

RNA-seq Report

Nathan Hughes (JIC)

May 17, 2019

Contents

1	Data setup	3
1.1	Helper funcs for pprinting	3
1.2	Load up counts and DE	3
1.3	Extracting gene names example	4
2	Inspect Samples	5
2.1	Creating a distance map of samples using normalised counts	5
2.1.1	Samples separated	5
2.1.2	Samples together	5
3	Simple Analysis	5
3.1	Largest/Lowest expression sum	5
3.2	Largest/Lowest TF expression	7
3.3	Largest/Lowest deviation in log2foldchanges	7
3.4	PCA on count data	7
3.5	PCA on expression data	10

1 Data setup

1.1 Helper funcs for pprinting

Lets go!

1.2 Load up counts and DE

Loaded data

1.3 Extracting gene names example

	incoming	name	description
0	AT1G73850	AT1G73850	Protein of unknown function (DUF1666)
1	AT2G35210	AGD10	RPA
2	AT5G23360	AT5G23360	GEM-like protein 7
3	AT3G04480	AT3G04480	Endoribonuclease
4	AT5G65165	SDH2-3	Succinate dehydrogenase
5	AT2G08635	AT2G08635	None
6	AT4G38810	AT4G38810	Calcium-binding EF-hand family protein
7	AT3G63530	BB	BB2
8	AT2G33000	AT2G33000	Ubiquitin-associated (UBA)/TS-N domain-containing protein-like protein
9	AT5G10336	AT5G10336	unknown protein
10	AT3G21620	AT3G21620	CSC1-like protein At3g21620
11	AT3G17225	AT3G17225	Plant invertase/pectin methylesterase inhibitor superfamily protein
12	AT2G30910	ARPC1A	Actin-related protein 2/3 complex subunit 1A
13	AT5G50520	AT5G50520	Major facilitator superfamily protein
14	AT5G19590	AT5G19590	At5g19590
15	AT3G19190	ATG2	Autophagy-related protein 2
16	AT5G01675	AT5G01675	None
17	AT1G27040	NPF4.5	Protein NRT1/ PTR FAMILY 4.5
18	AT3G49490	AT3G49490	Uncharacterized protein T9C5.90
19	AT1G02890	AT1G02890	AAA-type ATPase family protein

2 Inspect Samples

2.1 Creating a distance map of samples using normalised counts

2.1.1 Samples separated

<Figure size 720x720 with 4 Axes>

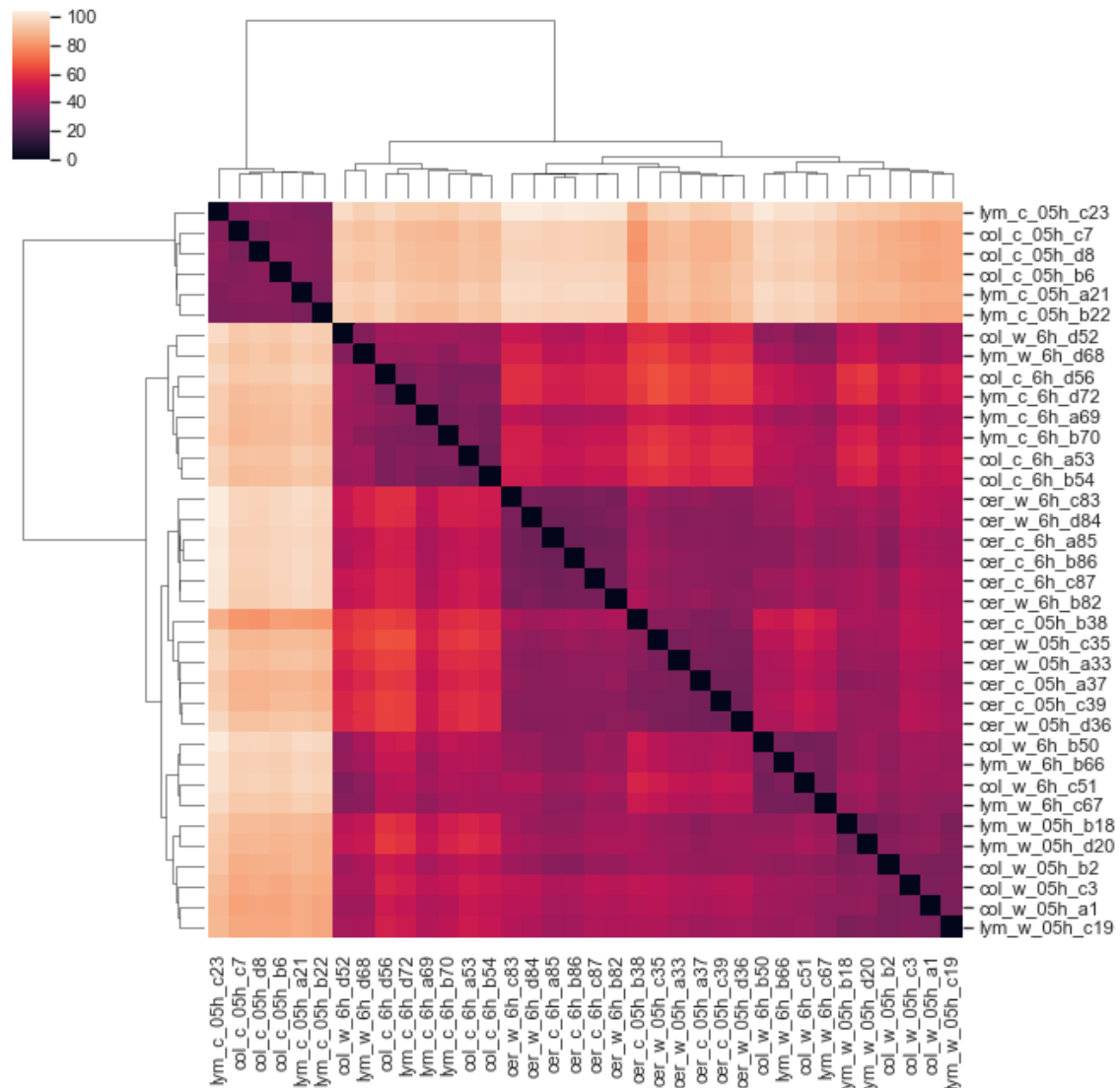


Figure 1: Distance map between samples

2.1.2 Samples together

<Figure size 720x720 with 5 Axes>

3 Simple Analysis

3.1 Largest/Lowest expression sum

<Figure size 720x720 with 4 Axes>

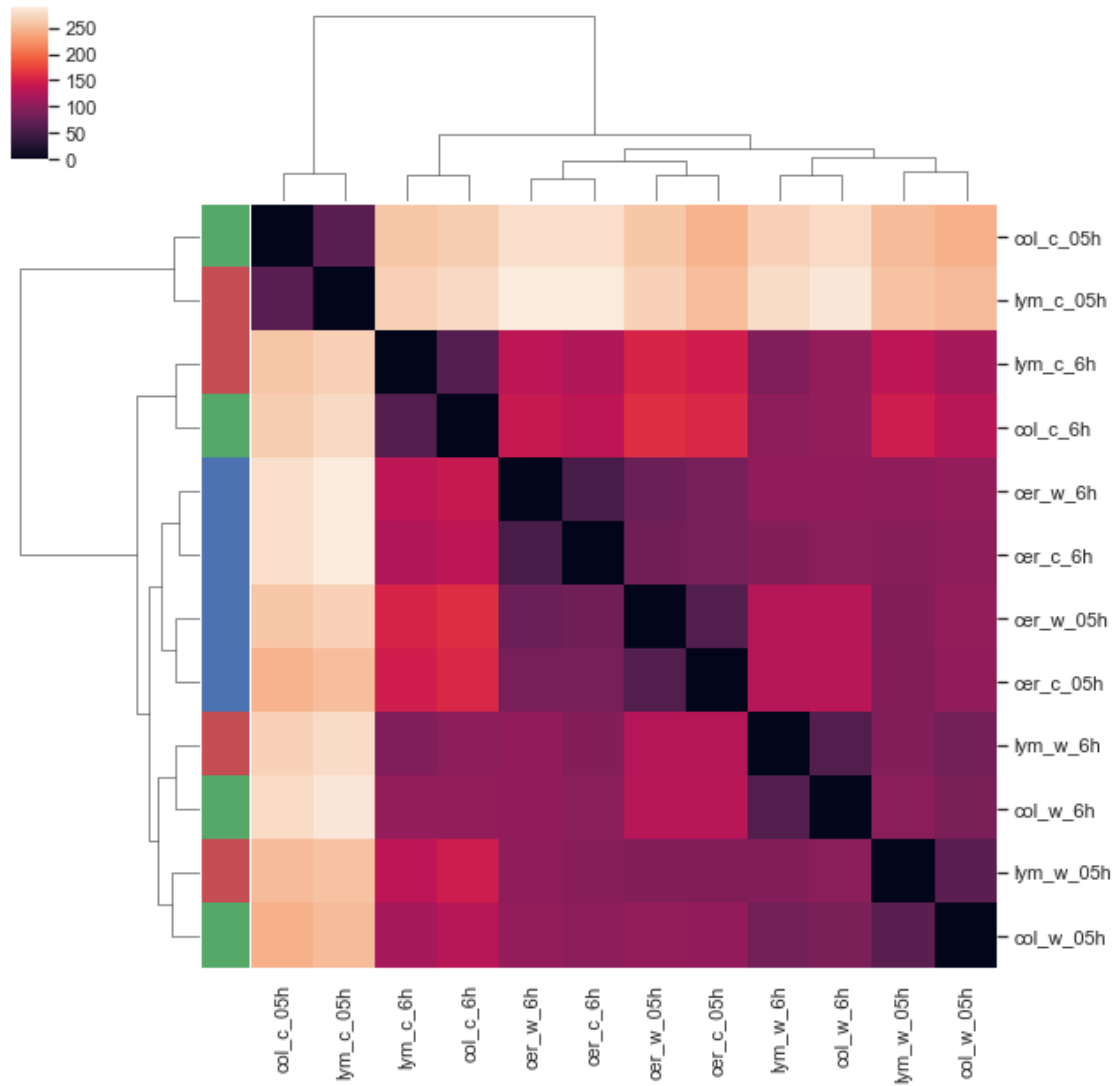


Figure 2: Distance map between samples, pooled together

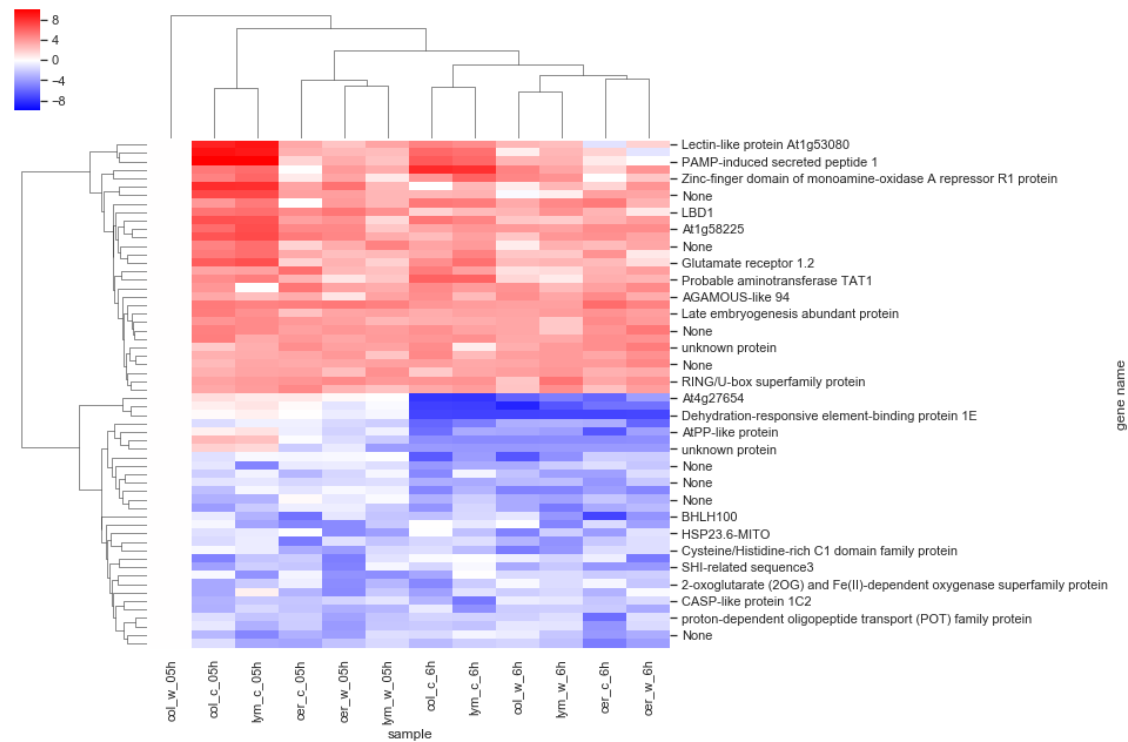


Figure 3: Largest and least DE genes, values given are log2 fold changes /i.e a change of -1 is equivalent to 2^{-1} or 0.5 the amount of control

3.2 Largest/Lowest TF expression

<Figure size 720x720 with 4 Axes>

3.3 Largest/Lowest deviation in log2foldchanges

<Figure size 720x720 with 4 Axes>

3.4 PCA on count data

Explained variance from PC1 & 2 respectively: [0.54031803 0.14242627]

<Figure size 483.925x648 with 6 Axes>

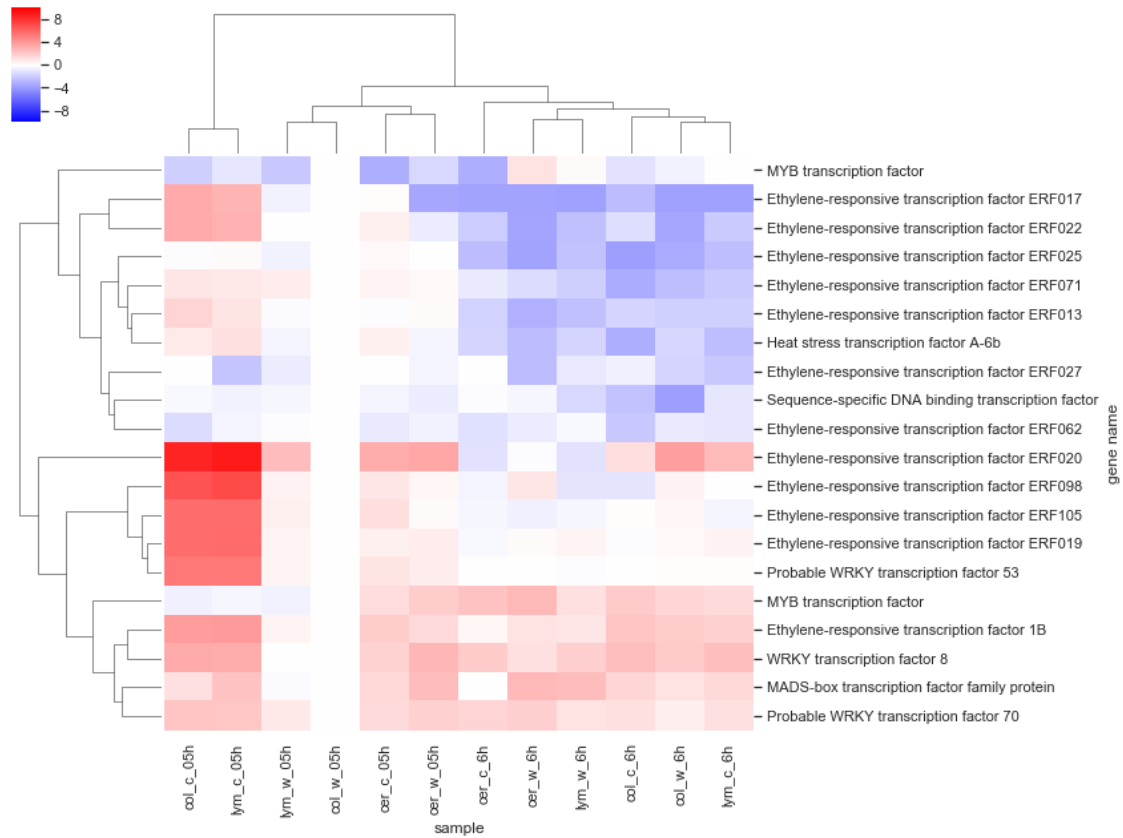


Figure 4: Largest and least DE TFs, values given are log2 fold changes /i.e a change of -1 is equivalent to 2^{-1} or 0.5 the amount of control

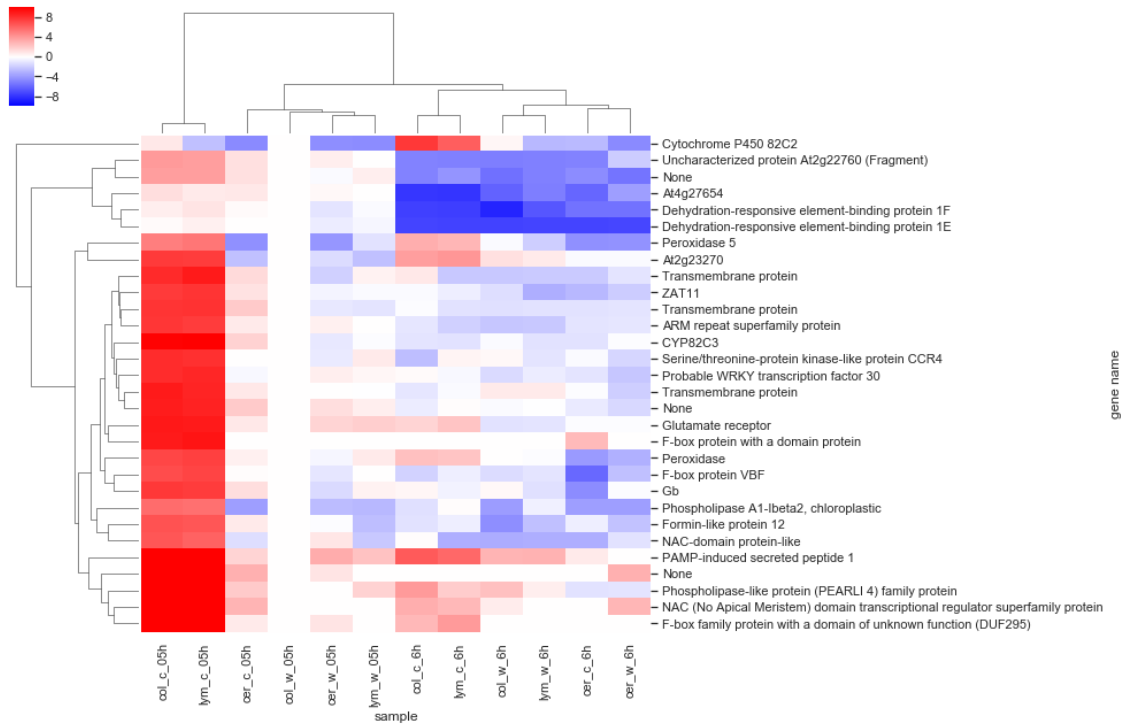


Figure 5: Largest and least std DE genes, values given are log2 fold changes /i.e a change of -1 is equivalent to 2^{-1} or 0.5 the amount of control

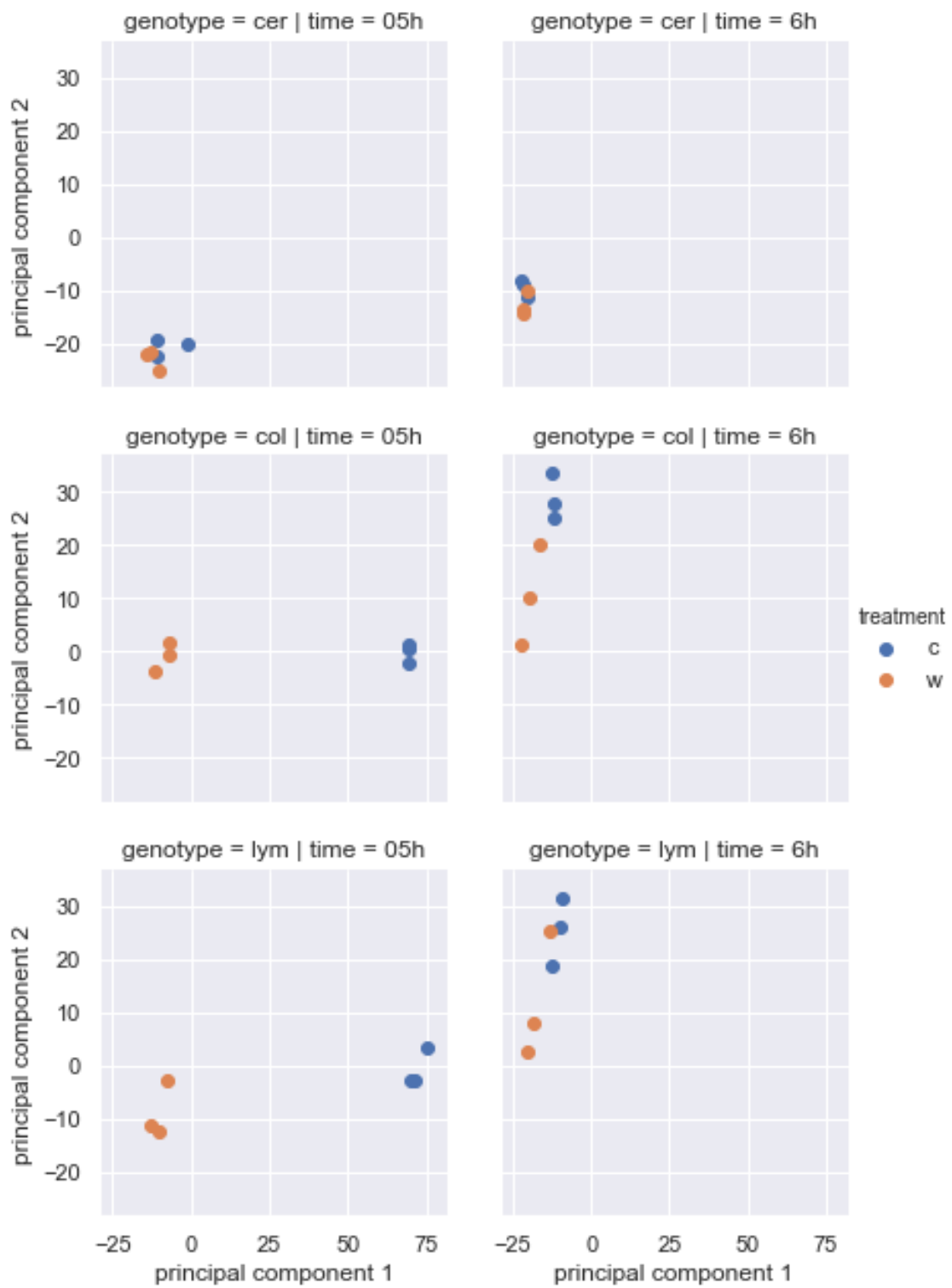


Figure 6: PCA of sample counts

3.5 PCA on expression data

Explained variance from PC1 & 2 respectively: [0.77401765 0.14661658]

<Figure size 483.925x648 with 6 Axes>

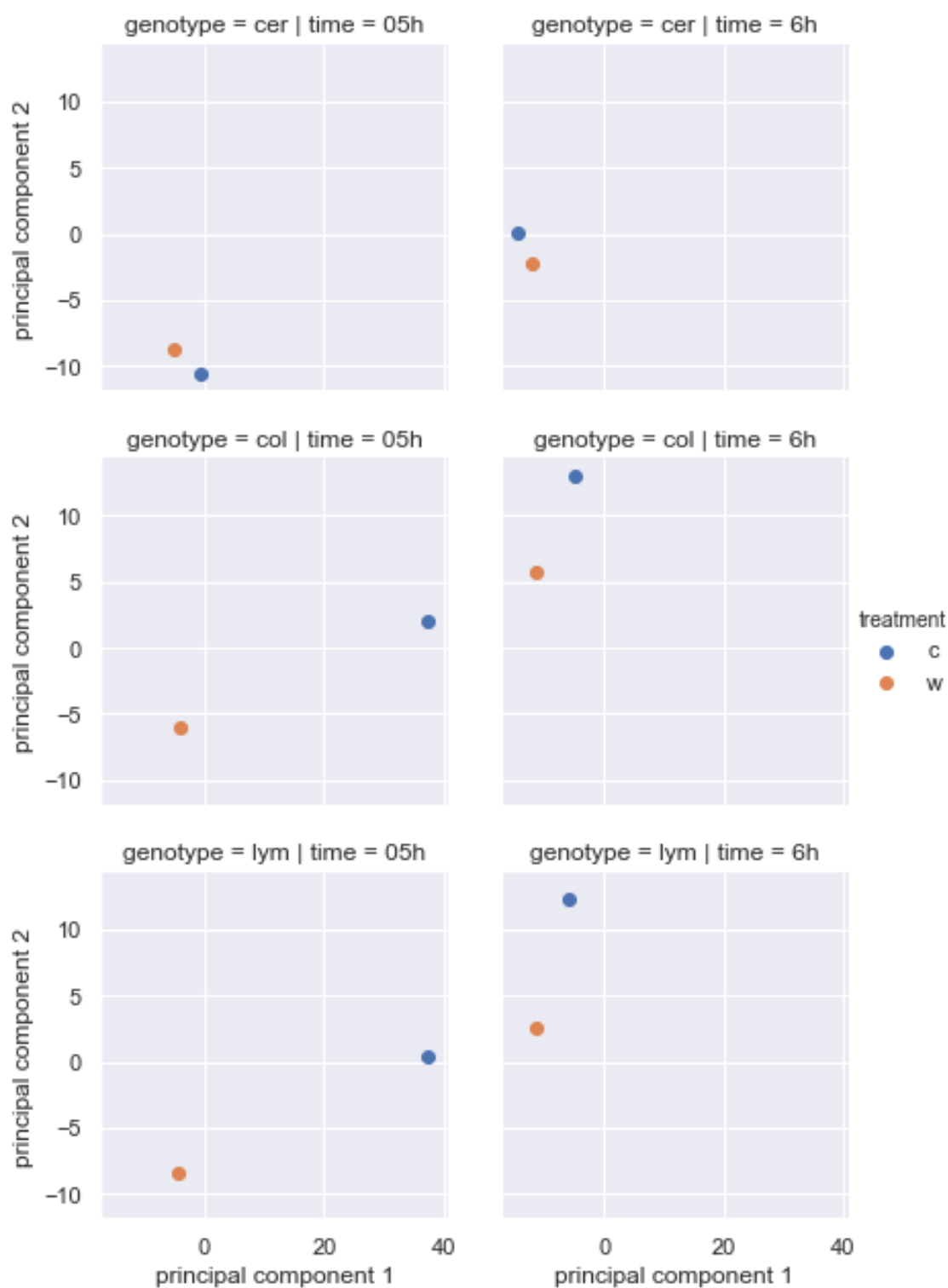


Figure 7: PCA of sample min, max expression