DoE: Design of Experiment: RNA-seq Library Prep & Target Enrichment (TE) Panel

Patrick Cherry

2023-05-24

Contents

Introduction and Background

I had the idea that, given the success Panel A's bioinformatic performance, *it could be useful to show that TE panels work well for an extension of said bioinformatic performance.*

To do so, I propose using multiplexed capture, 100 ng (and optionally 10 ng of RNA input), and technical replicates (3, perhaps 2). These are parameterized in this DoE script below and the resulting sample plan is exported to google sheets.

Procedure

```
file_pref <- "2023_05_24_RNA_TE__sensitivty_DoE"
```

```
panel_info <- tribble(
  ~panel,  ~panel_size,  ~needed_sequencing,
  "TE Panel A", 3.0, NA,
  "TE Panel B",  36.8, NA,
  "TE Panel C",  35.8, NA,
  "Whole Transcriptome", NA, NA,
)
```

DoE with blocking for multiple operators

I will block for the operators carrying out library prep, because operator is a known source of variation that is not relevant to understanding the effect of TE panel, RNA input mass, or concentration on performance.

Blocking is the non-random assignment of samples to groups to minimize differences in the sample composition between the groups such that any effect of the grouping can be determined by the model and ignored (modeled out quantitatively and precisely).

```
rna_TE__sensitivity_doe$D;
```

```
## [1] 0.3789291
```

```
rna_TE__sensitivity_doe$diagonality
```

```
## [1] 0.871
```

Diagonality is the degree to which the blocked variables are uncorrelated: a diagonality of 1.0 is perfectly uncorrelated. A value of 0.871 is moderate. We are getting values less than 1.0, because not every number of unique sample (2 * 4 * ~~5~~) factors to be processed is divisible by the number of blocking groups. We will see this effect illustrated in the "Check orthogonality of blocking" section.

```
rna_TE__blocking_df <- bind_rows(
 mutate(rna_TE__sensitivity_doe$Blocks$B1, "operator" = "Operator A"),
 mutate(rna_TE__sensitivity_doe$Blocks$B2, "operator" = "Operator B"),
) %>%
  arrange(panel, conc, mass_input)
```

```
(panels_to_join <- rna_TE__blocking_df %>%
  distinct(panel) %>%
   bind_cols(rename(panel_info, "panel_name" = 1)))
```

panel	panel_name	panel_size	needed_sequencing
1	TE Panel A	3.0	NA
2	TE Panel B	36.8	NA
3	TE Panel C	35.8	NA
4	Whole Transcriptome	NA	NA

```
(fusconcs_to_join <- rna_TE__blocking_df %>%
  distinct(conc) %>%
  bind_cols(concentrations) %>%
  rename("concentrations" = 2))
```

conc	concentrations
-2	0.027
-1	0.0027
0	0.00027
1	2.7e-05
2	2.7e-06

```
(massinput_to_join <- rna_TE__blocking_df %>%
  distinct(mass_input) %>%
  bind_cols(mass_inputs) %>%
  rename("mass_inputs" = 2))
```

mass_input	mass_inputs
-1	10
1	100

```
rna_TE__doe_blocked <- rna_TE__blocking_df %>%
  left_join(panels_to_join, by = "panel") %>%
  left_join(fusconcs_to_join, by = "conc") %>%
  left_join(massinput_to_join, by = "mass_input") %>%
```

```
  select("panel" = "panel_name", "conc" = "concentrations",
         "mass_input" = "mass_inputs", "operator", panel_size, needed_sequencing) %>%
  arrange(panel, desc(conc), mass_input) %>%
  mutate("replicate_num" = row_number(), .by = c(panel, conc, mass_input)) %>%
  relocate(replicate_num, .after = operator) %>%
  rename("LP_operator" = "operator") %>%
  arrange(panel, mass_input) %>%
  mutate("capture" = ceiling(row_number()/6) ) %>%
  relocate("capture", .after = replicate_num) %>%
  arrange(panel, desc(conc), mass_input)
head(rna_TE__doe_blocked, n = 10)
```

panel	conc	mass_input	LP_operator	replicate_num	capture	panel_size	needed_sequencing
TE Panel A	0.027	10	Operator A	1	1	3	NA
TE Panel A	0.027	10	Operator B	2	1	3	NA
TE Panel A	0.027	10	Operator B	3	1	3	NA
TE Panel A	0.027	100	Operator B	1	3	3	NA
TE Panel A	0.027	100	Operator B	2	3	3	NA
TE Panel A	0.027	100	Operator B	3	3	3	NA
TE Panel A	0.0027	10	Operator A	1	1	3	NA
TE Panel A	0.0027	10	Operator A	2	1	3	NA
TE Panel A	0.0027	10	Operator A	3	1	3	NA
TE Panel A	0.0027	100	Operator A	1	4	3	NA

Check orthogonality of blocking

```
rna_TE__doe_blocked %>% count(LP_operator, mass_input)
```

LP_operator	mass_input	n
Operator A	10	30
Operator A	100	30
Operator B	10	30
Operator B	100	30

```
rna_TE__doe_blocked %>% count(LP_operator, panel)
```

LP_operator	panel	n
Operator A	TE Panel A	15
Operator A	TE Panel B	15

LP_operator	panel	n
Operator A	TE Panel C	15
Operator A	Whole Transcriptome	15
Operator B	TE Panel A	15
Operator B	TE Panel B	15
Operator B	TE Panel C	15
Operator B	Whole Transcriptome	15

```
rna_TE__doe_blocked %>% count(LP_operator, conc)
```

LP_operator	conc	n
Operator A	2.7e-06	13
Operator A	2.7e-05	12
Operator A	0.00027	9
Operator A	0.0027	14
Operator A	0.027	12
Operator B	2.7e-06	11
Operator B	2.7e-05	12
Operator B	0.00027	15
Operator B	0.0027	10
Operator B	0.027	12

```
googlesheets4::write_sheet(rna_TE__doe_blocked, ss = "sheet_string_goes_here",
                           sheet = "rna_TE__sensitivity_doe");
write_csv(rna_TE__doe_blocked, paste("doe_design_of_experiment_with_library_prep_files/",
                                     file_pref,
                                     "_rna_TE__sensitivity_doe.csv"))
```

Analyze captures

```
rna_TE__doe_blocked %>%
  count(capture, panel, mass_input)
```

capture	panel	mass_input	n
1	TE Panel A	10	6
2	TE Panel A	10	6
3	TE Panel A	10	3
3	TE Panel A	100	3
4	TE Panel A	100	6
5	TE Panel A	100	6
6	TE Panel B	10	6
7	TE Panel B	10	6
8	TE Panel B	10	3
8	TE Panel B	100	3
9	TE Panel B	100	6
10	TE Panel B	100	6
11	TE Panel C	10	6
12	TE Panel C	10	6
13	TE Panel C	10	3

capture	panel	mass_input	n
13	TE Panel C	100	3
14	TE Panel C	100	6
15	TE Panel C	100	6
16	Whole Transcriptome	10	6
17	Whole Transcriptome	10	6
18	Whole Transcriptome	10	3
18	Whole Transcriptome	100	3
19	Whole Transcriptome	100	6
20	Whole Transcriptome	100	6

Conclusion

Great! We have an experiment design for testing the effect of these panels against each other and against whole transcriptome sequencing. The experiment had n = 3 replicates, and it is blocked for having two operators carry out the RNA-seq library preps.

Importantly, when I get feedback on the design of this sample layout, I can easily show my work. Even better, if changes are needed, the entire design is programmed, and can be changed in seconds.

Let's go!