

Bachelorproef

Studiegebied Gezondheidszorg
Bachelor Biomedische Laboratoriumtechnologie
Afstudeerrichting Farmaceutische en biologische laboratoriumtechnologie
Keuzetraject Bio-informatica
Academiejaar 2012-2013
Studente Liesbet Martens

Thema

Analyse van *Xbp1 knock-out* microarray data in conventionele dendritische cellen

Stageplaats

VIB, Gent
(Vlaams Instituut voor Biotechnologie)

Voor akkoord verklaring

Dit eindwerk is een examen; eventuele fouten die worden vastgesteld tijdens de eindwerkverdediging of erna werden niet gecorrigeerd. Het gebruik als referentie in publicaties is toegelaten na goedkeuring van de promotor en eindwerkbegeleider (van de stageplaats).

Paco Hulpiau

Prof. Dr. Bart Lambrecht

Stagementor

Stagegever

Raphaël Kiekens

Promotor

Woord vooraf

Een stage doorlopen, gekoppeld aan het schrijven van een eindwerk, is het laatste onderdeel van de opleiding Bachelor in Biomedische en Farmaceutische Laboratoriumtechnieken, afstudeerrichting Bio-informatica. Gedurende veertien weken heb ik meegewerkt op de afdeling *Department for Molecular Biomedical Research* (DMBR) in het Vlaams Instituut voor Biotechnologie (VIB) te Gent.

Mijn initiële opdracht was het analyseren van microarray data. Deze opdracht is uiteindelijk geëvolueerd tot het eveneens uitvoeren van netwerkanalyses en motiefanalyses.

Mijn resultaten en dit eindwerk zou ik niet hebben kunnen bekomen zonder de hulp van volgende mensen, die ik van harte wil bedanken.

Prof. Dr. Bart Lambrecht die me de mogelijkheid gaf stage te lopen in het VIB en meer specifiek in zijn onderzoeksgroep, *Immunoregulation and Mucosal Immunology*.

Mijn stagementor Paco Hulpiau voor de uitleg, de inzichten en het nalezen van dit eindwerk.

De volledige *BioIT Core Facility* voor de raad en de inzichten.

Dr.Sc. Martin Guilliams voor de steun, het nalezen van een deel van dit werk en het beantwoorden van mijn vragen met betrekking tot immunologie.

Mijn stagebegeleider Raphaël Kiekens voor de begeleiding en het nalezen van dit eindwerk.

Ten slotte wil ik ook mijn ouders bedanken omdat ze me de mogelijkheid boden te studeren. Ik wil hen ook bedanken voor de steun en de raad die ze me gaven gedurende mijn volledige opleiding.

Samenvatting

Xbp1 is een transcriptiefactor die upgereguleerd wordt tijdens stress in het endoplasmatisch reticulum (ER). Recent onderzoek heeft aangetoond dat *Xbp1* essentieel is voor de ontwikkeling en overleving van dendritische cellen (DC). DC vormen de link tussen het aangeboren, niet-specifieke immuunsysteem en het verworven, specifieke immuunsysteem. Bovendien zijn het de DC die beslissen over de differentiatie van T-cellen en zijn dus een belangrijke component in de beslissing om tolerantie te vertonen of juist inflammatie op te wekken. Er kunnen vier subsets van DC onderscheiden worden: $CD8\alpha^+$ conventionele DC ($CD8\alpha^+$ cDC), $CD11b^+$ conventionele DC ($CD11b^+$ cDC), plasmacytoïde DC (pDC) en *monocyte-derived* DC (moDC). Deze vier subtypes vertonen allen verschillen op vlak van omgang met een antigeen en interactie met T-cellen. Momenteel is het nog onduidelijk waarom een transcriptiefactor betrokken bij ER-stress zo'n grote invloed heeft op DC. Daarom worden microarray experimenten uitgevoerd waarbij *Xbp1* knock-out (KO)cDC vergeleken worden met *wild type* (WT) cDC.

Uit de analyse van de ruwe microarray data via Bioconductor blijkt dat 55 genen beïnvloed worden door de *Xbp1* KO in $CD8\alpha^+$ DC, terwijl in de $CD11b^+$ DC slechts 2 genen beïnvloed worden. De functies van deze genen worden bepaald via de *Gene Ontology* van de genen en door te achterhalen van welke netwerken de genen deel uitmaken (via *Ingenuity Pathway Analysis*[®]). Aangezien *Xbp1* een transcriptiefactor is, wordt ook nog geanalyseerd in hoeverre *Xbp1* de genen met een differentiële genexpressie direct of indirect reguleert. Daarvoor wordt gebruik gemaakt van de tool FIMO (*Find Individual Motif Occurrences*, onderdeel van MEME Suite) en Contra V2.

De resultaten van de microarray analyse tonen aan dat de *Xbp1* KO een grotere invloed heeft op $CD8\alpha^+$ DC dan op de $CD11b^+$ DC. In een poging dat te verklaren worden de WT $CD8\alpha^+$ DC microarrays vergeleken ten opzichte van de WT $CD11b^+$ cDC microarrays. Op die manier wordt een lijst bekomen van potentiële merkers voor $CD8\alpha^+$ DC en $CD11b^+$ DC. Wanneer deze lijst vergeleken wordt met de 55 genen die beïnvloed worden door de *Xbp1* KO in $CD8\alpha^+$ DC en de lijst met de genen die een *Xbp1*-motief bezitten, kunnen vier interessante genen geselecteerd worden. Deze vier genen zijn opmerkelijk aangezien ze verschillend tot expressie komen in $CD8\alpha^+$ DC en $CD11b^+$ DC en omdat ze mogelijks direct gereguleerd worden door *Xbp1*. De genen bezitten immers een evolutionair geconserveerd *Xbp1*-motief (volgens FIMO en Contra V2) en hun expressie wordt beïnvloedt door de KO van *Xbp1*. Deze vier genen kunnen dus niet alleen gebruikt worden om het verschil tussen $CD8\alpha^+$ DC en $CD11b^+$ DC verder uit te diepen, ze kunnen eveneens

Analyse van Xbp1 KO-microarray data in cDC

nieuwe, tot nu toe ongekende, targets van Xbp1 zijn. Verder experimenteel onderzoek is noodzakelijk om hierrond meer zekerheid te bekomen.

Lijst met afkortingen en symbolen

Afkorting	Betekenis
APC	Antigeenpresenterende Cel
Atf6	<i>Activating Transcription Factor 6</i>
BiP	<i>Binding Immunoglobulin Protein</i>
Ccl22	<i>Chemokine (C-C motif) ligand 22</i>
Ccr2	<i>Chemokine (C-C motif) receptor 2</i>
CD	<i>Cluster of Differentiation</i>
CD115	<i>Colony stimulating factor 1 receptor (Csfr1)</i>
CD11b	<i>Integrin alpha M (Itgam)</i>
CD172a	<i>Signal-regulatory protein alpha (Sirpa)</i>
CD205	<i>Lymphocyte antigeen 75 (Ly75)</i>
CD36	Cd36 antigeen
cDC	<i>Conventional Dendritic Cell</i>
cDNA	<i>Copy desoxyribonucleïnezuur</i>
CHOP	<i>CCAAT/enhancer-binding protein homologous protein</i>
Ckrox	<i>Zinc finger and BTB domain containing 7B (Zbtb7b)</i>
Clec4a4	<i>C-type lectin domain family 4, member 4a</i>
Clec9a	<i>C-type lectin domain family 9, member a</i>
Csf-1R	<i>Colony stimulating factor 1 receptor</i>
Cx3CR1	<i>Chemokine (C-X3-C) receptor 1</i>
Cy3 / Cy5	Cyanine 3 / Cyanine 5
DC	Dendritische cel
dChip	DNA chip
DNA	<i>Desoxyribonucleïnezuur</i>
Dnajc3	<i>Colony Stimulating Factor 1 Receptor</i>
DR5	<i>Death Receptor 5</i>
Edem1	<i>ER Degradation Enhancer, Mannosidase alpha-like 1</i>
Eif2ak3	<i>Eukaryotic translation initiation factor 2 alpha kinase 3</i>
eIF2 α	<i>Eukaryotic translation initiation factor 2 alpha</i>
ER	Endoplasmatisch Reticulum
ERAD	<i>ER-associated Degradation pathway</i>

Ero1α	<i>Endoplasmic Oxidoreductin 1 like</i>
Erp44	<i>Endoplasmic reticulum protein 44</i>
E-value	<i>Expected value</i>
F4/80	<i>EGF-like module containing, mucin-like, hormone receptor-like sequence 1 (Emr1)</i>
Fcer1a	<i>Fc receptor, IgE, high affinity I, alpha polypeptide</i>
Fcgr1	<i>Fc receptor, IgG, high affinity</i>
FDR	<i>False Discovery Rate</i>
FIMO	<i>Find Individual Motif Occurrences</i>
Flt3	<i>FMS-like tyrosine kinase 3</i>
Flt3l	<i>FMS-like tyrosine kinase 3 ligand</i>
GCRMA	<i>Guanine Cytosine Robust Multi-array Average</i>
GNU	<i>GNU's Not Unix</i>
GO	<i>Gene Ontology</i>
Herpud1	<i>Homocysteine-inducible, Endoplasmic Reticulum stress-inducible, Ubiquitin-like Domain member 1</i>
HSV-1	<i>Herpes Simplex Virus 1</i>
IFN-α	<i>Interferon alfa</i>
Igf2r	<i>Insulin-like Growth Factor 2 Receptor</i>
IL	<i>Interleukine</i>
iNOS	<i>Inducible Nitric Oxide Synthase</i>
IPA®	<i>Ingenuity Pathway Analysis®</i>
Ire1	<i>Inositol-Requiring kinase 1</i>
KEGG	<i>Kyoto Encyclopedia of Genes and Genomes</i>
Klf4	<i>Kruppel-like factor 4</i>
KO	<i>Knock-out</i>
Lamp2	<i>lysosomal-associated membrane protein 2</i>
Log	<i>Logaritme</i>
logFC	<i>Logaritme van fold change</i>
Ly6C	<i>Lymphocyte antigen 6 complex</i>
MAS 5.0	<i>Microarray Suite 5.0</i>
MAST	<i>Motif Alignment and Search Tool</i>
MEME	<i>Multiple EM for Motif Elicitation</i>
MHC	<i>Major histocompatibiliteitscomplex</i>

miRNA	Micro RNA
Mlec	<i>Malectin</i>
moDC	<i>Monocyte-derived Dendritic Cell</i>
mRNA	<i>Messenger RNA</i>
NCBI	<i>National Center for Biotechnology Information</i>
NUSE	<i>Normalized Unscaled Standard Error</i>
PBM	<i>Protein Binding Microarrays</i>
pDC	<i>Plasmacytoid Dendritic Cell</i>
PLIER	<i>Probe Logarithmic Intensity Error</i>
Ppp1r15a	<i>Protein Phosphatase 1 Regulatory subunit 15A</i>
PRRs	<i>Pattern Recognition Receptors</i>
PWMs	<i>Position Weight Matrices</i>
RLE	<i>Relative Log Expression</i>
RMA	<i>Robust Multi-array Analysis</i>
RNA	Ribonucleïnezuur
Sp1	<i>Trans-acting transcription factor 1</i>
Syvn1	<i>Synovial apoptosis inhibitor 1</i>
Tapbp	<i>TAP binding protein</i>
T _H 1	T-helpercel 1
T _H 17	T-helpercel 17
T _H 2	T-helpercel 2
Tip-DC	<i>TNF iNOS Producing Dendritic cells</i>
TLR	<i>Toll-like Receptor</i>
TNF	<i>Tumor Necrosis Factor</i>
TOMTOM	<i>Motif Comparison Tool</i>
Treg	Regulatorische T-cel
UPR	<i>Unfolded Protein Response</i>
UTR	<i>Untranslated Region</i>
Vs	Versus
WT	<i>Wild type</i>
Xbp1	<i>X-box Binding Protein 1</i>
Xcl1	<i>Chemokine (C motif) ligand 1</i>
Xcr1	<i>Chemokine (C motif) receptor 1</i>

α_j	gemiddelde expressie voor gen j
ε	<i>residuals</i>
X	<i>design</i> matrix
Y_j	gemeten expressie voor gen j

Verklarende woordenlijst

Begrip	Verklaring
25 ^{ste} percentiel	De getalswaarde die de laagste 25% van de getalswaarden onderscheidt van de hogere waarden. Synoniem van eerste kwartiel.
Algoritme	Een opeenvolging van instructies die een bepaald probleem oplossen.
Antigeen	Een molecule die in staat is een reactie van het immuunsysteem op te wekken.
Antigeenaffiniteit	De neiging om te binden aan een antigeen.
Antigeen presenterende cel	Een cel die een deel van een antigeen presenteert aan specifieke cellen van het immuunsysteem.
Apoptose	Het proces van een geprogrammeerde celdood.
Artefact	Een afwijking.
Autofosforylatie	Een kinase dat zichzelf voorziet van een fosfaatgroep.
Auto-immuunziekte	Een ziekte waarbij het immuunsysteem lichaamseigen cellen en stoffen als lichaamsvreemd aanziet.
<i>Bead</i> DNA-microarrays	Een DNA-microarray waarbij de probes gebonden zijn aan polystyrene bolletjes.
Besturingssysteem	Een programma dat, na het opstarten van een computer, in het geheugen geladen wordt en de hardware aanstuurt.
Chaperone	Een proteïne dat helpt bij het beter opvouwen van eiwitten.
Chemokine	Een synoniem voor signaalproteïnen.
<i>ChIP-sequencing</i>	Een experiment om interacties tussen DNA en proteïnen te onderzoeken.
Complementair	Synoniem voor aanvullend. Bij DNA is guanine complementair met cytosine en is thymine complementair met adenine.
Consensussequentie	Een sequentie die per positie het meest voorkomende nucleotide of aminozuur bevat van een ophijning van een aantal sequenties.
Cross-presentatie	Het presenteren van een lichaamsvreemd partikel op een MHC I receptor.

<i>Cut off</i>	Synoniem voor drempelwaarde.
Cytosol	Intracellulaire vloeistof waarin alle organellen zich bevinden.
<i>Design matrix</i>	Een matrix opgebouwd uit enen en nullen. Een <i>design matrix</i> wordt gebruikt tijdens een microarray analyse om aan te duiden tot welke conditie een bepaalde microarray behoort (bijvoorbeeld <i>knock-out</i> versus <i>wild type</i>).
Differentiatie	Synoniem voor ontwikkelen.
Dimerisatie	Twee moleculen die samen één grotere molecuule vormen.
DNA-microarray	Een plaatje waaraan de sequenties van duizenden genen (onder de vorm van probes) zijn vastgehecht.
Downregulatie	Een vermindering van de genexpressie.
Endocytose	De opname van stoffen in een cel via een insluiting van het celmembraan.
Epidermaal	Behorend tot de opperhuid.
Endoplasmatisch reticulum	Het organel waar de eiwitopvouwingen plaatsvinden.
E-value	Een waarde die aangeeft hoe groot de kans is dat het bekomen resultaat een toevalstreffer is. Hoe lager de e-value, hoe kleiner de kans dat het bekomen resultaat per toeval werd bekomen.
Exon	Een coderend deel van een gen.
Extracellulair	Het milieu buiten de cel.
Fagosoom	Vaste deeltjes kunnen door een cel opgenomen worden doordat het celmembraan zich omsluit rond die deeltjes. Op die manier ontstaat een holte, ook het fagosoom genoemd. Door secretie van proteasen naar het fagosoom kunnen de opgenomen deeltjes vernietigd worden.
<i>False Discovery Rate</i>	Proportie van de significante resultaten die eigenlijk vals-positief zijn.
Fenotype	Het totaal aan waarneembare eigenschappen van een organisme.
<i>Fold change</i>	Een getal dat aangeeft hoeveel de oorspronkelijke waarde verschilt van de uiteindelijke waarde. Stel dat de oorspronkelijke waarde gelijk is aan 2 en de finale waarde is

	gelijk aan 8, dan is de <i>fold change</i> gelijk aan 4 (8/2).
Genexpressie	Het proces waarbij DNA wordt omgezet naar mRNA dat vervolgens op zijn beurt omgezet wordt naar een proteïne.
Gespotte DNA-microarray	Een DNA-microarray waarbij een DNA-oplossing via <i>contact spotting</i> of via <i>piezo capillary jet spotting</i> op de drager wordt aangebracht.
Glycosylatie	Een enzymatisch proces waarbij suikergroepen gekoppeld worden aan een eiwit.
Golgi-apparaat	Het organel waar de eiwitten na het endoplasmatisch reticulum aankomen. In het Golgi-apparaat ondergaan de eiwitten de laatste bewerkingen om ten slotte naar de juiste bestemming getransporteerd te worden.
<i>Heatmap</i>	Een grafische representatie van data waarbij de waarden voorgesteld worden als kleuren.
Hematopoëtische cel	Synoniem voor een multipotente bloedcel.
<i>High-throughput</i>	De verwerking van grote hoeveelheden data op een korte tijd.
Homeostatisch	Synoniem voor in evenwicht zijn.
Hybridiseren	Het samenvoegen van twee complementaire DNA-strengen tot een dubbele helix.
Imflammatie	Synoniem voor ontstekingsreactie.
Immunomodulatie	Verandering in het immuunsysteem door de activatie of onderdrukking ervan.
In situ gehybridiseerde oligonucleotide-microarray	Een DNA-microarray waarbij de probes via fotolithografie op de drager worden aangebracht.
Induceren	Synoniem voor aansporen.
Inhiberen	Synoniem voor remmen.
Intracellulair	Het milieu binnenin een cel.
<i>Knock-out</i> (gen)	Een gen dat niet meer tot expressie komt.
Lymfeknoop	Een klein orgaan dat langs de lymfevaten voorkomt. Een synoniem is lymfeklier.
Macrofagen	Een mononucleaire cel die in staat is deeltjes zoals dode cellen op te nemen via fagocytose en te vernietigen.

Mediaan	De middelste waarde van een aantal waarden.
Methylatie	Een organische reactie waarbij een methylgroep toegevoegd wordt aan een molecule.
Major histocompatibiliteitscomplex	Antigenen die aanwezig zijn op alle lichaamscellen, behalve op de rode bloedlichaampjes. De menselijke variant wordt Humaan Leukocytenantigeen (HLA) genoemd.
Microtiterplaat	Een recipiënt met vele ruimten waarin reagentia kunnen gemengd en bewaard worden.
MicroRNA	Een korte niet-coderende RNA-sequentie die de expressie van genen reguleert. MicroRNA's binden vooral aan het 3'UTR van genen.
<i>Mismatch</i>	Een binding tussen twee DNA-moleculen of RNA-moleculen die niet complementair zijn aan elkaar.
Monocyten	Synoniem voor witte bloedcel.
Multimerizeren	Meerdere moleculen die samen één grotere molecule vormen.
Multiple testing	Er wordt over multiple testing gesproken wanneer er meerdere statistische toetsen uitgevoerd worden op dezelfde dataset.
Naïeve T-cellen	Synoniem voor onvolwassenen T-cellen, T-cellen die nog niet in contact zijn gekomen met een antigeen.
Necrose	Het eindstadium van celdood.
<i>Noise</i>	Synoniem voor ruis.
Normaalverdeling	Een continue kansverdeling met twee parameters, de verwachtingswaarde en de standaardafwijking. Synoniem voor Gaussverdeling.
Normalisatie	Het proces dat erop gericht is een zekere regelmaat of norm te verkrijgen in bepaalde data.
Nucleus	Synoniem voor celkern.
Oligonucleotide	Korte, enkelstrengige DNA-moleculen of RNA-moleculen.
Ontogenie	De wetenschap van de ontwikkelingsfysiologie.
Oxidatieve stress	Een stofwisselingstoestand waarbij meer reactieve zuurstofverbindingen aanwezig zijn of gevormd worden in

	een cel.
<i>Package</i>	Een apart software onderdeel dat de mogelijkheid biedt een specifieke taak te vervolledigen. Synoniem voor module.
Pathogeen	Synoniem voor ziekteverwekkend.
<i>Pathway</i>	Een reeks stofwisselingsprocessen in het inwendige van een cel onder invloed van enzymen.
Posttranslationele modificaties	Een enzymgekatalyseerde verandering van een eiwit na de translatie.
<i>Pattern Recognition Receptor</i>	Receptoren die patronen herkennen die geassocieerd worden met dode cellen of pathogeen.
Probes	Een DNA-molecule of RNA-molecule met lengte die varieert tussen 100 basen en 1000 basen. Probes worden gebruikt voor microarray experimenten.
Probeset	Een geheel van een aantal probes.
Progenitor	Synoniem voor voorloper.
Promotor	Het DNA-element voor een gen dat de expressie van het gen reguleert.
P-waarde	Geeft de betrouwbaarheid van het bekomen resultaat aan. Algemeen geldt: hoe kleiner de p-waarde, hoe kleiner de kans dat het gaat om een toevalstreffer.
<i>Position Weight Matrix</i>	Een veel gebruikte representatie van motieven in biologische sequenties.
Residual	Synoniem voor een foutenmarge.
Significant	Statistische term die uitdrukt dat het verschil tussen twee waarden waarschijnlijk niet berust op toeval.
Significantieniveau	De kans dat de nulhypothese ten onrechte wordt verworpen.
Spot	Een punt, een plek.
Standaarddeviatie	Een maat voor de spreiding van een verdeling.
Standaardfout	De standaardafwijking van het steekproefgemiddelde.
<i>Steady state</i>	De toestand waarbij er geen sprake is van inflammatie. <i>Steady state</i> is het tegengestelde van <i>inflammatory state</i> .
T-effectorcel	Een geactiveerde T-cel.

<i>Threshold</i>	Een signaalniveau waaronder of waarboven een bepaalde bewerking wordt gestart. Synoniemen zijn <i>cut-off</i> en drempelwaarde.
<i>Toll-like receptor</i>	Membraanreceptoren die functioneren als patroonherkenningsmoleculen voor micro-organismen.
Toxine	Synoniem voor een giftige stof.
Toxoplasmose	Een ziekte veroorzaakt door de eencellige parasiet <i>Toxoplasma gondii</i> .
Transcript	Het resultaat van transcriptie. Een bepaalde mRNA-sequentie.
Transcriptiefactor	Een proteïne dat bindt op specifieke DNA-sequenties en zo de transcriptie van (een) gen(en) stimuleert of inhibeert.
Transmembraan	Synoniem voor door het celmembraan heen.
Transcriptiestartplaats	De locatie waar de transcriptie start aan het 5' uiteinde van een gen.
T-test	Test die gebruikt wordt om na te gaan of twee gemiddelden significant verschillend zijn van elkaar.
Upregulatie	Een verhoging van de genexpressie.
Vals-negatief	Een resultaat dat volgens de test negatief is, terwijl het in werkelijkheid positief is.
Vals-positief	Een resultaat dat volgens de test positief is, terwijl het in werkelijkheid negatief is.
Variantie	Een maat voor de spreiding van waarden.
Vrijheidsgraad	Een onafhankelijke parameter waarmee een aspect van een fysisch systeem wordt vastgelegd. Het aantal vrijheidsgraden wordt meestal berekend door één af te trekken van het aantal metingen.
<i>Wild type</i>	Een fenotype dat voorkomt in de natuur.

Lijst van tabellen

Tabel 2-1: Overzicht van de verschillende subsets van DC.	27
Tabel 4-1: De tien meest <i>upregulated</i> (\nearrow) en <i>downregulated</i> (\searrow) genen in de CD8 α^+ DC dataset.	57
Tabel 4-2: Lijst van de meest voorkomende GO-termen in CD8 α^+ DC dataset (moleculaire functie)..	59
Tabel 4-3: Lijst van de meest voorkomende GO-termen in CD8 α^+ DC dataset (biologisch proces).	59
Tabel 4-4: Lijst van de meest voorkomende GO-termen in CD8 α^+ DC dataset (cellulair compartiment).	60
Tabel 4-5: Lijst van de genen die een differentiële genexpressie vertonen in de CD11b $^+$ DC dataset.	60
Tabel 4-6: Lijst van de meest voorkomende GO-termen in CD11b $^+$ DC dataset (moleculaire functie).	61
Tabel 4-7: Lijst van de meest voorkomende GO-termen CD11b $^+$ DC dataset (biologisch proces).	61
Tabel 4-8: Lijst van de meest voorkomende GO-termen CD11b $^+$ DC dataset (cellulair compartiment).	62
Tabel 4-9: Overzicht van de netwerken waarin de genen met differentiële genexpressie van de CD8 α^+ DC dataset betrokken zijn..	62
Tabel 4-10: Overzicht van de netwerken waarin de genen met differentiële genexpressie van de CD11b $^+$ DC dataset betrokken zijn..	63
Tabel 4-11: Lijst van de genen waarin het Xbp1-motief sterker geconserveerd aanwezig is in vergelijking met de andere genen die de FIMO-analyse als resultaat geeft.....	69
Tabel 4-12: Lijst van de genen die enig verschil vertonen in expressie tussen dataset 1 (CD8 α^+ DC WT vs CD8 α^+ DC KO) en dataset 2 (CD8 α^+ DC WT vs CD11b $^+$ DC WT).	71
Tabel 4-13: Lijst van genen die een Xbp1-motief in hun promotors bezitten en die een interessant verschil vertonen in expressie tussen dataset 1 (CD8 α^+ DC WT vs CD8 α^+ DC KO) en dataset 2 (CD8 α^+ DC WT vs CD11b $^+$ DC WT).	72

Lijst van figuren

Figuur 2-1: DC vormen de link tussen het aangeboren (<i>innate</i>) en verworven (<i>adaptive</i>) immuunsysteem.....	25
Figuur 2-2: De werking van een cytotoxische T-cel op de plaats van infectie.....	26
Figuur 2-3: Overzicht van de verschillende T-helpercellen.....	26
Figuur 2-4: Overzicht van de drie <i>pathways</i> van de UPR.....	33
Figuur 2-5: DNA-motief waaraan Xbp1 bindt.....	33
Figuur 2-6: Voorbeeld van een DNA-microarray.....	34
Figuur 2-7: De procedure van een microarray experiment.	35
Figuur 2-8: Voorbeeld van een microarray.	36
Figuur 3-1: Voorbeeld van een boxplot.	40
Figuur 3-2: De gemeten data voor een bepaalde probeset over vijf microarrays.....	41
Figuur 3-3: Een model fitten op gemeten data.....	42
Figuur 3-4: Voorbeeld van een RLE-plot.	43
Figuur 3-5: Voorbeeld van een NUSE-plot.	44
Figuur 3-6: De geobserveerde expressie is eigenlijk een som van het eigenlijke signaal en de achtergrond.....	45
Figuur 3-7: Resultaat van een \log_2 -transformatie.....	45
Figuur 3-8: Voorbeeld van lineair model fitten op data.....	47
Figuur 3-9: Screenshot IPA®.....	50
Figuur 4-1: De histogrammen van de ruwe microarrays van de CD8 α^+ DC dataset.	52
Figuur 4-2: De histogrammen van de ruwe microarrays van de CD11b $^+$ DC dataset.	53
Figuur 4-3: De boxplots van de ruwe microarray data van CD8 α^+ DC dataset.	54
Figuur 4-4: De boxplots van de ruwe microarray data van CD11b $^+$ DC dataset.	54
Figuur 4-5: De boxplots van de genormaliseerde microarray data van CD8 α^+ DC dataset.	56
Figuur 4-6: De boxplots van de genormaliseerde microarray data van CD11b $^+$ DC dataset..	56
Figuur 4-7: De heatmap van de genen met differentiële genexpressie van de CD8 α^+ DC dataset.	58
Figuur 4-8: De heatmap van de genen met differentiële genexpressie van de CD11b $^+$ DC dataset. ...	61
Figuur 4-9: Netwerk 1 van de CD8 α^+ DC dataset..	64
Figuur 4-10: Resultaat motiefanalyse via MEME Suite (TOMTOM) voor de downgereguleerde genen van de CD8 α^+ DC dataset (800 bp upstream).....	65
Figuur 4-11: Resultaat motiefanalyse via MEME Suite (TOMTOM) voor de upgereguleerde genen van de CD8 α^+ DC dataset (800 bp upstream).	65

Analyse van *Xbp1* KO-microarray data in cDC

Figuur 4-12: Het motief dat aanwezig is in de promotors van de downgereguleerde genen van de CD8 α^+ DC dataset.....	66
Figuur 4-13: Het motief dat aanwezig is in de promotors van de upgereguleerde genen van de CD8 α^+ DC dataset.....	66
Figuur 4-14: Resultaat analyse Xbp1-motief in promotor <i>Sp1</i>	67
Figuur 4-15: Resultaat analyse Xbp1-motief in promotor <i>Klf4</i>	67
Figuur 4-16: Resultaat analyse Xbp1-motief in promotor <i>Ckrox</i>	68
Figuur 4-17: Resultaat analyse Xbp1-motief in promotor <i>Entpd7</i>	70
Figuur 8-1: <i>Chip pseudo-image</i> gebaseerd op <i>residuals</i> van microarray 1 van de CD8 α^+ DC dataset .	94
Figuur 8-2: <i>Chip pseudo-image</i> gebaseerd op <i>weights</i> van microarray 1 van de CD8 α^+ DC dataset....	95
Figuur 8-3: RLE-plot van de ruwe microarray data van de CD8 α^+ DC dataset	95
Figuur 8-4: NUSE-plot van de CD8 α^+ DC dataset.....	96
Figuur 8-5: <i>Chip pseudo-image</i> gebaseerd op <i>residuals</i> van microarray 1 van de CD11b $^+$ DC dataset	97
Figuur 8-6: <i>Chip pseudo-image</i> gebaseerd op <i>weights</i> van microarray 1 van de CD11b $^+$ DC dataset..	98
Figuur 8-7: RLE-plot van de ruwe microarray data van de CD11b $^+$ DC dataset.	98
Figuur 8-8: NUSE-plot van de ruwe microarray data van de CD11b $^+$ DC dataset	99

Inhoudsopgave

Voor akkoord verklaring.....	2
Woord vooraf	3
Samenvatting	4
Lijst met afkortingen en symbolen.....	6
Verklarende woordenlijst.....	10
Lijst van tabellen	16
Lijst van figuren	17
1 Inleiding en probleemstelling.....	22
1.1 Dendritische cellen subsets.....	23
1.2 Transcriptiefactor Xbp1.....	23
2 Situering	25
2.1 Het immuunsysteem	25
2.2 Subsets dendritische cellen.....	25
2.2.1 <i>Conventional</i> DC.....	27
2.2.2 Andere subsets DC	30
2.3 Xbp1 en endoplasmatisch reticulum stress	31
2.3.1 IRE1.....	31
2.3.2 Eif2ak3.....	32
2.3.3 ATF6.....	32
2.4 Genregulatie.....	33
2.4.1 Transcriptiefactor	33
2.4.2 Netwerkindertacties.....	33
2.5 DNA-microarrays.....	34
2.5.1 Werking DNA-microarray	34
3 Materiaal en methoden	37
3.1 Personal Computer.....	37
3.2 R – Bioconductor	37
3.3 Contra V2.....	37
3.4 MEME Suite	38
3.4.1 TOMTOM.....	38
3.4.2 FIMO	38
3.5 Ingenuity Pathway Analysis® (IPA®)	38

Analyse van *Xbp1* KO-microarray data in cDC

3.6	Analyse microarray data via Bioconductor	39
3.7	Controle kwaliteit ruwe data via plots	39
3.7.1	Histogram	39
3.7.2	Boxplot	39
3.7.3	Chip pseudo-image.....	40
3.7.4	RLE (<i>Relative Log Expression</i>)	43
3.7.5	NUSE (<i>Normalized Unscaled Standard Errors</i>)	43
3.8	Normalisatie	44
3.8.1	Achtergrondcorrectie	44
3.8.2	Log ₂ -transformatie	45
3.8.3	Quantile normalisatie.....	45
3.8.4	Probenormalisatie	46
3.9	Identificatie van genen met differentiële genexpressie	46
3.9.1	Lineair model fitten op data.....	46
3.9.2	Standaard T-test.....	47
3.9.3	Moderated T-test	48
3.9.4	P-waarde aanpassen volgens Benjamini-Hochberg.....	48
3.9.5	Samenstellen lijst genen met differentiële genexpressie	49
3.10	Annotatie van genen met differentiële genexpressie.....	49
3.11	Pathway analyse.....	49
4	Resultaten	51
4.1	Analyse microarray data via Bioconductor	52
4.1.1	Controle kwaliteit ruwe data via plots	52
4.1.2	Visualisatie normalisatie	55
4.1.3	Lijst van genen met differentiële genexpressie	57
4.2	IPA [®] -analyse.....	62
4.3	Motiefanalyse.....	64
4.3.1	Directe of indirecte regulering van <i>Xbp1</i>	68
4.3.2	Verschil tussen CD8α ⁺ DC en CD11b ⁺ DC in wild type condities	70
5	Discussie.....	73
5.1	Functies genen met differentiële genexpressie	73
5.2	Motiefanalyse.....	74
5.2.1	Regulatie van Sp1, Klf4 en Ckrox door <i>Xbp1</i>	74
5.2.2	Directe of indirecte regulatie door <i>Xbp1</i>	75

Analyse van Xbp1 KO-microarray data in cDC

5.2.3	CD11b ⁺ DC versus CD8α ⁺ DC.....	76
5.2.4	Toekomstperspectieven.....	76
6	Besluit.....	77
7	Literatuurlijst.....	78
8	Bijlagen.....	81
8.1	Bijlage 1.....	81
8.2	Bijlage 2.....	83
8.3	Bijlage 3.....	92
8.4	Bijlage 4.....	94
8.4.1	CD8α ⁺ DC dataset.....	94
8.4.2	CD11b ⁺ DC dataset.....	96
8.5	Bijlage 5.....	99
8.6	Bijlage 6.....	101
8.6.1	CD8α ⁺ DC dataset.....	101
8.6.2	CD11b ⁺ DC dataset.....	104
8.7	Bijlage 7.....	106

1 Inleiding en probleemstelling

Het immuunsysteem wordt gedefinieerd als het verdedigingssysteem met als doel indringers of veranderde lichaamseigen cellen te vernietigen [1]. De belangrijkste functie van het immuunsysteem is dus te reageren op alle potentieel pathogene, vreemde partikels, terwijl er tegen lichaamseigen cellen tolerantie moet vertoond worden. Desondanks dit eenvoudig principe loopt het hierbij vaak mis. Wanneer het immuunsysteem toch reageert op lichaamseigen cellen, betekent dit dat de persoon in kwestie begint te vechten tegen zichzelf, met alle gevolgen van dien (auto-immuunziekte). Daarnaast is het ook mogelijk dat het immuunsysteem een immunologische reactie in gang zet tegen een vreemd partikel dat helemaal niet pathogeen is, zoals huisstofmijt of pollen (allergie) [2]. Vertrekkende vanuit dat concept is de long een zeer interessant weefsel. Aan de ene kant komt de long namelijk in contact met vele onschadelijke deeltjes afkomstig van de ingeademde lucht, terwijl aan de andere kant de long ook de ingang bij uitstek is voor heel wat pathogenen, zoals virussen en bacteriën.

Het immuunsysteem wordt onderverdeeld in een aangeboren, niet-specifiek immuunsysteem en een verworven, specifiek immuunsysteem. De aangeboren immuniteit wordt gekenmerkt door een aantal cellen die lokaal in staat zijn om het even welk pathogeen partikel te vernietigen. Deze natuurlijke, niet-specifieke immuniteit zorgt voor tijdswinst die nodig is om het verworven immuunsysteem op gang te krijgen. Het verworven immuunsysteem wordt op zijn beurt gekenmerkt door cellen die gericht zijn tegen één bepaald antigeen [2].

Een centrale rol in het immuunsysteem is weggelegd voor de dendritische cellen (DC). DC zijn professionele antigeenpresenterende cellen (APC). Dit betekent dat DC in staat zijn een antigeen op te nemen, te vernietigen en vervolgens een stuk van dit antigeen te presenteren op hun oppervlak. DC zullen dit stuk antigeen presenteren aan naïeve T-cellen in de dichtstbijzijnde lymfeknoop. Daarbij zal de dendritische cel beslissen of een naïeve T-cel zich moet ontwikkelen tot een T-cel die leidt tot tolerantie of een T-cel die leidt tot inflammatie. De dendritische cel is dus dé cel die beslist over tolerantie of inflammatie en is daarom een zeer interessante cel in het belang van klinische toepassingen, zoals vaccintechnologie en immunomodulatie [3][4].

1.1 Dendritische cellen subsets

Onderzoek heeft aangetoond dat de DC kunnen onderverdeeld worden in vier subgroepen [4]:

- *Conventional* DC (cDC);
 - CD11b⁺ DC
 - CD8α⁺ DC
- *Plasmacytoïd* DC (pDC);
- *Monocyt-derived* DC (moDC);
- Langerhanscellen (LC).

Deze vier subtypes vertonen allen verschillen op vlak van omgang met een antigeen en interactie met T-cellen, maar tegelijkertijd is er nog maar weinig gekend over de genen die aan de basis liggen van deze functionele verschillen. Momenteel is men echter wel in staat op basis van de expressie van een aantal receptoren een onderscheid te maken tussen de verschillende subsets, maar verder onderzoek is vereist [4].

1.2 Transcriptiefactor Xbp1

Xbp1 is een transcriptiefactor die upgereguleerd wordt tijdens stress in het endoplasmatisch reticulum (ER). Recente experimenten met *Xbp1 knock-out* (KO) muizen tonen echter aan dat het *Xbp1* gen heel belangrijk is voor de ontwikkeling en overleving van DC [5]. De afwezigheid van *Xbp1* leidt namelijk tot de vorming van abnormale DC, waardoor deze hun functies niet langer optimaal kunnen uitvoeren. Het is echter onduidelijk waarom een transcriptiefactor betrokken bij ER-stress zo'n grote invloed heeft op DC, omdat weinig gekend is over de genen die gereguleerd worden door Xbp1.

Om een antwoord te kunnen formuleren op die vraag werden microarray experimenten uitgevoerd waarbij *Xbp1* KO in CD8α⁺ DC vergeleken werden met *wild type* (WT) CD8α⁺ DC. Aangezien CD11b⁺ DC eveneens cDC zijn, werd een zelfde microarray experiment ook uitgevoerd voor de CD11b⁺ DC subset.

De analyse van de ruwe microarray data wordt uitgevoerd in Bioconductor, een gratis open-source softwarepakket beschikbaar voor de statistische programmeertaal R. Vervolgens wordt een *Ingenuity Pathway Analysis*[®] (IPA[®]) toegepast om te achterhalen in welke netwerken de genen met een differentiële genexpressie, betrokken zijn. Aangezien Xbp1 een transcriptiefactor is, wordt ook nog geanalyseerd in hoeverre Xbp1 de genen met een differentiële genexpressie

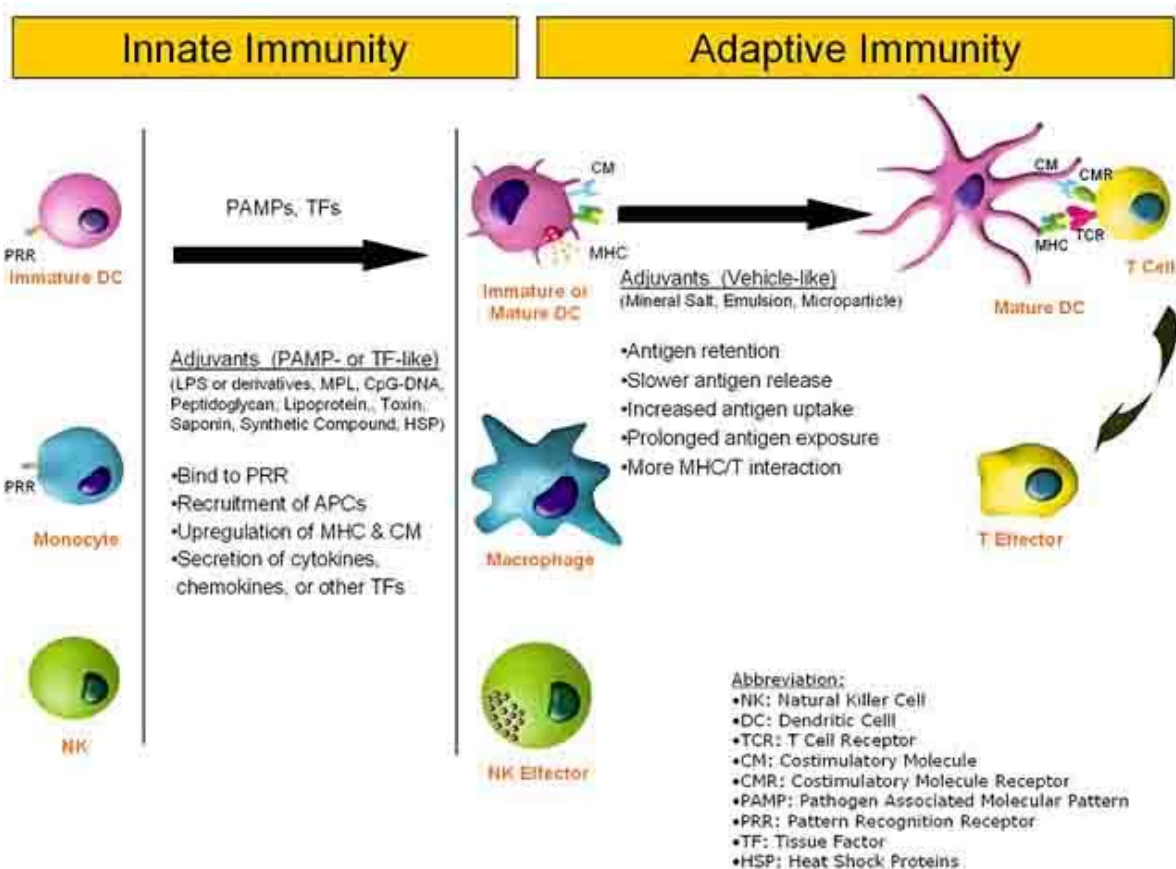
Analyse van Xbp1 KO-microarray data in cDC

direct of indirect reguleert. Daarvoor wordt gebruik gemaakt van verschillende tools, zoals MEME Suite en Contra V2. Op die manier wordt geprobeerd te achterhalen hoe zowel de CD8 α^+ DC als de CD11b $^+$ DC beïnvloed worden door de Xbp1 KO.

2 Situering

2.1 Het immuunsysteem

Het aangeboren immuunsysteem treedt op tegen om het even welk vreemd partikel, terwijl een verworven immuunrespons steeds gericht is tegen één bepaald antigeen. DC vormen de link tussen deze twee immuniteiten: DC zijn cellen van de aangeboren immuniteit die allerlei pathogenen herkennen, maar tegelijkertijd zijn DC ook een onderdeel van de verworven immuniteit, aangezien het de DC zijn die de T-cel activatie en differentiatie controleren (figuur 2-1).



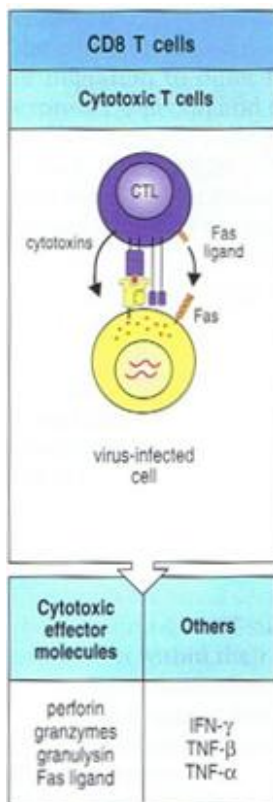
Figuur 2-1: DC vormen de link tussen het aangeboren (*innate*) en verworven (*adaptive*) immuunsysteem [6].

2.2 Subsets dendritische cellen

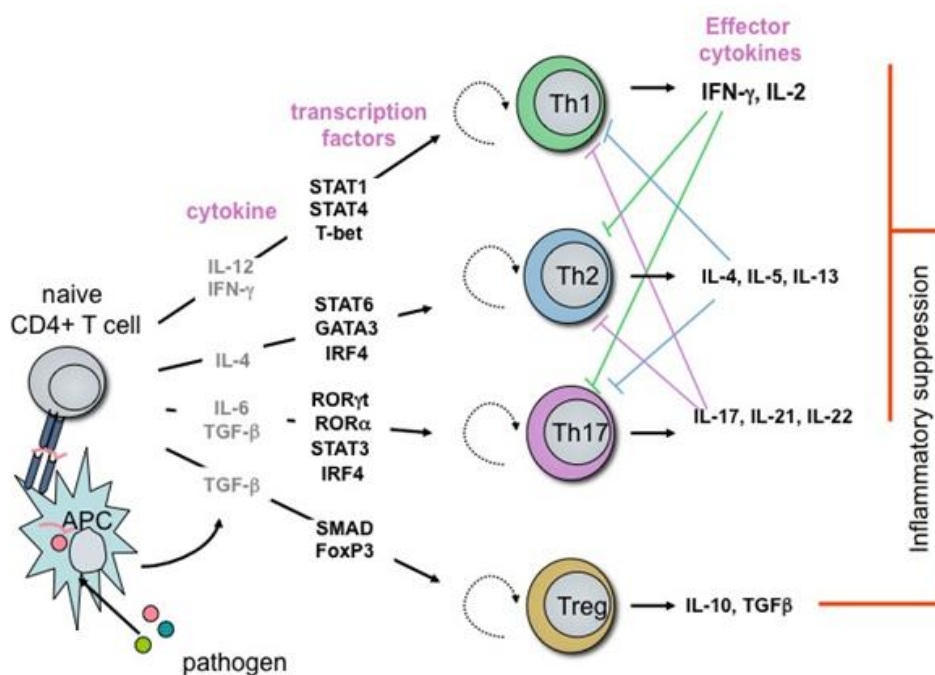
De werking van DC kan samengevat worden als het herkennen van om het even welk antigeen, het opnemen en verwerken ervan, migreren naar de dichtstbijzijnde lymfeknoop en het presenteren van het antigeen aan naïeve CD8 of CD4 T-cellen, zodat ten slotte ofwel cytotoxische

Analyse van *Xbp1* KO-microarray data in *cDC*

T-cellen (figuur 2-2), T_H1 cellen, T_H2 cellen, T_H17 cellen of T-suppressorcellen (Treg) gevormd worden (figuur 2-3).



Figuur 2-2: De werking van een cytotoxische T-cel op de plaats van infectie [3].



Figuur 2-3: Overzicht van de verschillende T-helpercellen [7].

Analyse van *Xbp1* KO-microarray data in cDC

T-suppressorcellen zijn in staat een immuunrespons stil te leggen, waardoor deze cellen een centrale rol spelen in tolerantie. De functie van T-cytotoxische cellen bestaat erin macrofagen en andere cellen van het immuunsysteem zodanig te activeren dat deze starten met de productie van toxines die het binnengedrongen pathogeen snel doden. T_H1 cellen komen vooral in actie voor intracellulaire pathogenen (bijvoorbeeld bacteriën), terwijl een T_H2 respons essentieel is om extracellulaire parasieten (bijvoorbeeld wormen) te overwinnen. T_H17 cellen zijn ten slotte nodig voor de eliminatie van extracellulaire bacteriën en fungi.

Uiteraard is één soort dendritische cel niet in staat een naïeve T-cel te doen differentiëren tot om het even welke T-effectorcel, maar bestaan er verschillende subsets van DC. Deze subsets tonen bijvoorbeeld verschillen in *Toll-Like Receptor* (TLR) expressie en vertonen dus verschillen in antigeenaffiniteit.

Op basis van de expressie van een combinatie van receptoren kunnen de DC opgesplitst worden in vier subsets: *Conventional DC* (cDC), *Plasmacytoid DC* (pDC), *Monocyt-derived DC* (moDC) en Langerhanscellen (LC). De cDC worden nogmaals opgesplitst in twee groepen. Tabel 2-1 stelt dit schematisch voor.

Tabel 2-1: Overzicht van de verschillende subsets van DC.

DC					
pDC	cDC			moDC	LC
	<i>Lymfoïd weefsel</i>		<i>Non-lymfoïd weefsel</i>		
	CD8α ⁺ cDC	CD11b ⁺ cDC	CD103 ⁺ CD11b ⁻ cDC	CD103 ⁻ CD11b ⁺ cDC	

De algemene eigenschappen, het fenotype en de functie van de verschillende DC subsets worden hieronder besproken.

2.2.1 *Conventional DC*

Deze subset van DC komt zowel voor in de meeste lymfoïde weefsels als in de meeste non-lymfoïde weefsels. cDC zijn een zeer belangrijke subset omwille van hun grote gevoeligheid voor weefselbeschadings en het opnemen, verwerken en presenteren van vreemde antigenen, terwijl er tolerant wordt omgegaan met lichaamseigen antigenen [8]. Deze functies worden zo goed beoefend door de cDC, omdat deze volgende eigenschappen bezitten:

Analyse van *Xbp1* KO-microarray data in cDC

- Ze vormen een lokaal verdedigingssysteem, want komen ook voor in non-lymfoid weefsel.
- Ze blinken uit in het verwerken en presenteren van antigenen.
- Ze zijn de beste in het migreren naar de lymfeknoop (zowel in *steady state* als in *inflammatory state*) ten opzichte van de andere DC subsets.
- Ze zijn zeer goed in staat naïeve T-cellen te doen differentiëren tot een bepaalde T-effectorcel [8].

2.2.1.1 *CD8α⁺* en *CD103⁺* DC

De $CD8\alpha^+$ DC en de $CD103^+$ DC hebben dezelfde progenitor en delen ook gelijkaardig fenotype en transcriptieprofiel. $CD8\alpha^+$ DC expresseren $CD8\alpha$, maar niet $CD8\beta$. $CD103^+$ DC expresseren daarentegen de receptor $CD103$ [8].

De $CD8\alpha^+$ DC en $CD103^+$ DC expresseren een aantal merkers die niet door de $CD11b^+$ DC worden geëxprimeerd [8]. Deze merkers zijn:

- $CD205$ = merker die bindt aan stervende cellen, herkent meer bepaald aan liganden die geëxprimeerd worden tijdens apoptose en necrose [9].
- $Clec9a$ = merker die bindt aan necrotische cellen [8].
- $Xcr1$ = chemokine receptor [10].
- $CD36$ = merker die bindt aan dode cellen [8].

Verder expresseren deze subsets ook genen gerelateerd aan MHC I (Major Histocompatibiliteitscomplex klasse 1) presentatie en hogere levels van $Flt3$ (*FMS-like tyrosine kinase 3*) ten opzichte van $CD11b^+$ DC. $CD8\alpha^+$ DC en $CD11b^+$ DC zijn eveneens de enige cDC die $TLR3$ en $TLR11$ expresseren. $TLR3$ herkent dubbelstrengig viraal RNA en $TLR11$ herkent *Toxoplasma gondii*, de parasiet die toxoplasmose veroorzaakt [8].

De $CD8\alpha^+$ DC en de $CD103^+$ DC ontbreken volgende merkers [8]:

- $Clec4a4$ = merker typisch voor $CD11b^+$ DC.
- $CD115$ = merker typisch voor macrofagen.
- $CD11b$ = merker typisch voor $CD11b^+$ DC en macrofagen.
- $CD172a$ = merker typisch voor $CD11b^+$ DC en macrofagen.
- $F4/80$ = merker typisch voor macrofagen.
- Cx_3CR1 = merker typisch voor macrofagen.

Analyse van *Xbp1* KO-microarray data in cDC

CD8 α^+ DC en CD103 $^+$ DC bezitten de uitmuntende eigenschap om antigenen via MHC I te presenteren aan CD8 α^+ T-cellen. CD8 α^+ DC en CD103 $^+$ DC blinken dus uit in cross-presentatie ten opzichte van de CD11b $^+$ DC. De reden hiertoe is dat de CD8 α^+ en CD103 $^+$ DC de twee eigenschappen bezitten die nodig zijn om goed te cross-presenteren [8]. Deze twee eigenschappen zijn:

- Ze bezitten een endocytische *pathway* met een laag afbrekend potentieel.
- Ze bezitten een fagosoom-naar-cytosol transport dat de transfer van het antigeen van het fagosoom naar het cytosol toestaat en het presenteren van het antigeen op MHC I toestaat.

Verder zijn deze DC ook de voornaamste bron van interleukine 12 (IL-12) en IL-15, twee cytokines die betrokken zijn in de differentiatie van cytotoxische T-cellen. CD8 α^+ DC en CD103 $^+$ DC zijn ook de enige hematopoëtische cellen die Xcr1 (*Chemokine (C motif) receptor 1*) expresseren. Het ligand van Xcr1, Xcl1 (*Chemokine (C motif) ligand 1*), wordt snel geproduceerd door CD8 T-cellen wanneer hen een antigeen gepresenteerd wordt door CD8 α^+ /CD103 $^+$ DC. Bovendien promoot Xcr1 ook de differentiatie van CD8 T-cellen tot cytotoxische T-cellen [8].

Deze subset lijkt dus gespecialiseerd in het activeren van CD8 T-cellen. Aangezien de meeste T-suppressorcellen differentiëren uit de CD8 T-cellen, zijn de CD8 α^+ DC en CD103 $^+$ DC van kritiek belang bij het vertonen van tolerantie. De rol van CD8 α^+ /CD103 $^+$ DC in de activatie van CD4 T-cellen is momenteel nog niet duidelijk [8].

2.2.1.2 CD11b $^+$ DC

De CD11b $^+$ DC ontbreken de CD103 merker en CD8 merker en expresseren in plaats daarvan CD11b en Clec4a4 (*C-type lectin domain family 4, member 4a*). Deze subset expresseert ook Flt3 en prolifereert dus in aanwezigheid van Flt3l (*FMS-like tyrosine kinase 3 ligand*), maar minder in vergelijking met de CD8 α^+ /CD103 $^+$ DC. Verder expresseren CD11b $^+$ DC *Pattern Recognition Receptors* (PRRs) die duidelijk verschillend zijn van de CD8 α^+ /CD103 $^+$ DC [8].

CD11b $^+$ DC blinken uit in de presentatie van antigenen via MHC II en activeren dus vooral CD4 T-cellen. Ex vivo onderzoek heeft echter aangetoond dat CD11b $^+$ DC wel degelijk in staat zijn een antigeen te presenteren aan CD8 T-cellen en dus met andere woorden wel degelijk in staat zijn te cross-presenteren, weliswaar in minder mate ten opzichte van de CD8 α^+ /CD103 $^+$ DC [8].

Analyse van *Xbp1* KO-microarray data in cDC

2.2.2 Andere subsets DC

2.2.2.1 Plasmacytoid DC

Plasmacytoid DC presenteren een kleine groep DC gehuisvest in de lymfeknopen. In de *steady state* vertonen de pDC een lage expressie van MHC II (Major Histocompatibiliteitscomplex klasse 2), costimulerende moleculen (bijvoorbeeld B7) en de transmembrane receptor CD11c. Daarentegen worden wel twee PRRs geëxprimeerd door deze subset, TLR7 en TLR9, die typische onderdelen van bacteriën en virussen herkennen [8].

Bij herkenning van vreemd DNA (Desoxyribonucleïnezuur) zullen pDC grote hoeveelheden interferon-alfa (IFN- α) produceren en de mogelijkheid verwerven om het vreemde antigeen te presenteren (door opregulatie van MHC II) [8]. Toch zijn pDC algemeen minder goed in het presenteren van antigenen en in migratie in vergelijking met cDC.

2.2.2.2 Monocyte-derived DC

Zoals hun naam al doet vermoeden, ontwikkelen de *monocyte-derived* DCs uit monocyten. In *steady state* is het aantal moDC zeer laag, terwijl deze subset tijdens infectie de meest voorkomende dendritische cel is. De functie van moDC bestaat erin pro-inflammatoire cytokines te produceren en antigenen te presenteren aan zowel CD4 T-cellen als CD8 T-cellen (cross-presentatie)[11][12][13].

De laatste jaren hebben al heel wat rapporten het bestaan van een moDC subset beschreven [13][14]. De zogenaamde *TNF iNOS Producing Dendritic cells* (Tip-DC) zijn vooral betrokken in het aangeboren immuunsysteem [15]. Tip-DC zijn namelijk de enige DC die in staat zijn *Tumor Necrosis Factor* (TNF) en *Inducible Nitric Oxide Synthase* (iNOS) te produceren en op die manier een pathogeen te doden zoals een macrofaag.

2.2.2.3 Langerhanscellen

De Langerhanscellen zijn de DC die aanwezig zijn in de epidermale laag van de huid. Deze DC staan apart van de andere DC omwille van hun unieke ontogenie en homeostatische eigenschappen [8].

Langerhanscellen worden gekenmerkt door een lagere expressie van MHC II, een gemiddelde CD11c expressie en een hoge expressie van CD207. In tegenstelling tot de meeste DC ontwikkelen Langerhanscellen onafhankelijk van Flt3 en Flt3l. Daarentegen is wel Csf-1R (*Colony Stimulating Factor 1 Receptor*) nodig net zoals bij de macrofagen [8].

2.3 *Xbp1* en endoplasmatisch reticulum stress

Het ER is het organel dat verantwoordelijk is voor de productie, het opvouwen en de modificatie van vele membraaneiwitten en eiwitten die uiteindelijk gesecreteerd worden. De opvouwing en modificatie van eiwitten is een proces dat zeer gevoelig is aan verschillende factoren, zoals bijvoorbeeld een tekort aan Ca^{2+} (calciumion), oxidatieve stress, inflammatie, verhoogde proteïnesynthese... Slecht opgevouwen eiwitten zijn minder functioneel of helemaal niet functioneel en worden daarom vernietigd via de *ER-associated Degradation pathway* (ERAD). Wanneer het aantal niet en/of slecht opgevouwen eiwitten stijgt, zal de zogenaamde *Unfolded Protein Response* (UPR) geactiveerd worden. De activering van de UPR leidt tot een verminderde proteïnesynthese, een verbeterde ERAD van de niet en/of slecht opgevouwen proteïnen en biedt ook hulp bij het opvouwen van de proteïnen. Wanneer de cel desondanks de activering van de UPR niet in staat is het aantal niet en/of slecht opgevouwen proteïnen te verminderen, zullen *pathways* geactiveerd worden die leiden tot celdood [16].

De activering en werking van de UPR berust op drie transmembrane receptoren: *Inositol-requiring kinase 1* (Ire1), *Eukaryotic translation initiation factor 2 alpha kinase 3* (Eif2ak3) en *Activating transcription factor 6* (Atf6) (figuur 2-4)[16][17].

2.3.1 IRE1

Ire1 vormt de meest geconserveerde *pathway* van de UPR. Zoogdieren bezitten twee paralogen van deze molecule, namelijk Ire1 α en Ire1 β . Wanneer er geen sprake is van ER-stress is Ire1 gebonden aan het chaperone *Binding Immunoglobulin Protein* (BiP) en op die manier inactief. Doordat BiP een grotere bindingsaffiniteit vertoont voor niet of slecht opgevouwen eiwitten, zal BiP Ire1 loslaten bij ER-stress. Op die manier is Ire1 in staat te dimeriseren en autofosforylatie te ondergaan. Er is echter ook nog een tweede model beschreven voor de activering van Ire1. Dit model stelt dat de niet of slecht opgevouwen eiwitten een directe binding aangaan met Ire1 om dimerisatie te induceren. Onderzoek heeft tot nu toe aangetoond dat dit laatste model enkel van toepassing is voor het gist Ire1 (Ire1p) en niet voor het humane Ire1 [16][17].

Eens Ire1 geactiveerd is, knipt het een exon van 26 basenparen weg van het *Xbp1* mRNA. Op die manier wordt *Xbp1* geactiveerd, zodat *Xbp1* vervolgens migreert naar de nucleus om daar de expressie van bepaalde genen te induceren [16][17].

Opmerkelijk is dat gist enkel deze *pathway* bezit en dat het signaleren van ER-stress dus volledig afhankelijk is van Ire1 [16][17]. De UPR werd het eerst beschreven in gist in 1996 [18].

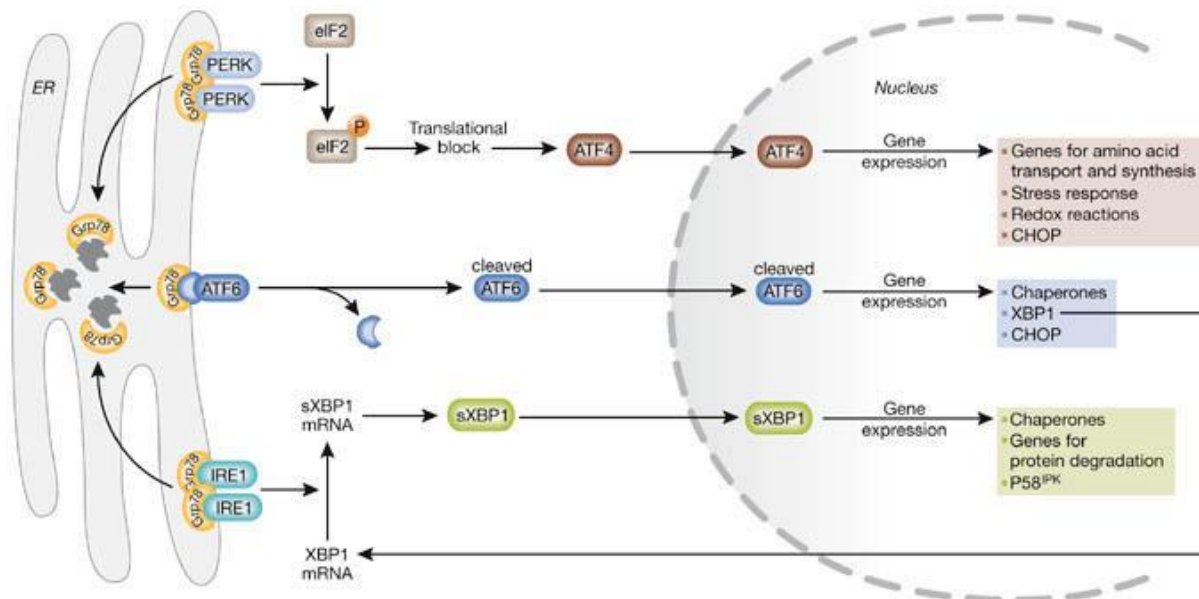
Analyse van *Xbp1* KO-microarray data in *cDC*

2.3.2 Eif2ak3

Net zoals Ire1 interageert Eif2ak3 (ook bekend als PERK) met BiP en blijft Eif2ak3 inactief zolang de eiwitopvouwing zonder problemen verloopt. Eens de UPR echter geactiveerd wordt, laat BiP Eif2ak3 los, zodat Eif2ak3 geactiveerd wordt via oligomerisatie en autofosforylatie. De functie van Eif2ak3 bestaat erin de transcriptiefactor eIF2 α (*Eukaryotic translation Initiation Factor 2 Alpha*) te fosforyleren, zodat deze op die manier de eiwitproductie kan verminderen. Naast de globale onderdrukking van de proteïnesynthese, verhoogt eIF2 α ook de translatie van een aantal specifieke mRNA's, zoals *Activating Transcription Factor 4 (ATF4)*, *Xbp1* en allerlei apoptose-inhibitoren. Eén van de bekendste apoptose-inhibitoren die downstream van eIF2 α ligt, is *CHOP (CCAAT/enhancer-binding protein homologous protein)*. CHOP activeert tal van pro-apoptotische factoren, zoals bijvoorbeeld *DR5 (Death Receptor 5)*. Verder induceert CHOP ook de transcriptie van *Ppp1r15a (Protein Phosphatase 1 Regulatory subunit 15A)*, wat op zijn beurt de proteïnesynthese verhoogt. CHOP verhoogt ook de expressie van *Ero1 α (Endoplasmic Oxidoreductin 1 like)*, een gen dat oxidatieve stress veroorzaakt en Ca²⁺ secreteert. Ca²⁺ is een vereiste opdat de ER chaperones hun functie zouden kunnen uitoefenen en oxidatieve stress is uit den boze tijdens eiwitopvouwingen [16][17].

2.3.3 ATF6

Tijdens de UPR wordt Atf6 geactiveerd door het loslaten van BiP. Geactiveerd Atf6 migreert onmiddellijk naar het Golgi-apparaat om daar geknipt te worden, zodat Atf6 p50 gevormd wordt. Atf6 p50 migreert op zijn beurt naar de nucleus om de expressie van bepaalde genen (bijvoorbeeld *Dnajc3, DnaJ (Hsp40) homolog, subfamily C, member 3*) te activeren. ATF kan ook heterodimeriseren met *Xbp1* tijdens ER-stress en op die manier de expressie van ERAD componenten, zoals *Edem1 (ER Degradation Enhancer, Mannosidase alpha-like 1)*, *Herpud1 (Homocysteine-inducible, Endoplasmic Reticulum stress-inducible, Ubiquitin-like Domain member 1)* en *Syvn1 (Synovial apoptosis inhibitor 1)* induceren. Net zoals bij Ire1 komen in zoogdieren twee paralogen van Atf6 voor, namelijk Atf6 α en Atf6 β [16][17].

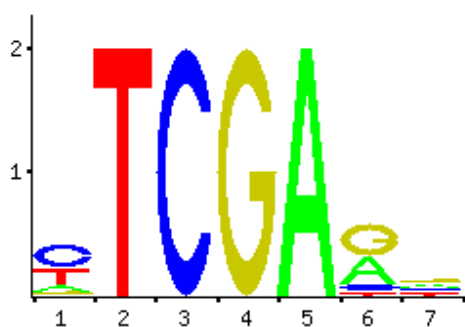


Figuur 2-4: Overzicht van de drie *pathways* van de UPR [19].

2.4 Genregulatie

2.4.1 Transcriptiefactor

Een transcriptiefactor wordt gedefinieerd als een proteïne dat bindt op specifieke DNA-sequenties om zo de transcriptie van een gen of enkele genen te stimuleren of te inhiberen [20]. Een transcriptiefactor kan dus pas binden aan een DNA-sequentie als die DNA-sequentie een bepaald motief bezit. Het motief waaraan Xbp1 bindt, is weergegeven in onderstaande figuur.



Figuur 2-5: DNA-motief waaraan Xbp1 bindt [21].

2.4.2 Netwerkkinteracties

De meeste proteïnen functioneren in samenwerking met andere proteïnen. Proteïnen vormen dus met andere woorden interacties met andere proteïnen. De verschillende proteïne-proteïne interacties vormen dan samen een bepaald netwerk dat een specifieke functie uitoefent. Dit heeft

Analyse van *Xbp1* KO-microarray data in *cDC*

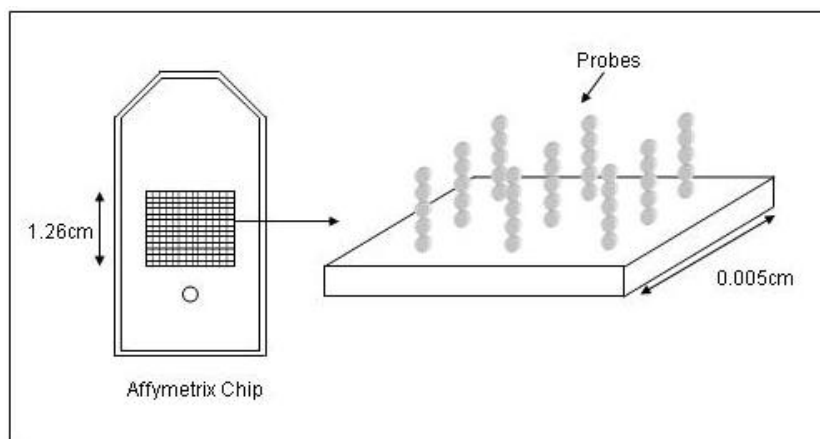
tot gevolg dat als men de functie wil kennen van een gen – of zijn proteïne – moet nagegaan worden tot welk netwerk dat gen/proteïne behoort [22].

2.5 DNA-microarrays

Bij een DNA-microarray experiment wordt de genexpressie in twee verschillende condities (bijvoorbeeld WT ten opzichte van KO) vergeleken [23][24][25]. Door dus een DNA-microarray te analyseren, kan achterhaald worden welke genen een andere expressie vertonen in bijvoorbeeld een KO conditie ten opzichte van een WT conditie.

2.5.1 Werking DNA-microarray

Een DNA-microarray is een plaatje waaraan de sequenties van duizenden genen vastgehecht zijn. In het eenvoudigste geval is dat plaatje een glazen microscoopglasje, maar het kan evengoed een *silicon chip* of een nylonmembraan zijn [24]. De sequenties die aan de plaat zijn vastgehecht, worden “probes” genoemd en worden in een groepje of een zogenaamde “spot” aangebracht op de plaat. Elke spot op de plaat komt dan dus overeen met een bepaald gen (figuur 2-6).

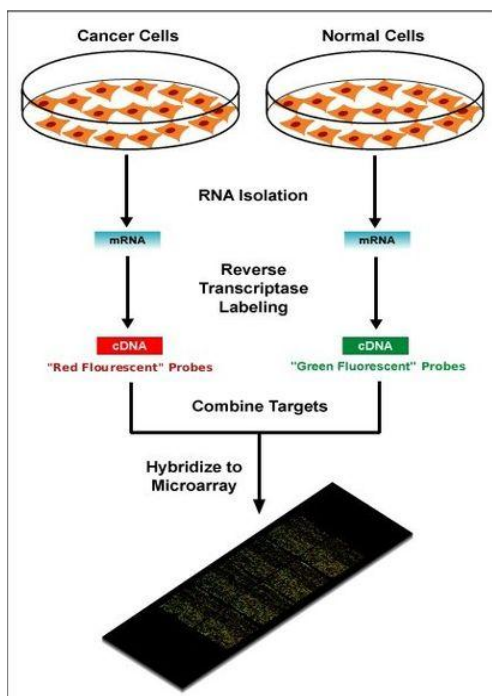


Figuur 2-6: Voorbeeld van een DNA-microarray. Een DNA-microarray is in het eenvoudigste geval een glazen microscoopglasje waaraan duizenden probes zijn gehecht [25].

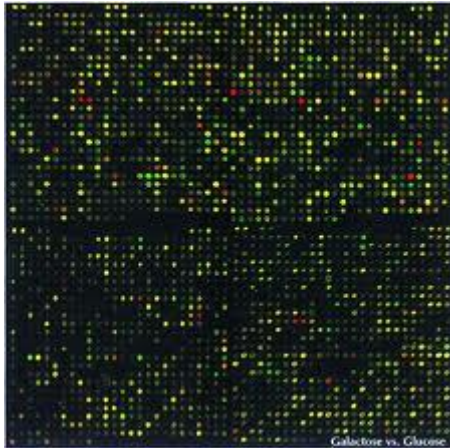
Stel een experiment waarbij de invloed van de KO van een bepaald gen wordt onderzocht via DNA-microarrays. Men wil dus met andere woorden via DNA-microarrays achterhalen welke genen een differentiële genexpressie (hogere of lagere expressie) vertonen in de KO cellen ten opzichte van de normale, gezonde cellen (WT). In zo’n experiment zal er sprake zijn van twee soorten stalen, namelijk een WT-staal en een staal waarbij het gen van interesse is uitgeschakeld, een KO-staal.

Analyse van *Xbp1* KO-microarray data in cDC

De eerste stap in een microarray experiment is de isolatie van een bepaalde hoeveelheid zuiver RNA voor beide soorten stalen. Vervolgens zal dat RNA omgezet worden naar cDNA (*copy* DNA) via het enzym *reverse transcriptase*. Daarna zal het cDNA van beide soorten stalen verschillend fluorescerend gelabeld worden. Meestal wordt het cDNA van het WT-staal groen gelabeld (via Cy3) en wordt het cDNA van het KO-staal rood gelabeld (via Cy5). De volgende stap houdt in dat het gelabelde cDNA van beide stalen vermengd en aangebracht wordt op de microarray. Op die manier komt het cDNA in contact met de verschillende probes en kan het cDNA hybridiseren aan zijn complementaire sequentie (figuur 2-7). Stel dat gen X in de KO sterk upgereguleerd wordt, dan betekent dit dat er dan veel mRNA (*messenger* RNA) van dit gen werd gevormd, wat dus ook betekent dat er voor dit gen veel meer “rood cDNA” aanwezig is dan “groen cDNA”. Doordat het “rode cDNA” veel talrijker is, zal er meer “rood cDNA” hybridiseren aan de probes dan “groen cDNA”. De spot die gen X voorstelt, zal dus rood kleuren. Verder kan een spot ook nog groen, geel of zwart kleuren. Een groene kleur betekent dat er meer “groen cDNA” aanwezig is, wat dus eigenlijk betekent dat het betreffende gen in de KO-conditie downgereguleerd werd. Een gele kleur betekent dat het betreffende gen in de WT-conditie en de KO-conditie op dezelfde manier tot expressie komt en een zwarte spot is een spot waaraan helemaal geen cDNA gehybridiseerd is (figuur 2-7).



Figuur 2-7: De procedure van een microarray experiment. De eerste stap is de isolatie van mRNA uit beide soorten cellen, zodat dit mRNA vervolgens kan omgezet worden in cDNA. Nadat het cDNA van beide soorten cellen verschillend fluorescerend gelabeld werd, wordt een mengsel van deze cDNA sequenties op de microarray gebracht, zodat hybridisatie kan plaatsvinden. Ten slotte zal de microarray ingelezen worden door een scanner [26].



Figuur 2-8: Voorbeeld van een microarray. Een microarray kan vier kleuren vertonen: rood, groen, geel en zwart [27].

Vervolgens zal de microarray, die allemaal gekleurde spots vertoont (figuur 2-8), “gelezen” worden door een scanner. Die scanner overloopt dus de microarray spot voor spot en bepaalt daarbij telkens de kleur van de spot. Hierbij is het belangrijk op te merken dat er veel variëteiten in de vier mogelijke kleuren kunnen voorkomen. Stel dat bijvoorbeeld gen X in de KO-conditie sterk upgereguleerd wordt, terwijl gen Y slechts een beetje upgereguleerd wordt, dan is het logisch dat de spot van gen X intenser rood zal gekleurd zijn dan de spot van gen Y. Ten slotte zal de scanner de gemeten kleur vergelijken met een standaard, zodat de kleur kan omgezet worden naar een bepaald getal.

Men kan drie soorten DNA-microarrays onderscheiden op basis van de manier waarop de microarray is gemaakt. De eerste soort DNA-microarray die ontwikkeld werd, was de gespote DNA-microarray. Later volgde de *in situ* gehybridiseerde oligonucleotide-microarray en de *bead* DNA-microarrays.

3 Materiaal en methoden

3.1 Personal Computer

DMBR029

- Besturingssysteem: Fedora release 15
- Model: Dell Precision desktop
- Aantal processoren: 2
- Geheugen: 6 GB
- Scherm: 24 inch

DMBR101

- Besturingssysteem: Fedora release 15
- Model: Dell Precision desktop
- Aantal processoren: 8
- Geheugen: 12 GB

3.2 R – Bioconductor

R is een softwarepakket en programmeertaal die gebruikt wordt voor statistische en data-analysedoeleinden. R is een GNU-poject (*GNU's Not Unix*) dat verwant is met de statistische programmeertaal S. R wordt meer bepaald aanzien als een implementatie van S. In 1993 werd R ontwikkeld door Ross Ihaka en Robert Gentleman aan de universiteit van Auckland (Nieuw-Zeeland) [28].

Bioconductor is een *open-source* softwarepakket dat tools voorziet voor een analyse van *high-throughput* genomische data (bijvoorbeeld microarrays). Bioconductor bestaat immers uit verschillende R *packages* die de gebruiker de mogelijkheid bieden bepaalde commando's uit te voeren op genomische data [29][30].

3.3 Contra V2

Contra V2 is een tool waarmee de bindingplaatsen van transcriptiefactoren kunnen gevisualiseerd en gedetecteerd worden. Het gebruik van Contra V2 wordt opgedeeld in vier stappen. In de eerste stap geeft de gebruiker het organisme en gen van interesse in. Contra V2 biedt negen organismen aan, gaande van mens tot gist. In de tweede stap selecteert de gebruiker het

Analyse van *Xbp1* KO-microarray data in cDC

transcript van het gen van interesse. In de volgende stap kan de gebruiker kiezen welke transcriptiefactoren hij wil onderzoeken. De *Position Weight Matrices* (PWMs) worden opgehaald uit vier databanken: JASPAR, TRANSFAC (TRANScription FACTor), phyloFACTS en PBM homeodomains (*Protein Binding Microarrays*). Contra V2 biedt de gebruiker eveneens de mogelijkheid eigen PWMs te uploaden. De laatste stap houdt het analyseren van de resultaten in [31]. Contra V2 is gratis beschikbaar via <http://bioit.dnbr.ugent.be/contrav2/>.

3.4 MEME Suite

3.4.1 TOMTOM

TOMTOM (*Motif Comparison Tool*) is een tool die gebruikt wordt om na te gaan welke motieven aanwezig zijn in bepaalde DNA sequenties. TOMTOM vergelijkt meer bepaald de ingegeven DNA sequenties met een databank van gekende motieven. Het resultaat is een lijst van de matchende motieven gerangschikt volgens q-waarde (zie 3.11.4) [32][33].

3.4.2 FIMO

Find Individual Motif Occurrences of kortweg FIMO is een tool die het de gebruiker mogelijk maakt te achterhalen welke sequenties uit een lijst van sequenties een bepaald motief bevatten. De gebruiker dient dus twee bestanden te voorzien. Ten eerste voorziet de gebruiker een bestand met de motieven die hij wil onderzoeken. Dit bestand moet in MEME (*Multiple EM for Motif Elicitation*) formaat zijn. Het tweede bestand dat nodig is, is een fasta bestand met alle sequenties die de gebruiker wenst te scannen op de aanwezigheid van de motieven uit het eerste bestand [34][35].

3.5 Ingenuity Pathway Analysis® (IPA®)

IPA® is betalende software ontwikkeld door het bedrijf Ingenuity Systems en is een grote hulp bij het verwerken van grote hoeveelheden data. Via IPA® zijn onderzoekers bijvoorbeeld in staat om bekomen microarray resultaten om te gieten in netwerken en deze vervolgens op een gebruiksvriendelijke manier te analyseren. Daarnaast biedt IPA® ook nog tal van andere mogelijkheden aan, zoals "*Comparison Analysis*" om bijvoorbeeld de meest significante *pathway* te achterhalen, "*Path Designer*" om de lay-out van de *pathways* naar eigen smaak te kunnen aanpassen, "*Gene and ChemView*" om snel gerichte informatie over een bepaald gen te kunnen opzoeken... IPA® is dus tegenwoordig niet weg te denken uit de *life science community* en het feit dat IPA® reeds geciteerd is in duizenden wetenschappelijk artikels getuigt daarvan [36].

3.6 Analyse microarray data via Bioconductor

De analyse van ruwe microarray data verloopt in verschillende stappen:

1. Controle kwaliteit ruwe data via plots
 2. Normalisatie data
 3. Identificatie genen met differentiële genexpressie
 4. Annotatie genen met differentiële genexpressie
 5. *Pathway* analyse
- } Bioconductor – R
} IPA®

Het doel van een microarray data-analyse is te achterhalen welke genen een significante differentiële genexpressie vertonen in de onderzochte situatie (bijvoorbeeld een KO) ten opzichte van de WT situatie. De functie(s) van deze genen wordt bepaald door vervolgens *Gene Ontology* (GO) data te koppelen aan de genen (annotatie). Als laatste stap worden de netwerken waarvan deze genen deel uitmaken, geanalyseerd, zodat op die manier kennis verworven wordt over welke functies beïnvloed worden in de onderzochte situatie.

De verschillende stappen van een microarray analyse worden hieronder besproken.

3.7 Controle kwaliteit ruwe data via plots

3.7.1 Histogram

Door een histogram te maken van elke microarray kan de verdeling van de probe-intensiteiten vergeleken worden tussen de verschillende microarrays. Het is vooral belangrijk te letten op de positie en de vorm van de verschillende histogramcurves. De meeste verschillen hierin zullen verdwijnen na normalisatie, omdat deze te wijten zijn aan *noise* die sowieso voorkomt. Een histogram dat echter een significante andere vorm en positie vertoont, duidt op een artefact [37].

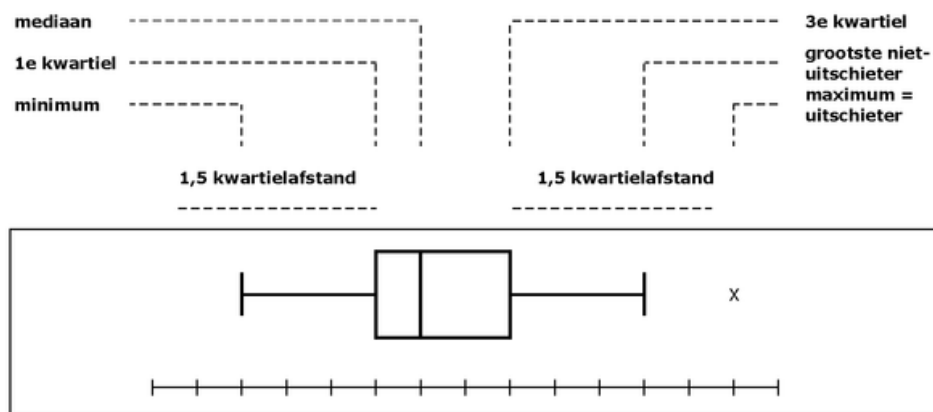
Meestal wordt de \log_2 van de intensiteiten genomen, omdat op die manier de data beter verdeeld worden over de x-as. Indien men dat niet zou doen, zouden de histogramcurves te veel tegen de y-as gedrukt worden.

3.7.2 Boxplot

Een boxplot geeft net zoals een histogram informatie over de verschillen in de probe-intensiteitenverdeling tussen de verschillende microarrays. Hierbij is het opnieuw belangrijk dat de vormen van de verschillende boxplots niet te veel van elkaar afwijken [37].

Analyse van *Xbp1* KO-microarray data in *cDC*

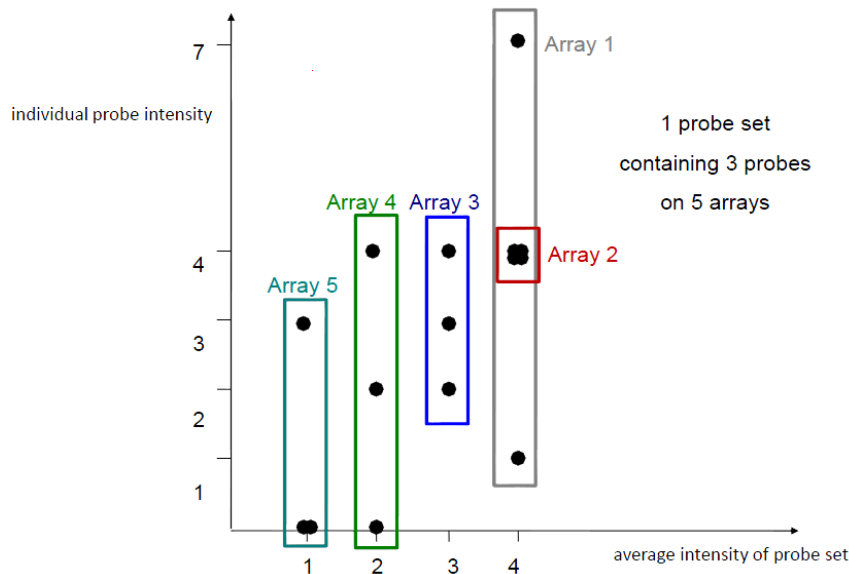
Een boxplot of een doosdiagram is een grafische weergave van de vijf-getallensamenvatting. Dit betekent dat de boxplot informatie geeft over vijf statistische waarden, namelijk de mediaan, het 25^{ste} percentiel, het 75^{ste} percentiel, de laagste waarde die geen uitschieter is en de grootste waarde die geen uitschieter is. Wanneer waarde X gelijk is aan het 25^{ste} percentiel betekent dit dat 25% van de data kleiner is dan waarde X. Op analoge wijze is 75% van de data kleiner dan het 75^{ste} percentiel (figuur 3-1) [38].



Figuur 3-1: Voorbeeld van een boxplot [39].

3.7.3 Chip pseudo-image

Vooraleer er een *chip pseudo-image* kan getekend worden, moet er een model gefit worden op de data. Stel dat men in de studie vijf microarrayexperimenten uitvoerde, waarbij men gebruik maakte van een probeset bestaande uit drie probes [37]. Een plot waarbij de individuele probe-intensiteiten uitgezet worden ten opzichte van de gemiddelde probesetintensiteiten, zou er als volgt kunnen uitzien:

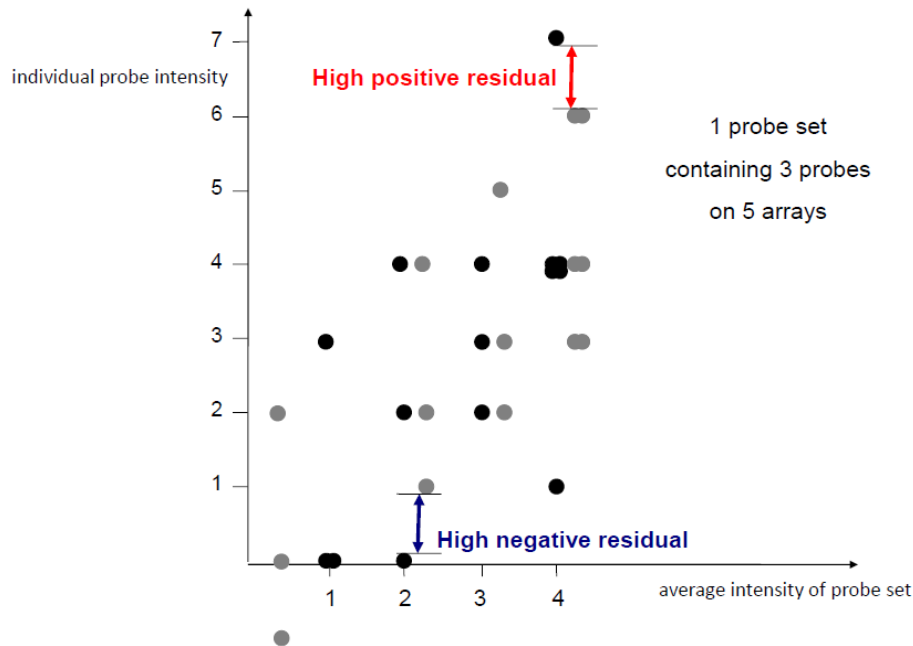


Figuur 3-2: De gemeten data voor een bepaalde probeset over vijf microarrays [40].

Figuur 3-2 stelt hetgeen voor dat gemeten wordt voor een bepaalde probeset over de verschillende microarrays. Daarbij is het belangrijk op te merken dat het gemeten signaal van een microarray eigenlijk een combinatie is van drie zaken:

- De expressiewaarde van het gen.
- De kwaliteit van de specifieke probe. Sommige probes hybridiseren immers zeer goed, andere veel minder.
- Een foutenmarge, die ook de *residual* genoemd wordt.

Vandaar dat een model gebouwd wordt dat ervan uitgaat dat de probes van een bepaalde probeset zich op dezelfde manier gedragen over de verschillende microarrays. Als de probes dus goed hybridiseren op de ene microarray, gaat het model ervan uit dat dit ook geldt voor de andere microarrays. Dit model wordt dan vervolgens gefit op de data om een idee te krijgen in hoeverre de gemeten waarden afwijken van het model. Er wordt dus met andere woorden achterhaald in hoeverre de probes van een bepaalde probeset zich op dezelfde manier gedragen over de verschillende microarrays. Ten slotte worden de verschillen tussen de gemeten data en het model berekend (de *residuals*). Het model gefit op de gemeten data wordt voorgesteld in figuur 3-3.



Figuur 3-3: Een model fitten op gemeten data. De zwarte punten stellen de gemeten data voor, terwijl de grijze punten het model voorstellen [40].

Een hoge negatieve of positieve *residual* duidt dus op een groot verschil tussen de gemeten waarde en de verwachte waarde. Hieruit volgt dus: hoe lager de *residual*, hoe beter. Vandaar dat er gewichten (*weights*) toegekend worden aan de *residuals*. De gewichten zijn gelegen tussen nul en één en hierbij komt een gewicht van één overeen met een kleine *residual* en een gewicht van nul overeen met een grote *residual* [37].

3.7.3.1 *Chip pseudo-image - residuals*

Er zijn verschillende opties om een *chip pseudo-image* van de *residuals* te tekenen. Een methode die echter veel toegepast wordt, is deze waarbij negatieve *residuals* afgebeeld worden in blauw en positieve *residuals* afgebeeld worden in rood. Het is wel belangrijk op te merken dat een *chip pseudo-image* gemaakt wordt voor de verschillende microarrays apart, zodat de kwaliteit van de verschillende microarrays apart kan beoordeeld worden. Deze plots zijn dus met andere woorden niet geschikt om te vergelijken tussen verschillende microarrays [37].

3.7.3.2 *Chip pseudo-image - weights*

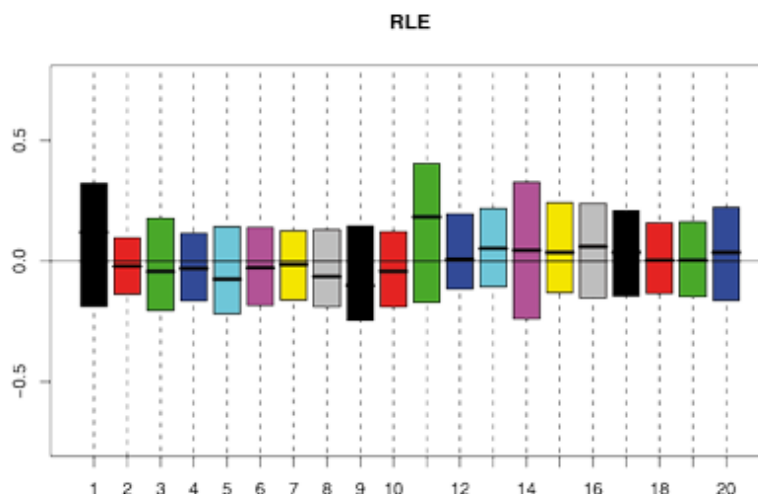
Op een *chip pseudo-image* van *weights* worden gewichten rond nul in groen afgebeeld, terwijl gewichten rond één in lichtgrijs afgebeeld worden. Aangezien de *weights* rond nul zoveel mogelijk moeten vermeden worden, betekent dit dat er zo weinig mogelijk groen moet voorkomen op de

Analyse van *Xbp1* KO-microarray data in *cDC*

chip pseudo-image. Aangezien microarrays niet exact zijn, zal er steeds wat groen aanwezig in deze plots. Het is vooral belangrijk dat er geen zones voorkomen die volledig groen kleuren [37].

3.7.4 RLE (*Relative Log Expression*)

Een RLE-plot is een boxplot van de RLE-waarden voor een bepaalde microarray. De RLE-waarden worden per probeset bepaald en berekend door de ratio te bepalen tussen het logaritme van de probesetexpressie en de mediaan van die probesetexpressies over alle microarrays. Ten eerste gaat men ervan uit dat de meeste genen geen verandering in expressie vertonen over de verschillende microarrays en ten tweede gaat men ervan uit dat het aantal genen met een verhoogde expressie gelijk is aan het aantal genen met een verlaagde expressie. Dit heeft tot gevolg dat de RLE-waarden grotendeels rond nul gelegen zijn. Een tweede manier om microarrays met een lagere kwaliteit op te sporen, is de grootte van de boxplot. De grootte van de boxplots zou ongeveer gelijk moeten zijn voor de verschillende microarrays (figuur 3-4) [41].



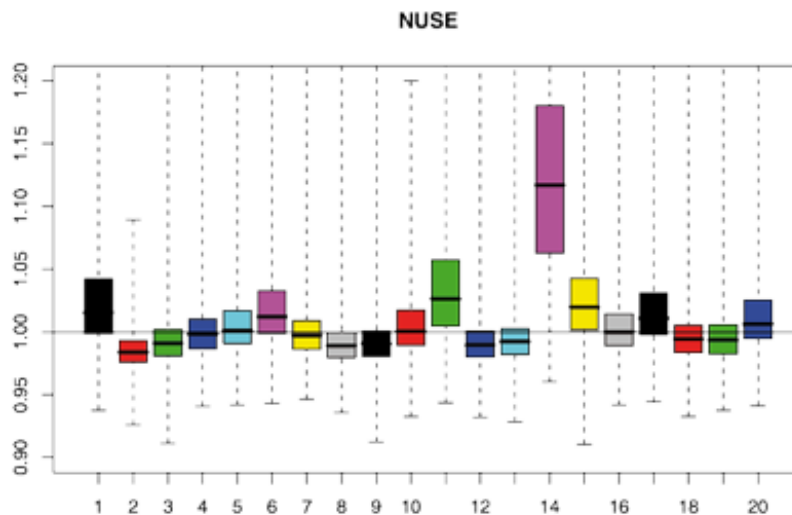
Figuur 3-4: Voorbeeld van een RLE-plot. De meeste boxplots zijn gelokaliseerd in de buurt van nul. De groottes van de boxplots zijn gelijkaardig voor de meeste microarrays. Enkel microarrays 1, 11 en 14 vertonen een afwijking [41].

3.7.5 NUSE (*Normalized Unscaled Standard Errors*)

Voor een NUSE-plot worden eerst de standaarddeviaties van de expressies van de verschillende probesets berekend. Vervolgens worden de standaarddeviaties zodanig genormaliseerd dat de mediaan van de standaarddeviaties over alle microarrays ongeveer gelijk is aan één voor elke probeset. De NUSE-waarden worden berekend door de ratio te bepalen tussen de standaarddeviatie van een bepaalde probeset ten opzichte van de mediaan [42]. Uiteindelijk wordt er een boxplot getekend van de NUSE-waarden voor de verschillende microarrays apart.

Analyse van *Xbp1* KO-microarray data in *cDC*

Voor de interpretatie van een NUSE-plot is het vooral belangrijk te kijken in hoeverre de boxplots gelokaliseerd zijn rond één. NUSE is wel zeer gevoelig aan *noise* of variabiliteit in de data, waardoor er steeds een zekere schommeling waarneembaar is. Meestal wordt de grens voor een afwijkende boxplot op 1,1 gezet (figuur 3-5) [41].



Figuur 3-5: Voorbeeld van een NUSE-plot. Microarray 14 vertoont een grote afwijking, wat betekent dat er in dit microarray experiment mogelijks een fout is gebeurd. Microarrays 1 en 11 vertonen net zoals in de RLE-plot een grotere afwijking, maar deze afwijking is niet alarmerend, aangezien de grens van 1,1 niet overschreden is [41].

3.8 Normalisatie

Het meest toegepaste algoritme voor normalisatie is het *Robust Multi-array Average* (RMA) algoritme. Alternatieven voor het RMA algoritme zijn: *Microarray Suite 5.0* (MAS 5.0), *Probe Logarithmic Intensity Error* (PLIER), DNA chip (dChip) en *Guanine Cytosine Robust Multi-array Average* (GCRMA) [37]. De normalisatie via het RMA algoritme kan onderverdeeld worden in vier stappen:

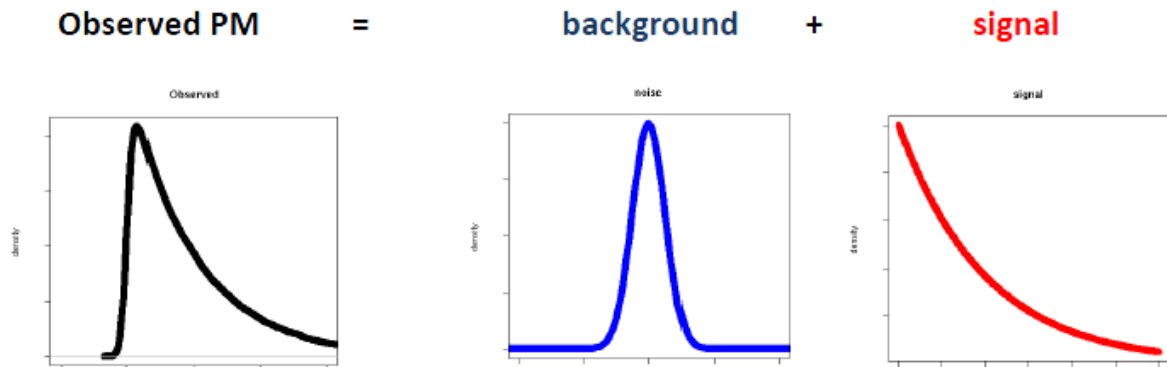
1. Achtergrondcorrectie
2. Log₂-transformatie
3. *Quantile* normalisatie
4. Probenormalisatie

3.8.1 Achtergrondcorrectie

Achtergrond of *background* wordt veroorzaakt doordat de microarray, los van de probes, ook een signaal geeft. Hierbij is het belangrijk op te merken dat de achtergrond niet constant is, maar een

Analyse van *Xbp1* KO-microarray data in *cDC*

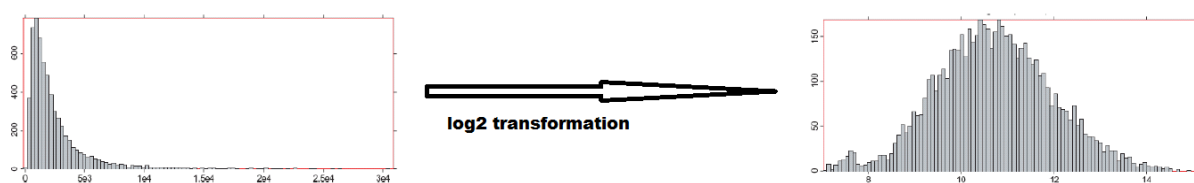
normaalverdeling volgt. Sommige zones op een microarray vertonen immers een grotere achtergrond dan andere zones. Dit betekent dus dat de geobserveerde expressie eigenlijk een som is van het werkelijke signaal en de achtergrond [37]. Figuur 3-6 stelt dit voor.



Figuur 3-6: De geobserveerde expressie is eigenlijk een som van het eigenlijke signaal en de achtergrond [40].

3.8.2 Log₂-transformatie

De log₂-transformatie is een noodzakelijke stap in de microarray analyse, omdat vele statistische procedures (bijvoorbeeld gemiddelde of standaarddeviatie berekenen) enkel kunnen toegepast worden op data die zo goed mogelijk de normaalverdeling volgen. Bovendien zorgt de log₂ transformatie er ook voor dat de data beter visualiseerbaar en interpreteerbaar worden (figuur 3-7) [43].



Figuur 3-7: Resultaat van een log₂-transformatie. Een log₂-transformatie leidt tot een normaalverdeling van de data [40].

3.8.3 Quantile normalisatie

Door het uitvoeren van een *quantile* normalisatie worden de datadistributies van de verschillende microarrays gelijkgesteld aan elkaar. Dit wordt bekomen door eerst per microarray de hoogste probe-intensiteit (die ondertussen achtergrondcorrectie en log₂ transformatie heeft ondergaan) te detecteren. Vervolgens wordt het gemiddelde van die probe-intensiteiten berekend en worden die hoogste probe-intensiteiten vervangen door hun gemiddelde. Dit proces wordt daarna herhaald voor de probe-intensiteiten die oorspronkelijk de tweede hoogste waarden vertoonden.

Analyse van *Xbp1* KO-microarray data in cDC

Zo blijft men doorgaan totdat de oorspronkelijke laagste probe-intensiteiten vervangen zijn door hun gemiddelde. Op die manier zullen uiteindelijk het gemiddelde en de standaarddeviatie van de probe-intensiteiten dezelfde waarde vertonen over de verschillende microarrays [37][43]. Een uitgewerkt voorbeeld van dit algoritme is terug te vinden in bijlage 1.

3.8.4 Probenormalisatie

Daar *quantile* normalisatie corrigeert voor variatie tussen de microarrays, corrigeert probenormalisatie voor de variatie van de probesets over de verschillende microarrays. Deze correctie wordt uitgevoerd, omdat de probes van een bepaalde probeset zich verschillend gedragen over de verschillende microarrays. Vandaar dat deze normalisatie noodzakelijk is, opdat het gedrag van de probesets over de verschillende microarrays gelijk is en op die manier vergelijkingen kunnen gemaakt worden [37][43].

Praktisch wordt dit verwezenlijkt door een lineair additief model te fitten op de data. Dit gebeurt op dezelfde manier als bij het maken van een *chip pseudo-image* (zie 3.7.3) [37][43].

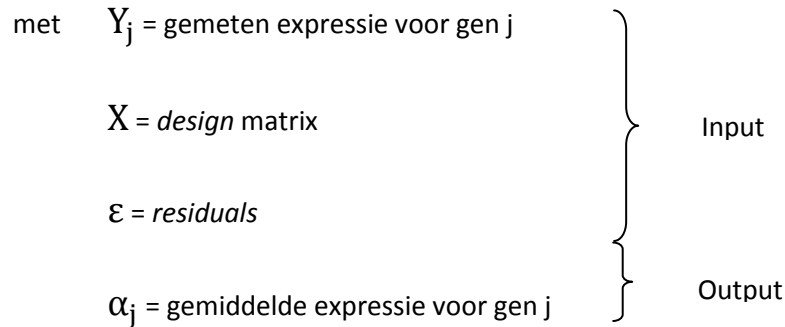
3.9 Identificatie van genen met differentiële genexpressie

Voor de identificatie van de genen die een differentiële genexpressie vertonen, wordt het “limma package” van Bioconductor gebruikt. Het centrale idee van limma is een lineair model te fitten op de (genormaliseerde) data om op die manier een schatting te bekomen van de gemiddelde expressie van elk gen. Eens men beschikt over de gemiddelde genexpressies kan er nagegaan worden of er expressiewaarden aanwezig zijn die significant verschillend zijn. Dit laatste gebeurt via een “*moderated T-test*” [40].

3.9.1 Lineair model fitten op data

Er wordt een lineair model gefit op de data zodat de gemiddelde expressie per gen kan bepaald worden. Om een lineair model te kunnen fitten op de genormaliseerde data is echter eerst een *design* matrix nodig (figuur 3-8). Een *design* matrix geeft aan welke waarden behoren tot de controle microarrays (bijvoorbeeld *wild type*) en welke waarden behoren tot de conditie die onderzocht wordt (bijvoorbeeld een *knock-out*) [40]. De gemiddelde expressie van een gen wordt via volgende formule berekend:

$$Y_j = X \cdot \alpha_j + \varepsilon \quad (3.1)$$



Bovenstaande formule toegepast op een voorbeeld is weergegeven in figuur 3-8.

$$\begin{array}{c}
 y_j \\
 \begin{bmatrix} 1.2 \\ 1.8 \\ 2.4 \\ 3 \\ 3.6 \\ 4.2 \end{bmatrix}
 \end{array}
 =
 \begin{array}{c}
 X \\
 \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}
 \end{array}
 \cdot
 \begin{array}{c}
 \alpha_j \\
 \begin{bmatrix} \mu_{wt} \\ \mu_{KO} \end{bmatrix}
 \end{array}
 +
 \begin{array}{c}
 \varepsilon \\
 \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix}
 \end{array}
 \quad
 \begin{array}{l}
 1.2 = \mu_{wt} + \varepsilon_1 \\
 1.8 = \mu_{wt} + \varepsilon_2 \\
 2.4 = \mu_{wt} + \varepsilon_3 \\
 3 = \mu_{KO} + \varepsilon_4 \\
 3.6 = \mu_{KO} + \varepsilon_5 \\
 4.2 = \mu_{KO} + \varepsilon_6
 \end{array}$$

Figuur 3-8: Voorbeeld van lineair model fitten op data [40].

3.9.2 Standaard T-test

Een standaard T-test wordt gebruikt om na te gaan of twee gemiddelden significant verschillend zijn van elkaar. Er zijn dus twee mogelijke hypothesen: de gemiddelden zijn niet significant verschillend van elkaar (nulhypothese) of de gemiddelden zijn significant verschillend van elkaar (alternatieve hypothese). Vervolgens wordt een significantieniveau vastgelegd. Meestal kiest men voor een significantieniveau van 0,05, wat betekent dat men uiteindelijk voor slechts 95% zeker is van de resultaten. Ten slotte wordt de T-waarde berekend op basis van onder andere de gemiddelden en standaarddeviaties. De bekomen T-waarde wordt vergeleken met een kritieke T-waarde (een vastgelegde T-waarde afhankelijk van het significantieniveau); indien de T-waarde groter is dan de kritieke T-waarde wordt de alternatieve hypothese aanvaard, in het andere geval wordt de nulhypothese aanvaard [44].

De standaard T-test kan echter niet toegepast worden om genen met een differentiële genexpressie op te sporen. De reden hiertoe is dat er onbetrouwbare varianties gevormd worden

Analyse van *Xbp1* KO-microarray data in *cDC*

door het enorme aantal genen in combinatie met een laag aantal vrijheidsgraden (varianties berekend op basis van een laag aantal vrijheidsgraden kan onbetrouwbare varianties opleveren). Het is belangrijk dat er gewerkt wordt met betrouwbare varianties, want zeer kleine varianties leiden tot een grote T-waarde, wat op zijn beurt leidt tot vals-positieve resultaten. Op analoge manier leiden grote varianties tot vals-negatieve resultaten [37][40][45]. Vandaar dat in plaats van de standaard T-test de *moderated* T-test toegepast wordt.

3.9.3 Moderated T-test

Bij een *moderated* T-test wordt gewerkt met een “*shrunken*” variantie. Dit betekent dat de variantie van elk gen omgevormd wordt tot een kleinere variantie. Het tweede verschil ten opzichte van de standaard T-test is dat de *moderated* T-waarde overeenkomt met de \log_2 van de *fold change* van een gen gedeeld door zijn standaardfout. De bekomen *moderated* T-waarde wordt op dezelfde manier geïnterpreteerd als de standaard T-waarde (zie 3.9.2) [37][40].

3.9.4 P-waarde aanpassen volgens Benjamini-Hochberg

De p-waarde geeft de betrouwbaarheid van het bekomen resultaat aan. Praktisch houdt dat in dat de nulhypothese verworpen wordt wanneer de bekomen p-waarde kleiner is dan het significantieniveau. Wanneer men dus bijvoorbeeld werkt bij een significantieniveau van 5% en men bekomt een kleinere p-waarde dan dit significantieniveau, dan betekent dit dat men voor minstens 95% zeker is dat het bekomen resultaat geen toevalstreffer is [44]. Dit betekent dus ook dat er een kans van 5% is dat de nulhypothese ten onrechte werd verworpen. Stel dat men bijvoorbeeld beschikt over 200 genen en het significantieniveau wordt op 0,05 gesteld, dan betekent dit dat er 10 genen per toeval een p-waarde zullen opleveren kleiner dan het significantieniveau. Er zal dus met andere woorden sprake zijn van 10 vals-positieven. Dit betekent dus ook dat een dataset van 20 000 genen bij een significantieniveau van 0,05 1000 vals-positieven zal opleveren. Het probleem bij de p-waarden is dus dat deze niet geschikt zijn voor *multiple testing* [46].

Een mogelijkheid om dit probleem op te lossen, is door de zogenaamde *false discovery rate* (FDR) te controleren. De FDR is de proportie van de significante resultaten die eigenlijk vals-positief zijn. Stel dat men beschikt over een dataset met 100 000 genen en men aanvaardt dat 10% van de genen met differentiële genexpressie vals-positieven zijn, dat betekent dit dat de FDR op 10% gezet wordt. Het grote verschil tussen de p-waarde en de FDR is dat de p-waarde betrekking heeft op de volledige dataset, terwijl de FDR enkel betrekking heeft op de significante resultaten [46]. De FDR analoog van de p-waarde wordt de q-waarde of de *adjusted P-value* genoemd [37].

3.9.5 Samenstellen lijst genen met differentiële genexpressie

Nadat de q-waarde *threshold* is ingesteld, kan een lijst samengesteld worden van de upgereguleerde en downgereguleerde genen. Dit gebeurt eveneens op basis van de *fold change*. De *fold change* wordt gedefinieerd als de ratio van de waarde in conditie A en de waarde in conditie B. Stel dat de initiële waarde gelijk is aan 30 (conditie A) en de uiteindelijke waarde is gelijk aan 60 (conditie B), dan is er sprake van een *fold change* van twee (60/30). Meestal werkt men met het logaritme van de *fold change* en wordt er gewerkt met een *threshold* één. In strengere condities wordt de *cut off* op twee gelegd. Er worden dus twee selecties doorgevoerd: eerst worden de genen geselecteerd op basis van statistische significantie (via q-waarde) en daarna worden de genen ook nog geselecteerd op basis van het verschil in expressie (via *fold change*) [37].

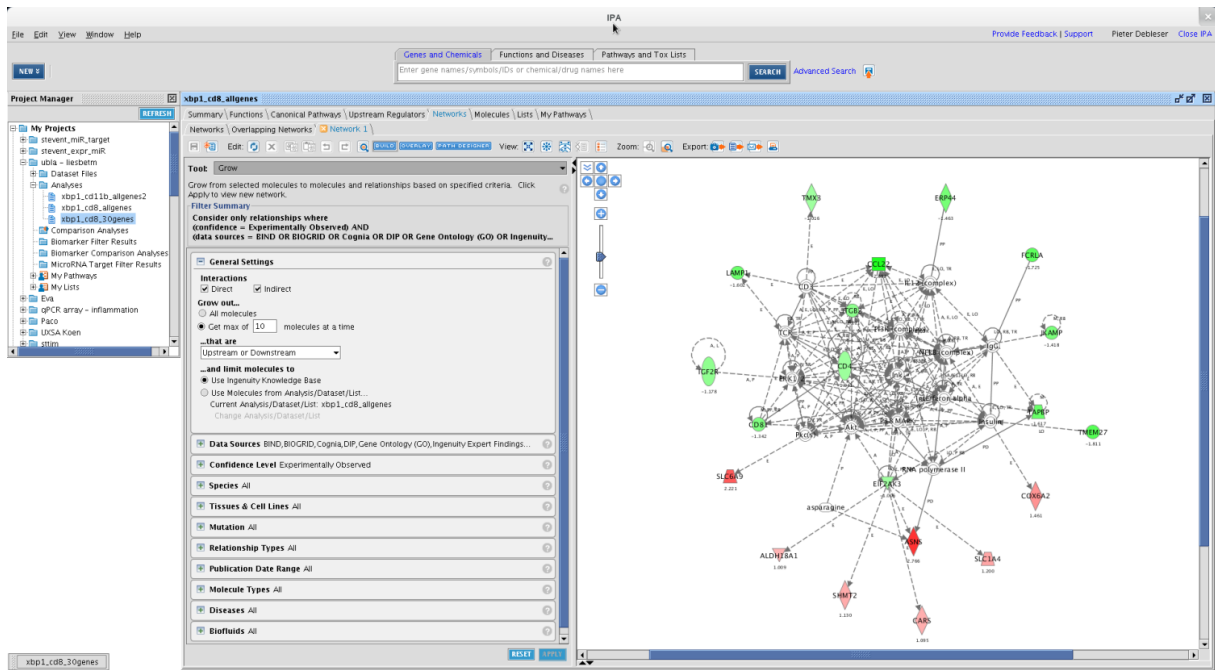
3.10 Annotatie van genen met differentiële genexpressie

Aangezien de uiteindelijk bekomen lijst van de genen met een differentiële genexpressie enkel een lijst is van de *probesetidentifiers*, is er ten slotte nog annotatie nodig. Annoteren betekent dat er biologische informatie gekoppeld wordt aan de data, hier dus aan de *probesetidentifier*. Op zich is men vrij welke biologische data er gekoppeld worden aan de *probesetidentifiers*, maar hetgeen steeds geannoteerd wordt, zijn de gennamen. Verder kan de annotatie ook nog uitgebreid worden met bijvoorbeeld de GO-data of de *Kyoto Encyclopedia of Genes and Genomes* (KEGG) data [37].

3.11 Pathway analyse

Eens de analyse in Bioconductor achter de rug is, beschikt men over de lijst van de genen die een differentiële genexpressie vertonen in de onderzochte situatie (bijvoorbeeld een KO) ten opzichte van de WT situatie. De volgende stap houdt in dat de functies van deze genen achterhaald worden. De functie van een bepaald eiwit achterhalen, komt eigenlijk neer op het achterhalen van de functie van het netwerk waarvan het betreffende eiwit deel uitmaakt. Een eiwit werkt immers nooit alleen, maar gaat steeds interacties aan met andere eiwitten en is dus op die manier werkzaam in een bepaald netwerk.

Ingenuity Pathway Analysis (IPA[®]) is momenteel één van de beste softwarepakketten om een antwoord te formuleren op bovenstaand bio-informatica probleem [47].



Figuur 3-9: Screenshot IPA®.

4 Resultaten

De data kunnen opgedeeld worden in twee datasets met elk zes microarrays:

CD8 α^+ DC dataset

- Hyb16266_(MoGene-1_0-s1-v1).CEL = WT 1
- Hyb16267_(MoGene-1_0-s1-v1).CEL = WT 2
- Hyb16268_(MoGene-1_0-s1-v1).CEL = WT 3
- Hyb16269_(MoGene-1_0-s1-v1).CEL = KO 1
- Hyb16270_(MoGene-1_0-s1-v1).CEL = KO 2
- Hyb16271_(MoGene-1_0-s1-v1).CEL = KO 3

CD11b $^+$ DC dataset

- Hyb16272_(MoGene-1_0-s1-v1).CEL = WT 1
- Hyb16273_(MoGene-1_0-s1-v1).CEL = WT 2
- Hyb16274_(MoGene-1_0-s1-v1).CEL = WT 3
- Hyb16275_(MoGene-1_0-s1-v1).CEL = KO 1
- Hyb16276_(MoGene-1_0-s1-v1).CEL = KO 2
- Hyb16277_(MoGene-1_0-s1-v1).CEL = KO 3

Eens het microarray experiment achter de rug is, wil men zo snel mogelijk de lijst bekomen van de genen met een differentiële genexpressie (via R, Bioconductor). Uiteraard wil men ook de functies van deze genen kennen en die kunnen achterhaald worden via de GO en de netwerkinteracties van de genen. Bij deze data werd de transcriptiefactor *Xbp1* uitgeschakeld. Daarom is het interessant via motiefanalyses na te gaan hoe de betreffende transcriptiefactor de genen met differentiële expressie reguleert.

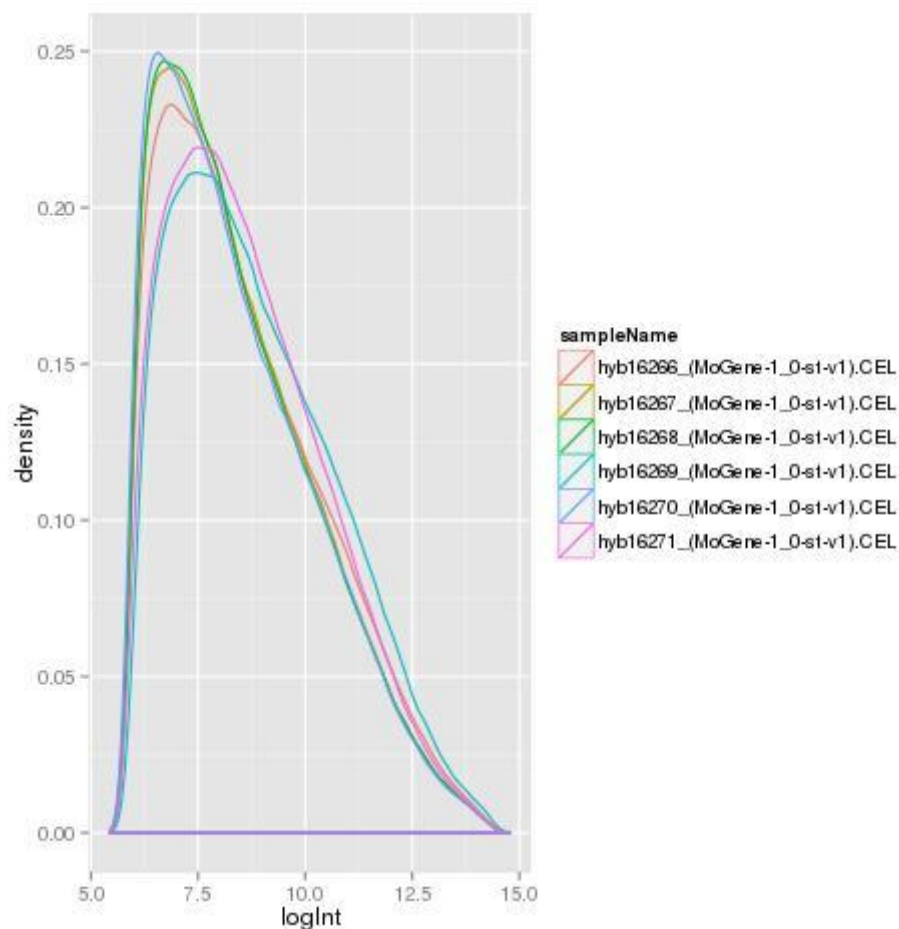
Het script voor de analyse in R van de CD8 α^+ DC dataset is terug te vinden in bijlage 2. Extra informatie over de verschillende microarrays is terug te vinden in bijlage 3.

4.1 Analyse microarray data via Bioconductor

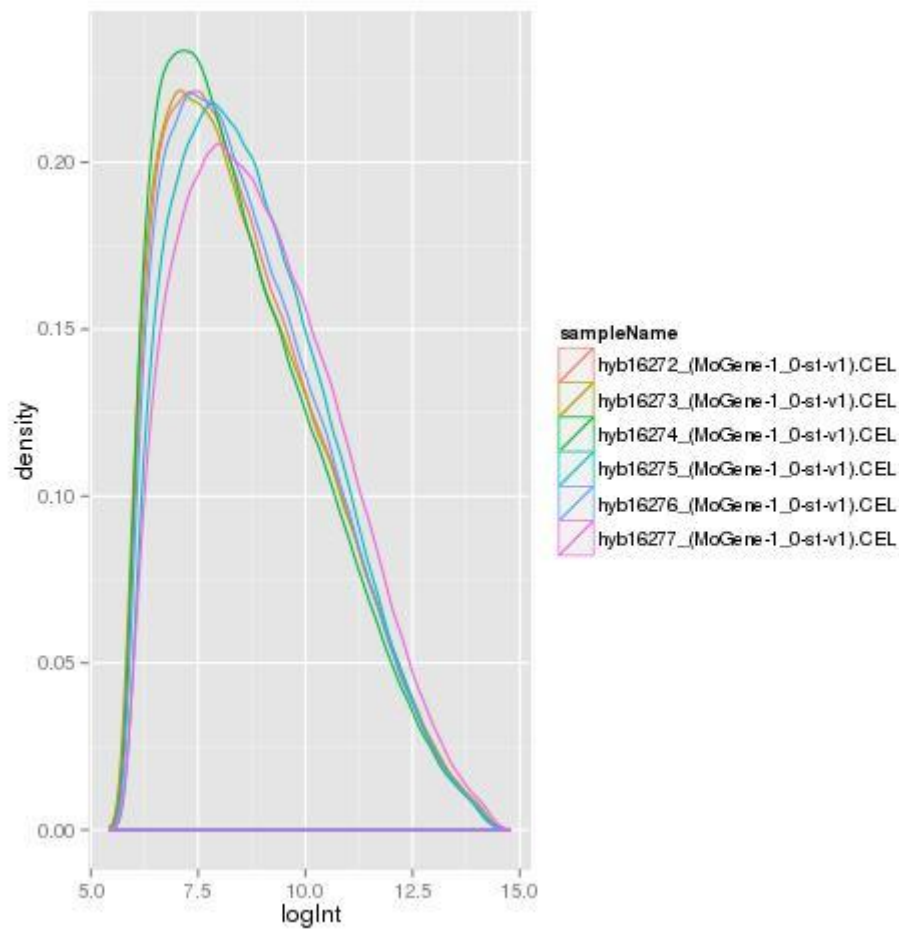
4.1.1 Controle kwaliteit ruwe data via plots

4.1.1.1 Histogram

Figuur 4-1 en figuur 4-2 tonen de histogrammen van de ruwe microarrays van respectievelijk de CD8 α^+ DC dataset en de CD11b $^+$ DC dataset. Bij de histogrammen van ruwe microarray data is het belangrijk dat de curves een gelijkaardige positie en vorm vertonen. Dit is voor beide datasets het geval, wat dus aanduidt dat de data van goede kwaliteit zijn.



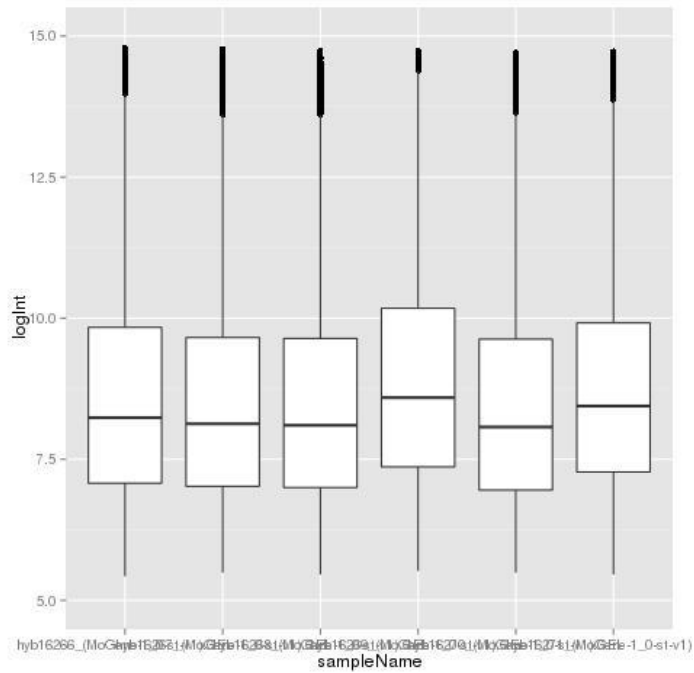
Figuur 4-1: De histogrammen van de ruwe microarrays van de CD8 α^+ DC dataset. De eerste drie “sampleNames” komen overeen met de WT-microarrays, terwijl de laatste drie “sampleNames” overeenkomen met de KO-microarrays.



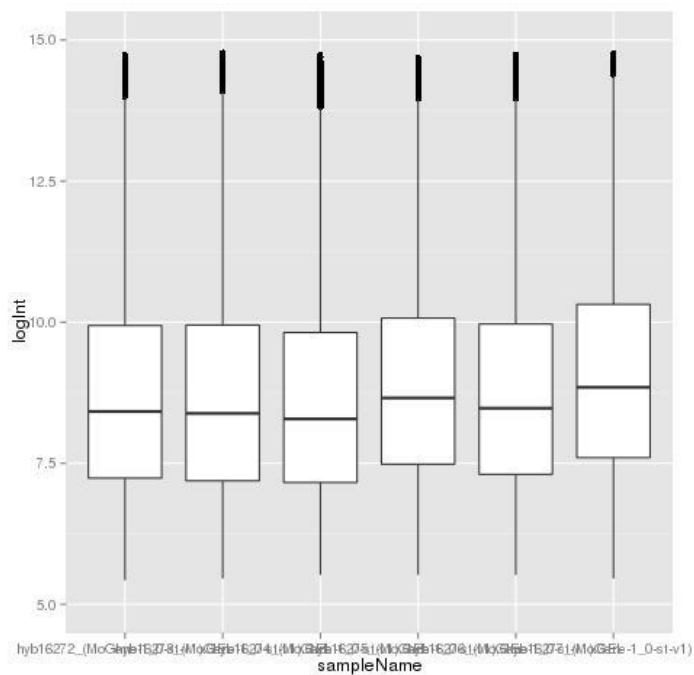
Figuur 4-2: De histogrammen van de ruwe microarrays van de CD11b⁺ DC dataset. De eerste drie “sampleNames” komen overeen met de WT-microarrays, terwijl de laatste drie “sampleNames” overeenkomen met de KO-microarrays.

4.1.1.2 Boxplot

Bij boxplots van ruwe microarrays is het belangrijk dat de verschillende boxplots een gelijkaardige vorm vertonen. Verder is het ook belangrijk dat er niet te veel verschillen aanwezig zijn tussen de medianen (middelste lijn) van de verschillende boxplots. Figuren 4-3 en 4-4 tonen de boxplots van de ruwe microarray data van respectievelijk de CD8α⁺ DC dataset en de CD11b⁺ DC dataset. Deze figuren duiden opnieuw aan dat de data van goede kwaliteit zijn.



Figuur 4-3: De boxplots van de ruwe microarray data van CD8 α^+ DC dataset. De eerste drie boxplots komen overeen met de WT-microarrays, terwijl de laatste drie boxplots overeen komen met de KO-microarrays.



Figuur 4-4: De boxplots van de ruwe microarray data van CD11b $^+$ DC dataset. De eerste drie boxplots komen overeen met de WT-microarrays, terwijl de laatste drie boxplots overeen komen met de KO-microarrays.

Analyse van *Xbp1* KO-microarray data in cDC

4.1.1.3 Andere plots

De kwaliteit van ruwe microarrays kan, naast via histogrammen en boxplots, ook nog gecontroleerd worden via volgende plots:

- *Chip pseudo-image* gebaseerd op *residuals*
- *Chip pseudo-image* gebaseerd op *weights*
- RLE-plot
- NUSE-plot

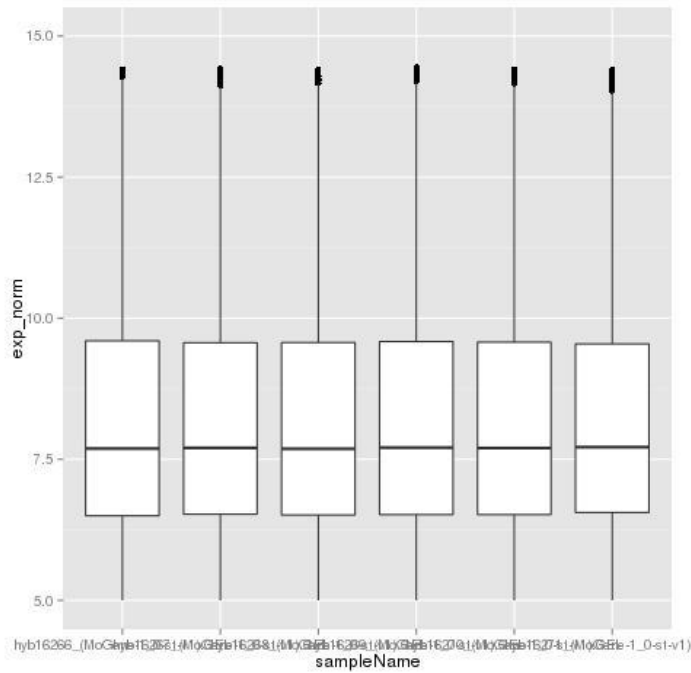
Deze plots zijn terug te vinden in bijlage 4 voor zowel de CD8 α^+ DC dataset als de CD11b $^+$ DC dataset.

4.1.2 Visualisatie normalisatie

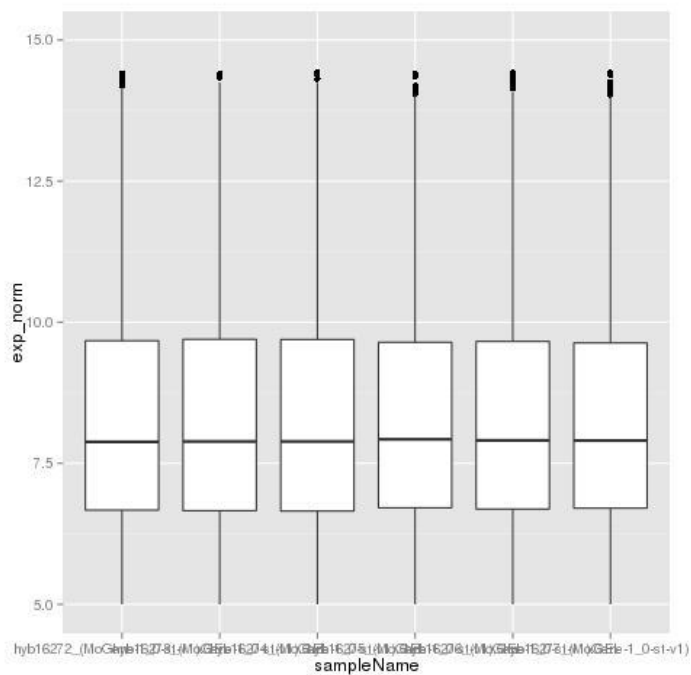
4.1.2.1 Boxplot

Normalisatie wordt toegepast om de verschillen tussen de microarrays te verwijderen. Eén van de beste manieren om dit te visualiseren is via boxplots. Boxplots van ruwe microarrays vertonen steeds kleine schommelingen in de mediaan. Na normalisatie zouden deze verschillen niet meer aanwezig mogen zijn en zouden dus de boxplots even groot moeten zijn en de verschillende medianen op één horizontale lijnen moeten liggen. Figuren 4-5 en 4-6 tonen de boxplots van de genormaliseerde microarrays van respectievelijk de CD8 α^+ DC dataset en de CD11b $^+$ DC dataset. Uit de figuren kan afgeleid worden dat de normalisatie goed is gelukt.

Analyse van *Xbp1* KO-microarray data in cDC



Figuur 4-5: De boxplots van de genormaliseerde microarray data van CD8 α^+ DC dataset. De eerste drie boxplots komen overeen met de WT-microarrays, terwijl de laatste drie boxplots overeen komen met de KO-microarrays.



Figuur 4-6: De boxplots van de genormaliseerde microarray data van CD11b $^+$ DC dataset. De eerste drie boxplots komen overeen met de WT-microarrays, terwijl de laatste drie boxplots overeen komen met de KO-microarrays.

Analyse van *Xbp1* KO-microarray data in *cDC*

4.1.3 Lijst van genen met differentiële genexpressie

4.1.3.1 *CD8α⁺* DC dataset

Uit de analyse van de microarrays van de *CD8α⁺* DC dataset volgt dat er in totaal 55 genen een differentiële genexpressie vertonen (p parameters= p -waarde < 0,05 en $\log_{2}FC < -1$ of > 1). Van die 55 genen zijn er 15 genen die een upregulatie vertonen en 30 genen die downregulatie vertonen. Tabel 4-1 geeft de top tien weer van de upgereguleerde en downgereguleerde genen van de *CD8α⁺* DC dataset.

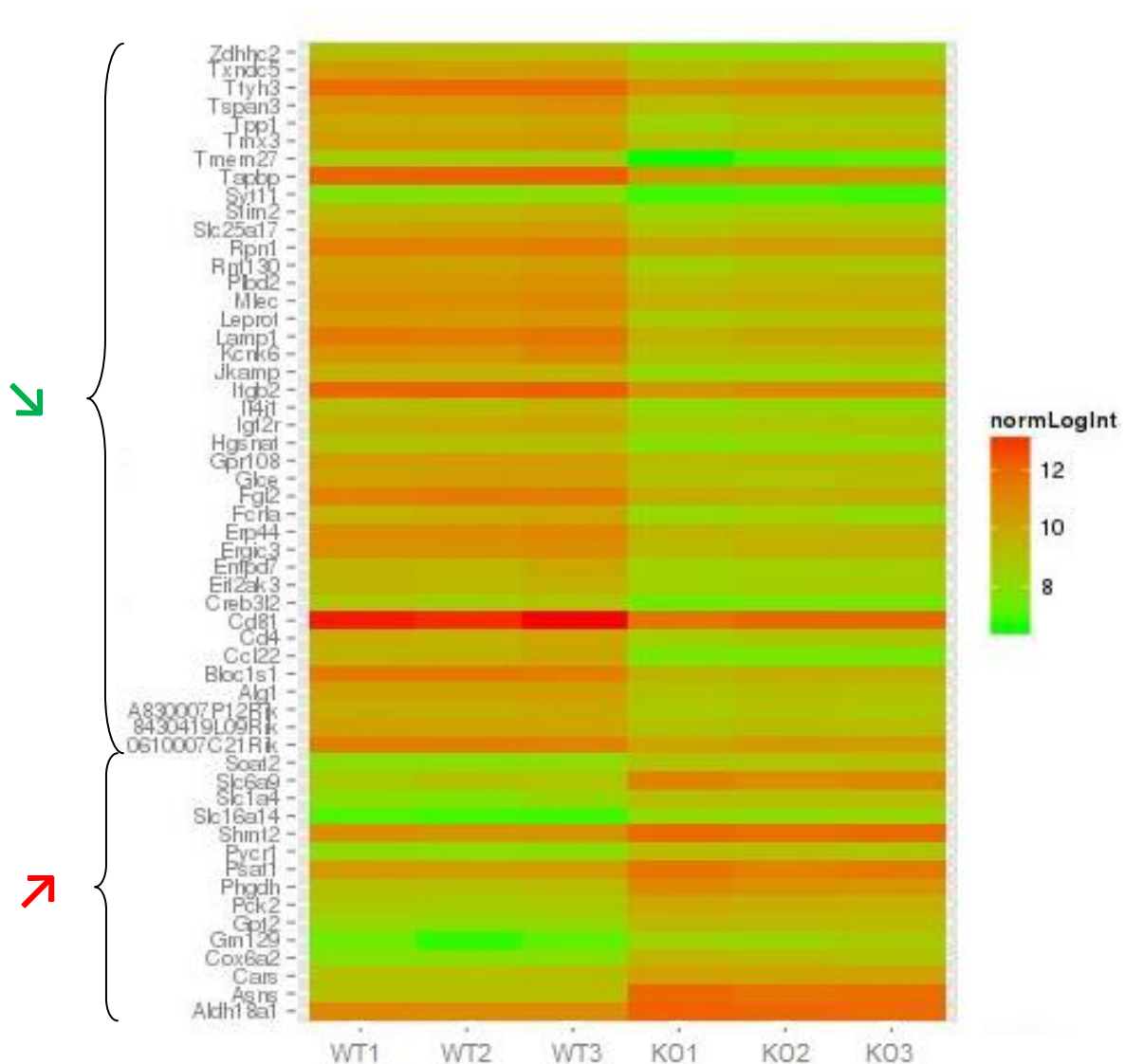
Tabel 4-1: De tien meest *upregulated* (↗) en *downregulated* (↘) genen in de *CD8α⁺* DC dataset.

Accessienummer	Gennaam	Log ₂ (fold change)
NM_012055	<i>Asns</i>	2.766231 ↗
NM_008135	<i>Slc6a9</i>	2.220586 ↗
NM_016966	<i>Phgdh</i>	1.696261 ↗
BC132471	<i>Gm129</i>	1.466857 ↗
NM_009943	<i>Cox6a2</i>	1.461136 ↗
NM_027921	<i>Slc16a14</i>	1.330624 ↗
NM_173866	<i>Gpt2</i>	1.254666 ↗
NM_144795	<i>Pycr1</i>	1.202696 ↗
NM_018861	<i>Slc1a4</i>	1.200493 ↗
NM_146064	<i>Soat2</i>	1.159205 ↗
NM_021540	<i>Rnf130</i>	-1.57019 ↘
NM_001033525	<i>Kcnk6</i>	-1.58496 ↘
NM_010684	<i>Lamp1</i>	-1.60165 ↘
NM_001025313	<i>Tapbp</i>	-1.61709 ↘
NM_175036	<i>Leprot</i>	-1.64292 ↘
NM_008013	<i>Fgl2</i>	-1.65727 ↘
NM_015740	<i>Bloc1s1</i>	-1.68447 ↘
NM_001160215	<i>Fcrla</i>	-1.72538 ↘
NM_020626	<i>Tmem27</i>	-1.81126 ↘
NM_009137	<i>Ccl22</i>	-2.38195 ↘

Analyse van *Xbp1* KO-microarray data in cDC

De volledige lijst van de genen die een differentiële genexpressie vertonen in de CD8 α^+ DC subset is terug te vinden in bijlage 5.

Figuur 4-7 stelt de heatmap van de genen met differentiële genexpressie van de CD8 α^+ DC subset.



Figuur 4-7: De heatmap van de genen met differentiële genexpressie van de CD8 α^+ DC dataset.

Onderstaande tabellen geven de GO-termen (muis) weer die overgerepresenteerd zijn in deze dataset. De GO-termen worden bekomen met behulp van een Perl-script dat gebruik maakt van de *E-utilities* van NCBI (*National Center for Biotechnology Information*).

Moleculaire functie

Tabel 4-2: Lijst van de meest voorkomende GO-termen in CD8 α^+ DC dataset (moleculaire functie).

GO-term	Aantal keer voorkomen
<i>transferase activity</i>	16
<i>protein binding</i>	11
<i>molecular_function</i>	10
<i>metal ion binding</i>	8
<i>nucleotide binding</i>	8
<i>ATP binding</i>	5
<i>glycoprotein binding</i>	5
<i>oxidoreductase activity</i>	5
<i>protein heterodimerization activity</i>	5

Biologisch proces

Tabel 4-3: Lijst van de meest voorkomende GO-termen in CD8 α^+ DC dataset (biologisch proces).

GO-term	Aantal keer voorkomen
<i>transport</i>	10
<i>biological_process</i>	8
<i>metabolic process</i>	8
<i>oxidation-reduction process</i>	8
<i>cell adhesion</i>	5
<i>cellular amino acid biosynthetic process</i>	5
<i>DNA-dependent</i>	4
<i>cell-cell adhesion</i>	4
<i>cellular extravasation</i>	4
<i>neutrophil chemotaxis</i>	4
<i>response to unfolded protein</i>	4

Cellulair compartiment



Tabel 4-4: Lijst van de meest voorkomende GO-termen in CD8 α^+ DC dataset (cellulair compartiment).

GO-term	Aantal keer voorkomen
<i>membrane</i>	38
<i>integral to membrane</i>	36
<i>mitochondrion</i>	14
<i>endoplasmic reticulum</i>	13
<i>plasma membrane</i>	12
<i>cell surface</i>	9
<i>cellular_component</i>	8
<i>integral to plasma membrane</i>	8
<i>lysosome</i>	8
<i>cytoplasm</i>	7
<i>endoplasmic reticulum membrane</i>	6
<i>extracellular region</i>	4
<i>mitochondrial inner membrane</i>	4
<i>multivesicular body</i>	4
<i>nucleus</i>	4

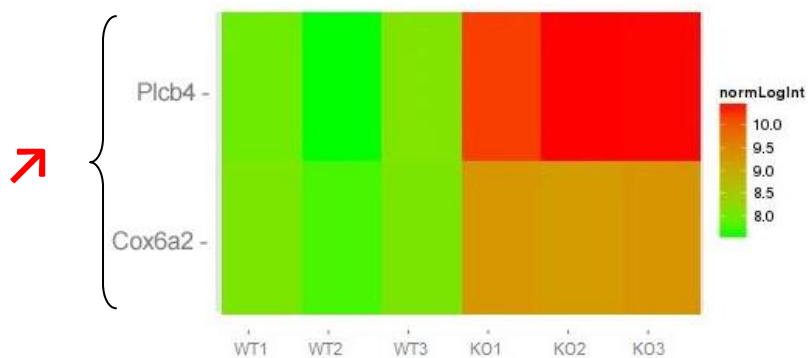
4.1.3.2 CD11b⁺ DC dataset

De CD11b⁺ DC dataset vertoont enkel twee upgereguleerde genen. Deze genen zijn weergegeven in tabel 4-5.

Tabel 4-5: Lijst van de genen die een differentiële genexpressie vertonen in de CD11b⁺ DC dataset.

Accessienummer	Gennaam	Log ₂ (fold change)
NM_013829	<i>Plcb4</i>	2.518736 
NM_009943	<i>Cox6a2</i>	1.372301 

De heatmap van de genen met differentiële genexpressie van de CD11b⁺ DC subset is voorgesteld in onderstaande figuur.



Figuur 4-8: De heatmap van de genen met differentiële genexpressie van de CD11b⁺ DC dataset.

De GO-termen (muis) van deze dataset zijn hieronder weergegeven.

Moleculaire functie

Tabel 4-6: Lijst van de meest voorkomende GO-termen in CD11b⁺ DC dataset (moleculaire functie).

GO-term	Aantal keer voorkomen
<i>Protein kinase binding</i>	1
<i>Protein binding</i>	1
<i>Signal transducer activity</i>	1
<i>Cytochrome-c oxidase activity</i>	1
<i>Mitogen-activated protein kinase binding</i>	1

Biologisch proces

Tabel 4-7: Lijst van de meest voorkomende GO-termen CD11b⁺ DC dataset (biologisch proces).

GO-term	Aantal keer voorkomen
<i>Signal transduction</i>	1
<i>Oxidation-reduction proces</i>	1
<i>Negative regulation of potassium ion transport</i>	1

Cellulair compartiment

Tabel 4-8: Lijst van de meest voorkomende GO-termen CD11b⁺ DC dataset (cellulair compartiment).

GO-term	Aantal keer voorkomen
<i>Smooth endoplasmic reticulum</i>	1
<i>Membrane</i>	1
<i>Cytoplasm</i>	1
<i>Dendrite</i>	1
<i>Nucleus</i>	1
<i>Postsynaptic density</i>	1
<i>Mitochondrial respiratory chain complex IV</i>	1
<i>Mitochondrion</i>	1

4.2 IPA[®]-analyse

De IPA[®]-analyse levert vier netwerken op voor de CD8α⁺ DC dataset en twee netwerken voor de CD11b⁺ dataset. Onderstaande tabellen geven een overzicht van deze netwerken met hun overeenkomstige score en geassocieerde functies.

CD8α⁺ DC dataset

Tabel 4-9: Overzicht van de netwerken waarin de genen met differentiële genexpressie van de CD8α⁺ DC dataset betrokken zijn. De geassocieerde functies met de verschillende netwerken zijn vermeld, alsook de score toegekend aan de netwerken. Omtrent de score geldt; hoe hoger de score, hoe meer de genen met differentiële genexpressie voorkomen in dit netwerk.

Netwerk	Geassocieerde netwerkfuncties	Score
1	Cel-cel signalisatie en interactie	45
	Hematologische systeemontwikkeling en hematologische functies	
	Inflammatoire respons	
2	Carbohydraatmetabolisme	29
	Vorming kleine moleculen	
	Orgaanmorfologie	
3	Lipidemetabolisme	27
	Moleculair transport	

	Vorming kleine moleculen	
4	Aminozuurmetabolisme	8
	Vorming kleine moleculen	
	Carbohydraatmetabolisme	

CD11b⁺ DC dataset

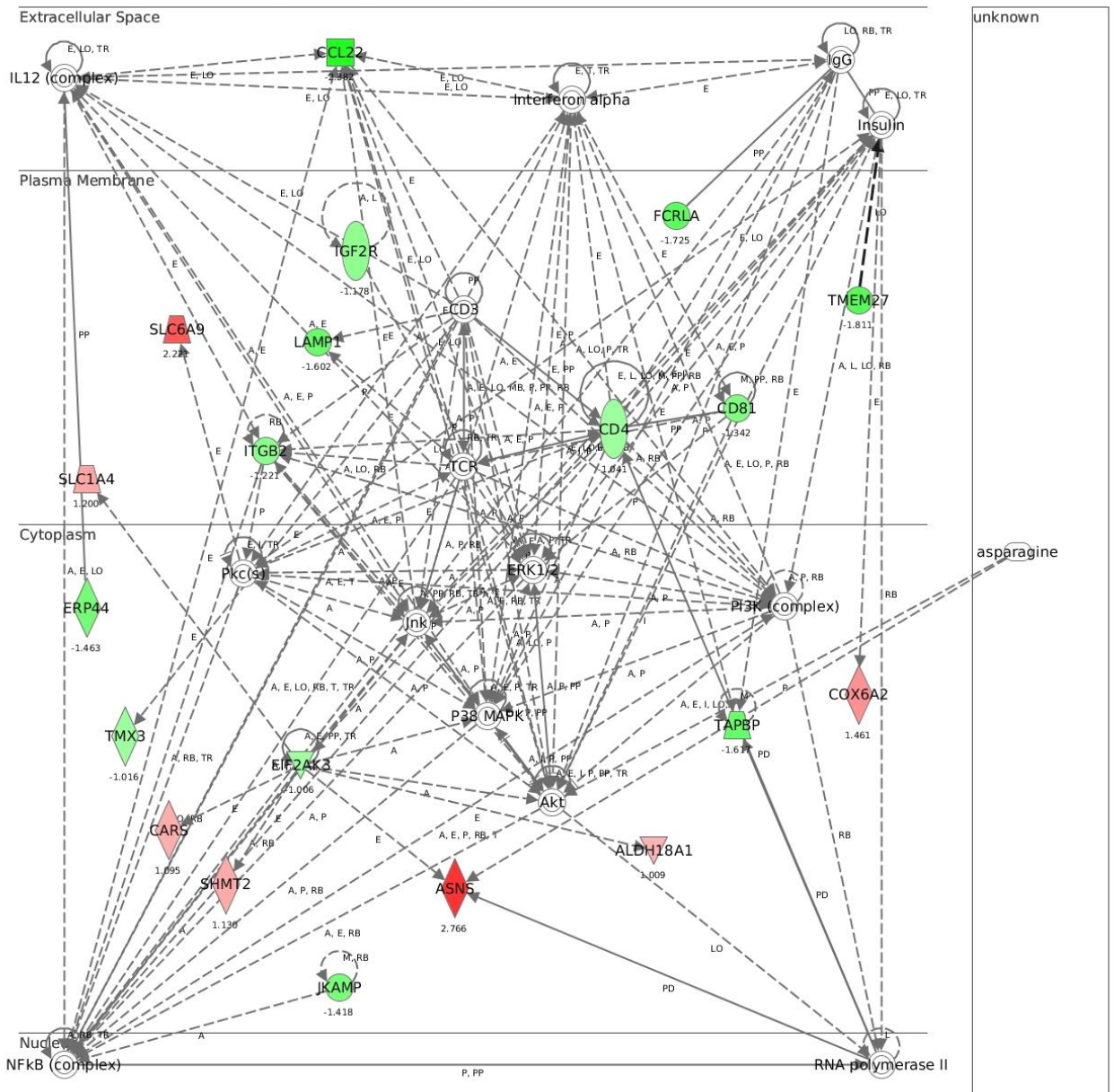
Tabel 4-10: Overzicht van de netwerken waarin de genen met differentiële genexpressie van de CD11b⁺ DC dataset betrokken zijn. De geassocieerde functies met de verschillende netwerken zijn vermeld, alsook de score toegekend aan de netwerken. Omtrent de score geldt; hoe hoger de score, hoe meer de genen met differentiële genexpressie voorkomen in dit netwerk.

Netwerk	Geassocieerde netwerkfuncties	Score
1	Cellulaire verzameling en organisatie	3
	Ontwikkeling zenuwstelsel	
	Weefselmorfologie	
2	Energieproductie	3
	Lipidemetabolisme	
	Vorming kleine moleculen	

Voor de CD11b⁺ DC dataset levert IPA[®] twee netwerken op met beide zeer lage scores. Dit komt overeen met de verwachtingen aangezien er slechts twee genen zijn die een differentiële genexpressie vertonen in CD11b⁺ DC met een *Xbp1* KO ten opzichte van CD11b⁺ DC met een WT fenotype. Hieruit kan dus afgeleid worden dat een *Xbp1* KO een veel grotere invloed heeft op de CD8α⁺ DC subset dan op de CD11b⁺ DC subset. Vandaar dat het verdere verloop van de analyse zich vooral toespitst op de CD8α⁺ DC dataset.

Figuur 4-9 stelt het netwerk voor met de hoogste score van de CD8α⁺ DC dataset. De downgereguleerde genen zijn met een groene kleur voorgesteld, terwijl de upgereguleerde genen met een rode kleur zijn voorgesteld. De moleculen zonder kleur komen overeen met de moleculen die eveneens deel uitmaken van het netwerk, maar die geen differentiële genexpressie vertonen.

Network 1 : xbp1_cd8_allgenes : genes_de_annotated_sorted.txt : xbp1_cd8_allgenes



© 2000-2013 Ingenuity Systems, Inc. All rights reserved.

Figuur 4-9: Netwerk 1 van de CD8 α^+ DC dataset. Dit netwerk is betrokken in (a) cel-cel signalisatie en interactie, (b) hematologische systeemontwikkeling en hematologische functies en (c) inflammatoire respons.

De overige drie netwerken van de CD8 α^+ DC dataset en de twee netwerken van de CD11b $^+$ DC dataset zijn terug te vinden in bijlage 6.

4.3 Motiefanalyse




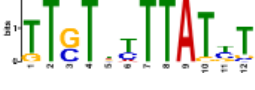
Aangezien Xbp1 een transcriptiefactor is, wordt verwacht het Xbp1-motief terug te vinden in de promotors van de differentieel geëxprimeerde genen. Dit zal echter enkel het geval zijn voor de genen die direct gereguleerd worden door Xbp1. Genen kunnen ook indirect gereguleerd worden

Analyse van Xbp1 KO-microarray data in cDC

door Xbp1, bijvoorbeeld via andere transcriptiefactoren die gereguleerd worden door Xbp1 of door interacties van eiwitten binnen een netwerk. Figuur 4-10 stelt het resultaat voor van de motiefanalyse via MEME Suite (TOMTOM) voor de downgereguleerde genen van de CD8 α^+ DC dataset. Figuur 4-11 stelt op zijn beurt de motiefanalyse voor van de upgereguleerde genen van de CD8 α^+ DC dataset.

Preview ?	Matches ?	List ?
	38	M00933 (Sp1) , MA0039.2 (Klf4) , MA0079.2 (SP1) , M01175 (CKROX) , M00649 (MAZ) , MA0283.1 (CHA4) , MA0338.1 (MIG2)
	6	MA0112.2 (ESR1) , MA0258.1 (ESR2) , M00191 (ER) , MA0066.1 (PPAR) , MA0159.1 (RXR::RAR_DR5)
	9	MA0212.1 (bcd) , MA0234.1 (oc) , MA0190.1 (Gsc) , M00623 (Crx) , M0402.1 (SWI5) , M01067 (Gfi1)
	11	MA0045.1 (HMG-I/Y) , M00772 (IRF) , MA0120.1 (id1) , MA0277.1 (AZF) , M01107 (RUSH-1alpha) , MA0109.1 (Hltf) , MA0152.1 (NFATC2) , MA00
	15	M00649 (MAZ) , M01122 (ZNF219) , MA0068.1 (Pax4) , M00444 (VDR) , MA0079.2 (SP1) , M01175 (CKROX) , M00933 (Sp1) , M00380 (Pax-4)

Figuur 4-10: Resultaat motiefanalyse via MEME Suite (TOMTOM) voor de downgereguleerde genen van de CD8 α^+ DC dataset (800 bp upstream).

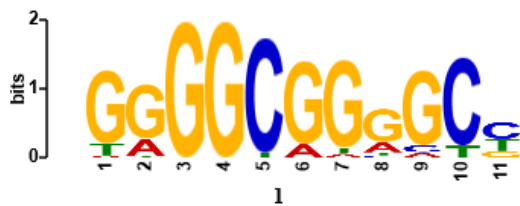
Preview ?	Matches ?	List ?
	36	MA0079.2 (SP1) , M01175 (CKROX) , M00933 (Sp1) , M00649 (MAZ) , MA034 , MA0056.1 (MZf1_1-4) , MA0341.1 (MSN2) , M00982 (KROX) , MA0039.2 (Klf4)
	4	M01169 (Ikaros) , M00419 (MEIS1) , M00777 (STAT) , MA0159.1 (RXR::RAR_D)
	17	M00526 (GCMF) , M01142 (LRH1) , MA0141.1 (Esrrb) , M01132 (SF1) , M0019 , M00979 (Pax-6) , MA0066.1 (PPARG) , MA0160.1 (NR4A2) , M00762 (PPAR)
	11	MA0151.1 (ARID3A) , MA0201.1 (Ptx1) , MA0190.1 (Gsc) , M00623 (Crx) , MA , M00512 (PPARgamma:RXRalpha) , MA0205.1 (Trl) , M01075 (PLZF)
	18	MA0042.1 (FOXI1) , M00203 (GATA-X) , MA0120.1 (id1) , MA0044.1 (HMG-1) , MA0045.1 (HMG-I/Y) , MA0127.1 (PEND) , M00129 (HFH1) , MA0082.1 (squam)

Figuur 4-11: Resultaat motiefanalyse via MEME Suite (TOMTOM) voor de upgereguleerde genen van de CD8 α^+ DC dataset (800 bp upstream).

Analyse van *Xbp1* KO-microarray data in *cDC*

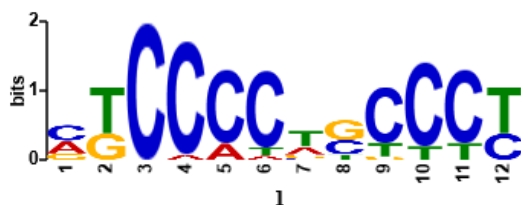
Het achterhalen van relevante motieven gebeurt op basis van het aantal matches en de *Expect values* (e-values) geassocieerd met de verschillende motieven. Daaruit volgt dat er één motief sterk aanwezig is in de promotors van de downgereguleerde en upgereguleerde genen.

Het motief dat voorkomt in de promotors van de downgereguleerde genen van de CD8 α^+ DC dataset, volgt een patroon zoals voorgesteld in figuur 4-12. Transcriptiefactoren die binden aan dit motief zijn Sp1, Klf4 en Ckrox (*Trans-acting transcription factor 1*, *Kruppel-like factor 4* en *Zinc finger and BTB domain containing 7B*).



Figuur 4-12: Het motief dat aanwezig is in de promotors van de downgereguleerde genen van de CD8 α^+ DC dataset. Sp1, Klf4 en Ckrox zijn voorbeelden van transcriptiefactoren die binden aan dit motief.

Figuur 4-13 stelt het motief voor dat voorkomt in de promotors van de upgereguleerde genen van de CD8 α^+ DC dataset. Transcriptiefactoren die gelinkt worden aan dit motief zijn Sp1, Ckrox en Klf4.

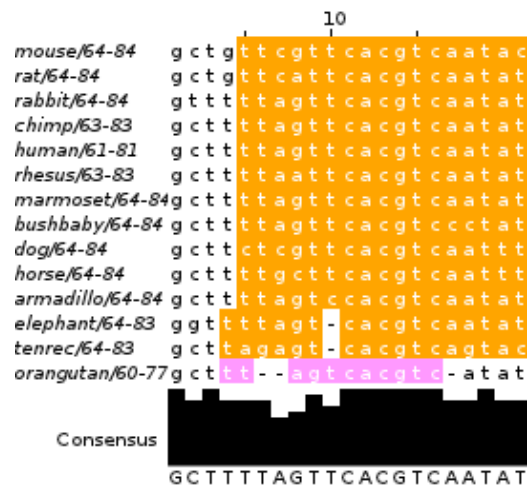


Figuur 4-13: Het motief dat aanwezig is in de promotors van de upgereguleerde genen van de CD8 α^+ DC dataset. Sp1, Ckrox en Klf4 zijn voorbeelden van transcriptiefactoren die binden aan dit motief. Dit motief is complementair aan het motief afgebeeld in figuur 4-12.

Het motief dat het sterkst aanwezig is in de promotors van de downgereguleerde en upgereguleerde genen is dus niet het *Xbp1*-motief, maar het motief voor transcriptiefactoren zoals Sp1, Klf4 en Ckrox (zie figuur 4-12 en figuur 4-13). Een mogelijke verklaring hiervoor is dat *Xbp1* de meeste genen met een differentiële genexpressie indirect reguleert. *Xbp1* reguleert dus met andere woorden transcriptiefactoren zoals Sp1, Klf4 en Ckrox, die op hun beurt de genen met differentiële genexpressie reguleren. Als deze hypothese klopt, dan moeten Sp1, Klf4 en Ckrox in de eerste plaats een *Xbp1*-motief vertonen in hun promotors. Dit wordt nagegaan via de tool Contra V2.

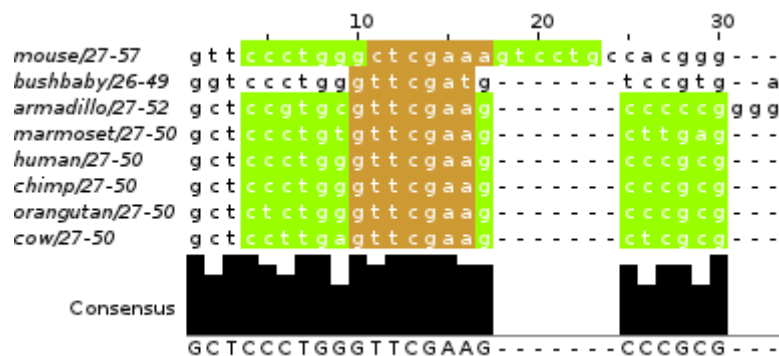
Analyse van *Xbp1* KO-microarray data in *cDC*

Figuur 4-14 stelt het resultaat voor van de analyse in Contra V2 van *Sp1*. Daaruit blijkt dat er een *Xbp1*-motief aanwezig is in de promotor van *Sp1*.



Figuur 4-14: Resultaat analyse *Xbp1*-motief in promotor *Sp1* (positie in muis = -545 tot -525 ten opzichte van transcriptiestartplaats). Volgende settings werden gebruikt voor Contra V2; species: muis, transcript: NM_013672, aantal basenparen upstream transcriptiestartplaats: 1000, stringency: core match = 0,95, matrix match = 0,85.

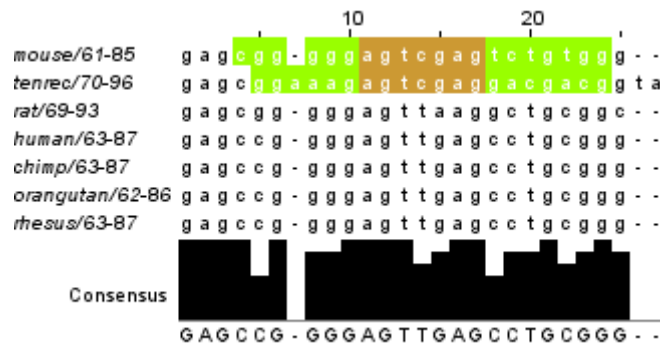
Er is eveneens een *Xbp1*-motief aanwezig in de promotor van *Klf4*. Dit wordt geïllustreerd in figuur 4-15.



Figuur 4-15: Resultaat analyse *Xbp1*-motief in promotor *Klf4* (positie in muis = -394 tot -365 ten opzichte van transcriptiestartplaats). Volgende settings werden gebruikt voor Contra V2; species: muis, transcript: NM_010637, aantal basenparen upstream transcriptiestartplaats: 1000, stringency: core match = 0,95, matrix match = 0,85.

Het resultaat van de analyse in Contra V2 van *Ckrox* is voorgesteld in figuur 4-16. Dit toont aan dat *Ckrox* (mogelijks) een *Xbp1*-motief bezit in zijn promotor.

Analyse van Xbp1 KO-microarray data in cDC



Figuur 4-16: Resultaat analyse Xbp1-motief in promotor *Ckrox* (positie in muis = -930 tot -907 ten opzichte van transcriptiestartplaats). Volgende settings werden gebruikt voor Contra V2; species: muis, transcript: NM_009565, aantal basenparen upstream transcriptiestartplaats: 1000, stringency: core match = 0,95, matrix match = 0,85.

4.3.1 Directe of indirecte regulering van Xbp1

Bovenstaande resultaten tonen aan dat de meeste genen met differentiële genexpressie indirect gereguleerd worden door Xbp1. Tegelijkertijd is het mogelijk dat sommige genen met differentiële genexpressie direct gereguleerd worden door Xbp1, ondanks dat de motiefanalyse via TOMTOM het Xbp1-motief niet oplevert. Om een beter inzicht te krijgen in welke genen met differentiële genexpressie direct gereguleerd worden door Xbp1 wordt een MEME Suite (FIMO) analyse uitgevoerd op de promotors van de genen met differentiële genexpressie. Om een FIMO-analyse te kunnen uitvoeren, heeft men twee bestanden nodig: een bestand met de motieven (PWMs) waarvan men de aanwezigheid wil onderzoeken (in MEME formaat) en een bestand met de sequenties die men wil onderzoeken (in fasta formaat). Er worden in totaal vijf PWMs voor Xbp1 (afkomstig van JASPAR en TRANSFAC) gebruikt en deze worden getest op de promotors (800 basenparen) en de 5' UTR (*Untranslated Region*) van alle 55 genen met differentiële genexpressie. Deze analyse geeft 27 genen terug (zie bijlage 7). Dit betekent dus dat er een Xbp1-motief aanwezig is in 27 van de 55 genen met differentiële genexpressie. Deze resultaten worden gecontroleerd via Contra V2 en daaruit blijkt dat bij 13 genen van die 27 genen het Xbp1-motief sterker geconserveerd aanwezig is in vergelijking met andere genen. Deze genen zijn in onderstaande tabel weergegeven. Bij 7 van die 13 genen is het Xbp1-motief heel sterk geconserveerd. Deze zeven genen zijn aangeduid met een oranje achtergrond in onderstaande tabel.

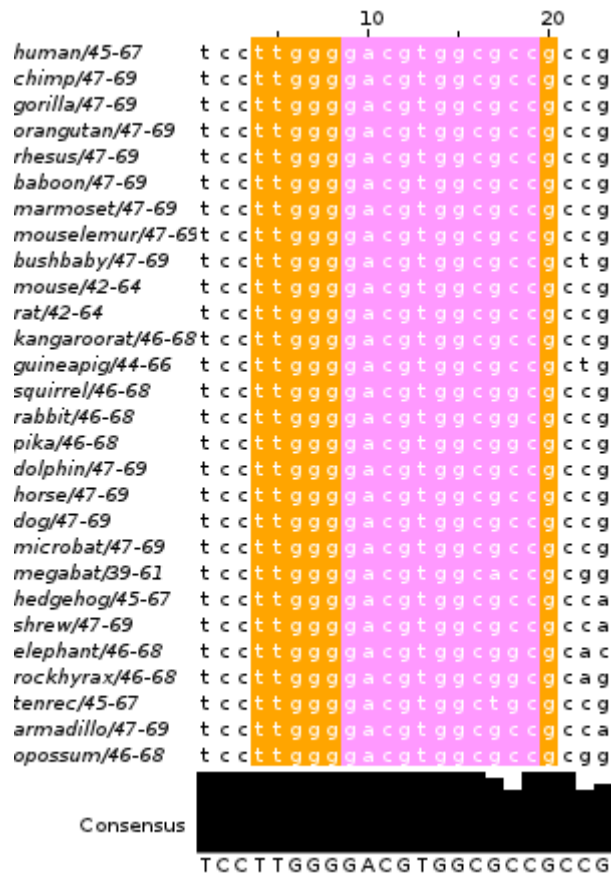
Analyse van *Xbp1* KO-microarray data in *cDC*

Tabel 4-11: Lijst van de genen waarin het *Xbp1*-motief sterker geconserveerd aanwezig is in vergelijking met de andere genen die de FIMO-analyse als resultaat geeft. Bij zeven genen is het *Xbp1*-motief heel sterk geconserveerd aanwezig. Deze genen zijn aangeduid met een oranje achtergrond. De genen zijn gerangschikt van meest downgereguleerd/upgereguleerd naar minst downgereguleerd/upgereguleerd.

Gennaam	Uppereguleerd of downgereguleerd
<i>Bloc1s1</i>	Down ↘
<i>Tapbp</i>	Down ↘
<i>Lamp1</i>	Down ↘
<i>Kcnk6</i>	Down ↘
<i>Erp44</i>	Down ↘
<i>Entpd7</i>	Down ↘
<i>Igf2r</i>	Down ↘
<i>Mlec</i>	Down ↘
<i>Syt11</i>	Down ↘
<i>Eif2ak3</i>	Down ↘
<i>Gm129</i>	Up ↗
<i>Psat1</i>	Up ↗
<i>Aldh18a1</i>	Up ↗

Het *Xbp1*-motief is heel sterk aanwezig in de promotoren van zeven genen. Dit is voorgesteld in figuur 4-17 met *Entpd7* als voorbeeld. Eén van deze zeven genen, *Eif2ak3* (*Eukaryotic translation initiation factor 2 alpha kinase 3*), is eveneens een gekend target van *Xbp1* [48].

Analyse van *Xbp1* KO-microarray data in *cDC*



Figuur 4-17: Resultaat analyse *Xbp1*-motief in promotor *Entpd7* (positie in mens = -37 tot -15 ten opzichte van transcriptiestartplaats). Volgende settings werden gebruikt voor Contra V2; species: mens, transcript: NM_020354, aantal basenparen upstream transcriptiestartplaats: 1000, stringency: core match = 0,95, matrix match = 0,85.

4.3.2 Verschil tussen CD8 α^+ DC en CD11b $^+$ DC in wild type condities

Wanneer het *Xbp1* gen uitgeschakeld wordt in een CD8 α^+ DC worden 55 genen daardoor beïnvloed. Wanneer het *Xbp1* gen uitgeschakeld wordt in een CD11b $^+$ DC worden slechts 2 genen beïnvloed (zie 4.1.3). Hieruit kan dus besloten worden dat de *Xbp1* een veel grotere invloed heeft op de CD8 α^+ DC dan op de CD11b $^+$ DC. Dit is opmerkelijk aangezien CD8 α^+ DC en CD11b $^+$ DC beide conventionele DC zijn. In een poging dit te verklaren, worden de WT CD8 α^+ microarrays vergeleken ten opzichte van de WT CD11b $^+$ microarrays.

In totaal zijn er 938 genen die een differentiële genexpressie vertonen. Er zijn 536 genen die een downregulatie vertonen, terwijl 402 genen een upregulatie vertonen. Wanneer de expressielevels van de 55 genen met differentiële genexpressie van de CD8 α^+ KO dataset vergeleken worden met hun expressielevels in deze dataset, zijn er 21 genen die upgereguleerd zijn in de ene dataset en downgereguleerd zijn in de andere dataset (of omgekeerd). Van deze 21 genen zijn er 8 genen

Analyse van *Xbp1* KO-microarray data in cDC

waarbij er sprake is van een groot verschil. Deze acht genen zijn met een oranje achtergrond aangeduid in onderstaande tabel.

Tabel 4-12: Lijst van de genen die enig verschil vertonen in expressie tussen dataset 1 (CD8 α^+ DC WT vs CD8 α^+ DC KO) en dataset 2 (CD8 α^+ DC WT vs CD11b $^+$ DC WT). Bij acht genen is er sprake van een groot verschil. Deze genen zijn weergegeven met een oranje achtergrond. Een positief getal in de kolom "Log₂(FC) dataset1" betekent dat het gen upgereguleerd is in de CD8 α^+ DC KO. Een positief getal in de kolom "Log₂(FC) dataset2" betekent dat het gen upgereguleerd is in een CD8 α^+ DC WT en downgereguleerd in een CD11b $^+$ DC WT.

Gennaam	Log ₂ (FC) dataset1	Log ₂ (FC) dataset2
<i>Slc16a14</i>	1.330624101	-1.76
<i>Soat2</i>	1.159204754	-0.401
<i>Gpr108</i>	-1.040152608	0.0165
<i>Txnbc5</i>	-1.065092167	0.936
<i>Rpn1</i>	-1.091861803	0.82
<i>Slc25a17</i>	-1.101252995	0.0838
<i>Mlec</i>	-1.118049586	1.44
<i>Zdhhc2</i>	-1.160888623	2.46
<i>Plbd2</i>	-1.167971072	0.319
<i>Igf2r</i>	-1.178135437	2.97
<i>Glce</i>	-1.266160318	1.96
<i>Ergic3</i>	-1.268226196	0.211
<i>Creb3l2</i>	-1.284850343	0.577
<i>Cd81</i>	-1.342455248	1.08
<i>Jkamp</i>	-1.417541688	0.00472
<i>Erp44</i>	-1.463048302	0.596
<i>Tapbp</i>	-1.617094129	0.14
<i>Leprot</i>	-1.642921404	0.398
<i>Fcrla</i>	-1.725381803	2.92
<i>Tmem27</i>	-1.811264643	3.19
<i>Ccl22</i>	-2.381954447	0.916

Een zeer interessant gen uit bovenstaande tabel is *Igf2r* (*Insulin-like Growth Factor 2 Receptor*). *Igf2r* is namelijk in een WT CD8 α^+ DC sterk upgereguleerd, terwijl het in een CD8 α^+ DC zonder *Xbp1* downgereguleerd is. Een mogelijke verklaring hiervoor is dat *Igf2r* direct gereguleerd wordt

Analyse van Xbp1 KO-microarray data in cDC

door Xbp1. Dit klopt met de resultaten uit paragraaf 4.3.1; *Igf2r* is één van de genen waarbij het Xbp1-motief heel sterk aanwezig is. Op analoge manier kunnen nog drie andere interessante genen teruggevonden worden (tabel 4-13).

Tabel 4-13: Lijst van genen die een Xbp1-motief in hun promotors bezitten en die een interessant verschil vertonen in expressie tussen dataset 1 (CD8 α^+ DC WT vs CD8 α^+ DC KO) en dataset 2 (CD8 α^+ DC WT vs CD11b $^+$ DC WT). De kolom "Conservatie" is bekomen via volgende settings voor Contra V2; species: mens, aantal basenparen upstream transcriptiestartplaats: 1000, stringency: core match = 0,95, matrix match = 0,85.

Gennaam	Log ₂ (FC) dataset1	Log ₂ (FC) dataset2	Xbp1-motief heel sterk geconserveerd?	Conservatie
<i>Igf2r</i>	-1.178135437	2.97	Ja	Meerdere malen geconserveerd in \pm 14 organismen
<i>Mlec</i>	-1.118049586	1.44	Ja	Geconserveerd in 30 organismen
<i>Erp44</i>	-1.463048302	0.596	Nee	Meerdere malen in \pm 7 organismen
<i>Tapbp</i>	-1.617094129	0.14	Nee	Geconserveerd in 9 organismen

5 Discussie

5.1 Functies genen met differentiële genexpressie

De uitschakeling van *Xbp1* beïnvloedt 55 genen in $CD8\alpha^+$ DC. Om inzicht te krijgen omtrent de functies van deze genen en dus de functies die beïnvloed worden door de KO, worden de GO-data van de 55 genen opgehaald en wordt achterhaald van welke netwerken de genen deel uit maken.

De GO-data van de 55 genen met differentiële genexpressie van de $CD8\alpha^+$ DC dataset komen overeen met de verwachtingen. Zo is transferase activiteit (*transferase activity*) de meest voorkomende moleculaire functie. Transferase activiteit heeft te maken met posttranslationele modificaties, denk maar aan bijvoorbeeld glycosyltransferasen die eiwitten voorzien van een suikerketen in het ER-lumen (glycosylatie) [20]. Transport is het meest voorkomend biologisch proces. Vele eiwitten moeten namelijk getransporteerd worden naar bijvoorbeeld het Golgi-apparaat, het lysosoom, het celoppervlak... De lijst van de meest voorkomende biologische processen bevat ook "*reaction to UPR*", wat uiteraard zeer logisch is aangezien *Xbp1* een transcriptiefactor is die upgereguleerd wordt tijdens de UPR. Ten slotte bevinden de meeste genen met differentiële genexpressie zich in het membraan, wat dus betekent dat een groot deel van de genen met differentiële genexpressie transmembraaneiwitten zijn. Het ER is de tweede cellulaire component waar veel genen met differentiële genexpressie voorkomen, wat perfect overeen komt met de verwachtingen.

De resultaten van de IPA[®]-analyse van de 55 genen met differentiële genexpressie van de $CD8\alpha^+$ DC dataset zijn goed, aangezien de bekomen netwerken hoge scores hebben en functies vertonen die gekoppeld zijn aan immunologie. Deze functies zijn echter zeer algemeen en geven dus geen duidelijkheid omtrent wat de gevolgen zijn van de *Xbp1* KO. De GO-data van enkele specifieke genen bieden wel een antwoord op die vraag. Zo vertoont *Ccl22* (*Chemokine (C-C motif) ligand 22*) functies in migratie en aantrekking van DC. *Tapbp* (*TAP binding protein*) reguleert MHC I moleculen en is dus betrokken bij antigeenpresentatie. *Ccl22* en *Tapbp* zijn beide sterk downgereguleerd in de *Xbp1* KO $CD8\alpha^+$ DC, wat dus betekent dat $CD8\alpha^+$ DC zonder *Xbp1* bijvoorbeeld minder goed in staat zijn te migreren naar de lymfeknoop en minder goed in staat zijn antigenen te presenteren aan naïeve T-cellen.

In vergelijking met de $CD8\alpha^+$ DC heeft de KO van *Xbp1* een veel kleinere invloed op de $CD11b^+$ DC. De uitschakeling van *Xbp1* beïnvloedt namelijk slechts twee genen in $CD11b^+$ DC. De IPA[®]-analyse van deze twee genen levert twee netwerken op, met telkens een score van drie. De verklaring

Analyse van *Xbp1* KO-microarray data in *cDC*

voor deze lage scores heeft te maken met hoe de netwerken gebouwd worden binnen IPA®. IPA® probeert de genen die werden opgegeven door de gebruiker zoveel mogelijk samen te plaatsen in een netwerk. Als er een cruciaal gen ontbreekt voor het netwerk wordt dit gen er zelf door IPA® bijgeplaatst. Hoe meer genen IPA® er zelf bij moet plaatsen, hoe lager de score. Aangezien er bij deze dataset slechts twee genen met differentiële genexpressie zijn, moet IPA® zelf veel genen toevoegen om een goed netwerk te bouwen, wat leidt tot lage scores. Uit de GO-data van deze twee genen kan er eveneens weinig afgeleid worden.

5.2 Motiefanalyse

Het motief dat het sterkst aanwezig is in de promotors van zowel de upgereguleerde genen als de downgereguleerde genen, is het motief waar transcriptiefactoren zoals Sp1, Klf4 en Ckrox binden (Sp1, Klf4 en Ckrox binden aan gelijkaardige motieven [49] [50]). Dit is vreemd, want op die manier is er geen onderscheid tussen de upgereguleerde genen en de downgereguleerde genen. Men zou namelijk verwachten dat de upgereguleerde genen en de downgereguleerde genen één of meerdere verschillende motieven vertonen. Op die manier zouden de upgereguleerde genen gereguleerd worden door bepaalde transcriptiefactoren en de downgereguleerde genen door andere transcriptiefactoren. De reden waarom een bepaald gen upgereguleerd wordt en niet downgereguleerd, zou dan dus kunnen te maken hebben met de transcriptiefactoren die al dan niet binden.

Een tweede mogelijkheid is dat naast het Sp1/Klf4/Ckrox-motief nog een tweede motief aanwezig is in de promotors van de upgereguleerde en downgereguleerde genen, waarbij dit tweede motief verschillend is bij de upgereguleerde en downgereguleerde genen. Dit is echter niet het geval.

Ter controle wordt de motiefanalyse ook eens uitgevoerd met behulp van FIMO (MEME Suite) en MotifLab. Deze tools leverden gelijkaardige resultaten op.

5.2.1 Regulatie van Sp1, Klf4 en Ckrox door Xbp1

De resultaten van de Contra V2 analyse van Sp1 en Klf4 voor een Xbp1-motief zijn overtuigend. Zowel *Sp1* als *Klf4* bezitten namelijk een Xbp1-motief in hun promotors dat op een goede afstand ligt ten opzichte van de transcriptiestartplaats en geconserveerd aanwezig is bij een goede strengheid. De Contra V2 resultaten van Ckrox voor een Xbp1-motief zijn minder overtuigend, want het bekomen motief ligt ver van transcriptiestartplaats en is zeer weinig geconserveerd (enkel muis en tenreks, zie figuur 4-16). Dit motief is echter wel nog aanwezig wanneer gewerkt

Analyse van *Xbp1* KO-microarray data in *cDC*

wordt met de hoogste strengheid, waardoor het mogelijk is dat *Ckrox* toch gereguleerd wordt door *Xbp1*.

De aanwezigheid van een *Xbp1*-motief in de promotors van *Sp1*, *Klf4* en *Ckrox* werd enkel getest via Contra V2. Dit zou zeker ook nog moeten getest worden via PhysBinder. PhysBinder is net zoals Contra V2 een tool waarmee bindingsplaatsen van transcriptiefactoren kunnen gevisualiseerd en gedetecteerd worden. Het grote verschil tussen PhysBinder en Contra V2 is dat PhysBinder ook de DNA-structuur in rekening brengt. PhysBinder gaat dus met andere woorden eerst na of de consensussequentie van de transcriptiefactor van interesse aanwezig is en bepaalt daarna ook nog de DNA-opvouwing. De manier waarop het DNA is opgevouwen, is namelijk van cruciaal belang opdat de transcriptiefactor kan binden aan die bepaalde DNA-conformatie of niet. Het grote nadeel van PhysBinder is dat het momenteel slechts een zestigtal verschillende transcriptiefactormodellen aanbiedt, waarvan het *Xbp1*-motief geen deel uitmaakt [51].

5.2.2 Directe of indirecte regulatie door *Xbp1*

De resultaten van de FIMO-analyse geven aan dat 27 genen van de 55 genen een *Xbp1*-motief bezitten in hun promotor. De p-waarden van deze 27 genen zijn allemaal lager dan 0,0001, wat zeer significant is (zie bijlage 7). Als extra controle worden de FIMO-resultaten gecontroleerd via Contra V2. Op die manier wordt een lijst van 13 genen bekomen waarbij het *Xbp1*-motief sterk aanwezig is. Bij 7 van deze 13 genen zijn de resultaten zeer overtuigend: het *Xbp1*-motief is zeer sterk geconserveerd of komt meerdere malen geconserveerd voor. Eén van deze zeven genen, *Eif2ak3*, is zelfs een gekend target van *Xbp1*. De overige zes genen kunnen dus nieuwe, tot nu toe ongekende, targets zijn van *Xbp1*. Meer zekerheid kan men daarover bekomen via *wet lab* experimenten.

De differentieel geëxprimeerde genen die geen *Xbp1*-motief vertonen, worden waarschijnlijk indirect beïnvloed door *Xbp1*. Mogelijks worden deze genen gereguleerd door andere transcriptiefactoren zoals *Sp1*, *Klf4* en *Ckrox* die op hun beurt gereguleerd worden door *Xbp1*. Een tweede mogelijkheid is dat deze genen beïnvloed zijn door de *Xbp1* KO omwille van netwerkinteracties. *Xbp1* kan bijvoorbeeld gen A direct reguleren en wanneer gen A een bepaalde interactie heeft met gen B, wordt gen B ook beïnvloed door een *Xbp1* KO. De expressie van genen wordt echter niet alleen gereguleerd door transcriptiefactoren of netwerkverbindingen, maar is afhankelijk van nog tal van andere factoren, zoals bijvoorbeeld methylatie, micro RNA's (miRNA's)... Daarom wordt de rol van miRNA's in de regulatie van de 55 genen met differentiële

Analyse van *Xbp1* KO-microarray data in cDC

genexpressie van de CD8 α^+ DC dataset onderzocht. Daarbij wordt verwacht meer miRNA-regulatie vast te stellen bij de genen die geen *Xbp1*-motief vertonen. Dit is niet het geval.

5.2.3 CD11b⁺ DC versus CD8 α^+ DC

Het is opmerkelijk dat de *Xbp1* KO een veel grotere invloed heeft op de CD8 α^+ DC dan op de CD11b⁺ DC. In een poging dit te verklaren, worden de WT CD8 α^+ microarrays vergeleken ten opzichte van de WT CD11b⁺ microarrays. Zo werd bijvoorbeeld verwacht dat de vele downgereguleerde genen in de KO CD8 α^+ DC een hogere expressie vertonen in een WT CD8 α^+ DC ten opzichte van een WT CD11b⁺ DC. Dit blijkt niet het geval te zijn.

Toch is deze analyse zeer interessant, zeker wanneer deze resultaten vergeleken worden met de genen die een *Xbp1*-motief vertonen. Op die manier kunnen vier interessante genen geselecteerd worden, aangezien deze een *Xbp1*-motief bezitten en hun expressies aantonen dat ze sterk beïnvloed worden door een *Xbp1* KO. Deze vier genen kunnen niet alleen nieuwe, tot nu toe ongekende, targets zijn van *Xbp1*, ze kunnen eveneens zeer interessante genen zijn om het verschil tussen de CD8 α^+ DC en de CD11b⁺ DC experimenteel verder uit te spitten.

5.2.4 Toekomstperspectieven

De vier genen *Igf2r*, *Mlec* (*Malectin*), *Erp44* (*Endoplasmic reticulum protein 44*) en *Tapbp* vertonen een hogere expressie in CD8 α^+ DC dan in CD11b⁺ DC en bezitten een *Xbp1*-motief (volgens MEME Suite FIMO en Contra V2). Deze vier genen zijn dus interessante genen om verder experimenteel te onderzoeken.

Momenteel zijn er ongeveer een vijftal *Xbp1*-motieven beschikbaar in de databanken TRANSFAC en JASPAR. Doordat het *Xbp1*-motief een kort motief is, is het moeilijk kwaliteitsvolle motieven te bekomen. Betere *Xbp1*-motieven zouden kunnen bekomen worden na een ChIP-sequencing experiment.

6 Besluit

De uitschakeling van *Xbp1* leidt tot minder functionele CD8 α^+ DC. Doordat *Xbp1* ontbreekt, kunnen CD8 α^+ DC bijvoorbeeld minder goed migreren naar de lymfeknoop en zijn ze minder goed in staat antigenen te presenteren aan naïeve T-cellen.

De uitschakeling van *Xbp1* heeft een grotere invloed op de CD8 α^+ DC dan op de CD11b $^+$ DC, wat opmerkelijk is aangezien CD8 α^+ DC en CD11b $^+$ DC beide conventionele DC zijn. In een poging dit te verklaren, worden WT CD8 α^+ microarrays vergeleken met WT CD11b $^+$ microarrays. De resultaten van die analyse verschaffen hierrond geen duidelijkheid.

Het motief van transcriptiefactoren zoals Sp1, Klf4 en CKrox is het sterkst aanwezig in de promotors van zowel de upgereguleerde als de downgereguleerde genen. Het is opmerkelijk dat er daarrond geen verschil is tussen de upgereguleerde en downgereguleerde genen.

Bij 13 van de 55 genen met differentiële genexpressie is het *Xbp1*-motief sterk aanwezig. Vier genen van deze dertien genen vertonen eveneens een groot verschil in expressie in CD8 α^+ DC en CD11b $^+$ DC. Deze vier genen zijn:

- *Igf2r*;
- *Mlec*;
- *Erp44*;
- *Tapbp*.

Bovenstaande genen kunnen niet alleen nieuwe, tot nu toe ongekende, targets zijn van *Xbp1*, ze kunnen eveneens zeer interessante genen zijn om het verschil tussen de CD8 α^+ DC en de CD11b $^+$ DC experimenteel verder uit te spitten.

Alle resultaten zijn gebaseerd op *in silico* onderzoek. Als men meer zekerheid wil omtrent de resultaten zijn *wet lab* experimenten noodzakelijk.

7 Literatuurlijst

- [1] “Immuunsysteem”. [Online]. Beschikbaar op: <http://www.vib.be/nl/mens-en-gezondheid/Pages/Immuunsysteem.aspx>. [Bezocht: 02-mrt-2013].
- [2] K. Van Oostveldt, *BLT_11 Immunologie en virologie. Partim: Immunologie theorie*. Brugge: Hogeschool West-Vlaanderen Departement Simon Stevin, 2012.
- [3] P. Parham, *The immune system*. New York en London: Garland Science, 2005.
- [4] L.-Y. Liou, A.L. Blasius, M.J. Welch, M. Colonna, M.B.A. Oldstone, E.I. Zuniga, “In vivo conversion of BM plasmacytoid DC into CD11b+ conventional DC during virus infection”, *The Journal of Immunology*, 2008.
- [5] N. N. Iwakoshi, M. Pypaert, en L. H. Glimcher, “The transcription factor XBP-1 is essential for the development and survival of dendritic cells”, *J Exp Med*, vol. 204, nr. 10, pp. 2267–2275, jan. 2007.
- [6] “[Frontiers in Bioscience 9, 2972-2988, September 1, 2004]”. [Online]. Beschikbaar op: <http://www.bioscience.org/2004/v9/af/1452/figures.htm>. [Bezocht: 17-mei-2013].
- [7] “B6: Role of interferon regulated factor (IRF) 4 for Th17 differentiation”. [Online]. Beschikbaar op: http://www.imt.uni-marburg.de/loewe/index0cc5.html?option=com_content&view=article&id=40&Itemid=26. [Bezocht: 17-mei-2013].
- [8] M. Merad, P. Sathe, J. Helft, J. Miller, en A. Mortha, “The dendritic cell lineage: ontogeny and function of dendritic cells and their subsets in the steady state and the inflamed setting”, *Annu. Rev. Immunol.*, vol. 31, pp. 563–604, mrt. 2013.
- [9] R. E. Shrimpton, M. Butler, A.-S. Morel, E. Eren, S. S. Hue, en M. A. Ritter, “CD205 (DEC-205): a recognition receptor for apoptotic and necrotic self”, *Mol. Immunol.*, vol. 46, nr. 6, pp. 1229–1239, mrt. 2009.
- [10] “XCR1 chemokine (C motif) receptor 1 [Homo sapiens (human)] - Gene - NCBI”. [Online]. Beschikbaar op: <http://www.ncbi.nlm.nih.gov/gene/2829>. [Bezocht: 17-mei-2013].
- [11] Y. Zhan en L. Wu, “Functional regulation of monocyte-derived dendritic cells by microRNAs”, *Protein Cell*, vol. 3, nr. 7, pp. 497–507, jul. 2012.
- [12] “Dendritic Cells”. [Online]. Beschikbaar op: <http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/D/DCs.html#id3400140648>. [Bezocht: 03-apr-2013].
- [13] P. M. Domínguez en C. Ardavín, “Differentiation and function of mouse monocyte-derived dendritic cells in steady state and inflammation”, *Immunol. Rev.*, vol. 234, nr. 1, pp. 90–104, mrt. 2010.
- [14] P. Dresing, S. Borkens, M. Kocur, S. Kropp, en S. Scheu, “A fluorescence reporter model defines ‘Tip-DCs’ as the cellular source of interferon β in murine listeriosis”, *PLoS ONE*, vol. 5, nr. 12, p. e15567, 2010.
- [15] N. V. Serbina, T. P. Salazar-Mather, C. A. Biron, W. A. Kuziel, en E. G. Pamer, “TNF/iNOS-producing dendritic cells mediate innate immune defense against bacterial infection”, *Immunity*, vol. 19, nr. 1, pp. 59–70, jul. 2003.
- [16] S. S. Cao en R. J. Kaufman, “Unfolded protein response”, *Current Biology*, vol. 22, nr. 16, pp. R622–R626, aug. 2012.
- [17] A. L. Shaffer, M. Shapiro-Shelef, N. N. Iwakoshi, A.-H. Lee, S.-B. Qian, H. Zhao, X. Yu, L. Yang, B. K. Tan, A. Rosenwald, E. M. Hurt, E. Petroulakis, N. Sonenberg, J. W. Yewdell, K. Calame, L. H. Glimcher, en L. M. Staudt, “XBP1, downstream of Blimp-1, expands the secretory apparatus and other organelles, and increases protein synthesis in plasma cell differentiation”, *Immunity*, vol. 21, nr. 1, pp. 81–93, jul. 2004.
- [18] J. S. Cox en P. Walter, “A novel mechanism for regulating activity of a transcription factor that controls the unfolded protein response”, *Cell*, vol. 87, nr. 3, pp. 391–404, nov. 1996.

- [19] E. Szegezdi, S. E. Logue, A. M. Gorman, en A. Samali, "Mediators of endoplasmic reticulum stress-induced apoptosis", *EMBO Rep*, vol. 7, nr. 9, pp. 880–885, sep. 2006.
- [20] M. Demeyere, *BLT_12 Moleculaire biologie. Partim: De genexpressie*. Brugge: Hogeschool West-Vlaanderen Departement Simon Stevin, 2012.
- [21] "MA0414.1 detailed information". [Online]. Beschikbaar op: http://jaspar.binf.ku.dk/cgi-bin/jaspar_db.pl?ID=MA0414.1&rm=present&collection=CORE. [Bezocht: 21-mei-2013].
- [22] A.-L. Barabási, N. Gulbahce, en J. Loscalzo, "Network medicine: a network-based approach to human disease", *Nat. Rev. Genet.*, vol. 12, nr. 1, pp. 56–68, jan. 2011.
- [23] F. Nollet, *BLT_17 Research and development. Partim: De nieuwste technologische ontwikkelingen*. Brugge: Hogeschool West-Vlaanderen Departement Simon Stevin, 2012.
- [24] "Microarrays Factsheet". [Online]. Beschikbaar op: <http://www.ncbi.nlm.nih.gov/About/primer/microarrays.html>. [Bezocht: 16-mrt-2013].
- [25] "Genome Resource Facility". [Online]. Beschikbaar op: <http://grf.lshtm.ac.uk/microarrayoverview.htm>. [Bezocht: 16-mrt-2013].
- [26] "File:Microarray-schema.jpg - Wikipedia, the free encyclopedia". [Online]. Beschikbaar op: <http://en.wikipedia.org/wiki/File:Microarray-schema.jpg>. [Bezocht: 16-mrt-2013].
- [27] D. A. Lashkari, J. L. DeRisi, J. H. McCusker, A. F. Namath, C. Gentile, S. Y. Hwang, P. O. Brown, en R. W. Davis, "Yeast microarrays for genome wide parallel genetic and gene expression analysis", *Proc. Natl. Acad. Sci. U.S.A.*, vol. 94, nr. 24, pp. 13057–13062, nov. 1997.
- [28] "The R Project for Statistical Computing". [Online]. Beschikbaar op: <http://www.r-project.org/>. [Bezocht: 20-apr-2013].
- [29] "Bioconductor - About". [Online]. Beschikbaar op: <http://www.bioconductor.org/about/>. [Bezocht: 20-apr-2013].
- [30] "R & Bioconductor - Manuals". [Online]. Beschikbaar op: http://manuals.bioinformatics.ucr.edu/home/R_BioCondManual#TOC-Introduction1. [Bezocht: 20-apr-2013].
- [31] S. Broos, P. Hulpiau, J. Galle, B. Hooghe, F. V. Roy, en P. D. Bleser, "ConTra v2: a tool to identify transcription factor binding sites across species, update 2011", *Nucl. Acids Res.*, mei 2011.
- [32] "Tomtom - Introduction". [Online]. Beschikbaar op: <http://meme.nbcr.net/meme/tomtom-intro.html>. [Bezocht: 20-apr-2013].
- [33] S. Gupta, J. A. Stamatoyannopoulos, T. L. Bailey, en W. S. Noble, "Quantifying similarity between motifs", *Genome Biol*, vol. 8, nr. 2, p. R24, 2007.
- [34] "FIMO - Introduction". [Online]. Beschikbaar op: <http://meme.nbcr.net/meme/fimo-intro.html>. [Bezocht: 20-apr-2013].
- [35] C. E. Grant, T. L. Bailey, en W. S. Noble, "FIMO: scanning for occurrences of a given motif", *Bioinformatics*, vol. 27, nr. 7, pp. 1017–1018, apr. 2011.
- [36] "Ingenuity IPA -Integrate and understand complex 'omics data". [Online]. Beschikbaar op: <http://www.ingenuity.com/products/ipa#/?tab=features>. [Bezocht: 24-mrt-2013].
- [37] J. Mathys, *Analysis of Affymetrix microarray data using Bioconductor. Tutorial: Comparing two groups of samples*. Bioinformatics training and service facility VIB, 2012.
- [38] "www.boxplot.nl - alles over de boxplot". [Online]. Beschikbaar op: <http://www.boxplot.nl/>. [Bezocht: 03-mrt-2013].
- [39] "File:MadynoBoxplot01.png - Wikimedia Commons". [Online]. Beschikbaar op: <http://commons.wikimedia.org/wiki/File:MadynoBoxplot01.png>. [Bezocht: 03-mrt-2013].
- [40] J. Mathys, *Analysis of Affymetrix microarray data using Bioconductor. Slide: Comparing two groups of samples*. Bioinformatics training and service facility VIB, 2012.
- [41] "Genevestigator Documentation". [Online]. Beschikbaar op: <https://www.genevestigator.com/userdocs/manual/qc.html>. [Bezocht: 03-mrt-2013].
- [42] "Index of /agolden/bioconductor/version1". [Online]. Beschikbaar op: <http://www3.it.nuigalway.ie/agolden/bioconductor/version1/>. [Bezocht: 31-mei-2013].

- [43] Onbekend, *RMAexplained*. www.plexdb.org.
- [44] P. Dedeckere, *BLT10_ Wetenschappelijke verslaggeving en communicatie. Partim: Statistiek en ICT II. Deel: Statistiek voor de laboratoriumtechnoloog II*. Brugge: Hogeschool West-Vlaanderen Departement Simon Stevin, 2011.
- [45] Onbekend, *Introduction to Empirical Bayes Models and Moderated t-test*. 2012.
- [46] "Handbook of Biological Statistics: Multiple comparisons". [Online]. Beschikbaar op: <http://udel.edu/~mcdonald/statmultcomp.html>. [Bezocht: 05-jun-2013].
- [47] "Ingenuity IPA -Integrate and understand complex 'omics data". [Online]. Beschikbaar op: <http://www.ingenuity.com/products/ipa>. [Bezocht: 05-jun-2013].
- [48] D. Acosta-Alvear, Y. Zhou, A. Blais, M. Tsikitis, N. H. Lents, C. Arias, C. J. Lennon, Y. Kluger, en B. D. Dynlacht, "XBP1 controls diverse cell type- and condition-specific transcriptional regulatory networks", *Mol. Cell*, vol. 27, nr. 1, pp. 53–66, jul. 2007.
- [49] [Online]. Beschikbaar op: <http://bioit.dnbr.ugent.be/contrav2/TRANSFAC20104/M01175.html>. [Bezocht: 03-mei-2013].
- [50] J. Kaczynski, T. Cook, en R. Urrutia, "Sp1- and Krüppel-like transcription factors", *Genome Biol.*, vol. 4, nr. 2, p. 206, 2003.
- [51] S. Broos, A. Soete, B. Hooghe, R. Moran, F. van Roy, en P. D. Bleser, "PhysBinder: improving the prediction of transcription factor binding sites by flexible inclusion of biophysical properties", *Nucl. Acids Res.*, apr. 2013.

8 Bijlagen

8.1 Bijlage 1

Uitgewerkt voorbeeld van quantile normalisatie.

ALGORITME

Probe	Microarray 1	Microarray 2	Microarray 3	Microarray 4
1	22	5	17	15
2	14	9	16	21
3	7	18	20	15
4	18	13	5	17
5	14	1	14	20
6	19	25	19	20

Average 22

Probe	Microarray 1	Microarray 2	Microarray 3	Microarray 4
1	22	5	17	15
2	14	9	16	22
3	7	18	22	15
4	18	13	5	17
5	14	1	14	20
6	19	22	19	20

Average 19

Probe	Microarray 1	Microarray 2	Microarray 3	Microarray 4
1	22	5	17	15
2	14	9	16	22
3	7	19	22	15
4	18	13	5	17
5	14	1	14	19
6	19	22	19	20

Average 17

Probe	Microarray 1	Microarray 2	Microarray 3	Microarray 4
1	22	5	17	15
2	14	9	16	22
3	7	19	22	15
4	17	17	5	17
5	14	1	14	19
6	19	22	19	17

Average 14

Analyse van Xbp1 KO-microarray data in cDC

Probe	Microarray 1	Microarray 2	Microarray 3	Microarray 4
1	22	5	17	15
2	14	14	14	22
3	7	19	22	15
4	17	17	5	14
5	14	1	14	19
6	19	22	19	17
	Average	12		

Probe	Microarray 1	Microarray 2	Microarray 3	Microarray 4
1	22	12	17	12
2	14	14	14	22
3	7	19	22	15
4	17	17	5	14
5	12	1	12	19
6	19	22	19	17
	Average	7		

Probe	Microarray 1	Microarray 2	Microarray 3	Microarray 4
1	22	12	17	12
2	14	14	14	22
3	7	19	22	7
4	17	17	7	14
5	12	7	12	19
6	19	22	19	17

RESULTAAT

Original data

Probe	Microarray 1	Microarray 2	Microarray 3	Microarray 4
1	22	5	17	15
2	14	9	16	21
3	7	18	20	15
4	18	13	5	17
5	14	1	14	20
6	19	25	19	20
average	15.7	11.8	15.2	18.0
st. dev.	5.2	8.8	5.4	2.7

Quantile normalization

Probe	Microarray 1	Microarray 2	Microarray 3	Microarray 4
1	22	12	17	12
2	14	14	14	22
3	7	19	22	7
4	17	17	7	14
5	12	7	12	19
6	19	22	19	17
average	15.2	15.2	15.2	15.2
st. dev.	5.3	5.3	5.3	5.3

8.2 Bijlage 2

R script voor analyse microarray data toegepast op de CD8 α ⁺ DC dataset.

```

library("Matrix")
library("rpart")
library("ggplot2")
library("affy")
library("affyPLM")
library("affydata")
library("simpleaffy")
library("ArrayExpress")
library("lattice")
library("limma")
library("Biobase")
library("genefilter")
library("GOstats")
library("GSEABase")
library("mogene10sttranscriptcluster.db")
library("annotate")

####DATA INLADEN####
data<-ReadAffy(cefile.path=~/.microarrayDataFabiola/exp1717_RawData/cd8")
ph<-pData(data)
ph[,2]<-c("wt", "wt", "wt", "ko", "ko", "ko")
colnames(ph)[2]<-"source"
names<-rownames(ph)

####KWALITEIT DATA CONTROLEREN####
####Histogram
###alle microarrays in 1 figuur
exp_pm<-pm(data)
sampleNames<-vector()
logInt<-vector()
for(i in 1:6)
{
  sampleNames<-c(sampleNames, rep(names[i],dim(exp_pm)[1]))
  logInt<-c(logInt, log2(exp_pm[,i]))
}
logData<-data.frame(sampleName=sampleNames, logInt=logInt)
#jpeg("~/microarrayDataFabiola/plots/cd8/histogram_rawData_1.jpeg")
histogram<-ggplot(logData, aes(logInt, colour=sampleName))
histogram+geom_density()
#dev.off()

###alle microarrays opgesplitst

```

Analyse van Xbp1 KO-microarray data in cDC

```

exp_pm<-pm(data)
sampleNames<-vector()
logInt<-vector()
for(i in 1:6)
{
  sampleNames<-c(sampleNames, rep(names[i],dim(exp_pm)[1]))
  logInt<-c(logInt, log2(exp_pm[,i]))
}
logData<-data.frame(sampleName=sampleNames, logInt=logInt)
#jpeg("~/microarrayDataFabiola/plots/cd8/histogram_rawData_2.jpeg")
histogram<-ggplot(logData, aes(logInt))
histogram+geom_histogram(binwidth=0.2,
aes(y=..density..) )+geom_density(colour="red")+facet_grid(.~sampleName)
#dev.off()

####Boxplot
exp_pm<-pm(data)
sampleNames<-vector()
logInt<-vector()
for(i in 1:6)
{
  sampleNames<-c(sampleNames, rep(names[i],dim(exp_pm)[1]))
  logInt<-c(logInt, log2(exp_pm[,i]))
}
logData<-data.frame(sampleName=sampleNames, logInt=logInt)
#jpeg("~/microarrayDataFabiola/plots/cd8/boxplot_rawData.jpeg")
boxplot<-ggplot(logData, aes(sampleName, logInt))
boxplot+geom_boxplot()+ylim(5,15)
#dev.off()

####Microarray picture
for(i in 1:6)
{
  #jpeg(paste("~/microarrayDataFabiola/plots/cd8/microarrayPicture_",i,".jpe
g", sep=""))
  image(data[,i], main=names[i])
  #dev.off()
}

####Chip pseudo-images based on weights
Pset<-fitPLM(data)
for(i in 1:6)
{
  #jpeg(paste("~/microarrayDataFabiola/plots/cd8/pseudoImage_weights_",i,".j
peg", sep=""))
  image(Pset, which=i, main=names[i], add.legend=TRUE)
  #dev.off()
}

####Chip pseudo-images based on residuals
for(i in 1:6)
{
  #jpeg(paste("~/microarrayDataFabiola/plots/cd8/pseudoImage_residuals_",i,"
.jpeg", sep=""))
  image(Pset, which=i, type="resids", main=names[i], add.legend=TRUE)
  #dev.off()
}

####RLE
#jpeg(paste("~/microarrayDataFabiola/plots/cd8/RLE.jpeg"))
RLE(Pset, main="RLE")
#dev.off()

####NUSE
#jpeg("~/microarrayDataFabiola/plots/cd8/NUSE.jpeg")
NUSE(Pset, main="NUSE")
#dev.off()

```

```
#####NORMALISATIE#####
dataRMA<-rma(data)
dataRMAMatrix<-exprs(dataRMA)

#####VISUALISATIE NORMALISATIE#####
#####Correlatie microarrays
eset<-exprs(dataRMA)
distance<-dist(t(eset), method="maximum")
clusters<-hclust(distance)
#jpeg(paste("~/microarrayDataFabiola/plots/cd8/samengangData.jpeg"))
plot(clusters)
#dev.off()

#####Plot raw data <-> bg-corrected log intensities
exp_pm<-pm(data)
sampleNames<-vector()
logInt<-vector()
exp_bgcorr<-pm(bg.correct(data, method="rma"))
corrLogInt<-vector()
for(i in 1:6)
{
  sampleNames<-c(sampleNames, rep(names[i],dim(exp_pm)[1]))
  logInt<-c(logInt, log2(exp_pm[,i]))
  corrLogInt<-c(corrLogInt, log2(exp_bgcorr[,i]))
}
corrData<-data.frame(sampleName=sampleNames, logInt=logInt,
corrLogInt=corrLogInt)
#jpeg("~/microarrayDataFabiola/plots/cd8/plot_rawData_bgCorrData.jpeg")
scatter<-ggplot(corrData, aes(logInt, corrLogInt))
scatter+geom_point()+facet_grid(.~sampleName)+geom_abline(intercept=0, slope=1,
colour="red")
#dev.off()

#####Boxplot normalized intensities
exp_norm<-vector()
sampleNames<-vector()
for(i in 1:6)
{
  sampleNames<-c(sampleNames, rep(names[i],dim(dataRMAMatrix)[1]))
  exp_norm<-c(exp_norm, dataRMAMatrix[,i])
}
normData<-data.frame(sampleName=sampleNames, exp_norm=exp_norm)
#jpeg("~/microarrayDataFabiola/plots/cd8/boxplot_normData.jpeg")
boxplot<-ggplot(normData, aes(sampleName, exp_norm))
boxplot+geom_boxplot()+ylim(5,15)
#dev.off()

#####Histogram normalized intensities
#####alle microarrays in 1 figuur
sampleNames<-vector()
normInt<-vector()
for(i in 1:6)
{
  sampleNames<-c(sampleNames, rep(names[i],dim(dataRMAMatrix)[1]))
  normInt<-c(normInt, dataRMAMatrix[,i])
}
normData<-data.frame(sampleName=sampleNames, normInt=normInt)
#jpeg("~/microarrayDataFabiola/plots/cd8/histogram_normData_1.jpeg")
histogram<-ggplot(normData, aes(normInt, colour=sampleName))
histogram+geom_density()
#dev.off()

#####alle microarrays opgesplitst
sampleNames<-vector()
normInt<-vector()
for(i in 1:6)
```

Analyse van *Xbp1* KO-microarray data in cDC

```
{
  sampleNames<-c(sampleNames, rep(names[i],dim(dataRMAmatrix)[1]))
  normInt<-c(normInt, dataRMAmatrix[,i])
}
normData<-data.frame(sampleName=sampleNames, normInt=normInt)
#jpeg("~/microarrayDataFabiola/plots/cd8/histogram_normData_2.jpeg")
histogram<-ggplot(normData, aes(normInt))
histogram+geom_histogram(binwidth=0.2,
aes(y=..density..))+geom_density(colour="red")+facet_grid(.~sampleName)
#dev.off()

####MAplot
for(i in 1:6)
{
  #jpeg(paste("~/microarrayDataFabiola/plots/cd8/MAplot_normData_",i,".jpeg"
, sep=""))
  MAplot(dataRMA, which=i)
  #dev.off()
}

####IDENTIFICATIE DE GENEN####
####Design matrix maken
groups<-ph$source
f<-factor(groups, levels=c("wt", "ko"))
design<-model.matrix(~f)
colnames(design)<-c("wt", "ko_vs_wt")

####Linear model fitten op data
data_fit<-lmFit(dataRMAmatrix, design)

####Moderate T-test met 'shrunken' se
data_fit_eb<-eBayes(data_fit)

####Visualisatie
##Q-Q plot conventional T-statistics
conventional_t<-data_fit$coef/data_fit$stdev.unscaled/data_fit$sigma
#jpeg("~/microarrayDataFabiola/plots/cd8/QQplot_conventionalT.jpeg")
qqt(conventional_t, df=data_fit$df.residual, main="Conventional
T")+abline(0,1,col="red")
#dev.off()

##Q-Q plot moderated T-statistics
#jpeg("~/microarrayDataFabiola/plots/cd8/QQplot_moderatedT.jpeg")
y<-data_fit_eb$t[,2]
qqt(y, df=data_fit_eb$df.residual+data_fit_eb$df.prior, main="Moderated
T")+abline(0,1,col="red")+abline(-1,1,lty=2)+abline(1,1,lty=2)
#dev.off()

##Volcanoplot
#jpeg("~/microarrayDataFabiola/plots/cd8/volcanoplot.jpeg")
volcanoplot(data_fit_eb, coef=2, highlight=15)+abline(v=-
1,col="red")+abline(v=1,col="red")
#dev.off()

####P-values aanpassen volgens Benjamini-Hochberg
options(digits=3)
tab<-topTable(data_fit_eb, coef=2, number=2000, adjust.method="BH")

####Identificatie upregulated en downregulated genes
topgenes<-tab[tab$adj.P.Val<0.05,]
genes_up<-topgenes[topgenes$logFC>1,]
genes_down<-topgenes[topgenes$logFC<-1,]

####ANNOTATIE DE GENEN####
####Probe ID linken aan genaam
##Upregulated genes
genes_up_id<-genes_up$ID
```

Analyse van *Xbp1* KO-microarray data in cDC

```

TAIR<-mogenel0sttranscriptclusterACCNUM
TAIRt<-toTable(TAIR)
genes_up_annotated<-matrix(ncol=2)
colnames(genes_up_annotated)<-colnames(TAIRt)
for (i in 1:length(genes_up_id))
{
  if (nrow(TAIRt[TAIRt$probe_id==genes_up_id[i], ])!=0)
  {
    genes_up_annotated<-
      rbind(genes_up_annotated,TAIRt[TAIRt$probe_id==genes_up_id[i], ])
  }
}

geneSymbols<-matrix(ncol=1)
colnames(geneSymbols)<-("Symbol")
for(i in 1:length(genes_up_id))
{
  geneSymbols<-rbind(geneSymbols,
                    getSYMBOL(genes_up_id[i],
                              "mogenel0sttranscriptcluster.db"))
}
genes_up_annotated_full<-cbind(genes_up_annotated, geneSymbols)
genes_up_annotated<-na.omit(genes_up_annotated_full)
#write.table(genes_up_annotated,
file="~/microarrayDataFabiola/results/cd8/genes_up.txt",
row.names=TRUE, col.names=TRUE)
sep="\t",

##Dubbele hits aanpassen
waarheidstabel<-duplicated(genes_up_annotated$Symbol)
names<-genes_up_annotated$Symbol[waarheidstabel]
for(i in 1:length(names))
{
  slice<-genes_up_annotated[genes_up_annotated$Symbol==names[1],]
expressies<-exprs(data)[rownames(slice),]
gemiddelde<-apply(expressies,1,mean)
gemiddelde_dataframe<-as.data.frame(gemiddelde)
gemiddelde_dataframe[,2]<-rownames(gemiddelde_dataframe)
colnames(gemiddelde_dataframe)[2]<- "id"
maximum<-max(gemiddelde)
id<-gemiddelde_dataframe[gemiddelde_dataframe$gemiddelde==maximum,]$id
w1<-genes_up_annotated$Symbol==names[1]
w2<-rownames(genes_up_annotated)==id
w3<-w1-w2
waarheidstabel2<-as.logical(w3)
genes_up_annotated$Symbol[waarheidstabel2]<-NA
}
genes_up_annotated<-na.omit(genes_up_annotated)

##Downregulated genes
genes_down_id<-genes_down$ID
TAIR<-mogenel0sttranscriptclusterACCNUM
TAIRt<-toTable(TAIR)
genes_down_annotated<-matrix(ncol=2)
colnames(genes_down_annotated)<-colnames(TAIRt)
for (i in 1:length(genes_down_id))
{
  if (nrow(TAIRt[TAIRt$probe_id==genes_down_id[i], ])!=0)
  {
    genes_down_annotated<-
      rbind(genes_down_annotated,TAIRt[TAIRt$probe_id==genes_down_id[i], ])
  }
}

geneSymbols<-matrix(ncol=1)
colnames(geneSymbols)<-("Symbol")
for(i in 1:length(genes_down_id))
{

```

```

        geneSymbols<-rbind(geneSymbols,
                           "mogene10sttranscriptcluster.db"))
    }
    genes_down_annotated_full<-cbind(genes_down_annotated, geneSymbols)
    genes_down_annotated<-na.omit(genes_down_annotated_full)
    #write.table(genes_down_annotated,
    file="~/microarrayDataFabiola/results/cd8/genes_down.txt",
    row.names=TRUE, col.names=TRUE)
    sep="\t",

#Dubbele hits aanpassen
waarheidstabel<-duplicated(genes_down_annotated$Symbol)
names<-genes_down_annotated$Symbol[waarheidstabel]
for(i in 1:length(names))
{
    slice<-genes_down_annotated[genes_down_annotated$Symbol==names[1],]
    expressies<-exprs(data)[rownames(slice),]
    gemiddelde<-apply(expressies,1,mean)
    gemiddelde_dataframe<-as.data.frame(gemiddelde)
    gemiddelde_dataframe[,2]<-rownames(gemiddelde_dataframe)
    colnames(gemiddelde_dataframe)[2]<-"id"
    maximum<-max(gemiddelde)
    id<-gemiddelde_dataframe[gemiddelde_dataframe$gemiddelde==maximum,]$id
    w1<-genes_down_annotated$Symbol==names[1]
    w2<-rownames(genes_down_annotated)==id
    w3<-w1-w2
    waarheidstabel2<-as.logical(w3)
    genes_down_annotated$Symbol[waarheidstabel2]<-NA
}
genes_down_annotated<-na.omit(genes_down_annotated)

####RESULTAAT
genes_de<-rbind(genes_up, genes_down)
genes_de_annotated<-rbind(genes_up_annotated, genes_down_annotated)
genes_de_annotated<-cbind(genes_de_annotated, 1)
colnames(genes_de_annotated)[4]<-"logFC"
for(i in 1:dim(genes_de_annotated)[1])
{
    logFC<-genes_de$logFC[genes_de$ID==genes_de_annotated$probe_id[i]]
    genes_de_annotated$logFC[i]<-logFC
}
genes_de_annotated_sorted<-genes_de_annotated[order(genes_de_annotated$logFC,
decreasing=TRUE),]
#write.table(genes_de_annotated_sorted,
file="~/microarrayDataFabiola/results/cd8/genes_de_annotated_sorted.txt",
sep="\t\t", row.names=TRUE, col.names=TRUE)
genes_up_annotated_sorted<-
genes_de_annotated_sorted[genes_de_annotated_sorted$logFC>0,]
genes_down_annotated_sorted<-
genes_de_annotated_sorted[genes_de_annotated_sorted$logFC<0,]

####HEATMAP UPREGULATED EN DOWNREGULATED GENES####
dataRMAMatrix<-dataRMAMatrix[unlist(genes_de_annotated_sorted$probe_id),]
namesMicroarray<-c("WT1", "WT2", "WT3", "KO1", "KO2", "KO3")
sampleNames<-vector()
featureNames<-rep(genes_de_annotated_sorted$Symbol, 6)
normLogInt<-vector()
for(i in 1:6)
{
    sampleNames=c(sampleNames,
rep(namesMicroarray[i],dim(genes_de_annotated_sorted)[1]))
normLogInt=c(normLogInt, dataRMAMatrix[,i])
}
heatData=data.frame(normLogInt=normLogInt, sampleName=sampleNames,
featureName=featureNames)
#jpeg("~/microarrayDataFabiola/plots/cd8/heatmap_allGenes.jpeg")
heatmap<-ggplot(heatData, aes(sampleName, featureName))

```


Analyse van Xbp1 KO-microarray data in cDC

```
heatmap+geom_tile(aes(fill=normLogInt))+scale_fill_gradient(low="green",
high="red")
#dev.off()

-----
EXTRA1 = selecting DE genes only on adj.P.Val
-----

topgenes<-tab[tab$adj.P.Val<0.01,]
#Annotation
topgenes_id<-topgenes$ID
TAIR<-mogene10sttranscriptclusterACCNUM
TAIRt<-toTable(TAIR)
topgenes_annotated<-matrix(ncol=2)
colnames(topgenes_annotated)<-colnames(TAIRt)
for (i in 1:length(topgenes_id))
{
  if (nrow(TAIRt[TAIRt$probe_id==topgenes_id[i], ])!=0)
  {
    topgenes_annotated<-
    rbind(topgenes_annotated,TAIRt[TAIRt$probe_id==topgenes_id[i], ])
  }
}

geneSymbols<-matrix(ncol=1)
colnames(geneSymbols)<-("Symbol")
for(i in 1:length(topgenes_id))
{
  geneSymbols<-rbind(geneSymbols,
                    getSYMBOL(topgenes_id[i],
                    "mogene10sttranscriptcluster.db"))
}
topgenes_annotated_full<-cbind(topgenes_annotated, geneSymbols)
topgenes_annotated<-na.omit(topgenes_annotated_full)
#Dubbele hits aanpassen
waarheidstabel<-duplicated(topgenes_annotated$Symbol)
names<-topgenes_annotated$Symbol[waarheidstabel]
for(i in 1:length(names))
{
  slice<-topgenes_annotated[topgenes_annotated$Symbol==names[1],]
  expressies<-exprs(data)[rownames(slice),]
  gemiddelde<-apply(expressies,1,mean)
  gemiddelde_dataframe<-as.data.frame(gemiddelde)
  gemiddelde_dataframe[,2]<-rownames(gemiddelde_dataframe)
  colnames(gemiddelde_dataframe)[2]<- "id"
  maximum<-max(gemiddelde)
  id<-gemiddelde_dataframe[gemiddelde_dataframe$gemiddelde==maximum,]$id
  w1<-topgenes_annotated$Symbol==names[1]
  w2<-rownames(topgenes_annotated)==id
  w3<-w1-w2
  waarheidstabel2<-as.logical(w3)
  topgenes_annotated$Symbol[waarheidstabel2]<-NA
}
topgenes_annotated<-na.omit(topgenes_annotated)
#Result
topgenes_annotated<-cbind(topgenes_annotated, 1)
colnames(topgenes_annotated)[4]<- "logFC"
for(i in 1:dim(topgenes)[1])
{
  logFC<-topgenes$logFC[topgenes$ID==topgenes_annotated$probe_id[i]]
  topgenes_annotated$logFC[i]<-logFC
}
topgenes_annotated_sorted<-topgenes_annotated[order(topgenes_annotated$logFC,
decreasing=TRUE),]
write.table(topgenes_annotated_sorted,
file="~/microarrayDataFabiola/results/cd8/genes_pVal_0_01.txt",
row.names=TRUE, col.names=TRUE)
sep="\t",
```

```
-----
EXTRA2 = getting logFC and adj.P.Val for a certain gene
-----
```

```
tab<-topTable(data_fit_eb,          coef=2,          number=dim(data_fit_eb)[1],
adjust.method="BH")
TAIR<-mogene10sttranscriptclusterACCNUM
TAIRt<-toTable(TAIR)
tab_annotated<-matrix(ncol=2)
colnames(tab_annotated)<-colnames(TAIRt)
tab_id<-tab$ID
for (i in 1:length(tab_id))
{
  if (nrow(TAIRt[TAIRt$probe_id==tab_id[i], ])!=0)
  {
    tab_annotated<-rbind(tab_annotated,TAIRt[TAIRt$probe_id==tab_id[i],
])
  }
}
tab_annotated<-tab_annotated[-1,]

geneSymbols<-matrix(ncol=1)
colnames(geneSymbols)<-("Symbol")
tab_id<-tab_annotated$probe_id
for(i in 1:length(tab_id))
{
  geneSymbols<-rbind(geneSymbols,          getSYMBOL(tab_id[i],
"mogene10sttranscriptcluster.db"))
}
geneSymbols<-geneSymbols[-1,]
tab_annotated_full<-cbind(tab_annotated, geneSymbols)
tab_annotated<-na.omit(tab_annotated_full)

gene_refseqID<-"NM_013470"
temp<-tab_annotated[tab_annotated$accession==gene_refseqID,]
gene_probeID<-temp$probe_id
tab[tab$ID==gene_probeID,]

#check if gene_probeID and gene_refseqID is correct
getSYMBOL(gene_probeID, "mogene10sttranscriptcluster.db")
```

```
-----
EXTRA3 = getting logFC and adj.P.Val for a list of genes
-----
```

```
tab<-topTable(data_fit_eb,          coef=2,          number=dim(data_fit_eb)[1],
adjust.method="BH")
TAIR<-mogene10sttranscriptclusterACCNUM
TAIRt<-toTable(TAIR)
tab_annotated<-matrix(ncol=2)
colnames(tab_annotated)<-colnames(TAIRt)
for (i in 1:length(tab_id))
{
  if (nrow(TAIRt[TAIRt$probe_id==tab_id[i], ])!=0)
  {
    tab_annotated<-rbind(tab_annotated,TAIRt[TAIRt$probe_id==tab_id[i],
])
  }
}
tab_annotated<-tab_annotated[-1,]

geneSymbols<-matrix(ncol=1)
colnames(geneSymbols)<-("Symbol")
tab_id<-tab_annotated$probe_id
for(i in 1:length(tab_id))
{
```

```
geneSymbols<-rbind(geneSymbols,
  "mogene10sttranscriptcluster.db"))
}
geneSymbols<-geneSymbols[-1,]
tab_annotated_full<-cbind(tab_annotated, geneSymbols)
tab_annotated<-na.omit(tab_annotated_full)

list_ids<-read.table(file="~/list_ids.txt", sep="\n", header=FALSE)
hit<-vector()
hits_result<-vector()
symbols<-vector()
for(i in 1:dim(list_ids)[1])
{
  gene_refseqID<-list_ids[i,]
  temp<-tab_annotated[tab_annotated$accession==gene_refseqID,]
  if(!length(temp$probe_id)==0)
  {
    gene_probeID<-temp$probe_id
    symbols<-c(symbols,
      "mogene10sttranscriptcluster.db"))
    for(j in 1:length(gene_probeID))
    {
      hit<-tab[tab$ID==gene_probeID[j],]
      hits_result<-rbind(hits_result, hit)
    }
  }
}
end_result<-cbind(hits_result, symbols)
write.table(end_result, file="~/end_result.txt", sep="\t", col.names=TRUE,
row.names=TRUE)
```

8.3 Bijlage 3

Extra informatie over de microarrays.

<p>Nucleomics Core VIB Herestraat 49, Box 816 3000 LEUVEN, Belgium</p>				<p>Contact info: ✉ nucleomics@vib.be ☎ +32/16/373126 ☎ +32/16/373129 🌐 www.nucleomics.be</p>			
<h2>Experiment 1717: XBPI deficient</h2>							
<p>Description: Mo Gene 1.0ST</p>		<p>Experiment Date: 2012-12-24 15:21:42</p>					
<p>Organism: Mm (Mouse)</p>		<p>Costs:</p>					
<p>Project Leader: Fabiola Osorio - [NA]</p>							
<p>Scientist: Fabiola Osorio - [NA]</p>							
<p>Comments: Ambion start 50 ng</p>							
Hyb Id	Label	Sample information Cust.No	Description	Slide	Design	Status	Date
16266	Lab.Frag ssDNA	1	FO 1-CD8 DC WT 1		Affy-Mo-Gener1.0ST	OK	2013-01-30 14:35:51
16267	Lab.Frag ssDNA	2	FO 2-CD8 DC WT 2		Affy-Mo-Gener1.0ST	OK	2013-01-30 14:35:56
16268	Lab.Frag ssDNA	3	FO 3-CD8 DC WT 3		Affy-Mo-Gener1.0ST	OK	2013-01-30 14:36:02
16269	Lab.Frag ssDNA	13	FO 4bis-CD8 DC KO		Affy-Mo-Gener1.0ST	OK	2013-01-30 14:36:09
16270	Lab.Frag ssDNA	5	FO 5-CD8 DC KO 2		Affy-Mo-Gener1.0ST	OK	2013-01-30 14:36:13
16271	Lab.Frag ssDNA	6	FO 6-CD8 DC KO 3		Affy-Mo-Gener1.0ST	OK	2013-01-30 14:36:17
							Continued on next page

Experiment 1717: XBPI deficient							
<p>Description: Mo Gene 1.0ST Experiment Date: 2012-12-24 15:21:42</p> <p>Organism: Mm (Mouse) Costs:</p> <p>Project Leader: Fabiola Osorio - [NA] Scientist: Fabiola Osorio - [NA]</p> <p>Comments: Ambion start 50 ng</p>							
Hyb Id	Label	Sample information		Slide	Design	Status	Date
		Cust.No	Description				
16272	Lab.Frag.ssDNA	7	FO 7-CD11B DC WT 1		Affy-Mo-Gene1.0ST	OK	2013-01-30 14:36:21
16273	Lab.Frag.ssDNA	8	FO 8-CD11B DC WT 2		Affy-Mo-Gene1.0ST	OK	2013-01-30 14:36:25
16274	Lab.Frag.ssDNA	9	FO 9-CD11B DC WT 3		Affy-Mo-Gene1.0ST	OK	2013-01-30 14:36:29
16275	Lab.Frag.ssDNA	14	FO 10bis-CD11B DC KO		Affy-Mo-Gene1.0ST	OK	2013-01-30 14:36:33
16276	Lab.Frag.ssDNA	11	FO 11-CD8 DC KO 2		Affy-Mo-Gene1.0ST	OK	2013-01-30 14:36:37
16277	Lab.Frag.ssDNA	12	FO 12-CD8 DC KO 3		Affy-Mo-Gene1.0ST	OK	2013-01-30 14:38:12

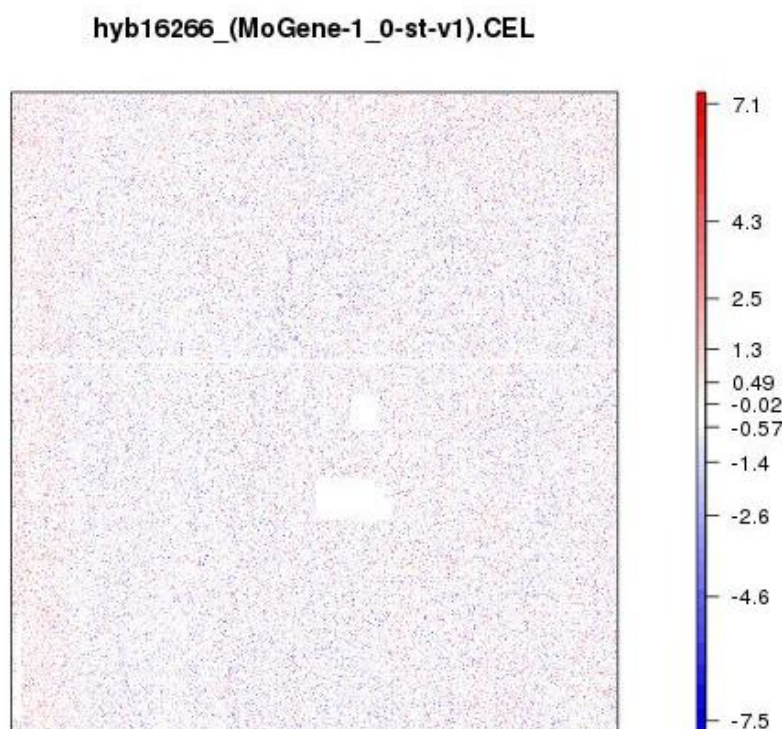
8.4 Bijlage 4

Controle van de kwaliteit van ruwe microarray data via chip pseudo-image gebaseerd op residuals, chip pseudo-image gebaseerd op weights, RLE-plot en NUSE-plot.

8.4.1 CD8 α^+ DC dataset

8.4.1.1 Chip pseudo-image - residuals

Opmerking: een *chip pseudo-image* gebaseerd op *residuals* wordt voor elke microarray apart gemaakt. Hier is als voorbeeld enkel de *chip pseudo-image* van de eerste microarray weergegeven. De overige vijf zijn terug te vinden op de CD-ROM.

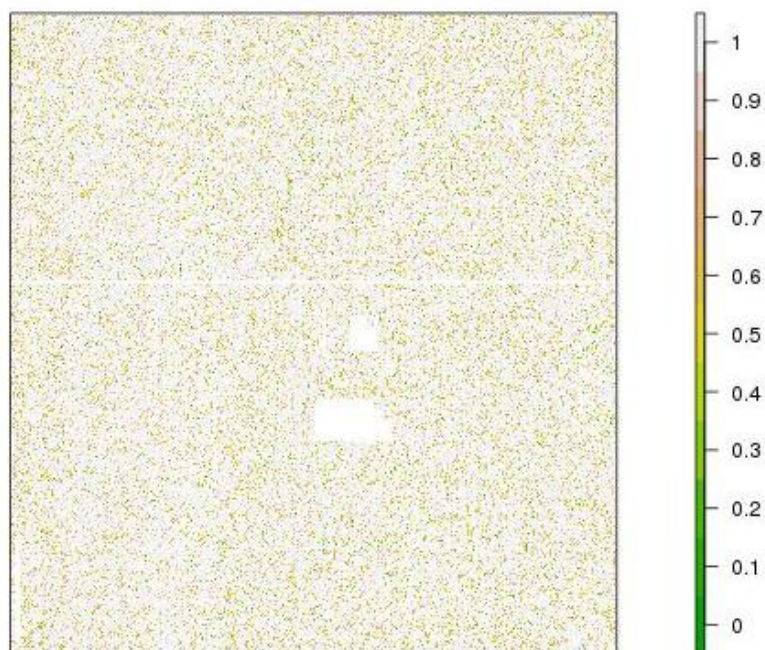


Figuur 8-1: *Chip pseudo-image* gebaseerd op *residuals* van microarray 1 van de CD8 α^+ DC dataset. Deze microarray komt overeen met een WT-microarray. Er zijn geen zones aanwezig die meer blauw of rood gekleurd zijn, waaruit kan afgeleid worden dat er geen artefacten aanwezig zijn in deze microarray.

8.4.1.2 Chip pseudo-image - weights

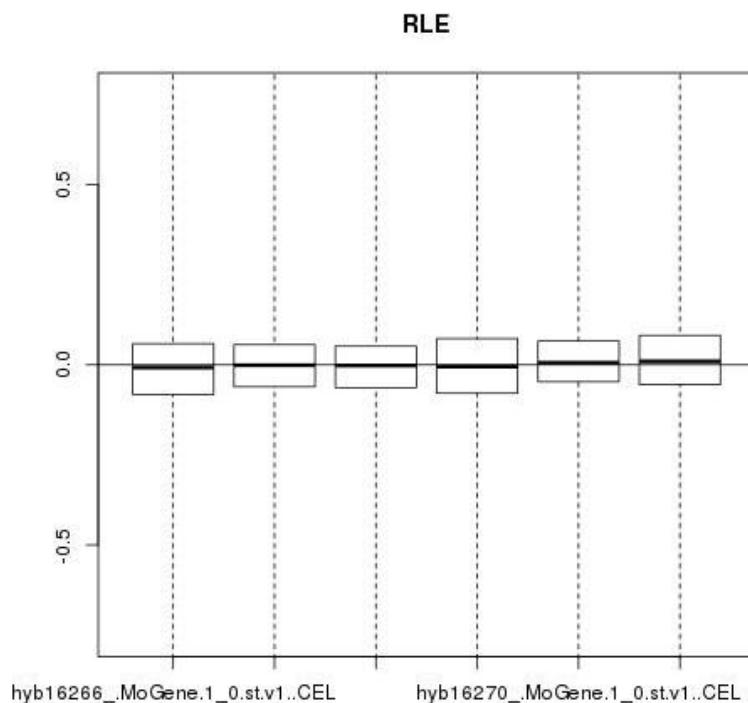
Opmerking: een *chip pseudo-image* gebaseerd op *residuals* wordt voor elke microarray apart gemaakt. Hier is als voorbeeld enkel de *chip pseudo-image* van de eerste microarray weergegeven. De overige vijf zijn terug te vinden op de CD-ROM.

hyb16266_(MoGene-1_0-st-v1).CEL



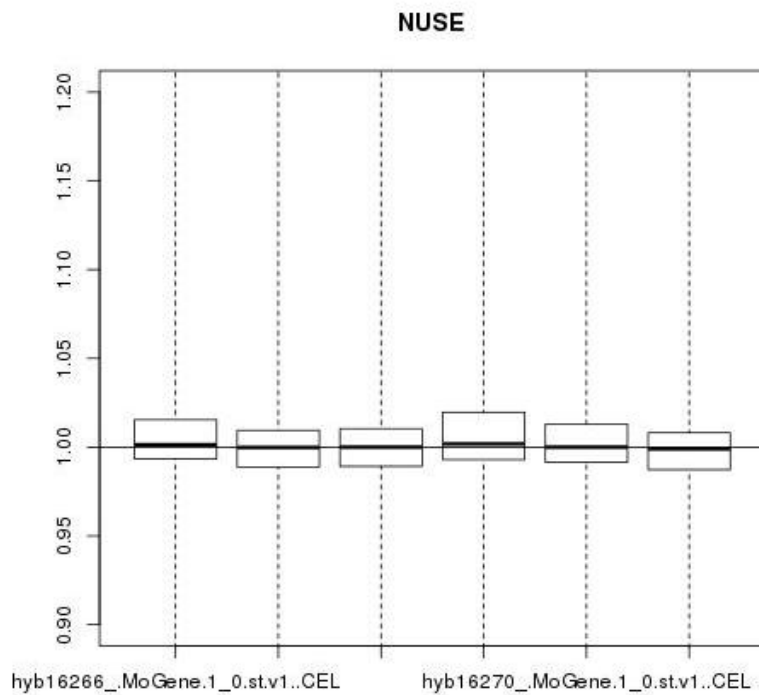
Figuur 8-2: *Chip pseudo-image* gebaseerd op *weights* van microarray 1 van de CD8 α^+ DC dataset. Deze microarray komt overeen met een WT-microarray. Er zijn geen zones aanwezig die sterk groen gekleurd zijn, waaruit kan afgeleid worden dat er geen artefacten aanwezig zijn in deze microarray.

8.4.1.3 RLE



Figuur 8-3: RLE-plot van de ruwe microarray data van de CD8 α^+ DC dataset. De grootte van de boxplots zijn gelijkaardig en de mediaan van de verschillende boxplots is steeds gelegen rond nul, wat betekent dat de data een goede kwaliteit vertonen.

8.4.1.4 NUSE

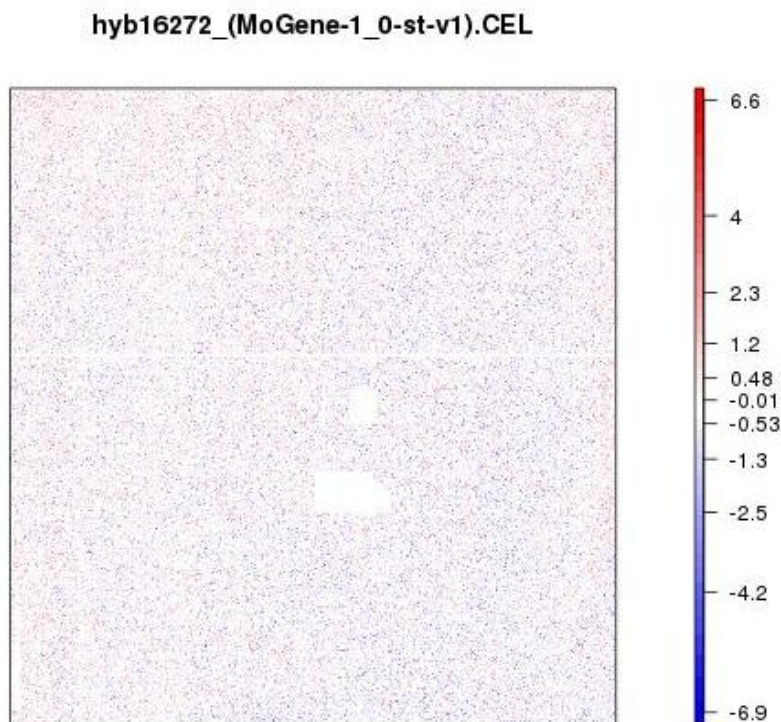


Figuur 8-4: NUSE-plot van de CD8 α^+ DC dataset . Bij een NUSE-plot is het vooral belangrijk te kijken naar het feit of de medianen van de verschillende boxplots rond één gelokaliseerd zijn. Aangezien NUSE zeer gevoelig is aan *noise* is de kans reëel dat er grote schommelingen kunnen waargenomen worden in de posities van de boxplots. Dit is hier niet het geval, wat bewijst de data van zeer goede kwaliteit zijn.

8.4.2 CD11b⁺ DC dataset

8.4.2.1 *Chip pseudo-image - residuals*

Opmerking: een *chip pseudo-image* gebaseerd op *residuals* wordt voor elke microarray apart gemaakt. Hier is als voorbeeld enkel de *chip pseudo-image* van de eerste microarray weergegeven. De overige vijf zijn terug te vinden op de CD.

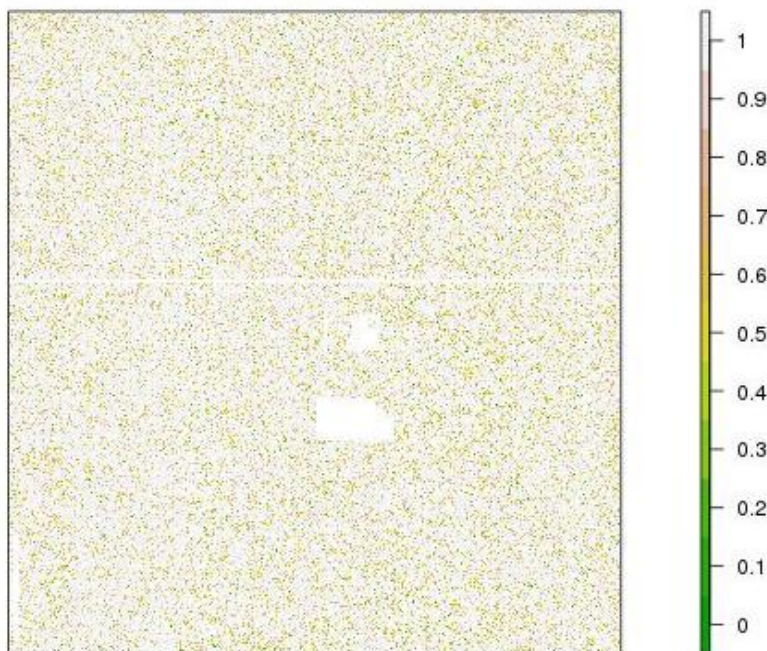


Figuur 8-5: *Chip pseudo-image* gebaseerd op *residuals* van microarray 1 van de CD11b⁺ DC dataset. Deze microarray komt overeen met een WT-microarray. Er zijn geen zones aanwezig die meer blauw of rood gekleurd zijn, waaruit kan afgeleid worden dat er geen artefacten aanwezig zijn in deze microarray.

8.4.2.2 *Chip pseudo-image - weights*

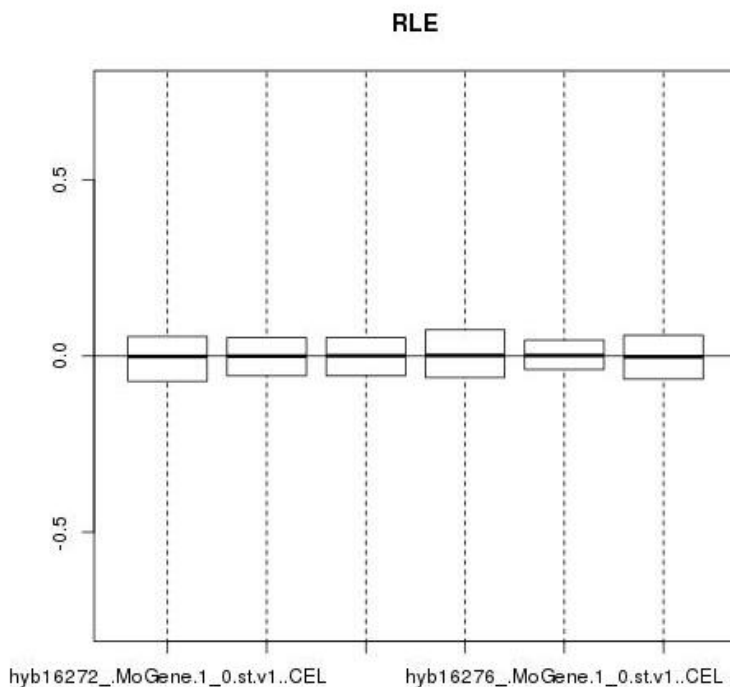
Opmerking: een *chip pseudo-image* gebaseerd op *weights* wordt voor elke microarray apart gemaakt. Hier is als voorbeeld enkel de *chip pseudo-image* van de eerste microarray weergegeven. De overige vijf zijn terug te vinden op de CD.

hyb16272_(MoGene-1_0-st-v1).CEL



Figuur 8-6: *Chip pseudo-image* gebaseerd op *weights* van microarray 1 van de CD11b⁺ DC dataset. Deze microarray komt overeen met een WT-microarray. Er zijn geen zones aanwezig die sterk groen gekleurd zijn, waaruit kan afgeleid worden dat er geen artefacten aanwezig zijn in deze microarray.

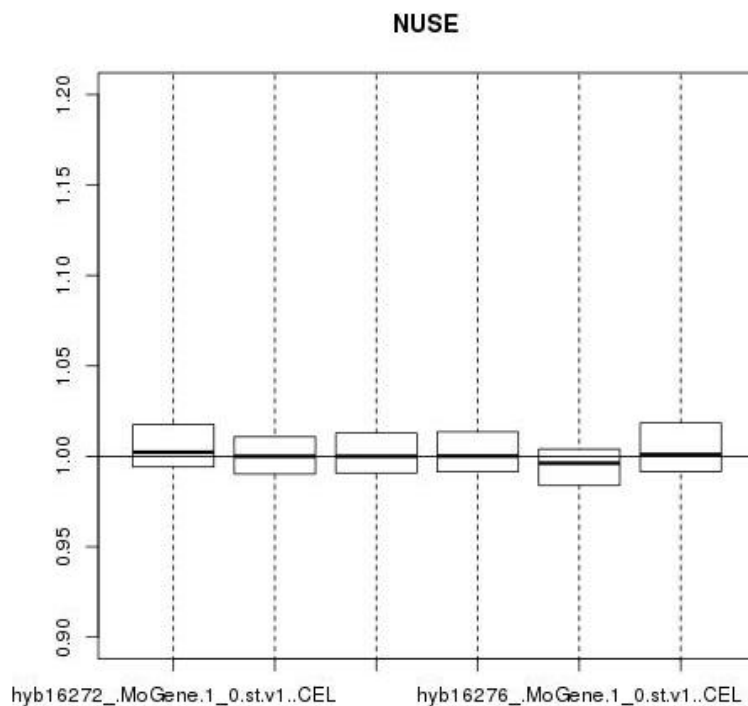
8.4.2.3 RLE



Figuur 8-7: RLE-plot van de ruwe microarray data van de CD11b⁺ DC dataset. De grootte van de boxplots zijn gelijkaardig en de mediaan van de verschillende boxplots is steeds gelegen rond nul, wat betekent dat de data een goede kwaliteit vertonen.

Analyse van *Xbp1* KO-microarray data in cDC

8.4.2.4 NUSE



Figuur 8-8: NUSE-plot van de ruwe microarray data van de CD11b⁺ DC dataset. Bij een NUSE-plot is het vooral belangrijk te kijken naar het feit of de medianen van de verschillende boxplot rond één gelokaliseerd zijn. Aangezien NUSE zeer gevoelig is aan noise is de kans reëel dat er grote schommelingen kunnen waargenomen worden in de positie van de boxplot. Dit is hier niet het geval, wat bewijst de data van zeer goede kwaliteit zijn.

8.5 Bijlage 5

De volledige lijst van de genen die een differentiële genexpressie vertonen in de CD8α⁺ DC subset.

Accession_number	Gene name	Log2(fold change)
NM_012055	<i>Asns</i>	2.766231
NM_008135	<i>Slc6a9</i>	2.220586
NM_016966	<i>Phgdh</i>	1.696261
BC132471	<i>Gm129</i>	1.466857
NM_009943	<i>Cox6a2</i>	1.461136
NM_027921	<i>Slc16a14</i>	1.330624
NM_173866	<i>Gpt2</i>	1.254666
NM_144795	<i>Pycr1</i>	1.202696
NM_018861	<i>Slc1a4</i>	1.200493
NM_146064	<i>Soat2</i>	1.159205
NM_028230	<i>Shmt2</i>	1.130133
NM_013742	<i>Cars</i>	1.095306
NM_028994	<i>Pck2</i>	1.094139
NM_177420	<i>Psat1</i>	1.024319
NM_019698	<i>Aldh18a1</i>	1.008829
NM_010121	<i>Eif2ak3</i>	-1.00565

NM_018804	<i>Syt11</i>	-1.01329
NM_198295	<i>Tmx3</i>	-1.01628
NM_028982	<i>8430419L09Rik</i>	-1.02532
NM_030084	<i>Gpr108</i>	-1.04015
NM_013488	<i>Cd4</i>	-1.04057
NM_146115	<i>A830007P12Rik</i>	-1.0483
NM_029884	<i>Hgsnat</i>	-1.06052
NM_145367	<i>Txndc5</i>	-1.06509
NM_027855	<i>0610007C21Rik</i>	-1.06824
NM_133933	<i>Rpn1</i>	-1.09186
NM_175274	<i>Ttyh3</i>	-1.09835
NM_011399	<i>Slc25a17</i>	-1.10125
NM_145362	<i>Alg1</i>	-1.10716
NM_175403	<i>Mlec</i>	-1.11805
NM_178395	<i>Zdhhc2</i>	-1.16089
NM_023625	<i>Plbd2</i>	-1.16797
NM_010515	<i>Igf2r</i>	-1.17814
NM_001081103	<i>Stim2</i>	-1.19671
NM_010215	<i>Il4i1</i>	-1.20838
NM_053103	<i>Entpd7</i>	-1.22067
NM_008404	<i>Itgb2</i>	-1.22134
NM_019793	<i>Tspan3</i>	-1.23032
NM_033320	<i>Glce</i>	-1.26616
NM_025516	<i>Ergic3</i>	-1.26823
NM_178661	<i>Creb3l2</i>	-1.28485
NM_133655	<i>Cd81</i>	-1.34246
NM_009906	<i>Tpp1</i>	-1.3593
NM_024205	<i>Jkamp</i>	-1.41754
NM_029572	<i>Erp44</i>	-1.46305
NM_021540	<i>Rnf130</i>	-1.57019
NM_001033525	<i>Kcnk6</i>	-1.58496
NM_010684	<i>Lamp1</i>	-1.60165
NM_001025313	<i>Tapbp</i>	-1.61709
NM_175036	<i>Leprot</i>	-1.64292
NM_008013	<i>Fgl2</i>	-1.65727
NM_015740	<i>Bloc1s1</i>	-1.68447
NM_001160215	<i>Fcrla</i>	-1.72538
NM_020626	<i>Tmem27</i>	-1.81126
NM_009137	<i>Ccl22</i>	-2.38195

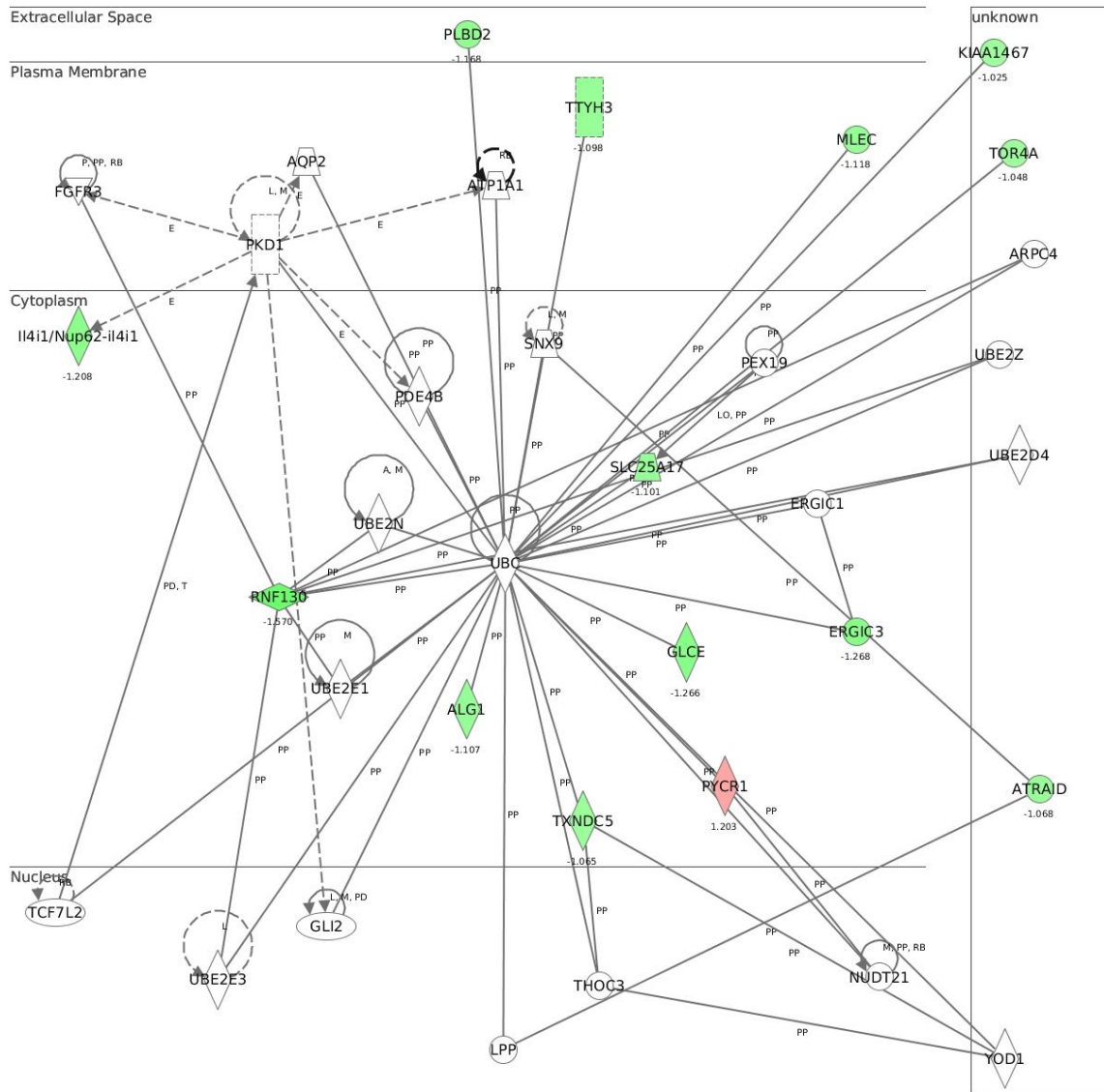
8.6 Bijlage 6

De overige drie netwerken van de CD8 α^+ DC dataset en de twee netwerken van de CD11b $^+$ DC dataset.

8.6.1 CD8 α^+ DC dataset

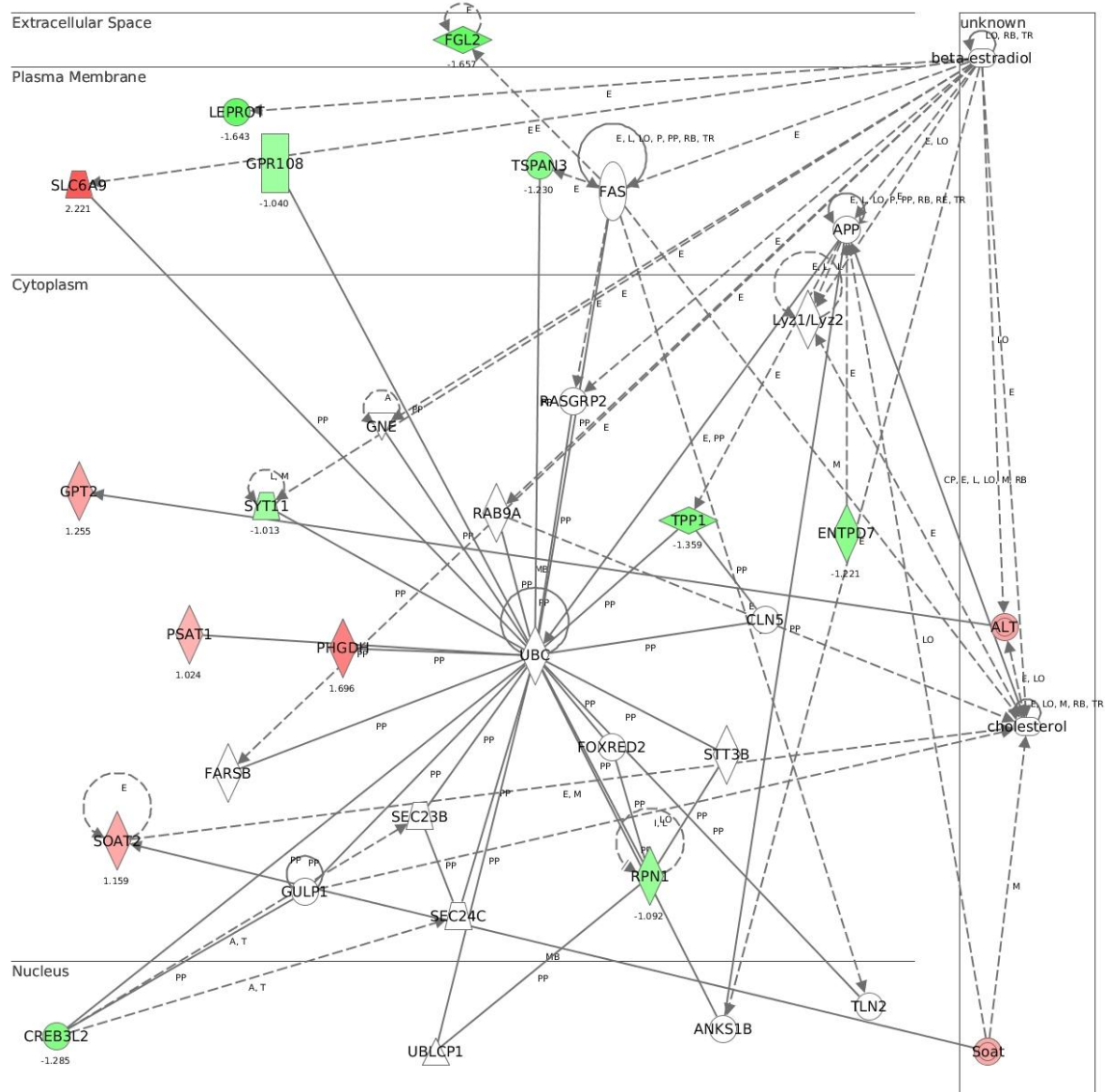
8.6.1.1 Network 2

Network 2 : xbp1_cd8_allgenes : genes_de_annotated_sorted.txt : xbp1_cd8_allgenes



8.6.1.2 Netwerk 3

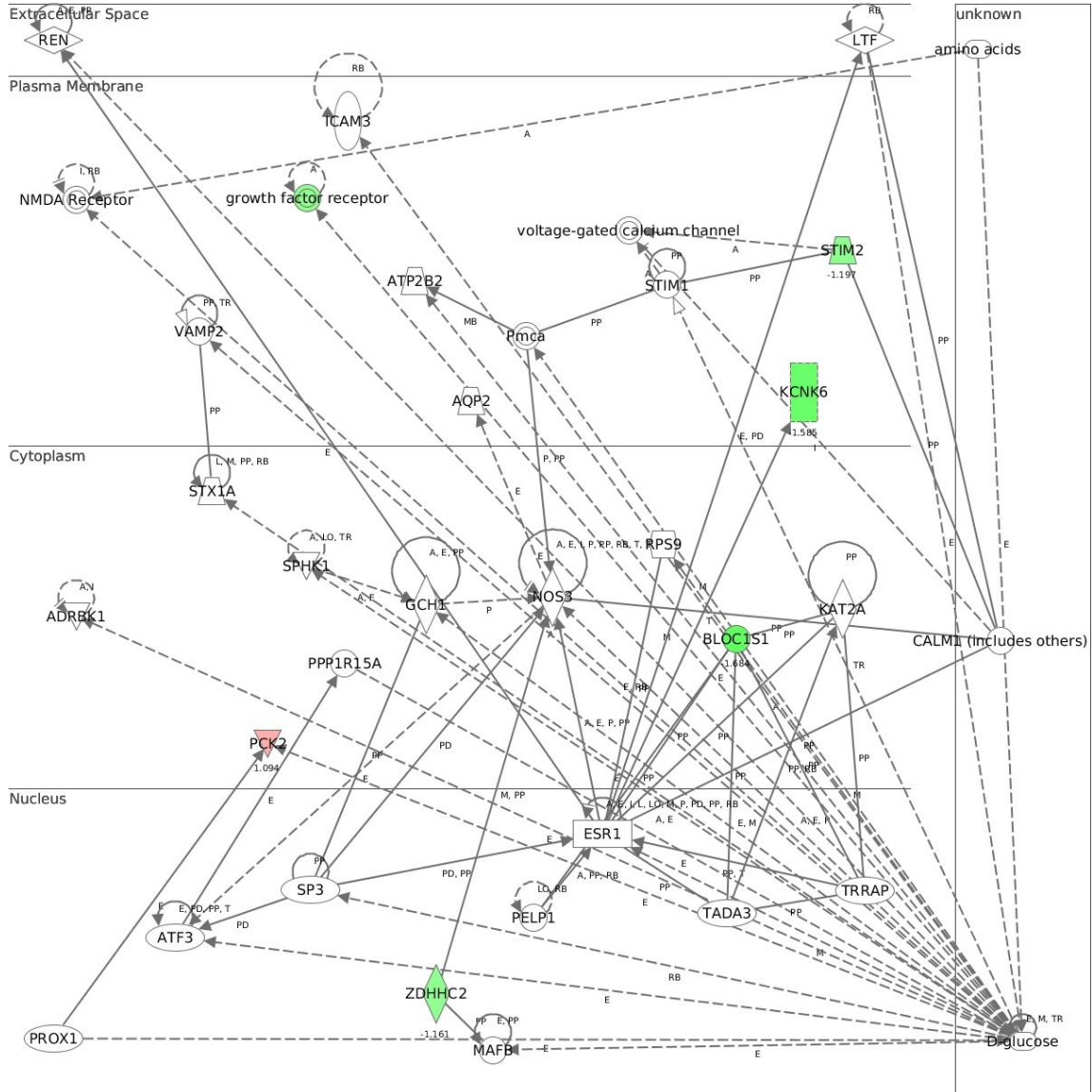
Network 3 : xbp1_cd8_allgenes : genes_de_annotated_sorted.txt : xbp1_cd8_allgenes



© 2000-2013 Ingenuity Systems, Inc. All rights reserved.

8.6.1.3 Netwerk 4

Network 4 : xbp1_cd8_allgenes : genes_de_annotated_sorted.txt : xbp1_cd8_allgenes

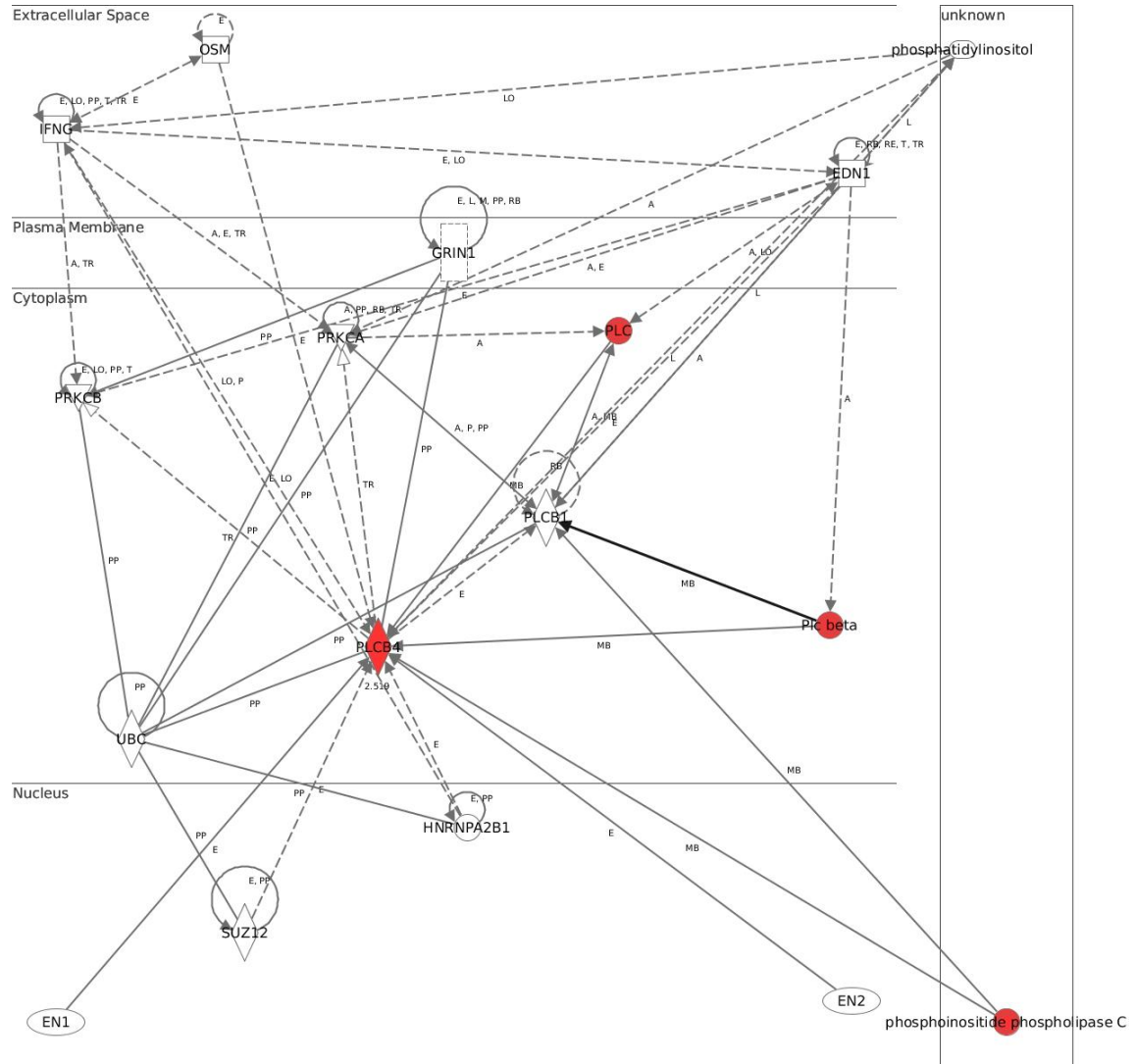


© 2000-2013 Ingenuity Systems, Inc. All rights reserved.

8.6.2 CD11b⁺ DC dataset

8.6.2.1 Netwerk 1

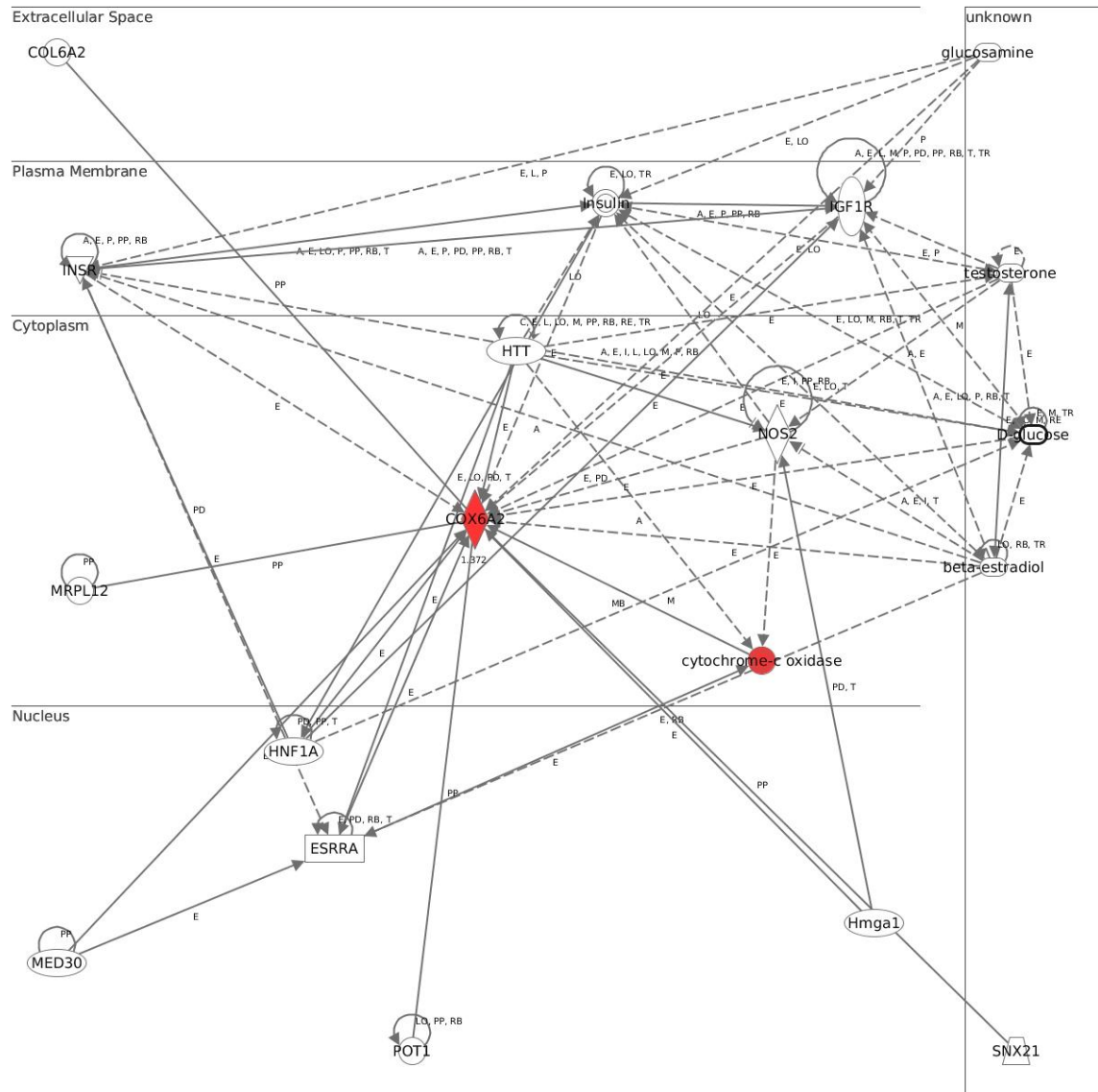
Network 1 : *xbp1_cd11b_allgenes2* : *xbp1_cd11b_allgenes2* : *xbp1_cd11b_allgenes2*



© 2000-2013 Ingenuity Systems, Inc. All rights reserved.

8.6.2.2 Netwerk 2

Network 1 : xbp1_cd11b_allgenes : xbp1_cd11b_allgenes : xbp1_cd11b_allgenes



© 2000-2013 Ingenuity Systems, Inc. All rights reserved.

8.7 Bijlage 7

Resultaten FIMO-analyse; lijst van de genen die een *Xbp1*-motief bezitten.

Gennaam	Uppereguleerd of downgereguleerd	P-waarde
<i>Tapbp</i>	Down ↓	1.18e-05
<i>Ttyh3</i>	Down ↓	1.71e-05
<i>Kcnk6</i>	Down ↓	1.92e-05
<i>Rnf130</i>	Down ↓	2.14e-06
<i>Entpd7</i>	Down ↓	2.1e-05
<i>Erp44</i>	Down ↓	2.26e-05
<i>Eif2ak3</i>	Down ↓	2.55e-05
<i>Cars</i>	Up ↑	2.85e-05
<i>Mlec</i>	Down ↓	2.85e-05
<i>Gpr108</i>	Down ↓	2.87e-05
<i>Ccl22</i>	Up ↑	3.06e-06
<i>Gm129</i>	Up ↑	3.9e-05
<i>Igf2r</i>	Down ↓	4.39e-05
<i>Itgb2</i>	Down ↓	4.39e-05
<i>Lamp1</i>	Down ↓	4.39e-05
<i>Zdhhc2</i>	Down ↓	4.39e-05
<i>Aldh18a1</i>	Up ↑	5.35e-05
<i>Fcrla</i>	Down ↓	5.35e-05
<i>Psat1</i>	Up ↑	5.35e-05
<i>Syt11</i>	Down ↓	5.35e-05
<i>Tmx3</i>	Down ↓	5.35e-05
<i>Plbd2</i>	Down ↓	5.66e-05
<i>Slc1a4</i>	Up ↑	8.03e-05
<i>Bloc1s1</i>	Down ↓	8.85e-05
<i>Atraid</i>	Down ↓	9.56e-05
<i>Cd4</i>	Down ↓	9.74e-05
<i>Cd81</i>	Down ↓	9.74e-05