

Top M Ranked Genes

Brian Jo Hsuan Lee

5/6/2022

Load the 3 lists of prioritized genes and the list of true genes

```
disnep = read_csv("/Users/beelee/Desktop/Columbia/Spring_2022/P8139-Statistical_Genetic_Modeling/Project/GW")
gw = read_csv("/Users/beelee/Desktop/Columbia/Spring_2022/P8139-Statistical_Genetic_Modeling/Project/GW")
control = read_csv("/Users/beelee/Desktop/Columbia/Spring_2022/P8139-Statistical_Genetic_Modeling/Project/GW")
true = read_csv("/Users/beelee/Desktop/Columbia/Spring_2022/P8139-Statistical_Genetic_Modeling/Project/GW")
```

Create and show a table of true gene counts for the 3 methods when m increases

```
top_m = matrix(rep(NA, 44), ncol = 11)

for(i in seq_len(11)){
  m = (i-1)*50
  d = length(intersect(disnep[1:m, ]$gene, true$PALB2))
  g = length(intersect(gw[1:m, ]$gene, true$PALB2))
  c = length(intersect(control[1:m, ]$gene, true$PALB2))
  top_m[1, i] = m
  top_m[2, i] = d
  top_m[3, i] = g
  top_m[4, i] = c
}

top_m = top_m %>% data.frame() %>% transpose()
rownames(top_m) = seq(0, 500, by = 50)
names(top_m) = c("m", "DiSNEP", "GeneWanderer", "Control")
knitr::kable(top_m)
```

	m	DiSNEP	GeneWanderer	Control
0	0	0	0	0
50	50	0	1	0
100	100	2	1	0
150	150	2	1	3
200	200	2	2	4
250	250	4	2	7
300	300	4	4	7
350	350	5	5	9
400	400	5	5	9
450	450	5	6	11
500	500	5	8	12

Visualize the curves

Saving 6.5 x 4.5 in image