1.

A)

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ALLB1234$BT	4	2.1053	0.52632	3.4829	0.01082 *
Residuals	90	13.6006	0.15112		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

H0: all disease stages have the same mean expression.

HA: at least one stage has a different gene expression.

The p value is much smaller than 0.05 so we reject H0 and accept HA

B)

Residuals:

Min	1Q	Median	3Q	Max
-1.09026	-0.27845	0.03999	0.26618	0.71532

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
ALLB1234$BTB	6.81021	0.17385	39.17	<2e-16 ***
ALLB1234$BTB1	6.57951	0.08918	73.78	<2e-16 ***
ALLB1234$BTB2	6.47503	0.06479	99.94	<2e-16 ***
ALLB1234$BTB3	6.68533	0.08106	82.48	<2e-16 ***
ALLB1234$BTB4	6.91417	0.11222	61.61	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3887 on 90 degrees of freedom

Multiple R-squared: 0.9967, Adjusted R-squared: 0.9966

F-statistic: 5513 on 5 and 90 DF, p-value: < 2.2e-16

The B3 stage patient mean value is 6.68533.

C)

Pairwise comparisons using t tests with pooled SD

data: y and ALLB1234$BT

	B	B1	B2	B3
B1	1.00	-	-	-
B2	0.52	1.00	-	-
B3	1.00	1.00	0.37	-
B4	1.00	0.20	0.01	0.61

So all groups do not have a p value smaller than 0.05, which means no significant different between any groups and B group.

D)

Pairwise comparisons using t tests with pooled SD

data: y and ALLB1234$BT

	B	B1	B2	B3
B1	0.40	-	-	-
B2	0.19	0.48	-	-
B3	0.57	0.48	0.15	-
B4	0.62	0.11	0.01	0.20

P value adjustment method: fdr

As shown in the table, group B2 and group B4 has a different mean value as the P value is 0.01 smaller than 0.05

E)

Shapiro-Wilk normality test

data: residuals(lm(y ~ ALLB1234$BT))

W = 0.97839, p-value = 0.1177

Breusch-Pagan test

data: lm(y ~ ALLB1234$BT)
BP = 1.1702, df = 4, p-value = 0.883

As the Shapiro test and the BP test shown, as the P-value for both of them are greater than 0.05, the variance follows a normal distribution with mean of 0 and there is no difference of variance between groups. So we do not need to conduct a non-parameter test.

2
A)
There are 423 genes express differently.
B)
For the first five with most significant small p value, their gene names are
1389_at   38555_at   40268_at 1866_g_at   40155_at

3
A)
Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Bcell	3	24.436	8.1453	19.1179	1.818e-09 ***
sex	1	0.032	0.0319	0.0748	0.7851
Bcell:sex	3	0.230	0.0768	0.1803	0.9095
Residuals	81	34.511	0.4261		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

As shown in the table, there is no interaction between sex and Bcell as the P-value of interaction is 0.9095 which is much greater than 0.05

Bcell has a significant effect as the P-value is 1.818e-09 which is much smaller than 0.05.

Sex do not play a role as the P-value is 0.7851 which is much greater than 0.05

B)

Shapiro-Wilk normality test

data: residuals(lm(y ~ Bcell * sex))
W = 0.96926, p-value = 0.03291

Breusch-Pagan test

data: lm(y ~ Bcell * sex)
BP = 6.7635, df = 7, p-value = 0.4539

As shown in the table, between groups there is no different as the p-value of BP test is 0.45 which is much greater than 0.05

But the variance do not follow a normal distribution as the p-value of Shapiro test is 0.03291 which is much smaller than 0.05

4)

A)

B)

H0: there is no difference between group mean of B1, B2, and B3.

HA: at least one group mean is different than the others.

0.511

As the permutation test p-value is much greater than 0.05, we accept H0 that there is no difference of mean between B1, B2, and B3.