|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Genes | Altered | Nobs | Mean ± SDa | Mean ± SDb | P-value\* |
| CD274|29126 | 0 | 348 | 150.83±270.01 | 4.36±1.11 | 0.221 |
|  | 1 | 141 | 157.95±174.56 | 4.49±1.15 |  |
| CD3D|915 | 0 | 348 | 191.83±230.93 | 4.68±1.14 | 0.163 |
|  | 1 | 141 | 159.68±198.31 | 4.52±1.15 |  |
| CD3E|916 | 0 | 348 | 426.40±464.28 | 5.51±1.11 | 0.157 |
|  | 1 | 141 | 320.88±295.26 | 5.35±1.02 |  |
| CD4|920 | 0 | 348 | 1076.58±989.07 | 6.57±0.95 | 0.366 |
|  | 1 | 141 | 941.42±757.51 | 6.49±0.93 |  |
| CD8A|925 | 0 | 348 | 338.94±457.87 | 5.03±1.36 | 0.210 |
|  | 1 | 141 | 273.72±381.14 | 4.86±1.33 |  |
| CD8B|926 | 0 | 348 | 100.63±162.23 | 3.65±1.48 | 0.199 |
|  | 1 | 141 | 67.68±103.18 | 3.47±1.30 |  |
| EGFR|1956 | 0 | 348 | 4897.54±8913.69 | 7.88±1.05 | 0.652 |
|  | 1 | 141 | 4734.29±6948.49 | 7.90±1.09 |  |
| EGF|1950 | 0 | 348 | 48.02±79.10 | 2.94±1.44 | 0.886 |
|  | 1 | 141 | 35.80±47.66 | 2.88±1.30 |  |
| ERBB2|2064 | 0 | 348 | 4458.63±13304.29 | 8.12±0.51 | 0.923 |
|  | 1 | 141 | 4518.04±8153.94 | 8.13±0.58 |  |
| ERBB3|2065 | 0 | 348 | 1763.64±1144.76 | 7.28±0.68 | 0.427 |
|  | 1 | 141 | 1594.15±853.09 | 7.23±0.55 |  |
| FGF1|2246 | 0 | 348 | 140.84±173.81 | 4.42±1.06 | **0.006** |
|  | 1 | 141 | 92.57±99.17 | 4.14±0.90 |  |
| FGF2|2247 | 0 | 348 | 146.09±249.06 | 4.27±1.17 | **0.009** |
|  | 1 | 141 | 97.69±133.59 | 3.96±1.14 |  |
| FOXP3|50943 | 0 | 348 | 172.29±165.27 | 4.70±1.03 | 0.128 |
|  | 1 | 141 | 147.92±133.21 | 4.53±1.14 |  |
| IFNG|3458 | 0 | 348 | 18.51±31.56 | 2.13±1.30 | 0.401 |
|  | 1 | 141 | 15.50±22.24 | 2.02±1.30 |  |
| IL2|3558 | 0 | 348 | 0.43±1.02 | 0.25±0.40 | **0.021** |
|  | 1 | 141 | 0.23±0.45 | 0.16±0.29 |  |
| IL6|3569 | 0 | 348 | 283.06±886.15 | 4.56±1.40 | 0.915 |
|  | 1 | 141 | 240.26±407.05 | 4.55±1.43 |  |
| NRG1|3084 | 0 | 348 | 578.34±825.50 | 5.69±1.27 | **0.000** |
|  | 1 | 141 | 911.97±2070.61 | 6.16±1.10 |  |
| PDCD1|5133 | 0 | 348 | 78.72±100.25 | 3.75±1.17 | 0.266 |
|  | 1 | 141 | 59.54±60.60 | 3.62±1.06 |  |
| TGFB1|7040 | 0 | 348 | 3261.26±1597.48 | 7.97±0.52 | **0.039** |
|  | 1 | 141 | 3470.87±1511.70 | 8.07±0.41 |  |
| TP53|7157 | 0 | 348 | 1537.49±1151.60 | 7.00±0.91 | 0.248 |
|  | 1 | 141 | 1313.41±897.08 | 6.89±0.86 |  |
| IL12A|3592 | 0 | 348 | 9.93±20.77 | 1.79±0.98 | 0.115 |
|  | 1 | 141 | 6.61±8.29 | 1.64±0.85 |  |
| IL12B|3593 | 0 | 348 | 4.36±6.29 | 1.24±0.89 | 0.494 |
|  | 1 | 141 | 3.81±5.13 | 1.18±0.84 |  |
| CASP8|841 | 0 | 348 | 602.69±262.71 | 6.31±0.43 | 0.331 |
| CASP8|841 | 1 | 141 | 637.85±287.74 | 6.36±0.47 |  |

Part 1 analyses for whole data

Table 1.1 Basic Descriptive statistics and t test of the genes

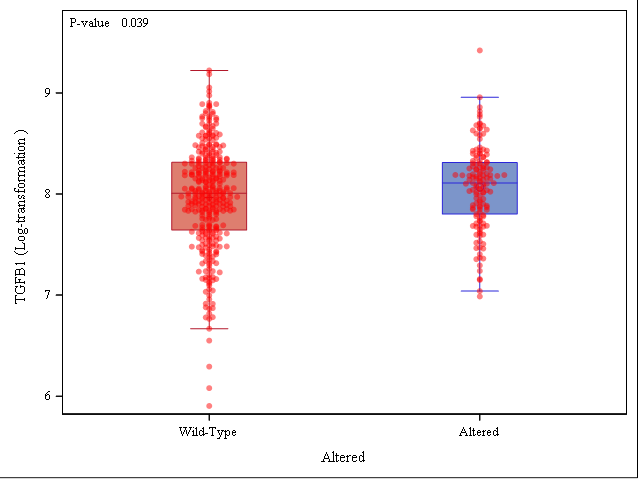
Altered: 0=”no” 1=”yes”

Mean ± SDa: were calculated by original data

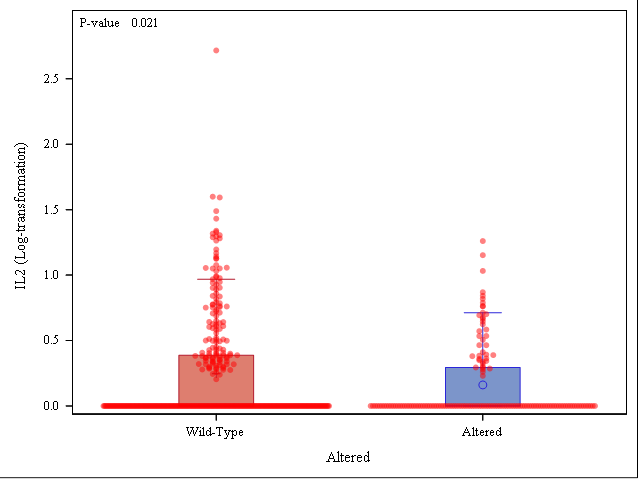
Mean ± SDb: were calculated with log-transformation data (log2(x+1)).

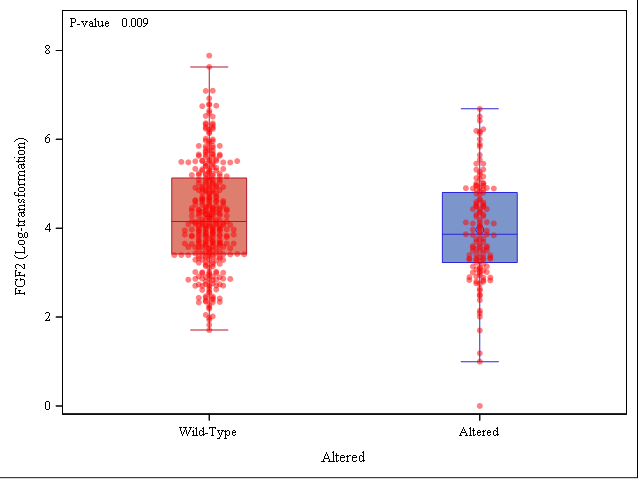
\* p-value was calculated by t-test for log-transformation data.

Bee-swarm plots for the significant genes











Part 2 Data analyses for HPV- data

Table 2.1 Descriptive statistics and t test of the genes

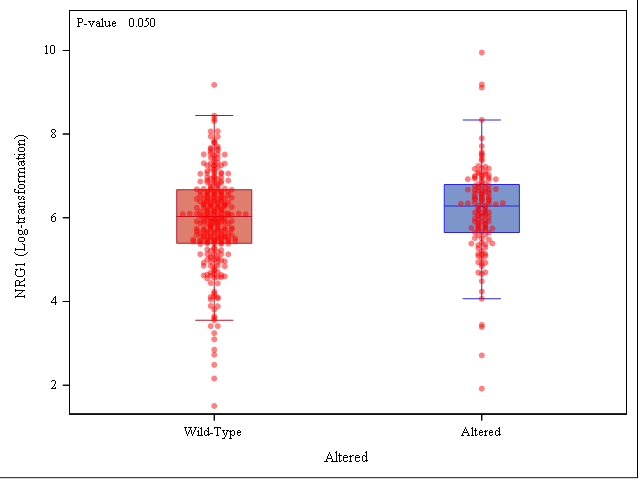
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Genes | Altered | Nobs | Mean ± SDa | Mean ± SDb | P-value\* |
| CASP8|841 | 0 | 280 | 609.90±279.25 | 6.32±0.45 | 0.330 |
| CASP8|841 | 1 | 136 | 642.75±289.84 | 6.36±0.47 |  |
| CD274|29126 | 0 | 280 | 145.48±209.41 | 4.32±1.14 | 0.060 |
| CD274|29126 | 1 | 136 | 162.66±175.90 | 4.55±1.12 |  |
| CD3D|915 | 0 | 280 | 146.60±161.43 | 4.47±1.07 | 0.303 |
| CD3D|915 | 1 | 136 | 165.07±199.90 | 4.59±1.10 |  |
| CD3E|916 | 0 | 280 | 333.97±361.32 | 5.31±1.05 | 0.270 |
| CD3E|916 | 1 | 136 | 331.28±295.50 | 5.42±0.96 |  |
| CD4|920 | 0 | 280 | 975.72±931.29 | 6.46±0.96 | 0.414 |
| CD4|920 | 1 | 136 | 969.39±756.58 | 6.54±0.90 |  |
| CD8A|925 | 0 | 280 | 259.76±357.52 | 4.79±1.32 | 0.309 |
| CD8A|925 | 1 | 136 | 283.02±384.96 | 4.93±1.30 |  |
| CD8B|926 | 0 | 280 | 65.14±103.88 | 3.34±1.37 | 0.188 |
| CD8B|926 | 1 | 136 | 69.93±104.38 | 3.52±1.29 |  |
| EGFR|1956 | 0 | 280 | 5572.41±9792.27 | 8.02±1.03 | 0.361 |
| EGFR|1956 | 1 | 136 | 4794.77±7057.96 | 7.92±1.03 |  |
| EGF|1950 | 0 | 280 | 52.36±77.10 | 3.15±1.36 | 0.105 |
| EGF|1950 | 1 | 136 | 36.66±48.16 | 2.93±1.26 |  |
| ERBB2|2064 | 0 | 280 | 3622.07±3333.05 | 8.07±0.45 | 0.315 |
| ERBB2|2064 | 1 | 136 | 4544.00±8301.33 | 8.12±0.59 |  |
| ERBB3|2065 | 0 | 280 | 1594.42±968.33 | 7.19±0.68 | 0.556 |
| ERBB3|2065 | 1 | 136 | 1602.13±862.83 | 7.23±0.56 |  |
| FGF1|2246 | 0 | 280 | 121.21±142.45 | 4.32±0.99 | 0.103 |
| FGF1|2246 | 1 | 136 | 93.26±99.99 | 4.16±0.85 |  |
| FGF2|2247 | 0 | 280 | 113.14±148.70 | 4.14±1.10 | 0.158 |
| FGF2|2247 | 1 | 136 | 98.63±135.57 | 3.97±1.13 |  |
| FOXP3|50943 | 0 | 280 | 150.21±150.41 | 4.55±1.03 | 0.909 |
| FOXP3|50943 | 1 | 136 | 151.43±134.10 | 4.56±1.14 |  |
| IFNG|3458 | 0 | 280 | 14.61±26.16 | 1.94±1.25 | 0.317 |
| IFNG|3458 | 1 | 136 | 16.00±22.48 | 2.07±1.29 |  |
| IL12A|3592 | 0 | 280 | 6.97±13.99 | 1.59±0.87 | 0.791 |
| IL12A|3592 | 1 | 136 | 6.28±7.96 | 1.62±0.82 |  |
| IL12B|3593 | 0 | 280 | 3.63±5.92 | 1.10±0.86 | 0.225 |
| IL12B|3593 | 1 | 136 | 3.92±5.19 | 1.21±0.84 |  |
| IL2|3558 | 0 | 280 | 0.26±0.58 | 0.17±0.32 | 0.805 |
| IL2|3558 | 1 | 136 | 0.24±0.46 | 0.16±0.29 |  |
| IL6|3569 | 0 | 280 | 315.72±972.90 | 4.71±1.37 | 0.305 |
| IL6|3569 | 1 | 136 | 242.61±411.05 | 4.56±1.43 |  |
| NRG1|3084 | 0 | 280 | 667.70±887.36 | 5.94±1.14 | **0.050** |
| NRG1|3084 | 1 | 136 | 922.96±2104.96 | 6.17±1.09 |  |
| PDCD1|5133 | 0 | 280 | 59.93±71.09 | 3.54±1.11 | 0.210 |
| PDCD1|5133 | 1 | 136 | 61.35±60.94 | 3.68±1.02 |  |
| TGFB1|7040 | 0 | 280 | 3538.57±1579.82 | 8.07±0.47 | 0.835 |
| TGFB1|7040 | 1 | 136 | 3490.51±1462.47 | 8.08±0.0.40 |  |
| TP53|7157 | 0 | 280 | 1213.98±896.55 | 6.79±0.86 | 0.293 |
| TP53|7157 | 1 | 136 | 1294.98±877.06 | 6.88±0.86 | 0.330 |

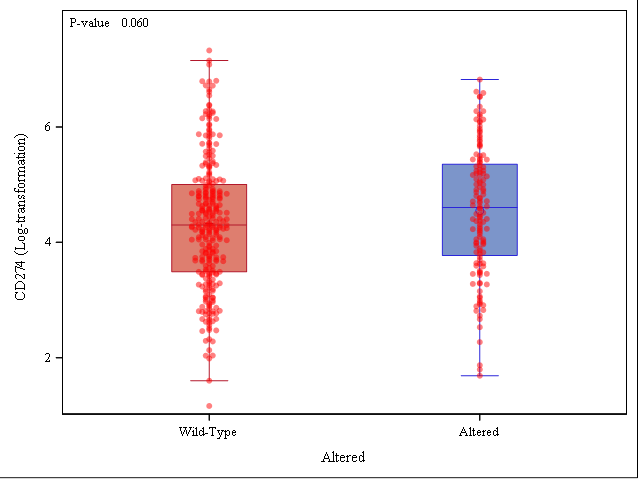
Altered: 0=”no” 1=”yes”

Mean ± SDa: were calculated by original data

Mean ± SDb: were calculated with log-transformation data (log2(x+1)).

\* p-value was calculated by t-test for log-transformation data.





**Statistical Analysis**

Statistical analysis was conducted using SAS Version 9.4 (Cary, NC). Descriptive statistics such as mean and standard deviations were reported for each variable. Each gene has larger variability, so log2(x+1) transformation was used to improve the interpretability and appearance of graphs for each gene. Independent t test was performed to detect the difference between altered and wild-type groups for log-transformation data. Bee-swarm plots were presented to show the basic statistics of altered and wild-type groups for significant genes. The significance level was set at 0.05.