DH-307 : Week 1&2

By:
Sahil Barbade

TRANSCRIPTOMICS DATASET

METABOLOMICS DATASET

PROTEOMICS DATASET

About Dataset

Worked first on transcriptomics dataset

Had meta-data giving description of each sample (like gender, age, part of brain, AD_10?)

And contained main gene expression data where for each sample gene expression of 20864 genes ,255 samples taken.

Analysis done over sample-wise to see how these sample's different genes expression varied over samples range and how these samples are correlated with each other.

20864

```
input_data.head(5)
```

	probeId	geneName	S1	S2	S3	S4	S5	S6	S7	S8	...	S244	S245	S246	S247	S248	S249	S250	S251	
0	224372_at	ND4	14.447004	14.488466	14.359237	14.447004	14.371623	14.309105	14.463468	14.371623	...	14.431147	14.471096	14.476100	14.405654	14.338739	14.477873	14.468459	14.514001	14.28
1	1553570_x_at	COX2	14.342658	14.319904	14.301045	14.311038	14.278164	14.181809	14.369584	14.274293	...	14.274278	14.387093	14.395561	14.318622	14.066148	14.314094	14.201326	14.315090	14.14
2	1553567_s_at	ATP6	14.274753	14.535888	14.169766	14.355387	14.137347	14.125699	14.493597	14.127482	...	14.276798	14.307191	14.399148	14.362724	13.931964	14.291478	14.155846	14.345779	13.95
3	1553538_s_at	COX1	14.131576	14.201428	14.109259	14.207664	14.102998	13.971966	14.338718	13.936151	...	14.066920	14.196323	14.213274	14.180875	13.870975	14.208501	13.761009	14.111776	13.82
4	211296_x_at	UBC	13.724726	13.681150	13.887249	13.771470	13.966789	14.064666	13.583540	13.836327	...	13.911063	13.706277	13.640640	13.752444	13.908008	13.948781	13.868106	13.796900	14.05

5 rows × 255 columns

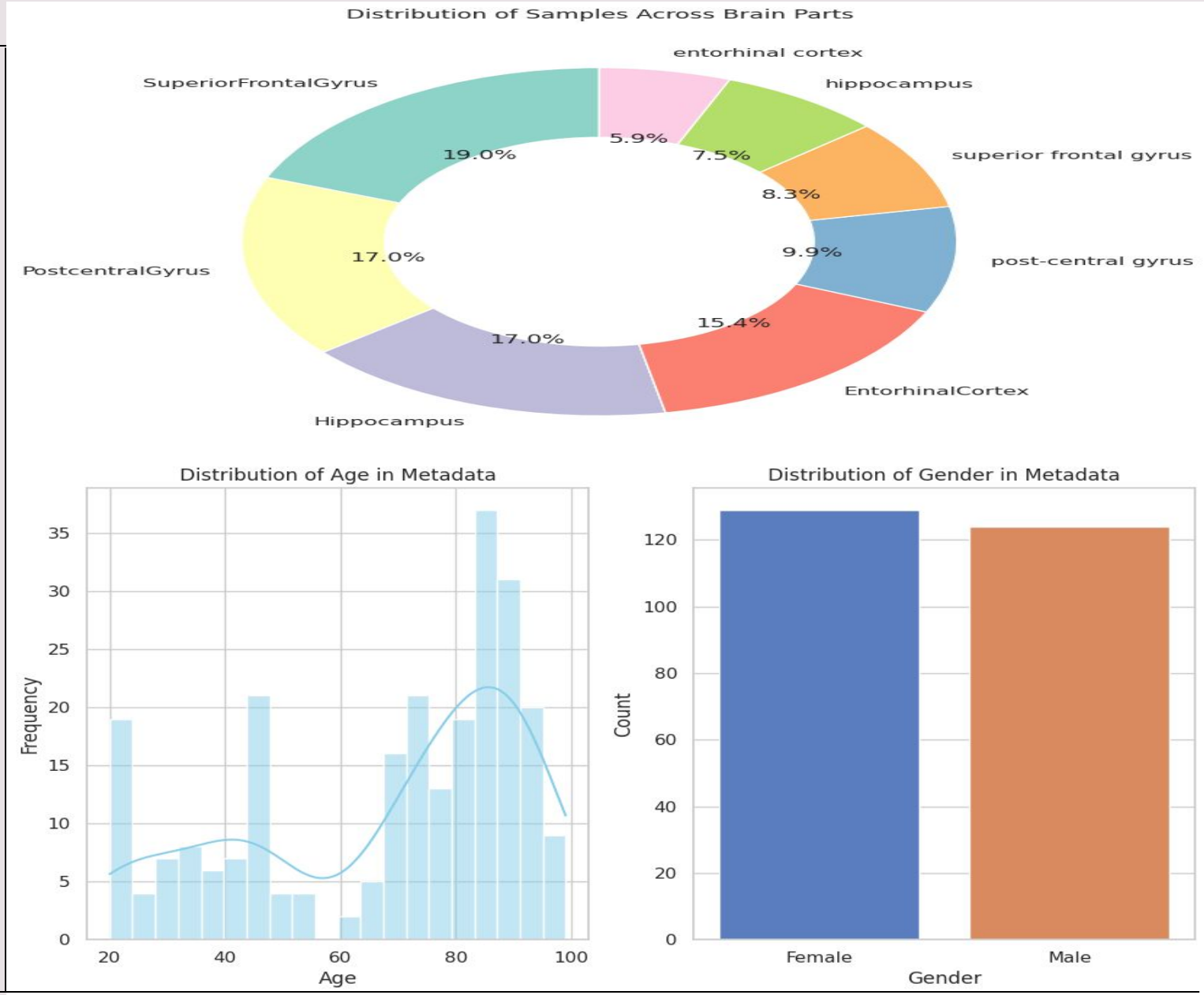
```
meta_data
```

	accession	sampleId	title	sourceName	gender	age	groupI	groupII
0	GSM300166	S1	PostcentralGyrus_female_91yrs_indiv10	brain, postcentral gyrus, female, 91 years	1	91	0	NaN
1	GSM300167	S2	SuperiorFrontalGyrus_female_91yrs_indiv10	brain, superior frontal gyrus, female, 91 years	1	91	0	NaN
2	GSM300168	S3	Hippocampus_female_96yrs_indiv105	brain, hippocampus, female, 96 years	1	96	0	NaN
3	GSM300169	S4	Hippocampus_male_82yrs_indiv106	brain, hippocampus, male, 82 years	0	82	0	NaN
4	GSM300170	S5	Hippocampus_male_84yrs_indiv108	brain, hippocampus, male, 84 years	0	84	0	NaN
...	...	...	...	...	...	...	...	...
248	GSM1176271	S249	superior frontal gyrus_male_85_AD_74	superior frontal gyrus_male_AD	0	85	1	NaN
249	GSM1176272	S250	superior frontal gyrus_male_86_AD_92	superior frontal gyrus_male_AD	0	86	1	NaN
250	GSM1176273	S251	superior frontal gyrus_male_94_AD_44	superior frontal gyrus_male_AD	0	94	1	NaN
251	GSM1176274	S252	superior frontal gyrus_male_94_AD_5	superior frontal gyrus_male_AD	0	94	1	NaN
252	GSM1176275	S253	superior frontal gyrus_male_94_AD_90	superior frontal gyrus_male_AD	0	94	1	NaN

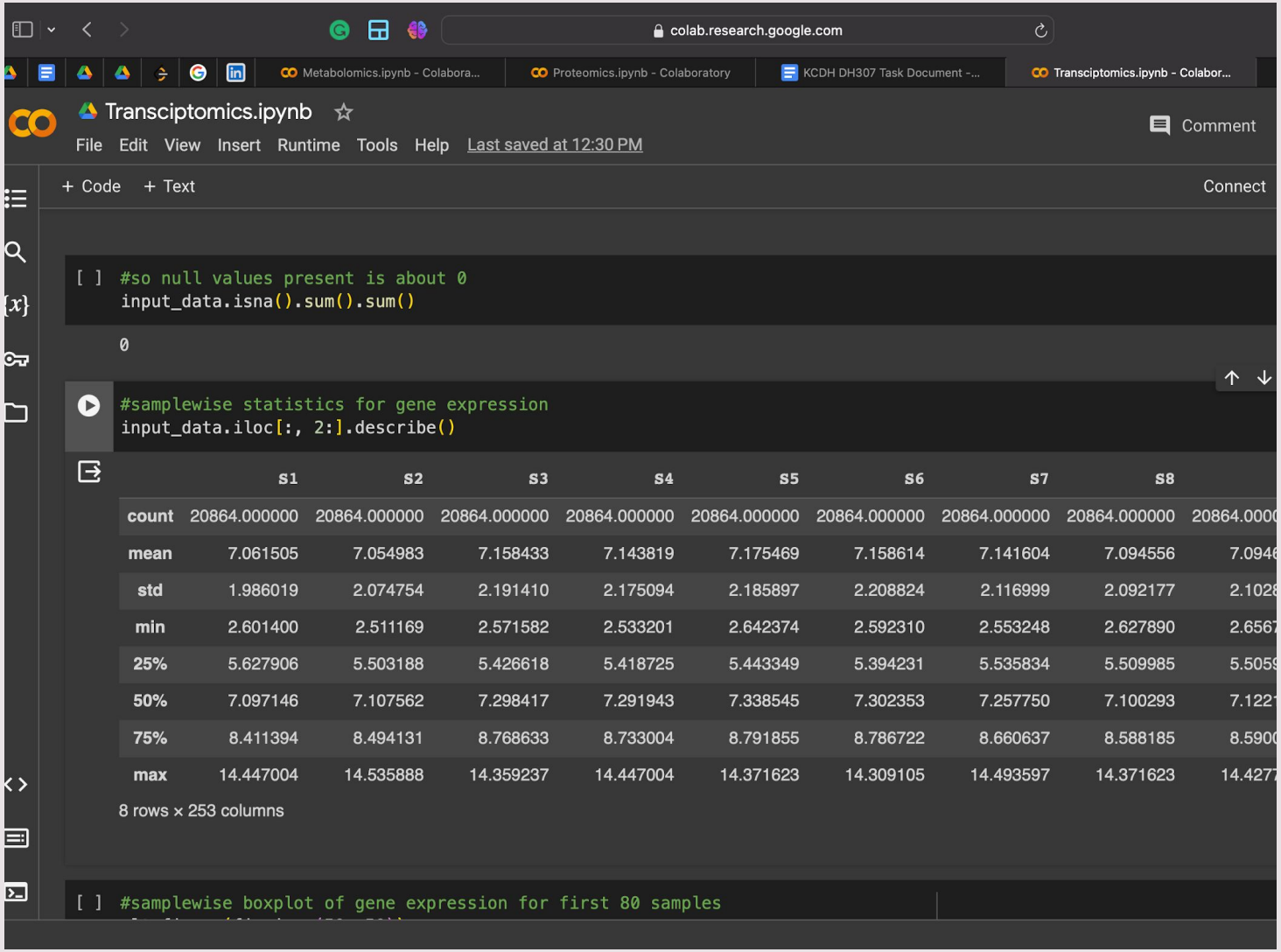
253 rows × 8 columns

Meta Data Analysis

- Sample is taken from wide range of ages (20-100 years) but mainly above 70 years patients.
- Distribution of Samples over gender is quite uniform and almost similar among males and females.
- These features could also be potential good indicator along with RNA-seq data and can be included later during feature selection process.



Descriptive Statistics of the Genes Data



Some inference on these Statistics

The mean values for each sample range from 7.061505 to 7.175469, indicating a relatively consistent central tendency across the samples.

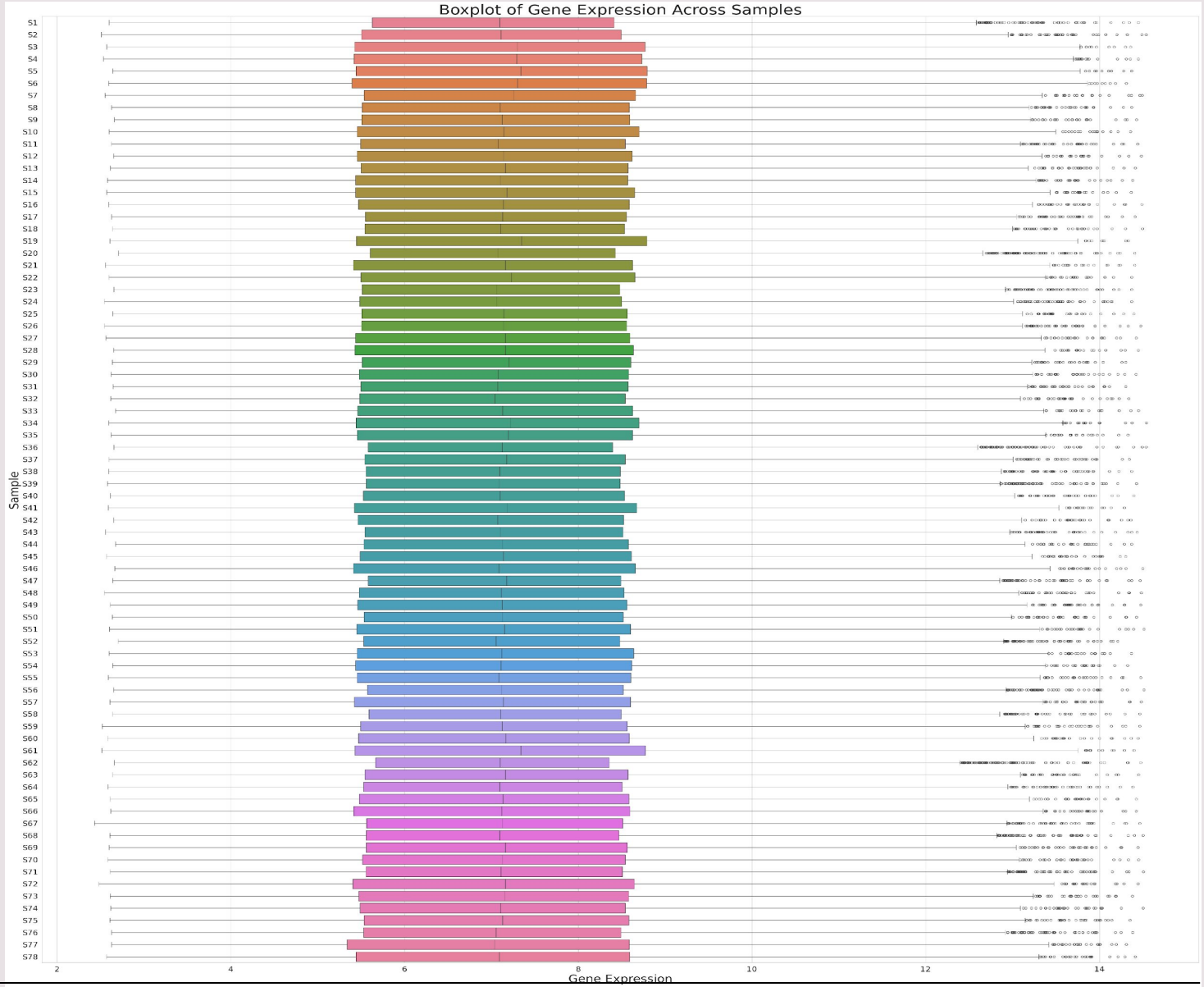
The median values (50th percentile) are close to the means, suggesting symmetric distributions.

The comparison of mean and median can provide insights into the skewness of the data. Since mean and median are close, the distribution is likely symmetric and less skewness presence.

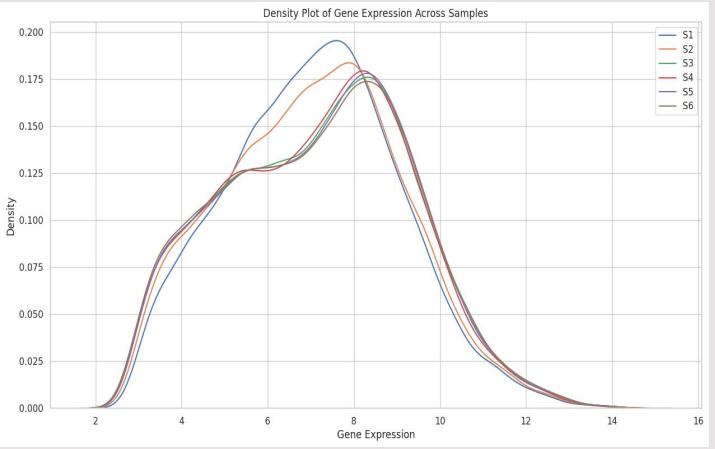
No sample exhibits extreme outliers based on a comparison of maximum values and the 75th percentile. We could say that there isn't a substantial technique difference and samples follow similar patterns in terms of central tendency and spread.

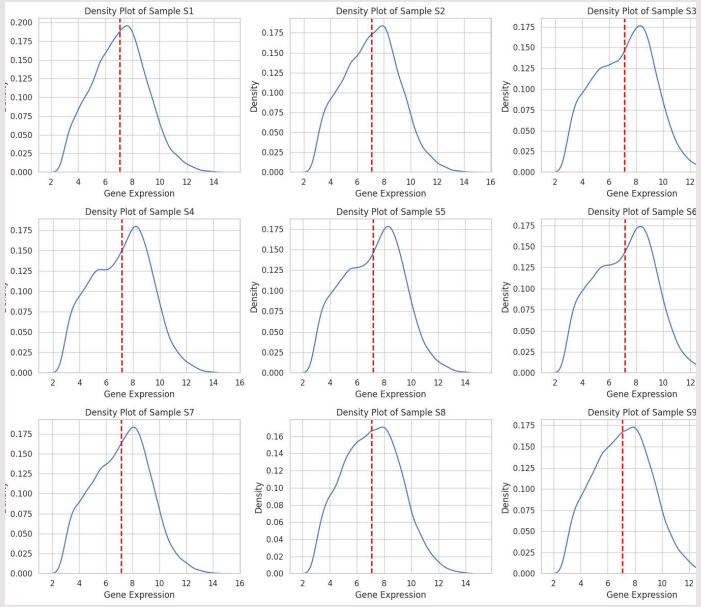
Sample Wise Distribution over gene expression (Box Plot)

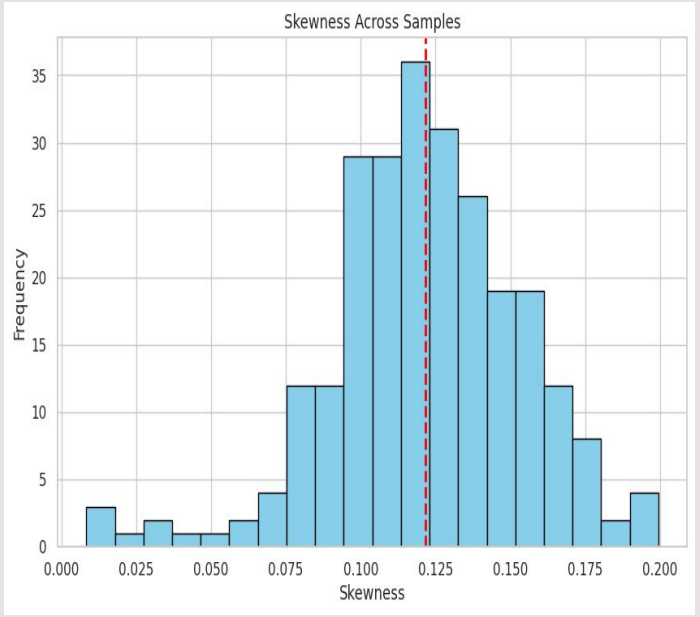
- Though large outliers are present but they aren't at extreme scale range.
- These outliers aren't removed since disease classifier can be possibly discovered in these outlier points or range and sample's data present is low.
- S19 & S61 have more distant mean and more spread due to some heterogenity from environment condition, experimental technique etc or it might represent subgroups with different characteristics or behavior



Sample Wise Gene Expression Density

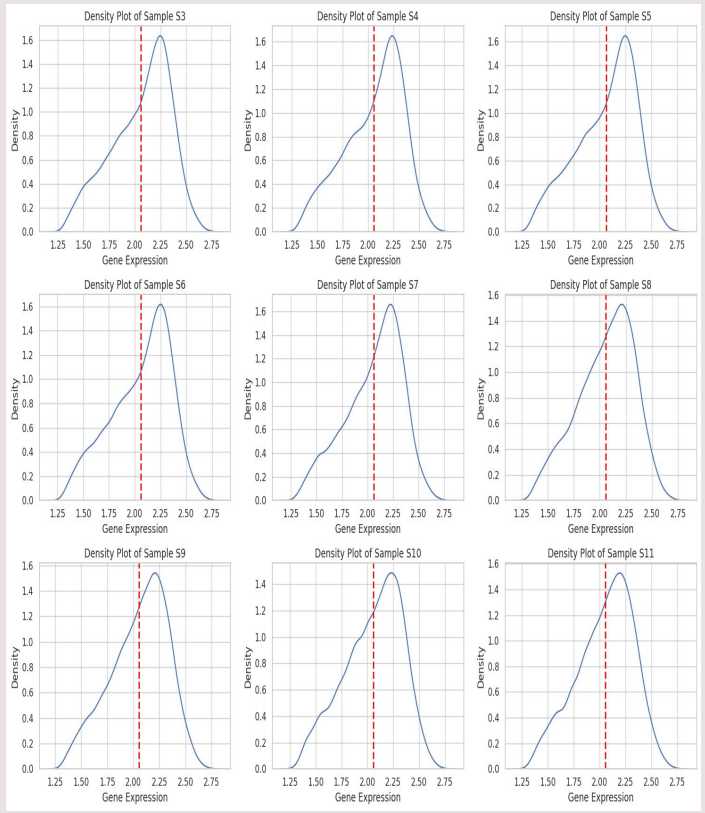




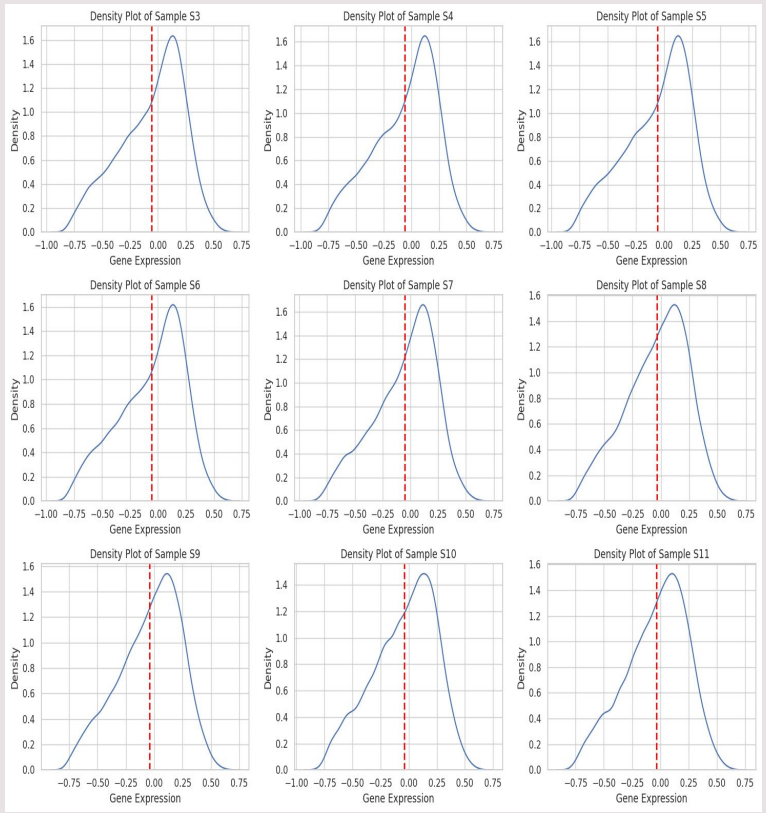


- Here median of all sample's density plot aren't aligned but vary so median normalization used.
- Additionally, though their central tendency are close but their scale over they vary aren't aligned and scaling is required (MAD Scaling).
- Though these plots look not much skewed but symmetric to some extent they are slightly positively skewed but with normal threshold i.e., Skewness <0.5 .

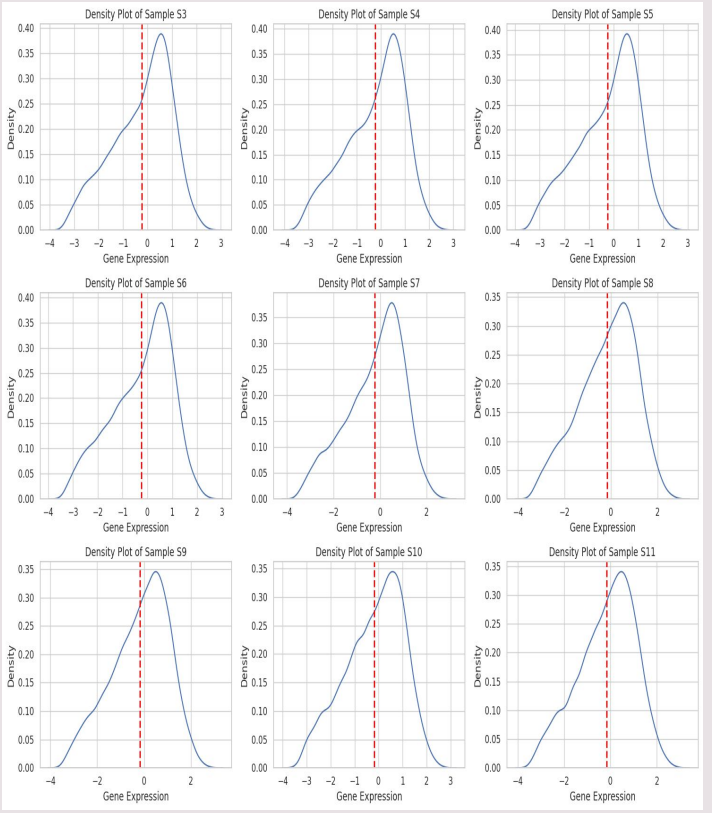
Sample Wise Gene Expression Preprocessing



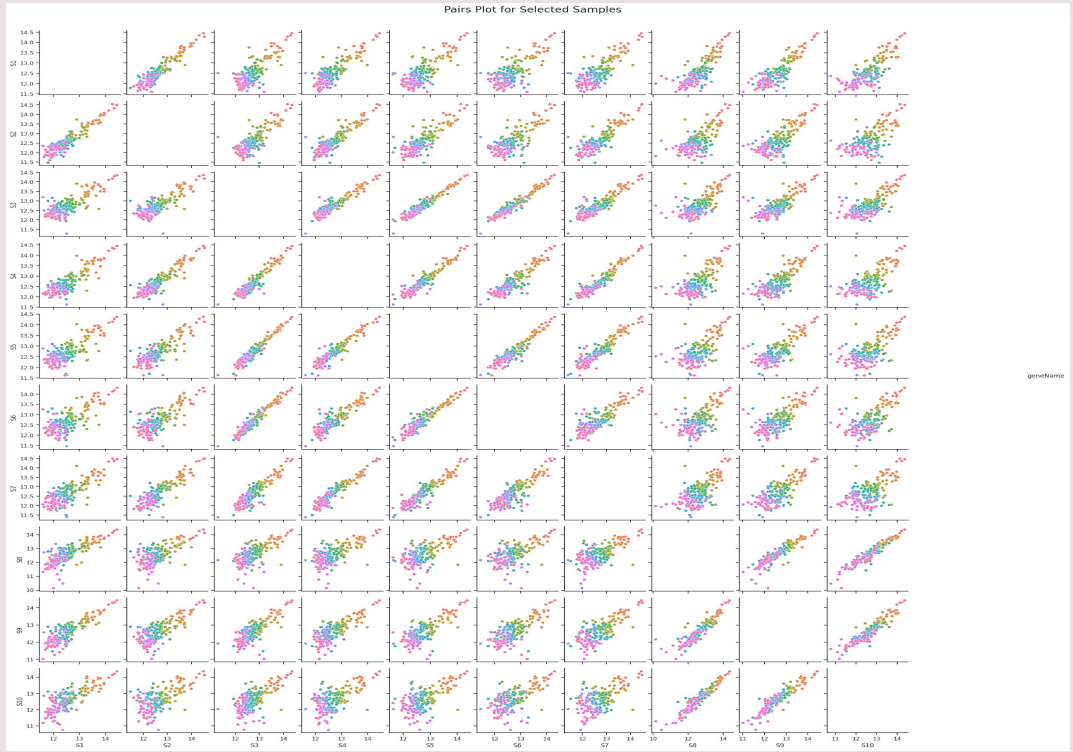
After Log transformation

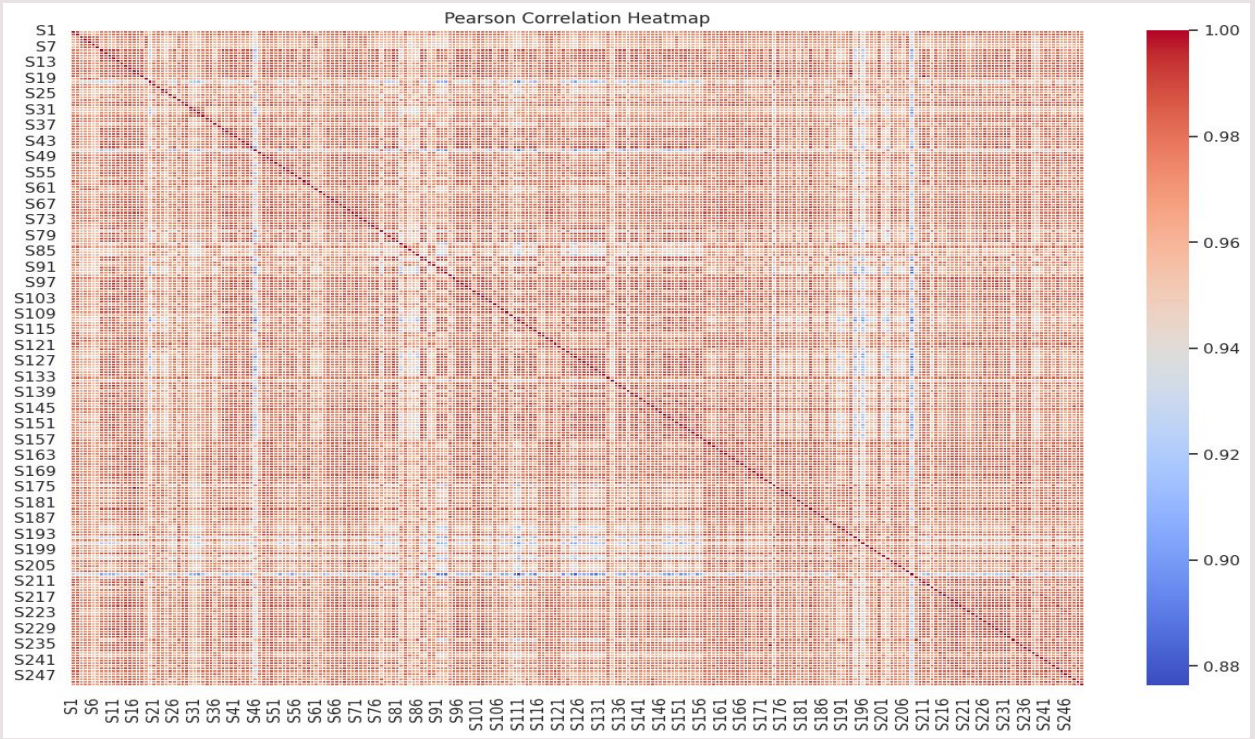


After Median Normalization



After MAD Scaling

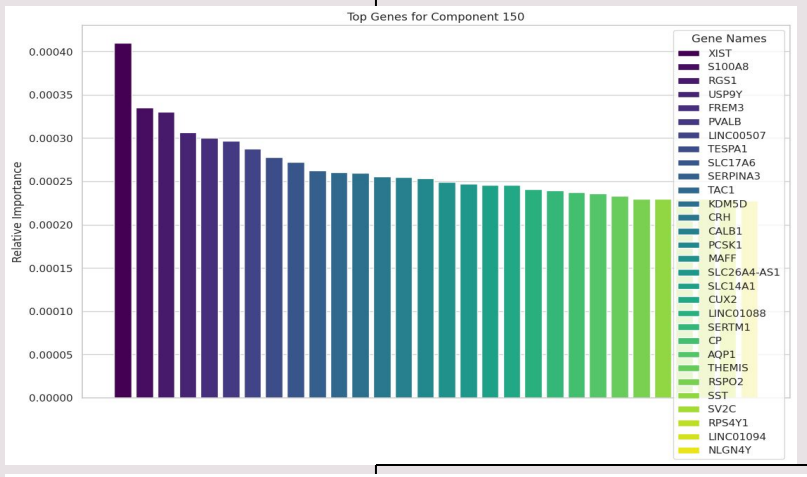


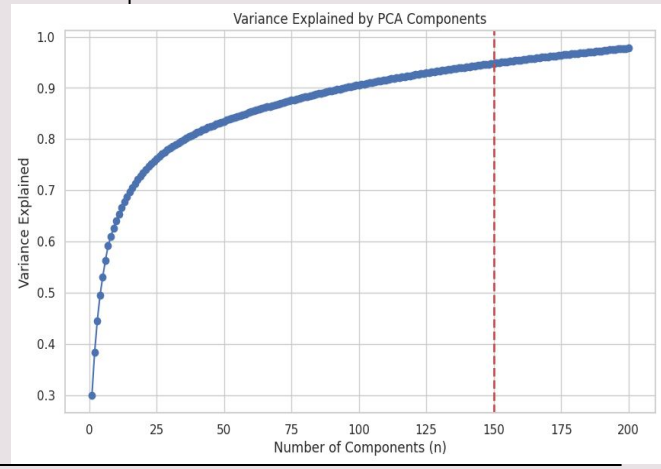


Relationships & Correlations Sample Wise

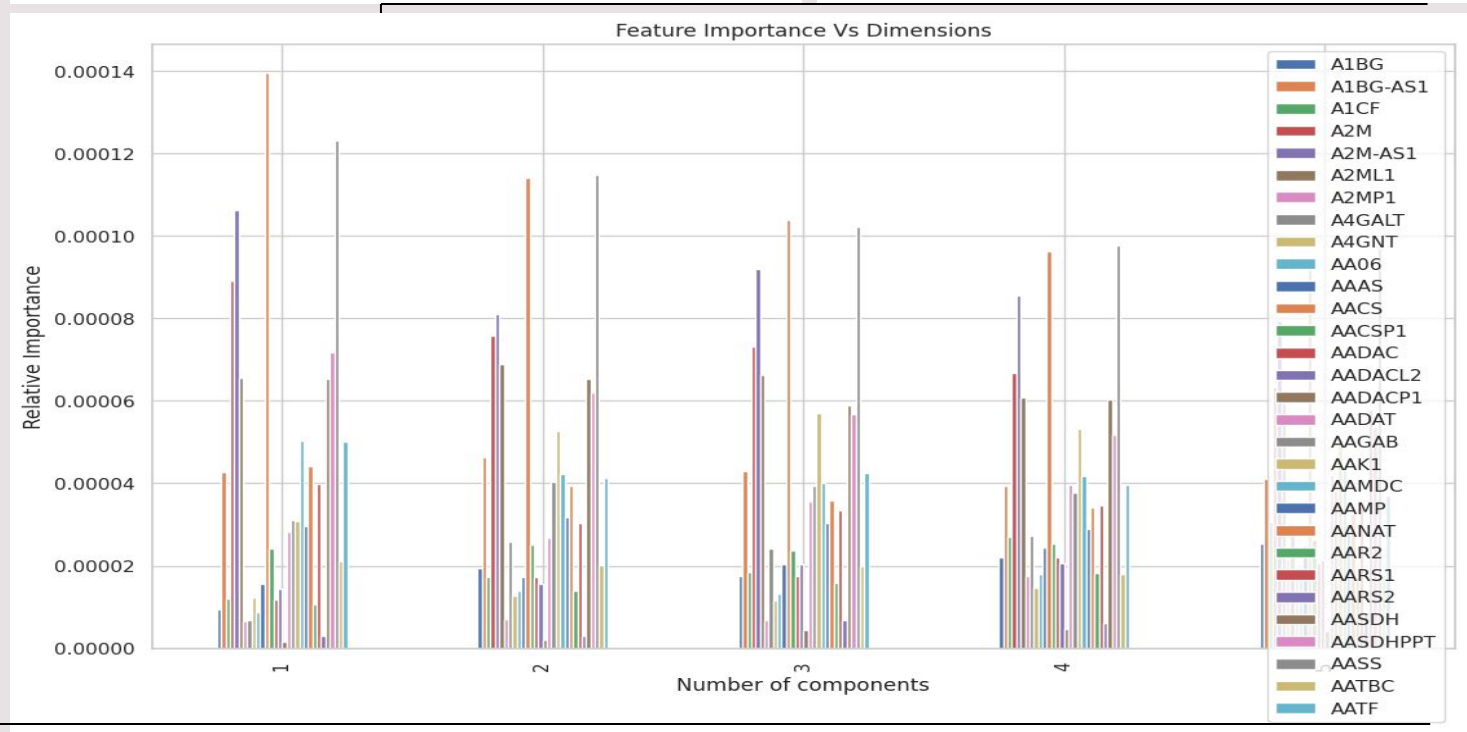
- Sample-wise they are highly correlated and are linear in relationship mostly with positive slope.
- Pearson was used since it is best measure for correlation of such linear relationship.
- From correlation heatmap one inference can be that neighbouring samples relate or correlate in more magnitude that those far way from it i.e., conditions use to produce or extract that neughbouring sample are similar in nature.

PCA Analysis on Genes Features



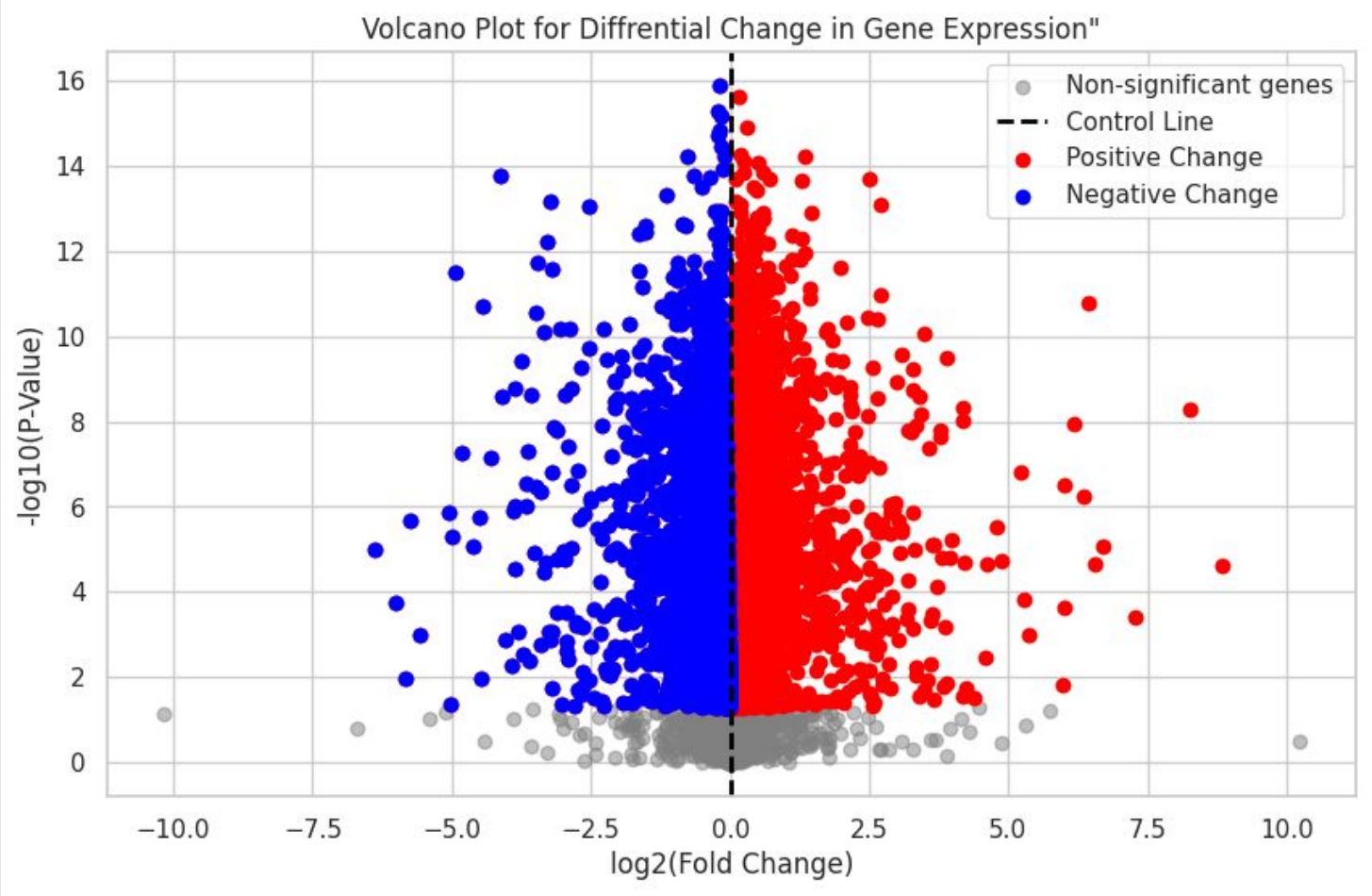


- PCA done to reduce the total number of gene expression for each sample to smaller component i.e., transforming the data into a lower-dimensional space while preserving most of the variability
- By 150th component it captures about 95% of total variance so reduction till this component is desirable.
- Each component linear combination of genes types and its weightage in it decides in priority or importance.
- Since our dataset has more features than sample as instance so PCA has to be applied to avoid dimensionality curse or overfitting while applying clustering or classifying algorithms like SVM, KNN etc.
- PCA technique used since others like PLSDA are sensitive to overfitting if feature size >> sample numbers and ICA requires these source to be completely independent which is not the case always or in this data.



Differential Expression Change

- To compare one experimental group versus a second one (or more) in order to find out which genes/transcripts change significantly between conditions this differential expression analysis is performed.
- Log2FC is used to quantify the change in expression levels between two conditions and bring them on same scale by taking log base to 2.
- P-value give statistical significance or probability to whether reject (>0.05) or accept that the observed difference of means is not the result of a real effect but just random noise for each gene.



- This plot shows that genes points are too close and slight divergent which infer that genes are changing with condition or noise, but the magnitude of change is relatively small.

About Dataset

Worked Second on metabolomics dataset

The samples were divided in two groups.

And contained main compounds expression data where for each sample compound expression of 5206 genes ,28 samples taken.

Analysis done over sample-wise to see how these sample's different compoun expression varied over samples range and how these samples are correlated with each other.

```
5206
```

```
[ ] input_data.head(5)
```

	Sample	S1	S2	S3	S4	S5	S6	S7	S8	S9	...	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28
0	Compound Annotation	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	2.000000	...	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000
1	Compound 468	7.475505	7.329756	6.811946	7.385248	7.626119	7.590575	7.403692	7.287125	7.355329	...	7.507965	7.435537	7.330367	7.367307	7.418969	7.351389	7.670616	7.412378	7.425305	7.367272
2	Compound 3382	6.989769	7.080236	6.889824	6.213607	6.722271	7.013836	7.060645	6.783460	7.001999	...	7.101507	6.887513	6.854224	7.062717	6.753966	6.863752	6.722227	6.822723	6.745978	6.775755
3	Compound 4793	6.819188	6.868873	6.843599	6.922629	6.833867	6.737740	6.826248	6.793886	6.845979	...	6.743371	6.835760	6.805242	6.758864	6.780618	6.735840	6.859521	6.907125	6.843397	6.905871
4	Compound 2186	6.858738	6.787733	6.856540	6.970408	6.947345	6.949742	7.394426	6.916902	6.692923	...	6.995888	6.905149	6.875869	6.914752	6.960787	6.756019	6.932341	6.792646	6.878472	6.970429

5 rows × 29 columns

```
orignal_data = input_data
input_data = input_data.drop(0,axis = 0)
input_data
```

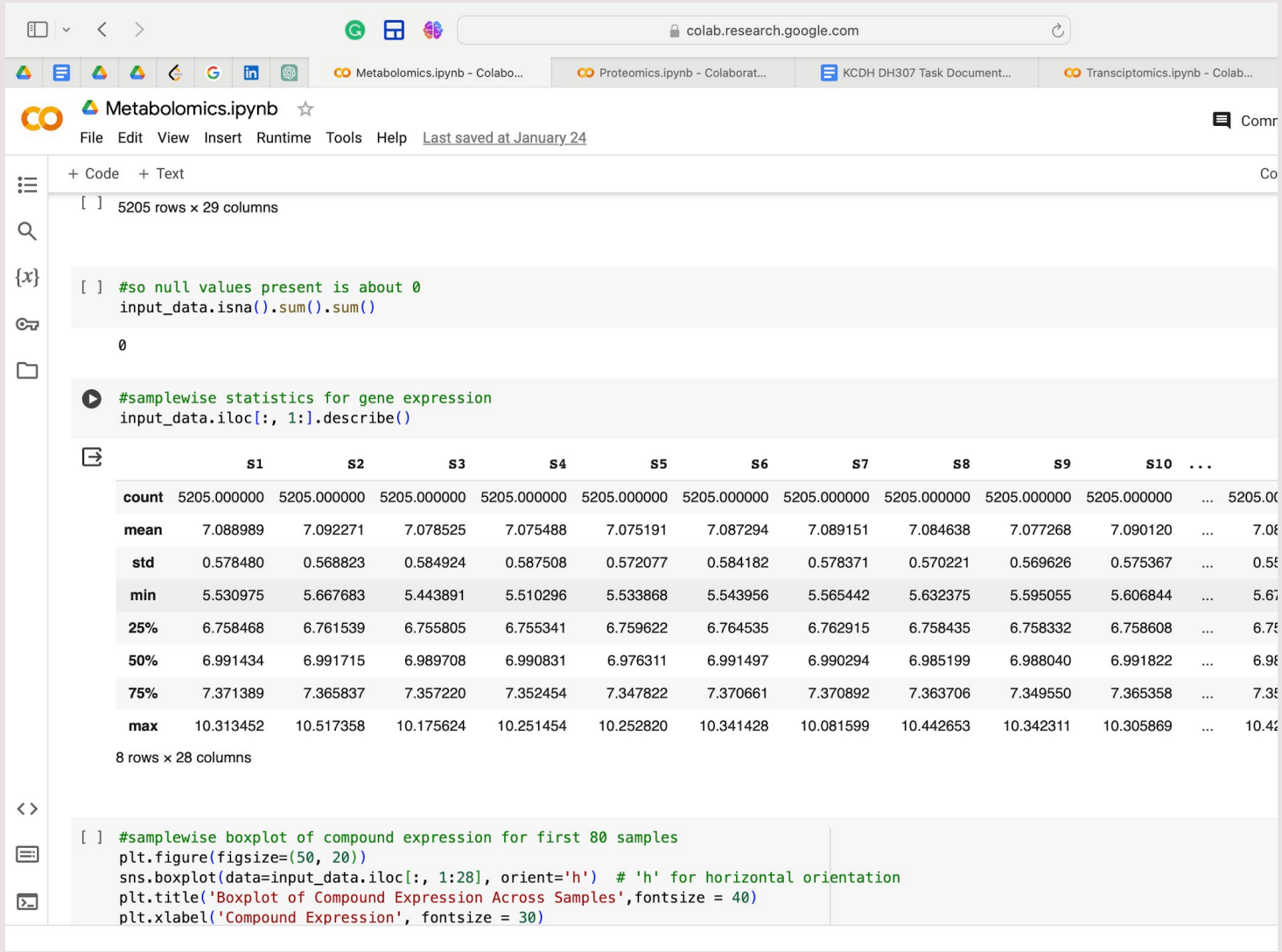
	Sample	S1	S2	S3	S4	S5	S6	S7	S8	S9	...	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28
1	Compound 468	7.475505	7.329756	6.811946	7.385248	7.626119	7.590575	7.403692	7.287125	7.355329	...	7.507965	7.435537	7.330367	7.367307	7.418969	7.351389	7.670616	7.412378	7.425305	7.367272
2	Compound 3382	6.989769	7.080236	6.889824	6.213607	6.722271	7.013836	7.060645	6.783460	7.001999	...	7.101507	6.887513	6.854224	7.062717	6.753966	6.863752	6.722227	6.822723	6.745978	6.775755
3	Compound 4793	6.819188	6.868873	6.843599	6.922629	6.833867	6.737740	6.826248	6.793886	6.845979	...	6.743371	6.835760	6.805242	6.758864	6.780618	6.735840	6.859521	6.907125	6.843397	6.905871
4	Compound 2186	6.858738	6.787733	6.856540	6.970408	6.947345	6.949742	7.394426	6.916902	6.692923	...	6.995888	6.905149	6.875869	6.914752	6.960787	6.756019	6.932341	6.792646	6.878472	6.970429
5	Compound 4227	7.067272	6.971488	6.868666	6.578207	6.821829	6.922966	7.076807	6.719544	7.012509	...	6.852931	6.973170	6.956334	6.918005	6.927555	7.117887	6.980225	6.942907	7.038279	6.960410
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
5201	Compound 5919	6.245255	6.194300	6.165928	6.390953	6.402920	6.258669	6.262649	6.370278	6.256610	...	6.269349	6.284218	6.166841	6.263467	6.387067	6.334254	6.145515	6.364408	6.246458	6.258794
5202	Compound 5930	6.265979	6.269323	6.296698	6.347539	6.403661	6.362765	6.222485	6.277444	6.303384	...	6.320016	6.411511	6.379362	6.353872	6.367754	6.400277	6.367340	6.348874	6.343532	6.375171
5203	Compound 5964	6.242246	6.124306	6.177051	6.082200	6.317291	6.257584	6.223415	6.282368	6.305714	...	6.201274	6.219690	6.144898	6.269480	6.191840	6.193191	6.295073	6.268855	6.180628	6.218116
5204	Compound 5978	6.157839	6.032436	6.130037	6.045991	6.050399	6.133175	6.040674	6.042695	6.117320	...	6.161811	6.124916	6.129828	6.095715	6.096871	6.076834	6.058739	6.199614	5.997863	5.999938
5205	Compound 5980	5.683136	5.884875	5.765027	5.616814	5.886334	5.653571	5.901391	5.905859	5.715543	...	5.938979	5.725710	5.471979	5.811799	5.761502	5.704077	5.918673	5.755731	5.892346	5.779499

5205 rows × 29 columns

```
[ ] #so null values present is about 0
    input_data.isna().sum().sum()
```

```
0
```

Descriptive Statistics of the Compounds Data



	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	...	
count	5205.000000	5205.000000	5205.000000	5205.000000	5205.000000	5205.000000	5205.000000	5205.000000	5205.000000	5205.000000	...	5205.00
mean	7.088989	7.092271	7.078525	7.075488	7.075191	7.087294	7.089151	7.084638	7.077268	7.090120	...	7.08
std	0.578480	0.568823	0.584924	0.587508	0.572077	0.584182	0.578371	0.570221	0.569626	0.575367	...	0.5
min	5.530975	5.667683	5.443891	5.510296	5.533868	5.543956	5.565442	5.632375	5.595055	5.606844	...	5.67
25%	6.758468	6.761539	6.755805	6.755341	6.759622	6.764535	6.762915	6.758435	6.758332	6.758608	...	6.75
50%	6.991434	6.991715	6.989708	6.990831	6.976311	6.991497	6.990294	6.985199	6.988040	6.991822	...	6.98
75%	7.371389	7.365837	7.357220	7.352454	7.347822	7.370661	7.370892	7.363706	7.349550	7.365358	...	7.35
max	10.313452	10.517358	10.175624	10.251454	10.252820	10.341428	10.081599	10.442653	10.342311	10.305869	...	10.42

Some inference on these Statistics

The mean values for each sample range from 7.07 to 7.09, indicating a moderate level of expression on average

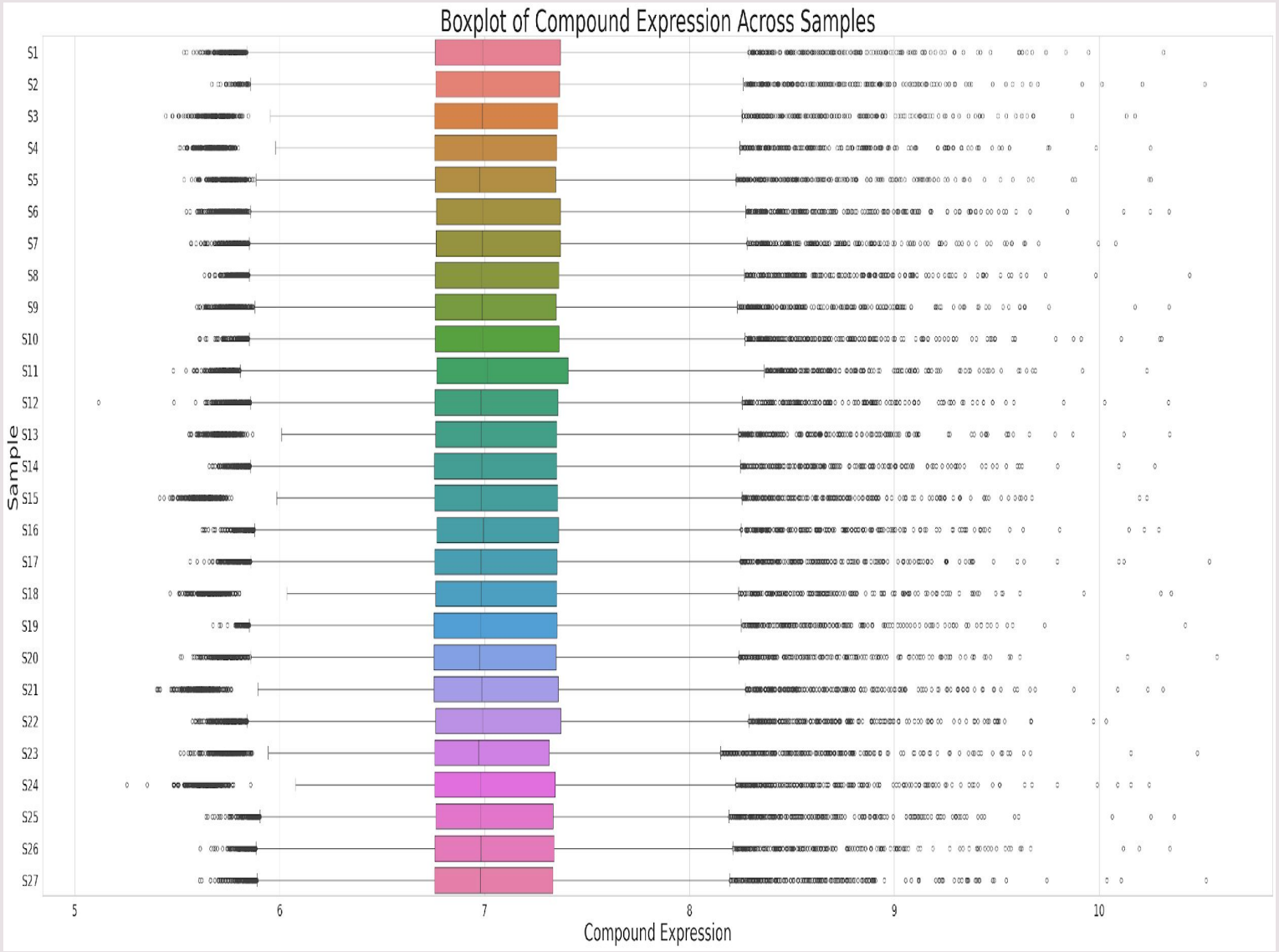
The median values (50th percentile) are close to the means, suggesting similar distributions.

The comparison of mean and median can provide insights into the distribution of the data. Since mean and median are close, the distribution are more similar.

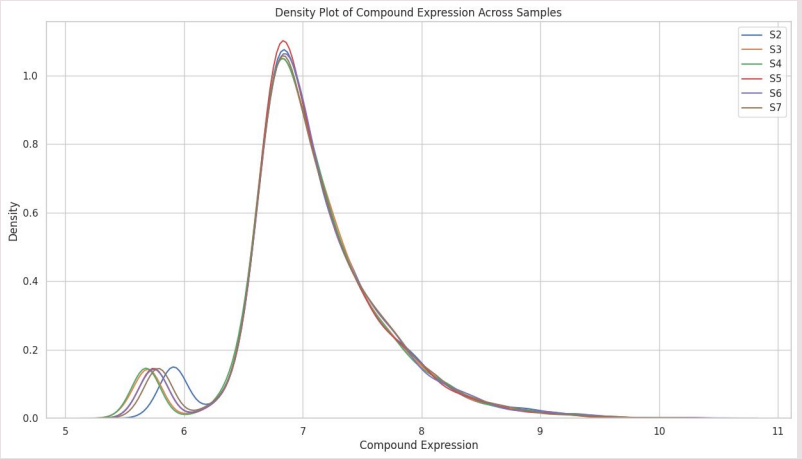
There may be potential extreme outliers, as suggested by the maximum values being notably higher than the 75th percentiles. This could be due to heterogenity present in data.

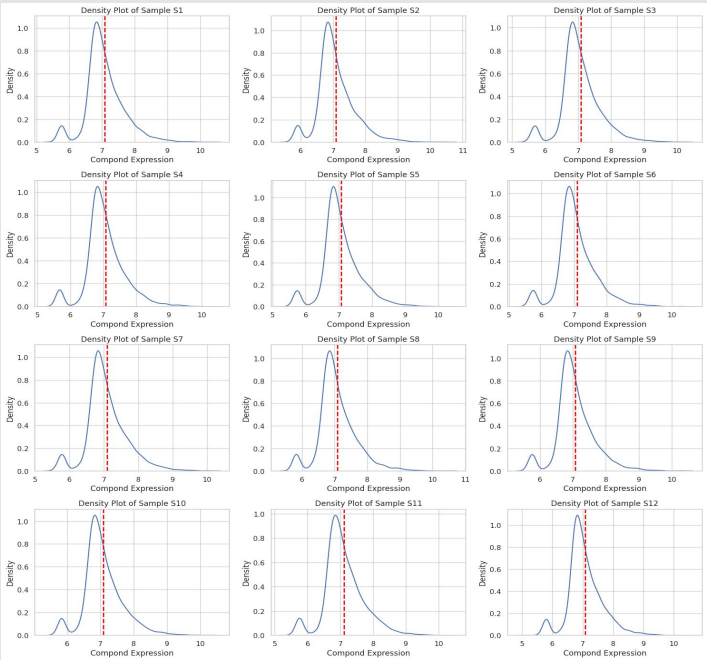
Sample Wise Distribution of Compound expression (Box Plot)

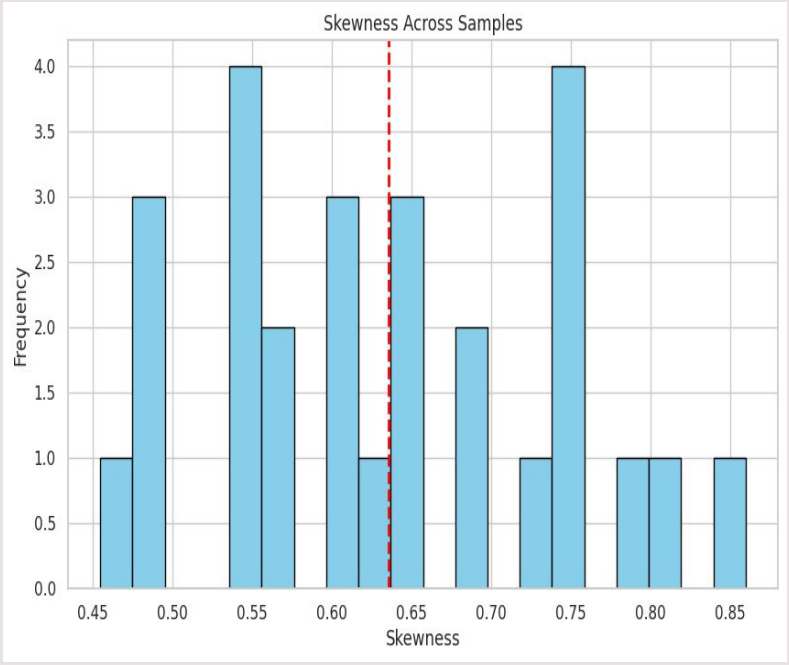
- Large outliers are present and some are at extreme scale range.
- These outliers aren't removed since they may highlight compounds that play a critical role in study and provide insights into their potential significance.
- These extreme outliers could be the result of experimental artifacts, technical errors, or these extreme expression levels may represent genuine biological variability .



Sample Wise Compound Expression Density

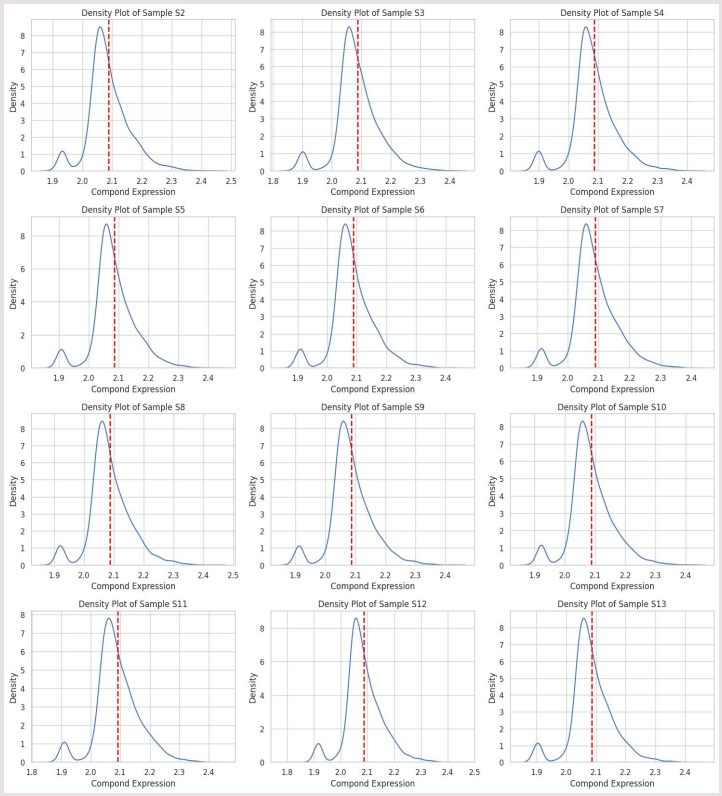




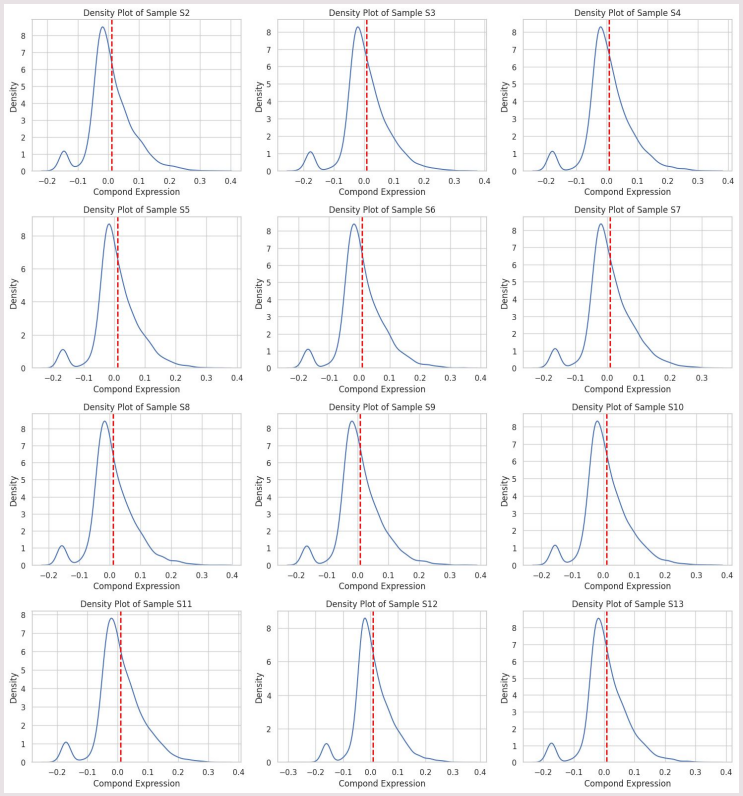


- Here median of most sample's density plot vary over small range so median normalization used to eliminate that.
- Additionally, though their central tendency are close but their scale over they vary aren't aligned and scaling is required (MAD Scaling).
- These plots look highly skewed since most exceed normal threshold i.e., Skewness >0.5 and normalization is highly required.

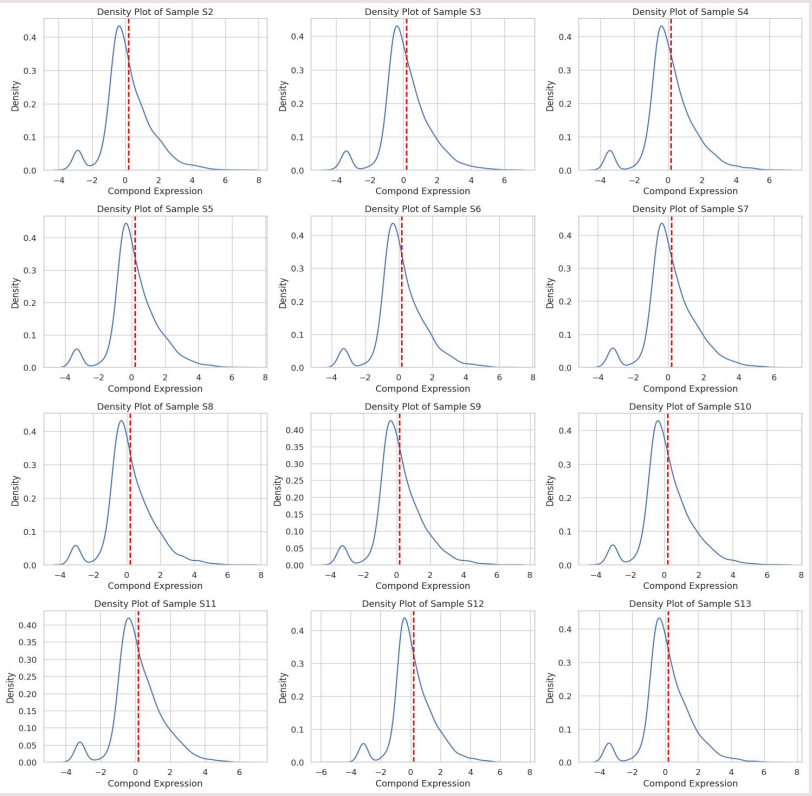
Sample Wise Compound Expression Preprocessing



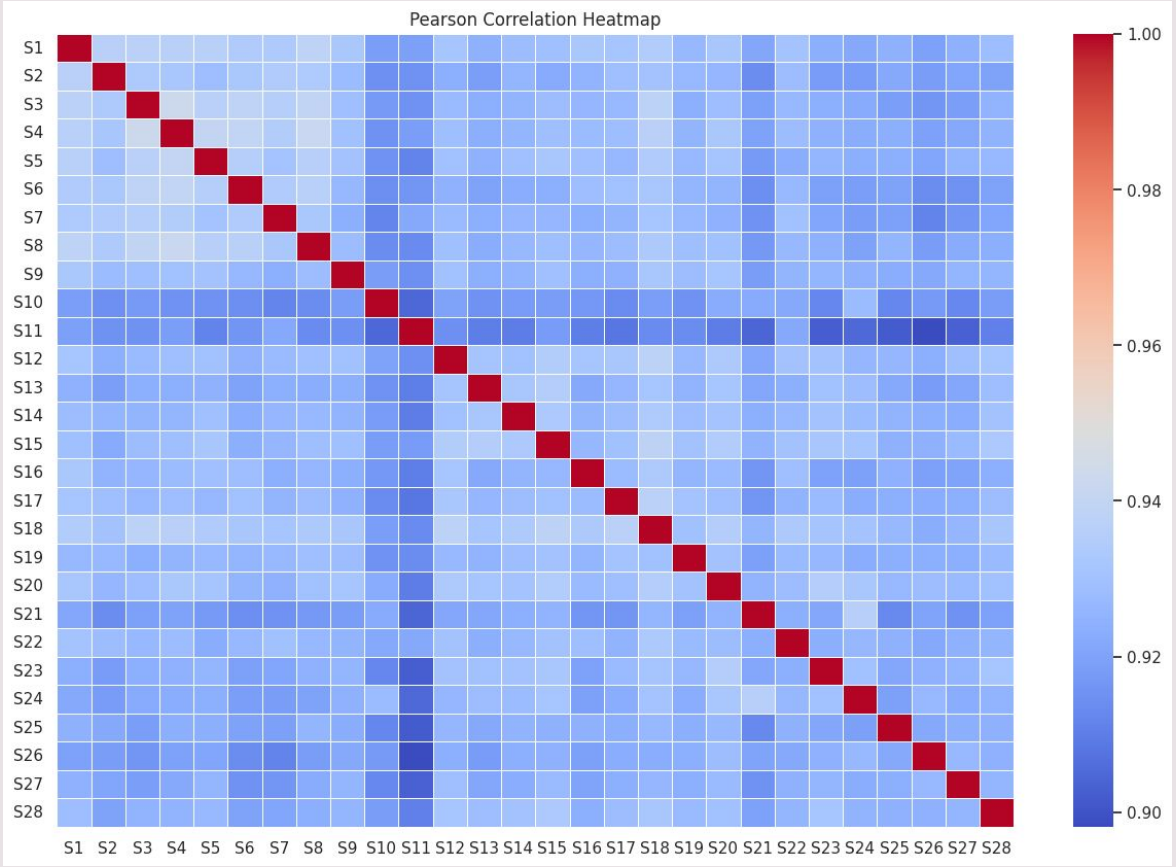
After Log transformation



After Median Normalization



After MAD Scaling

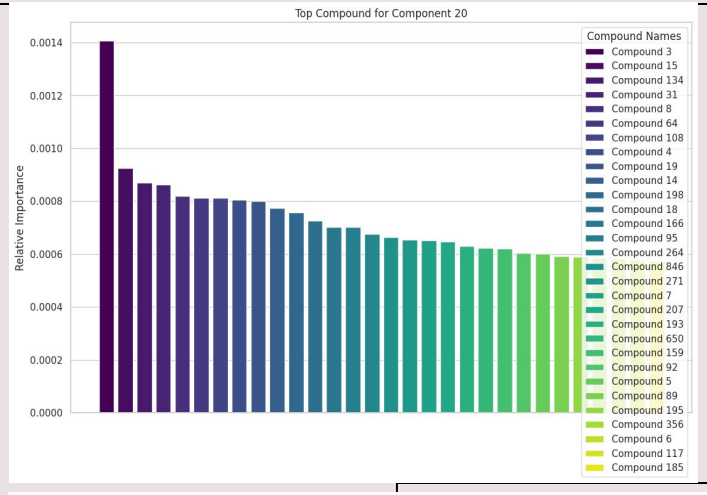


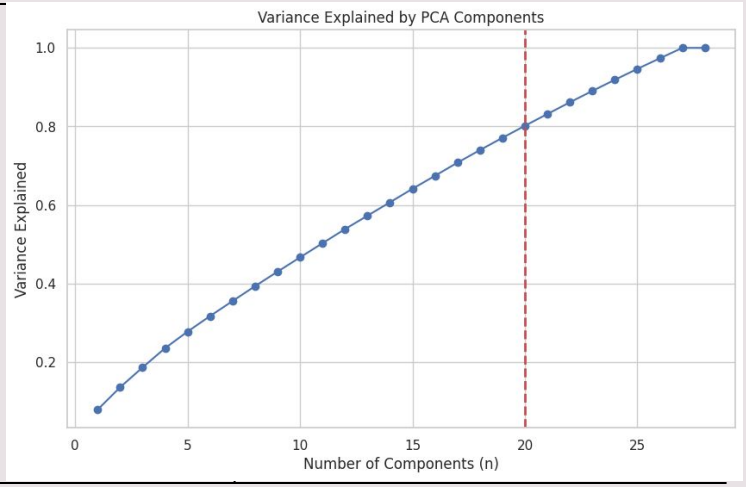
Relationships & Correlations Sample Wise

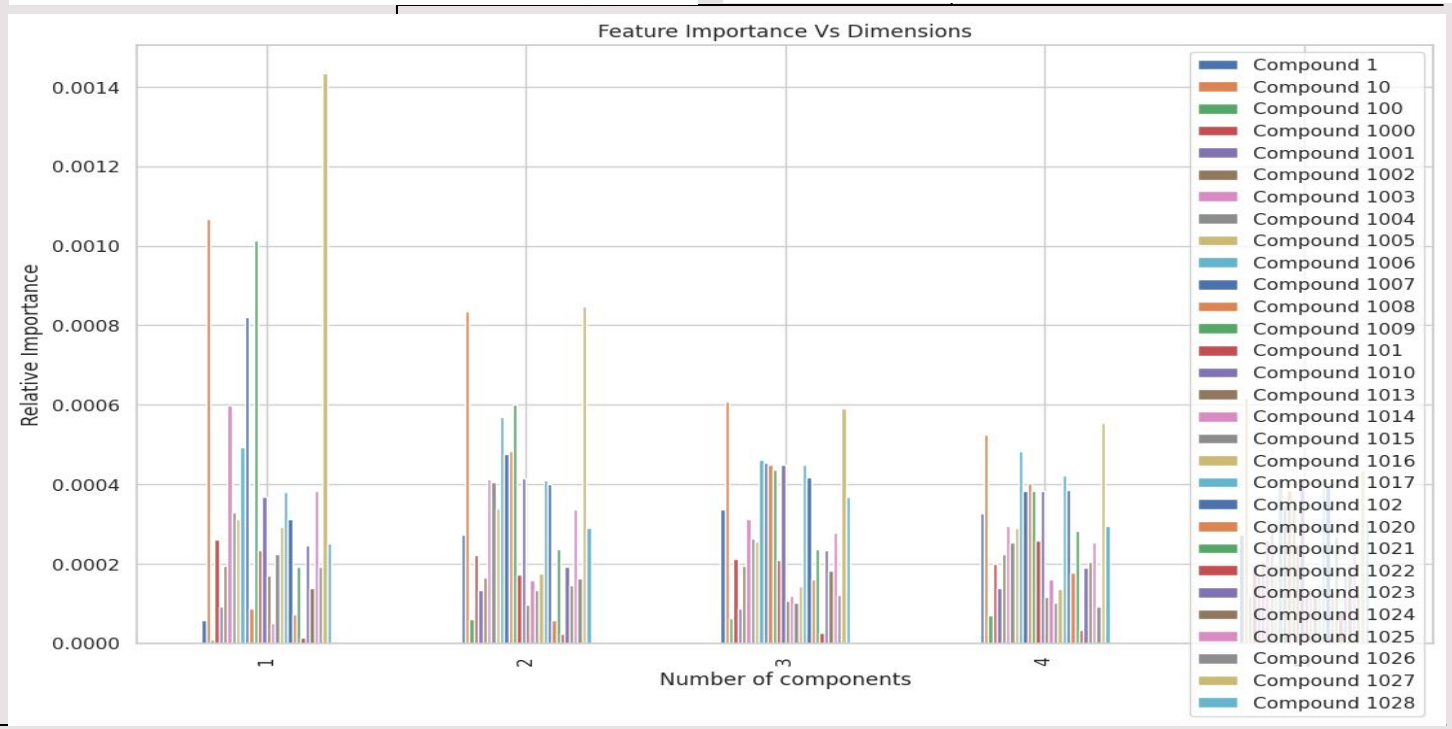
- Sample-wise they are highly correlated and are linear in relationship mostly with positive slope.
- Pearson was used since it is best measure for correlation of such linear relationship.
- From correlation heatmap one inference can be that S10 & S11 samples relate or correlate with every other in more magnitude that those other samples i.e., conditions use to produce or extract these samples with other samples could be similar in nature.

PCA Analysis on Compound Features

- PCA done to reduce the total number of compound expression for each sample to smaller component i.e., transforming the data into a lower-dimensional space while preserving most of the variability
- By 20th component it captures about 80% of total variance so reduction till this component is desirable.
- Each component linear combination of compound types and its weightage in it decides in priority or importance.
- Since our dataset has more features than sample as instance so PCA has to be applied to avoid dimensionality curse or overfitting while applying clustering or classifying algorithms like SVM, KNN etc.
- This helps in identifying key compounds that contribute most to the observed variability like Compound 3 for 20th component.





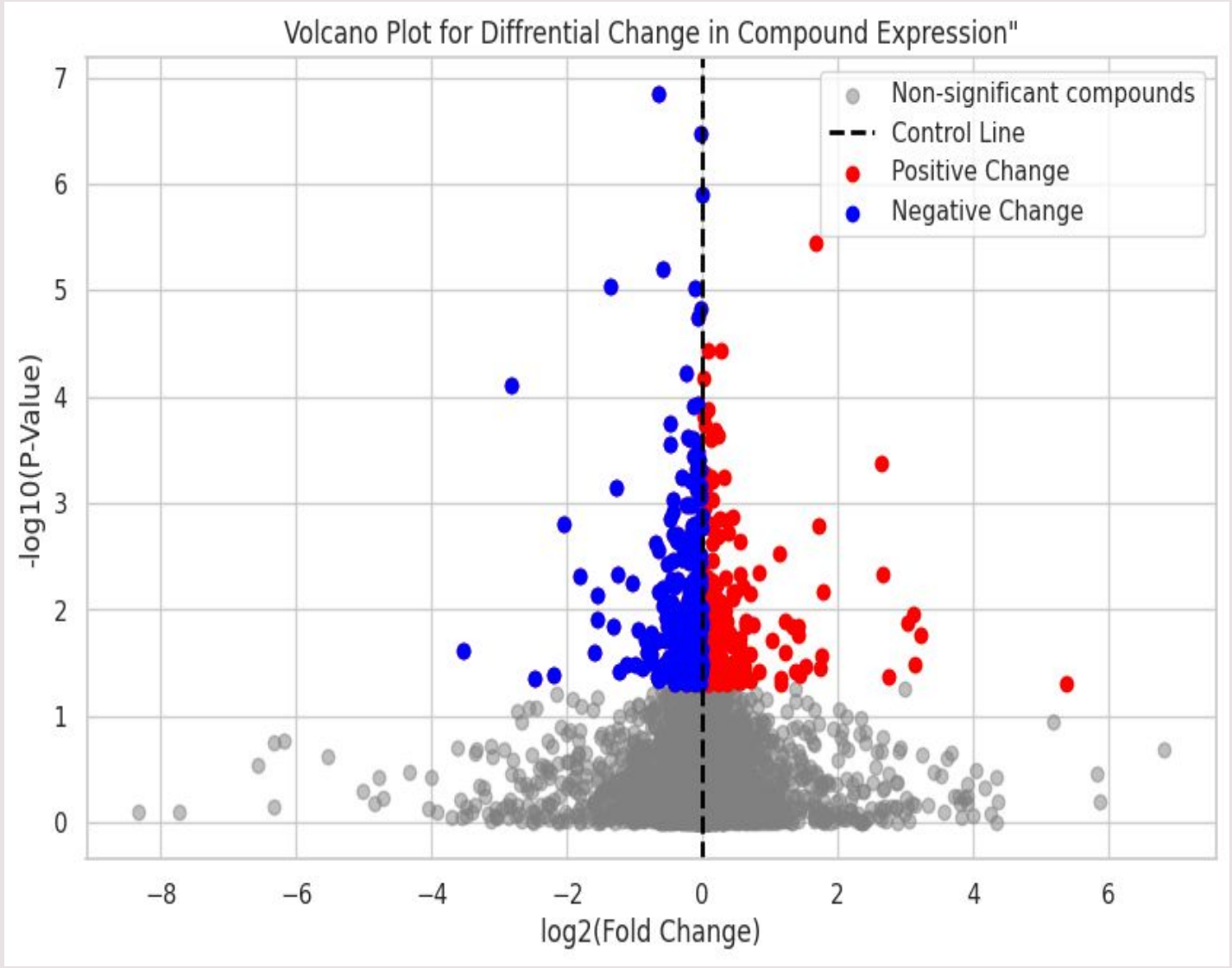


Differential Expression Change

- Compounds with significant differences in expression and substantial fold changes are often of particular interest as their difference are unlikely to be due to random chance alone but increasing confidence in the biological relevance of these genes.

- Compounds with smaller changes but high significance might also be biologically relevant since high significance implies that the observed differences are unlikely to occur by chance, even if the fold change is modest as their small but consistent changes could have functional implications.

- Plot shows quite less divergence and its closeness shows that compounds hardly changes with conditions or noise.



About Dataset

Worked thirdly on proteomics dataset

The samples were also divided in two groups as control and disease.

And contained main proteins expression data where for each sample protein expression of 3859 proteins ,16 samples taken.

Analysis done over sample-wise to see how these sample's different protein expression varied over samples range and how these samples are correlated with each other.

```
[ ] input_data.head(5)
```

	proteinId	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16
0	NaN	Disease	Control	Disease	Control	Disease	Control	Disease	Control	Disease	Control	Disease	Control	Disease	Control	Disease	Control
1	A0A0C4DH25	-0.4110154	-1.0761586	-0.8069974	-0.4348028	-0.725628	-0.9595164	-0.5867214	-1.025746	-0.8742558	-1.0567953	-0.986267	-1.0355096	-0.7404106	-0.9132753	-0.7551428	-1.389646
2	A0A0U1RRL7	-1.269271	-1.0159043	-1.2609829	-0.6592318	-0.53701	-0.8544569	-1.6614398	-0.6677959	-1.2161872	-1.0020238	-1.3610606	-1.0806478	-1.4885281	-0.8858549	-0.67001	-0.7622122
3	A0AVT1	0.04468818	0.41739166	0.30851027	0.43486232	0.20936986	0.52801891	0.07074149	0.10109937	-0.16425173	0.28496142	0.33446151	0.01832625	0.22050918	0.16477152	0.16178463	0.06954364
4	A0FGR8	0.35368785	0.22889006	0.33369563	0.16601788	0.39640282	0.15883829	0.32515478	0.18422908	0.46164821	0.11588105	0.19884589	0.21280111	0.32441772	0.06702933	0.39709315	0.1867064

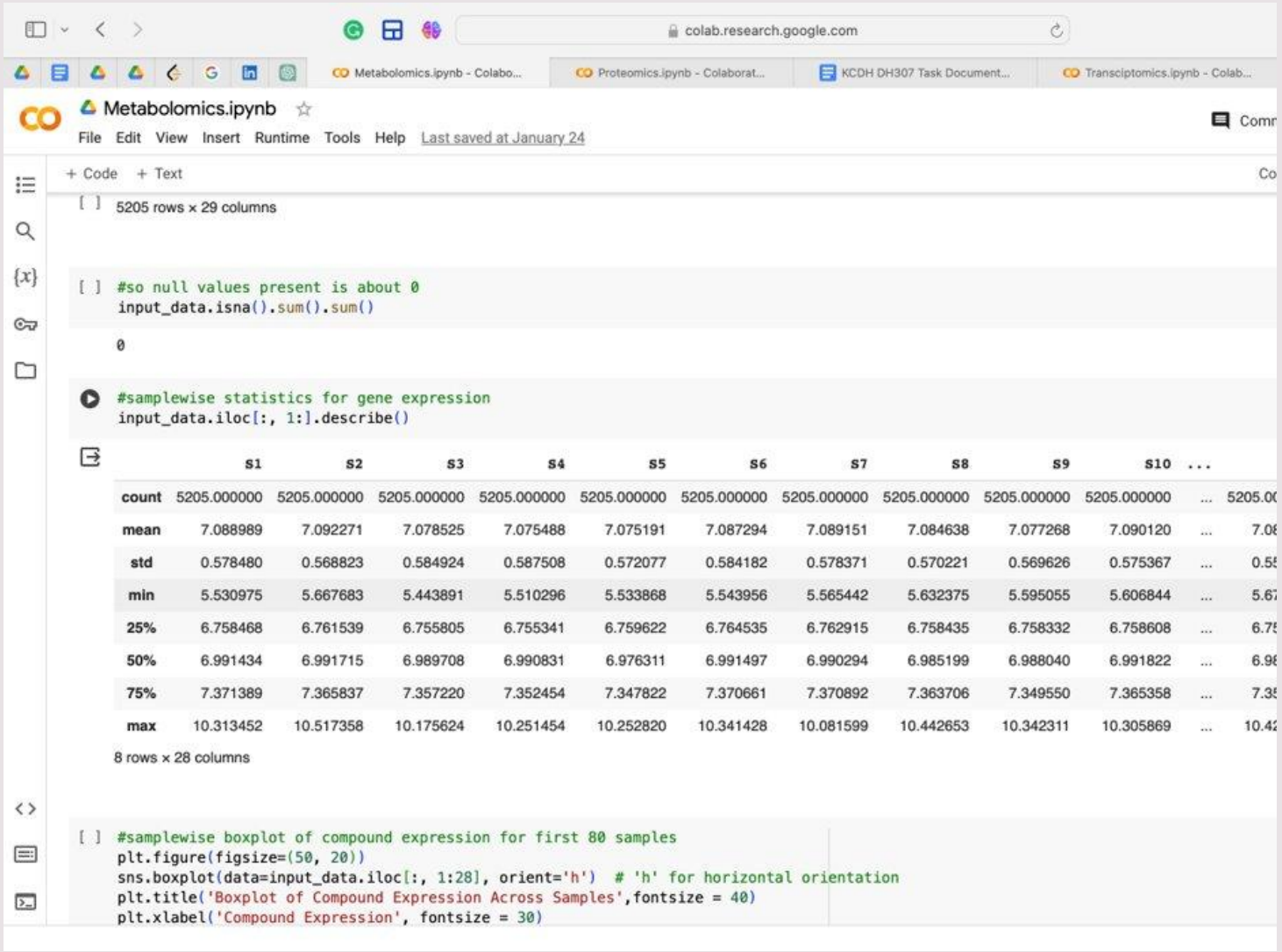
```
orignal_data = input_data
input_data = input_data.drop(0,axis = 0)
input_data
```

	proteinId	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	
1	A0A0C4DH25	-0.4110154	-1.0761586	-0.8069974	-0.4348028	-0.725628	-0.9595164	-0.5867214	-1.025746	-0.8742558	-1.0567953	-0.986267	-1.0355096	-0.7404106	-0.9132753	-0.7551428	-1.
2	A0A0U1RRL7	-1.269271	-1.0159043	-1.2609829	-0.6592318	-0.53701	-0.8544569	-1.6614398	-0.6677959	-1.2161872	-1.0020238	-1.3610606	-1.0806478	-1.4885281	-0.8858549	-0.67001	-0.7
3	A0AVT1	0.04468818	0.41739166	0.30851027	0.43486232	0.20936986	0.52801891	0.07074149	0.10109937	-0.16425173	0.28496142	0.33446151	0.01832625	0.22050918	0.16477152	0.16178463	0.06
4	A0FGR8	0.35368785	0.22889006	0.33369563	0.16601788	0.39640282	0.15883829	0.32515478	0.18422908	0.46164821	0.11588105	0.19884589	0.21280111	0.32441772	0.06702933	0.39709315	0.1
5	A0MZ66	0.19587454	0.55283494	0.22727856	0.84326458	0.2417761	0.87735259	-0.03300351	0.42275339	0.28360473	0.5880928	0.28346466	0.4662273	0.16698046	0.37043154	0.34400755	0.3
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3854	Q9Y6R7	-0.27421403	-2.50654387	-0.28831173	-1.86785784	-0.51404207	-0.63076717	-0.52071741	-0.88159577	-0.4604776	-1.8787921	-0.69229509	-0.96647469	-0.52982723	-1.09438839	-0.66272072	
3855	Q9Y6T7	-0.10440979	-0.99907169	-0.23987759	-0.30465856	0.1975486	-0.53641048	-0.01791051	0.1075084	-0.46913436	0.10921925	-1.07137009	-0.20175278	-0.27852009	0.07012497	-0.03417819	0.09
3856	Q9Y6U3	-1.1368852	-0.9074525	-0.3195858	-0.6394562	-0.6418592	-1.0493089	-0.5979857	-1.6171881	-0.8723723	-1.3672038	-0.6733125	-1.0543174	-0.1534924	-0.9309843	-0.6668135	-1.0
3857	Q9Y6V0	0.5475997	0.5938925	0.4278971	0.3736061	0.7467866	0.2886325	0.6230994	0.5665843	0.8126275	0.7534138	0.3072922	0.6990237	0.5125211	0.6033028	0.6068375	0.7
3858	Q9Y6X4	-0.47604107	-0.56113371	-0.35556967	-0.08255277	-0.34512998	-0.40411716	-0.49748451	-0.52179967	-0.4568023	-0.5507757	-1.05596188	-0.67623287	-0.28970971	-0.25971948	-0.33135098	-0.32

3858 rows × 17 columns

```
[ ] #so null values present is about 0
    input_data.isna().sum().sum()
```

Descriptive Statistics of the Proteins Data



	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	...	
count	5205.000000	5205.000000	5205.000000	5205.000000	5205.000000	5205.000000	5205.000000	5205.000000	5205.000000	5205.000000	...	5205.00
mean	7.088989	7.092271	7.078525	7.075488	7.075191	7.087294	7.089151	7.084638	7.077268	7.090120	...	7.08
std	0.578480	0.568823	0.584924	0.587508	0.572077	0.584182	0.578371	0.570221	0.569626	0.575367	...	0.5
min	5.530975	5.667683	5.443891	5.510296	5.533868	5.543956	5.565442	5.632375	5.595055	5.606844	...	5.67
25%	6.758468	6.761539	6.755805	6.755341	6.759622	6.764535	6.762915	6.758435	6.758332	6.758608	...	6.75
50%	6.991434	6.991715	6.989708	6.990831	6.976311	6.991497	6.990294	6.985199	6.988040	6.991822	...	6.98
75%	7.371389	7.365837	7.357220	7.352454	7.347822	7.370661	7.370892	7.363706	7.349550	7.365358	...	7.35
max	10.313452	10.517358	10.175624	10.251454	10.252820	10.341428	10.081599	10.442653	10.342311	10.305869	...	10.42

Some inference on these Statistics

The mean values for each sample range from −0.055 to 0.068, indicating a mean of all samples centerrd to 0.

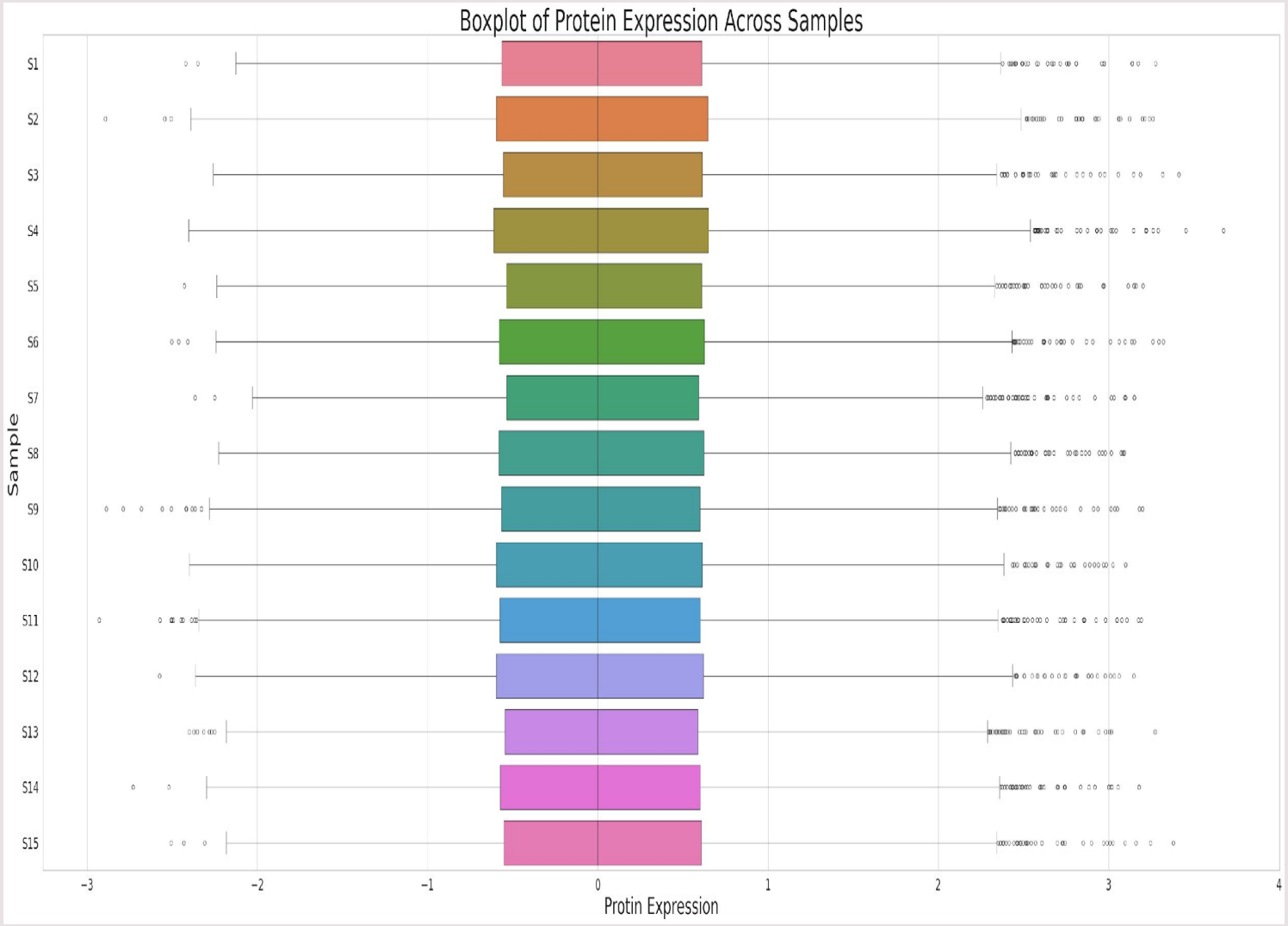
The median values (50th percentile) are close to the means, suggesting symmetric distributions.

The comparison of mean and median can provide insights into the skewness of the data. Since mean and median are almost same, the distribution is likely symmetric and very less skewness presence.

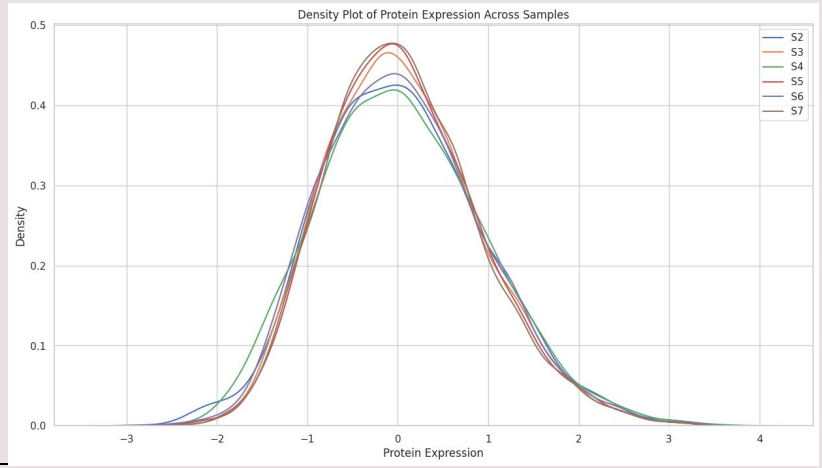
There may be potential extreme outliers, as suggested by the maximum values being notably higher than the 75th percentiles. This could be due to heterogenity present in data.

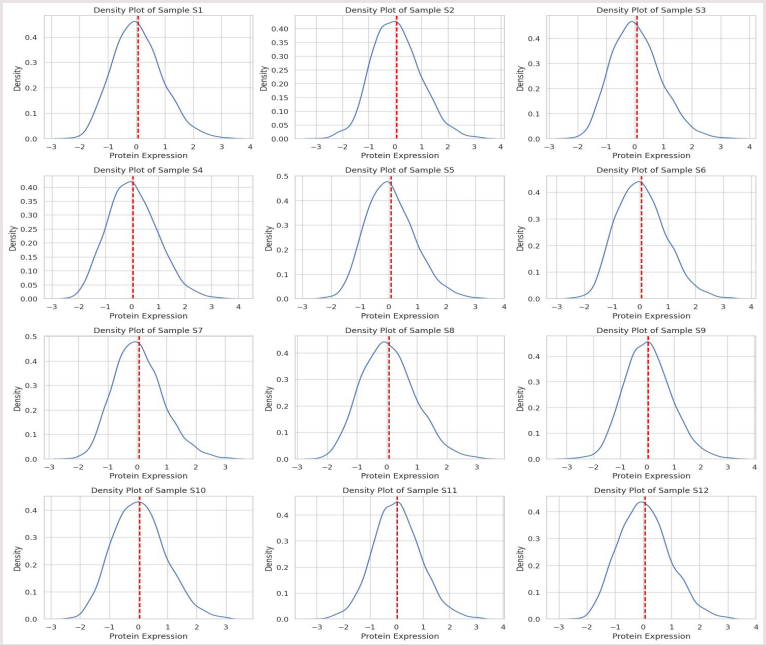
Sample Wise Distribution of Protein expression (Box Plot)

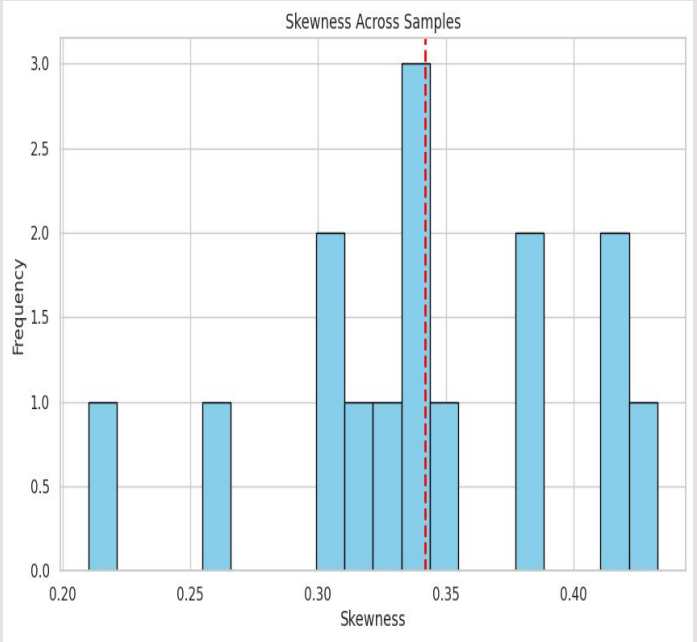
- Few outliers are present and some are at extreme scale range.
- These outliers aren't removed since they may highlight compounds that play a critical role in study and provide insights into their potential significance.
- Less outlier are present since the distribution of gene expression is almost symmetric and normalized i.e. preprocessing has already been done with less outlier presence.



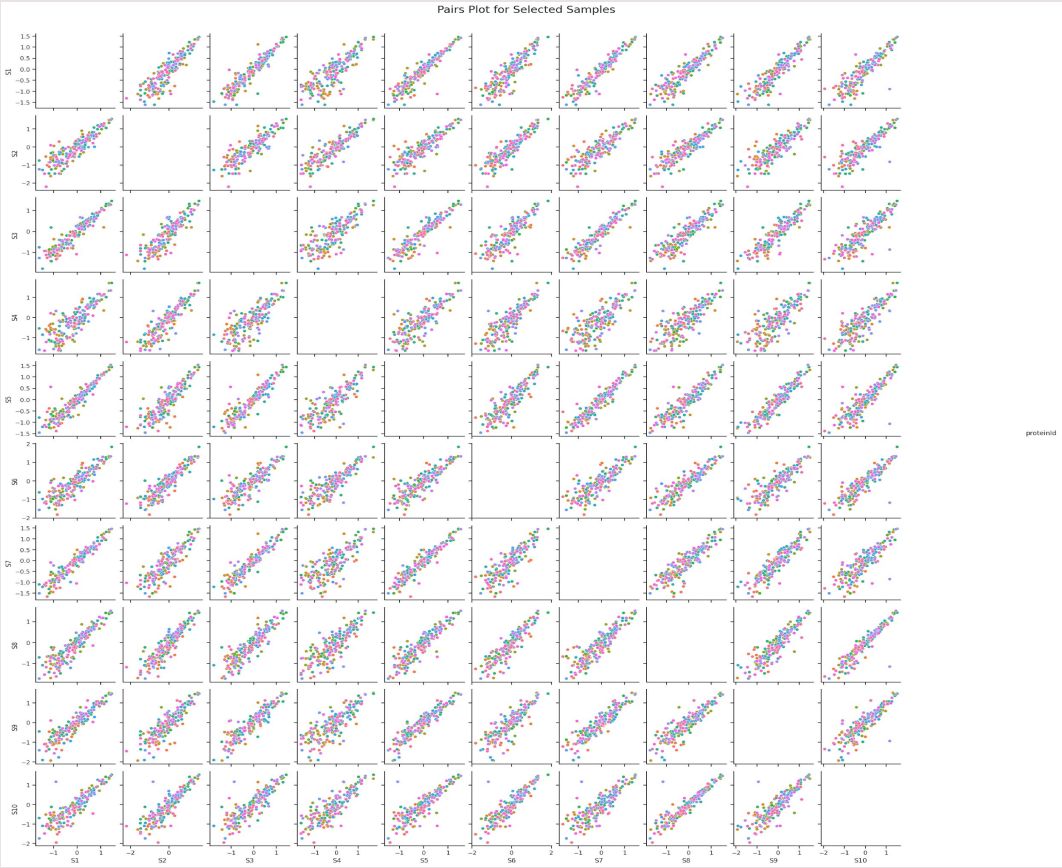
Sample Wise Protein Expression Density

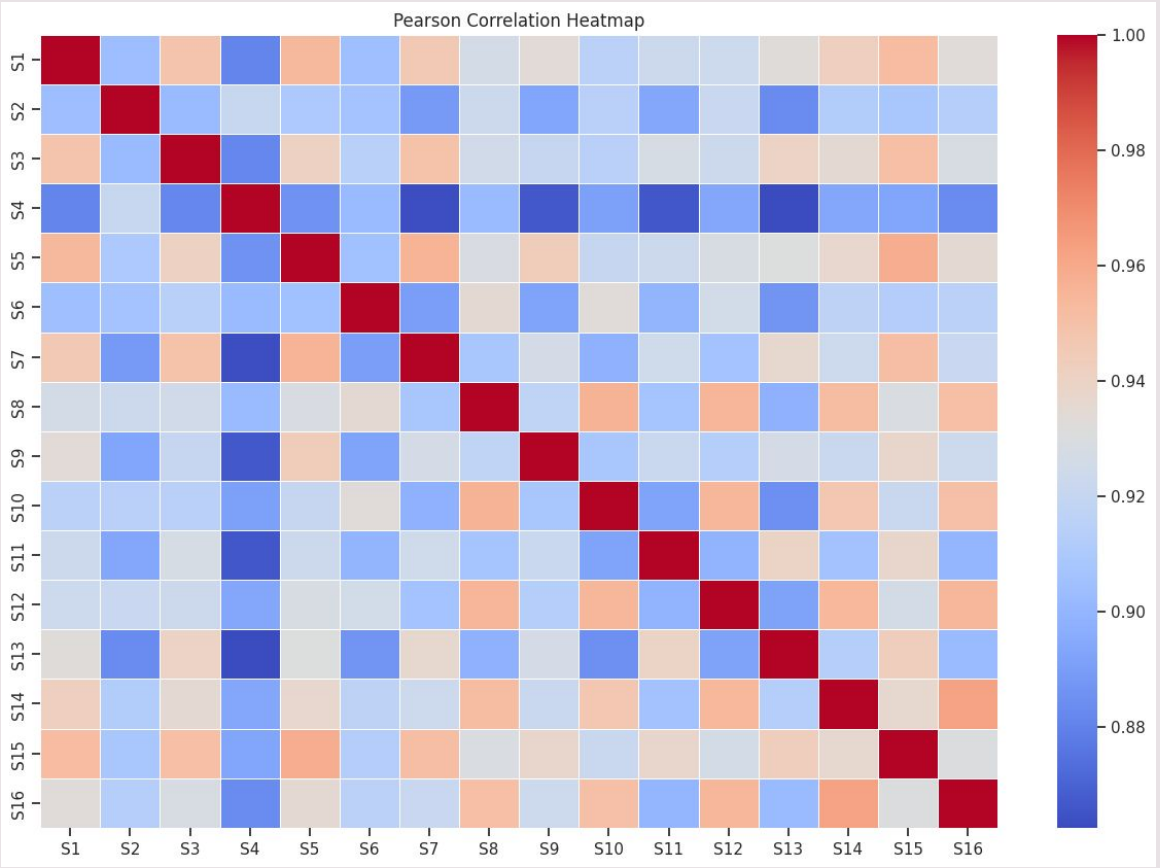






- Here median, mode and mean of most sample's density plot are similar vary over small range so no normalization required.
- These plots have very less skewness due to more symmetric and normalized distribution so no preprocessing required.



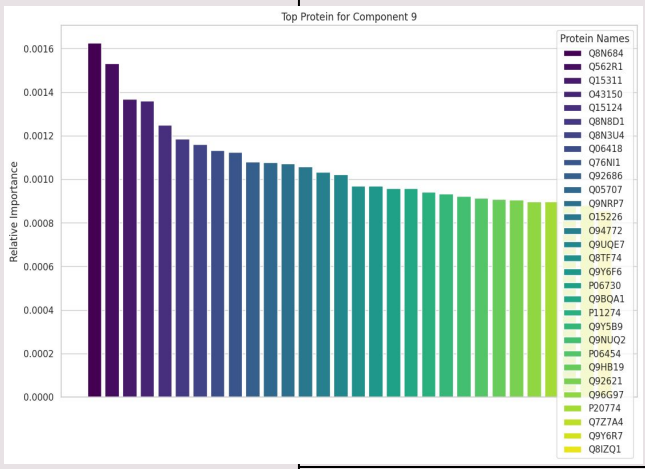


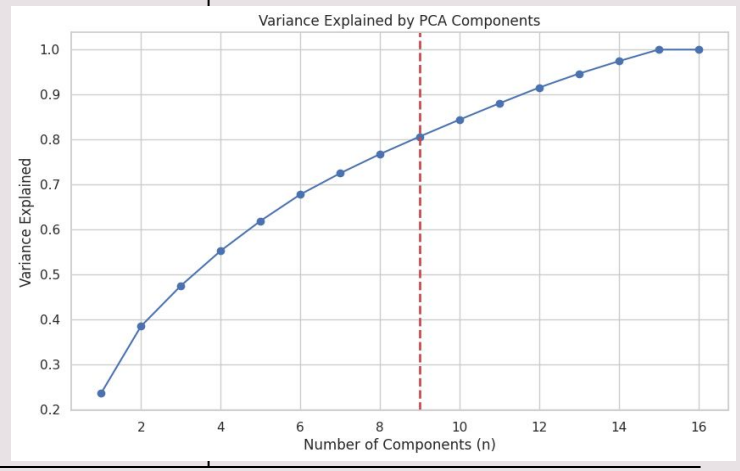
Relationships & Correlations Sample Wise

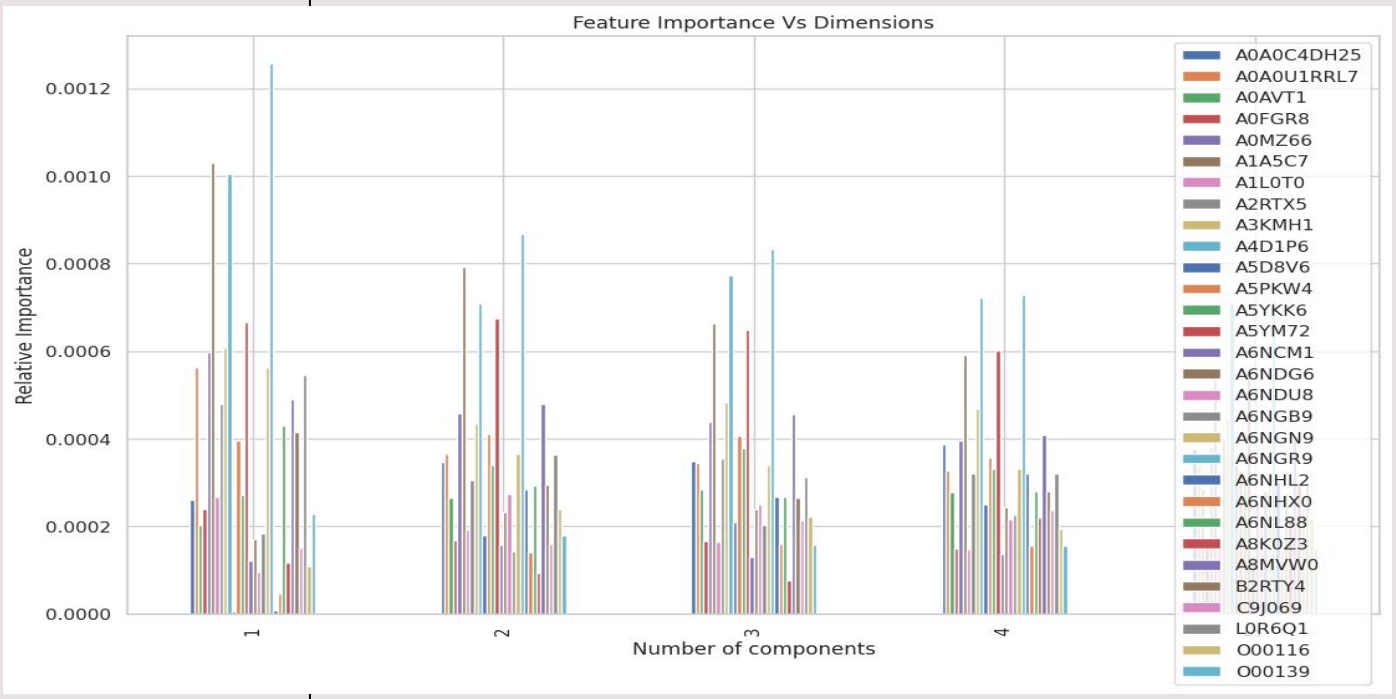
- Sample-wise they are highly correlated and are linear in relationship mostly with positive slope.
- Pearson was used since it is best measure for correlation of such linear relationship.
- From correlation heatmap one inference can be that S4 sample relate or correlate with every other in more magnitude that those other samples i.e., conditions use to produce or extract this sample with other samples could be similar in nature.

PCA Analysis on Protein Features

- PCA done to reduce the total number of compound expression for each sample to smaller component i.e., transforming the data into a lower-dimensional space while preserving most of the variability
- By 9th component it captures about 80% of total variance so reduction till this component is desirable.
- Each component linear combination of compound types and its weightage in it decides in priority or importance.
- Since our dataset has more features than sample as instance so PCA has to be applied to avoid dimensionality curse or overfitting while applying clustering or classifying algorithms like SVM, KNN etc.
- This helps in identifying key proteins that contribute most to the observed variability like Q8N684 protein Id for 9th component.

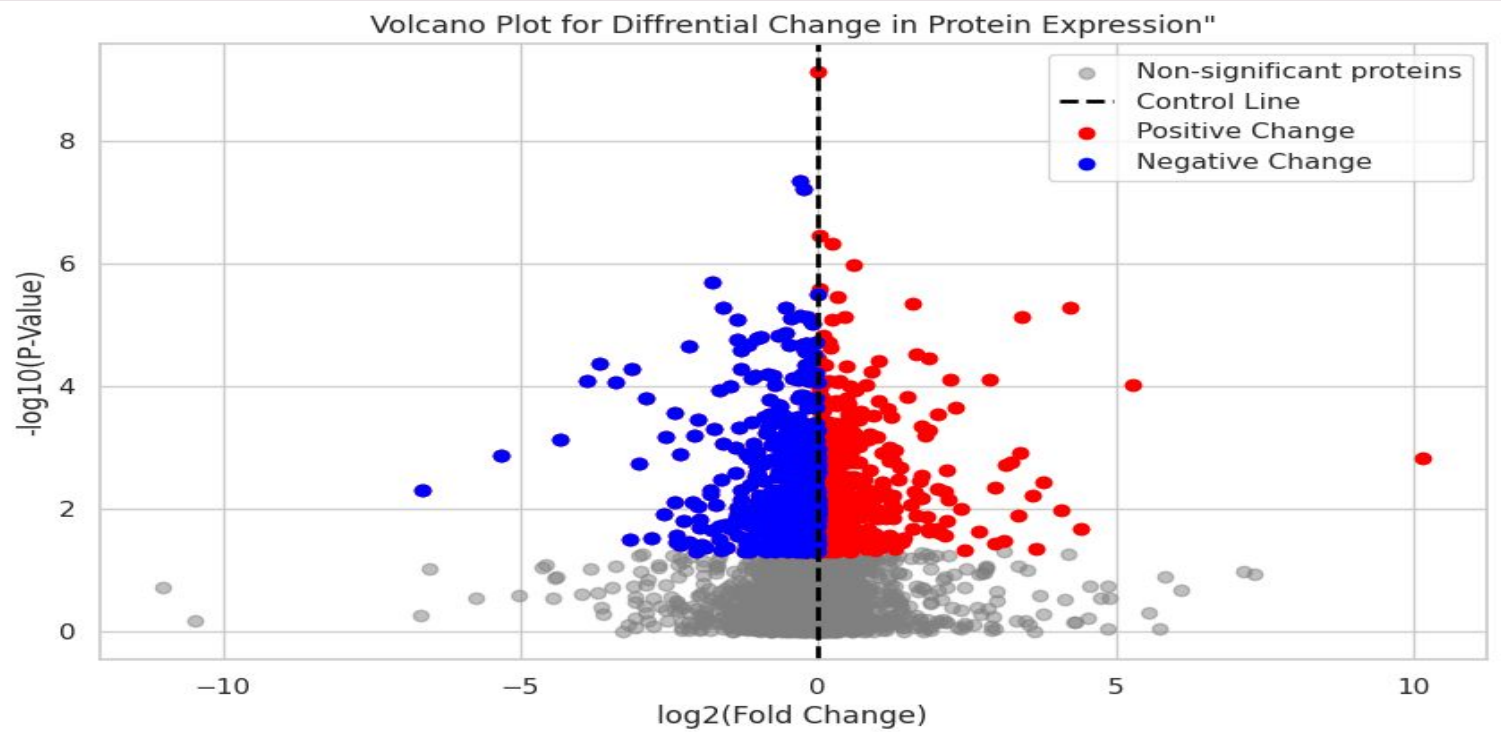


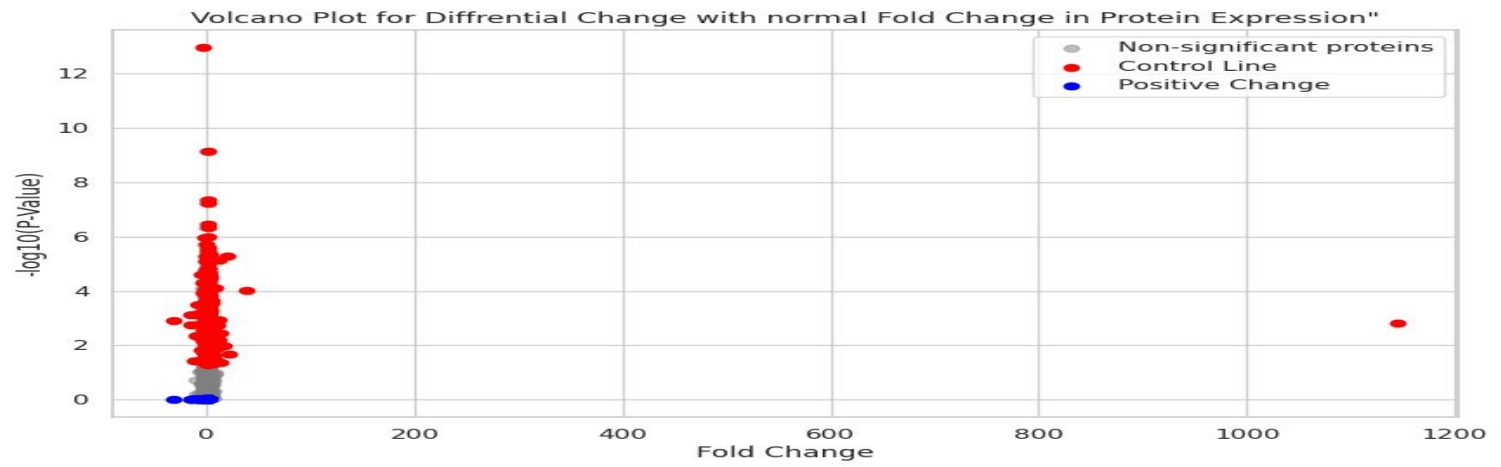




Differential Expression Change

- Plot shows quite less divergence and its closeness shows that proteins hardly changes with conditions or noise.
- However some proteins shows extreme fold change indicating external noises, outliers or protein abundance is very less in samples .
- Though this plot may show that all genes fold change don’t differ by much but with log scale removed their difference pretty much is significant and high.





Implementation

- Transcriptomics Dataset
  : https://colab.research.google.com/drive/1lzli2-nZtdPSyyguK-_LVtbZy6smHcCo?usp=sharing
- Metabolomics
  Dataset: https://colab.research.google.com/drive/1AUdBR5jVsk7tr7TXwfW1SH529LUgpvmR?usp=sharing
- Proteomics Dataset
  : https://colab.research.google.com/drive/1TMrg5tbwlxyEmPLWCKnLpGAEnmTf5EMh?usp=sharing


THANK YOU