Name: *Ankur Sharma*
ID: *2015A8PS0443G*

Differential Gene Expression Analysis

Introduction

Gene expression is the process by which the heritable information in a gene, the sequence of DNA base pairs, is made into a functional gene product, such as protein or RNA. The basic idea is that DNA is transcribed into RNA, which is then translated into proteins. Proteins make many of the structures and all the enzymes in a cell or organism.

Differential Expression has become popular with the development of microarray technology. In these experiments RNA transcript levels are determined by hybridization to a microarray of short DNA probes. Genes are represented by 10 to 20 probes on the array. From the signal intensities of these spots on the area the expression level of the gene can be determined. But these values aren't particularly interesting on their own, it is most interesting to look at differences between the expression levels and different samples. one possible comparison is between diseased and normal tissues.

One important aspect of experimental design and all of these types of experiments is the inclusion of biological replicates. We want to be confident that the genes we identify are really differentially expressed. Biological replicates are samples from different patients or animals or cell culture plates, they help show the normal variation between samples of the same type either due to biological noise or noise from experimental differences. This allows us to see if the difference between sample types is greater than the normal variance between experimental replicates.

In this study project, we will go through a dataset of 24 patients having primary breast cancer with tumor which are “de novo” resistance or have incomplete response to docetaxel and another patients which have tumors sensitive to docetaxel.

Some Concepts used in the project

T statistic:

After obtaining the microarray data how can we tell that whether two genes expression levels are different or not. For that we use significance tests. There are many significance tests. But for our analysis we will use t-statistics.

In our analysis we have two sample groups:

Group 1: $X_1, X_2, ..., X_{N1}$ (Resistant Tumor)
Group 2: $Y_1, Y_2, ..., Y_{N2}$ (Sensitive Tumor)

$X_i \sim \text{Normal}(\mu_1, \sigma_1^2)$ $Y_i \sim \text{Normal}(\mu_2, \sigma_2^2)$ and the t-parameter can be found by

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{X_1X_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where

$$S_{X_1X_2} = \sqrt{\frac{(n_1-1)S_{X_1}^2 + (n_2-1)S_{X_2}^2}{n_1+n_2-2}}.$$

between a sample 1 and a sample 2.

Here, my Null Hypotheses is $\mu_1 = \mu_2$.

Degree of freedom:
d.f. = $n_1 + n_2 - 2$

After finding a t statistic we will find corresponding p value to find whether our Null Hypotheses is holding up or not.

But sometimes computing a *t*-statistic can be problematic because the variance estimates can be skewed by genes having a very low variance. These genes are associated to a large *t*-statistic and falsely selected as differentially expressed.

MA Plot:

This plot visualises the differences between measurements taken in two samples, by transforming the data onto M (log ratio) and A (mean average) scales, then plotting these values.

$$M = \log_2(R/G) = \log_2(R) - \log_2(G)$$
$$A = \frac{1}{2}\log_2(RG) = \frac{1}{2}(\log_2(R) + \log_2(G))$$

In simpler terms A is the average value of all the genes expression intensity value in all of the samples after taking the logarithm with base 2, and M is the difference of mean of Group 1 genes expression intensity value and Group 2 genes expression intensity value after taking the logarithm with base 2.

By taking logarithm of base 2 we are normalizing the data so that there are no any systematic biases. It is a standard practice to normalize the data but we can find whether our data requires normalization or not by plotting the MA plot with itself. If the plot is not having a slope of 1 then normalization is required.

When plotting the genomic data in MA plot we can find which genes are differentially expressed and which are not. In the plot, genes closer to the line y=0 are not differentially expressed. But as some genes goes further away from y=0 line they become more differentially expressed.

So we take genes above the line y=1 and below the line y=-1 are highly differentially expressed.

Heat Maps:

It contains top genes which are differentially expressed and the colour of each cell tells us about the intensity of that gene in that particular sample.

Dendrogram and unsupervised hierarchical clustering heat map of cell types (2 replicates each), using uncentered Pearson correlation and centroid linkage. The vertical distances on each branch of the dendrogram represent the degree of similarity between cell types' gene expression profiles.

Refer to appendix for the R codes.

Libraries Used:

Biobase – contains functions to see the expression of a genomic data
genefilter – to plot MA plot
affy – for finding the top 150 differential genes and the plotting the heat map
limma – for finding the t-statistic and plotting the heatmap

About the Dataset:

There are 24 subjects from which samples are taken and been given for the biopsy. Subjects are from GSM4901 to GSM4924.

Phenotypic Data

	Patient	disease.state	Tumour.type..	IMC/IDC	Age..years.	Menopausal.status	Ethnic.origin	Bidimensional.tumour.size..cm.
GSM4901	1	docetaxel resistant tumor		IMC	37	Premenopausal	Hispanic	10x10
GSM4902	2	docetaxel resistant tumor		IDC	55	Postmenopausal	Hispanic	10x8
GSM4903	3	docetaxel sensitive tumor		IDC	41	Premenopausal	Black	6x5
GSM4904	4	docetaxel resistant tumor		IMC	43	Premenopausal	Black	15x13
GSM4905	5	docetaxel resistant tumor		IDC	50	Postmenopausal	Black	20x23
GSM4906	6	docetaxel resistant tumor		IDC	55	Postmenopausal	Black	11x11
.	.	.		.	.	.	.	.
.	.	.		.	.	.	.	.

Dimensions - 24x15

Expression Measurements of Genes in the Subjects

IDs	GSM4901	GSM4902	GSM4903	GSM4904	GSM4905	GSM4906	GSM4907	GSM4908	.....
1000_at	217.23500	497.50500	435.51200	659.98500	199.16000	370.33500	470.84400	511.41200	....
1001_at	19.29000	23.60270	29.01300	25.61380	10.67300	28.78240	44.68040	86.54450	....
1002_f_at	84.09120	83.97030	58.91600	59.18540	125.21400	83.11780	67.90290	76.01460	....
1003_s_at	743.85000	487.22200	359.84800	627.87500	750.30900	395.35500	312.68300	567.88800	....
.	.	.	.	.	.	.	.	.	....
.	.	.	.	.	.	.	.	.	....
.	.	.	.	.	.	.	.	.	....

Dimension – 12625x24

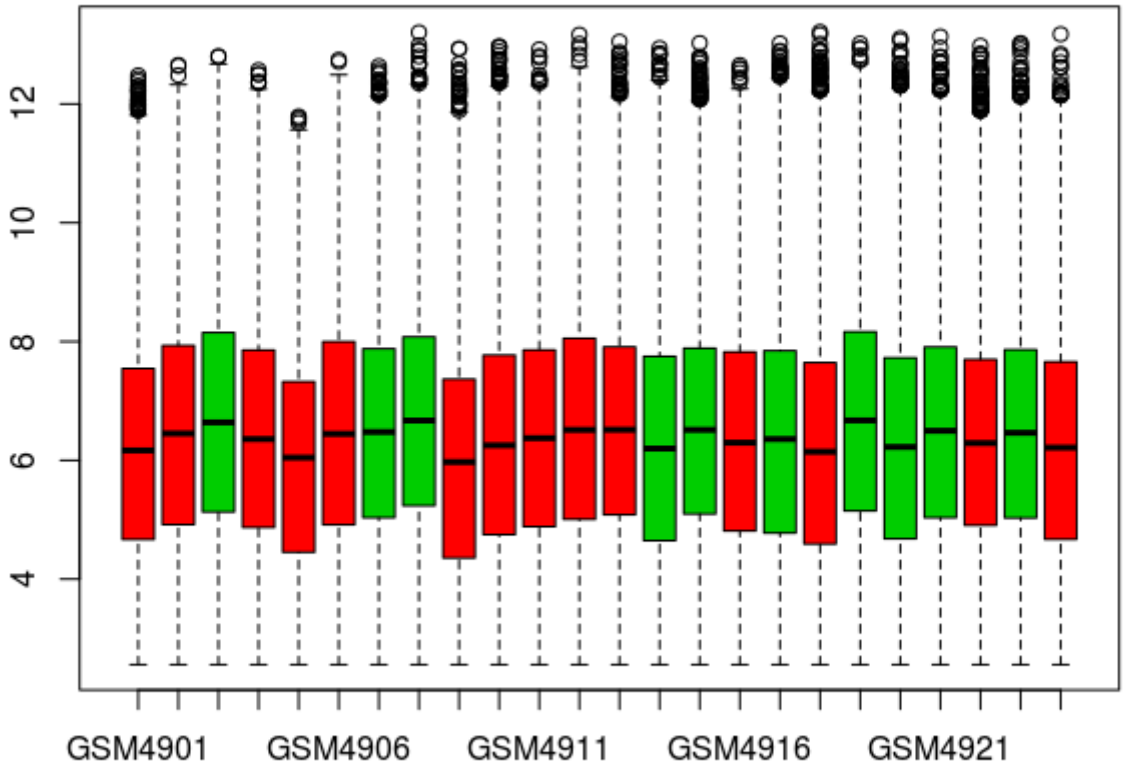
Number of different types of tumor

docetaxel resistant tumor	docetaxel sensitive tumor
14	10

Box Plot of Intensities of gene expressions of different Subjects

X axis : Subjects

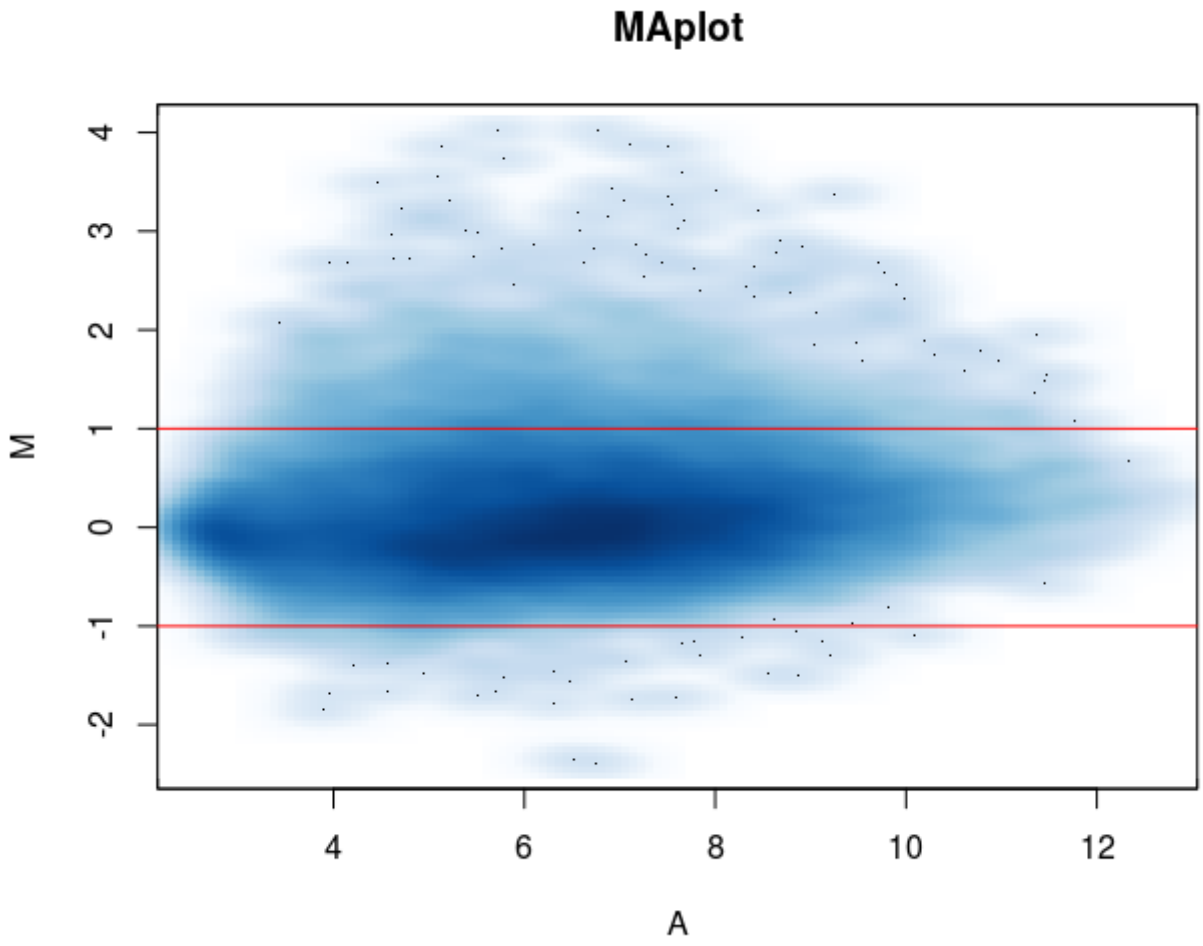
Y axis : Logarithm with base 2 of Intensities of genes



MA Plot

X axis: Average value of gene intensities of all the 24 subjects with Logarithmic base 2

Y axis: Difference of average values of genes intensities with Sensitive and Resistive Tumor with Logarithmic base 2



Here we can observe that some genes are densely populated near the line y=0. So therefore they are not differential. Whereas the genes which are above the line y=1 and y=-1 are differential genes.

Top 150 differential expressed genes with their t values:

	logFC	AveExpr	t	P.Value	adj.P.Val	B
36125_s_at	3.4850741	4.454124	6.653728	5.982479e-07	0.004487180	6.0419275
33781_s_at	2.6882577	3.949467	6.548652	7.730799e-07	0.004487180	5.8117976
40549_at	2.1311828	3.634828	6.372342	1.192014e-06	0.004487180	5.4221913
32523_at	2.6874932	4.146409	6.301000	1.421681e-06	0.004487180	5.2633437
37649_at	1.5074664	3.826876	6.106892	2.302449e-06	0.005813684	4.8277978
646_s_at	2.0080159	4.694693	5.810981	4.836420e-06	0.008029421	4.1549997
40090_at	2.7289659	4.790767	5.796914	5.011117e-06	0.008029421	4.1227692
39185_at	2.4124118	4.172302	5.684227	6.662699e-06	0.008029421	3.8638283
40439_at	1.6837762	3.325552	5.665401	6.988180e-06	0.008029421	3.8204405
35856_r_at	-1.0614755	4.475953	-5.651810	7.233107e-06	0.008029421	3.7890975
31331_at	3.0033322	5.377874	5.631550	7.614466e-06	0.008029421	3.7423395
40132_g_at	1.5912075	5.918078	5.630648	7.631925e-06	0.008029421	3.7402550
39812_at	3.1851058	5.284050	5.569204	8.920637e-06	0.008049677	3.5982028
40096_at	3.5486736	5.097301	5.568951	8.926375e-06	0.008049677	3.5976172
39076_s_at	1.9454016	5.243690	5.511735	1.032491e-05	0.008382223	3.4650217
40514_at	3.8542549	5.130584	5.480366	1.118370e-05	0.008382223	3.3922022
40867_at	1.4116427	7.217957	5.476760	1.128695e-05	0.008382223	3.3838250
38997_at	3.7356187	5.776386	5.394118	1.393617e-05	0.008522798	3.1915523
35763_at	1.8392718	4.170907	5.381633	1.438778e-05	0.008522798	3.1624563
40619_at	3.2345902	4.722115	5.381295	1.440020e-05	0.008522798	3.1616693
39180_at	4.0245668	5.727401	5.359486	1.522556e-05	0.008522798	3.1108134
39631_at	1.5033259	8.411806	5.357839	1.528979e-05	0.008522798	3.1069717
37313_at	2.1091645	4.680034	5.351825	1.552668e-05	0.008522798	3.0929410
39561_at	2.4720381	4.472510	5.310048	1.727784e-05	0.009088862	2.9953962
37768_at	1.7586714	4.948037	5.242709	2.053001e-05	0.010300854	2.8379035
1635_at	2.1074955	3.785590	5.208153	2.243197e-05	0.010300854	2.7569624
34157_f_at	2.7361816	5.466825	5.202763	2.274426e-05	0.010300854	2.7443300
40857_f_at	-0.6976574	6.736983	-5.183347	2.390592e-05	0.010300854	2.6988115
33404_at	2.5048070	4.556548	5.181674	2.400881e-05	0.010300854	2.6948868
1751_g_at	1.9658911	5.309627	5.174144	2.447728e-05	0.010300854	2.6772268
41528_at	2.1206943	5.009170	5.107250	2.906515e-05	0.011557569	2.5201875
38831_f_at	2.8701498	6.099273	5.104192	2.929443e-05	0.011557569	2.5130031

Showing 1 to 33 of 150 entries

	logFC	AveExpr	t	P.Value	adj.P.Val	B
34027_f_at	1.8532455	5.077245	5.082245	3.099450e-05	0.011848326	2.4614199
36830_at	1.8421043	5.297983	5.070943	3.190836e-05	0.011848326	2.4348465
41310_f_at	1.7814007	5.521797	5.031116	3.535021e-05	0.012426283	2.3411556
34082_at	1.5667983	6.487329	5.014959	3.685083e-05	0.012426283	2.3031249
39050_at	1.0224169	7.701295	5.004534	3.785289e-05	0.012426283	2.2785805
922_at	1.9269217	6.673194	4.990968	3.919805e-05	0.012426283	2.2466324
AFFX-HUMGAPDH/M33197_5_at	3.2735806	7.543266	4.986765	3.962452e-05	0.012426283	2.2367318
1053_at	1.5452673	3.450402	4.984610	3.984498e-05	0.012426283	2.2316552
36811_at	2.1192217	4.272943	4.967736	4.161426e-05	0.012426283	2.1919016
38618_at	3.1555736	4.991371	4.959143	4.254545e-05	0.012426283	2.1716514
318_at	1.3417354	5.603242	4.943766	4.426417e-05	0.012426283	2.1354103
197_at	1.4337150	7.059215	4.941837	4.448476e-05	0.012426283	2.1308610
38372_at	2.4901447	6.470307	4.937802	4.494944e-05	0.012426283	2.1213510
32331_at	2.1000006	6.141873	4.934993	4.527596e-05	0.012426283	2.1147269
31420_at	-1.1353072	4.051225	-4.925335	4.641658e-05	0.012468284	2.0919561
31638_at	1.7388529	6.415796	4.912210	4.801316e-05	0.012628461	2.0610038
36653_g_at	2.9642795	4.603955	4.894428	5.026446e-05	0.012950793	2.0190610
37329_at	2.5557159	6.436123	4.880783	5.206364e-05	0.013042289	1.9868685
38850_at	2.4032350	4.300344	4.871325	5.334853e-05	0.013042289	1.9645509
35844_at	2.5492067	5.597763	4.868643	5.371874e-05	0.013042289	1.9582203
35287_at	1.8815304	4.585307	4.860676	5.483349e-05	0.013061751	1.9394174
35695_at	1.9960517	4.444376	4.849148	5.648768e-05	0.013206610	1.9122075
41308_at	3.0700348	4.915820	4.840001	5.783575e-05	0.013262395	1.8906152
39030_at	1.8445763	8.109524	4.825960	5.996819e-05	0.013262395	1.8574649
36987_at	1.0542901	7.007192	4.820719	6.078415e-05	0.013262395	1.8450908
38784_g_at	2.7590338	7.282909	4.817725	6.125543e-05	0.013262395	1.8380192
39724_s_at	1.9293550	5.285213	4.813173	6.197872e-05	0.013262395	1.8272710
35807_at	2.6789481	6.624704	4.802567	6.369749e-05	0.013403014	1.8022243
AFFX-HSAC07/X00351_5_at	2.2089651	8.034107	4.791051	6.561799e-05	0.013580772	1.7750243
343_s_at	1.4230964	5.354417	4.784499	6.673648e-05	0.013589485	1.7595471
41075_at	-0.8822957	5.055686	-4.775200	6.835675e-05	0.013698476	1.7375801
1137_at	1.6582027	3.736683	4.764671	7.023890e-05	0.013836355	1.7127064

Showing 33 to 65 of 150 entries

	logFC	AveExpr	t	P.Value	adj.P.Val	B
39347_at	2.7041666	5.905443	4.754871	7.203731e-05	0.013836355	1.6895539
1590_s_at	1.8937010	4.222975	4.735203	7.578736e-05	0.013836355	1.6430787
41757_at	2.1596439	5.259715	4.734740	7.587787e-05	0.013836355	1.6419856
2011_s_at	1.9485791	3.931154	4.730602	7.669254e-05	0.013836355	1.6322050
40778_at	1.9254976	7.308201	4.722549	7.830286e-05	0.013836355	1.6131738
35015_at	-0.7246868	4.512886	-4.721916	7.843074e-05	0.013836355	1.6116793
1020_s_at	1.5797308	7.304913	4.718940	7.903545e-05	0.013836355	1.6046449
35655_at	1.2638364	3.593477	4.716081	7.962066e-05	0.013836355	1.5978884
41198_at	1.9831056	7.060662	4.700746	8.283448e-05	0.013836355	1.5616458
31936_s_at	2.3062745	6.417923	4.696355	8.377860e-05	0.013836355	1.5512657
762_f_at	1.9455143	6.101360	4.695589	8.394440e-05	0.013836355	1.5494548
38807_at	1.3402190	5.801953	4.694111	8.426502e-05	0.013836355	1.5459632
35733_at	2.0409529	7.605560	4.687953	8.561479e-05	0.013836355	1.5314080
40222_s_at	-0.7478771	7.324655	-4.685024	8.626440e-05	0.013836355	1.5244846
41261_at	1.1393748	3.367653	4.683610	8.657996e-05	0.013836355	1.5211401
40415_at	2.4953483	4.914822	4.671237	8.938919e-05	0.014106732	1.4918927
34413_at	1.7292668	3.617281	4.662444	9.144073e-05	0.014252336	1.4711083
330_s_at	1.4155602	5.462751	4.651762	9.399672e-05	0.014472055	1.4458559
32844_at	1.4535053	6.396888	4.634673	9.823527e-05	0.014942413	1.4054556
40060_r_at	2.1767398	5.054919	4.626196	1.004082e-04	0.014943559	1.3854149
36846_s_at	2.5797541	5.844983	4.624118	1.009480e-04	0.014943559	1.3805036
38613_at	2.1751467	7.079375	4.617768	1.026162e-04	0.014943559	1.3654898
38670_at	1.5526870	5.599604	4.616406	1.029774e-04	0.014943559	1.3622709
31340_at	-1.1506373	3.625392	-4.608873	1.049991e-04	0.015063786	1.3444616
40465_at	2.2313574	4.549728	4.587232	1.110304e-04	0.015750096	1.2932988
39846_at	1.9405003	3.949745	4.581398	1.127147e-04	0.015766737	1.2795064
36594_s_at	1.6569312	5.356382	4.578212	1.136454e-04	0.015766737	1.2719740
34536_g_at	-1.1155163	3.818305	-4.558285	1.196424e-04	0.016418321	1.2248666
36208_at	3.1082715	5.026413	4.551271	1.218280e-04	0.016473188	1.2082838
31488_s_at	1.9291538	3.947634	4.546432	1.233587e-04	0.016473188	1.1968454
41805_g_at	1.8367249	3.753416	4.544558	1.239567e-04	0.016473188	1.1924158
35626_at	0.6875344	8.191771	4.533412	1.275739e-04	0.016777300	1.1660664

Showing 65 to 98 of 150 entries

	logFC	AveExpr	t	P.Value	adj.P.Val	B
910_at	1.8371280	3.697196	4.523317	1.309405e-04	0.016969813	1.1422061
35217_at	0.8117210	7.306976	4.518498	1.325788e-04	0.016969813	1.1308153
1315_at	0.6653688	11.347872	4.511041	1.351546e-04	0.016969813	1.1131886
1355_g_at	-0.9254402	5.695798	-4.509319	1.357566e-04	0.016969813	1.1091174
1906_at	1.6918865	4.072959	4.509313	1.357585e-04	0.016969813	1.1091048
35889_at	-0.8530088	5.369999	-4.496703	1.402480e-04	0.017238257	1.0793016
32575_at	1.7681801	8.139508	4.495630	1.406369e-04	0.017238257	1.0767651
35934_at	1.9480837	3.769511	4.491628	1.420967e-04	0.017249720	1.0673055
38076_at	1.1053956	9.669974	4.484889	1.445886e-04	0.017322245	1.0513802
40912_s_at	1.0613053	7.006691	4.478414	1.470243e-04	0.017322245	1.0360777
40064_at	2.1650042	5.680532	4.477182	1.474923e-04	0.017322245	1.0331663
33697_at	0.8756105	6.855551	4.475372	1.481824e-04	0.017322245	1.0288907
37335_at	1.5311603	4.656950	4.468279	1.509189e-04	0.017480281	1.0121287
2085_s_at	1.9696862	6.642545	4.463566	1.527648e-04	0.017533237	1.0009923
38458_at	1.5549565	8.419487	4.458757	1.546716e-04	0.017560739	0.9896299
39018_at	2.0748806	7.128288	4.455973	1.557864e-04	0.017560739	0.9830514
38686_at	1.6610589	7.557517	4.449365	1.584642e-04	0.017647494	0.9674400
38373_g_at	1.7361709	6.133914	4.445425	1.600828e-04	0.017647494	0.9581312
41858_at	2.1307617	4.971581	4.436205	1.639349e-04	0.017647494	0.9363505
1486_at	2.2405409	7.309956	4.435504	1.642314e-04	0.017647494	0.9346949
1969_s_at	1.9034515	5.189406	4.433911	1.649077e-04	0.017647494	0.9309307
31527_at	1.1601365	11.165315	4.421608	1.702234e-04	0.017647494	0.9018708
40535_i_at	-1.4332343	6.302533	-4.419464	1.711671e-04	0.017647494	0.8968069
39580_at	1.4103912	3.648226	4.418747	1.714842e-04	0.017647494	0.8951116
33214_at	2.2658829	4.760019	4.417990	1.718192e-04	0.017647494	0.8933238
1358_s_at	-1.4993044	8.877121	-4.415262	1.730320e-04	0.017647494	0.8868811
33836_at	1.8804473	5.688008	4.414643	1.733081e-04	0.017647494	0.8854206
1861_at	0.9728303	4.367269	4.411066	1.749140e-04	0.017647494	0.8769725
33818_at	1.4989009	7.129169	4.409825	1.754746e-04	0.017647494	0.8740417
508_at	1.2331280	6.652281	4.408389	1.761255e-04	0.017647494	0.8706504
1272_at	1.8366976	3.730097	4.401320	1.793652e-04	0.017830591	0.8539556
31827_s_at	1.5467649	4.618071	4.392577	1.834539e-04	0.018094576	0.8333108

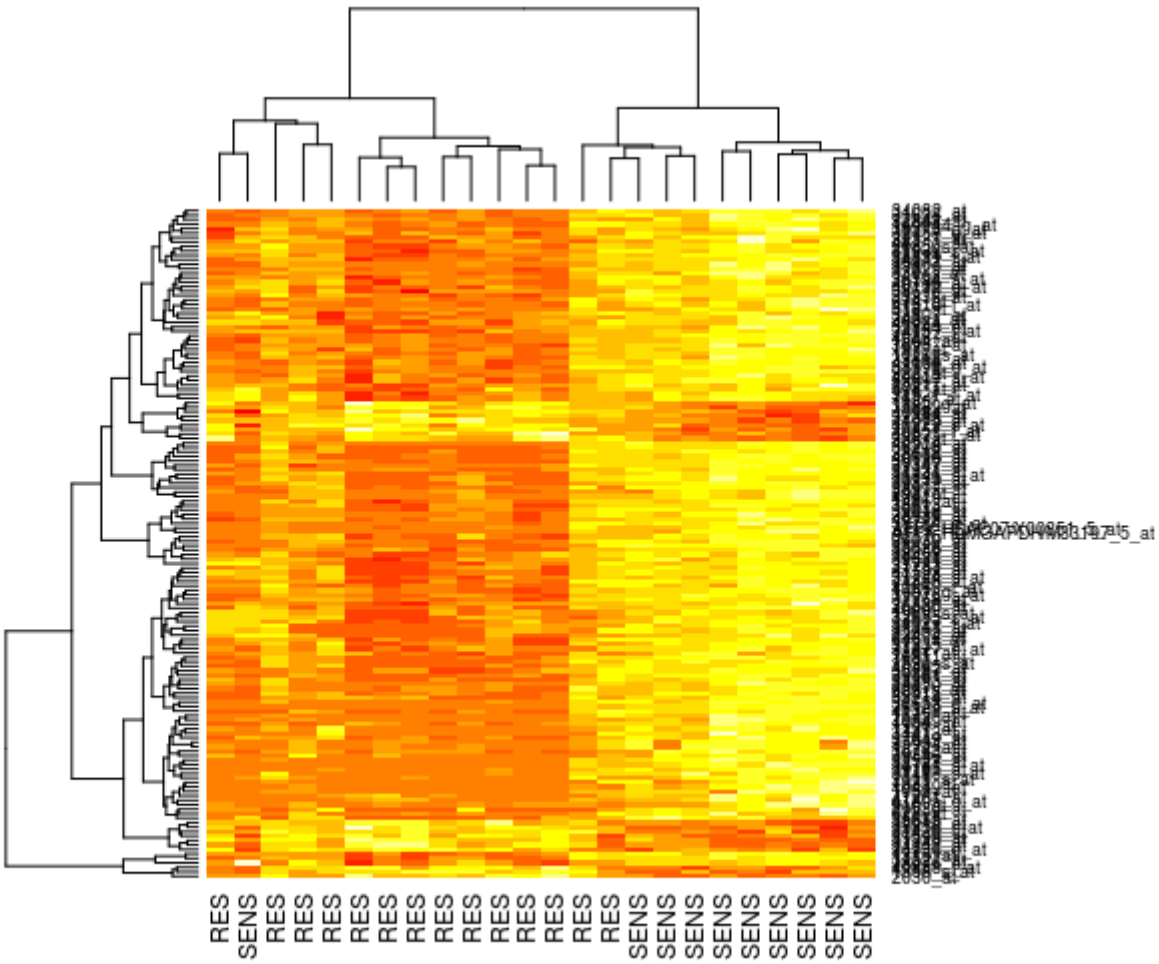
Showing 98 to 130 of 150 entries

32758_g_at	1.2537803	6.047572	4.356853	2.011478e-04	0.019385431	0.7489796
33393_at	2.2703296	4.554921	4.340319	2.099024e-04	0.019804434	0.7099632
336_at	-0.8300566	6.117102	-4.340002	2.100743e-04	0.019804434	0.7092135
31521_f_at	2.3024650	6.763744	4.339767	2.102015e-04	0.019804434	0.7086592
40136_at	1.0959193	5.546106	4.331607	2.146670e-04	0.019838914	0.6894083
2030_at	-0.7635970	8.341510	-4.330822	2.151016e-04	0.019838914	0.6875559
33133_at	2.9777699	5.512804	4.328283	2.165131e-04	0.019838914	0.6815664
40984_at	-0.6595235	5.595581	-4.327436	2.169858e-04	0.019838914	0.6795691
32383_at	-1.0697129	5.166963	-4.324871	2.184245e-04	0.019838914	0.6735174
1382_at	1.5070869	3.370051	4.312487	2.255039e-04	0.020070524	0.6443079
34343_at	-1.0291090	3.407013	-4.311840	2.258799e-04	0.020070524	0.6427822
634_at	1.5466439	5.865605	4.311051	2.263395e-04	0.020070524	0.6409211
32655_s_at	1.6821241	4.999455	4.309350	2.273335e-04	0.020070524	0.6369085
160044_g_at	1.8481548	6.009050	4.304619	2.301200e-04	0.020167271	0.6257523
39599_at	-1.0193418	3.942973	-4.301272	2.321118e-04	0.020167271	0.6178608
40888_f_at	-1.2948510	9.215957	-4.299420	2.332215e-04	0.020167271	0.6134933
38998_g_at	2.0193953	6.997941	4.282552	2.435734e-04	0.020919146	0.5737266
893_at	2.1001694	5.282143	4.276694	2.472744e-04	0.021093510	0.5599189
628_at	1.4303137	3.643075	4.255024	2.614568e-04	0.022153636	0.5088573
626_s_at	2.0388121	4.537869	4.241601	2.706449e-04	0.022652932	0.4772370

Showing 120 to 150 of 150 entries

Heat Map Plot of top 150 differential expressed genes:

The higher the gene expression intensity of the cell, darker is its color.



The dendrograms along the sides show how the variables and the rows are independently clustered. The heat map shows the data value for each row and column. Any patterns in the heat map indicate an association between the rows and the columns.

Appendix

- R code

```
library(Biobase)
library(genefilter)
library(affy)

load("~/Desktop/Bio Project/chang03.rda")   #Loads the Data
cat(abstract(experimentData(chang03)))      #Gives the intro of the data

dim(exprs(chang03))                         #Checking the dimensions
exprs(chang03)                              #Shows the genes expression intensity of all the 24 subjects
dim(pData(chang03))                         #Checking the dimensions of phenotype data
pData((chang03))                            #Displays the phenotype data
names(pData(chang03))                       #Displays the covariates of phenotype data
pData(chang03)$disease.state                #Displays the column disease.state in phenotype data
table(pData(chang03)$disease.state)         #Displays the freq of Sensitive and Resistant tumor
table(pData(chang03)$disease.state,pData(chang03)$Progest)
                                            #Displays the freq of  types of tumors against Progest
boxplot(exprs(chang03))                     #Shows the boxplot
y<-log2(exprs(chang03))                     #Normalizing the data
boxplot(y,col=as.numeric(pData(chang03)$disease.state)+1)
                                            #Plotting the Normalized data in box plot

Index<-as.numeric(pData(chang03)$disease.state)
d<-rowMeans(y[,Index==2])-rowMeans(y[,Index==1])
a<-rowMeans(y)
smoothScatter(a,d,main="MAplot",xlab="A",ylab="M")
abline(h=c(-1,1),col="red")                 #MA Plot

g<-y[23,]
m<-mean(g[Index==1])-mean(g[Index==2])
plot(jitter(Index),g,col=Index+1,xaxt="n",xlab="Patient Type")
axis(1,labels=c("RES","SENS"),at=1:2)
                                            #DifferentialExpressionAnalysis

library(limma)
design<-model.matrix(~factor(chang03$disease.state))
fit<-lmFit(y,design)
ebayes<-eBayes(fit)
tab<-topTable(ebayes,coef=2,adjust="fdr",n=150)
labCol<-c("RES","SENS")[Index]
heatmap(y[rownames(tab),],labCol=labCol)
                                            #DifferentialExpression
```

References

http://orange-bioinformatics.readthedocs.io/en/latest/widgets/maplot.html

http://www.nathalievilla.org/doc/pdf/tutorial-rnaseq.pdf

https://www.bioconductor.org/help/course-materials/2006/biocintro_oct/lectures/DifferentialGenes.pdf

https://www.ebi.ac.uk/training/online/course/functional-genomics-ii-common-technologies-and-data-analysis-methods/biological-0

https://stats.stackexchange.com/questions/53111/interpreting-cluster-heat-maps-from-r

https://www.nature.com/articles/srep32249/figures/2

https://en.wikipedia.org/wiki/T-statistic