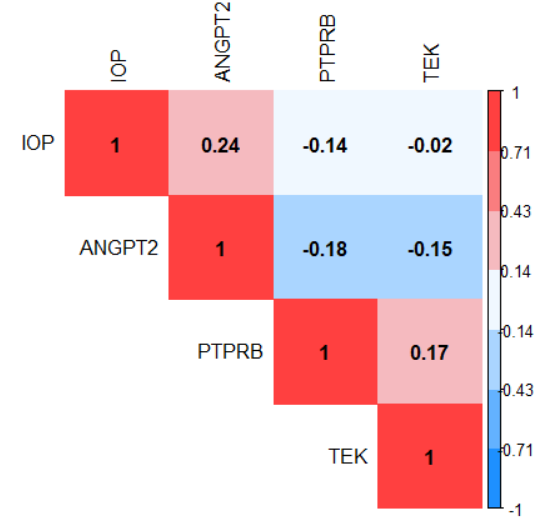
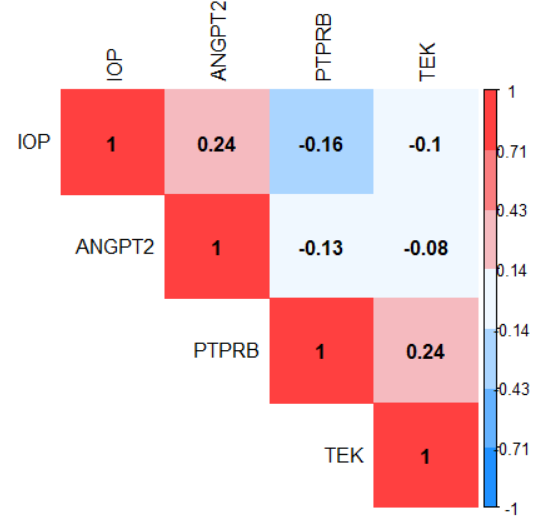
**Report date: March 17 Correlation Analysis Normalization method: Log2**

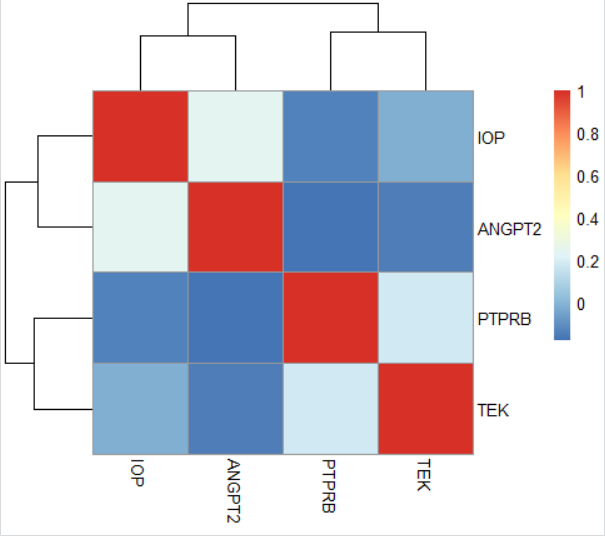
**Correlation method: Pearson**



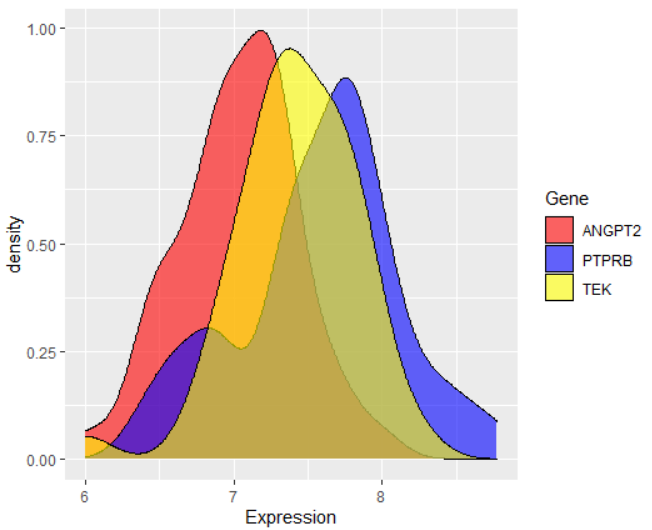
**Correlation method: Spearman**



**Clustering**

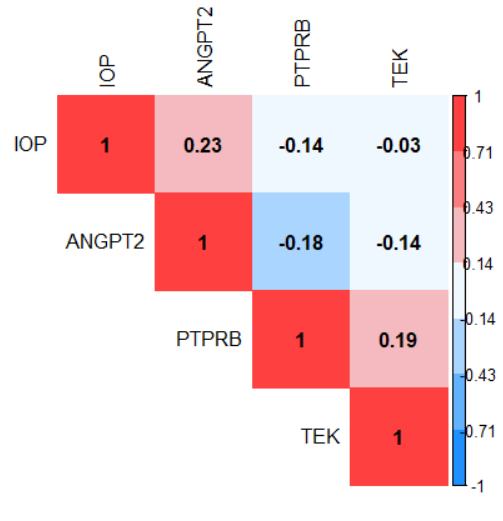


**Distribution**

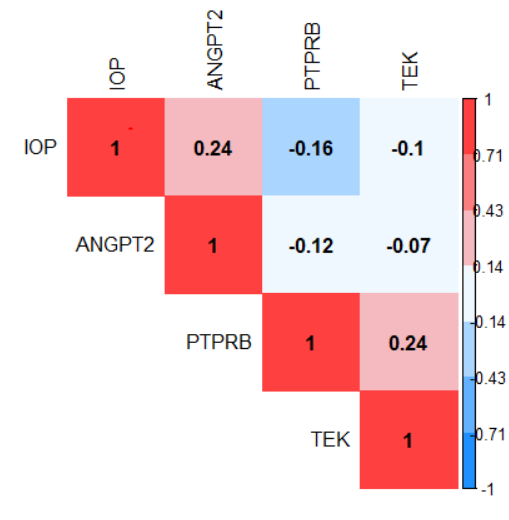


**Report date: March 17 Correlation Analysis Normalization method: rlog**

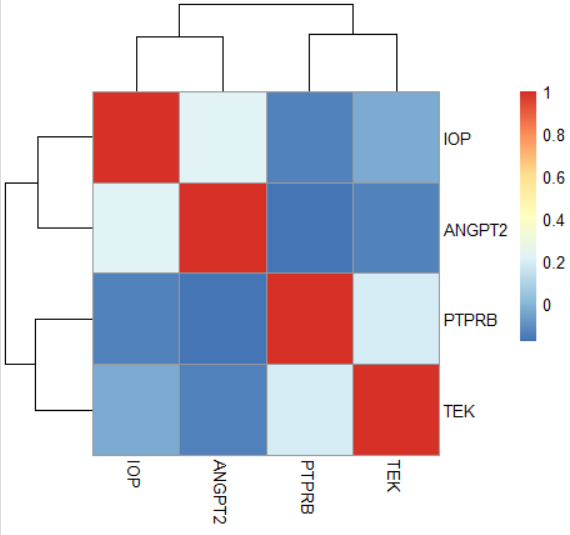
**Correlation method: Pearson**



**Correlation method: Spearman**



**Clustering**

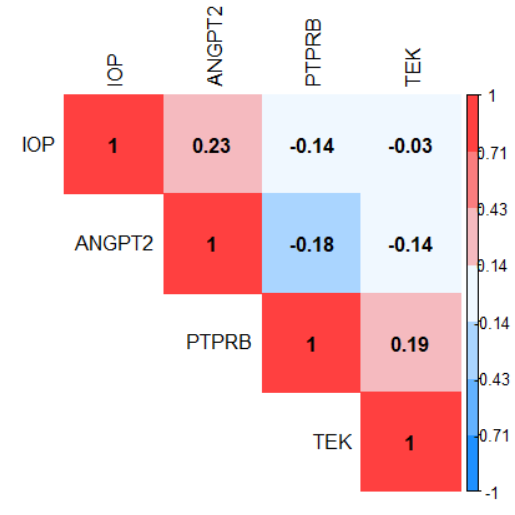


**Distribution**

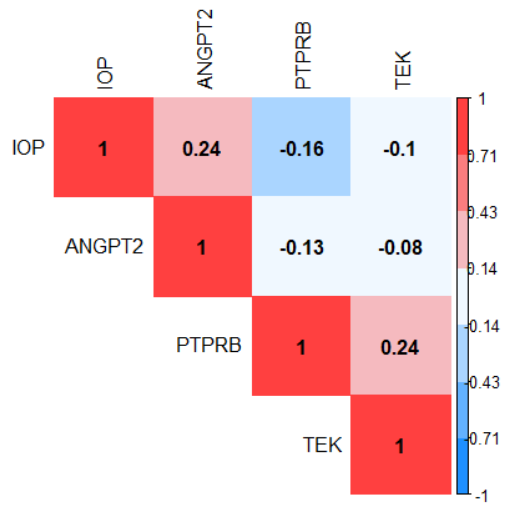


**Report date: March 17 Correlation Analysis Normalization method: vst**

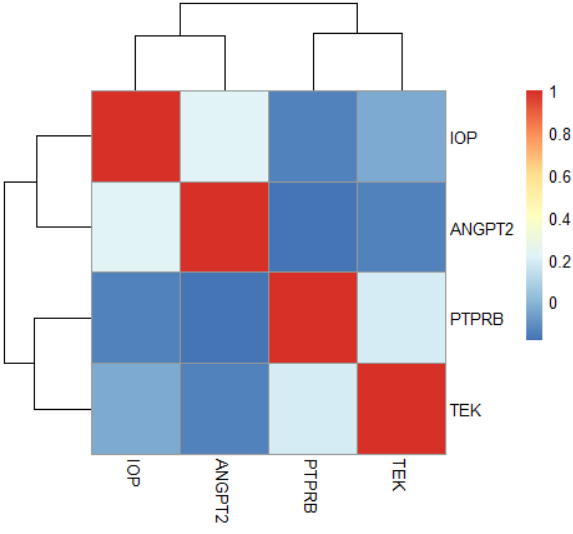
**Correlation method: Pearson**



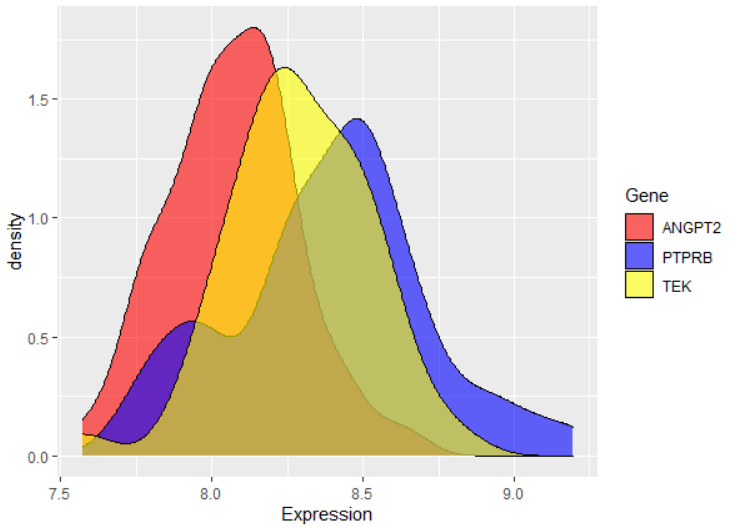
**Correlation method: Spearman**



**Clustering**

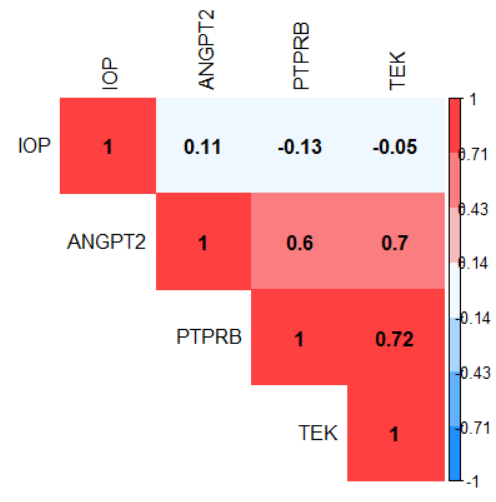


**Distribution**

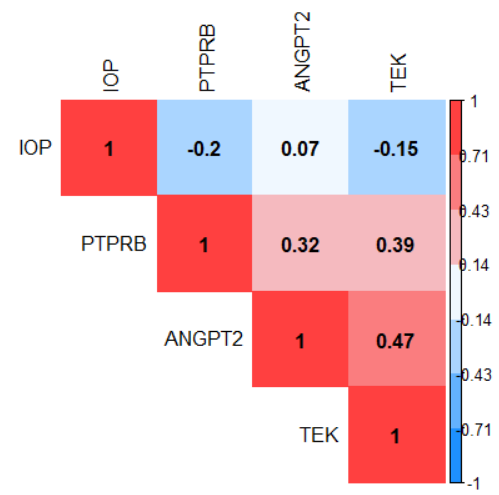


**Report date: March 17 Correlation Analysis No Normalization (Real Counts)**

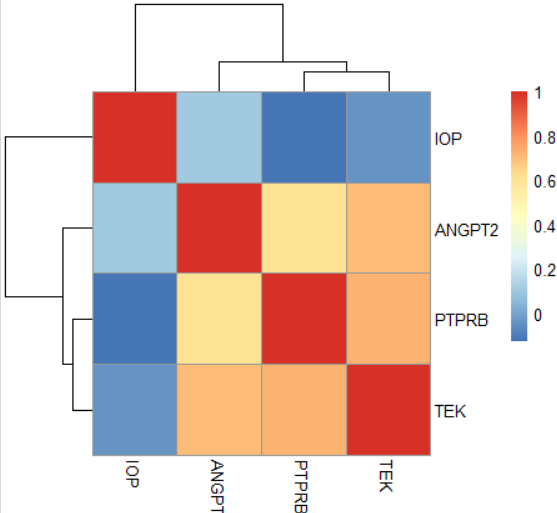
**Correlation method: Pearson**



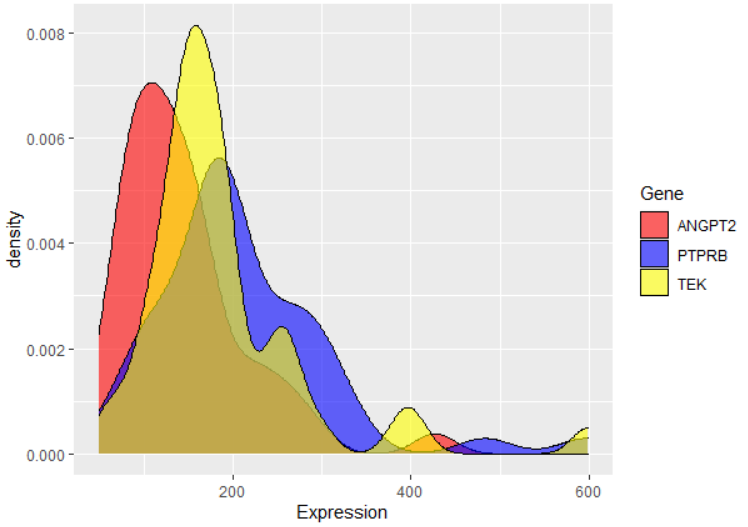
**Correlation method: Spearman**



**Clustering**



**Distribution**



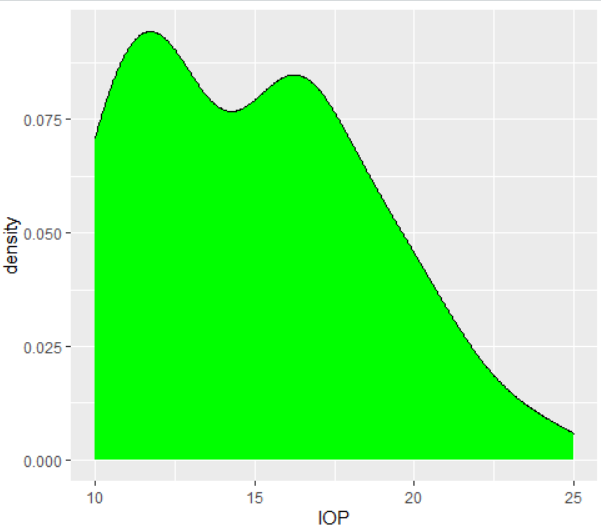
**Dsicusion:**

**Report date: March 17 Correlation Analysis**

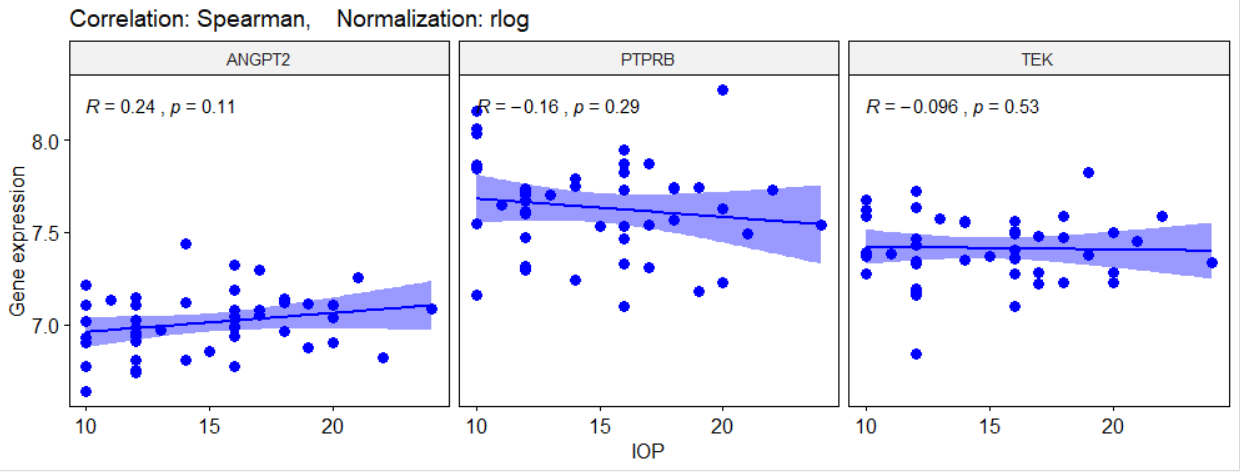
Based on the Clustering and Heatmap figures, I found that, two genes **PTPRB** and **TEK** act similarly. They have negative correlation with **IOP** and **ANGPT2**. Also, **IOP** has a weak positive correlation with **ANGPT2**. Compare three normalization methods: **Log2** (Log2 Normalized Counts Transformation), **rlog** (Regularized Log Transformation), and **vst** (Variance Stabilizing Transformation), and based on the distribution plots of three genes, it shows that **rlog** normalization gives good performance for our analysis in this step. Therefore, I considere **rlog** function as a normalization method for next analysis.

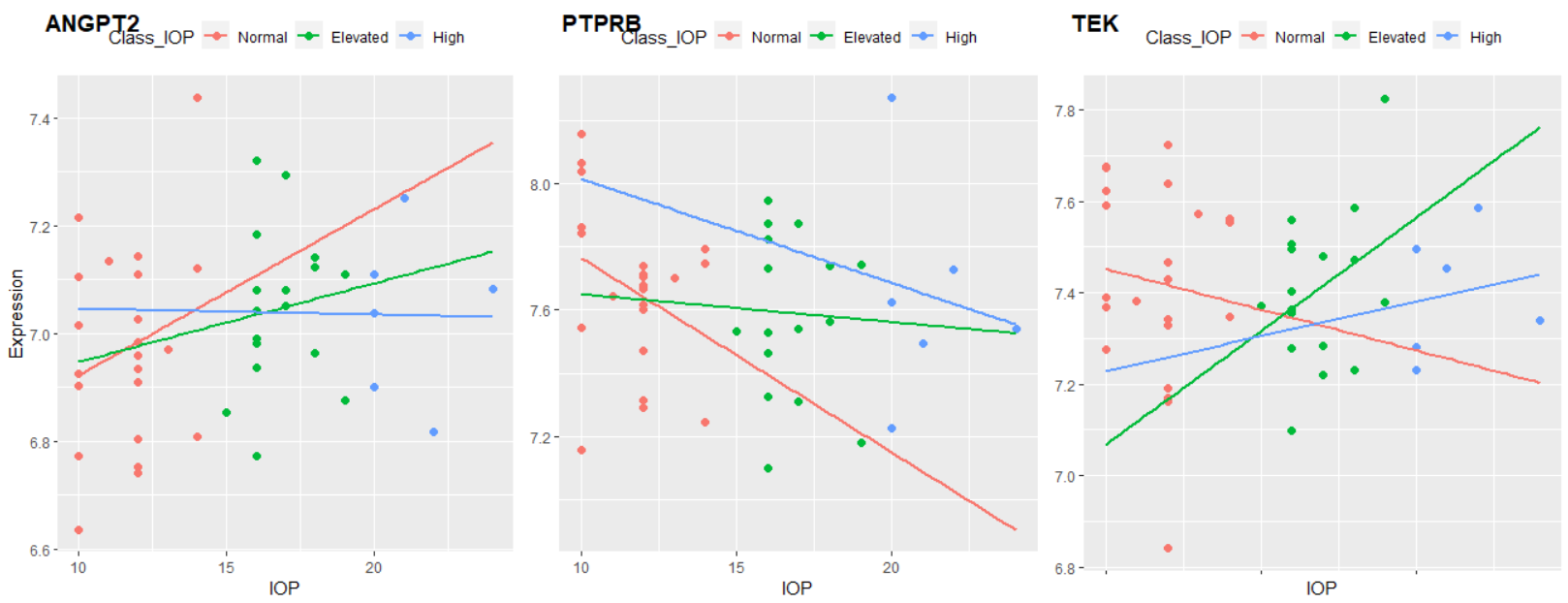
Also, between pearson (linear correlation) method and speaman (linear-skewed correlation) method, I found that the probability density function of TEK (yellow) is between ANGPT2 (red) and PTPRB (blue). It means that there is an important role or relation between ANGPT2 and TEK. I think, Spearman correlation is better than Pearson, and states that ANGPT2 and TEK are almost uncorrelated. I mean, I think thte ANGPT2 and PTPRB have negative correlation, PTPRB and TEK have positive correlation, and ANGPT2 and TEK are uncorrelated. Moreover, based on the density function of IOP, in figure below, there is almost negative correlation between two density function of IOP and PTPRB.

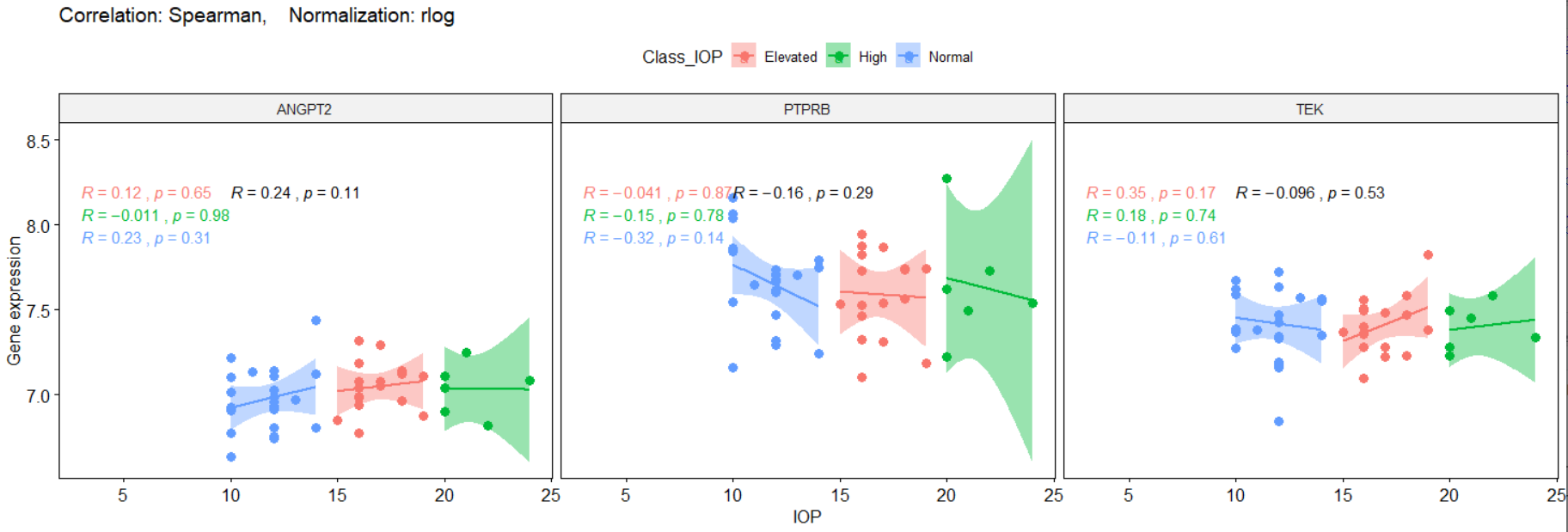
Therefore, I consider **rlog** as normalization method and **Spearman** as correlation method for further analysis.

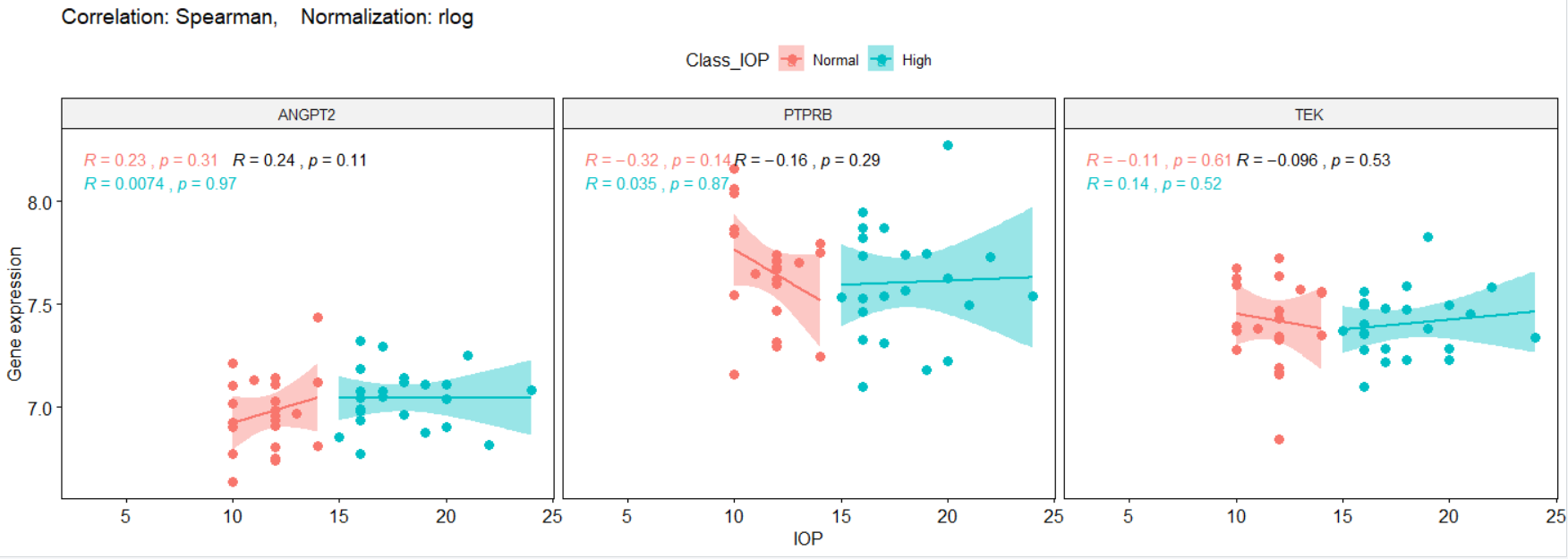
**Correlation between IOP and ANGPT2 Normalization: rlog Correlation: Spearman**







**Correlation between IOP and ANGPT2 Normalization: rlog Correlation: Spearman**



**Dsicusion:**

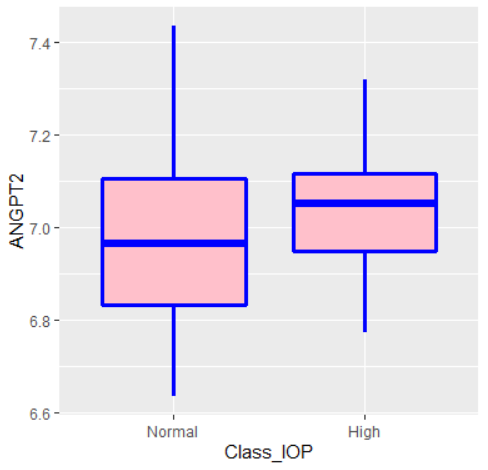
Based on figures above, ANGPT2 has not significant correlation with whole IOP data (R=0.24, P-value=0.11). Also, with two classified IOP (Normal, High) data and three classified IOP (Normal, Elevated, High) data has not significant correlation too. However, this gene has the maximum (week) correlation with IOP among other genes. To discuss statistically about them, I did t-test and ANOVA analysis too.

**Analysis of t-test:**

In this analysis, I want to test that are the mean of expression level of ANGPT2 is equal in two separated groups of Normal\_IOP and High\_IOP?

**H0: The mean of ANGPT2 expression level in Normal\_IOP group is equal to the mean of ANGPT2 expression level in High\_IOP group.**

**Ha: Both means are not equal.**



**Correlation between IOP and ANGPT2 Normalization: rlog Correlation: Spearman**

The results of two sided t-test with 95% confidential is:

t.test(ANGPT2 ~ Class\_IOP, data=corrTable\_IOP2, mu=0, alt="two.sided", conf=0.95, var.eq=F, paired=F)

Welch Two Sample t-test

data: ANGPT2 by Class\_IOP

t = -1.4124, df = 39.885, p-value = 0.1656

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.1705142 0.0302361

sample estimates:

mean in group Normal mean in group High

6.973136 7.043275

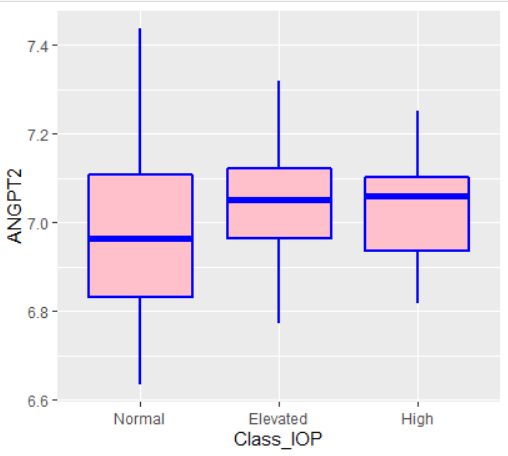
The P-value of the test is 0.1656, which is greater than the significance level alpha = 0.05. Therefore, I cannot reject the Null-Hypothesis and alternative hypothesis is accepted. **I conclude that** **the average expression level of ANGPT2 gene in samples with Normal\_IOP is not significantly different with the samples with High\_IOP. It means the the average expression level of ANGPT2 gene are almost same in two gropus.**

**Analysis of ANOVA:**

In this analysis, I want to run ANOVA analysis on all three IOP groups.

**H0: The mean of ANGPT2 expression level in all three groups of IOP (Normal, Elevated, and High) is equal.**

**Ha: They are not equal.**



Anova\_results <- aov(ANGPT2 ~ Class\_IOP, data=corrTable)

summary(Anova\_results)

Df Sum Sq Mean Sq F value Pr(>F)

Class\_IOP 2 0.0561 0.02806 1 0.377

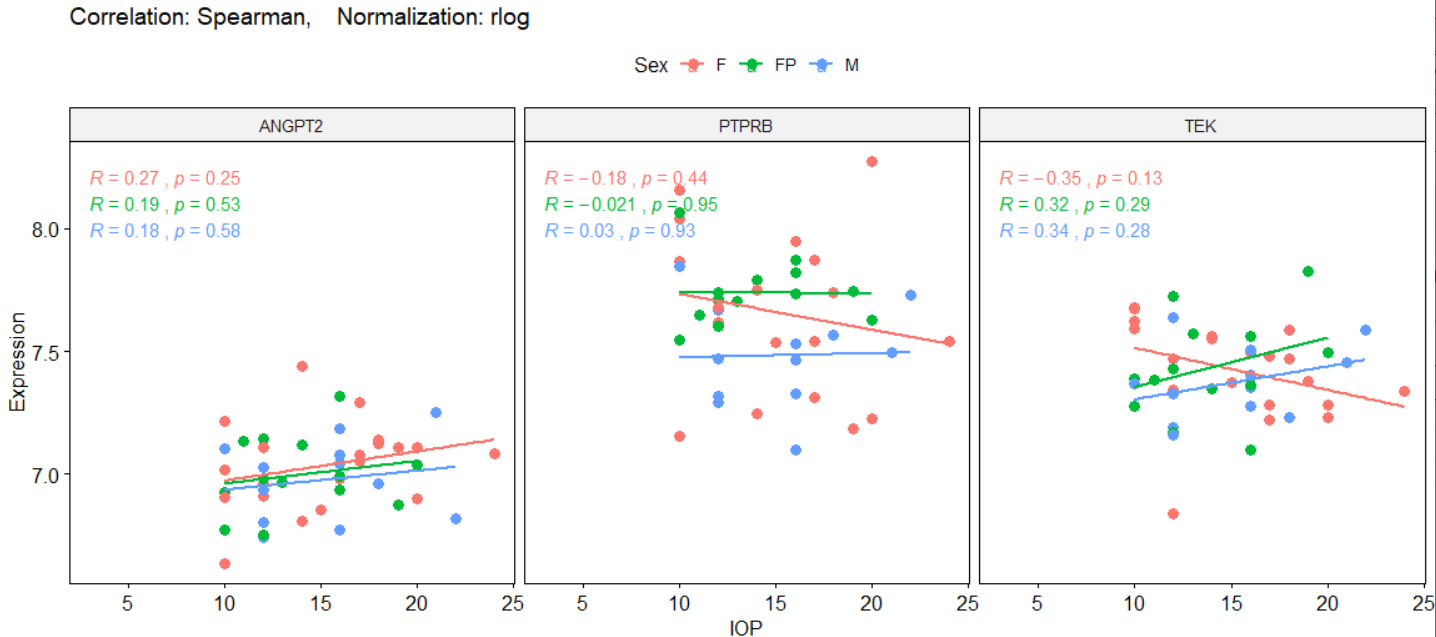
Residuals 42 1.1789 0.02807

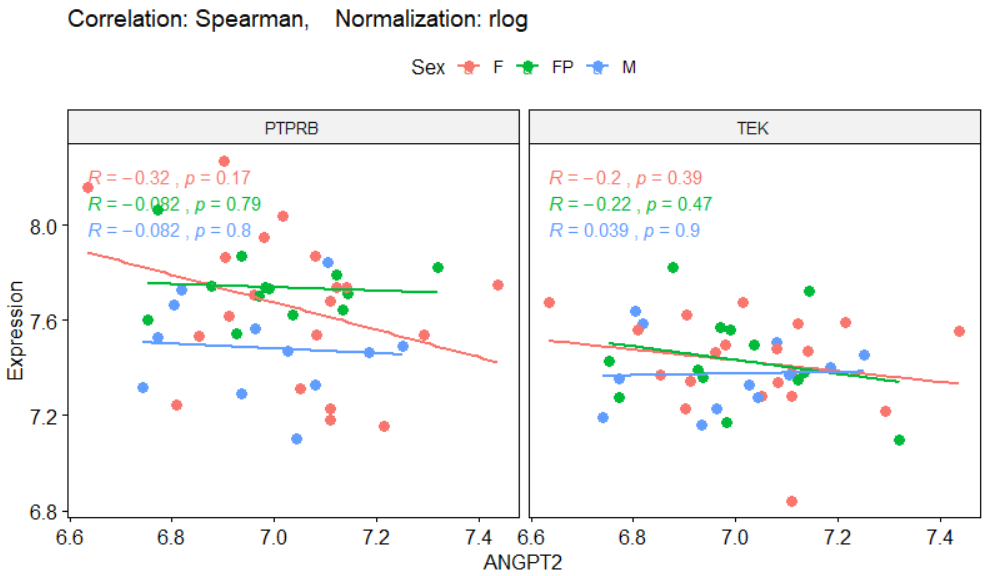
F(2,42) = 1, P-value>0.05

The P-value is greater thatn 0.05, so the null hypothesis is not rejected again. **Therefore, I cannot claime that the expression level of ANGPT2 is significantly different in three IOP groups.**

**Partial Correlation Analysis: Sex Normalization: rlog Correlation: Spearman**

**Sex**





**Partial Correlation Analysis: Age Normalization: rlog Correlation: Spearman**

**Age**

****

**346**

**238**

**183**

**154**

**135**

**113**

**Partial Correlation Analysis: Batch Normalization: rlog Correlation: Spearman**

**Batch**