Figure2

Fig. 2 Comparison of windowing scheme

Comparison of the number of identified peptides on DIA-LIT-based method on Normal scanning mode for WhisperTM 100 20 SPD from 1 ng of tryptic HeLa lysate with different ITs at fixed 40 isolation windows. Identified peptides with a coefficient of variation (CV) between 10% and 15% are colored with light red and those with a CV below 10% with dark red. The cycle times for the methods are indicated by their injection times: 38 ms 2.4 s, for 60 ms 3.28 s, for 80 ms 4.09 s, and for 100 ms 4.91 s.

Load packages

```
library("tidyverse")
library("ggpubr")
library("patchwork")
library("ggpointdensity")
#set bw theme as default theme
theme_set(theme_bw())
```

Load data

The path to every folder used is stored to keep everything compact. LIT data is stored in the folder located in lit_path . To reproduce analysis the path must be changed to corresponding path on the local computer.

```
lit_path <-"C:/Users/lukwolt/Documents/R/LIT/"
```

Data tables were exported from Spectronaut in csv files and need to be loaded in RStudio with the function `read.csv2()`.

20 SPD

1 ng

```
#34w
LITDIA_Normal_20SPD_34w_1 <- read.csv2(paste0(lit_path, "20220218_LIT_Normal_
DIA_1ng_20SPD_whisper100_1CV_auto_34w.csv"))
#40w
LITDIA_Normal_20SPD_40w_1 <- read.csv2(paste0(lit_path, "20220502_LIT_Normal_
DIA_1ng_20SPD_whisper100_1CV_auto.csv"))
#45w
LITDIA_Normal_20SPD_45w_1 <- read.csv2(paste0(lit_path, "20220218_LIT_Normal_
DIA_1ng_20SPD_whisper100_1CV_auto_45w.csv"))
```

100 ng

```
#34w
LITDIA_Normal_20SPD_34w_100 <- read.csv2(paste0(lit_path, "20220418_LIT_Normal_DIA_100ng_20SPD_whisper100_1CV_auto_34w.csv"))
#40w
LITDIA_Normal_20SPD_40w_100 <- read.csv2(paste0(lit_path, "20220413_LIT_Normal_DIA_100ng_20SPD_whisper100_1CV_auto_40w.csv"))
#45w
LITDIA_Normal_20SPD_45w_100 <- read.csv2(paste0(lit_path, "20220418_LIT_Normal_DIA_100ng_20SPD_whisper100_1CV_auto_45w.csv"))
```

40SPD

1 ng

```
#34w
LITDIA_Normal_34w_1 <- read.csv2(paste0(lit_path, "20220509_LIT_Normal_DIA_1ng_40SPD_whisper100_1CV_auto_34w.csv"))
#40w
LITDIA_Normal_40w_1 <- read.csv2(paste0(lit_path, "20220421_LIT_Normal_DIA_1ng_40SPD_whisper100_1CV_auto.csv"))
#45w
LITDIA_Normal_45w_1 <- read.csv2(paste0(lit_path, "20220509_LIT_Normal_DIA_1ng_40SPD_whisper100_1CV_auto_45w.csv"))
```

100 ng

```
#34w
LITDIA_Normal_34w_100 <- read.csv2(paste0(lit_path, "20220509_LIT_Normal_DIA_100ng_40SPD_whisper100_1CV_auto_34w.csv"))
#40w
LITDIA_Normal_40w_100 <- read.csv2(paste0(lit_path, "20220421_LIT_Normal_DIA_100ng_40SPD_whisper100_1CV_auto.csv"))
#45w
LITDIA_Normal_45w_100 <- read.csv2(paste0(lit_path, "20220509_LIT_Normal_DIA_100ng_40SPD_whisper100_1CV_auto_45w.csv"))
```

Clean data tables

Imported tables contain a lot of information, which are not useful for our analysis. The function `filter_pep()` only keeps the file names (replicates), quantitative values for each file name, and the peptide sequence. It is also calculating the mean between the four replicates. Spectronaut does some imputation by default , but we want to keep missing values. Imputed values are replaced by NA. Peptides missing in more than one replicate are highlighted as TRUE.

```
filter_pep <- function(x) {
  x$FG.Quantity[as.logical(x$EG.IsImputed)] <- NA
  y <- x %>%
```

```
  group_by(R.FileName, PEP.StrippedSequence) %>%
  summarise(mean_quantity = mean(FG.Quantity, na.rm = TRUE)) %>%
  pivot_wider(names_from = R.FileName,
              values_from = mean_quantity) %>%
    suppressMessages()
  z <- y %>%
  mutate(na_nbr = rowSums(is.na(y[-1]))) %>%
  mutate(mean_quantity = rowMeans(y[2:ncol(y)], na.rm = TRUE)) %>%
  mutate(filtered = !na_nbr <= 1) %>%
  select(-na_nbr) %>%
  suppressMessages("`summarise()` has grouped output by 'R.FileName'. You can
override using the `.groups` argument.")
  message("Highlighted ", sum(z$filtered), " peptide(s) found in less than ",
ncol(y) -2, " replicates")

 return(z) }
```

The `filter_pep()` function is used for every data-set. Furthermore a "filter" data-set, that has removed all peptides missing in more than one replicate, was created.

20 SPD 1 ng

```
#34w
LITDIA_Normal_20SPD_34w_1 <-  filter_pep(LITDIA_Normal_20SPD_34w_1)
```

```
## Highlighted 15 peptide(s) found in less than 3 replicates

LITDIA_Normal_20SPD_34w_1_filter <- LITDIA_Normal_20SPD_34w_1[LITDIA_Normal_2
0SPD_34w_1$filtered == FALSE, ]
#40w
LITDIA_Normal_20SPD_40w_1 <-  filter_pep(LITDIA_Normal_20SPD_40w_1)

## Highlighted 39 peptide(s) found in less than 3 replicates

LITDIA_Normal_20SPD_40w_1_filter <- LITDIA_Normal_20SPD_40w_1[LITDIA_Normal_2
0SPD_40w_1$filtered == FALSE, ]
#45w
LITDIA_Normal_20SPD_45w_1 <-  filter_pep(LITDIA_Normal_20SPD_45w_1)

## Highlighted 79 peptide(s) found in less than 3 replicates

LITDIA_Normal_20SPD_45w_1_filter <- LITDIA_Normal_20SPD_45w_1[LITDIA_Normal_2
0SPD_45w_1$filtered == FALSE, ]
```

100 ng

```
#34w
LITDIA_Normal_20SPD_34w_100 <-  filter_pep(LITDIA_Normal_20SPD_34w_100)

## Highlighted 39 peptide(s) found in less than 2 replicates
```

```
LITDIA_Normal_20SPD_34w_100_filter <- LITDIA_Normal_20SPD_34w_100[LITDIA_Norm
al_20SPD_34w_100$filtered == FALSE, ]
#40w
LITDIA_Normal_20SPD_40w_100 <-  filter_pep(LITDIA_Normal_20SPD_40w_100)

## Highlighted 99 peptide(s) found in less than 3 replicates

LITDIA_Normal_20SPD_40w_100_filter <- LITDIA_Normal_20SPD_40w_100[LITDIA_Norm
al_20SPD_40w_100$filtered == FALSE, ]
#45w
LITDIA_Normal_20SPD_45w_100 <-  filter_pep(LITDIA_Normal_20SPD_45w_100)

## Highlighted 56 peptide(s) found in less than 2 replicates

LITDIA_Normal_20SPD_45w_100_filter <- LITDIA_Normal_20SPD_45w_100[LITDIA_Norm
al_20SPD_45w_100$filtered == FALSE, ]
```

40 SPD 1 ng

```
LITDIA_Normal_34w_1 <-  filter_pep(LITDIA_Normal_34w_1)

## Highlighted 11 peptide(s) found in less than 3 replicates

LITDIA_Normal_34w_1_filter <- LITDIA_Normal_34w_1[LITDIA_Normal_34w_1$filtere
d == FALSE, ]
LITDIA_Normal_40w_1 <-  filter_pep(LITDIA_Normal_40w_1)
```

```
## Highlighted 41 peptide(s) found in less than 3 replicates

LITDIA_Normal_40w_1_filter <- LITDIA_Normal_40w_1[LITDIA_Normal_40w_1$filtere
d == FALSE, ]
LITDIA_Normal_45w_1 <-  filter_pep(LITDIA_Normal_45w_1)

## Highlighted 19 peptide(s) found in less than 3 replicates

LITDIA_Normal_45w_1_filter <- LITDIA_Normal_45w_1[LITDIA_Normal_45w_1$filtere
d == FALSE, ]
```

100 ng

```
LITDIA_Normal_34w_100 <-  filter_pep(LITDIA_Normal_34w_100)

## Highlighted 204 peptide(s) found in less than 3 replicates

LITDIA_Normal_34w_100_filter <- LITDIA_Normal_34w_100[LITDIA_Normal_34w_100$f
iltered == FALSE, ]
LITDIA_Normal_40w_100 <-  filter_pep(LITDIA_Normal_40w_100)

## Highlighted 316 peptide(s) found in less than 3 replicates

LITDIA_Normal_40w_100_filter <- LITDIA_Normal_40w_100[LITDIA_Normal_40w_100$f
iltered == FALSE, ]
LITDIA_Normal_45w_100 <-  filter_pep(LITDIA_Normal_45w_100)
```

```
## Highlighted 405 peptide(s) found in less than 3 replicates

LITDIA_Normal_45w_100_filter <- LITDIA_Normal_45w_100[LITDIA_Normal_45w_100$f
iltered == FALSE, ]
```

Number of identified peptides

Empty vectors to store the number of identified peptides in each replicates for each data-set are created.

```
id_LITDIA_Normal_20SPD_34w_1 <- id_LITDIA_Normal_20SPD_40w_1 <- id_LITDIA_Nor
mal_20SPD_45w_1 <-
  id_LITDIA_Normal_20SPD_34w_100 <- id_LITDIA_Normal_20SPD_40w_100 <- id_LITD
IA_Normal_20SPD_45w_100 <- id_LITDIA_Normal_34w_1 <- id_LITDIA_Normal_40w_1 <
- id_LITDIA_Normal_45w_1 <-
  id_LITDIA_Normal_34w_100 <- id_LITDIA_Normal_40w_100 <- id_LITDIA_Normal_45
w_100 <- rep(NA, 4)
```

The identified peptides, that are not NA, are summed for each replicate and replace the empty vectors.

20 SPD 1 ng 34w

```
for(i in 2:5){
  id_LITDIA_Normal_20SPD_34w_1[i-1] <- sum(!is.na(LITDIA_Normal_20SPD_34w_1[,
i]))}
```

40w

```
for(i in 2:5){
  id_LITDIA_Normal_20SPD_40w_1[i-1] <- sum(!is.na(LITDIA_Normal_20SPD_40w_1[,
i]))}
```

45w

```
for(i in 2:5){
  id_LITDIA_Normal_20SPD_45w_1[i-1] <- sum(!is.na(LITDIA_Normal_20SPD_45w_1[,
i]))}
```

100 ng 34w

```
for(i in 2:5){
  id_LITDIA_Normal_20SPD_34w_100[i-1] <- sum(!is.na(LITDIA_Normal_20SPD_34w_1
00[, i]))}
```

40w

```
for(i in 2:5){
  id_LITDIA_Normal_20SPD_40w_100[i-1] <- sum(!is.na(LITDIA_Normal_20SPD_40w_1
00[, i]))}
```

45w

```
for(i in 2:5){
  id_LITDIA_Normal_20SPD_45w_100[i-1] <- sum(!is.na(LITDIA_Normal_20SPD_45w_1
00[, i]))}
```

id 40 SPD 1 ng 34w

```
for(i in 2:5){
  id_LITDIA_Normal_34w_1[i-1] <- sum(!is.na(LITDIA_Normal_34w_1[, i]))}
```

40w

```
for(i in 2:5){
  id_LITDIA_Normal_40w_1[i-1] <- sum(!is.na(LITDIA_Normal_40w_1[, i]))}
```

45w

```
for(i in 2:5){
  id_LITDIA_Normal_45w_1[i-1] <- sum(!is.na(LITDIA_Normal_45w_1[, i]))}
```

100 ng 34w

```
for(i in 2:5){
  id_LITDIA_Normal_34w_100[i-1] <- sum(!is.na(LITDIA_Normal_34w_100[, i]))}
```

40w

```
for(i in 2:5){
  id_LITDIA_Normal_40w_100[i-1] <- sum(!is.na(LITDIA_Normal_40w_100[, i]))}
```

45w

```
for(i in 2:5){
  id_LITDIA_Normal_45w_100[i-1] <- sum(!is.na(LITDIA_Normal_45w_100[, i]))}
```

Compute CV

The function below takes the data-sets, already filtered by `filter_pep()`, for entry and compute the coefficient of variation (CV) for the replicates of each method. Most of our data-sets contain 4 replicates but some contains only 3. In this case an empty column is added with `compute_CV()` instead of a 4th replicate, to keep the same format for every data-set.

```
compute_CV <- function(x){
  y <- x %>%
    mutate(sd = apply(x[, 2:(ncol(x)-2)],
                      FUN = sd, MARGIN =  1, na.rm = TRUE),
           CV = sd/mean_quantity)
  colnames(y)[2:(ncol(x)-2)] <- paste0("S", 1:(ncol(x)-3))
  if(ncol(x) <= 6){
  y$S4 <- NA
 }
```

```
  return(y)
}
```

Information about the method, the number of samples per day and the input are included afterwards.

20 SPD

```
LITDIA_Normal_20SPD_34w_1_cv <- compute_CV(LITDIA_Normal_20SPD_34w_1)
LITDIA_Normal_20SPD_34w_1_cv$method <- "34w"
LITDIA_Normal_20SPD_34w_1_cv$spd<- "20SPD"
LITDIA_Normal_20SPD_34w_1_cv$input<- "1 ng"

LITDIA_Normal_20SPD_34w_100_cv <- compute_CV(LITDIA_Normal_20SPD_34w_100)
LITDIA_Normal_20SPD_34w_100_cv$method <- "34w"
LITDIA_Normal_20SPD_34w_100_cv$spd<- "20SPD"
LITDIA_Normal_20SPD_34w_100_cv$input<- "100 ng"

LITDIA_Normal_20SPD_40w_1_cv <- compute_CV(LITDIA_Normal_20SPD_40w_1)
LITDIA_Normal_20SPD_40w_1_cv$method <- "40w"
LITDIA_Normal_20SPD_40w_1_cv$spd<- "20SPD"
LITDIA_Normal_20SPD_40w_1_cv$input<- "1 ng"

LITDIA_Normal_20SPD_40w_100_cv <- compute_CV(LITDIA_Normal_20SPD_40w_100)
LITDIA_Normal_20SPD_40w_100_cv$method <- "40w"
LITDIA_Normal_20SPD_40w_100_cv$spd<- "20SPD"
LITDIA_Normal_20SPD_40w_100_cv$input<- "100 ng"

LITDIA_Normal_20SPD_45w_1_cv <- compute_CV(LITDIA_Normal_20SPD_45w_1)
LITDIA_Normal_20SPD_45w_1_cv$method <- "45w"
LITDIA_Normal_20SPD_45w_1_cv$spd<- "20SPD"
LITDIA_Normal_20SPD_45w_1_cv$input<- "1 ng"

LITDIA_Normal_20SPD_45w_100_cv <- compute_CV(LITDIA_Normal_20SPD_45w_100)
LITDIA_Normal_20SPD_45w_100_cv$method <- "45w"
LITDIA_Normal_20SPD_45w_100_cv$spd<- "20SPD"
LITDIA_Normal_20SPD_45w_100_cv$input<- "100 ng"
```

40 SPD

```
LITDIA_Normal_34w_1_cv <- compute_CV(LITDIA_Normal_34w_1)
LITDIA_Normal_34w_1_cv$method <- "34w"
LITDIA_Normal_34w_1_cv$spd<- "40SPD"
LITDIA_Normal_34w_1_cv$input<- "1 ng"



LITDIA_Normal_34w_100_cv <- compute_CV(LITDIA_Normal_34w_100)
LITDIA_Normal_34w_100_cv$method <- "34w"
LITDIA_Normal_34w_100_cv$spd<- "40SPD"
LITDIA_Normal_34w_100_cv$input<- "100 ng"
```

```
LITDIA_Normal_40w_1_cv <- compute_CV(LITDIA_Normal_40w_1)
LITDIA_Normal_40w_1_cv$method <- "40w"
LITDIA_Normal_40w_1_cv$spd<- "40SPD"
LITDIA_Normal_40w_1_cv$input<- "1 ng"

LITDIA_Normal_40w_100_cv <- compute_CV(LITDIA_Normal_40w_100)
LITDIA_Normal_40w_100_cv$method <- "40w"
LITDIA_Normal_40w_100_cv$spd<- "40SPD"
LITDIA_Normal_40w_100_cv$input<- "100 ng"

LITDIA_Normal_45w_1_cv <- compute_CV(LITDIA_Normal_45w_1)
LITDIA_Normal_45w_1_cv$method <- "45w"
LITDIA_Normal_45w_1_cv$spd<- "40SPD"
LITDIA_Normal_45w_1_cv$input<- "1 ng"

LITDIA_Normal_45w_100_cv <- compute_CV(LITDIA_Normal_34w_100)
LITDIA_Normal_45w_100_cv$method <- "45w"
LITDIA_Normal_45w_100_cv$spd<- "40SPD"
LITDIA_Normal_45w_100_cv$input<- "100 ng"
```

Join

Data frames containing CV are combined with the function `rbind()` and arranged with the function `factor()`.

```
join_fig2 <-
  rbind(LITDIA_Normal_20SPD_34w_1_cv, LITDIA_Normal_20SPD_34w_100_cv,
        LITDIA_Normal_20SPD_40w_1_cv, LITDIA_Normal_20SPD_40w_100_cv,
        LITDIA_Normal_20SPD_45w_1_cv, LITDIA_Normal_20SPD_45w_100_cv,
        LITDIA_Normal_34w_1_cv, LITDIA_Normal_34w_100_cv,
        LITDIA_Normal_40w_1_cv, LITDIA_Normal_40w_100_cv,
        LITDIA_Normal_45w_1_cv, LITDIA_Normal_45w_100_cv)
join_fig2$method %>% table()
```

```
## .
##   34w   40w   45w
## 28249 35745 31095
```

```
join_fig2$input %>% table()
```

```
## .
##   1 ng 100 ng
##  27863  67226
```

```
join_fig2$spd %>% table()
```

```
## .
## 20SPD 40SPD
## 55878 39211
```

```
join_fig2$method <- factor(join_fig2$method,
                           levels = unique(join_fig2$method))

join_fig2$input <- factor(join_fig2$input,
                           levels = unique(join_fig2$input))

join_fig2$spd <- factor(join_fig2$spd,
                           levels = unique(join_fig2$spd))
```

A new column called “CV_filter” is added to identify peptides with a CV over 0.15 (1), CV between 0.1 and 0.15 (2) and CV under 0.1 (3).

```
join_fig2$CV_filter <- NA
join_fig2$CV_filter[join_fig2$CV < 0.1] <- "3"
join_fig2$CV_filter[join_fig2$CV >= 0.1 & join_fig2$CV <= 0.15] <- "2"
join_fig2$CV_filter[join_fig2$CV > 0.15] <- "1"
```

Peptides without CV (found in just one replicate) are filtered out and the number of identified peptides for each replicate and each CV_filter is calculated by summing the number of non-missing peptides. The mean and the standard deviation of the replicates are also calculated for each method.

```
join_fig2 <- join_fig2 %>%
  filter(!is.na(CV_filter)) %>%
  group_by(method, spd, input, CV_filter) %>%
  summarise(id_S1 = sum(!is.na(S1)),
            id_S2 = sum(!is.na(S2)),
            id_S3 = sum(!is.na(S3)),
            id_S4 = sum(!is.na(S4)))
```

```
## `summarise()` has grouped output by 'method', 'spd', 'input'. You can over
ride
## using the `.groups` argument.

join_fig2$id_S4[join_fig2$id_S4 == 0] <- NA

join_fig2$mean_id <- rowMeans(join_fig2[5:8], na.rm = TRUE)
join_fig2$sd_id <- apply(join_fig2[5:8], FUN =  sd, MARGIN = 1, na.rm = TRUE)
```

Plot

The mean and the standard deviation of identified peptides for each CV_filter and method are plotted in a stacked col plot using ggplot().The function geom_errorbar() was used to compute the length of the errorbars from the mean and the standard deviation. geom_col(), labs(), theme_bw(), theme() scale_fill_manual() have just aesthetic effects. With scale_y_continuous() a manual limit of the y axis was set.

```
join_fig2$y_errorbar <- join_fig2$mean_id
join_fig2$y_errorbar[join_fig2$CV_filter == "1"] <-
```

```
  join_fig2$y_errorbar[join_fig2$CV_filter == "1"] +
  join_fig2$y_errorbar[join_fig2$CV_filter == "2"]+
  join_fig2$y_errorbar[join_fig2$CV_filter == "3"]
join_fig2$y_errorbar[join_fig2$CV_filter == "2"] <-
  join_fig2$y_errorbar[join_fig2$CV_filter == "2"]+
  join_fig2$y_errorbar[join_fig2$CV_filter == "3"]

figure2 <- join_fig2 %>%
  ggplot(aes(x = method, y = mean_id, fill = CV_filter))+
  geom_col(width = 0.7) +
 facet_grid(input~spd)+
  labs(x = "Number of Windows",
       y = "Identified peptides",
       fill = "CV") +
  theme_bw() +
  theme(axis.text = element_text(size = 24),
        axis.title = element_text(size = 27),
        strip.text = element_text(size = 27),
        strip.background = element_rect(fill = "#E5E4F7"),
        axis.title.y = element_text(vjust = 2.2),
        axis.text.x = element_text(angle = 90, vjust = 0.3, hjust = 1),
        legend.title = element_text(size = 27),
        axis.title.x = element_text(vjust = 0.2),
        legend.text = element_text(size = 27),
        legend.key.size = unit(0.9, "cm"))+
  scale_fill_manual(values = c("grey82", "#FF9090", "#FF3030"),
                    labels = c("\U2265 15%", "10 - 15%", "< 10%" ))+
  scale_y_continuous(limits = c(0, 16000))+
  geom_errorbar(aes(ymin = y_errorbar - sd_id,
                          ymax = y_errorbar + sd_id),
                    width = 0.2,
                    size = 1)
```

Figure 2 was saved as a PDF.

```
cairo_pdf(filename = "fig2.pdf", width = 20.83, height = 14.58)
figure2
dev.off()

## png
##   2
```