Summarize Peptide Level Measurements

WEW@FGCZ.ETHZ.CH

2019-01-24

Summarize dataset

Table 1: Sample Name to Raw file mapping

Replicate.Name	sampleName	Time
32_S165043_0896	T0	T0
34_S165044_0897	T0~1	T0
2_S165042_0895	T0~2	T0
37_S165059_0933	T168	T168
39_S165058_0932	T168~1	T168
3_S165057_0931	T168~2	T168
8_S165047_0921	T2	T2
14_S165045_0919	T2~1	T2
33_S165046_0920	T2~2	T2
17_S165052_0926	T24	T24
18_S165053_0927	T24~1	T24
28_S165051_0925	T24~2	T24
27_S165062_0936	T240	T240
4_S165061_0935	T240~1	T240
9_S165063_0937	T336	T336
19_S165065_0939	T336~1	T336
22_S165064_0938	T336~2	T336
29_S165055_0929	T72	T72
38_S165056_0930	T72~1	T72
12_S165049_0923	T8	T8
23_S165048_0922	T8~1	T8
24_S165050_0924	T8~2	T8

Table 2: Number of peptides and proteins

NR.Isotope.Label.Type	NR.protein_Id	NR.peptide_Id	NR.precursor_Id	NR.fragment_Id
light	37	117	120	912

Table 3: Number of quantified peptides and proteins per sample.

Isotope.Label.Type	sampleName	protein_Id
light	T0	37
light	T0~1	37
light	T0~2	37
light	T168	37
light	T168~1	37
light	T168~2	37

Isotope.Label.Type	sampleName	protein_Id
light	T2	37
light	T2~1	37
light	T2~2	37
light	T24	37
light	T24~1	37
light	T24~2	37
light	T240	37
light	T240~1	37
light	T336	37
light	T336~1	37
light	T336~2	37
light	T72	37
light	T72~1	37
light	T8	37
light	T8~1	37
light	T8~2	37

```
## [1] "sampleName"
```

Table 4: nr of proteins with more than on peptide.

protein_with	n
one	10
two and more	27

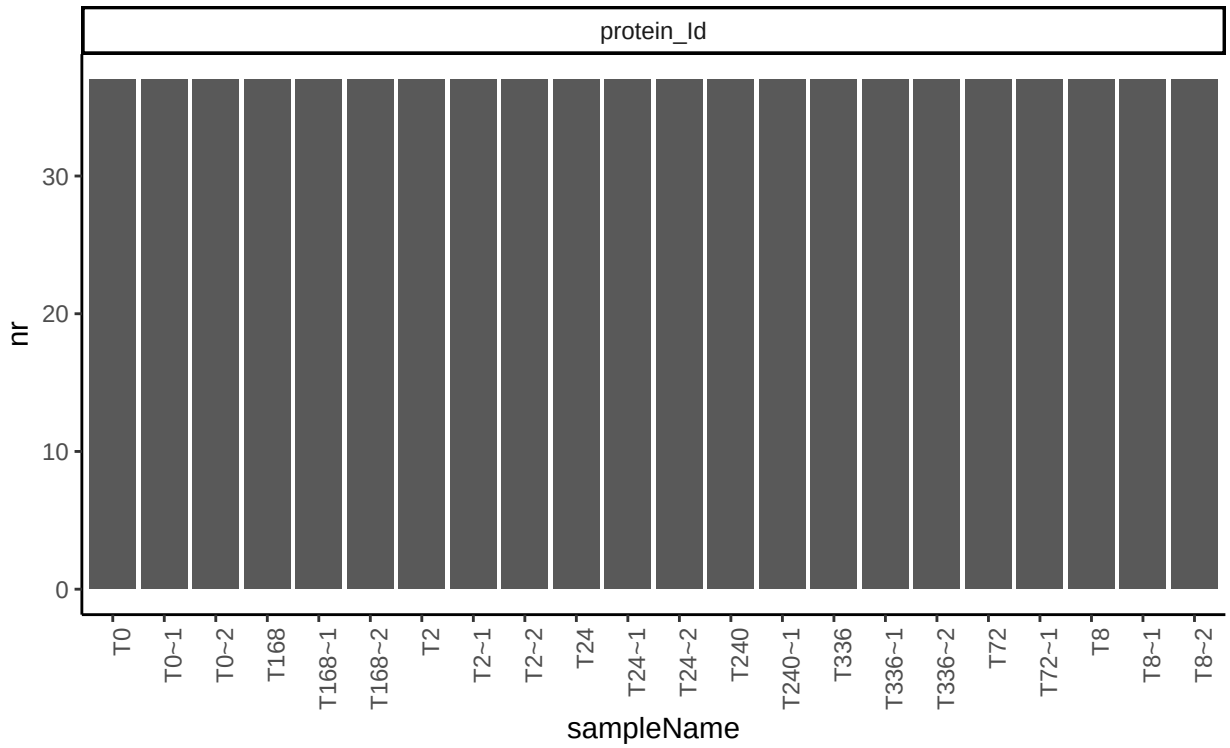

Figure 1: Number of quantified peptides and proteins per sample.

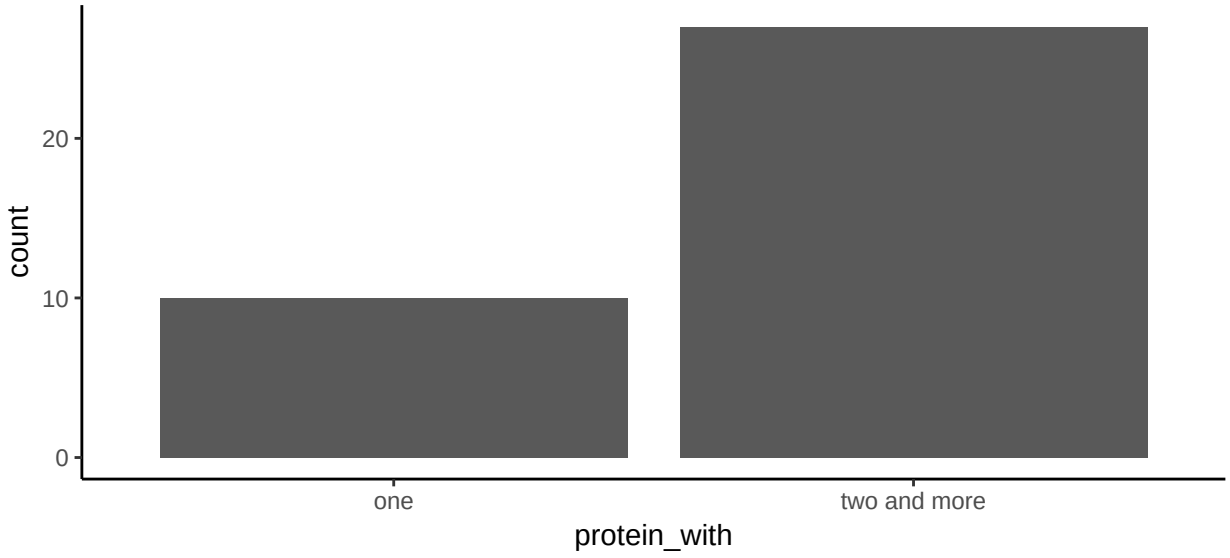

Figure 2: Number of proteins with one or more peptides.

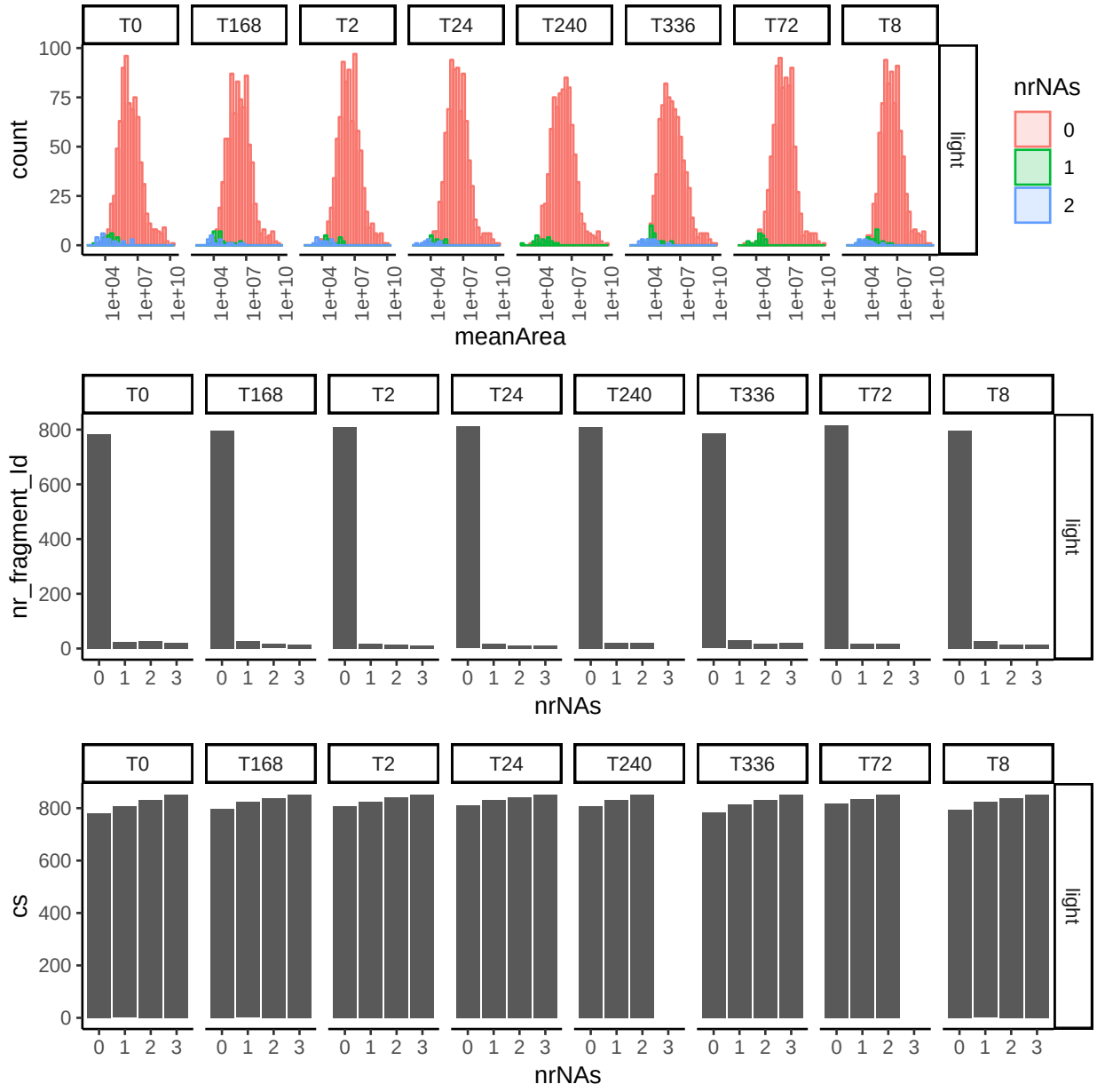

Figure 3: Top - intensity distribution of peptides with 0, 1 etc. missing values. B - number of peptides with 0,1,2 etc. mssing value.

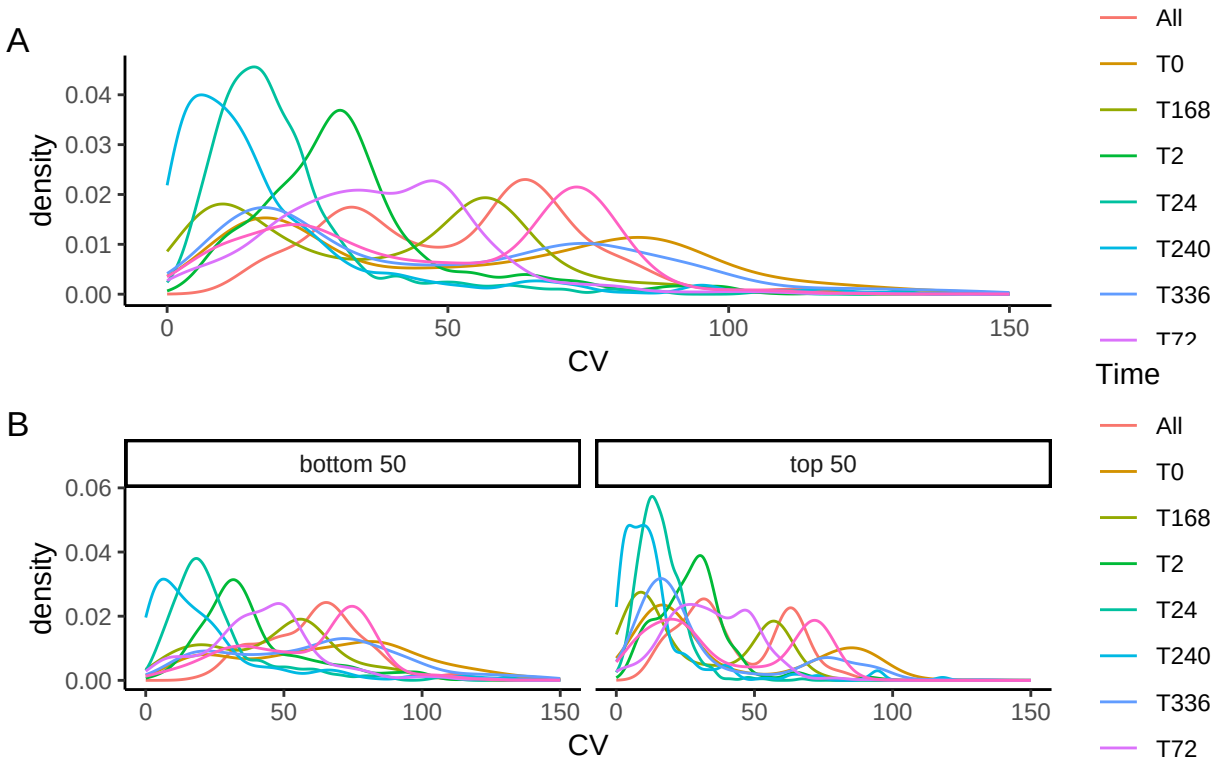

Figure 4: Density plot of peptide level Coefficient of Variations (CV).

Summarize Quantification

Raw Intensity Distribution

Table 5: CV at 0.1, 0.25, 0.5, 0.75, 0.9 quantiles.

probs	All	T0	T168	T2	T24	T240	T336	T72	T8
0.10	25.37090	12.74441	6.264258	13.79001	7.702428	2.395230	11.62694	16.17168	11.17375
0.25	33.66106	19.68010	12.393065	22.08158	11.929849	5.643386	18.39062	25.27044	23.09587
0.50	57.69672	54.38895	40.868328	30.29810	17.162201	12.333292	42.64204	37.17821	50.89922
0.75	67.13014	83.59988	57.580415	37.91759	23.748520	22.841515	75.12487	48.67743	72.63123
0.90	79.82608	96.66267	66.261026	56.95118	34.395925	45.954875	92.39457	57.48119	77.47036

Transformed Intensity Distribution

We apply the `vsn::justvsn` transformation to the data. This transformation transfroms and scales the data to reduce the variance. Because of this, we can't report CV anymore but report standard deviations.

Table 6: CV at 0.1, 0.25, 0.5, 0.75, 0.9 quantiles.

probs	All	T0	T168	T2	T24	T240	T336	T72	T8
0.10	0.4407612	0.2195275	0.2329857	0.0861232	0.0727729	0.0258206	0.2218088	0.0358502	0.3716141
0.25	0.5228299	0.3873940	0.4157805	0.1345618	0.1193328	0.0863851	0.3431818	0.0997736	0.6228423
0.50	0.6362648	0.5192583	0.5393999	0.2028809	0.1970787	0.1844301	0.4609000	0.1910500	0.8068618
0.75	0.8277867	0.7249180	0.6547826	0.3291582	0.3092970	0.3272676	0.6540983	0.3067420	0.9641441
0.90	1.2119262	0.9192457	0.9887650	0.6387087	0.4858342	0.6872021	0.8788957	0.4936039	1.2248992

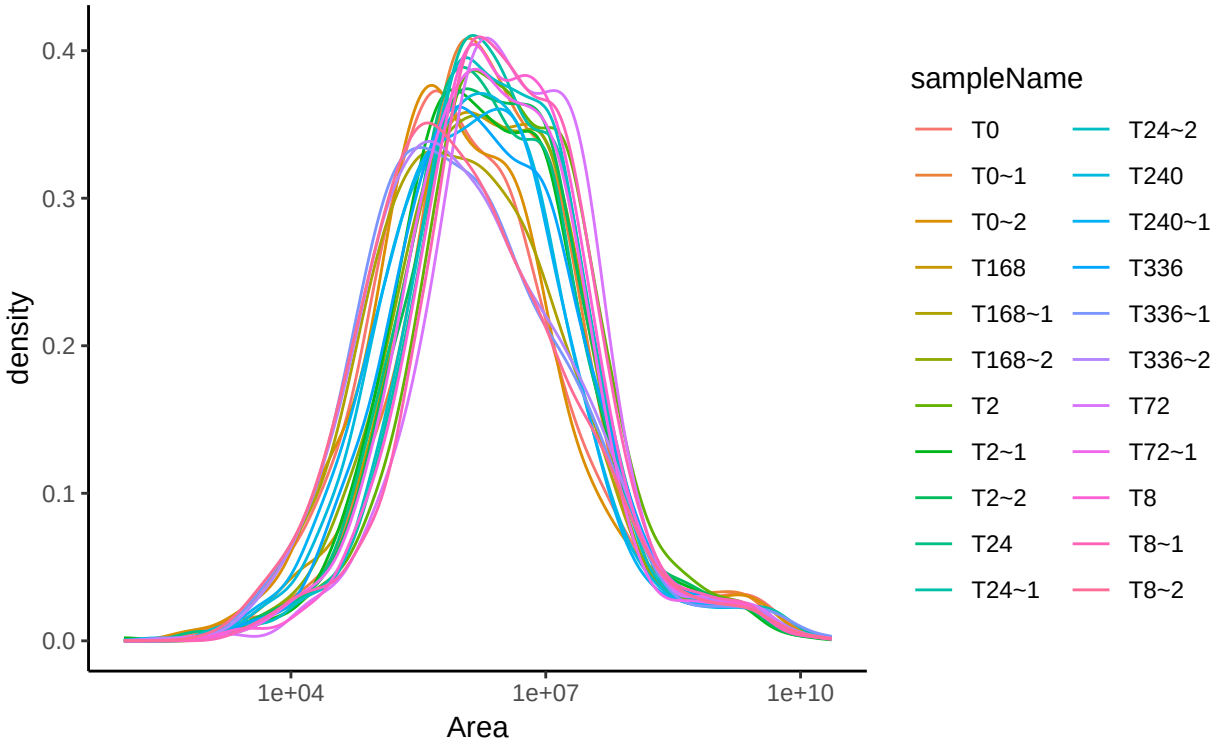

Figure 5: Not normalized intensity distribution.

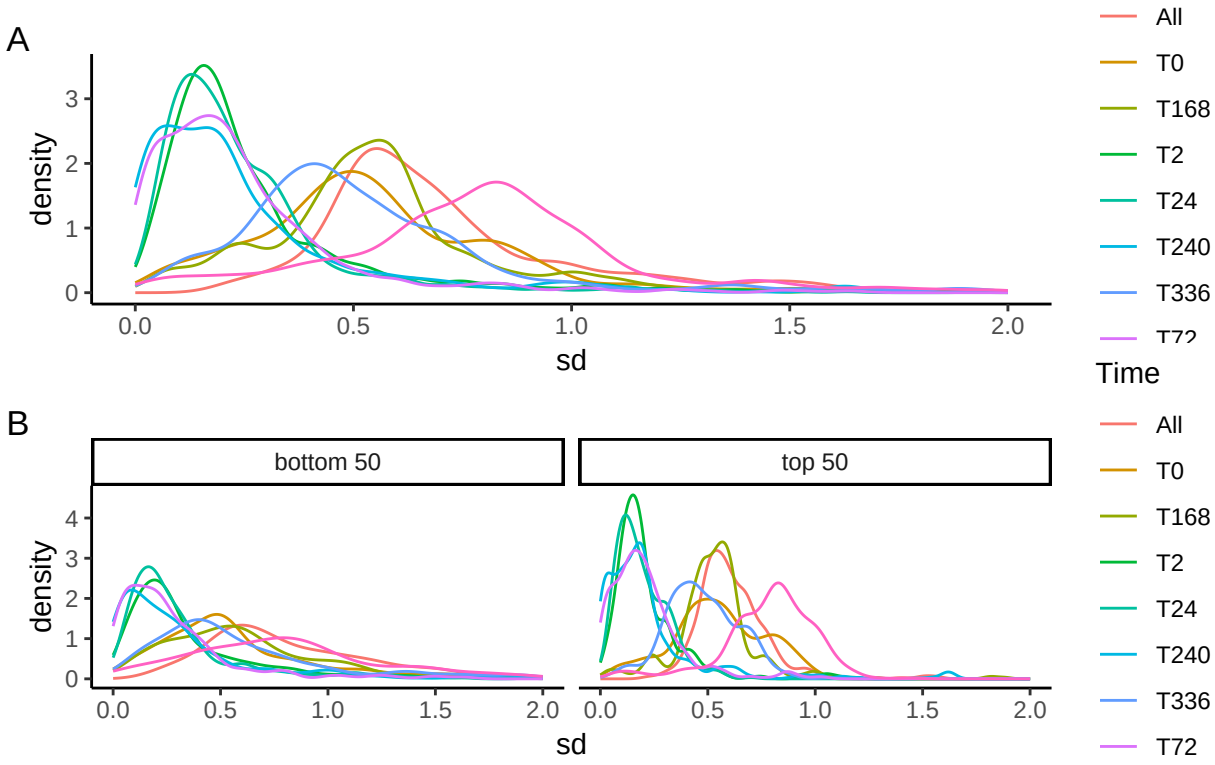

Figure 6: Visualization of peptide standard deviations. A) all. B) - for low (bottom 50) and high intensity peptides (top 50).

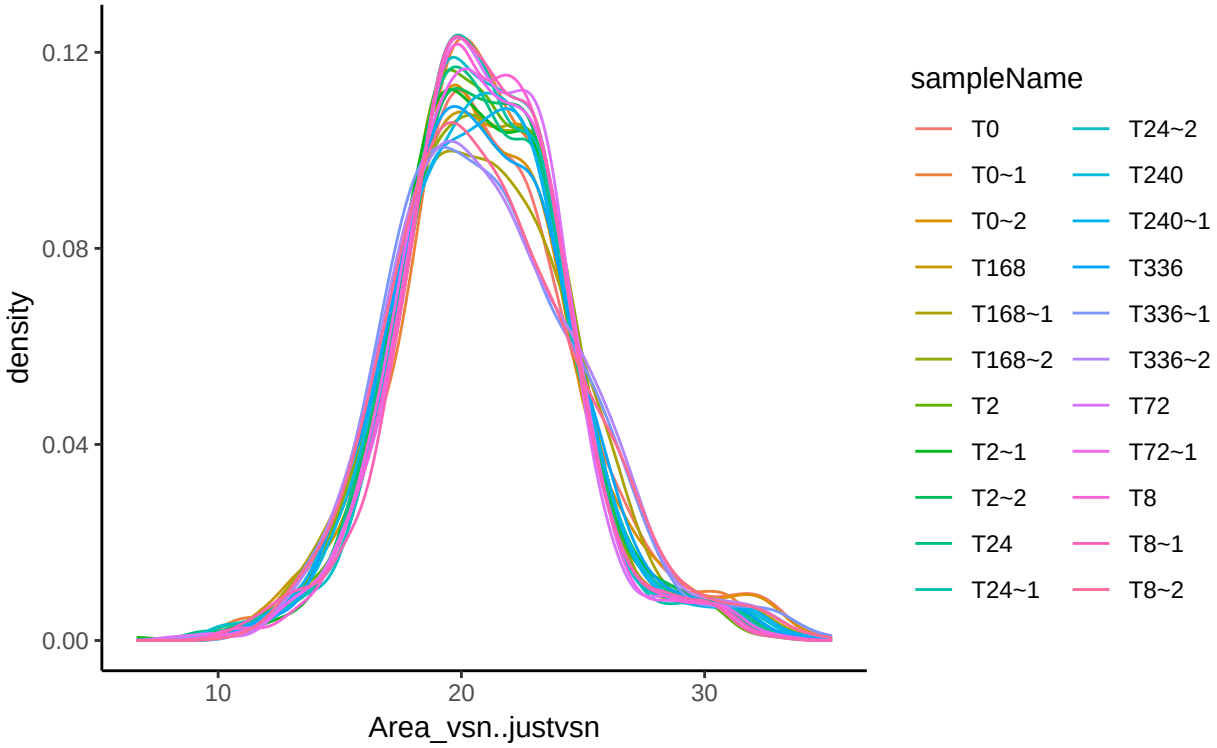

Figure 7: Peptide intensity distribution after transformation.

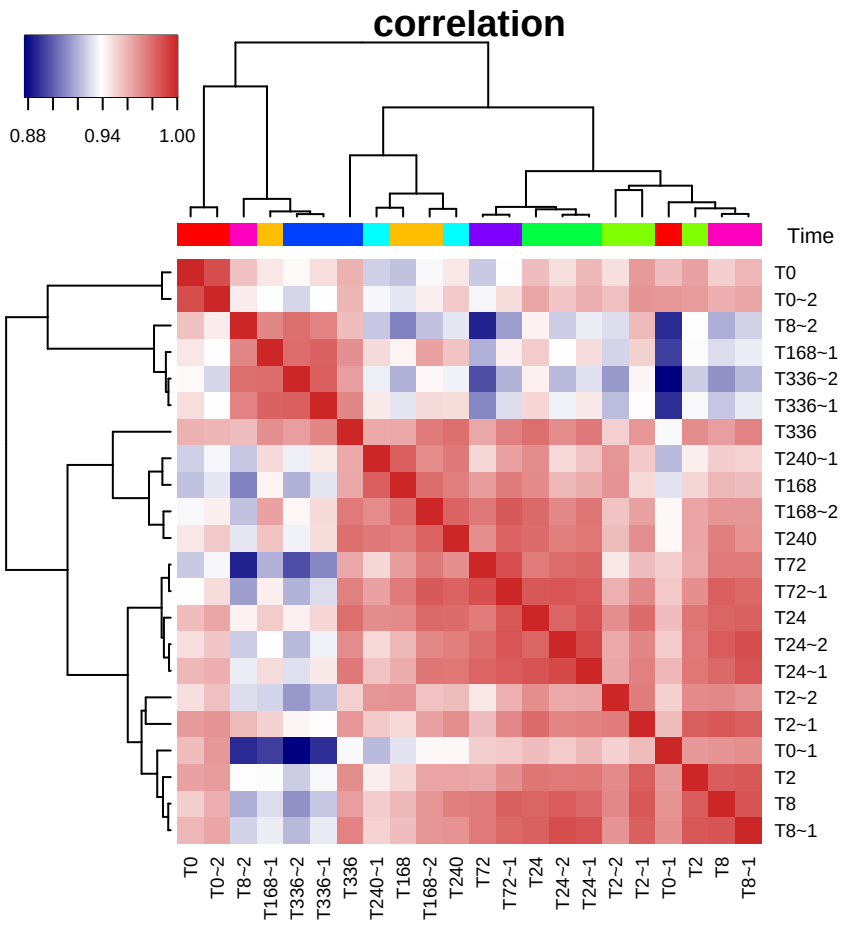

Figure 8: Heatmap of peptide intensity correlation between samples

Sample Size Calculation for peptide level data.

We estimated the *Variance* of the measurement using the QC samples. For each protein we can compute the standard deviation based on the 4 QC samples. The distribution of the standard deviations is shown in Figure @ref(fig:sdviolinplots).

This allows us to calculate the required sample size n. The table @ref(sampleSize) summarizes how many samples are needed to detect a fold change of 2 at a significance level of $1 - 0.05$ and power of 0.8 for $25, 50, 75, 90$ percent of the proteins.

Table 7: Sample size needed to report a fold change greater than 2 with a significance level of 0.05 and power 0.8 when using t-test to compare means.

quantile	sd	N_exact	N
20%	0.145	1.572	2
30%	0.201	1.677	2
40%	0.272	1.817	2
50%	0.362	2.020	3
60%	0.462	2.293	3
70%	0.564	2.638	3
80%	0.709	3.273	4
90%	0.926	4.558	5
100%	4.631	85.146	86

The *power* of a test is $1-\beta$, where *beta* is the probability of a Type 2 error (failing to reject the null hypothesis when the alternative hypothesis is true). In other words, if you have a 20% chance of failing to detect a real difference, then the power of your test is .8.

Significance is equal to $1 - \alpha$, where α is the probability of making a Type 1 Error. That is, alpha represents the chance of a falsely rejecting H0 and picking up a false-positive effect. Alpha is usually set at 0.05, for a 95% significance.

Fold change : Suppose you are comparing a treatment group to a placebo group, and you will be measuring some continuous response variable which, you hope, will be affected by the treatment. We can consider the mean response in the treatment group, μ_1, and the mean response in the placebo group, μ_2. We can then define $\Delta = \mu_1$â£ıμ_2. The smaller the difference you want to detect, the larger the required sample size.