Is Gene Expression Affected by Treatment?

Joan Shu Ting Lim

May 2024

1 Introduction

The dataset examines the effect of a new treatment on the effect of growth factor on the gene expression. The data was obtained from our collaborator, Dr Karl Berator.

The data consists of 88 rows with information on the cell line, type of treatment, name of cell line (gene line), concentration of treatment used and gene expression. The cell line had 2 types of data, which are "Cell-type 101" or "Wild-type". The type of treatment had 2 types of data, "Placebo" or "Activating Factor 42", where "Placebo" represents a saline solution being used as a placebo and "Activating Factor 42" represents the new treatment being used. The concentration of treatment goes from 0 to 10. There are 8 types of gene lines, those being "GL-cDZ", "GL-cwN", "GL-kYH", "GL- MFA", "GL-rjS", "GL-XIb", "GL-Xik" and "GL-ZHw".

The key research questions are to find if gene expression is affected by treatment and the concentration of treatment, and if treatment affects different cell line.

2 Methods

The analysis was performed using R 4.3.1 (R Core Team, 2024) using RStudio. The variables, cell line, type of treatment and gene line showed inconsistent labelling, to ensure consistency in the data, the variable was recoded.

- In the cell line variable, "WILD-TYPE" was recoded to "Wild-type" and "CELL-TYPE 101" was recoded to "Cell-type 101"
- In the treatment variable, "activating factor 42" was recoded to "Activating factor 42" and "placebo" was recoded to "Placebo".
- In the name variable, "Gl-Cwn" was recoded to "GL-cwN", "Gl-Rjs" was recoded to "GL-rjS", "Gl-Xib" was recoded to "GL-XIb", and "Gl-Zhw" was recoded to "GL-ZHw".

A scatter plot was then plotted to investigate the relationship over different concentrations of treatments, the gene expression on each cell line and their gene line.

To investigate the effects of treatment, concentration, and cell line on gene expression, a mixed-effects model was fitted. This model includes fixed effects for each treatment, their concentrations and type of cell line along with their interaction terms. The random effect for gene line was included to account for variability between different samples.

The model was then evaluated to obtain the parameter estimates (coefficients) and perform hypothesis testing on the fixed effects. The significance of the random effects was also assessed.

3 Results

Figure 1 gives a scatter-plot of gene expression against the concentration of treatment for each gene line. The color of the points denotes the treatment type: blue for Activating Factor 42 (new treatment) and brown for the placebo (saline solution). Each point is labelled with the corresponding gene line.

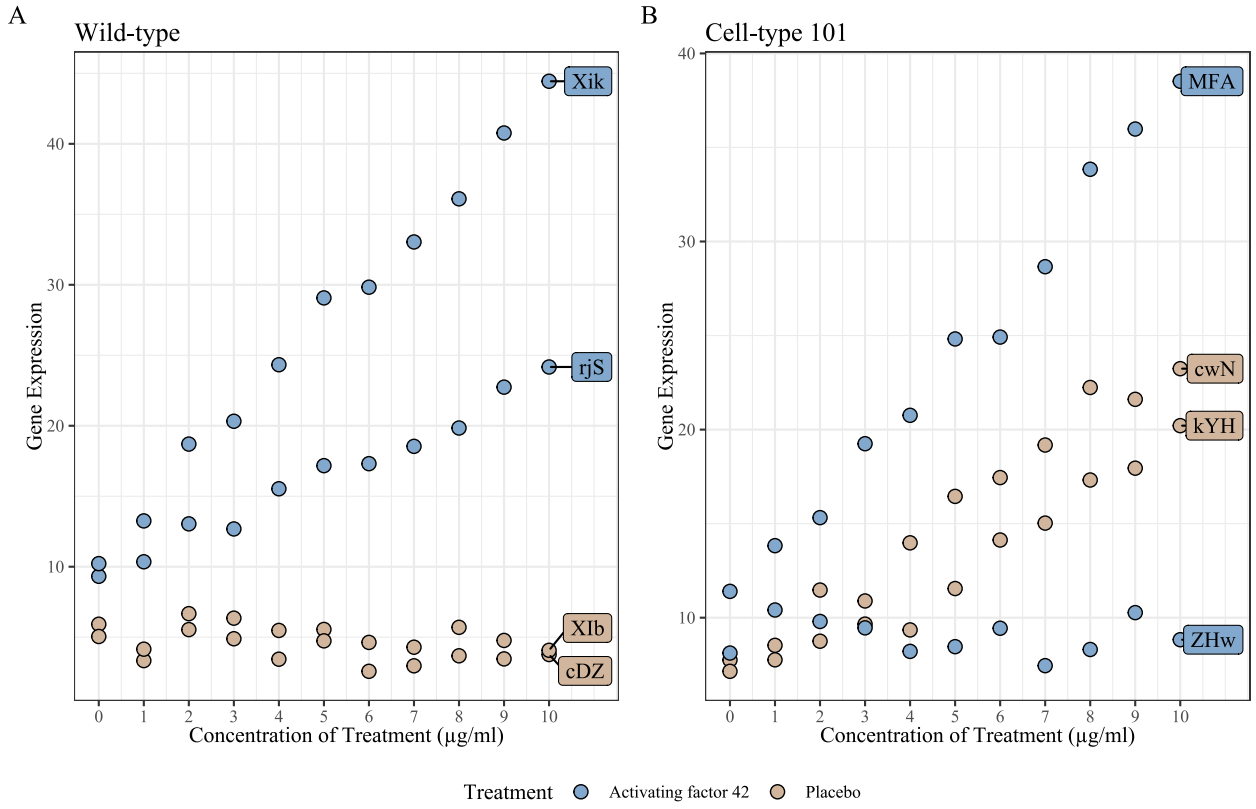

Figure 1: Scatter plot of gene expression over the concentration of each treatment for different cell lines.

From Figure 1, in the Wild-type cell line, both gene lines, Xik and rjS, exhibit an increasing trend in gene expression with the concentration of the new treatment. In contrast, with the placebo, there is no evident growth, and a slight decreasing trend may even be observed for both gene lines, XIb and cDZ. The gene expression for the MFA gene line increases with higher concentrations of the new treatment. Conversely, the ZHw gene line exhibits a decrease in gene expression as the concentration of the new treatment increases. On the other hand, when the placebo is used, both the cwN and kYH gene lines show an increasing trend in gene expression as the concentration of the placebo increases.

After fitting the mixed-effects model, the coefficient table of the fixed effects of the model is obtained in Table 1. The p-values indicate that concentration (conc), the interaction term between concentration and Wild-type cell line (conc:cell_lineWild-type), and the interaction term between treatment with Placebo, concentration and the Wild-type cell line (treatment-Placebo:conc:cell_lineWild-type) are considered significant in predicting the gene expression as all the associated p-values are less than the significant value at 0.05, whereas the intercept, treatment in Placebo and the remaining interaction terms are not considered useful in predicting gene expression.

From Table 1, concentration has a positive effect on gene expression. As concentration increases by 1 $\mu g/ml$, the gene expression for Cell-type 101 will increase by 1.3567, on average, when Activating Factor 42 is used; if the placebo is used, then the gene expression of Cell-type 101 decreases by $-3.1766 + 1.3567 + 0.1407 = -1.6792$. For Wild-type cell line, if Activating Factor is used, then, on average, gene expression will increase by $1.3567 + 0.1327 + 1.0162 = 2.5056$. On average, if placebo is used then gene expression will increase by $-3.1766 + 1.3567 + 0.1327 + 0.1407 - 1.5180 + 1.0162 - 2.6538 = 1.6511$ as concentration of treatment increases by 1 $\mu g/ml$.

Term	Estimate	Std.Error	Test Statistic	p.value
(Intercept)	9.8557	4.7573	2.072	0.09981
treatmentPlacebo	-3.1766	6.7279	-0.472	0.65900
conc	1.3567	0.2194	6.184	0.00000
cell_lineWild-type	0.1327	6.7279	0.020	0.98511
treatmentPlacebo:conc	0.1407	0.3103	0.453	0.65155
treatmentPlacebo:cell_lineWild-type	-1.5180	9.5147	-0.160	0.88021
conc:cell_lineWild-type	1.0162	0.3103	3.275	0.00159
treatmentPlacebo:conc:cell_lineWild-type	-2.6538	0.4388	-6.048	0.00000

Table 1: Coefficient table for the fixed effects of the model.

Table 2 shows the coefficient estimates for the random effects of the model. From this table it can be noted that all the coefficients are the same for the assumed random effects, which in this case are the gene lines, the only difference is in the estimated intercepts for the random effects.

Term	(Intercept)
GL-cDZ	9.9730
GL-cwN	11.2540
GL-kYH	8.4574
GL-MFA	17.0530
GL-rjS	4.5512
GL-XIb	9.7384
GL-Xik	15.1601
GL-ZHw	2.6583

Table 2: Coefficient table for the random effects of the mixed-effects model. The full table can be observed in Table 3 in the Appendix 5.1

The different estimated intercepts in Table 2 indicates that, given the other independent variables included in the model, the gene line GL-cDz an average increase of 9.9730 in gene expression. For the gene line GL-cwN, an average increase of 11.2540 in gene expression, the gene line GL-kYH will have an increase of 8.4574, on average, in gene expression, GL-MFA will increase, on average, by 17.0530 in gene expression. The gene lines GL-rjS, GL-XIb, GL-Xik, and GL-ZHw will have an average increase of 4.5512, 9.7384, 15.6101 and 2.6583 in gene expression given the other independent variables included in the model.

4 Discussion

In this study, the effects of treatment, concentration, and cell line on gene expression was investigated. A mixed-effects model was fitted, revealing that concentration, the two-way interaction term with concentration and the cell line Wild-type, three-way interaction between treatment with Placebo, concentration, and cell line Wild-type had significant effects on gene expression. With this we can conclude that gene expression is affected by the type of treatment and the concentration of treatment, and treatment does affect different cell lines.

5 Appendix

5.1 Coefficient table for random effects of the mixed effects model

Term	(Intercept)	treatmentPlacebo	conc	cell_lineWild-type
GL-cDZ	9.9730	-3.1766	1.3567	0.1327
GL-cwN	11.2540	-3.1766	1.3567	0.1327
GL-kYH	8.4574	-3.1766	1.3567	0.1327
GL-MFA	17.0530	-3.1766	1.3567	0.1327
GL-rjS	4.5512	-3.1766	1.3567	0.1327
GL-XIb	9.7384	-3.1766	1.3567	0.1327
GL-Xik	15.1601	-3.1766	1.3567	0.1327
GL-ZHw	2.6583	-3.1766	1.3567	0.1327

Term	treatmentPlacebo:conc	treatmentPlacebo:cell_lineWild-type
GL-cDZ	0.1407	-1.5180
GL-cDZ	0.1407	-1.5180
GL-cwN	0.1407	-1.5180
GL-kYH	0.1407	-1.5180
GL-MFA	0.1407	-1.5180
GL-rjS	0.1407	-1.5180
GL-XIb	0.1407	-1.5180
GL-Xik	0.1407	-1.5180
GL-ZHw	0.1407	-1.5180

Term	conc:cell_lineWild-type	treatmentPlacebo:conc:cell_lineWild-type
GL-cDZ	1.0162	-2.6538
GL-cDZ	1.0162	-2.6538
GL-cwN	1.0162	-2.6538
GL-kYH	1.0162	-2.6538
GL-MFA	1.0162	-2.6538
GL-rjS	1.0162	-2.6538
GL-XIb	1.0162	-2.6538
GL-Xik	1.0162	-2.6538
GL-ZHw	1.0162	-2.6538

Table 3: Coefficient table for the random effects of the model