

---

# Report For CSCI 5461 HW2

---

Yangyang Li  
Department of Computer Science  
University of Minnesota  
li002252@edu.com

## 1 Question 1

### 1.1 a

How many genes are in the RNAseq gene expression profiles?  
**Answer: 20530**

### 1.2 b

How many patient samples are in the RNAseq gene expression profiles?  
**Answer: 308**

### 1.3 c

How many genes are in the microarray gene expression profiles?  
**Answer: 12042**

### 1.4 d

How many patient samples are in the microarray gene expression profiles?  
**Answer: 593**

## 2 Question 2

### 2.1 a

**T-test for RNA-Seq data**

Table 1: T-test for RNA-Seq

<i>Top Pvalues</i>	<i>Top Genes</i>
7.347984e-06	CYTH3
4.626091e-05	ATP1A2
1.786655e-05	RGS14
4.176600e-05	TSPAN9
4.639776e-05	LOC728606
4.334056e-05	CHUK
3.814459e-05	FBXO4
5.971342e-05	BTRC
8.269875e-05	C12orf5
9.403408e-05	LOC257358

### Rank-sum for RNA-Seq data

NOTE: I use *scipy.stats.ranksums* to conduct the Rank-sum test.

Table 2: Rank-sum for RNA-Seq

<i>Top Pvalues</i>	<i>Top Genes</i>
6.039218e-05	TEX261
5.656294e-05	LOC728606
7.215013e-05	SMTNL2
7.374517e-06	ATP1A2
1.216938e-05	CYTH3
4.247306e-05	RGS14
5.046463e-05	BTRC
7.533007e-05	ASAP3
9.532514e-05	CAMK2G
1.129098e-04	EIF4E3

### T-test for Microarray data

Table 3: T-test for Microarray

<i>Top Pvalues</i>	<i>Top Genes</i>
8.051437e-06	AXL
6.598877e-05	LRIG1
2.988205e-05	PART1
3.274241e-05	RABAC1
9.725890e-06	GALNT10
6.275392e-05	TEC
9.134666e-06	NOTCH4
2.820660e-05	SLC37A4
2.636916e-06	NDRG3
4.852395e-05	OLFML3

### Rank-sum for Microarray data

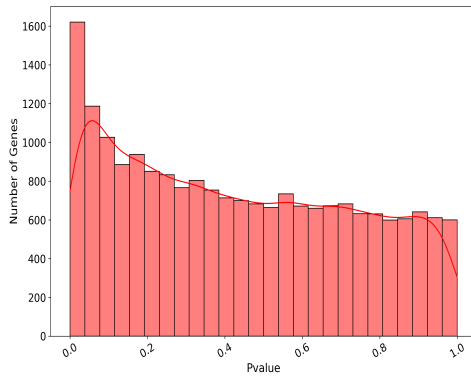
Table 4: Rank-sum for Microarray

<i>Top Pvalues</i>	<i>Top Genes</i>
2.503692e-07	NDRG3
3.252723e-05	RABAC1
5.281790e-05	C20orf3
1.534259e-05	GALNT10
2.033305e-05	AXL
5.298787e-05	ANXA4
7.591666e-05	SLC37A4
5.631749e-05	CTSA
8.554170e-05	LRIG1
1.040827e-04	AKAP12

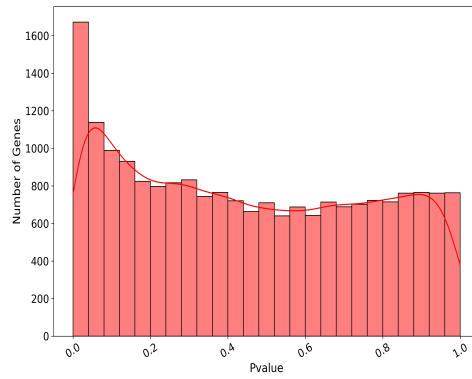
## 2.2 b

Table 5: Number of selected genes

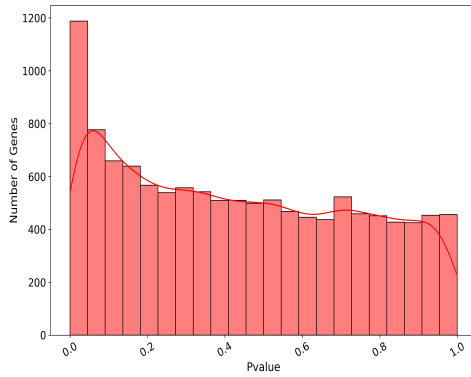
<i>Test</i>	<i>Genes No.</i>
T-test for RNA-Seq data	2000
Rank-sum for RNA-Seq data	1985
T-test for Microarray data	1268
Rank-sum for Microarray data	1257



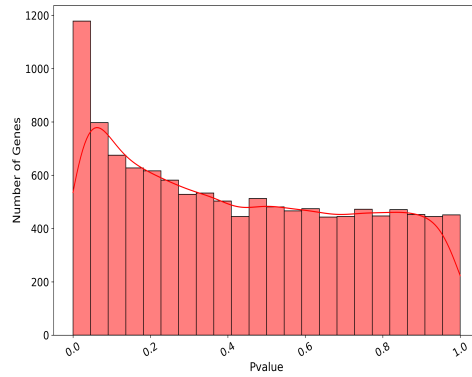
(a) T-test for RNA-Seq data



(b) Rank-sum for RNA-Seq data



(c) T-test for Microarray data



(d) Rank-sum for Microarray data

Figure 1: Histogram of the p-values of all genes

## 3 Question 3

### 3.1 a

Table 6: Number of selected genes

<i>Test</i>	<i>Genes No.</i>
Bonferroni correction for RNA-Seq data	0
Bonferroni correction for Microarray data	1

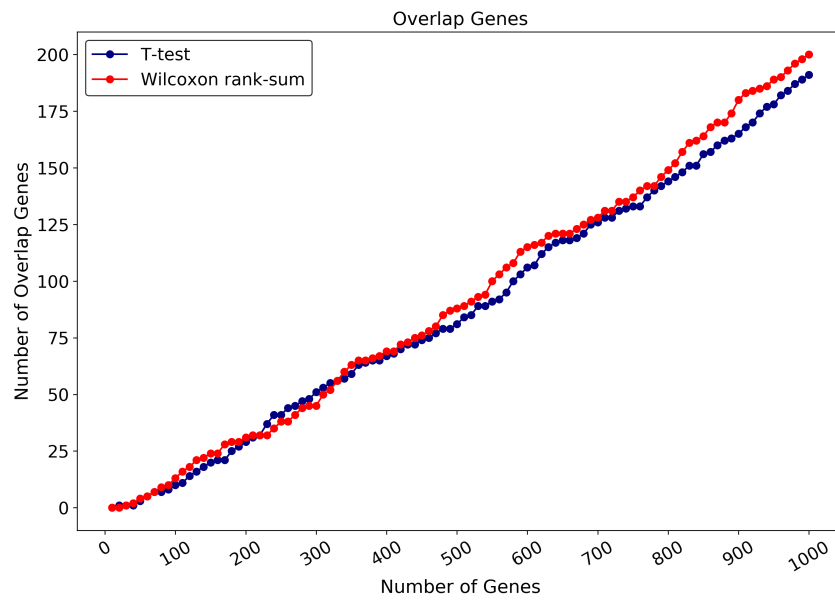
Table 7: FDR Values

Gene No.	FDR Value for RNA-Seq	FDR Value for microarray
20	0.1723868295115724	0.1324712603446682
50	0.2123598414113462	0.17366641103505984
100	0.25079808129875447	0.19131731961025905
200	0.31974860295072466	0.2781789817248434

3.2 b

#### 4 Question 4

Figure 2: Number of Overlap Genes



#### 5 Appendix

##### List of Tables

1	T-test for RNA-Seq . . . . .	1
2	Rank-sum for RNA-Seq . . . . .	2
3	T-test for Microarray . . . . .	2
4	Rank-sum for Microarray . . . . .	2
5	Number of selected genes . . . . .	3
6	Number of selected genes . . . . .	3
7	FDR Values . . . . .	4

##### List of Figures

1	Histogram of the p-values of all genes . . . . .	3
2	Number of Overlap Genes . . . . .	4