Correlation of Signals for Normal and Fibrotic Liver

Justin Creeden

2021-04-20

Purpose

Compare the individual kinase and kinase family intensities between mouse and human normal and fibrotic liver.

Data

```
library("ggpubr")
```

```
## Loading required package: ggplot2
```

```
figure_loc = here::here("data", "outputs", "figures")
```

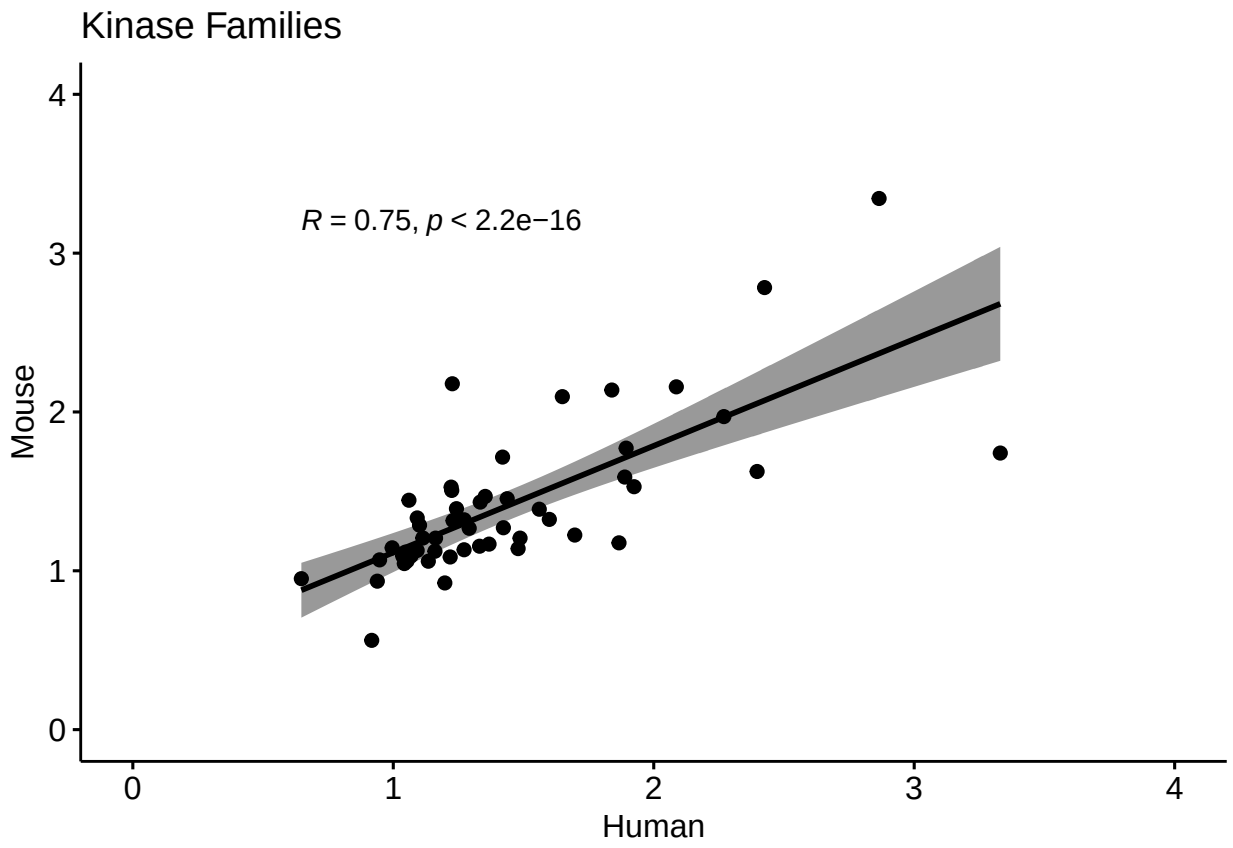
```
#load data
jfc_individual_kinase_comparison_data <- read.delim(here::here("data", "inputs", "differential_reports"
jfc_family_kinase_comparisone_data <- read.delim(here::here("data", "inputs", "differential_reports", "
```

Check Linearity

visualize data with scatter plot, to determine linearity

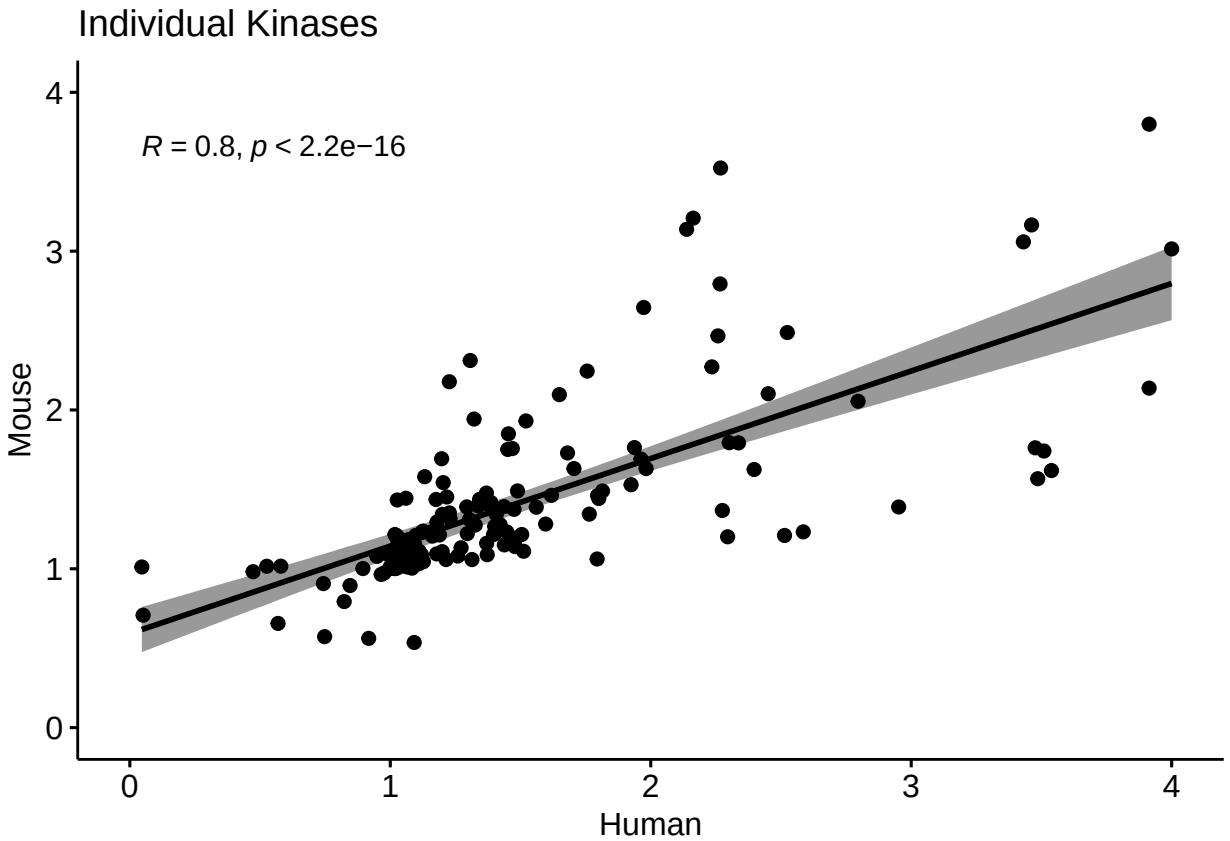
```
##### Family - Determine Linearity
ggscatter(jfc_family_kinase_comparisone_data, x = "human_avg_mean_final_score", y = "mouse_avg_mean_fina
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "spearman",
          xlim = c(0,4),
          ylim = c(0,4),
          xlab = "Human",
          ylab = "Mouse",
          title = "Kinase Families"
          )
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
##### Individual - Determine Linearity
ggscatter(jfc_individual_kinase_comparison_data, x = "human_mean_final_score", y = "mouse_mean_final_sc
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "spearman",
          xlim = c(0,4),
          ylim = c(0,4),
          xlab = "Human",
          ylab = "Mouse",
          title = "Individual Kinases",
          )
```

```
## `geom_smooth()` using formula 'y ~ x'
```



From the plots above, we can determine if the covariation (i.e.) relationship is linear. In the situation where the scatter plots show curved patterns, we are dealing with nonlinear association between the two variables.

Evaluate Normality

Shapiro Wilks

```
#Family
shapiro.test(jfc_family_kinase_comparisone_data$human_avg_mean_final_score)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  jfc_family_kinase_comparisone_data$human_avg_mean_final_score
## W = 0.84641, p-value = 8.825e-06
```

```
shapiro.test(jfc_family_kinase_comparisone_data$mouse_avg_mean_final_score)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  jfc_family_kinase_comparisone_data$mouse_avg_mean_final_score
## W = 0.82763, p-value = 2.822e-06
```

```
#Individual
shapiro.test(jfc_individual_kinase_comparison_data$human_mean_final_score)
```

```
##
##  Shapiro-Wilk normality test
##
```

```
## data:  jfc_individual_kinase_comparison_data$human_mean_final_score
## W = 0.82247, p-value = 1.585e-12
```

```
shapiro.test(jfc_individual_kinase_comparison_data$mouse_mean_final_score)
```

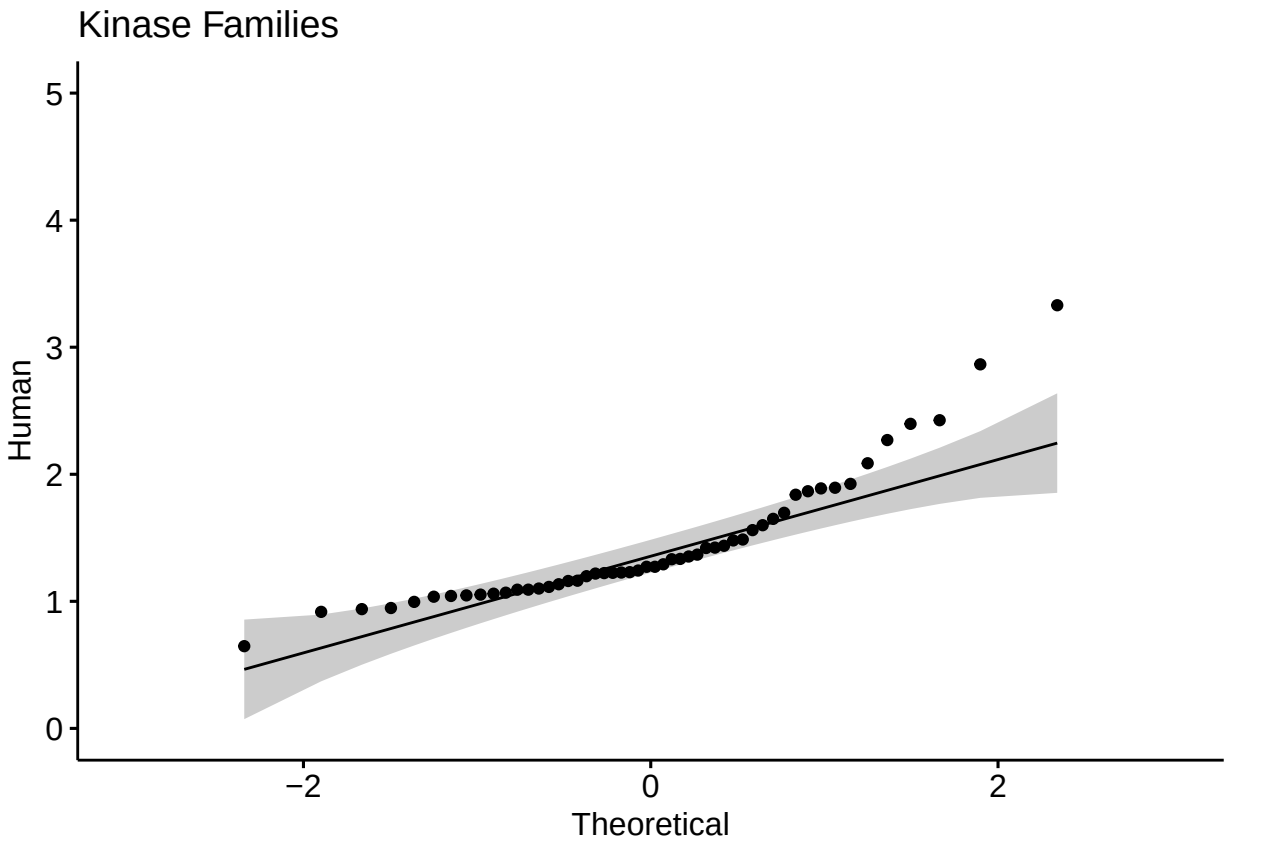
```
##
##  Shapiro-Wilk normality test
##
## data:  jfc_individual_kinase_comparison_data$mouse_mean_final_score
## W = 0.79478, p-value = 1.446e-13
```

From the output, if the two p-values are greater than the significance level 0.05 its implying that the distribution of the data are not significantly different from normal distribution. In other words, we can assume the normality.

Q-Q Plots

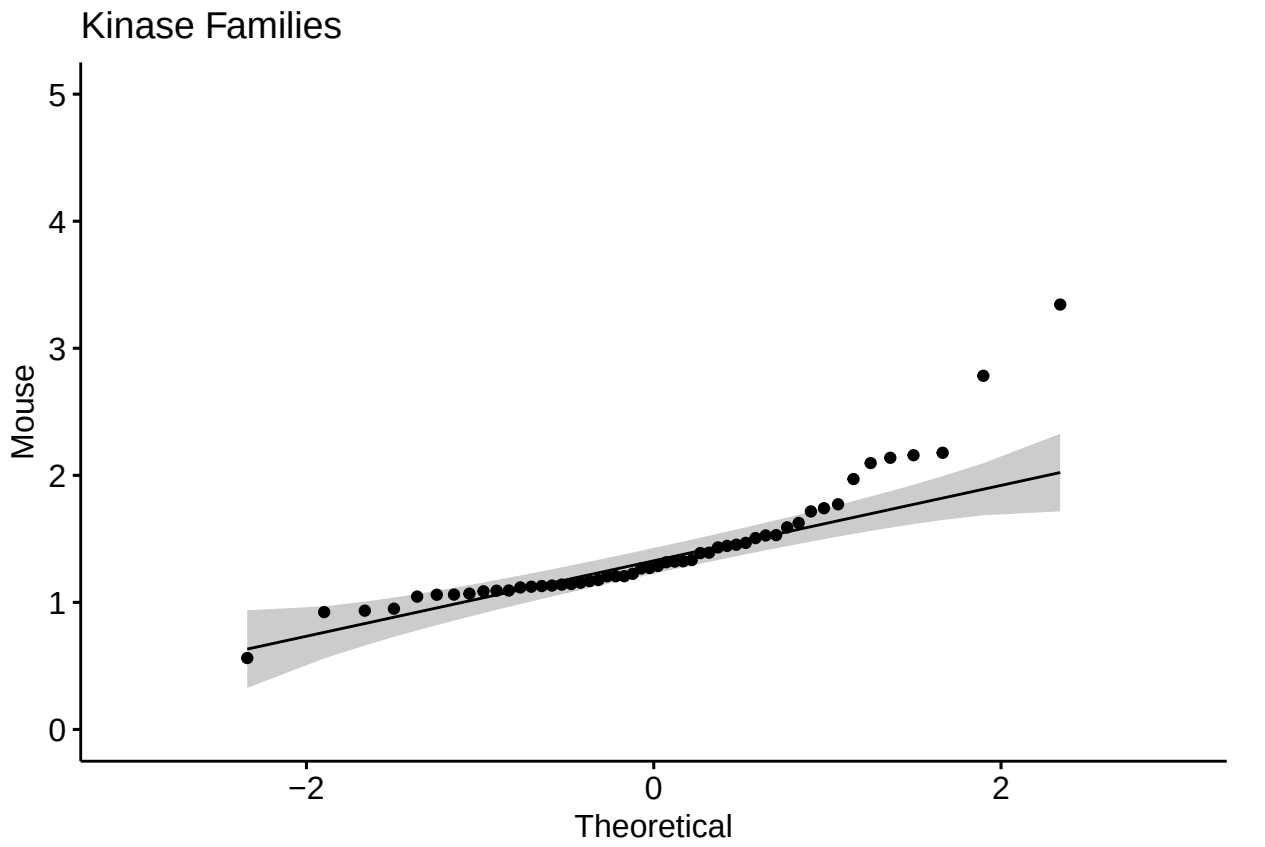
Visual inspection of the data normality using Q-Q plots (quantile-quantile plots).

```
#Family
ggpubr::ggqqplot(jfc_family_kinase_comparisone_data$human_avg_mean_final_score,
                 xlim = c(-3,3),
                 ylim = c(0,5),
                 ylab = "Human",
                 title = "Kinase Families")
```

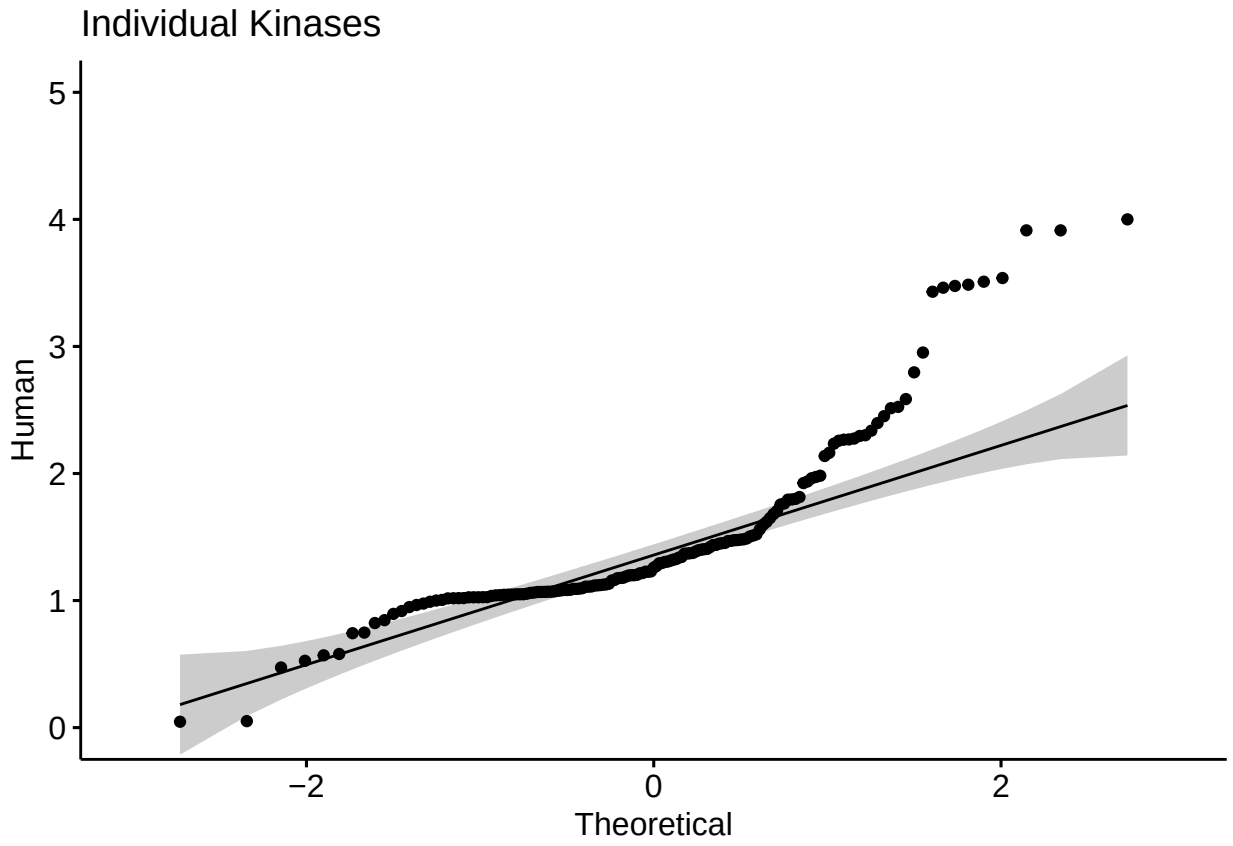



```
ggpubr::ggqqplot(jfc_family_kinase_comparisone_data$mouse_avg_mean_final_score,
                 xlim = c(-3,3),
                 ylim = c(0,5),
                 ylab = "Mouse",
```

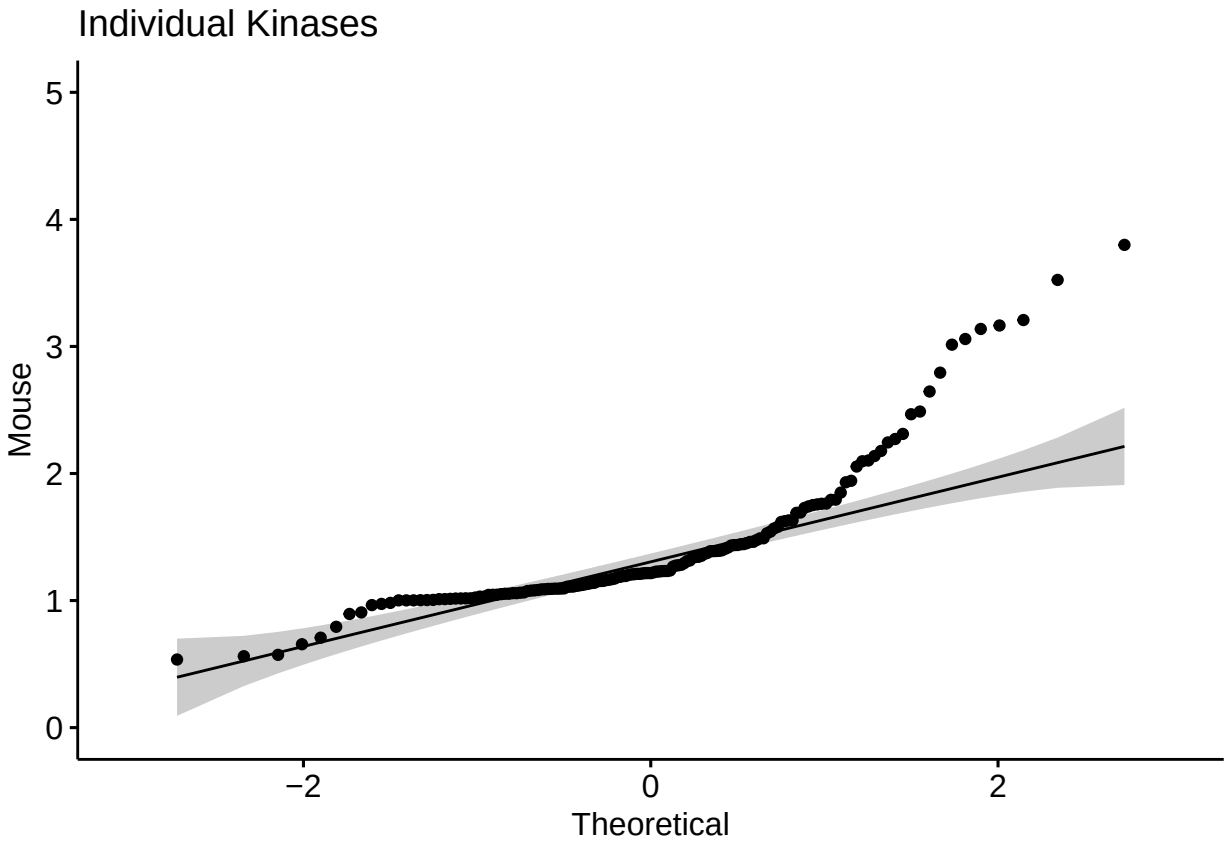
```
                  title = "Kinase Families")
```




```
#Individual
ggpubr::ggqqplot(jfc_individual_kinase_comparison_data$human_mean_final_score,
                  xlim = c(-3,3),
                  ylim = c(0,5),
                  ylab = "Human",
                  title = "Individual Kinases")
```




```
ggpubr::ggqqplot(jfc_individual_kinase_comparison_data$mouse_mean_final_score,
                 xlim = c(-3,3),
                 ylim = c(0,5),
                 ylab = "Mouse",
                 title = "Individual Kinases")
```



Q-Q plot draws the correlation between a given sample and the normal distribution. Note that, if the data are not normally distributed, it's recommended to use the non-parametric correlation, including Spearman and Kendall rank-based correlation tests.

Correlation Tests

Pearson

Pearson correlation (r), which measures a linear dependence between two variables (x and y). It's also known as a parametric correlation test because it depends to the distribution of the data. It can be used only when x and y are from normal distribution. The plot of y = f(x) is named the linear regression curve.

```
#Family
cor.test(x = jfc_family_kinase_comparisone_data$human_avg_mean_final_score, y = jfc_family_kinase_compa
```

```
##
##  Pearson's product-moment correlation
##
## data:  jfc_family_kinase_comparisone_data$human_avg_mean_final_score and jfc_family_kinase_comparison
## t = 7.3598, df = 50, p-value = 1.644e-09
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5580274 0.8305645
## sample estimates:
##       cor
## 0.7211125
```

```
#Individual
cor.test(x = jfc_individual_kinase_comparison_data$human_mean_final_score, y = jfc_individual_kinase_co
```

```
##
##  Pearson's product-moment correlation
##
## data:  jfc_individual_kinase_comparison_data$human_mean_final_score and jfc_individual_kinase_compari
## t = 12.308, df = 155, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6140344 0.7744087
## sample estimates:
##       cor
## 0.7030521
```

In the result above: t is the t-test statistic value (t=), df is the degrees of freedom (df=), p-value is the significance level of the t-test (p-value=). conf.int is the confidence interval of the correlation coefficient at 95% (conf.int = [,]); sample estimates is the correlation coefficient (Cor.coeff=).If the p-value of the test is less than the significance level alpha = 0.05. We can conclude that x and y are significantly correlated with a correlation coefficient of (cor.coeff=) and p-value of (p-value=).

Kendall-tau

Kendall tau is a rank-based correlation coefficient (non-parametric).

The Kendall rank correlation coefficient or Kendall's tau statistic is used to estimate a rank-based measure of association. This test may be used if the data do not necessarily come from a bivariate normal distribution.

The Kendall correlation method measures the correspondence between the ranking of x and y variables. The total number of possible pairings of x with y observations is n * (n - 1) / 2, where n is the size of x and y.The procedure is as follow: Begin by ordering the pairs by the x values. If x and y are correlated, then they would have the same relative rank orders. Now, for each yi, count the number of yj>yi (concordant pairs (c)) and the number of yj<yi (discordant pairs (d)).

```
#Family
cor.test(x = jfc_family_kinase_comparisone_data$human_avg_mean_final_score, y = jfc_family_kinase_compa
```

```
##
##  Kendall's rank correlation tau
##
## data:  jfc_family_kinase_comparisone_data$human_avg_mean_final_score and jfc_family_kinase_compariso
## z = 5.871, p-value = 4.333e-09
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## 0.561086
```

```
#Individual
cor.test(x = jfc_individual_kinase_comparison_data$human_mean_final_score, y = jfc_individual_kinase_co
```

```
##
##  Kendall's rank correlation tau
##
## data:  jfc_individual_kinase_comparison_data$human_mean_final_score and jfc_individual_kinase_compari
## z = 11.343, p-value < 2.2e-16
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## 0.610265
```

In the result above: t is the t-test statistic value (t=), df is the degrees of freedom (df=), p-value is the

significance level of the t-test (p-value=). conf.int is the confidence interval of the correlation coefficient at 95% (conf.int = [,]); sample estimates is the correlation coefficient (Cor.coeff=).If the p-value of the test is less than the significance level alpha = 0.05. We can conclude that x and y are significantly correlated with a correlation coefficient of (cor.coeff=) and p-value of (p-value=).

Spearman

Spearman rho is a rank-based correlation coefficient (non-parametric).

Spearman's rho statistic is also used to estimate a rank-based measure of association. This test may be used if the data do not come from a bivariate normal distribution.

The Spearman correlation method computes the correlation between the rank of x and the rank of y variables.

```
#Family
cor.test(x = jfc_family_kinase_comparisone_data$human_avg_mean_final_score, y = jfc_family_kinase_compa
```

```
##
##  Spearman's rank correlation rho
##
## data:  jfc_family_kinase_comparisone_data$human_avg_mean_final_score and jfc_family_kinase_comparison
## S = 5812, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.7518996
```

```
#Individual
cor.test(x = jfc_individual_kinase_comparison_data$human_mean_final_score, y = jfc_individual_kinase_co
```

```
## Warning in cor.test.default(x =
## jfc_individual_kinase_comparison_data$human_mean_final_score, : Cannot compute
## exact p-value with ties
```

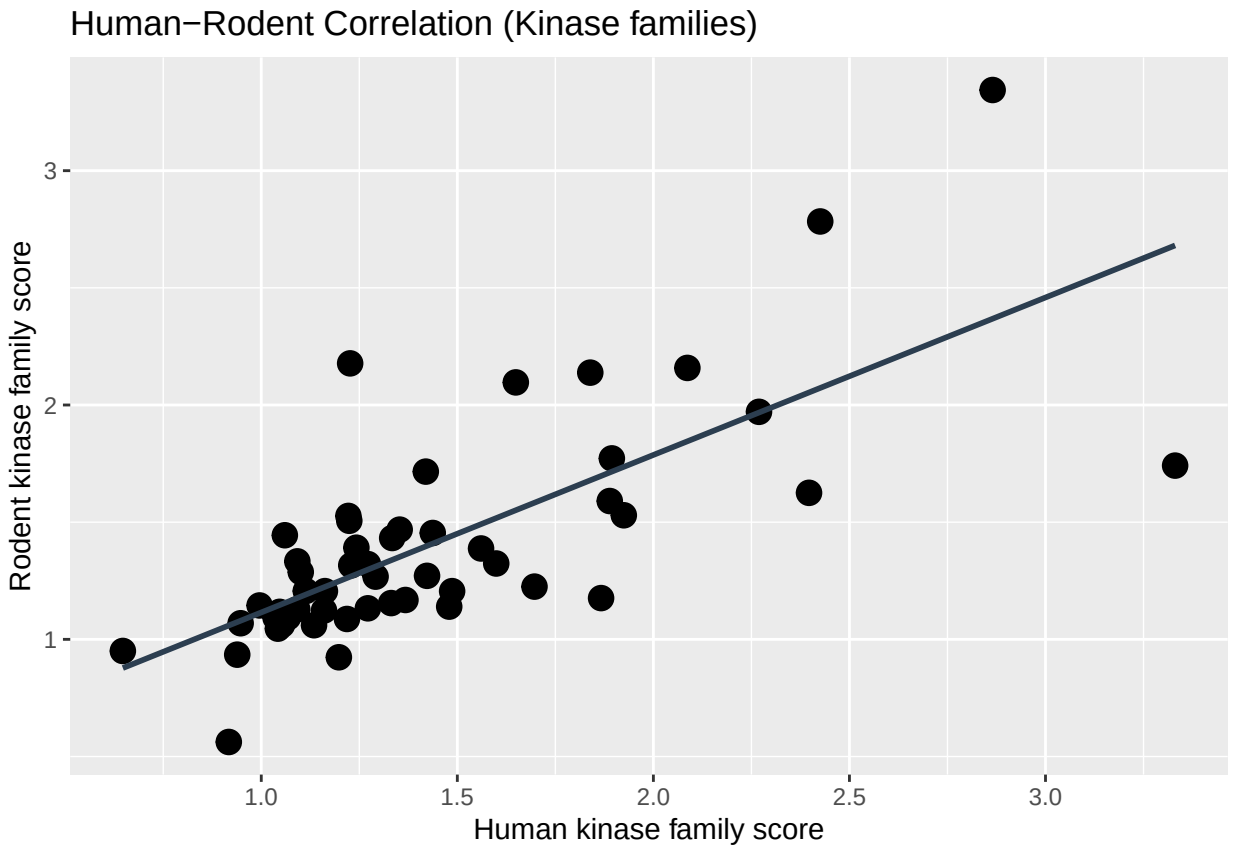
```
##
##  Spearman's rank correlation rho
##
## data:  jfc_individual_kinase_comparison_data$human_mean_final_score and jfc_individual_kinase_compari
## S = 126471, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.8039082
```

In the result above: t is the t-test statistic value (t=), df is the degrees of freedom (df=), p-value is the significance level of the t-test (p-value=). conf.int is the confidence interval of the correlation coefficient at 95% (conf.int = [,]); sample estimates is the correlation coefficient (Cor.coeff=).If the p-value of the test is less than the significance level alpha = 0.05. We can conclude that x and y are significantly correlated with a correlation coefficient of (cor.coeff=) and p-value of (p-value=).

```
#Family
plot_kinase_family_human_mouse_correlation <- ggplot(jfc_family_kinase_comparisone_data, aes(x = human_a
  geom_point(size=4) +
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE, color="#2C3E50")

plot_kinase_family_human_mouse_correlation + ggtitle("Human-Rodent Correlation (Kinase families)") + xla
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
pdf(file = file.path(figure_loc, "labeled_plot_kinase_family_human_mouse_correlation.pdf"), useDingbats
plot_kinase_family_human_mouse_correlation + ggtitle("Human-Rodent Correlation (Kinase families)") + xla
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
dev.off()
```

```
## pdf
##   2
```

```
png(file = file.path(figure_loc, "labeled_plot_kinase_family_human_mouse_correlation.png"), width = 200
plot_kinase_family_human_mouse_correlation + ggtitle("Human-Rodent Correlation (Kinase families)") + xla
```

```
## `geom_smooth()` using formula 'y ~ x'
```

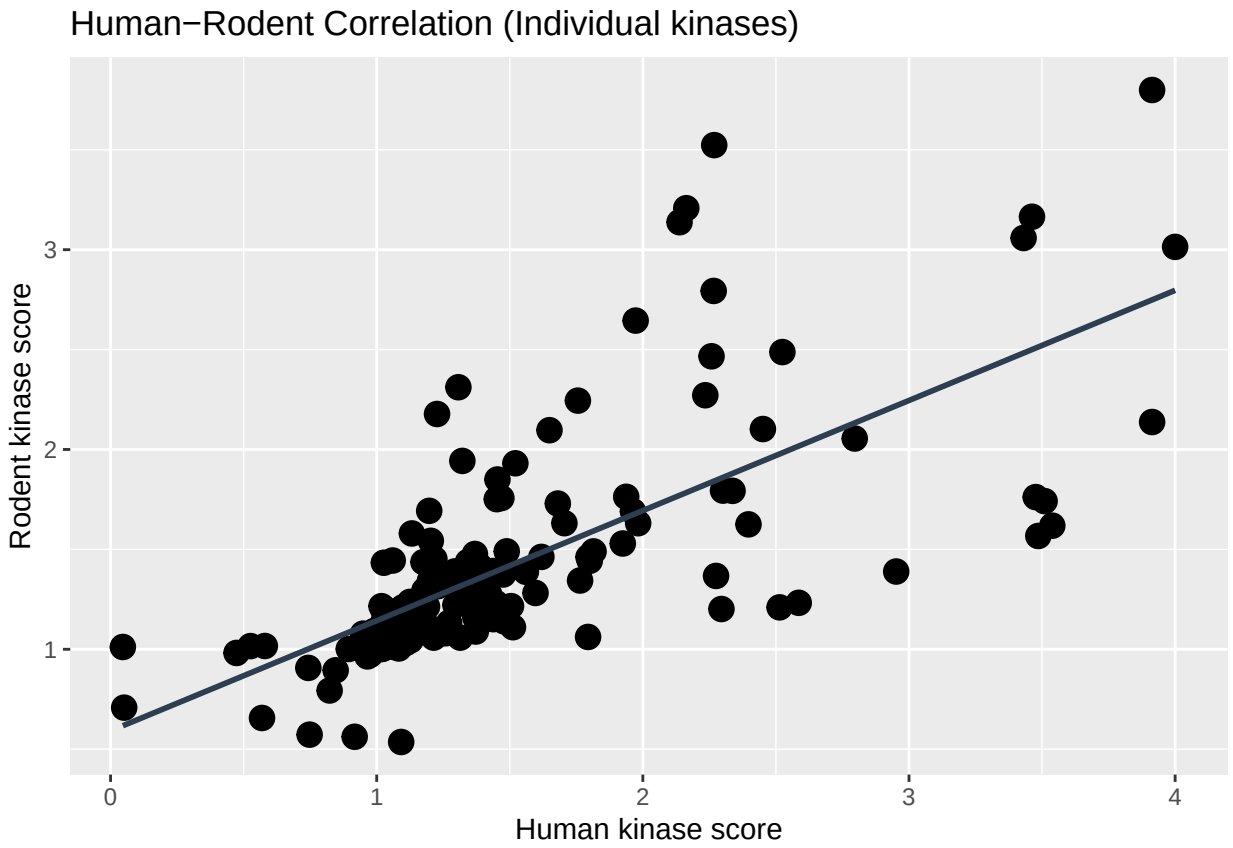
```
dev.off()
```

```
## pdf
##   2
```

```
#Individual
plot_individual_kinase_human_mouse_correlation<- ggplot(jfc_individual_kinase_comparison_data, aes(x = 
  geom_point(size=4) +
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE, color="#2C3E50")

plot_individual_kinase_human_mouse_correlation + ggtitle("Human-Rodent Correlation (Individual kinases)
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
pdf(file = file.path(figure_loc, "labeled_plot_individual_kinase_human_mouse_correlation.pdf") ,useDingb
plot_individual_kinase_human_mouse_correlation + ggtitle("Human-Rodent Correlation (Individual kinases)")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
dev.off()
```

```
## pdf
##   2
```

```
png(file = file.path(figure_loc, "labeled_plot_individual_kinase_human_mouse_correlation.png"), width =
plot_individual_kinase_human_mouse_correlation + ggtitle("Human-Rodent Correlation (Individual kinases)")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
dev.off()
```

```
## pdf
##   2
```