Mass Spectrometry Lemierre - Supplementary plots

Gustav Torisson

Descripton of raw data

- Raw data consists of a wide dataframe with 654 rows with protein names and 24 columns (1 with protein names and 23 samples).
- These were grouped into 8 with Lemierre syndrome ("LS"), 15 other sepsis ("Sepsis").
- There was one measurement per sample.

Initial data management

- Proteins with several names, separated with semicolon(;) were renamed to only the first name (left of semicolon)
- Datapoints labelled as "Filtered" were re-labelled as NA.
- Data was converted to "numeric" format, as it was "character" from the Excel import
- Datapoints labelled as Nan (Not a Number) were also labelled as NA

Log2 transformation

- all measurements were log2 transformed

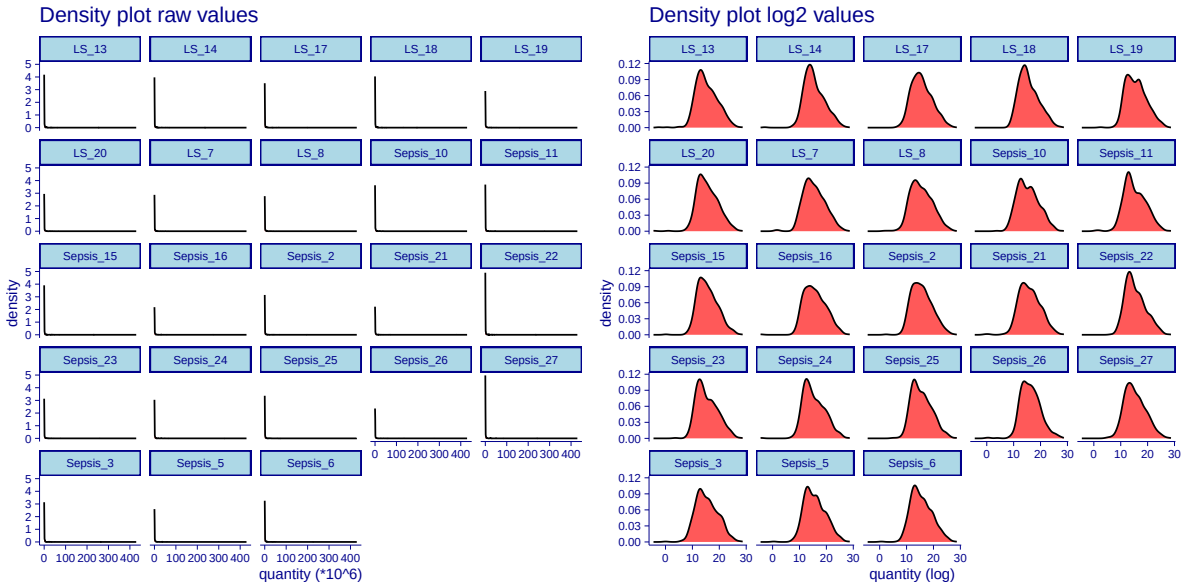

Figure 1: Density plots of raw and log2 transformed values

Filtering

- Of 654 proteins, 341 (52.1%) had complete data, in all 23 samples
- 205 proteins (31.3%) were missing in >= 7 (30% of all) samples.
- These were filtered, leaving 449 proteins

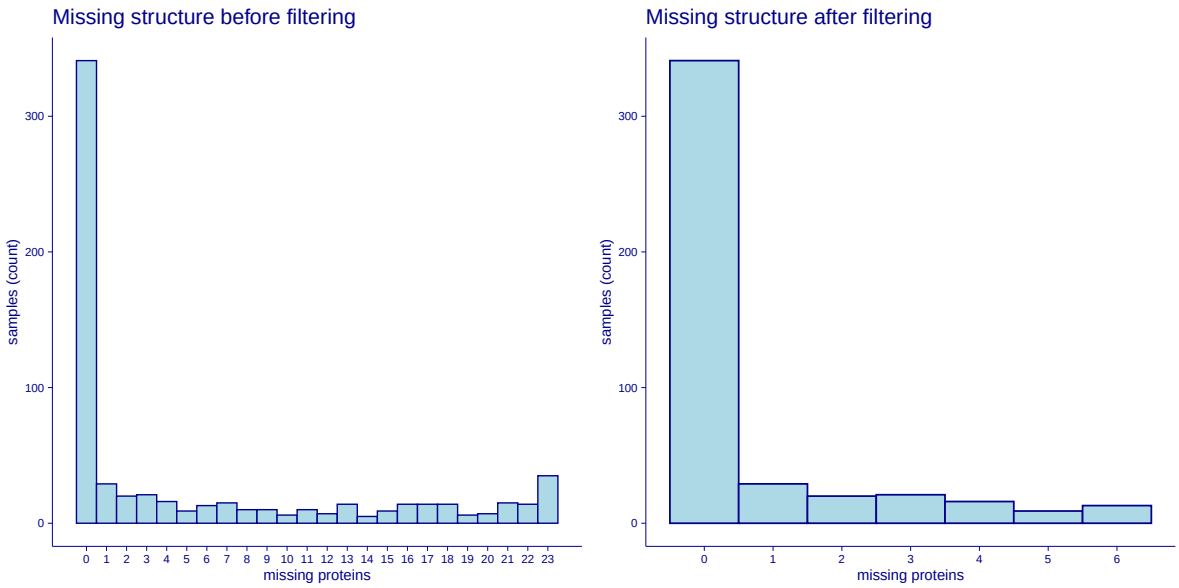

Figure 2: Missing structure before and after filterina

Normalisation

- all values were normalised by sample by subtracting the sample median from the Log2 intensity values

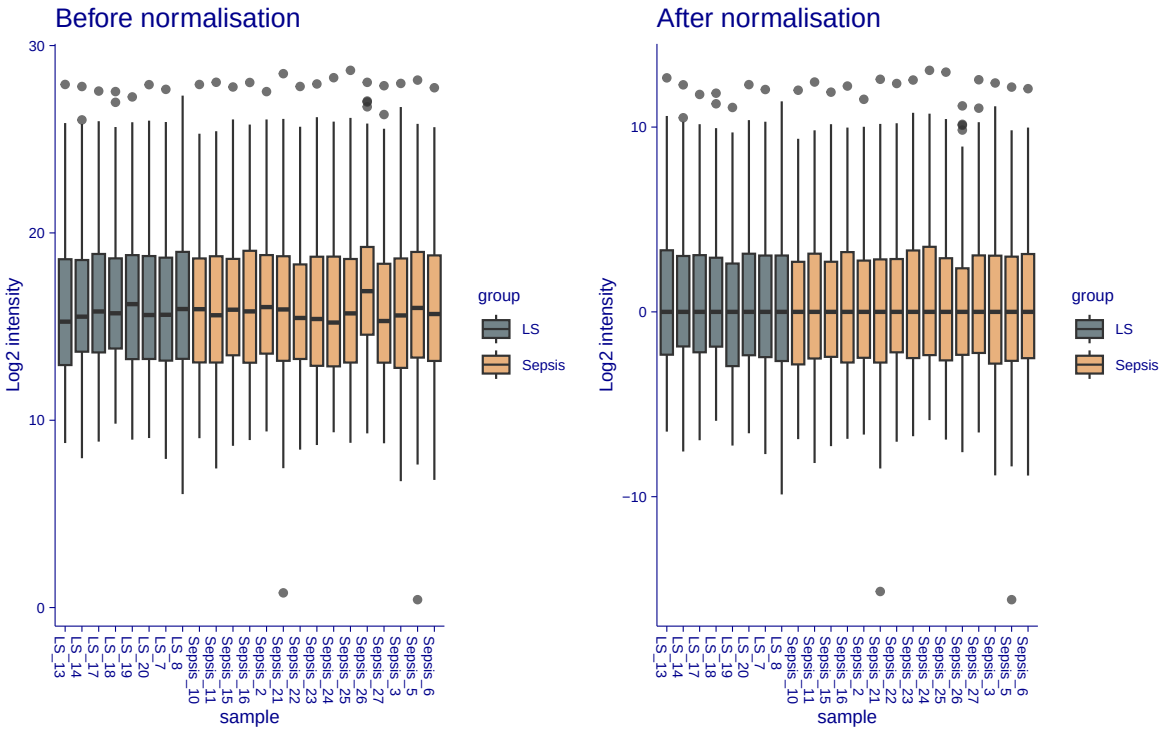

Figure 3: Before and after by-sample normalisation.

Missing per sample and group

- All NAs were considered to represent low intensities and to represent MNAR (missing not at random)

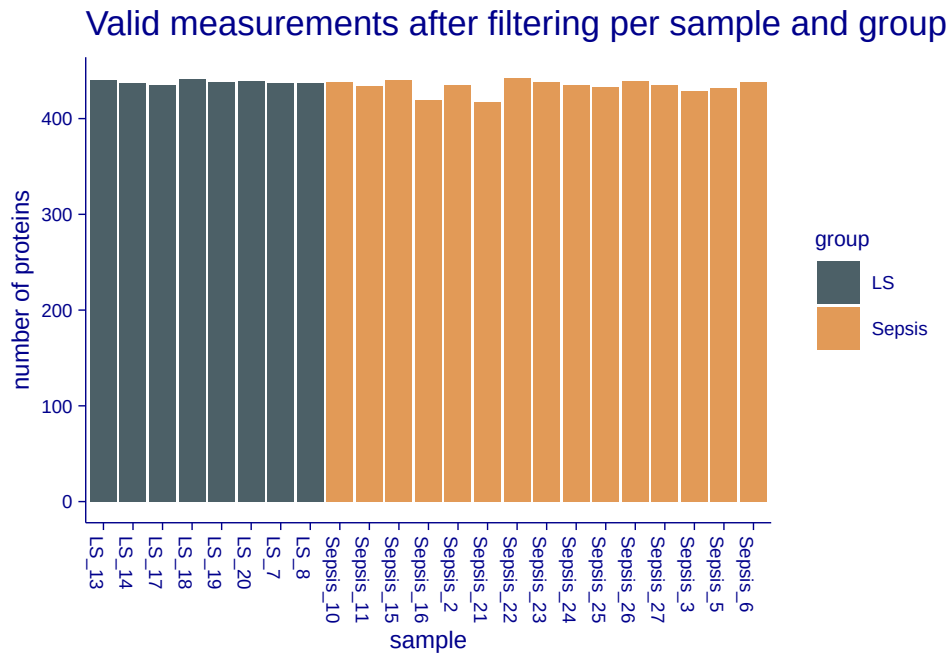

Figure 4: Valid measurements after filtering per sample and group

Imputation

- NAs were imputed using single imputation, assuming MNAR
- For each sample, the sample mean and sample sd were determined
- Then imputations were performed, using a random draw from a Gaussian distribution
- The mean for imputations was donwshifted with -1.8 sample sd and the width 0.3 * sample sd

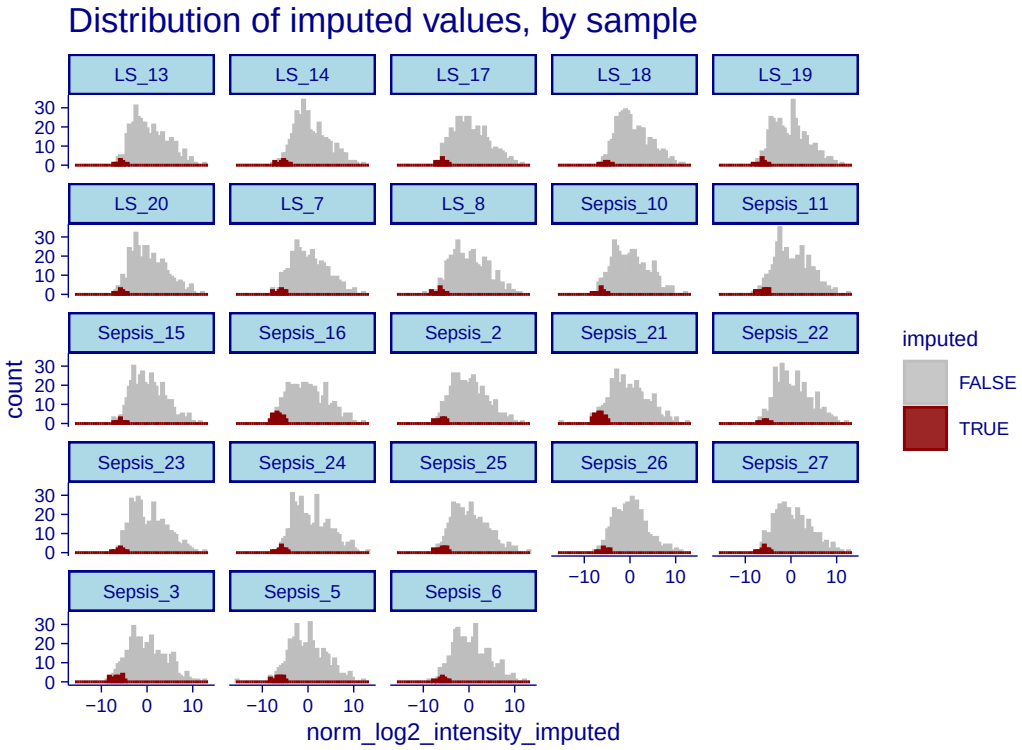

Figure 5: Distribution of imputed values.

Differential expression

- 449 t-tests (students t-test) were performed, one-at-a-time for each protein, between LS and Sepsis groups
- Results are presented as:
  - Log2FC = mean(log(LS)) - mean(log(Sepsis))
  - p values from t-test
  - q values using Benjamini-Hochberg corrections
  - FC(Fold change) = 2^Log2FC
  - Values with Log2FC ± 1.0 and q value < 0.05 were considered significant

Table 1: Differential expression between LS and Sepsis.

protein	log2FC	pval	qval	FC	sign
P05362	2.435316	0.00007	0.0075	5.41	+
P01833	2.235880	0.00203	0.0380	4.71	+
Q8NBJ4	2.211931	0.00013	0.0096	4.63	+
P22897	2.193893	0.00046	0.0173	4.58	+
Q8WWZ8	1.960625	0.00065	0.0224	3.89	+
Q8N6C8	1.954227	0.00021	0.0105	3.88	+
P59665	1.865810	0.00076	0.0243	3.64	+
Q9Y6R7	1.803332	0.00002	0.0047	3.49	+
O43493	1.746654	0.00127	0.0304	3.36	+
P13987	1.732519	0.00018	0.0100	3.32	+
P13796	1.528998	0.00000	0.0013	2.89	+
P18065	1.462872	0.00107	0.0301	2.76	+
P01011	1.234761	0.00032	0.0145	2.35	+
P16070	1.215531	0.00007	0.0075	2.32	+
P08637	1.211105	0.00213	0.0382	2.32	+
P02763	1.198712	0.00039	0.0160	2.30	+
P19652	1.125478	0.00015	0.0096	2.18	+
P02649	1.010606	0.00085	0.0254	2.01	+
P02647	-1.091806	0.00122	0.0304	0.47	+
P80108	-1.759293	0.00012	0.0096	0.30	+

Volcano plot

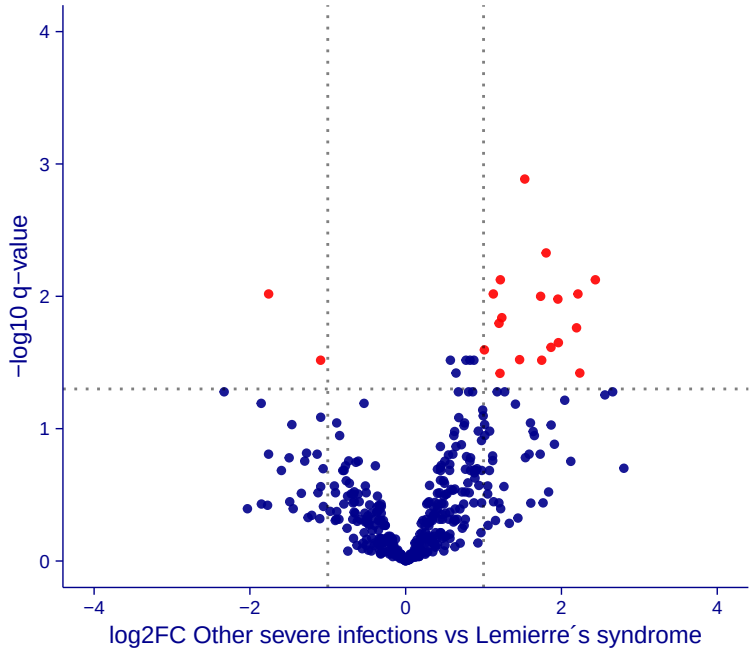

Figure 6: Volcano plot. Proteins that are differentially expressed between LS and Sepsis = red

Heatmap

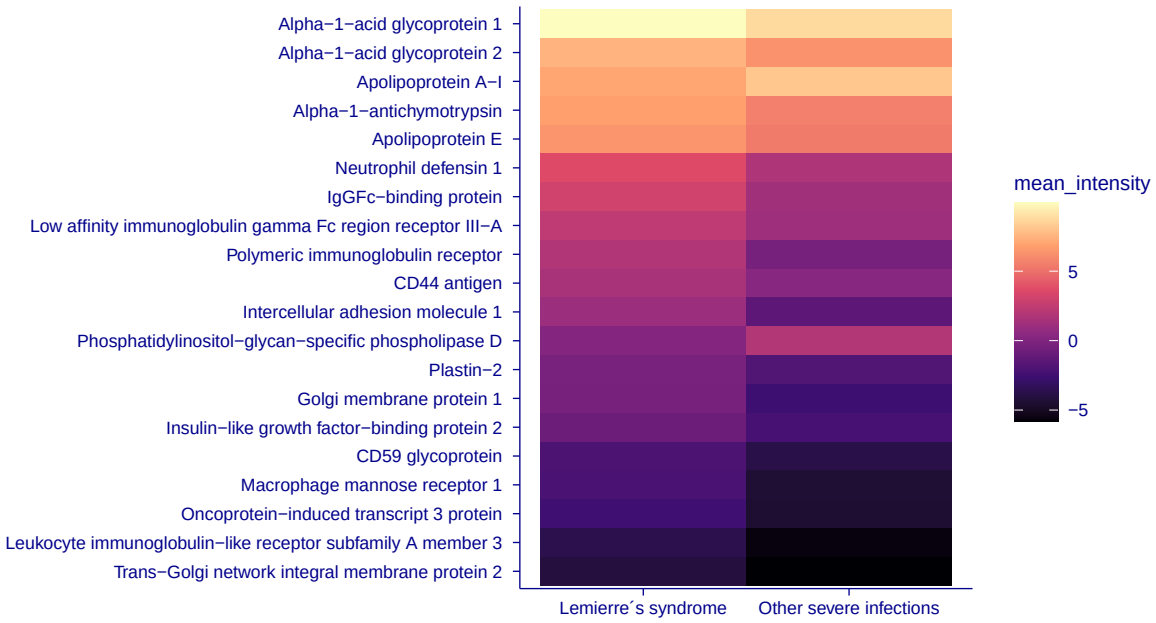

Figure 7: heatmap of differentially expressed proteins between LS and Other severe infections

Table 2: Table underlying the heatmap above. LS and Sepsis are mean normalized log2-transformed intensities

protein_id	protein_name	LS	Sepsis
P02763	Alpha-1-acid glycoprotein 1	9.92	8.72
P19652	Alpha-1-acid glycoprotein 2	7.43	6.31
P02647	Apolipoprotein A-I	6.99	8.08
P01011	Alpha-1-antichymotrypsin	6.78	5.55
P02649	Apolipoprotein E	6.38	5.37
P59665	Neutrophil defensin 1	3.62	1.75
Q9Y6R7	IgGFc-binding protein	3.05	1.25
P08637	Low affinity immunoglobulin gamma Fc region receptor III-A	2.34	1.13
P01833	Polymeric immunoglobulin receptor	1.82	-0.42
P16070	CD44 antigen	1.52	0.31
P05362	Intercellular adhesion molecule 1	1.02	-1.42
P80108	Phosphatidylinositol-glycan-specific phospholipase D	0.14	1.90
P13796	Plastin-2	-0.36	-1.89
Q8NBJ4	Golgi membrane protein 1	-0.40	-2.61
P18065	Insulin-like growth factor-binding protein 2	-0.84	-2.30
P13987	CD59 glycoprotein	-2.07	-3.80
P22897	Macrophage mannose receptor 1	-2.18	-4.38
Q8WWZ8	Oncoprotein-induced transcript 3 protein	-2.54	-4.50
Q8N6C8	Leukocyte immunoglobulin-like receptor subfamily A member 3	-3.66	-5.61
O43493	Trans-Golgi network integral membrane protein 2	-4.06	-5.81