Bootstrap Assessment of Age-Associated Enrichment

Derek Chiu, Steven McKinney

January 23, 2019

Introduction

To assess age-associated enrichment for intClust groups, we conducted a bootstrap analysis. We chose 50, 75, 100, 150, 200, and 300 bootstrap samples from each of the 10 intClust groups, and repeated this process 20 times. Each iteration repeat fitted a set of linear models predicting each gene target (48803 total) from age at diagnosis. An FDR of 0.01 was used for BH-adjusted p-values.

Tests of Independence

Because of a mix of small and large expected counts in several of the tables, we obtained p-values for Fisher's exact test using Monte Carlo simulation with B = 100000 replicates. For each `NiBoot` size, we obtained 20 p-values, one for each bootstrap replicate. The contingency tables are cross-tabulations between intClust group (1-10) and probes exhibiting age dependent status (yes/no).

Pearson Chi-Squared Tests

The distribution of p-values from Pearson's chi-square test for independence is shown below each bootstrap size table for the 20 replicates.

Fisher's Exact Tests

The distribution of p-values from Fisher's exact test for independence is shown below for the 20 replicates. Because the p-values were obtained using simulation with B = 100000 replicates, the minimum p-value will be 0.0000100

Raw data iClust enrichment test

Table 1: intClust number of cases

1	2	3	4	5	6	7	8	9	10
140	72	294	344	191	86	193	300	146	226

Table 2: Age associated gene counts (max(Abs(FC)) > 1.25) within intClust groups.

Group	max(Abs(FC)) > 1.25
intClust 1	8
intClust 2	0
intClust 3	878
intClust 4	536
intClust 5	4
intClust 6	0
intClust 7	12
intClust 8	636
intClust 9	1
intClust 10	0

Fisher's exact test $p < 0.0000100$

There is clear evidence of age associated enrichment within some of the intClust groups.

intClust groups 3, 4 and 8 appear to show the most age associated genes. However, those intClust groups also have the largest sample sizes, so detection of trends therein is more likely. Caution is warranted with regards to attributing age association tendencies to those intClust groups relative to others.

To assess the effect of sample size, we performed bootstrap analyses, selecting equally sized cohorts with replacement from within each intClust group, and counting the number of age-associated genes.

NB 50 Results

Table means

Table 3: Average trend counts across bootstrap replications for each intClust group of size 50

group	mean_trend	mean_no_trend
1	10.2	48793
2	349	48454
3	1.8	48801
4	0.15	48803
5	4.55	48798
6	174.2	48629
7	3.2	48800
8	1.5	48802
9	10.05	48793
10	0.25	48803
mean	55.49	48748

Table Medians

Table 4: Median trend counts across bootstrap replications for each intClust group of size 50

group	median_trend	median_no_trend
1	1.5	48802
2	57.5	48746
3	0	48803
4	0	48803
5	1	48802
6	12.5	48790
7	0.5	48802
8	0	48803
9	1	48802
10	0	48803
median	0.75	48802

Top Probe Counts

FC = 1.25

Table 5: Number of Probes chosen in 10, 15, 19, 20 times out of 20 bootstrap replicates of size 50 for FC 1.25

group	10	15	19	20
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0

group	10	15	19	20
4	0	0	0	0
5	0	0	0	0
6	0	0	0	0
7	0	0	0	0
8	0	0	0	0
9	0	0	0	0
10	0	0	0	0

FC = 2

Table 6: Number of Probes chosen in 10, 15, 19, 20 times out of 20 bootstrap replicates of size 50 for FC 2

group	10	15	19	20
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
6	0	0	0	0
7	0	0	0	0
8	0	0	0	0
9	0	0	0	0
10	0	0	0	0

FC = 4

Table 7: Number of Probes chosen in 10, 15, 19, 20 times out of 20 bootstrap replicates of size 50 for FC 4

group	10	15	19	20
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
6	0	0	0	0
7	0	0	0	0
8	0	0	0	0
9	0	0	0	0
10	0	0	0	0

Pearson Chi-Square Tests

```
## pvalue
##         0 6.15e-140 5.99e-102  7.43e-91  4.59e-63  1.19e-49  1.48e-45
##         9         1         1         1         1         1         1
##  3.51e-41  8.98e-36  1.09e-15  2.99e-15     5e-04
```

```
##              1           1           1           1           1
```

Fisher's Exact Tests

```
## pvalue
##   1e-05 0.00082
##      19       1
```

NB 75 Results

Table Means

Table 8: Average trend counts across bootstrap replications for each intClust group of size 75

group	mean_trend	mean_no_trend
1	158.9	48644
2	885.5	47918
3	19.9	48783
4	6.5	48796
5	44.9	48758
6	636.9	48166
7	4.6	48798
8	49.2	48754
9	120.7	48682
10	23.85	48779
mean	195.1	48608

Table Medians

Table 9: Median trend counts across bootstrap replications for each intClust group of size 75

group	median_trend	median_no_trend
1	77.5	48726
2	727	48076
3	2	48801
4	0	48803
5	3.5	48800
6	619	48184
7	1.5	48802
8	19	48784
9	2.5	48800
10	0	48803
median	3	48800

Top Probe Counts

FC = 1.25

Table 10: Number of Probes chosen in 10, 15, 19, 20 times out of 20 bootstrap replicates of size 75 for FC 1.25

group	10	15	19	20
1	0	0	0	0
2	25	1	0	0
3	0	0	0	0

group	10	15	19	20
4	0	0	0	0
5	0	0	0	0
6	30	1	0	0
7	0	0	0	0
8	0	0	0	0
9	0	0	0	0
10	0	0	0	0

FC = 2

Table 11: Number of Probes chosen in 10, 15, 19, 20 times out of 20 bootstrap replicates of size 75 for FC 2

group	10	15	19	20
1	0	0	0	0
2	11	1	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
6	21	1	0	0
7	0	0	0	0
8	0	0	0	0
9	0	0	0	0
10	0	0	0	0

FC = 4

Table 12: Number of Probes chosen in 10, 15, 19, 20 times out of 20 bootstrap replicates of size 75 for FC 4

group	10	15	19	20
1	0	0	0	0
2	6	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
6	3	1	0	0
7	0	0	0	0
8	0	0	0	0
9	0	0	0	0
10	0	0	0	0

Pearson Chi-Square Tests

```
## pvalue
##      0 2.03e-278 5.14e-125
##      18         1         1
```

Fisher's Exact Tests

```
## pvalue
## 1e-05
##    20
```

NB 100 Results

Table Means

Table 13: Average trend counts across bootstrap replications for each intClust group of size 100

group	mean_trend	mean_no_trend
1	486.1	48317
2	2308	46495
3	106.8	48696
4	11.25	48792
5	100.7	48702
6	1212	47590
7	219.2	48584
8	63.6	48739
9	232.3	48571
10	7	48796
mean	474.8	48328

Table Medians

Table 14: Median trend counts across bootstrap replications for each intClust group of size 100

group	median_trend	median_no_trend
1	282	48521
2	1898	46905
3	27.5	48776
4	1.5	48802
5	16	48787
6	1254	47549
7	80	48723
8	40.5	48762
9	31	48772
10	3	48800
median	35.75	48767

Top Probe Counts

FC = 1.25

Table 15: Number of Probes chosen in 10, 15, 19, 20 times out of 20 bootstrap replicates of size 100 for FC 1.25

group	10	15	19	20
1	20	0	0	0
2	477	91	17	6
3	2	1	0	0

group	10	15	19	20
4	0	0	0	0
5	0	0	0	0
6	155	16	2	0
7	0	0	0	0
8	5	0	0	0
9	0	0	0	0
10	0	0	0	0

FC = 2

Table 16: Number of Probes chosen in 10, 15, 19, 20 times out of 20 bootstrap replicates of size 100 for FC 2

group	10	15	19	20
1	13	0	0	0
2	201	41	14	5
3	2	1	0	0
4	0	0	0	0
5	0	0	0	0
6	79	6	0	0
7	0	0	0	0
8	4	0	0	0
9	0	0	0	0
10	0	0	0	0

FC = 4

Table 17: Number of Probes chosen in 10, 15, 19, 20 times out of 20 bootstrap replicates of size 100 for FC 4

group	10	15	19	20
1	3	0	0	0
2	57	16	4	0
3	1	1	0	0
4	0	0	0	0
5	0	0	0	0
6	7	1	0	0
7	0	0	0	0
8	2	0	0	0
9	0	0	0	0
10	0	0	0	0

Pearson Chi-Square Tests

```
## pvalue
##  0
## 20
```

Fisher's Exact Tests

```
## pvalue
## 1e-05
##    20
```

NB 150 Results

Table Means

Table 18: Average trend counts across bootstrap replications for each intClust group of size 150

group	mean_trend	mean_no_trend
1	1210	47593
2	5652	43151
3	502	48301
4	320.4	48483
5	358.5	48444
6	3255	45548
7	373.4	48430
8	266.4	48537
9	657	48146
10	132.8	48670
mean	1273	47530

Table Medians

Table 19: Median trend counts across bootstrap replications for each intClust group of size 150

group	median_trend	median_no_trend
1	1080	47722
2	5804	42999
3	452	48351
4	155	48648
5	355.5	48448
6	3327	45476
7	282.5	48520
8	193.5	48610
9	361.5	48442
10	55.5	48748
median	358.5	48444

Top Probe Counts

FC = 1.25

Table 20: Number of Probes chosen in 10, 15, 19, 20 times out of 20 bootstrap replicates of size 150 for FC 1.25

group	10	15	19	20
1	393	101	16	6
2	3327	1199	321	173
3	139	30	8	5

group	10	15	19	20
4	14	1	0	0
5	14	1	0	0
6	1348	396	81	40
7	39	2	1	0
8	63	20	6	1
9	12	1	0	0
10	0	0	0	0

FC = 2

Table 21: Number of Probes chosen in 10, 15, 19, 20 times out of 20 bootstrap replicates of size 150 for FC 2

group	10	15	19	20
1	176	56	11	5
2	1000	391	109	61
3	74	18	6	5
4	12	1	0	0
5	5	1	0	0
6	423	128	27	10
7	21	2	1	0
8	44	15	4	1
9	2	0	0	0
10	0	0	0	0

FC = 4

Table 22: Number of Probes chosen in 10, 15, 19, 20 times out of 20 bootstrap replicates of size 150 for FC 4

group	10	15	19	20
1	19	5	2	1
2	188	77	18	12
3	8	3	0	0
4	5	0	0	0
5	1	1	0	0
6	46	14	4	0
7	3	0	0	0
8	14	6	2	1
9	0	0	0	0
10	0	0	0	0

Pearson Chi-Square Tests

```
## pvalue
##  0
## 20
```

Fisher's Exact Tests

```
## pvalue
## 1e-05
##     20
```

NB 200 Results

Table Means

Table 23: Average trend counts across bootstrap replications for each intClust group of size 200

group	mean_trend	mean_no_trend
1	2211	46592
2	8462	40341
3	1051	47752
4	569.3	48234
5	765.3	48038
6	4538	44265
7	1037	47766
8	688.5	48114
9	2162	46641
10	352.1	48451
mean	2184	46619

Table Medians

Table 24: Median trend counts across bootstrap replications for each intClust group of size 200

group	median_trend	median_no_trend
1	2145	46658
2	8521	40282
3	913	47890
4	444.5	48358
5	687	48116
6	4568	44236
7	724.5	48078
8	640.5	48162
9	1320	47483
10	129	48674
median	818.8	47984

Top Probe Counts

FC = 1.25

Table 25: Number of Probes chosen in 10, 15, 19, 20 times out of 20 bootstrap replicates of size 200 for FC 1.25

group	10	15	19	20
1	1249	462	131	59
2	6665	3209	1174	632
3	484	185	49	23

group	10	15	19	20
4	146	16	0	0
5	106	15	2	0
6	3401	1655	583	338
7	218	59	14	5
8	264	97	30	16
9	418	75	5	1
10	3	0	0	0

FC = 2

Table 26: Number of Probes chosen in 10, 15, 19, 20 times out of 20 bootstrap replicates of size 200 for FC 2

group	10	15	19	20
1	362	153	48	26
2	1693	832	305	177
3	174	72	18	9
4	80	10	0	0
5	28	5	1	0
6	812	391	149	94
7	95	27	8	4
8	139	49	17	10
9	114	18	2	0
10	1	0	0	0

FC = 4

Table 27: Number of Probes chosen in 10, 15, 19, 20 times out of 20 bootstrap replicates of size 200 for FC 4

group	10	15	19	20
1	27	15	5	3
2	284	126	42	23
3	18	10	4	0
4	15	1	0	0
5	2	1	1	0
6	78	38	9	5
7	19	3	0	0
8	26	14	7	5
9	19	1	0	0
10	1	0	0	0

Pearson Chi-Square Tests

```
## pvalue
##  0
## 20
```

Fisher's Exact Tests

```
## pvalue
## 1e-05
##     20
```

NB 300 Results

Table Means

Table 28: Average trend counts across bootstrap replications for each intClust group of size 300

group	mean_trend	mean_no_trend
1	4615	44188
2	11910	36893
3	1943	46860
4	1218	47585
5	2620	46182
6	7937	40866
7	2238	46565
8	1892	46911
9	3611	45192
10	1349	47454
mean	3933	44870

Table Medians

Table 29: Median trend counts across bootstrap replications for each intClust group of size 300

| group	median_trend	median_no_trend
1	4669	44134
2	12014	36789
3	1826	46976
4	1186	47617
5	2518	46286
6	7808	40996
7	2255	46548
8	1694	47108
9	3380	45423
10	1224	47579
median	2386	46417

Top Probe Counts

FC = 1.25

Table 30: Number of Probes chosen in 10, 15, 19, 20 times out of 20 bootstrap replicates of size 300 for FC 1.25

group	10	15	19	20
1	3707	1927	806	518
2	11186	7299	3904	2692
3	1128	549	204	107

group	10	15	19	20
4	557	144	13	6
5	1244	389	75	37
6	7446	4549	2077	1389
7	1126	403	123	62
8	1140	560	208	141
9	2133	812	202	92
10	319	49	7	0

FC = 2

Table 31: Number of Probes chosen in 10, 15, 19, 20 times out of 20 bootstrap replicates of size 300 for FC 2

group	10	15	19	20
1	665	352	168	112
2	2392	1518	803	533
3	242	118	46	24
4	174	48	7	1
5	160	66	11	6
6	1202	685	283	190
7	240	104	38	22
8	308	173	78	50
9	435	145	36	14
10	37	4	0	0

FC = 4

Table 32: Number of Probes chosen in 10, 15, 19, 20 times out of 20 bootstrap replicates of size 300 for FC 4

group	10	15	19	20
1	50	28	11	7
2	369	228	111	77
3	24	13	4	2
4	16	6	1	0
5	12	4	1	1
6	89	43	20	9
7	29	8	3	2
8	51	26	13	8
9	60	17	4	2
10	2	1	0	0

Pearson Chi-Square Tests

```
## pvalue
##  0
## 20
```

Fisher's Exact Tests

```
## pvalue
## 1e-05
##     20
```

Boxplots

We can also visualize the results using boxplots. Each of the following boxplots shows the distribution of the proportion of age-associated gene targets within each intClust group across the 20 replicates. There is a figure for each of the `NiBoot` sizes chosen.
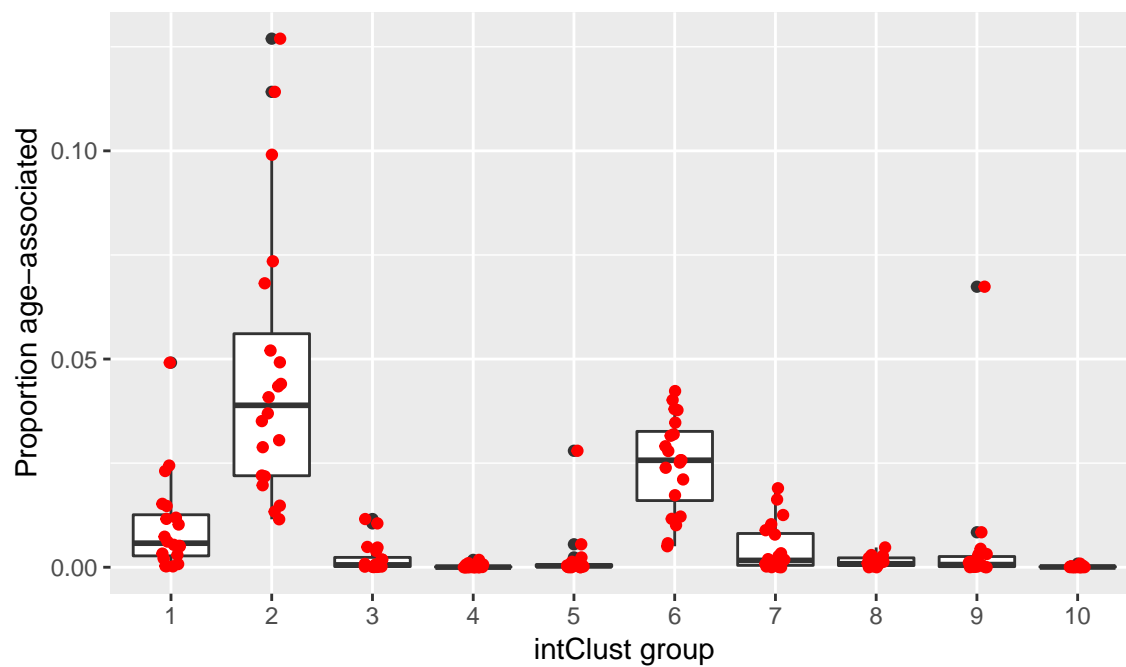
NiBoot = 50

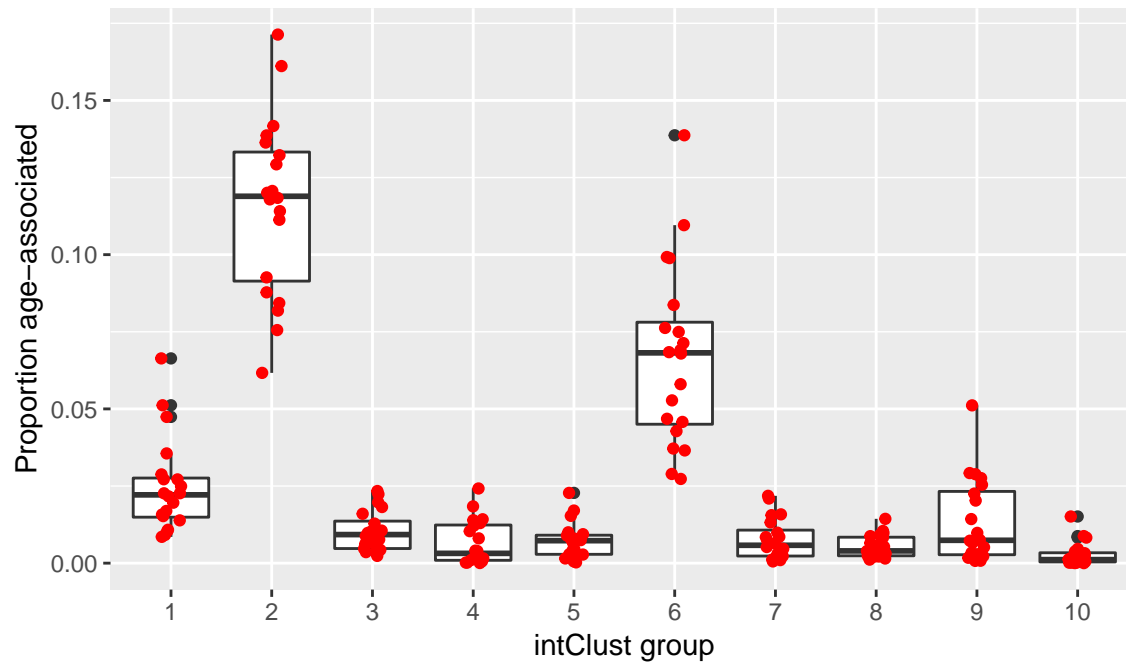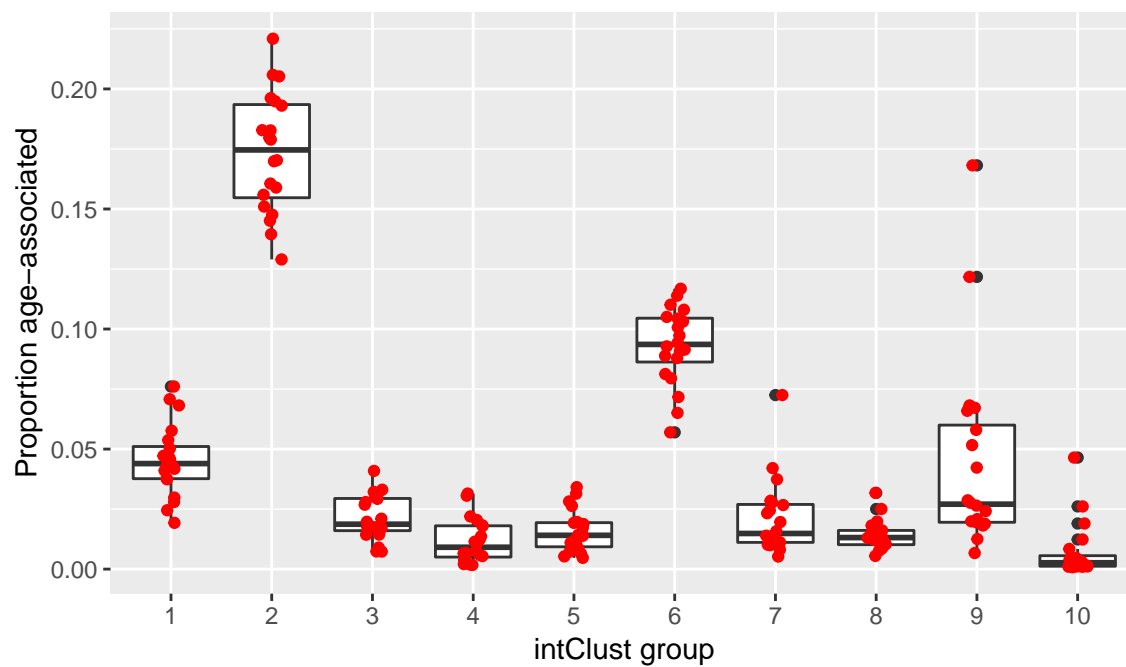Proportion of age−associated gene targets per intClust group on 20 replicates of bootstrapped samples of size 50
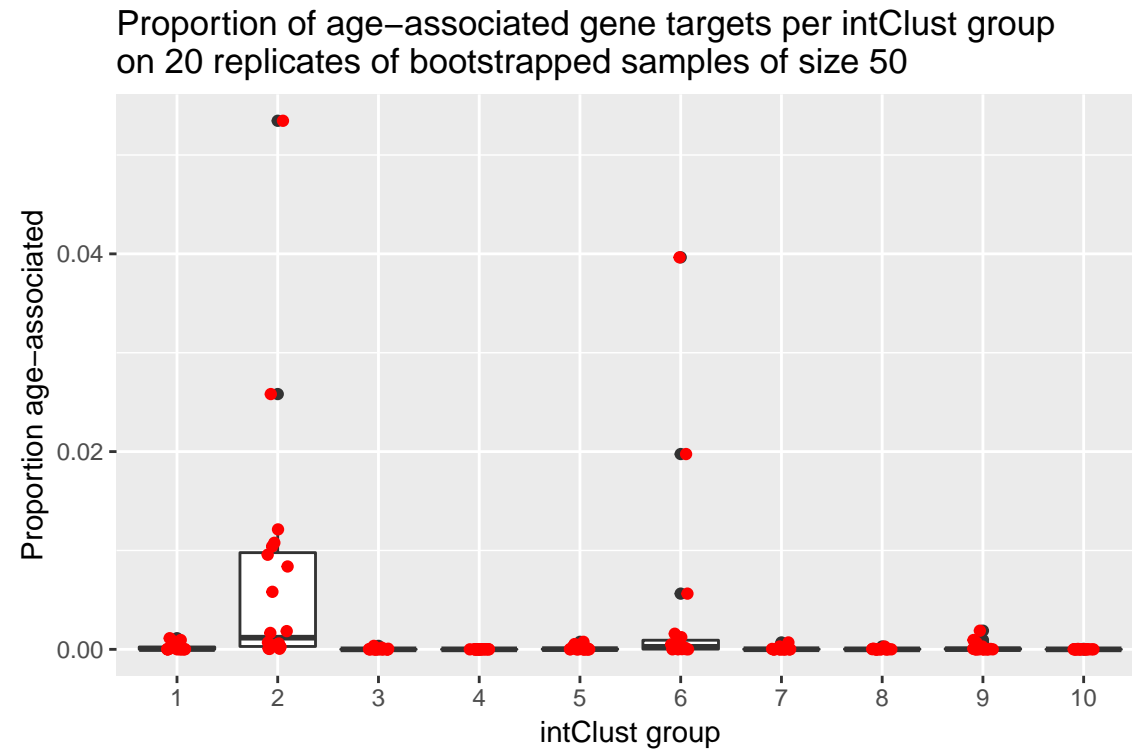
NiBoot = 75

Proportion of age−associated gene targets per intClust group
on 20 replicates of bootstrapped samples of size 75
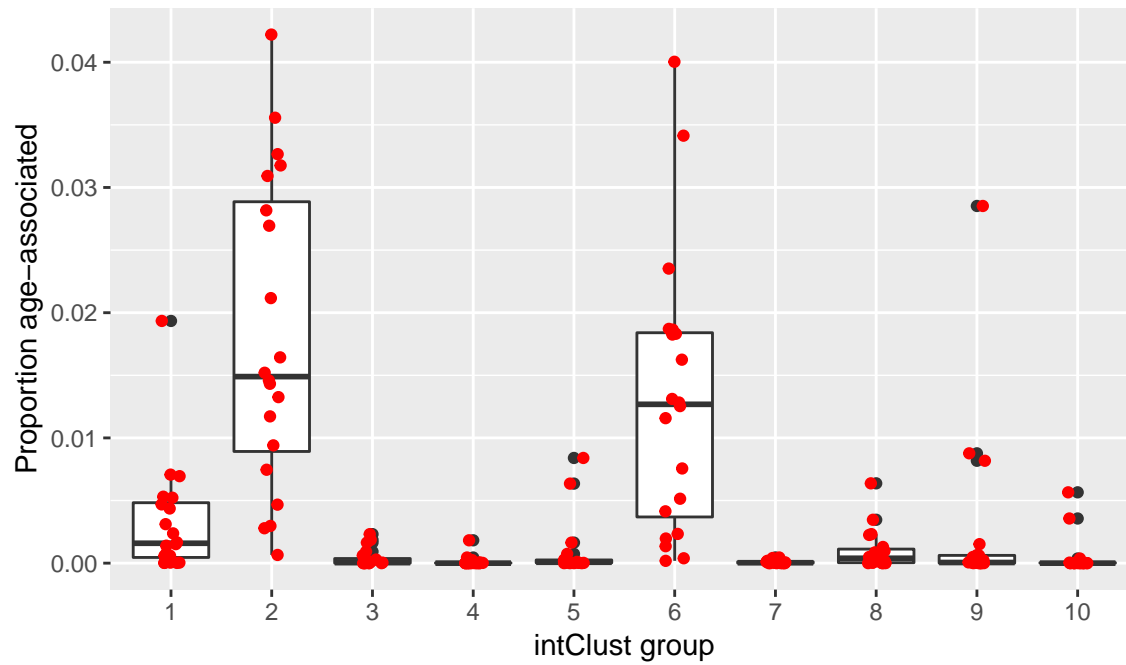
NiBoot = 100

Proportion of age−associated gene targets per intClust group
on 20 replicates of bootstrapped samples of size 100

Proportion of age−associated gene targets per intClust group
on 20 replicates of bootstrapped samples of size 150

Proportion of age−associated gene targets per intClust group
on 20 replicates of bootstrapped samples of size 200

Proportion of age−associated gene targets per intClust group
on 20 replicates of bootstrapped samples of size 300
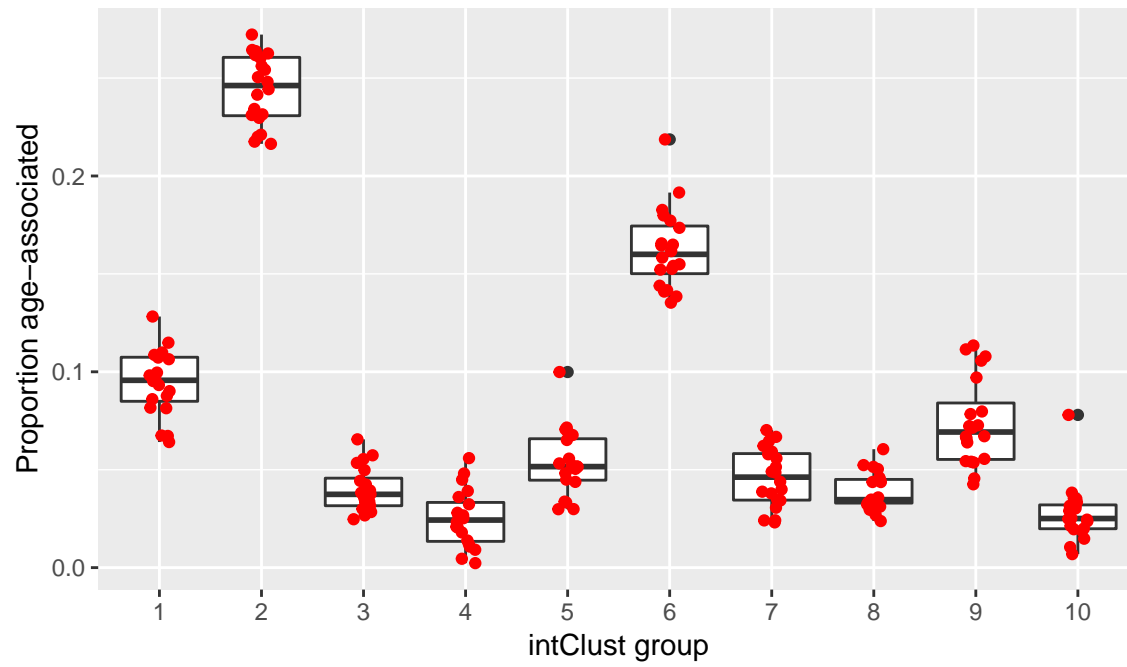
Conclusions

Keeping sample sizes equal across the intClust groups, intClust 2 and intClust 6 show the largest number of age associated genes followed by intClust 1. These are groups with smaller case counts, so will have larger variance associated with them in general. Recall that intClusts 3, 4 and 8 showed large numbers of age-associated genes in the original cohort.

intClust 10 shows few age associated genes even though it is one of the larger sample size groups, both in the original data and in the bootstrap runs. intClusts 5 and 7 exhibit similar behaviour. Hence for some of the intClust groups we can infer a lack of age associated behaviour.

Thus caution is warranted in interpreting reasons associated with the number of age-associated genes within intClust groups. Further assessment of this outcome will need to be performed on future data sets with more equal intClust group sizes.