Summary statistics of post-processed MaxQuant output submitted to Expression Atlas.

Ananth

19/02/2020

Summary

Summary of output from MaxQuant (proteinGroups.txt).
PXDID: PXD002029
Protein reference: Human Reference Proteome (UniProt, May 2019)
Dataset: /Users/ananth/Documents/MaxQuant_Bechmarking/Human/PXD002029/proteinGroups.txt

Table 1. Summary

	Pre-processed	Post-processed
Number of samples	8	8
Number of protein groups	4728	3317
Number of reverse decoysˆ	74	0
Number of contaminantsˆ	46	0
Total number of proteins§	22412	16239
....SwissProt	10651	7565
....Trembl	11651	8662
Total number of non-isoforms[a]	16864	12132
....SwissProt.	5103	3458
....Trembl.	11651	8662
Total number of peptides¶	25887	21807
Total number of unique peptides†	22420	19296
Protein groups with 2 or more mapped peptides*	3748	3317
Protein groups mapped to unique gene •	NA	4147

ˆ Some protein groups containing reverse decoys and/or contaminant entries present within are not marked as "+" in "Potential.contaminant" or "Reverse" fields in the output and therefore the total number of contaminants and/or reverse decoys counted here may be less than those present within the protein groups.
§ The total number of proteins present across all protein groups. Number of SwissProt and Trembl entries within are counted. Total number of proteins in preprocessed data includes contaminants, reverse decoys, SwissProt and Trembl proteins. In post-processed data, total number of proteins includes only SwissProt and Trembl entries.
[a] Total number of proteins excluding isoforms present across all protein groups. Total number of non-isoform SwissProt and Trembl entries within.
¶ Sum of peptides that are mapped across all protein groups.
† The total number of unique peptides associated with the protein group (i.e. these peptides are not shared with another protein group).
* Protein groups consisting proteins to which 2 or more peptides are mapped.
• Protein groups consisting proteins which are mapped to an unique Ensembl Gene ID.

Pre-processed data

Output from MaxQuant without any downstream processing

Post-processed data

Procecssed MaxQuant output
Filters applied:
(i) Remove reverse decoys
(ii) Remove contaminants
(iii) Include protein groups that have 2 or more peptides mapped to protein
(iv) Include protein groups wherein all protein IDs in the protein group are mapped to an unique Ensembl Gene ID

Normalisation method: Fraction Of Total.
In this normalisation method each protein iBAQ intensity value is scaled to the total amount of signal in a given MS run (column) and transformed to parts per billion (ppb)

Including Plots

```
## Warning: Transformation introduced infinite values in continuous x-axis
```
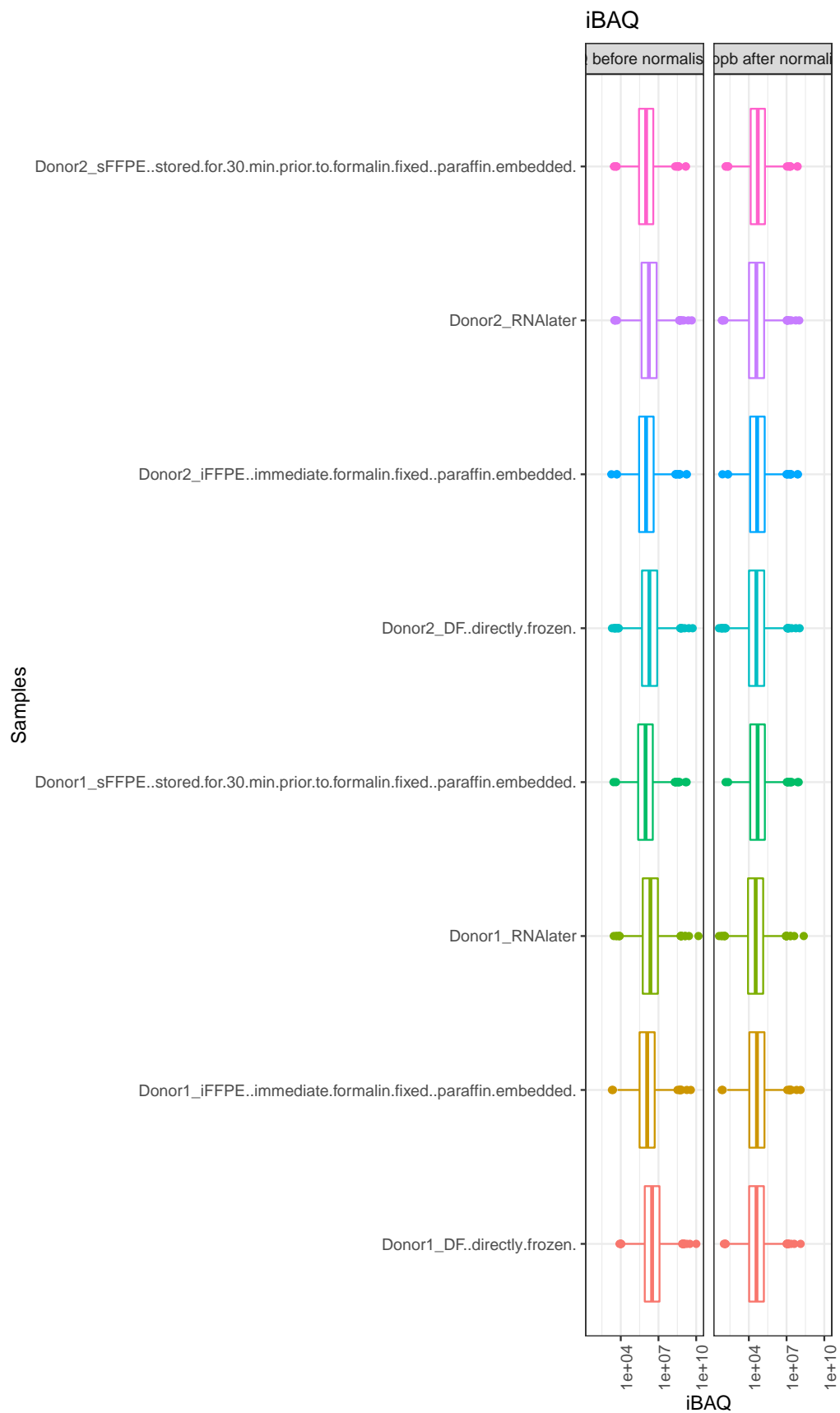
```
## Warning: Removed 2604 rows containing non-finite values (stat_density).
```
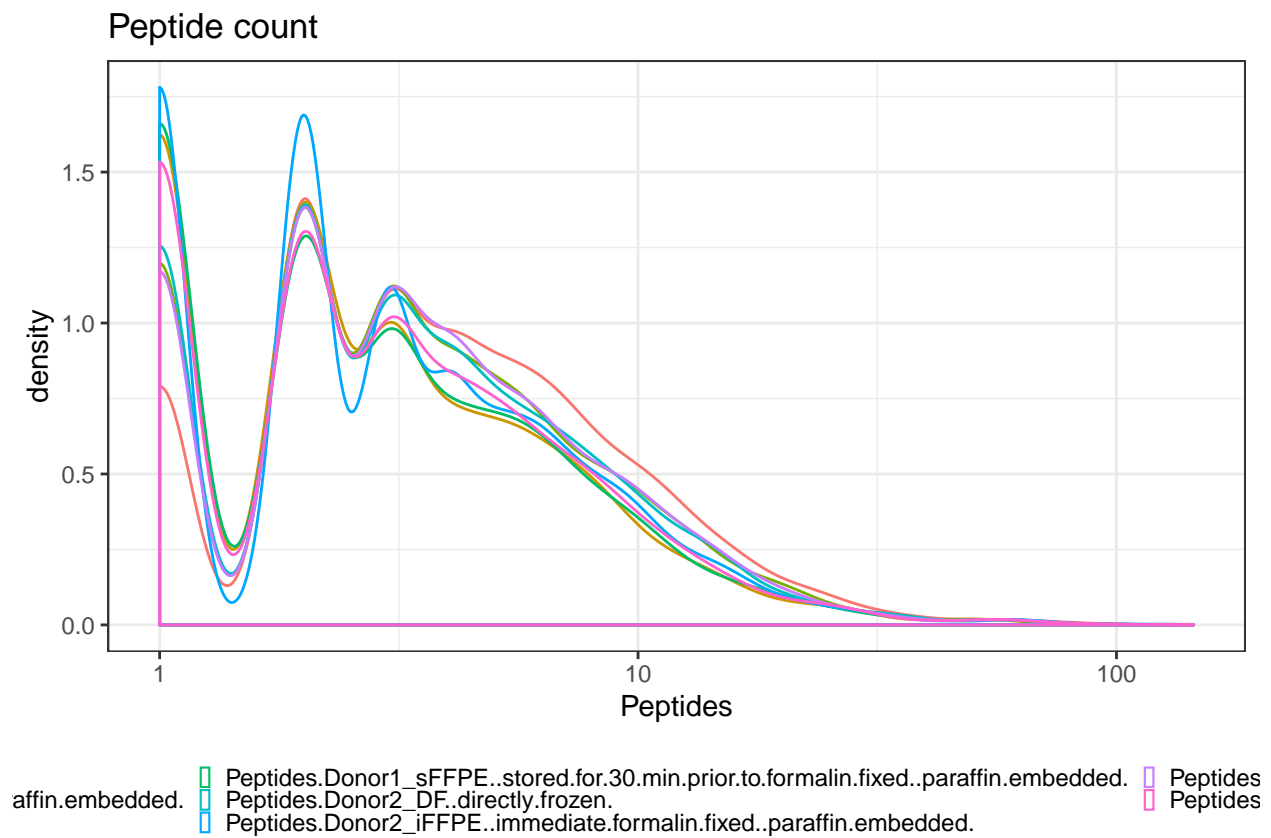

Boxplots with distribution of iBAQ values for each sample before and after FOT normalisation.

```
## Warning: Transformation introduced infinite values in continuous y-axis
```

```
## Warning: Removed 7684 rows containing non-finite values (stat_boxplot).
```
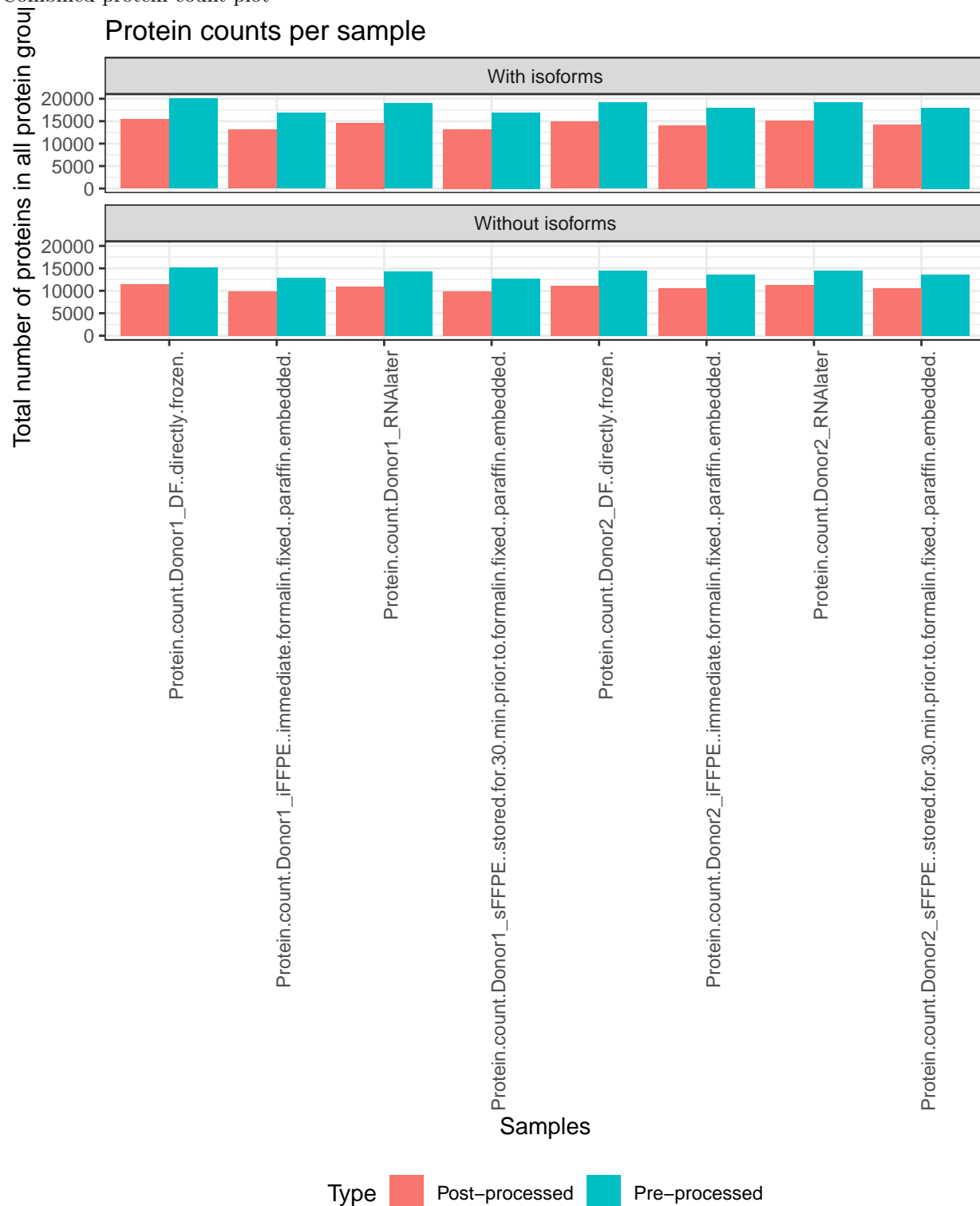
iBAQ

.
Combined protein count plot

Protein counts per sample

¶ Total number of proteins (including isoforms) in post-proccessed samples
† Total number of proteins (without isoforms) in pre-processed samples
€ Total number of proteins (without isoforms) in post-proccessed samples

	Samples	Pre. ALL§	Post. ALL¶	Pre. non-iso
1	Donor1_DF..directly.frozen.	20029	15399	1512
2	Donor1_iFFPE..immediate.formalin.fixed..paraffin.embedded.	16891	13150	1286
3	Donor1_RNAlater	18986	14616	1436
4	Donor1_sFFPE..stored.for.30.min.prior.to.formalin.fixed..paraffin.embedded.	16894	13205	1279
5	Donor2_DF..directly.frozen.	19138	14890	1449
6	Donor2_iFFPE..immediate.formalin.fixed..paraffin.embedded.	17901	14082	1356
7	Donor2_RNAlater	19159	15028	1450
8	Donor2_sFFPE..stored.for.30.min.prior.to.formalin.fixed..paraffin.embedded.	18010	14132	1361

. Combined peptide count plot
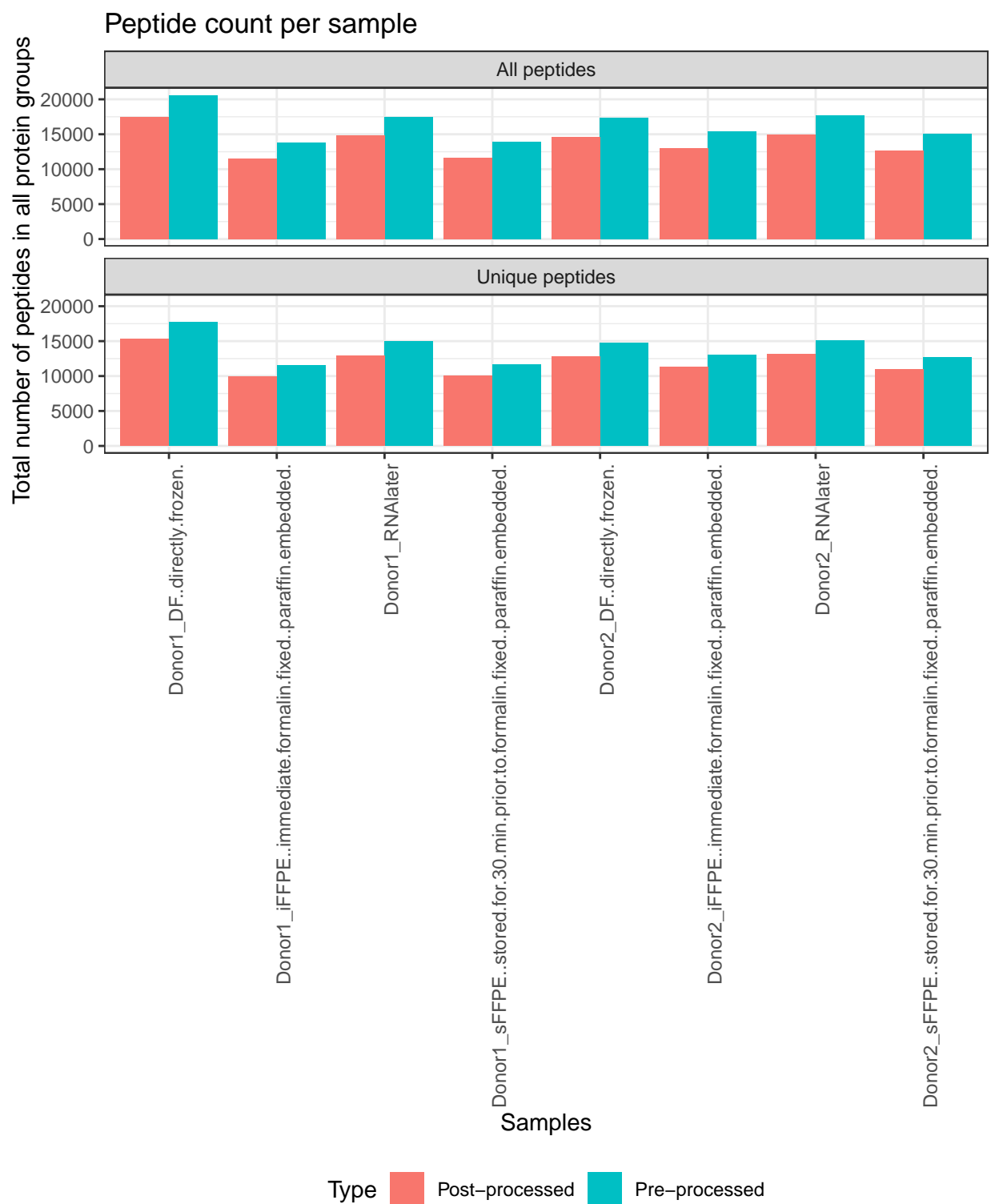
Peptide count per sample

Table 3. Peptide counts.

Total number of all peptides before and after downstream processing

§ Total number of all peptides in pre-processed samples

¶ Total number of peptides in post-proccessed samples

† Total number of unique peptides in pre-processed samples

€ Total number of unique peptides in post-proccessed samples

	Samples	Pre. ALL.§	Post. ALL.¶	Pre. uniqu
1	Donor1_DF..directly.frozen.	20623	17481	176
2	Donor1_iFFPE..immediate.formalin.fixed..paraffin.embedded.	13802	11518	115
3	Donor1_RNAlater	17520	14791	149
4	Donor1_sFFPE..stored.for.30.min.prior.to.formalin.fixed..paraffin.embedded.	13897	11639	116
5	Donor2_DF..directly.frozen.	17320	14628	147
6	Donor2_iFFPE..immediate.formalin.fixed..paraffin.embedded.	15455	13055	130
7	Donor2_RNAlater	17677	14960	150
8	Donor2_sFFPE..stored.for.30.min.prior.to.formalin.fixed..paraffin.embedded.	15057	12715	126

PCA plot.

.paraffin.embedded.

● iBAQ.Donor1_sFFPE..stored.for.30.min.prior.to.formalin.fixed..paraffin.embedded.   ● iBAQ.D<
● iBAQ.Donor2_DF..directly.frozen.   ● iBAQ.D<
● iBAQ.Donor2_iFFPE..immediate.formalin.fixed..paraffin.embedded.