en Mismatch Repair (MMR) zijn twee herstelmechanismen die betrokken zijn bij het repareren van DNA-schade, zoals de door AID veroorzaakte deaminatie van cytosine naar uracil. Beide mechanismen zijn betrokken op een manier waarbij fout-gevoelige DNA-polymerasen betrokken zijn. Het moleculaire mechanisme van BER

bevat UNG2 of SMUG, die delen van het DNA waarin uracil voorkomt, herkennen en losmaken uit het dubbelstrengs DNA waarna uracil weggeknipt kan worden. De deletie die daardoor ontstaat activeert andere componenten van BER, waarbij voor kleine en grote deleties verschillende herstel-routes worden bewandeld. Polymerasen zoals Pol- $\zeta$  en Pol- $\iota$  vullen de deletie op met nieuwe basen maar maken daarbij fouten. Een ander mechanisme dat uracil-guanine mismatches corrigeert is MMR. De uracil base wordt herkend door de eiwitten MSH2/MSH6 waarna exonuclease EXO1 een grote deletie veroorzaakt. Vervolgens wordt PCNA gemonoubiquitineerd, wat vervolgens foutgevoelige translesie synthese (TLS) polymerasen zoals Pol- $\eta$ , Pol- $\zeta$  en REV1 bij de reactie betreft die de deletie opvullen. Bij normale immuun reacties zijn de fouten die BER en MMR tijdens dit DNA herstel achterlaten essentieel voor de B-cel antilichaam rijping en diversificatie van het BCR repertoire.

## HEMATOLOGISCHE NEOPLASMA'S

Folliculair lymfoom (FL) is een indolente hematologische B-cel maligniteit met een incidentie van 2-4 per 100.000 mensen per jaar. Hoewel de term 'indolent' inhoudt dat progressie van de ziekte een traag verloop kent, blijft FL een ongeneselijke vorm van lymfeklierkanker. Daarnaast transformeert de ziekte op jaarbasis bij 1-2% van de patiënten naar een moeilijk behandelbaar agressief lymfoom. Een kenmerk van FL is de translocatie van BCL2 op chromosoom 18 naar het immunoglobuline zware keten locus op chromosoom 14, die op kan treden tijdens de V(D)J recombinitie. Deze translocatie leidt tot overexpressie van BCL2 waardoor de cel ontsnapt aan normale mechanismen die bedoeld zijn om ongewenste cellen te elimineren (apoptose). Op zichzelf is dit niet voldoende om de ziekte te veroorzaken maar door de aanwezigheid in kiemcentrum reacties en de doorlopende activiteit van AID, heeft FL een hoog niveau van SHM. Hierdoor blijven B-cellen met een lage tot gemiddelde immunoglobuline-affiniteit, die normaal gesproken in apoptose zouden gaan, in leven. Zo dragen ze bij aan de vroege vorming van lymfoom. De doorlopende activiteit van AID resulteert ook in mutaties in andere genen dan immunoglobulinen, wat in feite abnormale SHM is. Deze mutaties treffen vaak genen die geassocieerd zijn met signaling pathways, immuun herkenning, celcyclus regulatoren en epigenetica en met name de opeenstapeling deze afwijkingen is kenmerkend voor FL.

Een andere belangrijke factor bij de ontwikkeling van FL is het tumor micro-milieu (tumor micro-environment, TME). In tegenstelling tot andere lymfomen behoudt FL de normale 3-dimensionale morfologische kenmerken van een kiemcentrum reactie zonder zich diffuus te verspreiden door de lymfeklier. FL lijkt daarmee sterk afhankelijk van componenten van het TME zoals andere lymfoïde cellen, stromale cellen en de extracellulaire matrix. Het samenspel van genetische afwijkingen en het TME dragen bij aan het ontstaan van FL.

Chronische lymfatische leukemie (CLL) is de meest voorkomende leukemie bij volwassenen in westerse landen, met een incidentie van 6,8 mannen en 3,5 vrouwen per 100.000 mensen per jaar. CLL wordt geclassificeerd in twee groepen op basis van de mutatiestatus van de BCR, waarbij de ongemuteerde groep (98-100% identiteit met het ongemuteerde zware

keten V-gen) een agressiever klinisch verloop kent in vergelijking met de gemuteerde groep (identiteit < 98%). De ontwikkeling van CLL is vaak gekoppeld aan een fase die bekend staat als monoclonale B-cel lymfocytose (MBL) met cellen die fenotypisch niet te onderscheiden zijn van het CLL-fenotype CD19+CD5+CD20lowCD19low. Hoewel het een voorloper stadium lijkt, zal slechts 1-2% van de MBL transformeren tot CLL. MBL wordt geclassificeerd in laag-aantal MBL en hoog-aantal MBL op basis van het aantal cellen met CLL-fenotype in bloed.

De factoren die betrokken zijn bij de progressie van MBL naar CLL zijn nog onbegrepen en bieden de mogelijkheid nauwgezet te zoeken naar verschillen en overeenkomsten tussen normale en maligne cellen. Het feit dat FL en CLL beide ontstaan uit rijpe B-cellen maakt het bovendien interessant om FL en CLL in parallel te analyseren.

CLL heeft genetische veranderingen die betrokken kunnen zijn bij progressie van de ziekte. Door middel van associatiestudies zijn 45 kiembaan varianten in genomische loci geïdentificeerd die het risico op het ontwikkelen van de ziekte met 2-8 keer verhogen. Somatische mutaties zijn zeer variabel in CLL, met uitzondering van mutaties in 5 genen die frequent gemuteerd zijn (SF3B1, ATM, TP53, POT1 en NOTCH1). Copy number varianten komen daarentegen voor bij 80% van de patiënten en zijn nuttig bij de diagnostiek van de ziekte en zijn relevant voor het bepalen van de optimale behandeling. Bovendien vertoont CLL een specifiek oncogeen mechanisme dat wordt veroorzaakt door de BCR: autonome signalering. Dit mechanisme, dat een continue activerend signaal veroorzaakt zonder dat er sprake is van stimulatie door een extern antigeen, is de enige eigenschap die gezien wordt bij alle gevallen van CLL en is daarom zeer waarschijnlijk onontbeerlijk voor de ontwikkeling van deze ziekte.

## DIT PROEFSCHRIFT

Ons immuunsysteem omvat een groot repertoire aan B cellen dat in staat is om een vrijwel onbeperkt scala aan antigenen te herkennen. Dit wordt bereikt door de stochastische combinatie van genen en het fysiologisch proces van somatische hypermutatie wat mutaties in de B-cel receptor veroorzaakt. De deaminatie van cytosine naar uracil door het enzym AID speelt hierbij een centrale rol. AID richt zich bij voorkeur op specifieke motieven in het DNA. Het reparatieproces ten gevolge van deze deaminaties wordt uitgevoerd door middel van BER of MMR. Het gebruikelijke resultaat van dit fout gevoelige reparatieproces is een enkele nucleotide substitutie, maar het is ook mogelijk om tandem mutaties te produceren, waarbij twee of meer aangrenzende nucleotiden gelijktijdig worden gemuteerd.

In **hoofdstuk 2** hadden we als doel om deze tandem dinucleotide substituties (TDNS) in immuunglobulinegenen te identificeren en de achterliggende mechanismen te karakteriseren bij gezonde individuen en patiënten met deficiënties in DNA herstelmechanismen. Hiervoor hebben we eerst alle mutaties in BCR repertoires vastgesteld door middel van ‘massive parallel sequencing’ op populaties van gezuiverde B-cellen van 12 gezonde donoren en 5 patiënten met DNA-reparatie deficiëntie waarvan 4 in MMR en 1 BER. Door eerst het resultaat van AID mutagenese te simuleren op de ongemuteerde counterparts van deze BCR repertoires konden we onderscheid maken tussen twee toevallig naastgelegen mutaties als gevolg van sequentiële

mutatie, en werkelijke TDNS als gevolg van één enkele AID-hit. Dit leidde tot het verwijderen van 46.2% van de kandidaat TDNS waardoor 5775 mutaties overbleven voor verdere analyse. Hiervan was 6% een tandem mutatie, en 4.1% een TDNS. Er was geen verschil tussen bepaalde V-genen als we de frequentie van TDNS vergeleken met de frequentie van 'single' nucleotide substituties (SNS). Bijna alle gevallen van TDNS (98,6%) hadden een aminozuur verandering tot gevolg, tegenover 71,6% van de SNS. In een groot deel (89,6%) van de TDNS werd de kiembaansequentie gedeeltelijk of helemaal gekopieerd, wat het frequent voorkomen van inverterende tandem substituties kan verklaren. Bij de analyse van BCR-sequenties van patiënten met de DNA-reparatie deficiëntie MMR vonden we een vergelijkbaar patroon van TDNS. Bij de patiënt met BER-deficiëntie daarentegen ontbraken TDNS volledig, terwijl we bij deze patiënt een vergelijkbaar niveau van SNS vonden. Dit suggereert dat tandem mutaties in BCR gerelateerd zijn aan BER-activiteit, maar omdat we dit in slechts 1 patiënt zagen dient dit verder gevalideerd te worden in een grotere patiëntengroep.

Na het bestuderen van mutagene mechanismen die ten grondslag liggen aan veranderingen in de BCR sequentie, was onze volgende vraag of het mogelijk was om het bestaan van twee immunoglobuline transcript populaties binnen één cel te detecteren die verschillen door één of enkele nucleotiden. Immers, onmiddellijk nadat AID een mutatie heeft veroorzaakt, kunnen tegelijkertijd BCR transcripten van vóór de mutatie, en ook al nieuwe transcripten mét de recent verkregen mutatie(s) in de cel aanwezig zijn. In **hoofdstuk 3** hebben we deze B-cellen gedetecteerd en gekwantificeerd. Hiervoor hebben we resultaten van single-cell RNA sequencing (scRNA-seq) en single-cell B-cell receptor sequencing gebruikt van 12 FL-patiënten en 5 CLL-patiënten als controle. In een zelf ontwikkeld bio informatica proces hebben we na zeer strenge filtering 364 cellen geïdentificeerd die inderdaad twee verschillende immunoglobuline transcripten bevatten. Deze cellen vonden we in 7 van de 12 FL-patiënten, terwijl geen enkele cel met 2 BCR transcripten werd gedetecteerd bij de CLL-patiënten. In 85% van de gevallen konden we bepalen welk transcript de oorspronkelijke en welk transcript de nieuwe varianten was. Deze mutaties bleken in 41% C-T-transities te zijn en hadden in 27% plaatsgevonden in het canonieke WRCY-motief van AID, wat consistent is met bevindingen uit exoom-brede sequencing in een onafhankelijke cohort. Daarna hebben we de volledige transcriptomen van dezelfde cellen verkregen door scRNA-seq vergeleken: cellen met 2 BCR-transcripten tegenover cellen met slechts 1 BCR transcript. Hier vonden we dat 5766 genen verschillend tot expressie kwamen. Het meest in het oog sprong de hogere expressie van AID in cellen met twee BCR-transcript varianten. Door de 5766 genen te associëren met groepen van genen waarvan in de literatuur bekend is dat deze betrokken zijn bij bepaalde cellulaire processen konden we vaststellen dat MMR prominenter is dan BER in de FL-cellen met 2 BCR-transcripten. Deze analyses, die aantonen dat FL-cellen twee verschillende BCR-transcripten kunnen bevatten waarvan er één het directe gevolg van AID is, tonen voor het eerst aan dat in FL voortdurende somatische hypermutatie plaatsvindt.

In het tweede deel van het proefschrift hebben wij ons verder verdiept in het SHM-proces, met focus op de mechanismen die normale cellen kunnen transformeren tot maligne cellen. Bij AID maken we onderscheid tussen de normale activiteit van AID om immunoglobuline

genen te muteren, en de afwijkende AID activiteit buiten de immunoglobuline regio's. In **hoofdstuk 4** hebben wij ons gericht op het identificeren van de afwijkende activiteit van AID in het totale landschap van mutaties in FL en CLL. Hoewel FL en CLL een vergelijkbaar patroon vertoonden met overwegend C-naar-T mutaties, vonden we deze vaker in een AID-motief bij FL dan bij CLL (42,7% versus 33,6%). Naast AID zijn er vele veroorzakers van mutaties, zoals chemicaliën, uv-licht en tabaksrook, en deze laten ieder hun eigen mutatie signatuur achter. Bij vergelijking van afwijkende FL en CLL mutatie signaturen met een database waarin al deze signaturen gerubriceerd zijn, bleek inderdaad dat bij FL meestal AID-gerelateerde signaturen betrokken zijn. Bij CLL daarentegen hadden kanker gerelateerde signaturen de overhand. AID-mutaties kunnen alleen plaatsvinden in actief DNA dat afgerold is en wordt afgelezen. Met de zogenaamde HiC-techniek kan onderscheid gemaakt worden tussen mutaties in actief en niet-actief DNA. In data verkregen met deze HiC-techniek troffen we het canonieke AID signatuur en ook activiteit van DNA reparatie mechanismes vaker aan in actief DNA bij FL dan bij CLL. Foutieve correctie van DNA schade met mutaties tot gevolg kan het resultaat zijn van mutaties in de reparatiegenen zelf of in genen die onderdeel uitmaken van ketens waarbinnen de reparatiegenen werken. Mutaties bij FL in deze genen vonden we vaker in de DNA reparatie ketens Fanconi anemie en BER. In CLL kwamen de meeste mutaties in de DNA damage response keten maar het niveau ervan verschilde niet met FL. Deze bevindingen kunnen verklaren waarom FL en CLL een verschillend landschap aan mutaties vertonen.

In het laatste hoofdstuk van dit proefschrift hebben wij geprobeerd de dynamiek tussen de genetische componenten en BCR-siginaaloverdracht in de progressie van MBL naar CLL te bepalen. Het autonome BCR signaal bij CLL is een uniek oncogeen mechanisme dat niet ontstaat door binding aan een externe antigeen, maar wordt veroorzaakt door de herkenning van de eigen BCR. In **hoofdstuk 5** hebben we van 191 broers en zussen van CLL-patiënten een bloedmonster verkregen en gescreend op aanwezigheid van populaties van MBL cellen die herkenbaar zijn aan het CLL-fenotype (CD19+CD5+CD20lowCD19low). Wij vonden deze populaties in 34 monsters met een grote variatie in percentage CLL-fenotype cellen. In 17 gevallen konden we de volledige sequentie van de BCR achterhalen waarvan er vijf een sterke overeenkomst hadden met BCR's die we frequent in CLL vinden. Alle BCR's van MBL CLL-fenotype cellen hadden de aminozuur patronen waarvan aangenomen wordt dat deze betrokken zijn bij het ontstaan van het autonome signaal. Van 11 BCR's van MBL-individueen hebben we de sterkte van het autonome signaal gekwantificeerd. Het signaal was in alle experimenten aanwezig maar was structureel lager dan het signaal dat we in BCR's van CLL konden meten. Daarnaast vonden we een verband tussen de grootte van de MBL populatie en de sterkte van het autonome signaal. Door middel van genetische analyses vonden we een link met 12 kiembaan genotypes. Daarnaast waren 24 polygenetische risico factoren verhoogd in MBL en CLL vergeleken met de algemene bevolking. Daarnaast werd een copy number variatie (CNV) analyse uitgevoerd op 11 CLL en 15 MBL monsters. In alle monsters werden CNV's aangetroffen waarbij de CNV's in CLL voornamelijk klonaal waren en die in MBL subklonaal, wat suggereert dat de MBL populaties kunnen bestaan uit meerdere, verschillende

expansies van MBL cellen. Tot slot hebben we gekeken naar de somatische varianten binnen de gepaarde CLL-MBL samples. Het totaal aantal somatische varianten was gelijk in beide groepen, maar in MBL vonden we wederom een hoger aandeel aan sub-klonale somatische varianten. Deze bevindingen laten zien dat naast de fenotypische overeenkomsten tussen MBL en CLL het autonome BCR signaal ook aanwezig is in MBL en dat geëxpandeerde MBL klonen eveneens genetische afwijking bevatten die kenmerkend zijn voor CLL.

Samenvattend hebben wij mutagene mechanismen bestudeerd die geassocieerd zijn met antilichaam rijping en diversificatie. Wij analyseerden tandem mutaties bij gezonde donoren en bij patiënten met DNA-reparatie deficiëntie. Daarbij hebben wij een moleculair mechanisme voorgesteld voor het ontstaan van deze mutaties. Daarnaast hebben wij somatische hypermutatie op enkel-celniveau onderzocht, waarbij FL als model voor voortdurende SHM werd gebruikt. Dit stelde ons voor het eerst in staat om SHM op het niveau van één enkele B-cel te detecteren en te kwantificeren. Het tweede gedeelte van het proefschrift behandelde pathologische mechanismen, waaronder de afwijkende somatische hypermutatie in FL en CLL. Wij hebben het globale landschap en het effect van de pathologische activiteit van AID gekarakteriseerd en onderscheidende patronen tussen FL en CLL gedetecteerd. Ten slotte hebben we de relatie tussen genetische varianten en autonome signaaloverdracht bij individuen met MBL bestudeerd. MBL vertoonde een hoger aandeel sub-klonale somatische varianten in vergelijking met CLL. Interessant is dat alle geanalyseerde gevallen autonome signaaloverdracht vertoonden, maar met minder signaalsterkte dan bij CLL, wat de belangrijke rol van deze oncogene mechanisme in de progressie van MBL naar CLL aantoont.