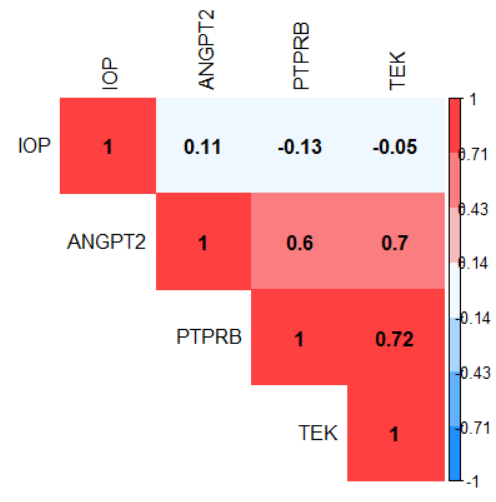
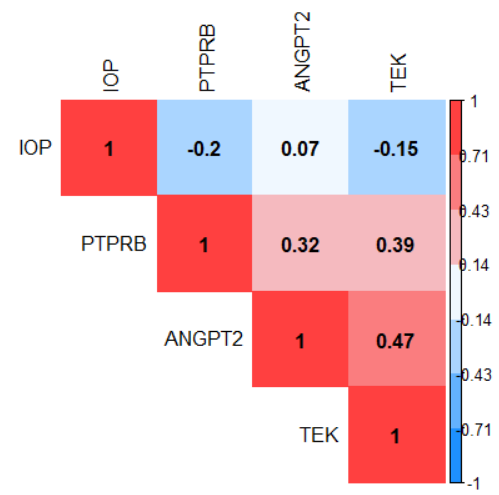
**Report date: March 24 Heatmap and Distribution No Normalization (Real Counts)**

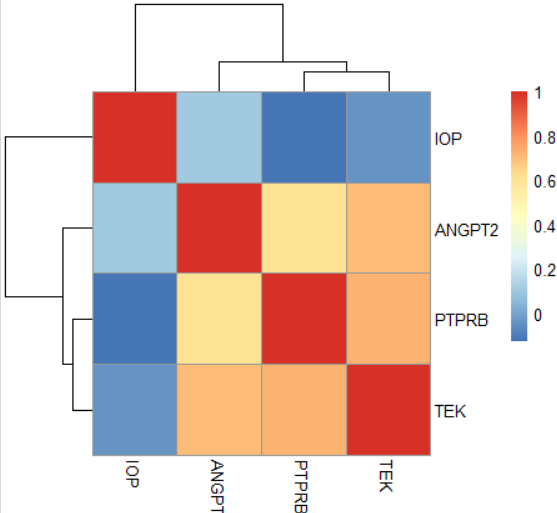
**Correlation method: Pearson**



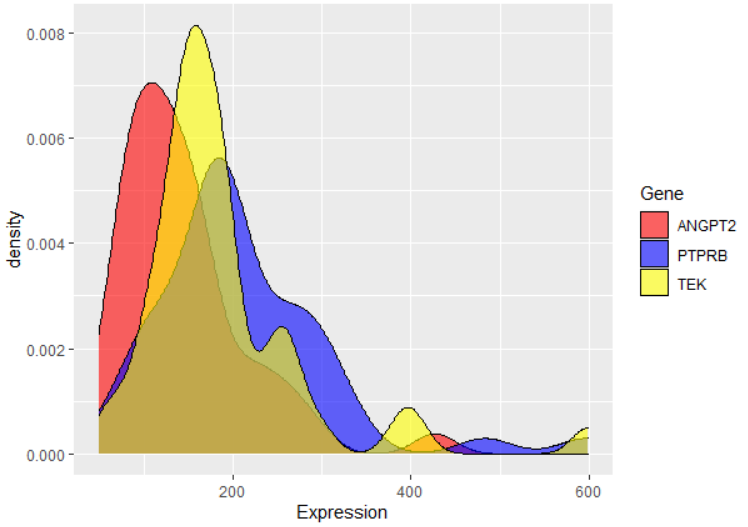
**Correlation method: Spearman**



**Clustering**

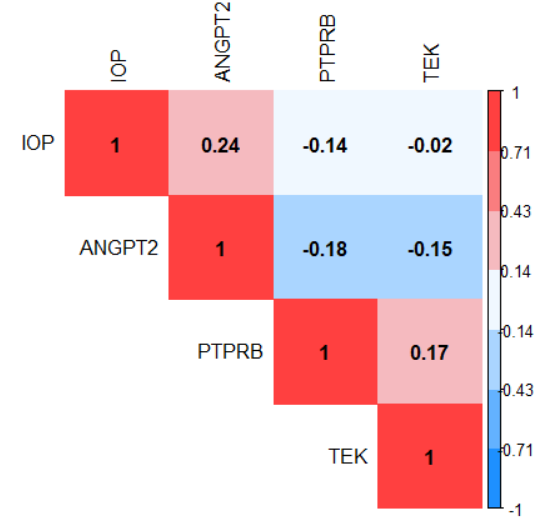


**Distribution**

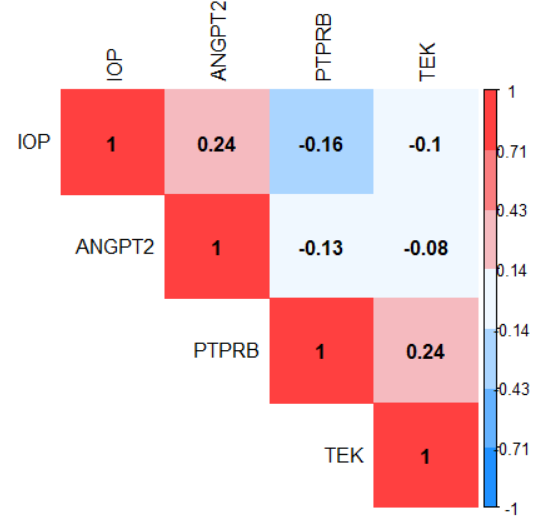


**Report date: March 24 Heatmap and Distribution Normalization method: Log2**

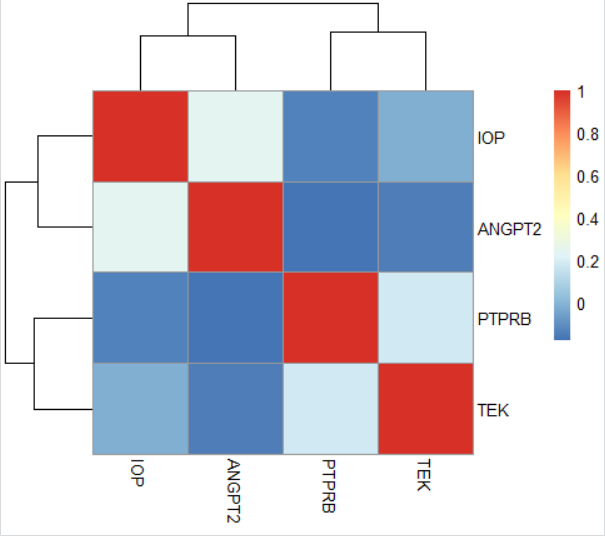
**Correlation method: Pearson**



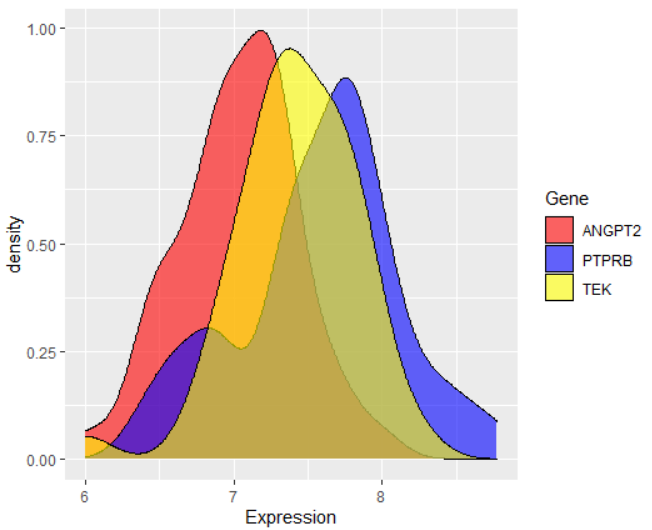
**Correlation method: Spearman**



**Clustering**

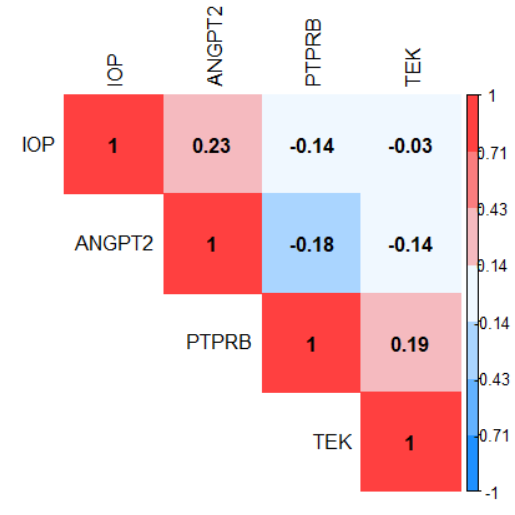


**Distribution**

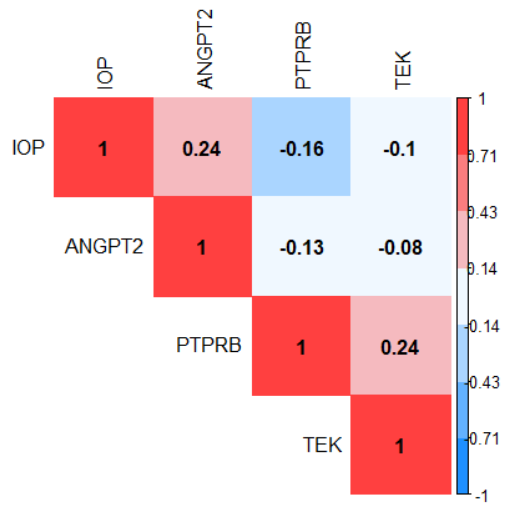


**Report date: March 24 Heatmap and Distribution Normalization method: vst**

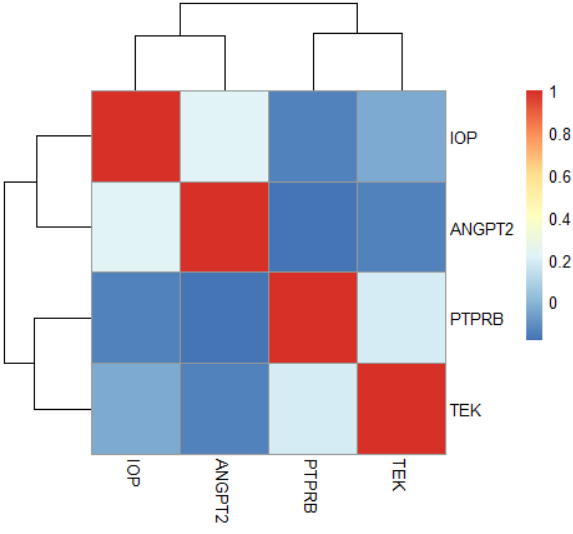
**Correlation method: Pearson**



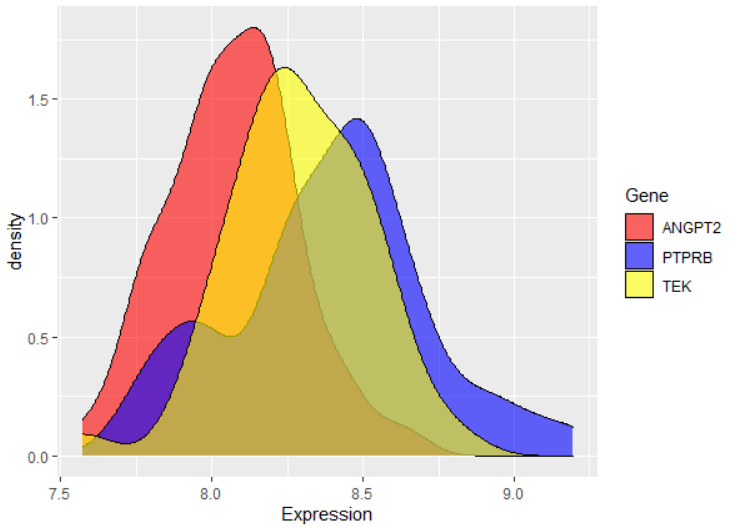
**Correlation method: Spearman**



**Clustering**

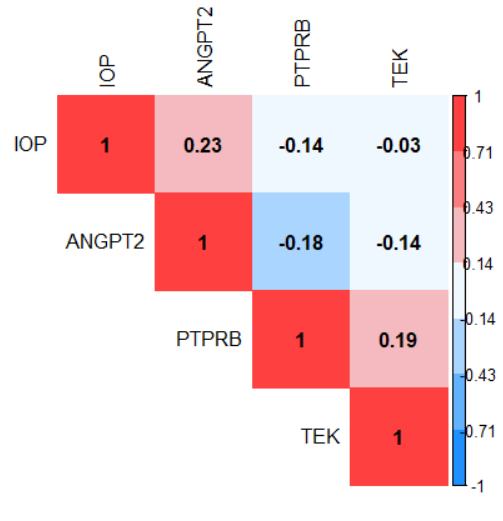


**Distribution**

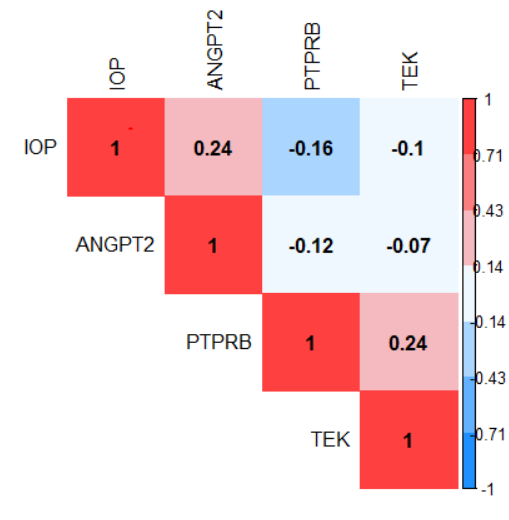


**Report date: March 24 Heatmap and Distribution Normalization method: rlog**

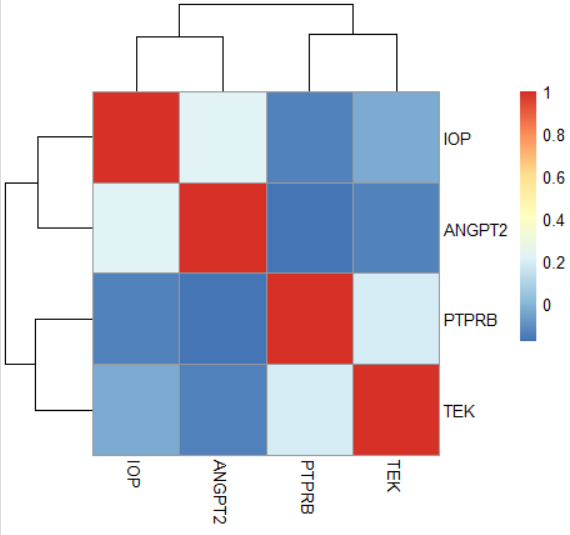
**Correlation method: Pearson**



**Correlation method: Spearman**



**Clustering**



**Distribution**

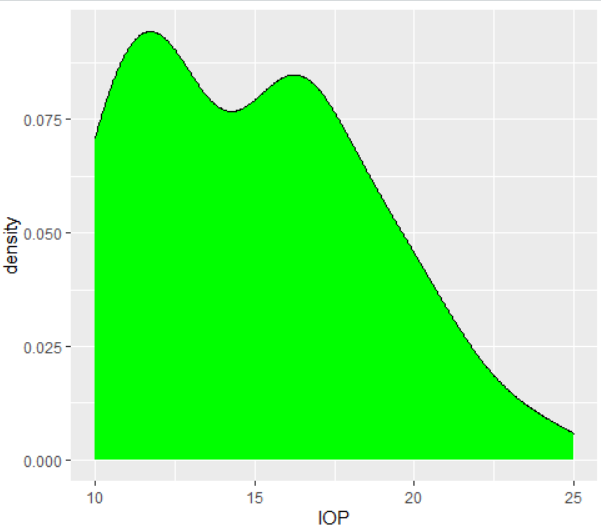


**Heatmap figures results**

Based on the Clustering and Heatmap figures, I found that two genes **PTPRB** and **TEK** act similarly. They have a negative correlation with **IOP** and **ANGPT2**. Also, **IOP** has a weak positive relationship 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 consider the **rlog** function as a normalization method.

Also, between Pearson (linear correlation) method and Spearman (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 believe that ANGPT2 and PTPRB have a negative relationship, PTPRB and TEK have a positive correlation, and ANGPT2 and TEK are uncorrelated.

Based on the density function of IOP, in the figure below, I found that **maybe** its distribution is not normal, and I need to normalize it first, then compare it with the other three genes. There are several ways to check the normality of distribution, like plotting the probability density function (like the figure below), plotting the quartile-quartile plot (Q-Q plot), and do Shapiro-Wilk normality test. Among them, I preferred the reliable Shapiro-Wilk normality test. In the output of this test, the p-value > 0.05, implying that the distribution of the data is not significantly different from the normal distribution. In other words, I can assume the normality for p-value > 0.05. After running this test for three normalized genes and IOP data, I found that for sure, the IOP data has not normal distribution and needs to normalization. In fact, it is right-skewed data. Common transformations of this data include **square root, cube root, and log.**



data: corrTable$ANGPT2

W = 0.9904, p-value = 0.9686 -------- as p-value > 0.05 then ANGPT2 is normal.

data: corrTable$PTPRB

W = 0.97918, p-value = 0.587 -------- as p-value > 0.05 then PTPRB is normal.

data: corrTable$TEK

W = 0.98158, p-value = 0.6841 -------- as p-value > 0.05 then TEK is normal.

data: corrTable$IOP

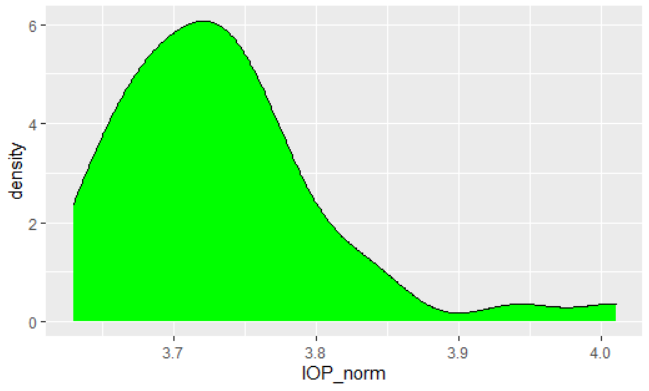
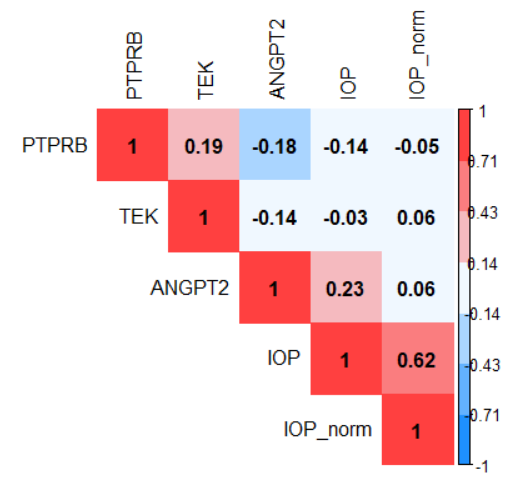
W = 0.93549, p-value = 0.01474 ------- as p-value < 0.05 then IOP is not normal.

1. **Normalization of IOP using rlog**

First, I tried to normalize the IOP with other 32883 genes and used the logarithmic geometry mean method. However, after checking the Shapiro-Wilk normality test, I found that it is not normal again. It has some skewness in the right side. So, as I thought, this is not a suitable method for normalization.

data: corrTable$IOP\_norm

W = 0.88815, p-value = 0.0004137 ------- as p-value < 0.05 then IOP\_norm is not normal.

1. **Normalization of IOP using Standard method**

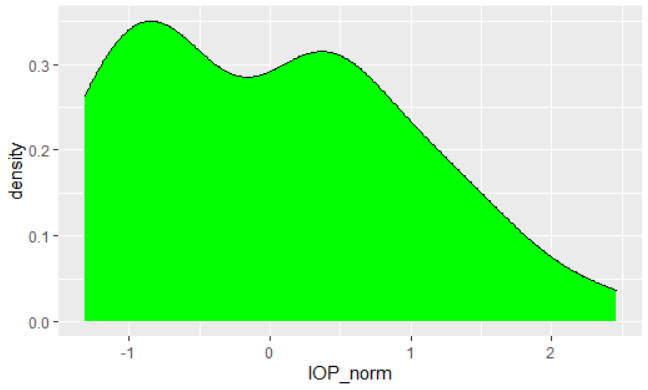
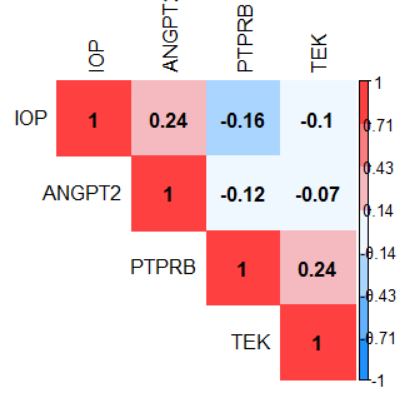
Then uses the standard method for normalization of IOP data to convert its distribution into the normal distribution.

(x - mean(x)) / std(x)

data: corrTable$IOP

W = 0.93549, p-value = 0.01474 ------- as p-value < 0.05 then IOP is not normal.

skewness(corrTable$IOP) = 0.4156063

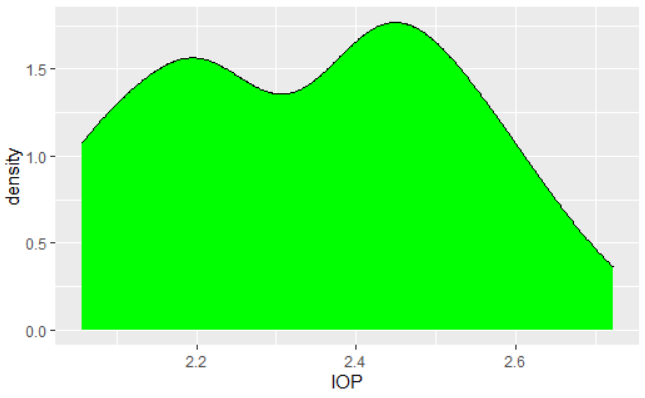
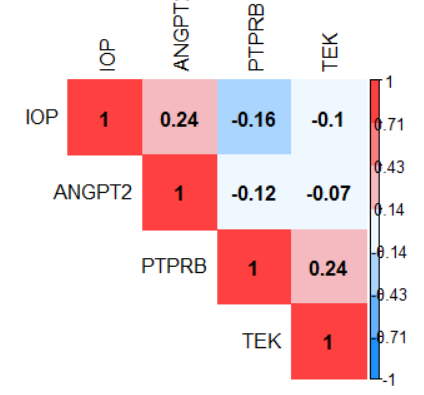
1. **Normalization of IOP using Box-Cox power transformation**

Lambda = -0.1

data: corrTable$IOP

W = 0.93785, p-value = 0.01795 ------- as p-value < 0.05 then IOP is not normal.

skewness(corrTable$IOP) = 0.01833074

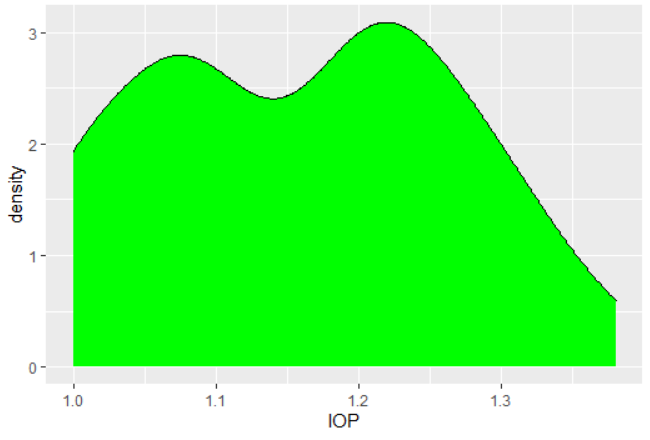
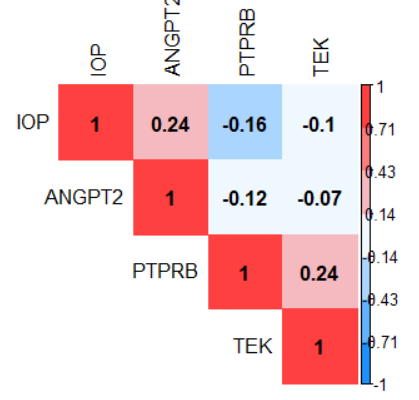
1. **Normalization of IOP using Log transformation**

Log(x), Log2(x), Log10(x)

data: corrTable$IOP

W = 0.93874, p-value = 0.01934 ------- as p-value < 0.05 then IOP is not normal.

skewness(corrTable$IOP) = 0.05192562

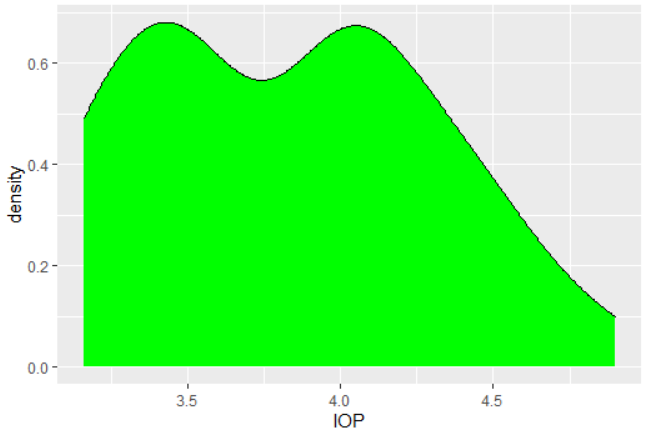
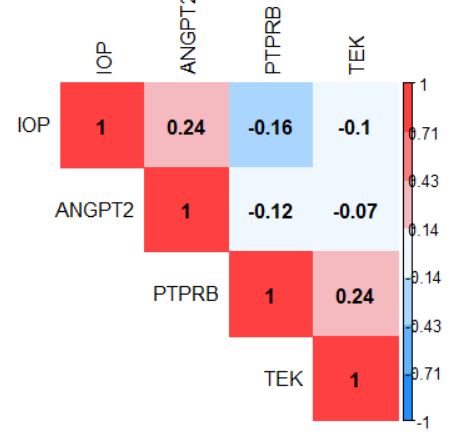
1. **Normalization of IOP using Square root transformation**

sqrt(x)

data: corrTable$IOP

W = 0.93998, p-value = 0.02147 ------- as p-value < 0.05 then IOP is not normal.

skewness(corrTable$IOP) = 0.2272393

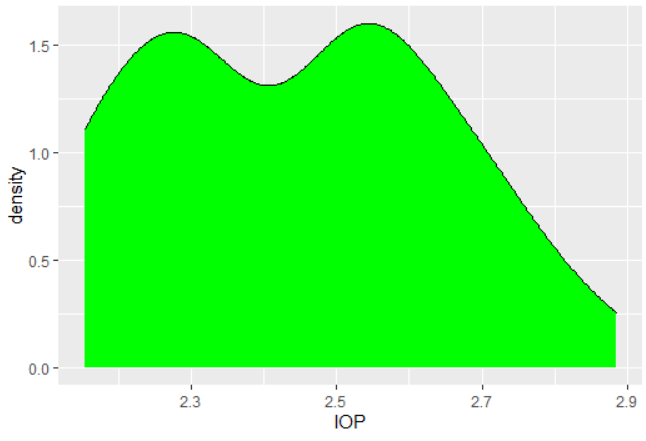
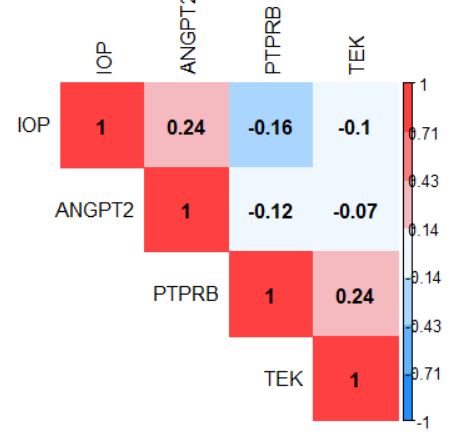
1. **Normalization of IOP using Cube root transformation**

cuberoot(x)

data: corrTable$IOP

W = 0.94018, p-value = 0.02184 ------- as p-value < 0.05 then IOP is not normal.

skewness(corrTable$IOP) = 0.1674028

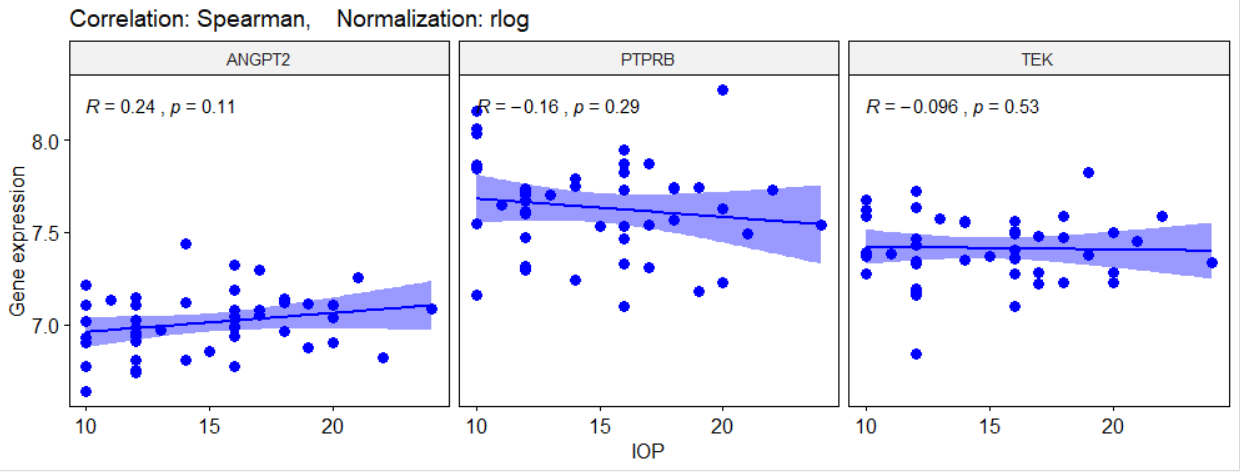
 

**Compare IOP normalization methods**

Among these six normalization methods, none of them can convert the distribution of IOP to an excellent normal distribution. Except for rlog method, the correlation matrix is not changed during normalization. Therefore, we cannot accept the rlog normalization method, as I thought before. Almost all of them skewed right. I have chosen the **Log10 transformation normalization** as the best normalization method. Its density is semi-normal compare to other methods and has relatively low skewness. However, the goal is a correlation calculation between ANGPT2 and IOP. The transformation normalization of IOP does not affect the correlation analysis. Therefore, I have considered **rlog** as a normalization method for genes and **No** normalization for IOP. Also, considered **Spearman** as a correlation method for further analysis.

**Correlation between IOP and ANGPT2**

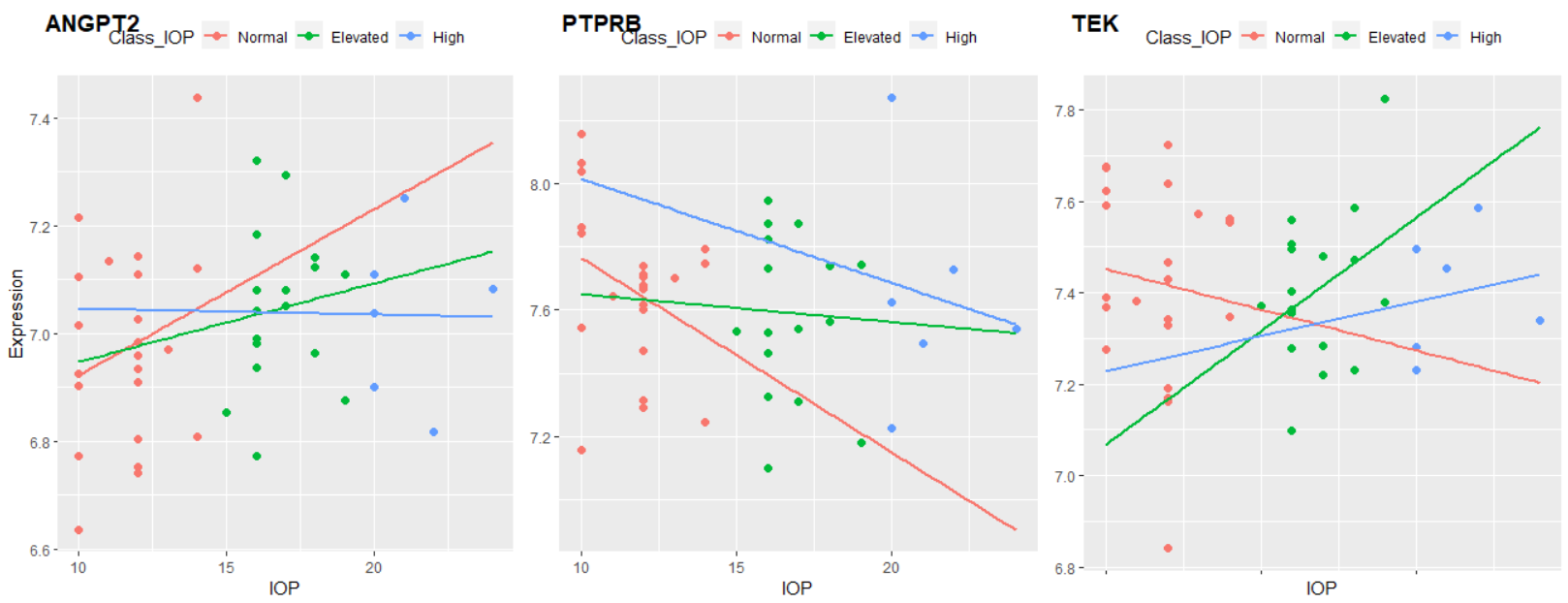
In this section, I want to calculate the **Spearman** correlation of real IOP (without normalization) with ANGPT2 gene expression (based on the **rlog** normalization method). Based on the correlation coefficient R and P-values in the figure below, I found that IOP has a positive relationship with ANGPT2 and a negative relationship with PTPRB and almost not correlated with TEK. However, none of them are significantly correlated. Based on the best R and P-values, I focused on the week relationship of IOP and ANGPT2.

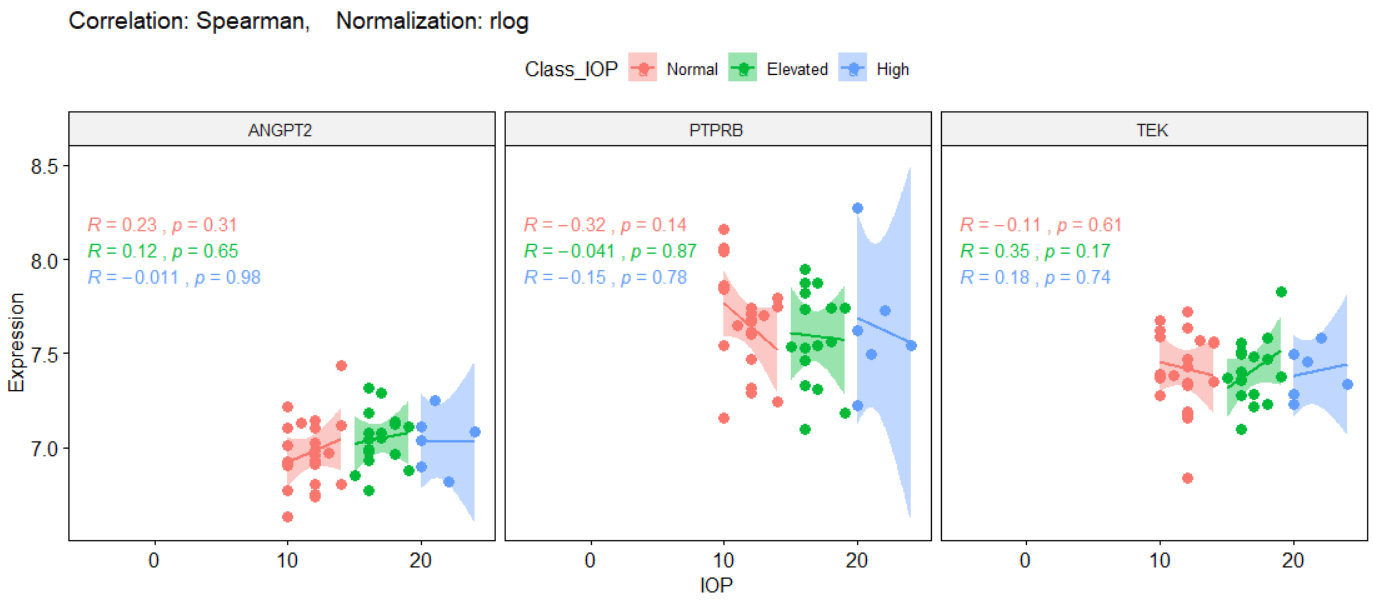


**Partial Correlations**

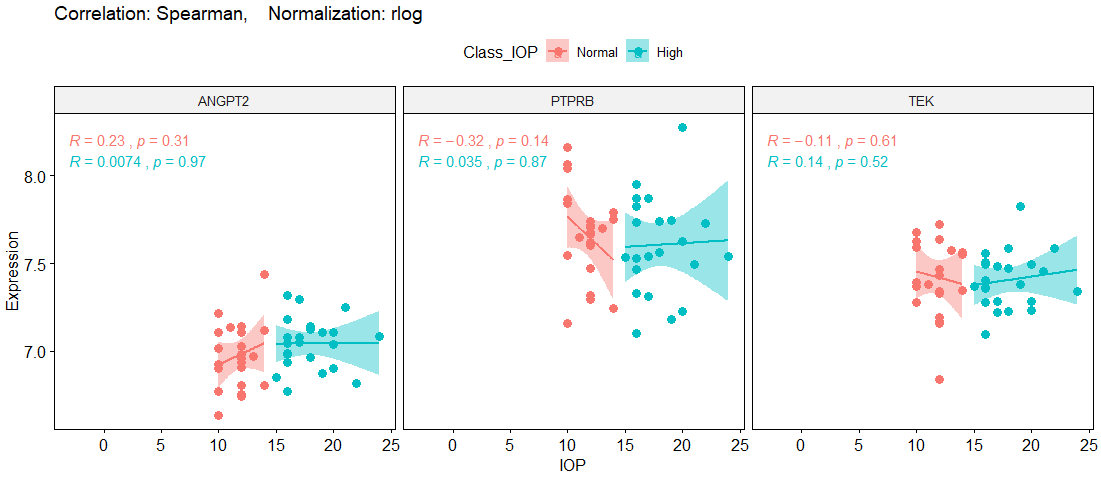
**Partial-Correlation between IOP and ANGPT2 based on IOP subgroups**

In the two figures below, I plot the correlation of three genes with three subgroups of IOP (Normal, Elevated, and High). As I can see, ANGPT2 has the highest positive correlation with Normal\_IOP (R = 0.23) subgroup. Also, PTPRB has the highest negative correlation with Normal\_IOP (R = -0.32) subgroup. It means that the Normal\_IOP subgroup has a major role in the IOP dataset in the association with ANGPT2 and PTPRB.





Although I am looking for a high correlation between the High\_IOP subgroup and ANGPT2, however, as there are low samples in the high\_IOP subgroup, it is not a significant relationship. Therefore, I decided to classified the IOP feature into two subgroups: Normal and High. In the figure below, I look for the correlation of IOP subgroups with three genes. In this scenario, half of the samples are High\_IOP; around 50% of samples (23 samples). However, the High\_IOP subgroup is not correlated with ANGPT2 and PTPRB again. I mean, Normal\_IOP has an important role in the correlation of IOP with ANGPT2 or PTPRB again.



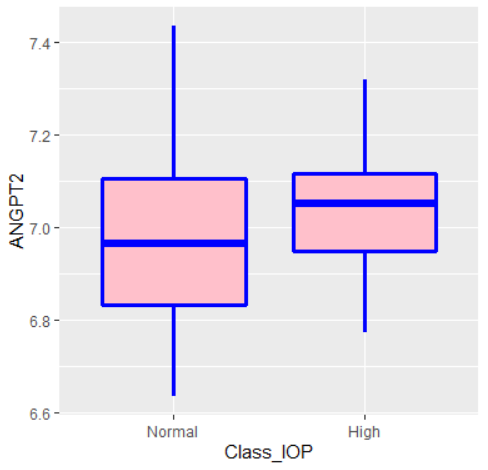
As a conclusion and based on figures above, ANGPT2 has a week-correlation, but not significant, with whole IOP data (R=0.24, P-value=0.11), which related to the Normal\_IOP (R=0.23, P-value=0.31). However, the ANGPT2 gene has the maximum week-correlation with IOP, among other genes. Therefore, we need to look at this gene deeply. Maybe other features like Age, Sex, or Batch give some clues. To discuss the correlation between ANGPT2 and IOP subgroups statistically, I did a t-test and ANOVA analysis too.

**Analysis of t-test for IOP subgroups**

In this analysis, I want to test that the mean expression level of ANGPT2 is equal in two separate groups of Normal\_IOP and High\_IOP (with almost the same sample sizes). If not, then the expression level of ANGPT2 in each subgroup of IOP are not similar together, and the expression level is varied in different subgroups of IOP.

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

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



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)

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

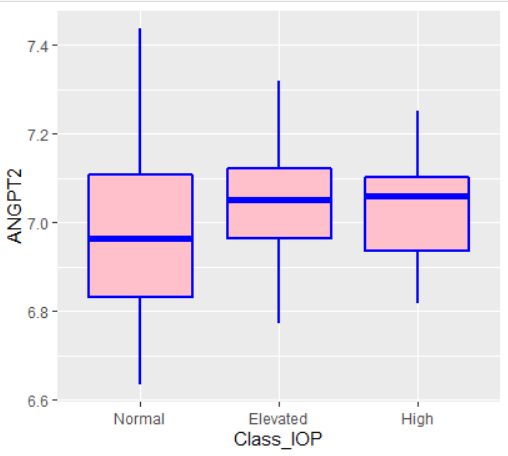
The P-value of the test is 0.1656, which is higher than the significant level, 0.05. Therefore, I cannot reject the Null-Hypothesis to accept the alternative hypothesis. I conclude that the average expression level of the ANGPT2 gene in samples with Normal\_IOP is not significantly different from the samples with High\_IOP; the expression level of the ANGPT2 gene is almost the same in two groups (6.97 ≈ 7.04). It means that different IOP subgroups have not an important role in the expression level of ANGPT2.

**Analysis of ANOVA for IOP subgroups**

As in the real state, we have three subgroups of IOP, so we need to run the ANOVA analysis instead of a t-test for them.

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

**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 higher than 0.05, so the null hypothesis is not rejected again. Therefore, I cannot claim that the expression level of ANGPT2 is significantly varied in three different IOP groups.

**Partial-Correlation between IOP and ANGPT2 based on Age feature**

In this section, I am going to calculate the correlation between IOP and ANGPT2 based on different Age subgroups. I classified samples into four subgroups: **Adolescent** (113<=Age<135), **Adult** (135<=Age<154), **Middle\_Aged**(154<=Age<183), and **Aged** (183<=Age<=346). Then, plot the correlation based on each subgroups:

****

346

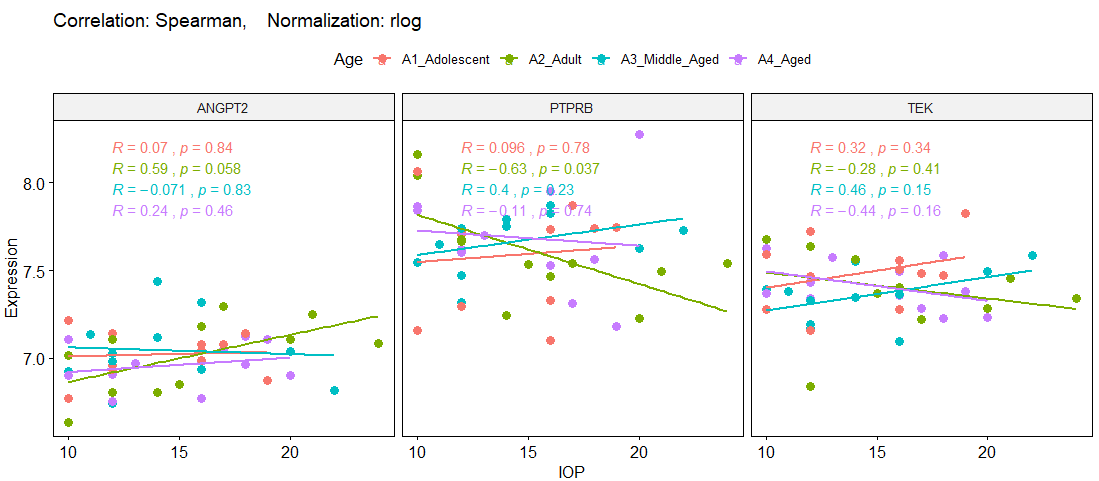
238

183

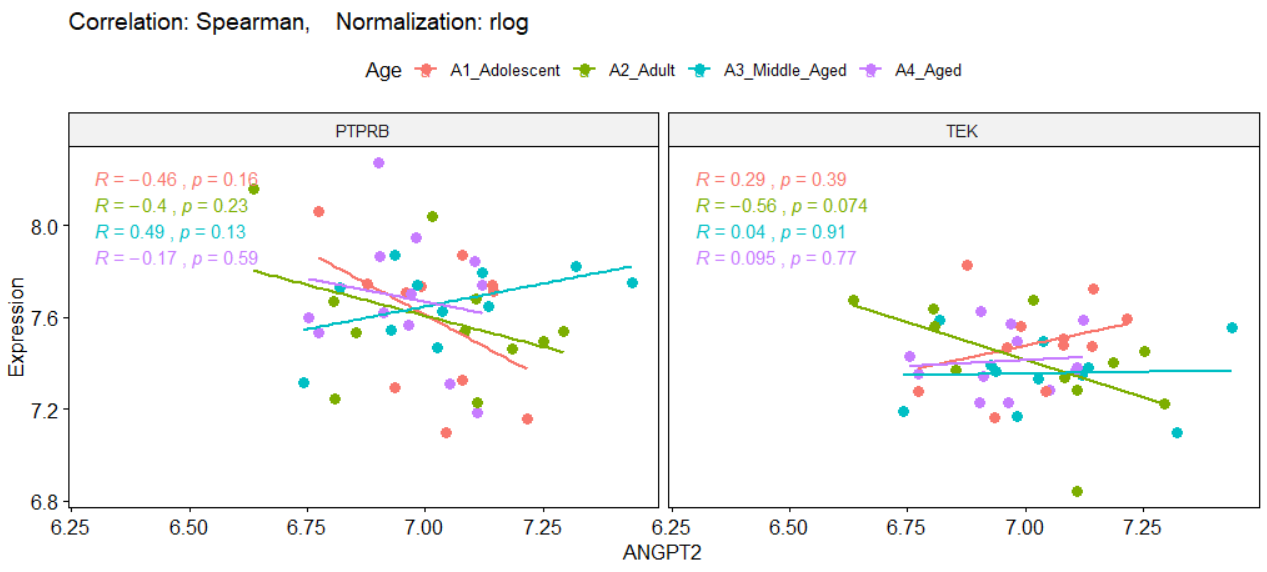
154

135

113



**Good news:** Based on the results, I can see there is a **significant relationship** in the **Adult** subgroup between IOP and ANGPT2 (R=+0.59, P-value=0.05), and PTPRB (R=-0.63, P-value=0.03). Also, there is semi-significant relationship between ANGPT2 and TEK genes in this subgroup (Adult); figure below.



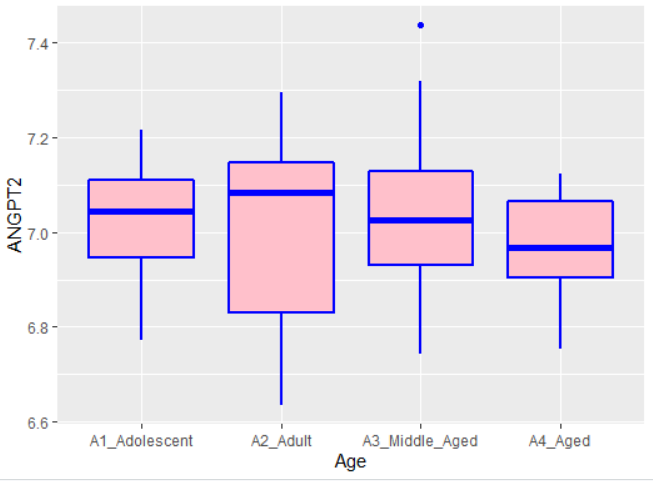
Therefore, I am currious about the Adult subgroup. To test these observation, I am going to do a t-test and ANOVA analysis.

**Analysis of ANOVA and lm considering Age**

There is four subgroups of Age, and I run ANOVA analysis in for ANGPT2 expression along these subgroups.

**H0:** The mean of ANGPT2 expression level in all four subgroups of Age is equal together.

**Ha:** They are not equal.



Anova\_results <- aov(ANGPT2 ~ Age, corrTable); summary(Anova\_results)

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

Age 3 0.0415 0.01384 0.476 0.701

Residuals 41 1.1935 0.02911

Anova\_results <- aov(ANGPT2 ~ IOP + Age, corrTable); summary(Anova\_results)

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

IOP 1 0.0642 0.06420 2.290 0.138

Age 3 0.0494 0.01647 0.587 0.627

Residuals 40 1.1214 0.02804

Linear\_model <- lm(ANGPT2 ~ IOP + Age, corrTable); summary(Linear\_model)

Call:

lm(formula = ANGPT2 ~ IOP + Age, data = corrTable)

Residuals:

Min 1Q Median 3Q Max

-0.33588 -0.07945 0.00237 0.10007 0.44388

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 6.9309303 0.1335659 51.891 <2e-16 \*\*\*

IOP 0.0109180 0.0067533 1.617 0.113

Age -0.0005102 0.0005813 -0.878 0.385

---

Residual standard error: 0.1655 on 42 degrees of freedom

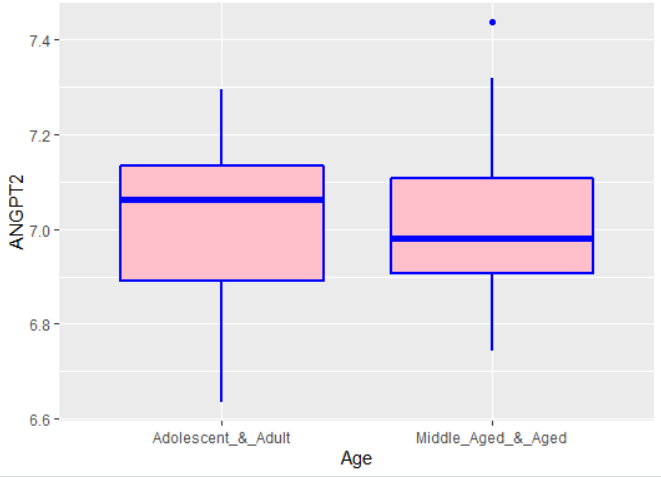
Multiple R-squared: 0.06906, Adjusted R-squared: 0.02473

F-statistic: 1.558 on 2 and 42 DF, p-value: 0.2225

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 varied in four different Age groups.

Then, I decided to devide the Age into just two subgroups as belows and run the t-test. In eanch division, I followed a logical rull: 1) devided ages half by half (Adolocent and Adults in first group and Middle-Ageed and Aged in the second group). 2) separate Adults from others. 3) separate Aged (as the oldest samples) from others. 4) separate Adolocent (as the youngest samples) from others.

**Division 1): Devided ages half by half (Adolocent and Adults in first group and Middle-Ageed and Aged in the second group)**



Anova\_results <- aov(ANGPT2 ~ Age, corrTable\_Age2); summary(Anova\_results)

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

Age 1 0.003 0.002999 0.105 0.748

Residuals 43 1.232 0.028651

Anova\_results <- aov(ANGPT2 ~ IOP + Age, corrTable\_Age2); summary(Anova\_results)

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

IOP 1 0.0642 0.06420 2.308 0.136

Age 1 0.0024 0.00239 0.086 0.771

Residuals 42 1.1684 0.02782

t.test(ANGPT2 ~ Age, corrTable\_Age2, mu=0, alt="two.sided", conf=0.95, var.eq=F, paired=F)

Welch Two Sample t-test

data: ANGPT2 by Age

t = 0.32342, df = 42.84, p-value = 0.748

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

95 percent confidence interval:

-0.08551491 0.11817765

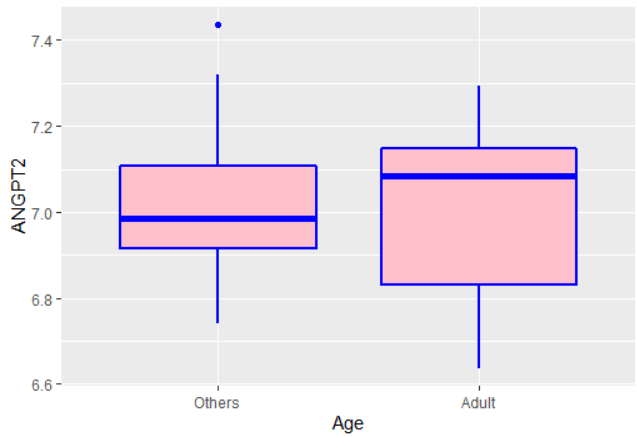
sample estimates:

mean in group Adolescent\_&\_Adult: 7.017332

mean in group Middle\_Aged\_&\_Aged: 7.001000

In all tests and analysis P-values are greater than 0.05 and I cannot refect the null hypothesis.

**Division 2): separate Adults from others**



Anova\_results <- aov(ANGPT2 ~ Age, corrTable\_Age2); summary(Anova\_results)

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

Age 1 0.0003 0.00027 0.009 0.923

Residuals 43 1.2347 0.02872

Anova\_results <- aov(ANGPT2 ~ IOP + Age, corrTable\_Age2); summary(Anova\_results)

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

IOP 1 0.0642 0.06420 2.303 0.137

Age 1 0.0001 0.00011 0.004 0.951

Residuals 42 1.1707 0.02787

t.test(ANGPT2 ~ Age, corrTable\_Age2, mu=0, alt="two.sided", conf=0.95, var.eq=F, paired=F)

Welch Two Sample t-test

data: ANGPT2 by Age

t = -0.082964, df = 13.696, p-value = 0.9351

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

95 percent confidence interval:

-0.1534463 0.1420401

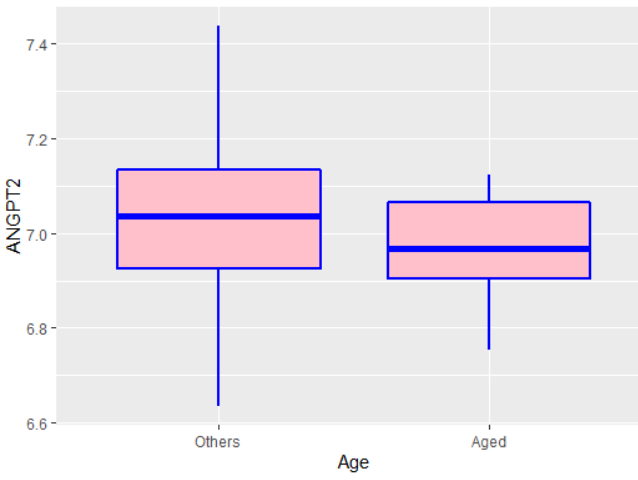
sample estimates:

mean in group Others mean in group Adult

7.007590 7.013293

In all tests and analysis P-values are greater than 0.05 and I cannot refect the null hypothesis.

**Division 3): separate Aged (as the oldest samples) from others**



Anova\_results <- aov(ANGPT2 ~ Age, corrTable\_Age2); summary(Anova\_results)

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

Age 1 0.0361 0.03613 1.296 0.261

Residuals 43 1.1989 0.02788

Anova\_results <- aov(ANGPT2 ~ IOP + Age, corrTable\_Age2); summary(Anova\_results)

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

IOP 1 0.0642 0.06420 2.384 0.130

Age 1 0.0397 0.03969 1.474 0.232

Residuals 42 1.1311 0.02693

t.test(ANGPT2 ~ Age, corrTable\_Age2, mu=0, alt="two.sided", conf=0.95, var.eq=F, paired=F)

Welch Two Sample t-test

data: ANGPT2 by Age

t = 1.3563, df = 28.837, p-value = 0.1855

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

95 percent confidence interval:

-0.03257421 0.16072912

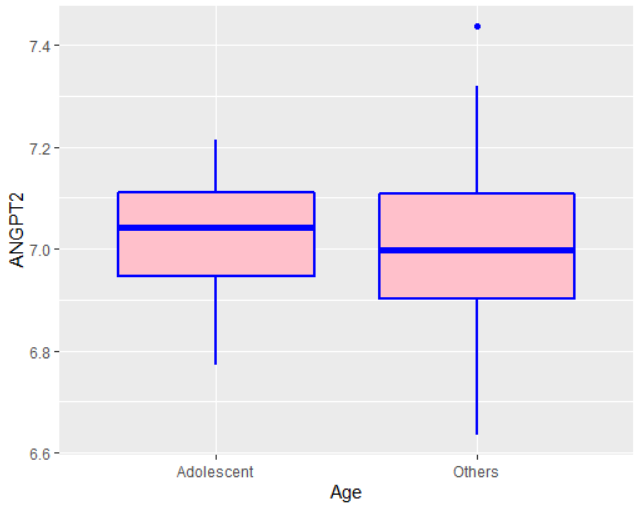
sample estimates:

mean in group Others mean in group Aged

7.026072 6.961994

In all tests and analysis P-values are greater than 0.05 and I cannot refect the null hypothesis.

**Division 4): separate Adolocent (as the youngest samples) from others**



Anova\_results <- aov(ANGPT2 ~ Age, data=corrTable\_Age2)

summary(Anova\_results)

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

Age 1 0.0022 0.002233 0.078 0.782

Residuals 43 1.2328 0.028669

Anova\_results <- aov(ANGPT2 ~ IOP + Age, data=corrTable\_Age2)

summary(Anova\_results)

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

IOP 1 0.0642 0.06420 2.312 0.136

Age 1 0.0045 0.00452 0.163 0.689

Residuals 42 1.1663 0.02777

t.test(ANGPT2 ~ Age, data=corrTable\_Age2, mu=0, alt="two.sided", conf=0.95, var.eq=F, paired=F)

Welch Two Sample t-test

data: ANGPT2 by Age

t = 0.32893, df = 23.455, p-value = 0.7451

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

95 percent confidence interval:

-0.08658991 0.11937447

sample estimates:

mean in group Adolescent mean in group Others

7.021370 7.004977

In all tests and analysis P-values are greater than 0.05 and I cannot refect the null hypothesis.

All analysis above indicates that there is no significant relation between IOP and ANGPT2 based on different subgroups of Age feature.

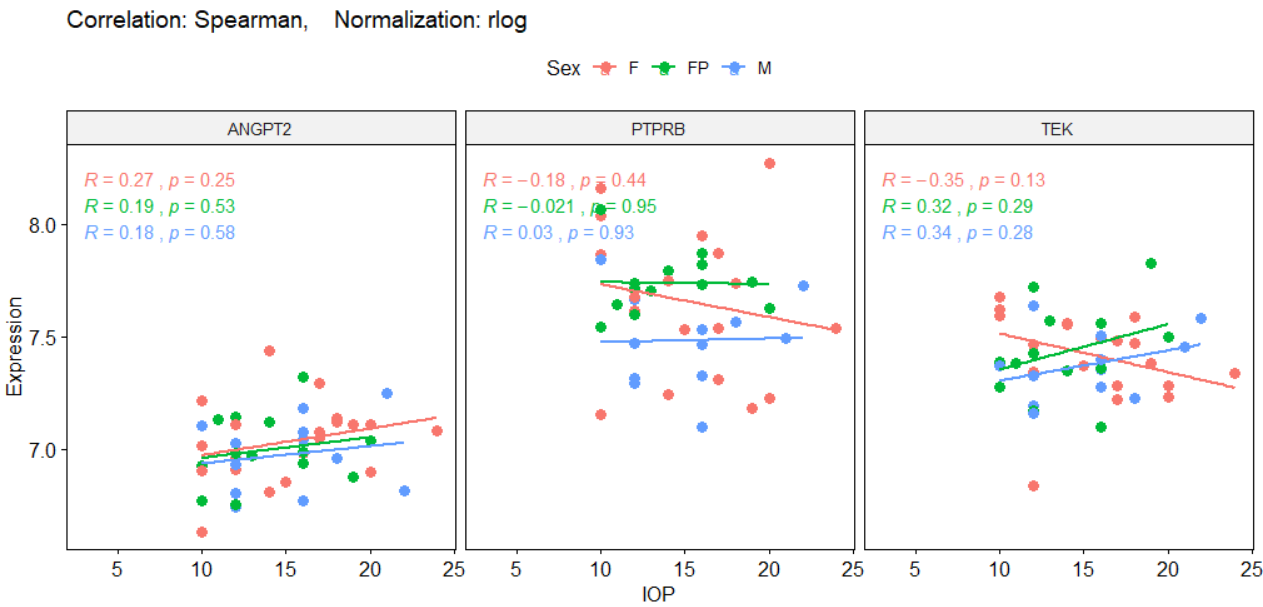
**Partial-Correlation between IOP and ANGPT2 based on Sex feature**

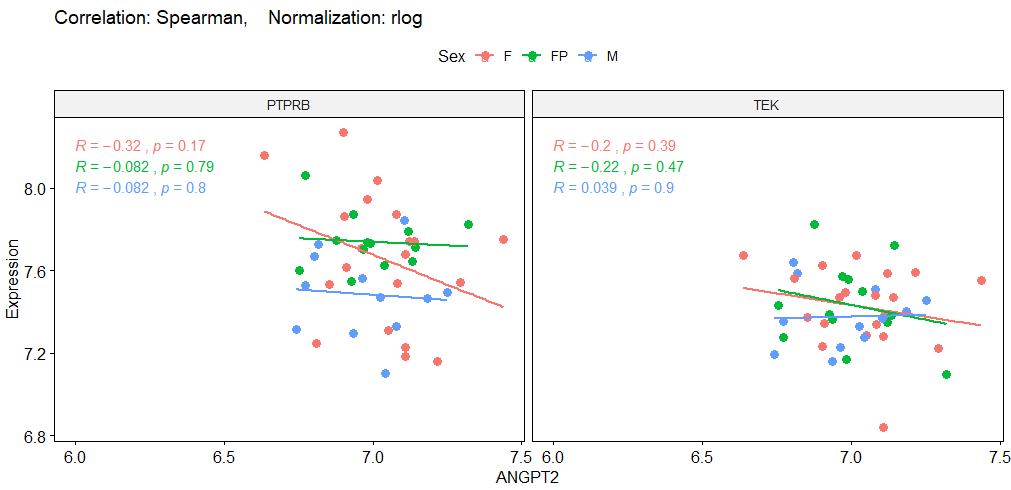
In this section, I am going to calculate the correlation between IOP and ANGPT2 based on different Sex subgroups. Among 45 samples, 12 (26%) of them are Male and 33 (73%) are Female. Aslo, there are 13 (29%)Pregnant samples.

summary(corrTable$Sex)

F FP M

20 13 12





There is no significant correlation among Sex subgroups and genes. Almost all three Sex subgroups have similar role in the relationship of IOP and ANGPT2.

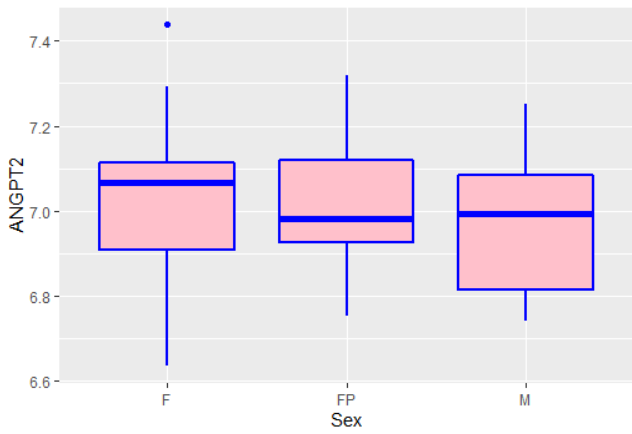
Now, I am going to do a t-test and ANOVA analysis.

**Analysis of ANOVA and lm considering Sex**

There is three subgroups of Sex, and I run ANOVA analysis in for ANGPT2 expression along these subgroups.

**H0:** The mean of ANGPT2 expression level in all three subgroups of Sex is equal together.

**Ha:** They are not equal.



Anova\_results <- aov(ANGPT2 ~ Sex, corrTable); summary(Anova\_results)

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

Sex 2 0.0286 0.01430 0.498 0.611

Residuals 42 1.2064 0.02872

Anova\_results <- aov(ANGPT2 ~ IOP + Sex, corrTable); summary(Anova\_results)

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

IOP 1 0.0642 0.06420 2.300 0.137

Sex 2 0.0261 0.01306 0.468 0.630

Residuals 41 1.1447 0.02792

Linear\_model <- lm(ANGPT2 ~ IOP + Sex, corrTable); summary(Linear\_model)

Call:

lm(formula = ANGPT2 ~ IOP + Sex, data = corrTable)

Residuals:

Min 1Q Median 3Q Max

-0.34689 -0.09229 -0.00910 0.09460 0.41424

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 6.880054 0.111283 61.825 <2e-16 \*\*\*

IOP 0.010219 0.006874 1.487 0.145

SexFP -0.025231 0.060223 -0.419 0.677

SexM -0.058909 0.061013 -0.966 0.340

---

Residual standard error: 0.1671 on 41 degrees of freedom

Multiple R-squared: 0.07313, Adjusted R-squared: 0.005313

F-statistic: 1.078 on 3 and 41 DF, p-value: 0.3689

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 varied in three different Sex subgroups.

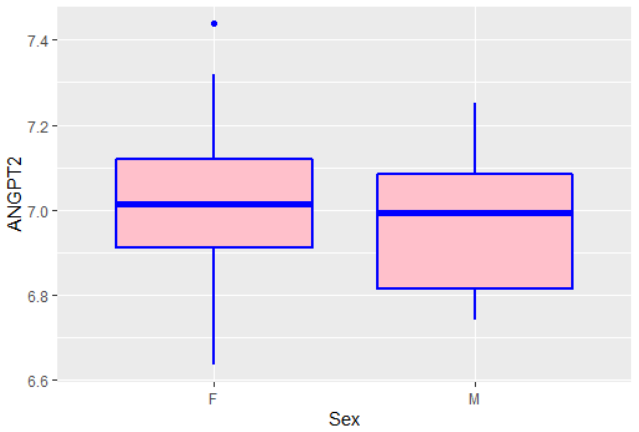
Then, I decided to devide the Sex group into just two subgroups as belows and run the t-test. Division 1) convert FP to F. Division 2) convert F and M to NonP.

**Division 1): Convert FP to F. So, we have just F and M**

summary(corrTable\_Sex2$Sex)

F M

33 12



Anova\_results <- aov(ANGPT2 ~ Sex, corrTable\_Sex2); summary(Anova\_results)

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

Sex 1 0.0167 0.01675 0.591 0.446

Residuals 43 1.2183 0.02833

Anova\_results <- aov(ANGPT2 ~ IOP + Sex, corrTable\_Sex2); summary(Anova\_results)

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

IOP 1 0.0642 0.06420 2.346 0.133

Sex 1 0.0212 0.02122 0.775 0.384

Residuals 42 1.1496 0.02737

Anova\_results <- aov(ANGPT2 ~ IOP + Sex + Age, corrTable\_Sex2); summary(Anova\_results)

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

IOP 1 0.0642 0.06420 2.279 0.139

Sex 1 0.0212 0.02122 0.753 0.391

Age 3 0.0511 0.01703 0.605 0.616

Residuals 39 1.0985 0.02817

t.test(ANGPT2 ~ Sex, corrTable\_Sex2, mu=0, alt="two.sided", conf=0.95, var.eq=F, paired=F)

Welch Two Sample t-test

data: ANGPT2 by Sex

t = 0.77303, df = 19.763, p-value = 0.4487

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

95 percent confidence interval:

-0.07418849 0.16144413

sample estimates:

mean in group F mean in group M

7.020618 6.976991

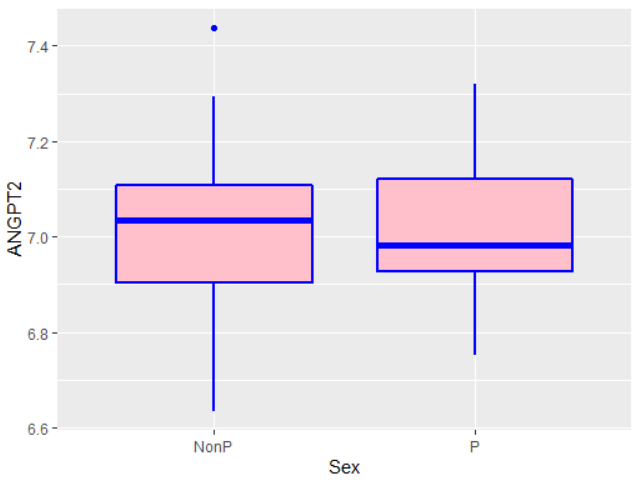
In all tests and analysis P-values are greater than 0.05 and I cannot reject the null hypothesis.

**Division 2): Convert F and M to NonP. So, we have just Prognant (P) and Non-Prognant (NonP)**

summary(corrTable\_Sex2$Sex)

NonP P

32 13



Anova\_results <- aov(ANGPT2 ~ Sex, corrTable\_Sex2); summary(Anova\_results)

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

Sex 1 0.0026 0.002578 0.09 0.766

Residuals 43 1.2324 0.028661

Anova\_results <- aov(ANGPT2 ~ IOP + Sex, corrTable\_Sex2); summary(Anova\_results)

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

IOP 1 0.0642 0.06420 2.303 0.137

Sex 1 0.0001 0.00009 0.003 0.955

Residuals 42 1.1707 0.02787

Anova\_results <- aov(ANGPT2 ~ IOP + Sex + Age, corrTable\_Sex2); summary(Anova\_results)

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

IOP 1 0.0642 0.06420 2.254 0.141

Sex 1 0.0001 0.00009 0.003 0.956

Age 3 0.0600 0.01999 0.702 0.557

Residuals 39 1.1108 0.02848

t.test(ANGPT2 ~ Sex, corrTable\_Sex2, mu=0, alt="two.sided", conf=0.95, var.eq=F, paired=F)

Welch Two Sample t-test

data: ANGPT2 by Sex

t = 0.31355, df = 24.603, p-value = 0.7565

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

95 percent confidence interval:

-0.09308437 0.12648488

sample estimates:

mean in group NonP mean in group P

7.013809 6.997109

In all tests and analysis P-values are greater than 0.05 and I cannot reject the null hypothesis.

All analysis above indicates that there is no significant relation between IOP and ANGPT2 based on different subgroups of Sex feature.

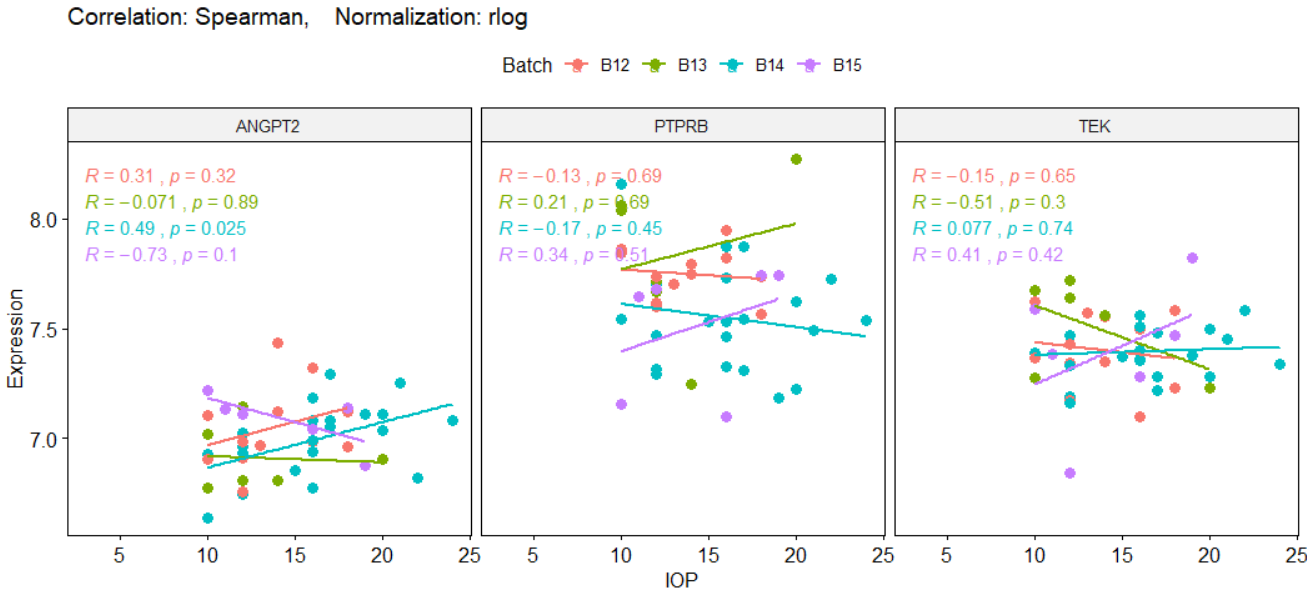
**Partial-Correlation between IOP and ANGPT2 based on Batch feature**

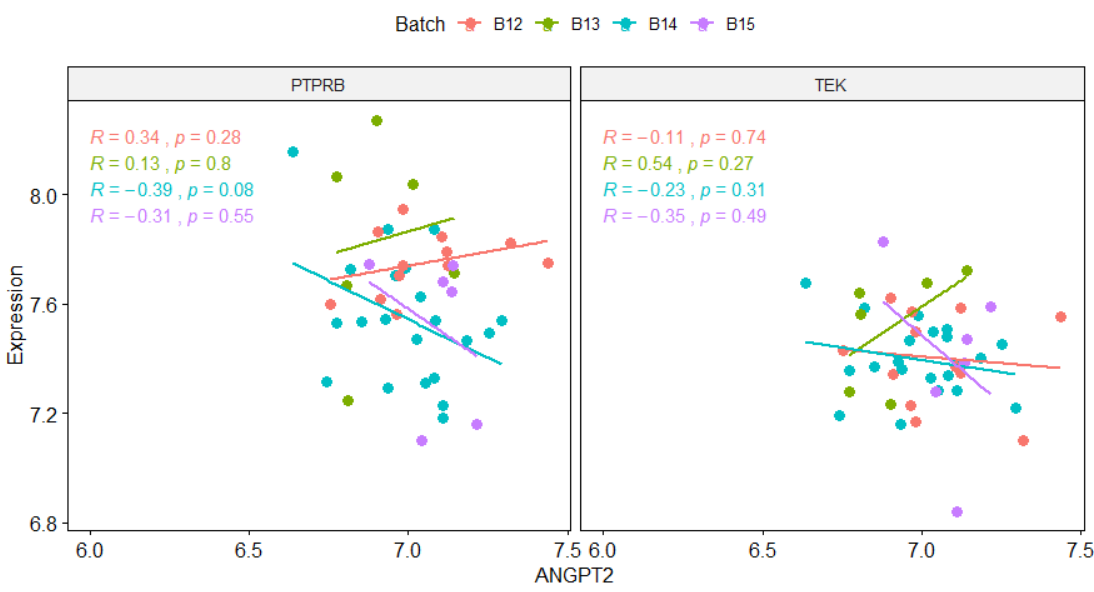
In this section, I am going to calculate the correlation between IOP and ANGPT2 based on different Batch subgroups. Our 45 samples have four batch numbers: B12 (12 samples), B13 (6 samples), B14 (21 samples), and B15 (6 samples).

> summary(corrTable$Batch)

B12 B13 B14 B15

12 6 21 6





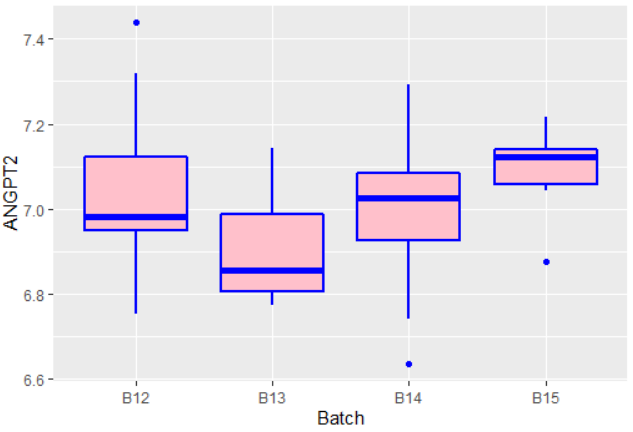
**Good news:** Based on the results, I can see there is a **significant relationship** in the **B14** subgroup between ANGPT2 and IOP (R=+0.49, P-value=0.025), and PTPRB (R=-0.39, P-value=0.08). Now, I am going to do a t-test and ANOVA analysis.

**Analysis of ANOVA and lm considering Batch**

There is four subgroups of Batch, and I run ANOVA analysis in for ANGPT2 expression along these subgroups.

**H0:** The mean of ANGPT2 expression level in all three subgroups of Batch is equal together.

**Ha:** They are not equal.



Anova\_results <- aov(ANGPT2 ~ Batch, corrTable); summary(Anova\_results)

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

Batch 3 0.120 0.0400 1.471 0.237

Residuals 41 1.115 0.0272

Anova\_results <- aov(ANGPT2 ~ IOP + Batch, corrTable); summary(Anova\_results)

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

IOP 1 0.0642 0.06420 2.468 0.124

Batch 3 0.1301 0.04336 1.666 0.190

Residuals 40 1.0407 0.02602

Anova\_results <- aov(ANGPT2 ~ IOP + Batch + Sex + Age, corrTable); summary(Anova\_results)

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

IOP 1 0.0642 0.06420 2.587 0.117

Batch 3 0.1301 0.04336 1.747 0.175

Sex 2 0.0200 0.00999 0.403 0.672

Age 3 0.1522 0.05073 2.044 0.125

Residuals 35 0.8685 0.02482

Linear\_model <- lm(ANGPT2 ~ IOP + Batch, corrTable); summary(Linear\_model)

Call:

lm(formula = ANGPT2 ~ IOP + Batch, data = corrTable)

Residuals:

Min 1Q Median 3Q Max

-0.28542 -0.09903 -0.00121 0.08289 0.38699

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 6.885063 0.106772 64.484 <2e-16 \*\*\*

IOP 0.011808 0.006988 1.690 0.0989 .

BatchB13 -0.130901 0.080821 -1.620 0.1132

BatchB14 -0.082366 0.060811 -1.354 0.1832

BatchB15 0.031996 0.080754 0.396 0.6940

---

Residual standard error: 0.1613 on 40 degrees of freedom

Multiple R-squared: 0.1573, Adjusted R-squared: 0.07304

F-statistic: 1.867 on 4 and 40 DF, p-value: 0.1352

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 varied in four different Batch subgroups.

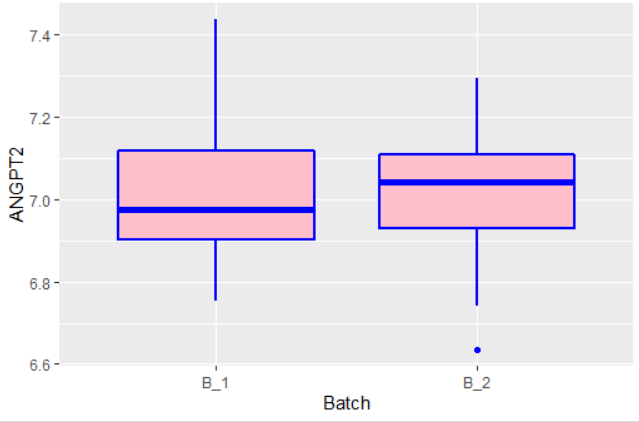
Then, I decided to devide the Batch group into just two subgroups as belows and run the t-test. Division 1) devided Batch half by half (B12 and B13 in first group and B14 and B15 in the second group). 2) separate B14 from others.

**Division 1): Devided Batch half by half (B12 and B13 in first group and B14 and B15 in the second group)**

summary(corrTable\_Batch2$Batch)

B\_1 B\_2

18 27



Anova\_results <- aov(ANGPT2 ~ Batch, corrTable\_Batch2); summary(Anova\_results)

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

Batch 1 0.002 0.001991 0.069 0.793

Residuals 43 1.233 0.028675

Anova\_results <- aov(ANGPT2 ~ IOP + Batch, corrTable\_Batch2); summary(Anova\_results)

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

IOP 1 0.0642 0.06420 2.305 0.136

Batch 1 0.0012 0.00115 0.041 0.840

Residuals 42 1.1697 0.02785

Anova\_results <- aov(ANGPT2 ~ IOP + Batch + Age, corrTable\_Batch2); summary(Anova\_results)

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

IOP 1 0.0642 0.06420 2.273 0.140

Batch 1 0.0012 0.00115 0.041 0.841

Age 3 0.0681 0.02269 0.803 0.500

Residuals 39 1.1016 0.02825

t.test(ANGPT2 ~ Batch, corrTable\_Batch2, mu=0, alt="two.sided", conf=0.95, var.eq=F, paired=F)

Welch Two Sample t-test

data: ANGPT2 by Batch

t = -0.25607, df = 32.962, p-value = 0.7995

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

95 percent confidence interval:

-0.12145519 0.09430056

sample estimates:

mean in group B\_1 mean in group B\_2

7.000838 7.014415

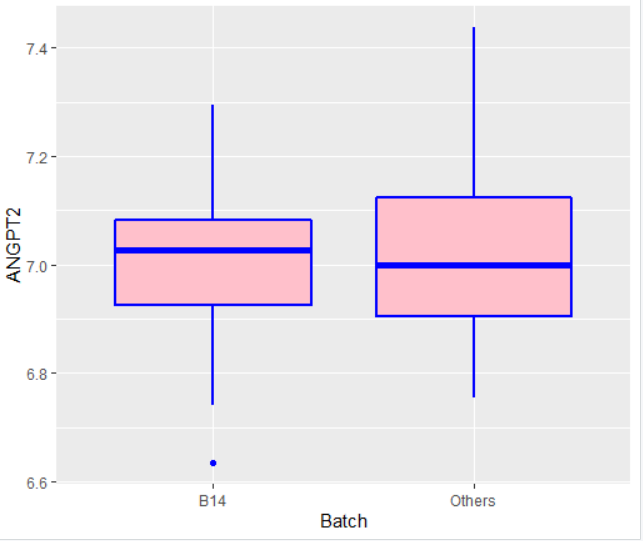
In all tests and analysis P-values are greater than 0.05 and I cannot reject the null hypothesis.

**Division 2): Separate B14 from others**

summary(corrTable\_Batch2$Batch)

B14 Others

21 24



Anova\_results <- aov(ANGPT2 ~ Batch, corrTable\_Batch2); summary(Anova\_results)

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

Batch 1 0.009 0.00899 0.315 0.577

Residuals 43 1.226 0.02851

Anova\_results <- aov(ANGPT2 ~ IOP + Batch, corrTable\_Batch2); summary(Anova\_results)

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

IOP 1 0.0642 0.06420 2.378 0.131

Batch 1 0.0366 0.03665 1.357 0.251

Residuals 42 1.1342 0.02700

Anova\_results <- aov(ANGPT2 ~ IOP + Batch + Age + Sex, corrTable\_Batch2); summary(Anova\_results)

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

IOP 1 0.0642 0.06420 2.388 0.131

Batch 1 0.0366 0.03665 1.363 0.250

Age 3 0.0798 0.02661 0.990 0.408

Sex 2 0.0597 0.02986 1.111 0.340

Residuals 37 0.9946 0.02688

t.test(ANGPT2 ~ Batch, corrTable\_Batch2, mu=0, alt="two.sided", conf=0.95, var.eq=F, paired=F)

Welch Two Sample t-test

data: ANGPT2 by Batch

t = -0.56261, df = 42.503, p-value = 0.5767

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

95 percent confidence interval:

-0.1299189 0.0732570

sample estimates:

mean in group B14 mean in group Others

6.993875 7.022205

In all tests and analysis P-values are greater than 0.05 and I cannot reject the null hypothesis.

All analysis above indicates that there is no significant relation between IOP and ANGPT2 based on different subgroups of Batch feature.