Multivariate Data Analysis of CHO Cell Batch Cultures

Anirudh Rao

Batch Culture Data

The dataset contains time course data for two CHO cell lines, CHO-TZ and CHO-S7, which are grown in two different media A and B. These cell lines have been engineered to produce an IgG antibody. The dataset contains data about the viable cell density, IgG titer, and the concentrations of amino acids, glucose, and lactate in the cell culture media, which were measured on Day 0, 4, 7, 10, and 15.

Visualizing the data allows us to obtain some insights about the performance of the cell lines and the media they are being cultured in.

Figure 1 shows the viable cell density, measured in cells/mL, over the duration of 15 days. All four conditions start off with a similar concentration of around 10^6 cells/mL. They then follow a similar exponential growth pattern, with the viable cell density rapidly increasing to 10^7 cells/mL by Day 4. They reach their peak density by Day 7. All conditions, except CHO-TZ in media A, show a drastic decrease in cell growth between Days 7 and 10, and then showing relatively stationary growth until Day 15. CHO-TZ in media A maintains its cell density between its peak at Day 7 and Day 10, and then shows a rapid decrease by Day 15. CHO-TZ in media B experiences the sharpest decline in growth. At the end of 15 days, both the CHO-S7 cell lines have higher cell densities than the CHO-TZ clones, with the CHO-S7 clone in media A growing better than the one in media B. Thus, from the growth data alone, it appears that the CHO-S7 cell line has better growth characteristics than CHO-TZ, and media A promotes cell growth more than media B.

Figure 2 shows the IgG titer (in mg/mL) over time. This steadily increases in all four conditions. There is a rapid increase in titer between Days 4 and 7, with production slowly reaching a stationary phase between Days 10 and 15. CHO-S7 in media A has the highest yield, followed by CHO-TZ in media B. CHO-S7 in media B has the lowest yield. Although CHO-S7 in media A initially lags behind the others at the Day 4 mark, it quickly takes the lead by Day 7, and maintains its high titer. Thus, both cell line type and media have an influence on IgG titer.

Figure 3 shows the change in concentrations (in μM) of various amino acids in the culture media. Asparagine and glutamine are rapidly depleted by all cell lines within the first 7 days. Most of the other amino acids show an initial fast consumption, followed by a slow secretion during the later time points. The consumption of aspartate, glutamate, phenylalanine, threonine, tryptophan, and cystine, is higher when the cell lines are in media A than when they are in media B. In the case of histidine, isoleucine, leucine, lysine, and valine, the CHO-S7 cells consume more than the CHO-TZ cells, irrespective of media. In the case of arginine, proline, and tyrosine, the consumption is higher when media B is used, irrespective of the cell type. In the case of serine, the consumption is higher in CHO-TZ cells, irrespective of the media used. For alanine, methionine, and cysteine, all the cell lines show similar behaviour, with the exception of CHO-TZ in media A, which consistently has a higher secretion. Glycine is the only amino acid that is continuously secreted, with the cells in media B having higher secretion, irrespective of the cell type. Thus, amino acid metabolism shows high variation from cell line to cell line and from media to media.

Figure 4 shows the consumption of glucose (in g/L) over time. The cells in media A consistently show a more rapid consumption of glucose than those in media B. In both media, CHO-S7 cells consume more glucose than CHO-TZ cells in the same media. This may explain the growth characteristics observed.

Figure 5 shows the lactate concentration (in mM) in the culture media over the 15 day period. All cells initially produce lactate, with the CHO-TZ producing more, indicating poorer utilization of glucose. Between Days 4 and 7, there is a decline in lactate concentration. This decline is not substantial in CHO-TZ cells in media A. By Day 15, there is no lactate in media B. However, both cell types in media A start producing lactate once again.

In general, the observed data seems to indicate that CHO-S7 cells in media A are best suited for IgG production. Media A seems to provide a better environment for both cell lines to be metabolically active.

Specific Rates

Nutrient consumption and antibody production seem to be most rapid between Days 4 and 7, which constitutes the exponential phase. Following this, there is a stationary phase until Day 15.

For the exponential phase, we compute the specific growth rates, specific productivities, and specific rates of the substrates, for all cell lines.

$$\text{Specific growth rate} = \mu = \frac{\ln\left(\frac{x}{x_0}\right)}{\Delta t}\ \text{day}^{-1}$$

$$\text{Specific productivity} = \mu \frac{\Delta[\text{Product}]}{\left(\frac{x+x_0}{2}\right)}\ \text{pg cell}^{-1}\ \text{day}^{-1}$$

$$\text{Specific rate} = \mu \frac{\Delta[\text{Metabolite}]}{\left(\frac{x+x_0}{2}\right)}\ \text{pmol cell}^{-1}\ \text{day}^{-1}$$

Here, $\left(\frac{x+x_0}{2}\right)$ is the average biomass concentration (in cells mL^{-1}) during the time period.

The results of the computation of specific rates is tabulated in Table 1. For glucose, the molecular weight was taken to be $180\ \text{g mol}^{-1}$.

For each of the specific rates computed, we highlight the best performing clone. We can clearly observe that cells in media A have better performance than those in media B. CHO-S7 cells in media A have the best performance. However, CHO-S7 cells in media B have the poorest performance during the exponential phase. By performance, we refer to the combination of growth rate, productivity, and nutrient consumption / secretion.

Principal Component Analysis

To better interpret the specific rates, we perform dimensionality reduction using principal component analysis (PCA). We first have to center each of the rates to have a mean of zero across cell types. PCA is then carried out using Python's `sklearn` library. This yields 4 principal components.

PC1 explains 62% of the variance in the data, followed by PC2 at 25%, PC3 at 13%, and PC4 explaining none of the variance. This is visualized in Figure 6.

We can then visualize the results of the PCA by plotting PC1 against PC2 in a score plot, shown in Figure 7. This reveals that the cells cultured in media B cluster together. The cells in media A are far apart from each other but lie on the same side of the plot, which is opposite to the side of the plot on which the cells in media B lie.

To better understand the results, we can plot the absolute values of PC1 and PC2 on separate 1D plots. This can be seen in Figure 8. We can observe that PC1 clusters cell types together, irrespective of the media they are grown in. The CHO-TZ cells form a separate cluster from the CHO-S7 cells when only PC1 is considered. Contrary to this, PC2 causes the points to cluster based on the media alone, irrespective of the cell type. The cells in media A form a separate cluster from the cells in media B when only PC2 is considered.

An explanation for the factors contributing to the separation of the 4 cell types can be obtained from the biplot in Figure 9, which visualizes the contributions (or loadings) of the various specific rates to PC1 and PC2. The specific growth rate only contributes to PC1. The line drawn orthogonal to its vector cleanly separates the cells based on cell type, pointing towards CHO-S7, which has better growth characteristics. The line drawn orthogonal to the specific productivity separates CHO-S7 in media A from the rest of the cells. This makes sense as it has the highest productivity.

We can also observe that substrates with similar consumption or secretion characteristics also have similar loadings. The clusters formed by these are (1) glutamine, asparagine, and serine, (2) arginine, proline, methionine, histidine, and lysine, (3) phenylalanine, aspartate, tyrosine, threonine, and glucose, (4) valine, leucine, isoleucine, tryptophan, and glutamate, and (5) glycine, alanine, and lactate. Cysteine does not appear to contribute to either PC1 or PC2. Lactate and glutamine almost entirely contribute to PC2. It also appears that the specific rate of cystine is correlated with the specific growth rate.

Conclusion

By analysing the specific rates of the batch culture data using PCA, we are able to better explain the performance characteristics of the cell types in different media.

Annexure

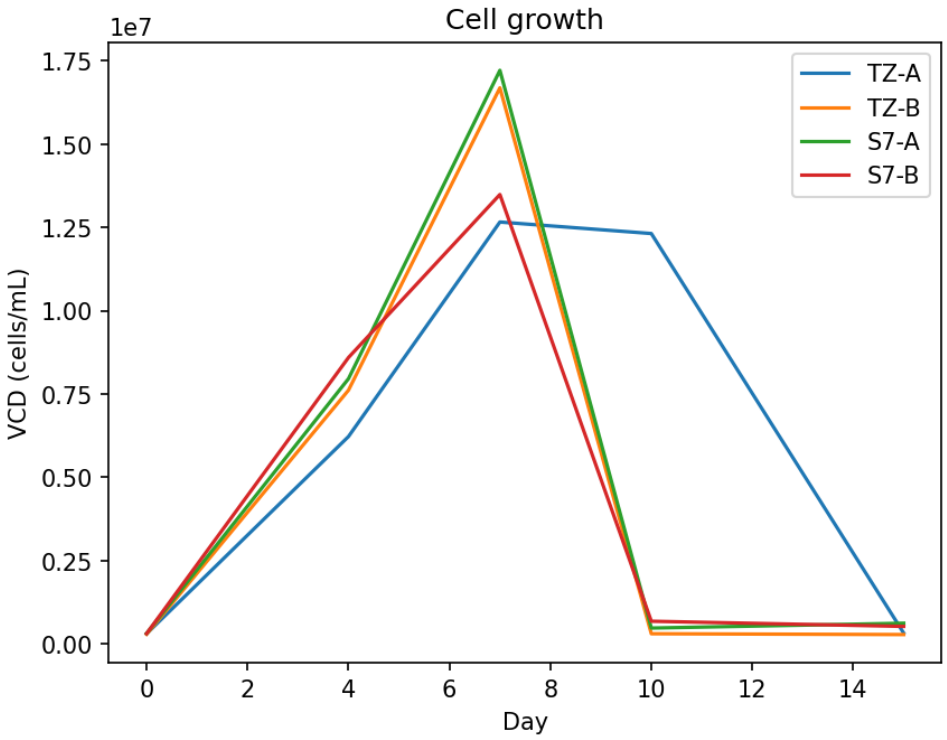

Figure 1: Cell growth over time

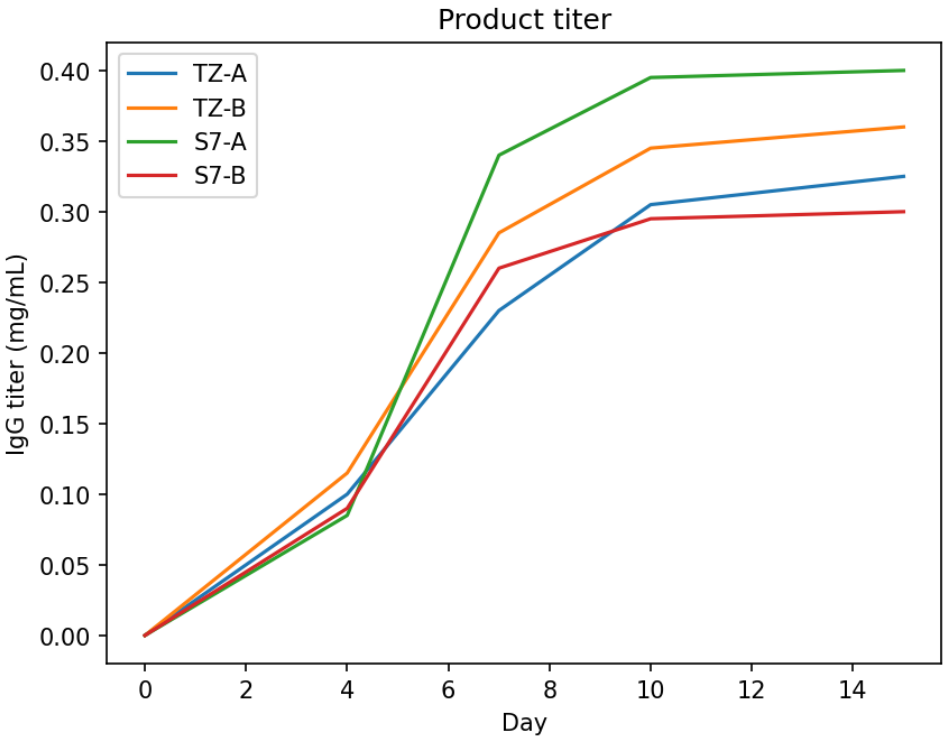

Figure 2: Antibody titer over time

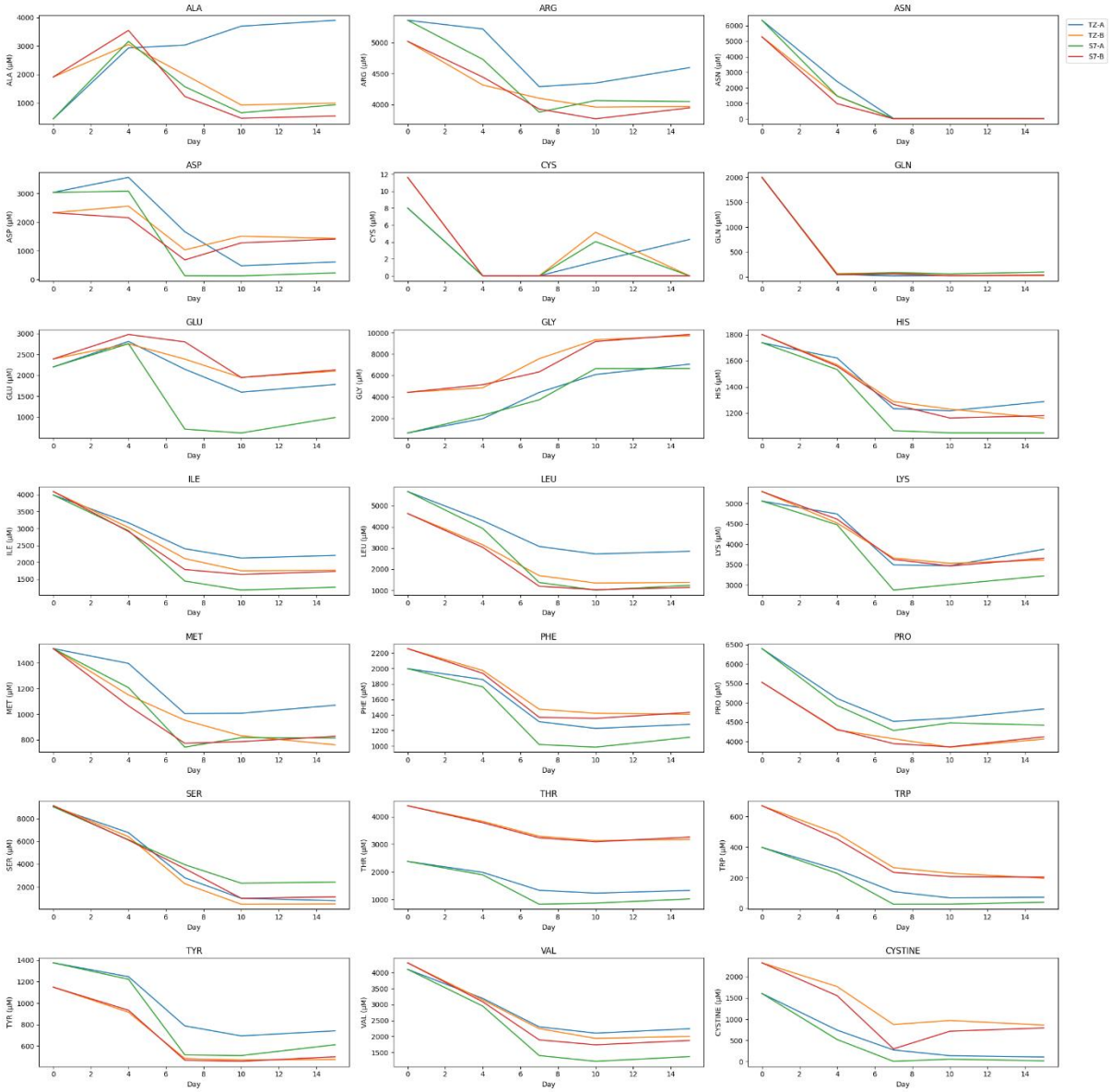

Figure 3: Nutrient concentrations over time

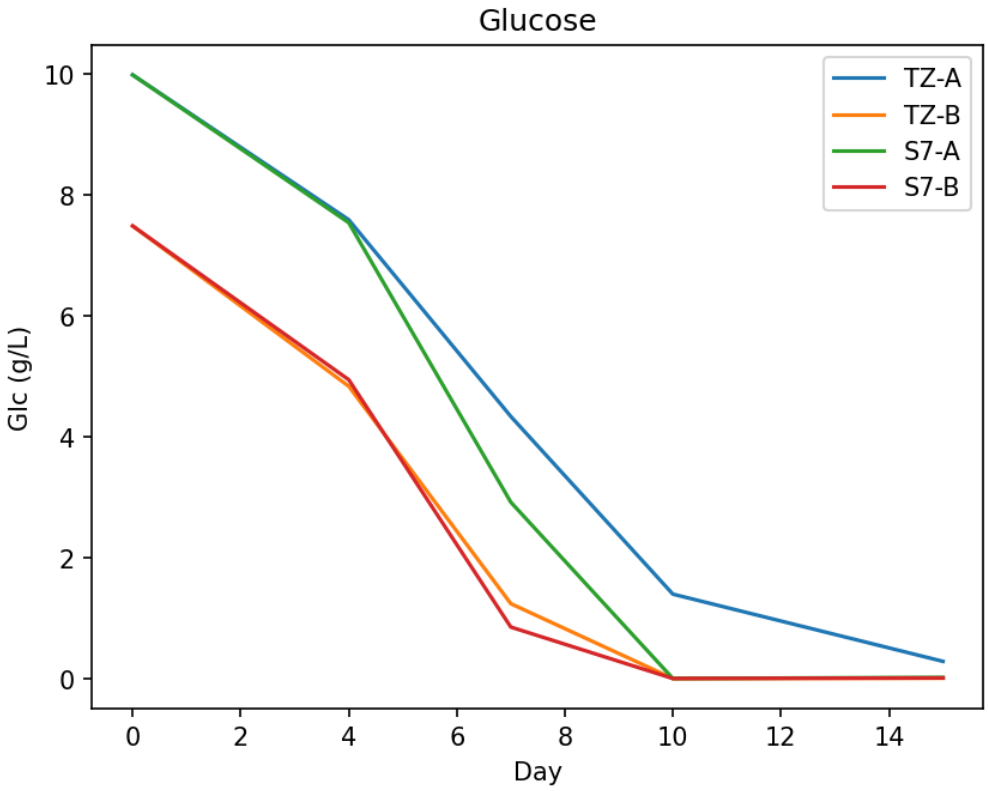

Figure 4: Glucose consumption over time

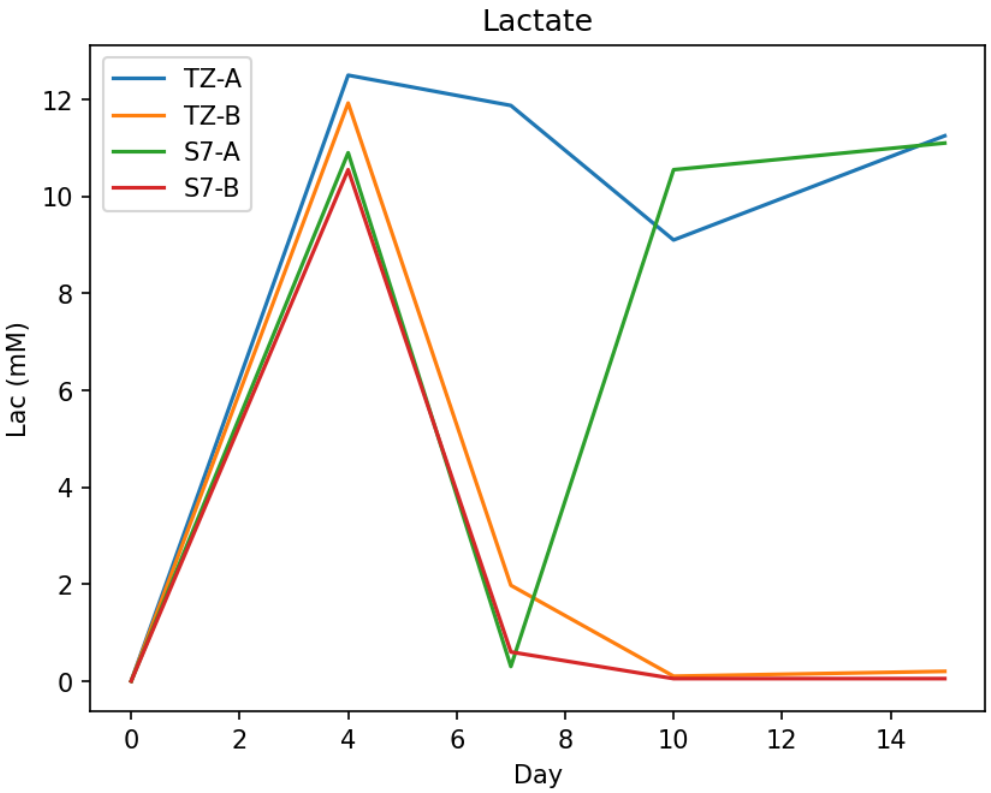

Figure 5: Lactate production over time

Table 1: Specific rates during the exponential phase

		TZ-A	TZ-B	S7-A	S7-B
	Specific growth rate (day^{-1})	0.236625	0.261463	0.257021	0.150132
	Specific productivity ($pg\ cell^{-1}\ day^{-1}$)	3.25774	3.658335	5.207806	2.311822
Specific rate ($pmol\ cell^{-1}\ day^{-1}$)	ALA	0.002541	-0.02273	-0.03255	-0.03159
	ARG	-0.0234	-0.0046	-0.01747	-0.00704
	ASN	-0.05959	-0.03134	-0.02968	-0.01322
	ASP	-0.04757	-0.0329	-0.0603	-0.02003
	CYS	0	0	0	0
	GLN	-0.00074	0.000269	0.000873	0.000258
	GLU	-0.01668	-0.00788	-0.04175	-0.00242
	GLY	0.061541	0.05863	0.029546	0.016317
	HIS	-0.00972	-0.00605	-0.00955	-0.00398
	ILE	-0.01923	-0.01979	-0.0304	-0.01529
	LEU	-0.03055	-0.03129	-0.05201	-0.02493
	LYS	-0.03126	-0.01844	-0.03269	-0.0134
	MET	-0.00981	-0.00426	-0.00948	-0.00397
	PHE	-0.01365	-0.01075	-0.0152	-0.00772
	PRO	-0.01484	-0.00484	-0.01321	-0.00499
	SER	-0.09947	-0.08903	-0.0447	-0.03412
	THR	-0.01622	-0.01152	-0.0217	-0.00742
	TRP	-0.00361	-0.00478	-0.0041	-0.00295
	TYR	-0.01149	-0.00928	-0.01439	-0.00633
	VAL	-0.02226	-0.01931	-0.03155	-0.01619
	CYSTINE	-0.01168	-0.01923	-0.0105	-0.01703
	Glucose	-0.45295	-0.42985	-0.52435	-0.30904
	Lactate	-0.01566	-0.21412	-0.21648	-0.13531

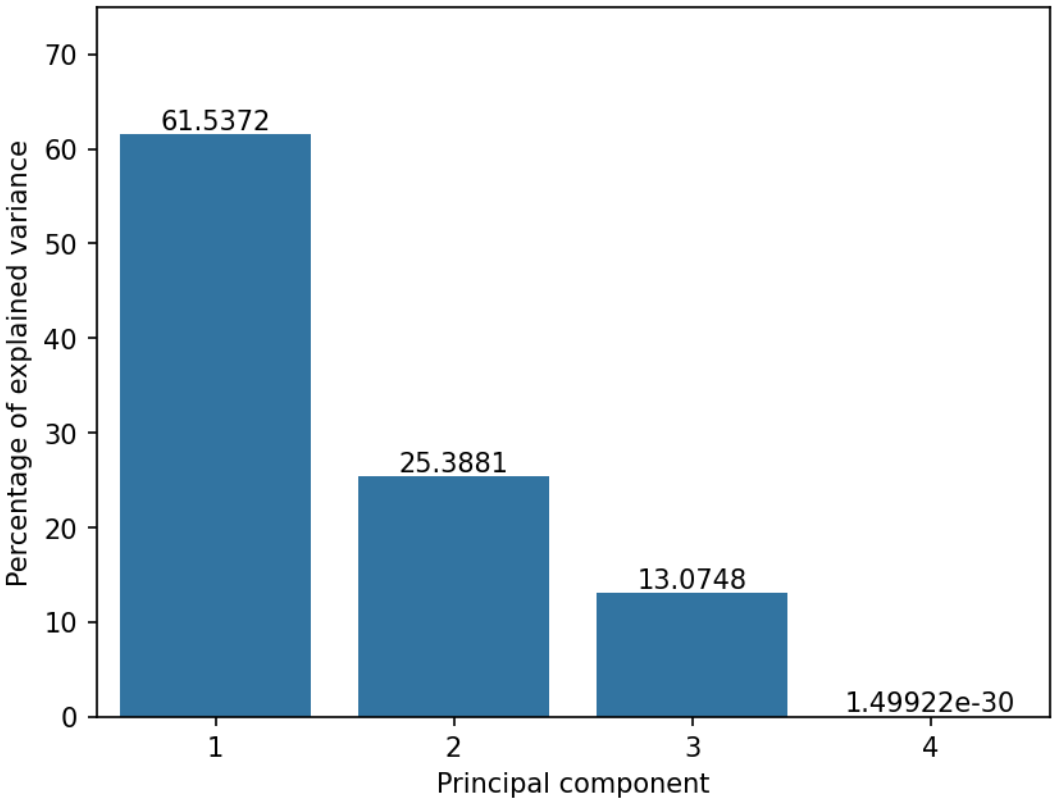

Figure 6: Explained variance of PCA

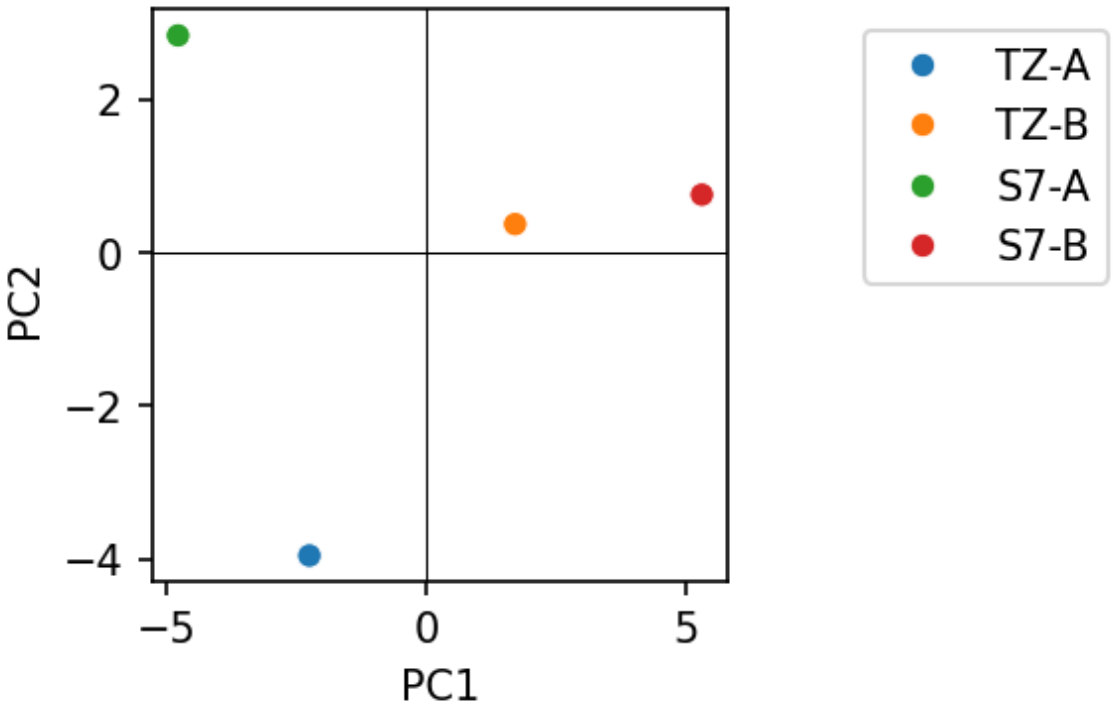

Figure 7: PCA score plot

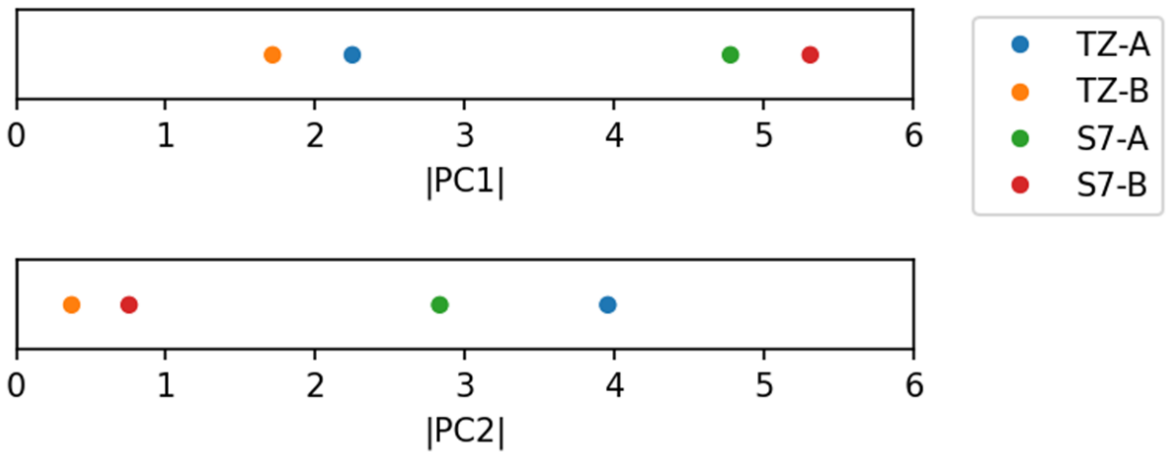

Figure 8: 1D visualization of PCA scores

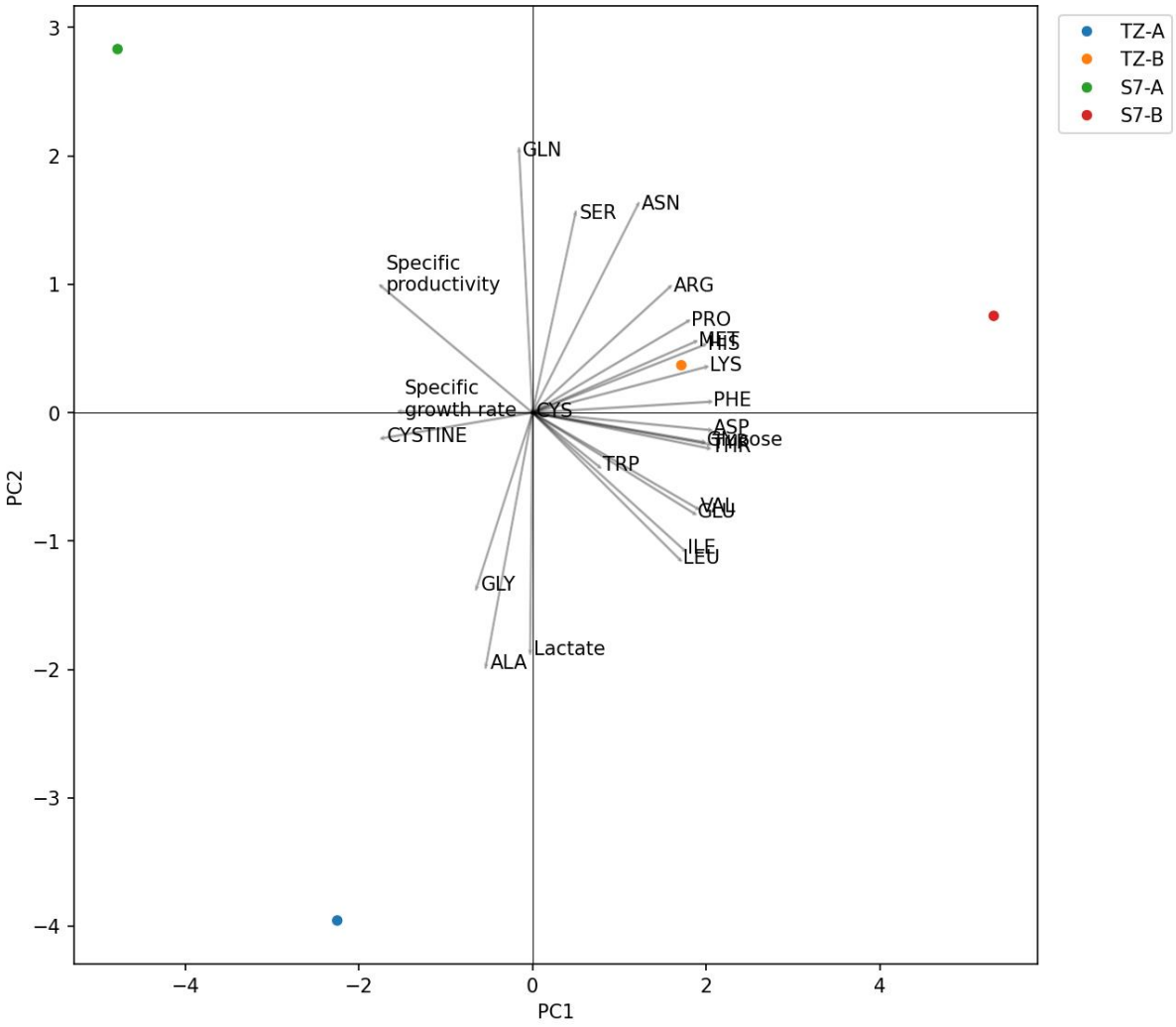

Figure 9: PCA biplot

All supplementary files associated with this assignment can be found here –
https://drive.google.com/drive/folders/1ua9zq52Qvat6dCvAwmPCV8kVqjiBq8g-?usp=sharing