**Report of Correlation between 1st Diagnosis and 2nd Diagnosis**

Jieqi Tu and Zhengjia (Nelson) Chen

Date: 03/25/2021

***Statistical methods:***

A total of 127 subjects were included in the study. For each subject, the locus, genes, AA change, and values for variants (before and after) were recorded. Variant 1 (the 5th column) was assumed to be values for cytology, and Variant2 (the 6th column) was assumed to be values for resection. The goal of this analysis was to check correlation between Variant 1 and 2 (Diagnosis 1 and Diagnosis 2). The n/a values were excluded for correlation tests.

The structure of the analysis includes: 1) Pearson and Spearman correlations were measured between variants (before and after). 2) Parametric paired t-test was performed to compare the means of variants (before and after). 3) Non-parametric paired test (Wilcoxon Signed Rank Test) was used to compare the medians of variants (before and after).

The significance levels were set at 0.05 for all tests. The SAS 9.4 Version (SAS Institute, Inc., Cary, North Carolina) was used for data managements and analyses.

***Conclusions:***

1. There are significant Spearman (p < 0.0001) and Pearson (p < 0.0001) correlation between the 1st and the 2nd diagnosis. The Spearman correlation is 0.65056 and the Pearson correlation is 0.54763.
2. Paired t-test and Wilcoxon Signed-rank test suggested that 1st diagnosis and 2nd diagnosis provided similar results (No significant difference in mean and median).

***Results***

Table 0. Reference table of correlation

|  |  |
| --- | --- |
| Correlation Values | Interpretation |
| Exactly –1 | A perfect downhill (negative) linear relationship |
| –0.70 | A strong downhill (negative) linear relationship |
| –0.50 | A moderate downhill (negative) relationship |
| –0.30 | A weak downhill (negative) linear relationship |
| 0 | No linear relationship |
| +0.30 | A weak uphill (positive) linear relationship |
| +0.50 | A moderate uphill (positive) relationship |
| +0.70 | A strong uphill (positive) linear relationship |
| Exactly +1 | A perfect uphill (positive) linear relationship |

1. Spearman Correlation

**Table 1. N Mean, SD, Median and range for Variant 1 and 2 at each time points.**

**Table

Description automatically generated**

**Table 2. Estimate for Spearman correlation.**

**Table

Description automatically generated**

Comments: The Spearman correlation between 1st diagnosis and 2nd diagnosis is 0.65056 with p-value less than 0.0001. There is significant moderate correlation coefficient between 1st diagnosis and 2nd diagnosis.

1. Pearson Correlation

**Table 3. Estimate for Pearson correlation.**

**Table

Description automatically generated**

Comments: The Pearson correlation between 1st diagnosis and 2nd diagnosis is 0.54763 with p-value less than 0.0001. There is significant moderate correlation coefficient between 1st diagnosis and 2nd diagnosis.

1. Paired t-test

**Table 4. Result for paired t-test.**

**Chart, histogram

Description automatically generated**

Comments: The p-value for the paired t-test is greater than 0.05, which implies that there is no significant difference between 1st diagnosis and 2nd diagnosis. The 1st diagnosis and 2nd diagnosis give similar results.

1. Wilcoxon Signed Rank Test

**Table 5. Result of Wilcoxon signed-rank test.**

**Table

Description automatically generated**

Comments: The p-value of the Wilcoxon Signed-rank test is greater than 0.05, which implies that there is no significant difference in medians between 1st diagnosis and 2nd diagnosis. The 1st diagnosis and 2nd diagnosis give similar results.