

Due: Fri April 24, 2009

Name: \_\_\_\_\_

## **Algebra 2<sup>nd</sup> Semester Project: 150 pts Microarray (statistics applications)**

### **I. Vocabulary (20 PTS)**

\* DNA/RNA:

\*DNA oligonucleotides

\* Gene chip:

\* Genome:

\* Nucleus:

\* Restriction Enzymes:

\* Single Nucleotide Polymorphisms:.

\*Hybridize

\*Microarrays

\*Bioinformatics

\*Probe

\*Tiling Array

\*Gene expression

\*P-value

\*False positives

\*False negatives

\*False Discovery Rate, (FDR)

II. **Compose an essay** of at least 300–500 words to explain what microarrays are, how they are made, and how data is collected from them. **(30 PTS)**

III. **Multiple choice questions:** Go to the following website and answer the following questions (use the information links) **(25 PTS)**

[http://www.fastol.com/~renkwitz/microarray\\_chips.htm](http://www.fastol.com/~renkwitz/microarray_chips.htm)

Part 1

|    |    |    |    |    |    |    |    |    |     |
|----|----|----|----|----|----|----|----|----|-----|
| 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. |
|    |    |    |    |    |    |    |    |    |     |

|     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 11. | 12. | 13. | 14. | 15. | 16. | 17. | 18. | 19. | 20. |
|     |     |     |     |     |     |     |     |     |     |

|    |    |    |    |    |
|----|----|----|----|----|
| 21 | 22 | 23 | 24 | 25 |
|    |    |    |    |    |

Part 2:

|    |    |    |    |    |    |    |    |    |     |
|----|----|----|----|----|----|----|----|----|-----|
| 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. |
|    |    |    |    |    |    |    |    |    |     |

|     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 11. | 12. | 13. | 14. | 15. | 16. | 17. | 18. | 19. | 20. |
|     |     |     |     |     |     |     |     |     |     |

IV **Powerpoint:**. The presentation should focus on the affect of microarray technology on the diagnosis, or treatment, of a disease. Support your position with the facts you learned from the previous section (section 3 part 2) **(25 PTS)**

Guidelines for powerpoint: minimum 7 slides– max 20 slides. Do not use a font size too small. (No smaller than 20 pt) Powerpoint should be more of an outline form. Do not use too much text. Be careful of background selection.(You will need to adjust font color of text so that it is easy to read.

V. **Statistics:** Use the following data and answer the questions below.

(Use graphing calculator for these ) (50 PTS)

| Gene # | Gene Name | logFC       | AveExpr     | adj.P.Val   | B value     |
|--------|-----------|-------------|-------------|-------------|-------------|
| 1      | TC258447  | 2.834021668 | 10.89918665 | 4.05E-05    | 8.221371218 |
| 2      | AW119963  | 2.39050806  | 10.53675103 | 0.000163999 | 7.624477357 |
| 3      | CD966451  | 1.492238579 | 11.67218004 | 0.000163999 | 7.566603083 |
| 4      | CD960633  | 1.23471765  | 12.75649625 | 0.000163999 | 7.514140459 |
| 5      | TC275102  | 1.133137518 | 9.526662945 | 0.000214301 | 7.344151738 |
| 6      | CF075306  | 1.263209793 | 10.55041141 | 0.000446337 | 6.984283668 |
| 7      | TC258261  | 1.495944655 | 12.8869253  | 0.000628622 | 6.764847662 |
| 8      | CB331535  | 1.749674368 | 12.81081752 | 0.000767873 | 6.567144919 |
| 9      | CD986101  | 1.322522764 | 10.46402228 | 0.000767873 | 6.54997961  |
| 10     | TC258202  | 1.315505669 | 14.48142999 | 0.000927118 | 6.397016746 |
| 11     | TC258394  | 1.270228394 | 12.51215623 | 0.000927118 | 6.352771316 |
| 12     | TC258433  | 2.838944043 | 10.66687677 | 0.001440897 | 6.022180462 |
| 13     | AW499234  | 1.348733131 | 12.07503394 | 0.001553039 | 5.936938561 |
| 14     | CD966910  | 1.574380503 | 10.83721183 | 0.002399693 | 5.636438734 |
| 15     | AI649871  | 1.556660239 | 14.02447323 | 0.002986383 | 5.455552395 |

- Mean, median & mode: Define and find each for the following various genes using the column average expression. Why is the median preferred for microarrays?
- Scatter plot and histogram. Create a scatter plot using B value as your independent value (x-value) and adjusted p-value for your dependent value (y-axis). Create a histogram using gene # and average expressions. Best to use excel spreadsheet.
- Standard deviation. Find the standard deviation of the average expression of the genes.
- Linear regression. Find the linear regression using the B value and adjusted p-value data.

\*\*P-value was already determined in the computer program for the above data. The following problems represent on a smaller scale

- p-value
  - Define p-value

- ii. A gene has an expression level mean of 2.45 with a standard deviation of 0.69. A biologist believes that “G” oligos have a higher gene expression than “S” oligos. A sample of 20 G oligos had a mean expression of 2.65.
  - a. Test the biologist claim at the 0.01 level of significance.
  - b. Find the p-value. The critical value = 2.326. For  $z=1.30$  the data from the chart of normal curves is equal to 0.4032. This follows  $p=0.5 + \text{chart value}$  (0.4032).
  
- f. Confidence interval. Define confidence interval.
  - i. Confidence interval:
  
  - ii. Find the 95% confidence interval for gene expression mean if the sample mean is 146 with a standard deviation of 12 for a sample of 18.
  
  - iii. Find the 90% confidence interval for the population proportion if a sample of 106 had a sample proportion of 21%.
  
  - iv. BONUS: the 90% confidence interval for the average expression of the genes.
  
- G. BONUS The data shows the adjusted p-value. This is adjusted due to Bonferroni corrections. Please explain the need for these corrections and give an example using Bonferroni corrections.

General overviews of how a microarray experiment works

<http://www.youtube.com/watch?v=VNstHmNjKhM>

<http://www.bio.davidson.edu/courses/genomics/chip/chip.html>

<http://www.dnai.org/d/index.html>

<http://www.youtube.com/watch?v=8YCuIsDidC8&feature=related>

<http://learn.genetics.utah.edu/units/biotech/microarray/>

<http://www.youtube.com/watch?v=ePFE7yg7LvM&NR=1>

Credit for assistance goes to Dr. Jack Gardiner and other Iplant personnel.