Joint Embedding - Pilot Analysis

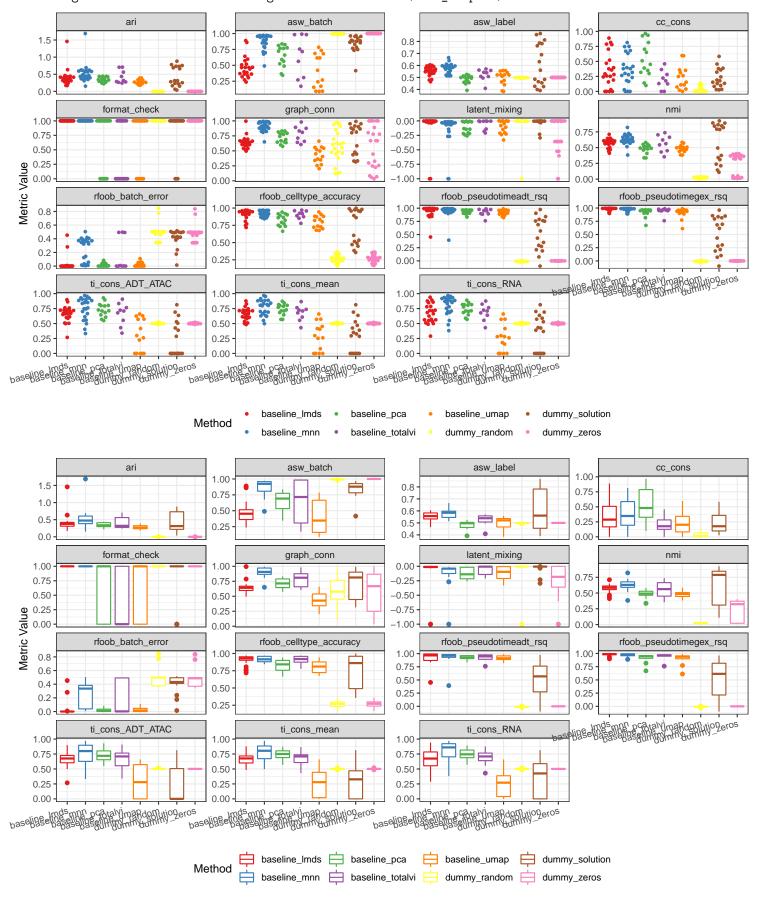
1

Visualise results

Colour by method.

Warning: Removed 22 rows containing missing values (position_quasirandom).

Warning: Removed 22 rows containing non-finite values (stat_boxplot).



Colour by dataset source.

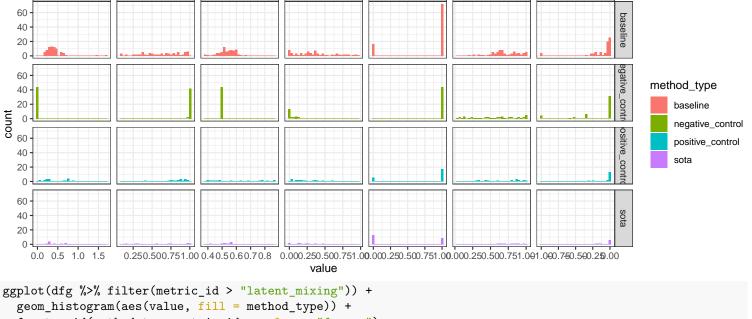
Warning: Removed 22 rows containing missing values (position_quasirandom).

Warning: Removed 22 rows containing non-finite values (stat_boxplot).



```
Comparing most interesting metrics.
# unique(dfq$metric_id)
patchwork::wrap_plots(
  ggplot(dfg %>% filter(metric_id %in% c("ari", "nmi", "asw_label", "ti_cons_mean", "cc_cons", "asw_batch", "gra
    geom_histogram(aes(value, fill = method_type)) +
    facet_grid(method_type~metric_id, scales = "free_x") +
    labs(title = "Metrics included in paper") +
    theme bw(),
  ggplot(dfg %>% filter(!metric_id %in% c("format_check", "ari", "nmi", "asw_label", "ti_cons_mean", "cc_cons",
    geom_histogram(aes(value, fill = method_type)) +
    facet_grid(method_type~metric_id, scales = "free_x") +
    labs(title = "Metrics not included in paper") +
    theme_bw(),
 ncol = 1
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
## Warning: Removed 22 rows containing non-finite values (stat_bin).
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
    Metrics included in paper
                   asw_batch
                                asw_label
                                                                                   ti_cons_mean
                                              cc cons
                                                          graph_conn
                                                                                                   method_type
                                                                                                      baseline
  和
count
                                                                                                      negative_control
  錮
                                                                                                      positive_control
                  0.0 0.5 1.0 1.5
                                              value
    Metrics not included in paper
        latent mixing
                         rfoob batch error
                                         rfoob_celltype_accuracy
                                                          rfoob_pseudotimeadt_rsq
                                                                            rfoob_pseudotimegex_rsq
                                                                                                   method_type
                                                                                            sel
                                                                                                      baseline
 #
                                                                                             _e
                                                                                                      negative_control
 #
                                                                                             (H)
                                                                                                      positive_control
                                                                                             őt
    -1.00-0.75-0.50-0.25 0.000.00 0.25 0.50
                                  0.75
                                          0.25
                                              0.50 0.75 1.00 0.00 0.25 0.50 0.75 1.00 0.00 0.25 0.50 0.75 1.00
ggplot(dfg %>% filter(metric_id <= "latent_mixing")) +</pre>
  geom_histogram(aes(value, fill = method_type)) +
  facet_grid(method_type~metric_id, scales = "free_x") +
  theme_bw()
```

- ## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
- ## Warning: Removed 22 rows containing non-finite values (stat_bin).



format_check

graph_conn

latent_mixing

```
facet_grid(method_type~metric_id, scales = "free_x") +
theme_bw()
```

'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.

asw_label

cc_cons

asw_batch

