# Analyse translocatie GR

**Celbiologie jaar 1 1516**

**Kwantificatie van eGFP-hGR translocatie**

* Condities:
	+ Geen GR-ligand => negatieve controle
	+ Cortisol (translocatie GR naar nucleus) => positieve controle
	+ Corticosteron (translocatie GR naar nucleus)
	+ Mifepreston (translocatie naar nucleus, maar geen transcriptie)
	+ 17-AAG of geldanamycine (geen translocatie naar nucleus)
* Localisatie GFP signaal:

Type 1: Nucleus

Type 2: Partieel (zowel in cytosol als in nucleus)

Type 3: Cytosol

**Statististische analyse bespreken**

* We tellen aantal type 1, 2 en 3 cellen op een dekglaasje (3 foto’s) en hiervan kunnen we proporties uitrekenen
* Bespreek hoe we de data in gaan voeren (kolommen A-C : P-groep, tafel, duo; kolom D: conditie 1/2/3; kolommen E-G: # cellen voor type 1, 2, 3)
* Toetskeuzeschema met de studenten doorlopen (output (afhankelijke) variabele (proportie/aantal cellen met translocatie) is continu en de predictor (onafhankelijke) variabele (conditie) is categoriaal. Je komt dan uit op een ANOVA (F-waarde (wordt niet gegeven door ons R script) met post-hoc Tukey tests (t-waardes).
* Uitleggen waarom we voor non-parametrische statistiek kiezen (omdat spreiding groot is in absolute tellingen (niet iedereen telt 50 cellen), wellicht niet normaal verdeeld)
* Hoe denken ze dat het figuur van de data er uit zal zien?
* Aan studenten uitleggen waarom ze een boxplot maken (non-parametrische statistiek)

**Data verzamelen**

* Data wordt op Google docs verzameld:
	+ Log in op Google Drive en ga naar het mapje Celbiologie

Google Drive account naam: celbiologiepb2015@gmail.com

Google Drive account wachtwoord: celbio2015

**Script**

* Het script geeft 2 figuren (onder plots vind je pijltjes waarmee je naar de eerste kunt navigeren)

Figuur 1: Alle condities

Figuur 2: Experimentele conditie die je zelf invuld + pos. en neg. Controle

* conditie invullen (1 = cort, 2 = mif, 3 = 17AAG) default “1”
* referentie locatie invullen (nucleus = type 1, partieel = type 2, cytoplasma = type 3) default “nucleus”
* simpele\_output = “ja” => default “ja”

* + “ja” geeft enkel de t en p-waardes van de vergelijkingen aantal type 1 cellen (je kunt ook kiezen voor type 2 of 3) in de experimentele conditie (zelf invullen welke) versus positieve ctrl en negatieve ctrl
	+ “nee” geeft de ingewikkelde output (t/p-waardes e.d.) met alle posthoc vergelijkingen
* Script runnen: CTRL-A (select all) > CTRL-R (run)
* Bij runnen script:
	+ Vraagt op een gegeven moment een file (browse window wordt geopend (taakbalk)) 🡪 selecteer de csv file
* Mogelijke errors
	+ Iets verkeerd invullen
	+ Installeren van packages (in v4.0.2 wordt eerst gecheckt of de vereiste packages al geïnstalleerd zijn, indien niet, dan worden ze geïnstalleerd. R39-41 => FALSE is niet geïnstalleerd.)
* Outliers: blijven er in, want als je die er uit wil halen moet je weten waarom je die er uit haalt.
	+ Als je die er uit haalt: noemen in discussie en beide waardes geven
* **Figuur:**
	+ Mediaan
	+ 1e + 3e kwartiel grenzen (25-75%
	+ Error barrs: 1,5 x (25%-75%) en dan de min/max waarde daarbinnen => outliers liggen dus daarbuiten

**Verdere uitleg script**

****

Hier kunnen de studenten aangeven welke experimentele conditie ze willen bekijken en welke proportie (van welk type cel: 1, 2 of 3) ze willen vergelijken. En tot slot de keus voor simpele output of volledige output (dus alle posthoc Tukey tests).



Als de studenten per ongeluk iets anders invullen dan aangegeven, dan geeft R een error message maar runt wel, dus als iets niet klopt eerst even de console\* output lezen (scroll scroll scroll), zoals hier onder.



\* console is het kwadrant linksonder

(vervolg script)



Hier worden 3 packages gedownload die we nodig hebben voor onze analyse (zonder werkt het niet). Het kan zijn dat je R moet runnen in administrator mode (rechtermuisknop op programma), zie afbeelding hier onder.



(vervolg script)



Hier wordt de excel data ingelezen (met een browse venster) en in een nieuwe tabel gekopieerd op een manier die R graag wil voor de toets die het zo gaat uitvoeren. Vrij simpele tabel met de naam data\_selectie. Klik maar op data\_selectie in de global environment (het rechterboven kwadrant in R-studio) om deze te bekijken.



De uiteindelijke toets. Een generalized linear model (glm functie in R), dit is een regressie-analyse als ANOVA (deze twee zijn hetzelfde zie: <http://www.theanalysisfactor.com/why-anova-and-linear-regression-are-the-same-analysis/>), als ik het goed begrepen heb van de maker van dit script (Boris Berkhout) is deze glm analyse geschikt voor zowel parametrisch als non-parametrisch toetsen. De toets maakt gebruik van de absolute aantallen cellen (niet de proporties) en vergelijkt alle medianen met de mediaan van de referentielocatie (dit is default de nucleus (type 1 cel)), het zogenaamde intercept. De glm functie geeft geen F-waarde (verander glm op regel 70 in lm en je krijgt deze wel te zien mocht je dat willen) maar enkel de posthoc Tukey tests, dit zijn t-waarden.



Hier staan twee zogenaamde if statements. De eerste if zegt als simpele output = “ja” laat dan de volgende waarden zien in de console output (script versie 4.0.3):



Als je bij simpele output “nee” hebt gekozen dan zie de volledige summary van het glm model, dus alle Tukey posthoc tests.

(vervolg script)



Hier worden de twee plots gemaakt. Op regel 105 zie je Conditie in het groen, dit is een variabele en die heb je aan het begin gedefinieerd (1 = cort, 2 = mif, 3 = 17AAG).

(vervolg script)



En tot slot worden de plots onder de juiste naam in de huidige folder opgeslagen (je vind de huidige folder in het kwadrant rechtsonder bij Files). Maar zoals je bij regel 121 ziet geeft R dit ook in de console weer.