Lab 11

```
library(readr)
X373531_SampleGenotypes_Homo_sapiens_Variation_Sample_rs8067378 <- read_csv("373531-SampleGenotypes-Homo_sapiens_Variati
```

```
Rows: 64 Columns: 5
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (5): Sample (Male/Female/Unknown), Genotype (forward strand), Population...

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
View(X373531_SampleGenotypes_Homo_sapiens_Variation_Sample_rs8067378)
```

```
gg_count <- sum(X373531_SampleGenotypes_Homo_sapiens_Variation_Sample_rs8067378$`Genotype (forward strand)` == 'G|G')

gg_count/64*100
```

```
[1] 14.0625
```

> Q13: Read this file into R and determine the sample size for each genotype and their corresponding median expression levels for each of these genotypes. Hint: The read.table(), summary() and boxplot() functions will likely be useful here. There is an example R script online to be used ONLY if you are struggling in vein. Note that you can find the medium value from saving the output of the boxplot() function to an R object and examining this object. There is also the medium() and summary() function that you can use to check your understanding.

```
data <- "
sample geno exp
1 HG00367 A/G 28.96038
2 NA20768 A/G 20.24449
3 HG00361 A/A 31.32628
4 HG00135 A/A 34.11169
```

```
5 NA18870 G/G 18.25141
6 NA11993 A/A 32.89721
7 HG00256 A/G 31.48736
8 NA18498 A/A 47.64556
9 HG00327 G/G 17.67473
10 HG00115 A/G 33.85374
11 NA20806 A/G 16.29854
12 HG00278 A/G 19.7345
13 NA20585 A/A 30.71355
14 NA19137 A/G 13.96175
15 HG00235 A/A 25.44983
16 NA20798 A/A 34.24915
17 NA12546 G/G 18.55622
18 NA19116 A/A 35.15014
19 HG00381 A/G 18.40351
20 NA18488 G/G 23.10383
21 HG00259 A/G 34.21985
22 HG00177 A/G 23.32404
23 NA19214 G/G 30.94554
24 NA19247 A/A 24.54684
25 NA19098 A/G 23.18606
26 NA20589 A/G 18.15997
27 NA19207 A/A 49.39612
28 HG00112 G/G 21.14387
29 NA20518 G/G 18.39547
30 HG00335 A/A 28.20755
31 NA19119 G/G 12.02809
32 HG00247 G/G 17.44761
33 NA12155 A/G 28.0358
34 NA20771 A/G 30.6527
35 NA20758 G/G 29.82254
36 HG00121 A/G 20.51327
37 NA20759 A/A 28.56199
38 NA20816 A/G 29.72309
39 NA20542 A/G 22.50789
40 NA18511 A/G 31.68959
41 NA12249 G/G 23.01983
42 NA11830 A/G 28.76435
```

```
43 NA19159 A/G 35.85543
44 NA20778 A/G 37.62403
45 NA18908 A/G 20.54885
46 HG00320 G/G 13.4247
47 NA11843 G/G 22.65437
48 HG00105 A/A 51.51787
49 NA20588 G/G 11.07445
50 NA20510 G/G 28.35841
51 NA12342 A/G 31.04941
52 HG00249 A/G 18.94583
53 NA11894 A/A 38.10956
54 HG00240 A/G 32.29483
55 HG00132 A/A 31.13741
56 HG00118 G/G 28.79371
57 NA18520 G/G 27.08956
58 NA18508 A/G 27.81775
59 HG00353 A/G 19.89903
60 NA20792 A/G 48.0341
61 NA12234 G/G 16.11138
62 HG00377 A/A 39.12999
63 NA19143 A/G 27.90313
64 NA20787 A/G 36.47949
65 NA20513 A/G 20.03116
66 HG00243 A/G 29.65063
67 NA19172 A/A 32.44173
68 NA06994 A/G 34.92257
69 NA18510 A/G 16.71385
70 HG00337 A/G 16.68151
71 NA20503 A/G 25.71008
72 NA19152 G/G 26.61928
73 NA20761 G/G 30.18323
74 NA19235 A/G 11.60808
75 HG00382 A/G 19.30953
76 NA20544 A/A 34.0326
77 NA18923 G/G 19.4079
78 HG00313 A/G 20.4904
79 HG00238 G/G 19.52301
80 NA20754 A/G 22.37224
```

```
81 NA11918 A/G 15.20045
82 NA18868 A/A 36.27151
83 NA06986 A/G 20.07459
84 HG00263 A/G 35.42982
85 NA12058 G/G 26.56808
86 NA20507 A/G 19.10884
87 NA12777 A/G 24.81087
88 NA12144 A/G 33.22193
89 HG00129 G/G 17.34076
90 HG00123 A/G 33.40835
91 NA12814 A/G 22.38996
92 HG00183 G/G 10.74263
93 HG00109 G/G 16.66051
94 NA20505 A/G 31.31626
95 NA12273 A/G 9.36055
96 HG00174 A/A 26.10355
97 HG00324 A/A 19.48106
98 HG00365 A/G 23.17937
99 NA20520 A/A 38.77623
100 NA19189 A/G 30.63079
101 HG00155 A/G 19.1042
102 HG00111 A/A 40.82922
103 NA12827 A/G 25.70962
104 NA18517 G/G 29.0172
105 NA20801 G/G 20.69333
106 NA20529 G/G 21.15677
107 NA18909 A/G 38.34531
108 HG00173 A/G 19.03976
109 HG00349 G/G 18.58691
110 HG00234 G/G 19.04962
111 NA19248 G/G 22.81974
112 NA20810 A/A 46.50527
113 HG00255 A/G 28.8177
114 NA12813 G/G 32.01142
115 NA20537 G/G 21.12823
116 NA18912 A/G 42.75662
117 HG00332 G/G 18.61268
118 HG00152 G/G 19.37093
```

```
119 NA20783 G/G 31.42162
120 NA12154 A/G 25.61662
121 HG00236 A/A 33.0732
122 NA19146 A/A 25.47283
123 HG00312 A/G 26.48467
124 HG00148 A/G 28.02486
125 HG00364 A/G 24.23377
126 HG00311 A/G 21.03717
127 NA11881 A/A 29.50655
128 HG00185 G/G 16.67764
129 NA20807 A/G 33.51752
130 NA19184 A/G 20.73493
131 HG00133 A/G 33.5565
132 NA20531 G/G 19.08659
133 NA19138 A/A 27.48438
134 NA19206 A/G 36.62034
135 HG00277 G/G 21.55001
136 NA18858 A/G 40.06318
137 HG00375 A/G 33.92744
138 HG00127 A/G 21.02084
139 NA19099 A/G 29.95687
140 HG00336 G/G 8.29591
141 HG00097 A/G 25.80393
142 HG00267 A/G 21.49924
143 NA20581 G/G 12.58869
144 NA12286 A/G 34.79575
145 NA20797 A/G 34.57705
146 NA12872 A/G 30.03549
147 HG00360 A/G 16.59638
148 NA20530 A/G 27.223
149 NA12348 A/G 24.35621
150 NA20538 G/G 17.34109
151 NA12760 A/G 22.86793
152 NA12763 A/G 23.19511
153 NA20814 G/G 28.23642
154 NA19222 A/A 35.69719
155 NA06989 A/A 32.42236
156 NA19171 G/G 19.99979
```

```
157 NA11829 A/G 33.74015
158 NA11992 A/G 24.08401
159 HG00141 G/G 25.55413
160 NA19150 A/G 26.39419
161 NA20828 A/G 32.33359
162 NA12749 A/A 28.91526
163 NA19190 G/G 24.45672
164 NA06985 A/G 11.36287
165 HG00178 A/G 21.16515
166 NA10851 G/G 23.53572
167 HG00371 A/A 19.14544
168 NA20541 A/G 17.21277
169 NA12004 A/A 22.85572
170 HG00116 G/G 22.48273
171 NA12272 G/G 14.66862
172 NA19096 G/G 33.95602
173 NA20800 A/G 22.73049
174 HG00102 A/A 31.17067
175 NA19236 G/G 18.26466
176 HG00264 A/G 25.57669
177 NA20521 A/A 27.87464
178 HG00345 G/G 16.06661
179 NA20509 A/A 27.9158
180 HG00329 A/A 16.8678
181 NA12830 A/G 11.9759
182 HG00359 A/A 23.66127
183 NA07051 A/G 25.35846
184 NA20516 A/G 33.32411
185 HG00128 A/G 22.09122
186 NA20534 A/G 25.19977
187 NA11892 A/A 28.03403
188 NA20804 A/A 36.51922
189 NA11994 A/G 30.83577
190 HG00156 G/G 17.32504
191 NA12843 A/G 23.63709
192 HG00180 A/G 19.66773
193 HG00282 G/G 19.14766
194 HG00343 G/G 12.57599
```

```
195 HG00139 G/G 22.28749
196 HG01789 A/G 24.6487
197 HG00321 A/G 17.03159
198 HG00306 A/A 27.43637
199 HG00232 G/G 17.29261
200 NA20528 A/G 22.27101
201 HG00122 G/G 24.18141
202 NA07037 A/A 35.63983
203 NA07056 A/G 15.92557
204 HG00151 A/G 32.5415
205 NA19129 A/A 38.85161
206 NA20517 A/G 22.40203
207 NA19149 G/G 16.07627
208 HG00341 A/G 27.41638
209 HG00274 A/G 31.99645
210 HG00106 A/G 30.05415
211 HG00189 G/G 14.80495
212 HG00252 A/G 20.01602
213 NA11832 A/G 34.47373
214 HG00323 A/A 22.44576
215 NA18916 A/A 37.06379
216 NA18867 A/G 28.75978
217 HG00100 A/A 35.67637
218 HG00126 G/G 23.46573
219 NA20813 A/G 29.91249
220 NA20504 A/G 15.71646
221 NA20532 A/G 21.7661
222 NA12812 A/G 9.62656
223 HG00244 A/G 28.53965
224 HG00265 G/G 28.97074
225 HG00378 G/G 27.78837
226 NA20790 A/A 50.16704
227 NA20512 A/A 37.94544
228 HG00268 A/A 29.15536
229 HG00380 A/A 28.85309
230 NA12761 A/A 38.57101
231 HG00384 A/G 29.49417
232 NA20796 G/G 23.92355
```

```
233 NA12399 G/G 9.55902
234 HG00310 A/G 29.5552
235 HG00096 A/A 30.89365
236 NA19147 A/G 19.44178
237 NA20752 A/G 21.43751
238 NA19107 A/G 30.40382
239 HG00099 G/G 12.35836
240 NA07048 A/A 39.31537
241 NA19114 G/G 22.5391
242 HG00376 A/A 31.43743
243 NA19092 A/A 35.26739
244 HG00130 A/G 28.50982
245 HG00158 A/A 22.37043
246 HG00269 A/A 28.46943
247 NA19210 G/G 21.98118
248 HG00258 A/A 30.15636
249 NA19256 A/G 21.48847
250 HG00276 G/G 16.40569
251 HG00331 A/G 31.10134
252 NA12751 A/G 35.99067
253 HG00181 G/G 25.21931
254 HG00346 G/G 24.32857
255 NA11920 A/G 26.42877
256 HG00326 A/G 26.28329
257 NA12347 A/A 35.88457
258 NA12716 A/G 20.72639
259 HG00142 G/G 19.42882
260 HG00309 A/G 21.0914
261 HG00315 G/G 26.56993
262 HG00338 A/G 23.79292
263 NA11995 A/A 32.59723
264 NA19209 A/A 36.02549
265 NA20540 A/A 23.86454
266 NA12890 A/A 28.38114
267 HG00250 G/G 13.34557
268 NA20769 G/G 16.60507
269 HG00138 A/A 25.14243
270 NA19200 A/A 51.3017
```

```
271 NA19144 G/G 24.85165
272 NA12815 G/G 21.56943
273 NA12043 A/G 18.79569
274 HG00350 A/G 29.54042
275 NA12383 A/A 28.14811
276 NA19201 A/G 18.787
277 HG00187 A/G 21.41071
278 NA06984 A/A 29.1839
279 NA20508 A/G 21.29782
280 NA19175 G/G 23.95528
281 NA20815 A/G 33.91853
282 NA12044 A/G 27.20808
283 NA18519 G/G 16.18962
284 NA20799 A/G 17.14895
285 NA20535 G/G 22.5372
286 NA19141 A/G 28.72738
287 HG00260 G/G 26.04123
288 HG00372 G/G 6.67482
289 NA07347 A/G 37.7384
290 NA07357 A/A 27.0976
291 NA20543 A/G 34.14567
292 HG00261 G/G 20.07363
293 HG00273 G/G 19.76527
294 NA12341 A/G 15.36874
295 HG00245 A/G 29.5035
296 NA19198 A/G 25.704
297 NA20757 A/G 20.07219
298 NA11930 A/A 33.89656
299 HG00358 G/G 18.50772
300 NA18933 A/G 24.53928
301 HG00242 A/G 17.84487
302 NA20773 A/G 23.35766
303 NA12282 A/G 15.71243
304 NA19131 A/A 33.48253
305 NA18499 A/A 15.43178
306 HG00117 A/A 29.45277
307 NA19121 G/G 20.14146
308 NA20515 G/G 18.07151
```

```
309 HG00355 A/G 19.89034
310 NA12775 A/G 25.37234
311 NA12005 A/G 16.12745
312 NA11893 A/G 24.18529
313 NA20808 A/G 21.97051
314 NA10847 G/G 6.9439
315 NA19102 A/G 13.08172
316 NA12400 G/G 22.14277
317 NA18487 A/G 32.00764
318 NA19093 A/G 30.59653
319 HG00342 G/G 14.23742
320 NA19160 A/G 29.74443
321 NA19095 A/G 27.88354
322 HG00160 A/A 26.80283
323 NA20766 A/G 11.12451
324 NA12717 A/G 7.07505
325 HG00125 A/G 23.13726
326 HG00171 A/G 21.09331
327 NA12873 A/G 8.20002
328 NA20525 A/G 20.62572
329 NA20826 A/G 18.24345
330 HG00136 G/G 19.85388
331 HG00272 A/G 11.13478
332 NA12340 A/A 43.51943
333 HG00251 A/G 24.43943
334 HG00369 A/G 22.24289
335 NA20803 A/G 24.67325
336 NA12842 A/G 41.03924
337 HG00146 A/A 45.80808
338 HG01790 A/G 33.31795
339 NA20809 A/G 27.98844
340 NA20765 G/G 27.73467
341 HG00362 A/A 26.55972
342 HG00114 A/G 31.57994
343 NA18917 A/A 24.8733
344 NA18502 G/G 19.02064
345 HG00150 A/G 36.73337
346 NA20527 A/A 29.99549
```

```
347 HG00179 A/G 18.45322
348 NA20805 A/A 26.68589
349 NA19117 A/G 23.60431
350 HG00285 A/G 24.33489
351 NA20772 G/G 14.49816
352 NA19213 A/G 35.74662
353 HG00344 A/G 22.75684
354 NA12156 A/A 39.37193
355 HG00257 G/G 26.7894
356 NA18486 G/G 20.84709
357 HG00188 G/G 10.77316
358 HG00366 A/G 34.42403
359 HG00157 A/A 38.39523
360 HG00262 A/A 41.23635
361 HG00280 G/G 12.82128
362 HG00308 G/G 16.90256
363 NA11831 A/G 25.34866
364 NA18910 G/G 29.60045
365 NA20795 A/G 25.06486
366 HG00231 A/G 36.78028
367 NA19197 A/G 30.67131
368 HG00101 A/A 27.13936
369 HG00281 G/G 14.81945
370 NA20760 A/A 36.55643
371 HG00176 A/A 28.34688
372 NA18489 A/G 37.8286
373 NA12275 G/G 17.46326
374 NA20514 A/A 15.42908
375 HG00351 G/G 23.26922
376 HG00186 G/G 21.39806
377 NA20586 A/G 25.44086
378 HG00275 G/G 18.0632
379 HG00325 G/G 15.91528
380 NA19118 G/G 24.80823
381 HG00124 G/G 26.04514
382 NA20785 A/A 47.50579
383 HG02215 G/G 18.28089
384 HG00253 A/A 30.15754
```

```
385 HG00134 G/G 23.24907
386 HG00339 A/A 34.88439
387 NA20519 A/G 29.49548
388 NA12778 A/G 23.27255
389 NA18861 A/A 29.29955
390 NA20539 A/A 32.87767
391 NA11931 G/G 17.91118
392 NA20812 A/G 28.69506
393 HG00120 G/G 21.09502
394 HG00103 A/G 26.52036
395 HG00328 A/G 27.49975
396 NA20774 A/G 24.66196
397 NA18873 A/G 25.81562
398 NA20502 A/G 22.49429
399 HG00143 A/G 26.88264
400 HG00145 A/A 43.43665
401 NA19225 A/A 26.5605
402 NA12829 A/G 28.982
403 HG00137 A/G 34.31875
404 NA20524 A/G 26.40231
405 HG00379 A/A 21.87746
406 NA18505 A/G 21.67621
407 HG01334 A/G 27.56805
408 NA18907 A/A 33.42582
409 NA19204 A/A 25.38406
410 NA12874 A/G 16.16277
411 NA20506 A/G 18.28963
412 NA20770 A/A 18.20442
413 NA12776 A/G 30.55183
414 NA18934 A/G 20.70871
415 NA19153 A/G 17.66476
416 HG00356 A/G 22.79543
417 NA12283 A/G 24.03419
418 HG00284 A/G 18.02351
419 NA12489 A/G 21.63102
420 HG00104 A/A 21.62336
421 NA20582 G/G 24.74366
422 NA11840 A/G 27.54976
```

```
423 HG00383 A/G 14.79717
424 NA20786 A/A 35.80093
425 NA20802 A/G 25.34921
426 NA20756 A/A 32.26844
427 NA19113 A/G 21.34916
428 NA12889 G/G 27.40521
429 NA12718 A/G 21.2008
430 HG00266 A/G 28.36006
431 NA12287 A/G 22.43773
432 HG00319 A/G 25.56306
433 NA12762 A/A 34.40756
434 HG00334 A/G 19.50634
435 NA12006 G/G 24.85772
436 NA19108 G/G 23.08482
437 NA19185 A/G 28.93651
438 HG00246 A/G 31.79897
439 NA12045 A/G 30.80067
440 NA19257 A/G 33.95134
441 NA12413 A/G 39.43243
442 HG00159 A/A 23.99631
443 NA20811 A/A 11.39643
444 HG00149 A/G 23.91465
445 NA19223 A/G 20.9756
446 NA07346 G/G 16.56929
447 NA20536 A/G 20.02507
448 HG01791 A/A 35.24632
449 HG00271 A/G 33.4417
450 HG00373 A/G 17.32813
451 HG00182 A/A 23.38376
452 HG00110 A/G 32.61856
453 NA20819 A/G 36.77906
454 HG00154 G/G 16.69044
455 HG00330 A/G 16.84776
456 NA12750 A/A 34.94395
457 HG00233 G/G 25.0888
458 HG00131 G/G 32.78519
459 HG00108 A/A 31.92036
460 HG00119 A/G 31.53069
```

```
461 NA19130 A/A 44.27738
462 HG00239 A/G 23.1825
"

# Convert to data frame
df <- read.table(text = data, header = TRUE)

# Check the data frame
head(df)
```

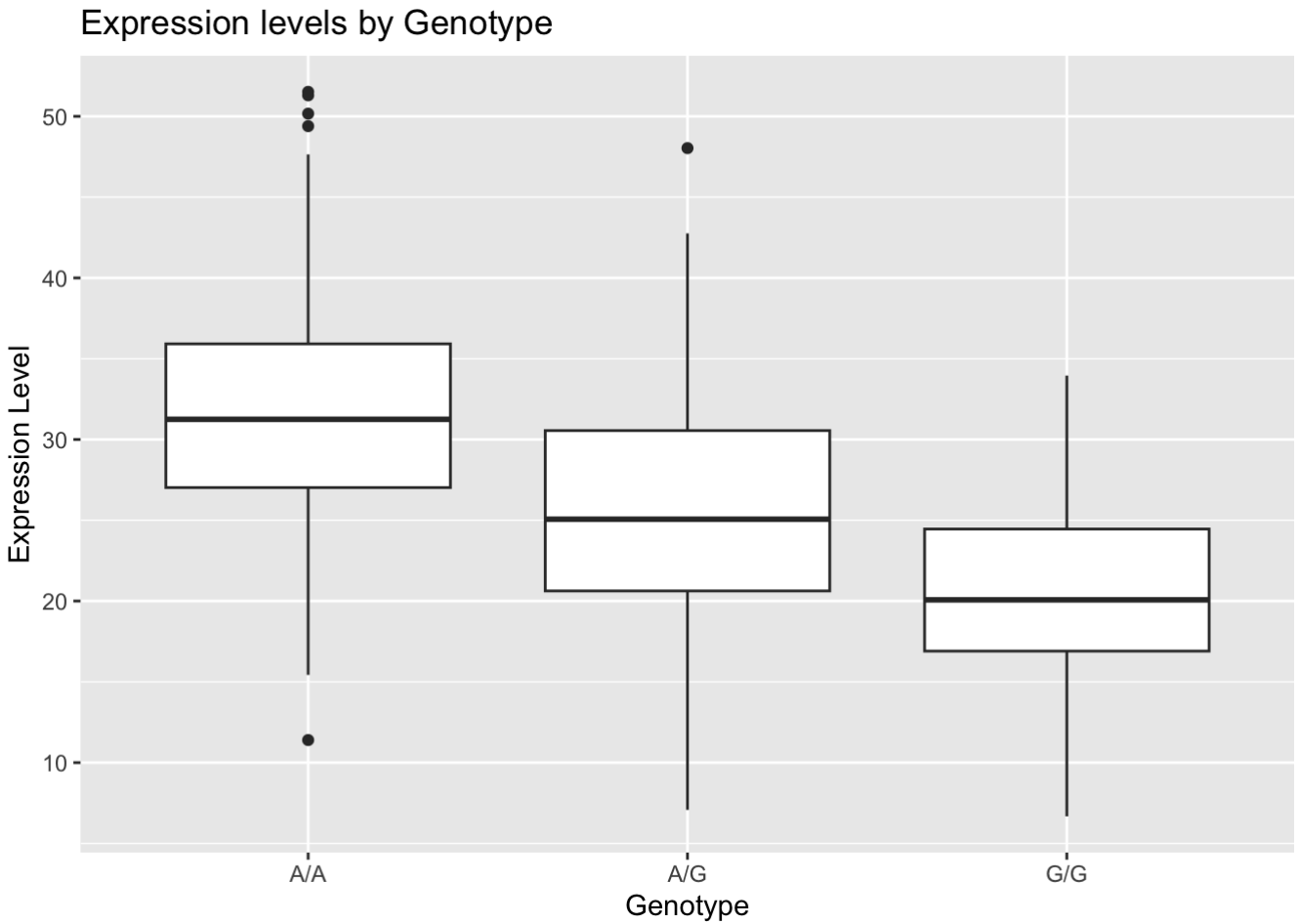
```
   sample geno      exp
1 HG00367  A/G 28.96038
2 NA20768  A/G 20.24449
3 HG00361  A/A 31.32628
4 HG00135  A/A 34.11169
5 NA18870  G/G 18.25141
6 NA11993  A/A 32.89721
```

```
# Calculate median expression levels for each genotype
medians <- aggregate(df$exp ~ df$geno, FUN = median)
colnames(medians) <- c('geno', 'exp')

# Calculate the sample size for each genotype
sample_sizes <- table(df$geno)
sample_sizes <- as.data.frame(sample_sizes)
colnames(sample_sizes) <- c('geno', 'sample_size')

# Join median expression levels and sample size data frames
result <- merge(medians, sample_sizes, by = 'geno')

# Print the result
print(result)
```

```
  geno      exp sample_size
1  A/A 31.24847         108
2  A/G 25.06486         233
3  G/G 20.07363         121
```

> Q14: Generate a boxplot with a box per genotype, what could you infer from the relative expression value between A/A and G/G displayed in this plot? Does the SNP effect the expression of ORMDL3? Hint: An example boxplot is provided overleaf – yours does not need to be as polished as this one.

```
# Load ggplot2
library(ggplot2)

# Generate a boxplot
ggplot(df, aes(x = geno, y = exp)) +
  geom_boxplot() +
  labs(title = "Expression levels by Genotype", x = "Genotype", y = "Expression Level")
```



Yes, the median (the horizontal line in the box) of the A/A genotype is significantly higher than the G/G genotype, inferring that the A/A genotype leads to higher expression of ORMDL3 compared to the G/G genotype