Supplemental Material

Uncovering transposable element variants and their potential adaptive impact in urban populations of the malaria vector *Anopheles coluzzii*

Carlos Vargas-Chavez[1], Neil Michel Longo Pendy[2,3], Sandrine E. Nsango[4], Laura Aguilera[1], Diego Ayala[2,5]*, Josefa González[1]*

[1]Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona, Spain.

[2]CIRMF, Franceville, Gabon.

[3] Ecole doctorale en infectiologie tropicale (EDR), Franceville, Gabon.

[4] Faculté de Médecine et des Sciences Pharmaceutiques, Université de Douala, BP 2701, Douala, Cameroun

[5]MIVEGEC, Univ Montpellier, CNRS, IRD, Montpellier, France

Emails: carlos.vargas@ibe.upf-csic.es, longo2michel@gmail.com, nsango2013@yahoo.fr, mlaura.aguilera@gmail.com, diego.ayala@ird.fr, josefa.gonzalez@csic.es

*Corresponding authors

Contents

Supplemental Figures

Figure S1. Number of TE copies identified when using the TE libraries from an increasing number of genomes. Number of TE copies identified when using the TE library of a single genome or when using all possible combinations of more than one genome.

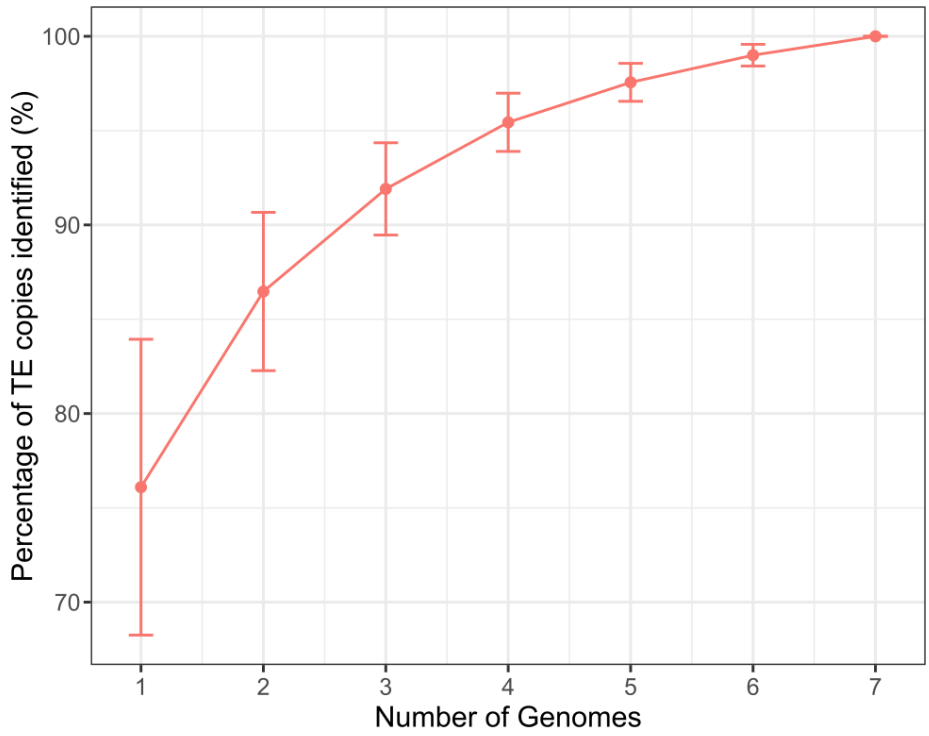
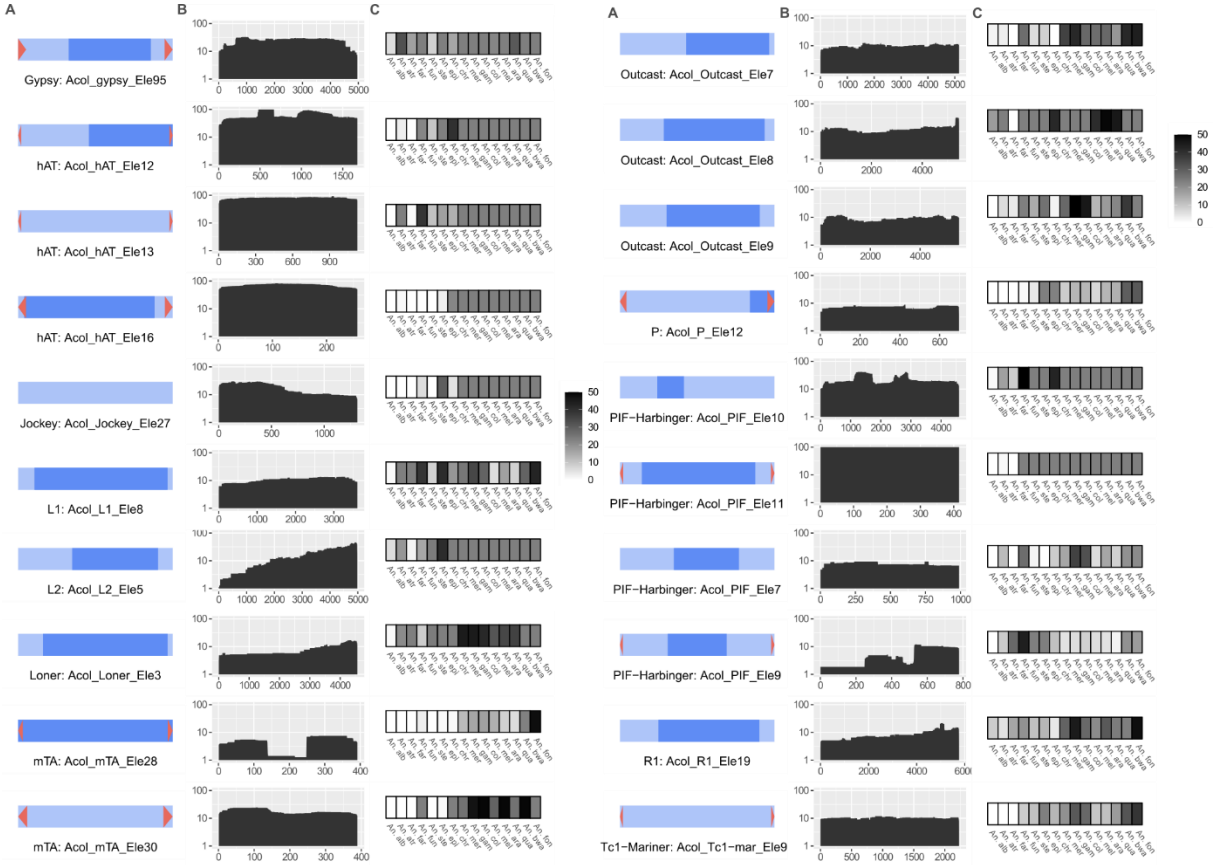


Figure S2. Novel TE families. A) The structure of each new family is displayed: the light blue box represents the full extension of the TE and the red arrows represent LTRs. B) All insertions for each TE family are shown as a coverage plot where each line represents a copy in a genome. C) Phylogenetic distribution of the TE family insertions in 15 members of the *Anopheles* genus, *Culex quinquefasciatus*, *Aede aegypti* and *D. melanogaster*. The number of insertions with more than 80% identity and spanning at least 80% of the consensus sequence, in each species is shown using a black and white gradient: species with no insertions are shown in white while species with 50 or more insertions are shown in black.



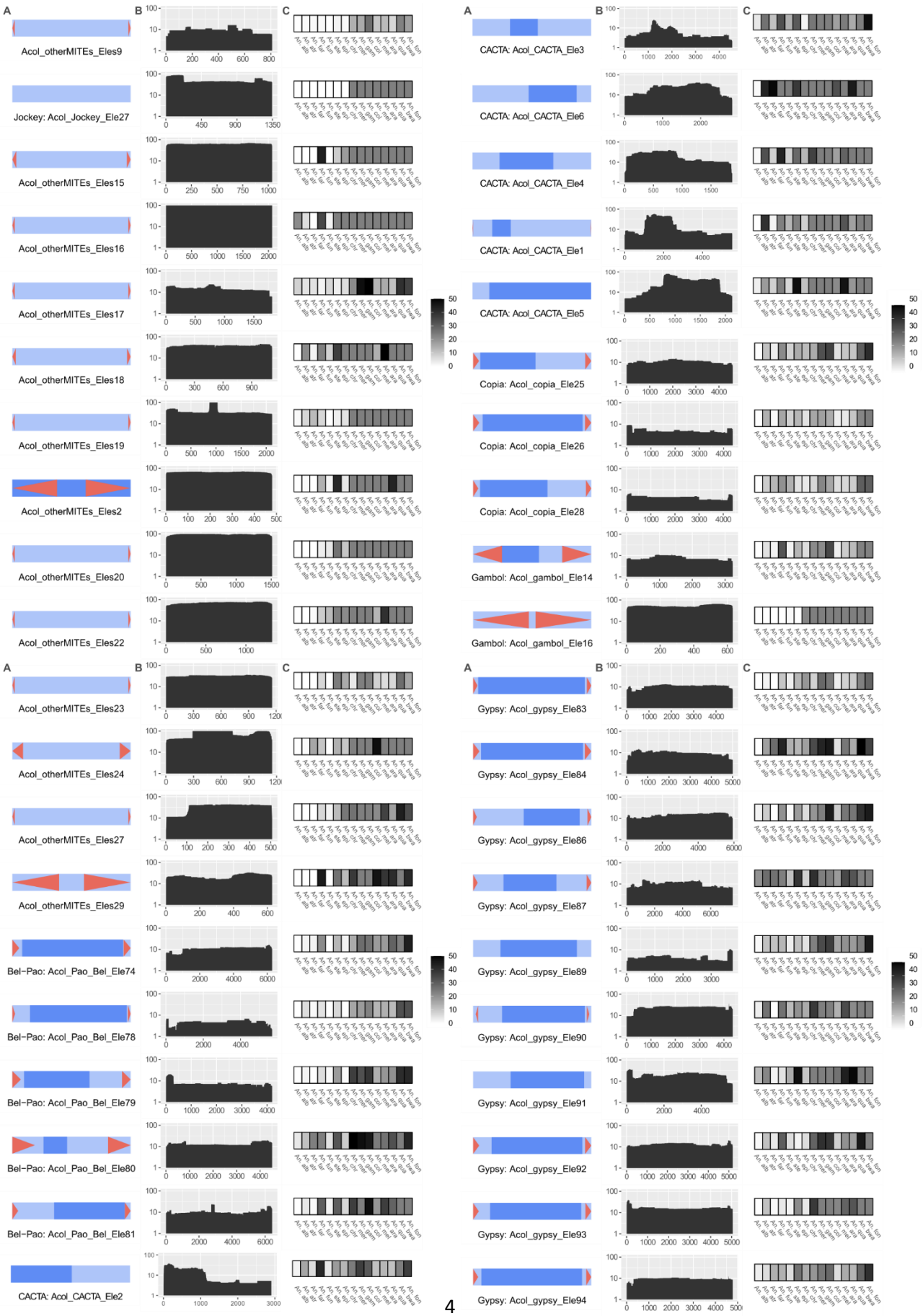

A
B
C
Acol_otherMITEs_Eles9
Jockey: Acol_Jockey_Ele27
Acol_otherMITEs_Eles15
Acol_otherMITEs_Eles16
Acol_otherMITEs_Eles17
Acol_otherMITEs_Eles18
Acol_otherMITEs_Eles19
Acol_otherMITEs_Eles2
Acol_otherMITEs_Eles20
Acol_otherMITEs_Eles22
CACTA: Acol_CACTA_Ele3
CACTA: Acol_CACTA_Ele6
CACTA: Acol_CACTA_Ele4
CACTA: Acol_CACTA_Ele1
CACTA: Acol_CACTA_Ele5
Copia: Acol_copia_Ele25
Copia: Acol_copia_Ele26
Copia: Acol_copia_Ele28
Gambol: Acol_gambol_Ele14
Gambol: Acol_gambol_Ele16
Acol_otherMITEs_Eles23
Acol_otherMITEs_Eles24
Acol_otherMITEs_Eles27
Acol_otherMITEs_Eles29
Bel−Pao: Acol_Pao_Bel_Ele74
Bel−Pao: Acol_Pao_Bel_Ele78
Bel−Pao: Acol_Pao_Bel_Ele79
Bel−Pao: Acol_Pao_Bel_Ele80
Bel−Pao: Acol_Pao_Bel_Ele81
CACTA: Acol_CACTA_Ele2
Gypsy: Acol_gypsy_Ele83
Gypsy: Acol_gypsy_Ele84
Gypsy: Acol_gypsy_Ele86
Gypsy: Acol_gypsy_Ele87
Gypsy: Acol_gypsy_Ele89
Gypsy: Acol_gypsy_Ele90
Gypsy: Acol_gypsy_Ele91
Gypsy: Acol_gypsy_Ele92
Gypsy: Acol_gypsy_Ele93
Gypsy: Acol_gypsy_Ele94

Figure S3. Number of TE insertions *vs*. genome size. Comparison of the bases spanned by TEs in each genome with their full genome sizes.

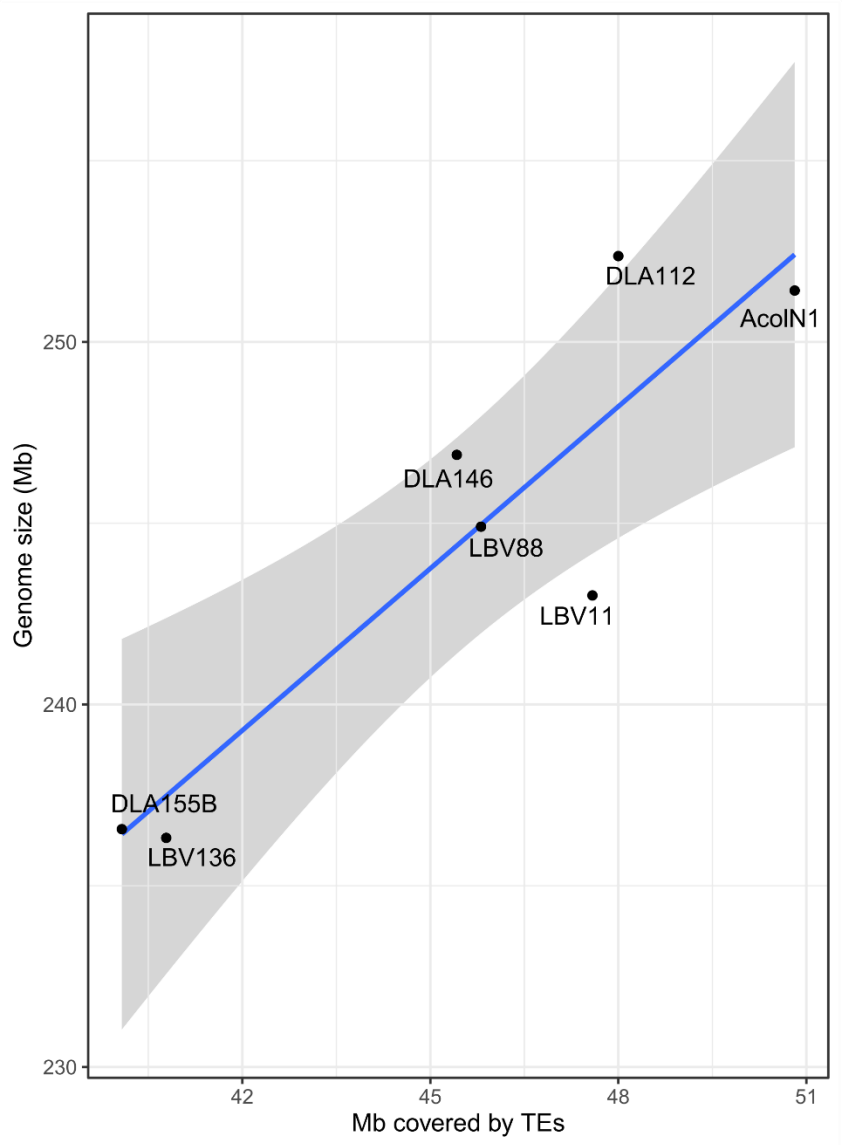
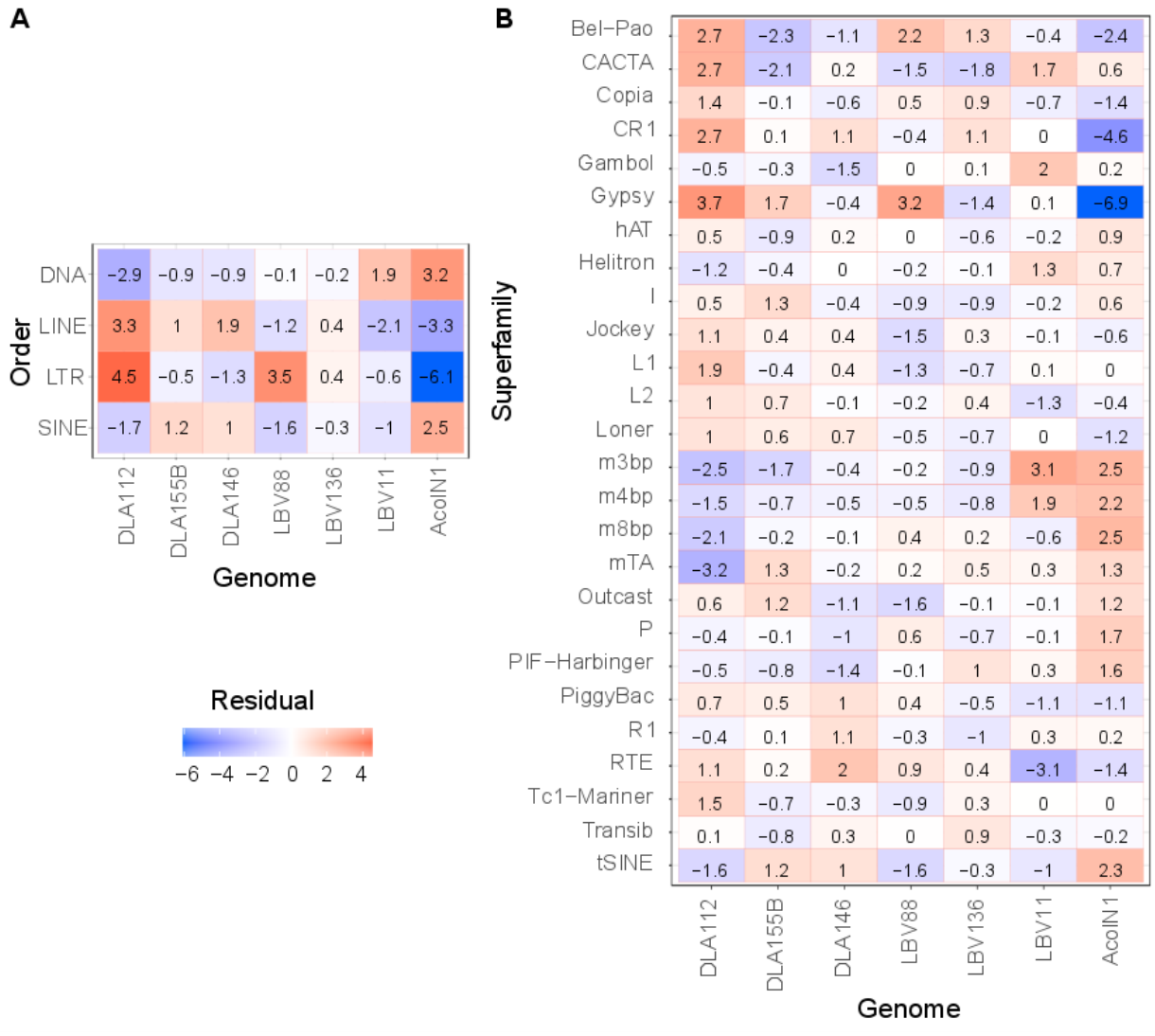


Figure S4. Differences in TE content between the seven *An. coluzzii* genomes.

Differences are shown at the (A) order and (B) superfamily levels. $\chi 2$ tests were performed for the number of insertions and the Pearson's residuals are shown. Note that MITEs are divided into the *m3bp, m4bp, m8bp* and *mTA* superfamilies. (C) PCA using the SNPs identified using the Illumina data selecting (i) those that were found on introns (and thus most likely to be neutral and to reflect the population structure) and (ii) those located on chromosome 3, as this chromosome is free from polymorphic inversions and regions of reduced recombination (Miles et al 2017, Science)



C

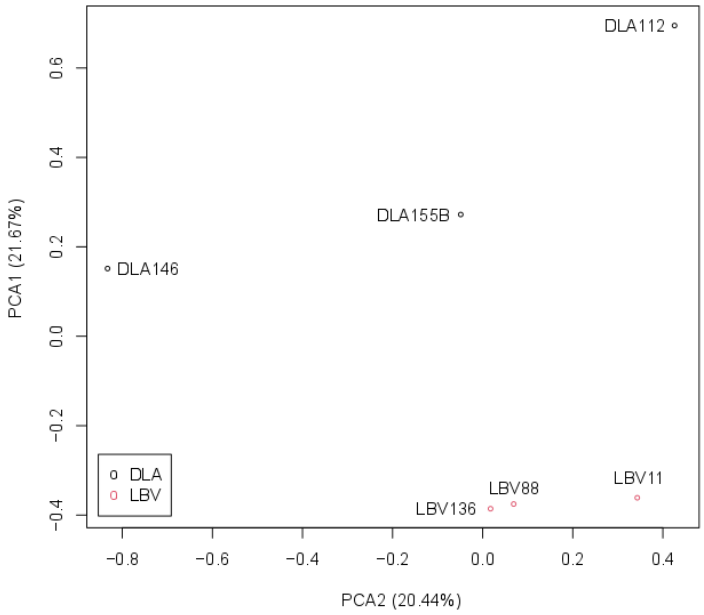

DLA112
DLA155B
DLA146
LBV88
LBV11
LBV136
DLA
LBV
PCA1 (21.67%)
PCA2 (20.44%)
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.4

Figure S5. TE landscapes. TE landscapes for the six genomes sequenced in this work generated using dnaPipeTE.

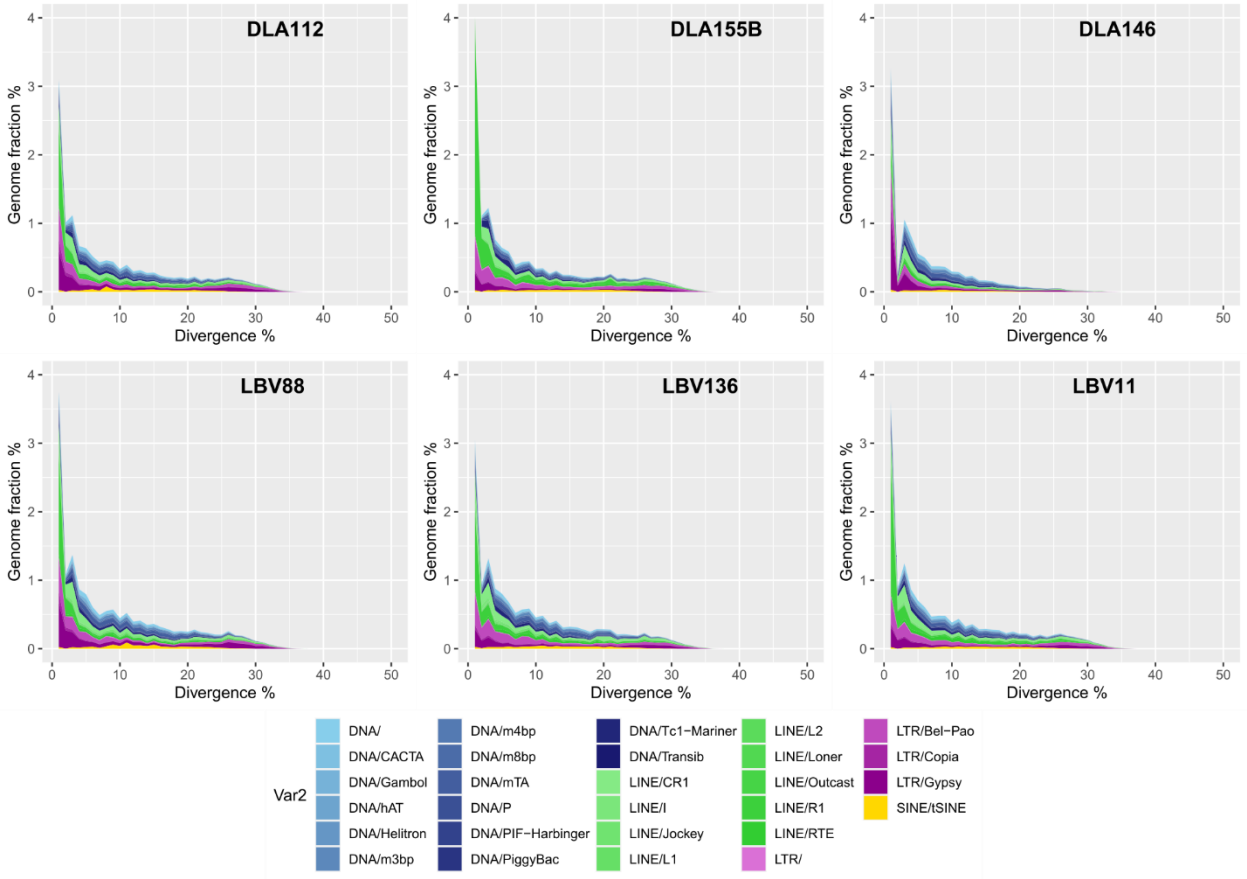
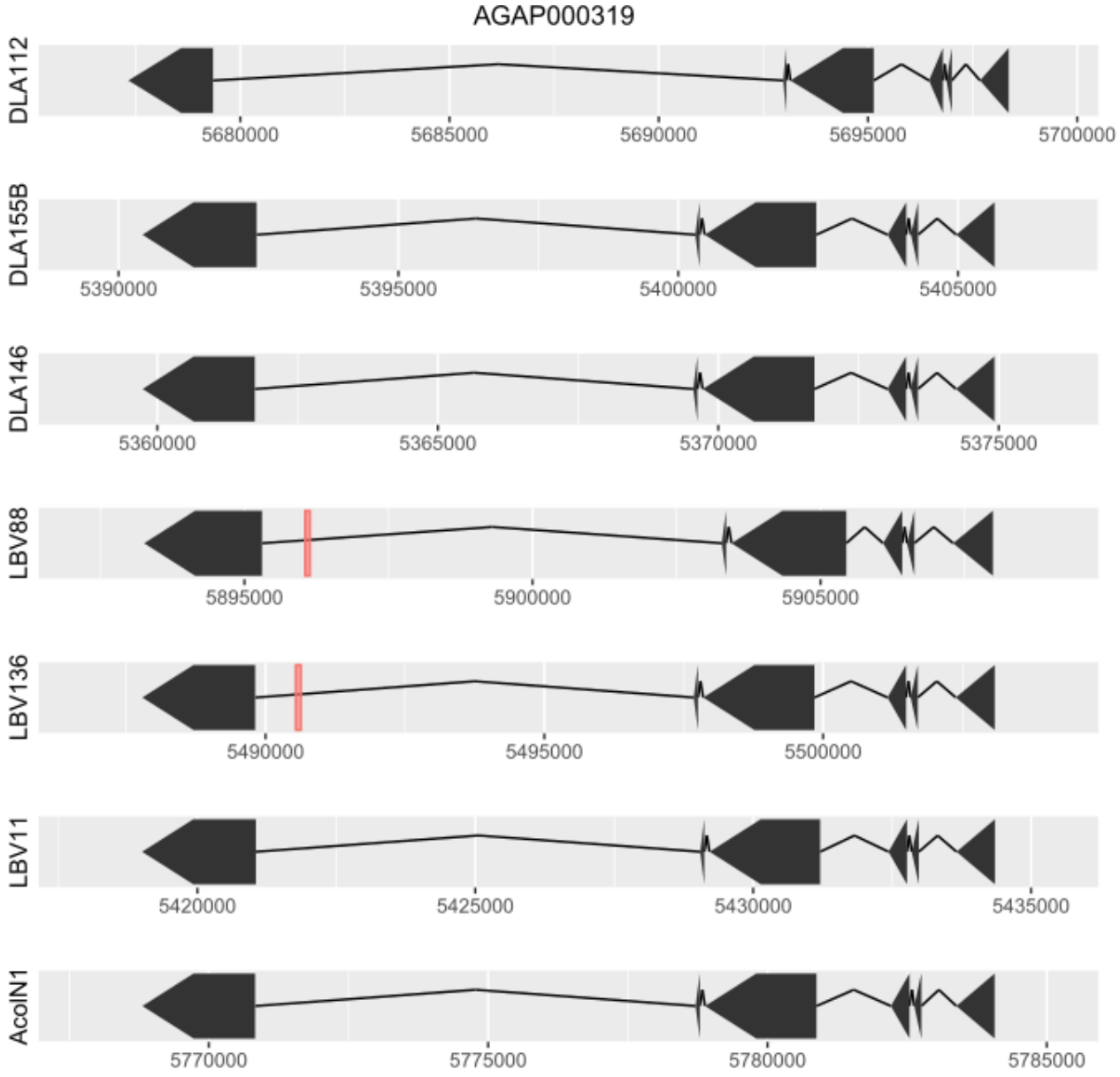
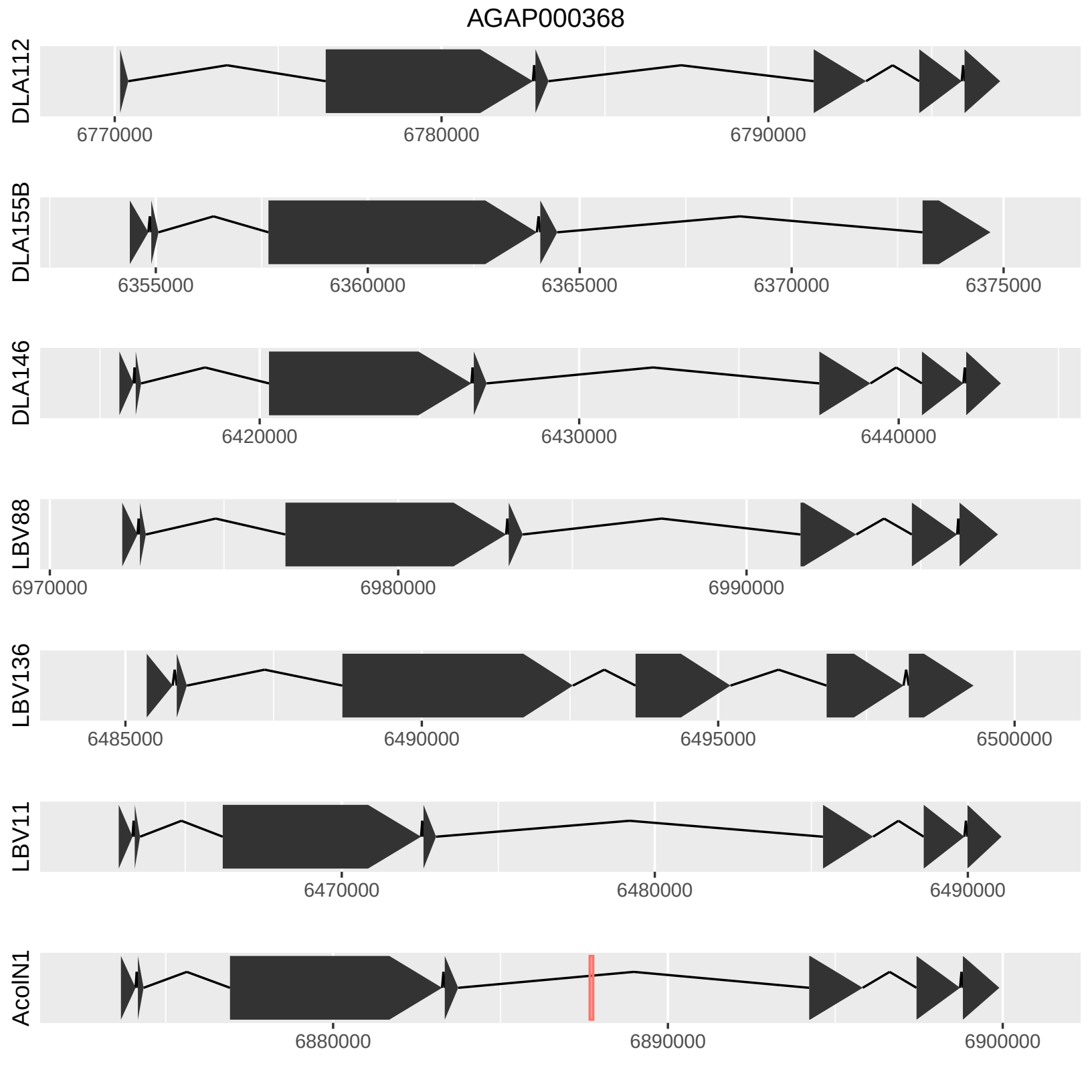
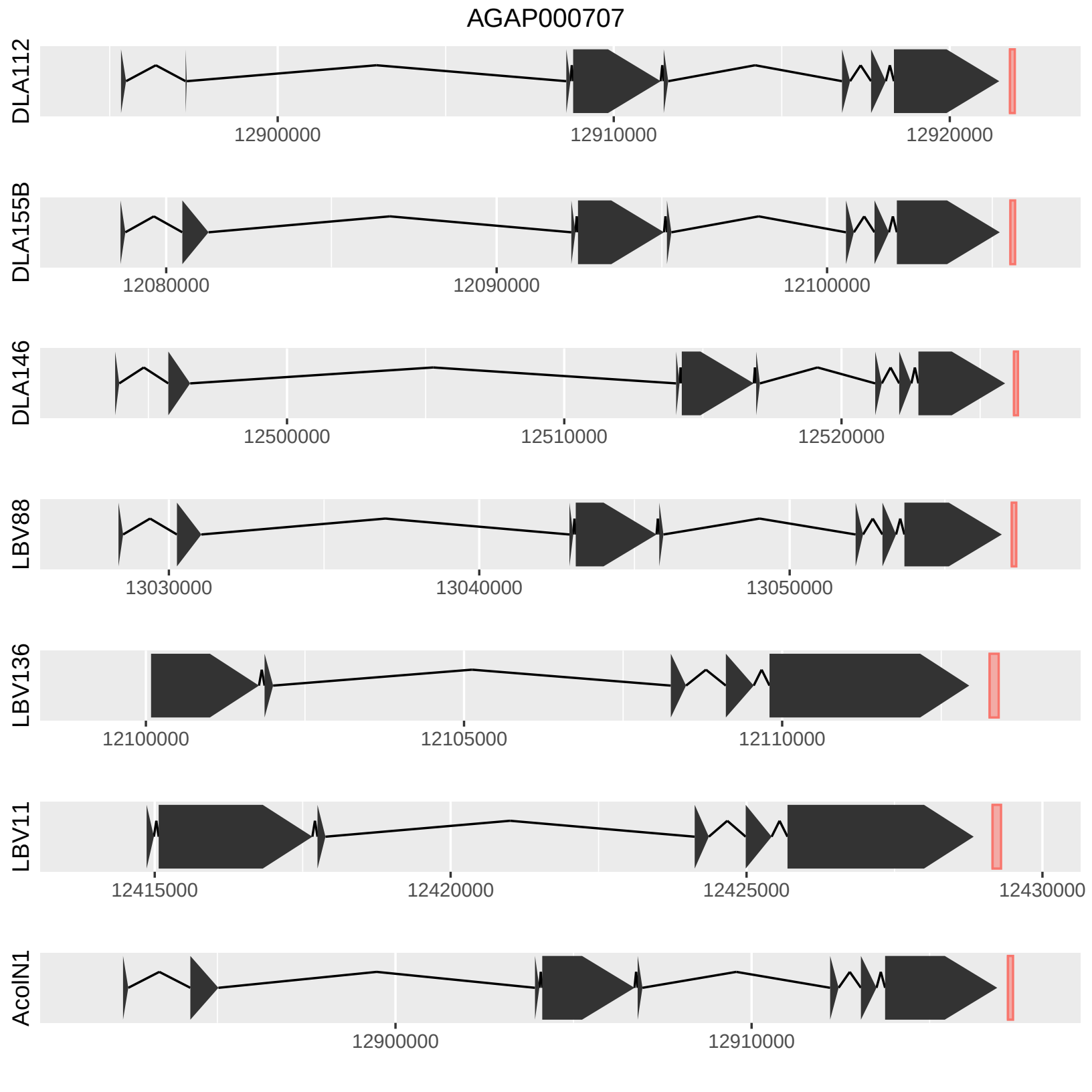


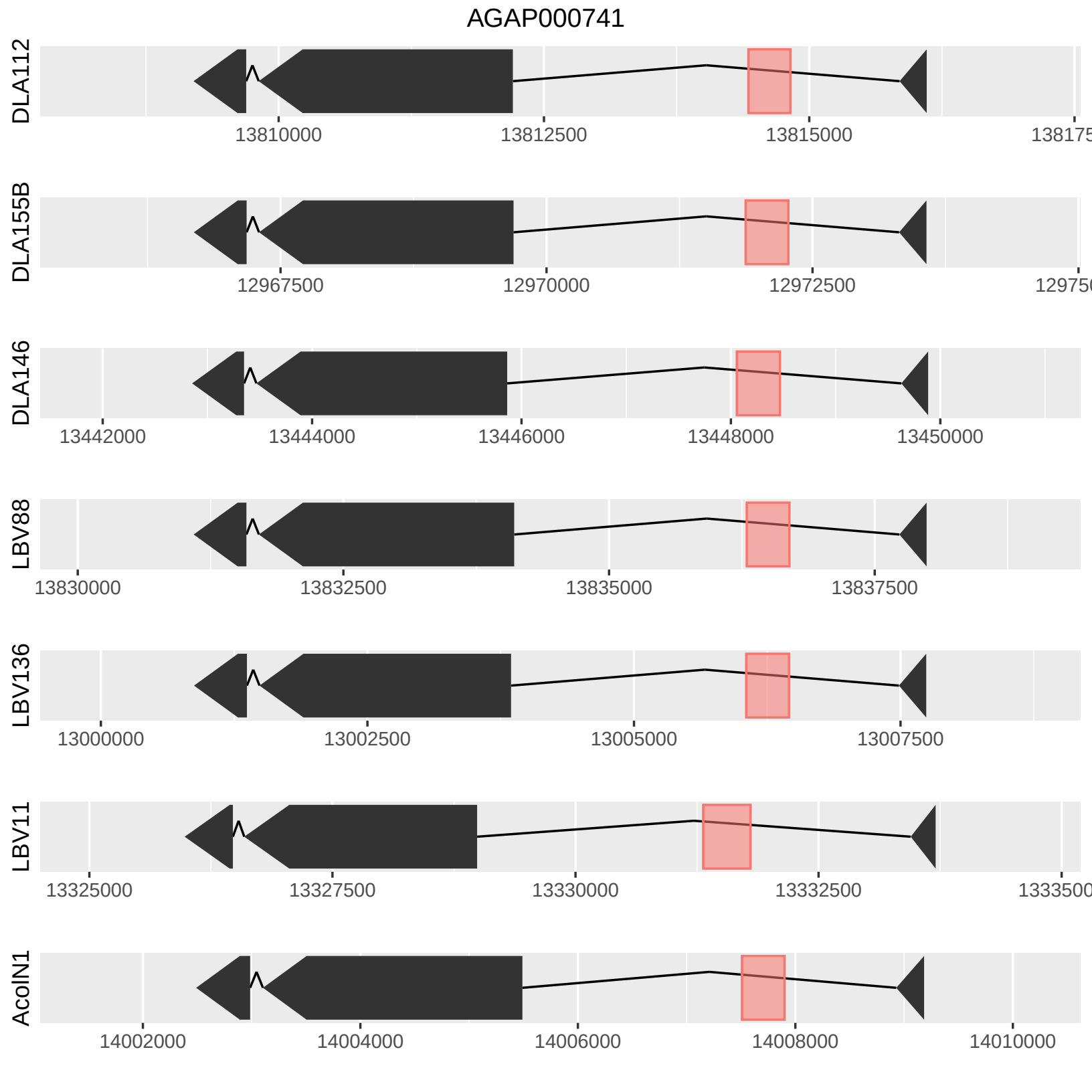
Figure S6. Genes with TE insertions from active families. Diagrams of TE insertions located inside genes or less than 1 kb from a gene showing the gene structure and the TE insertion (in red).

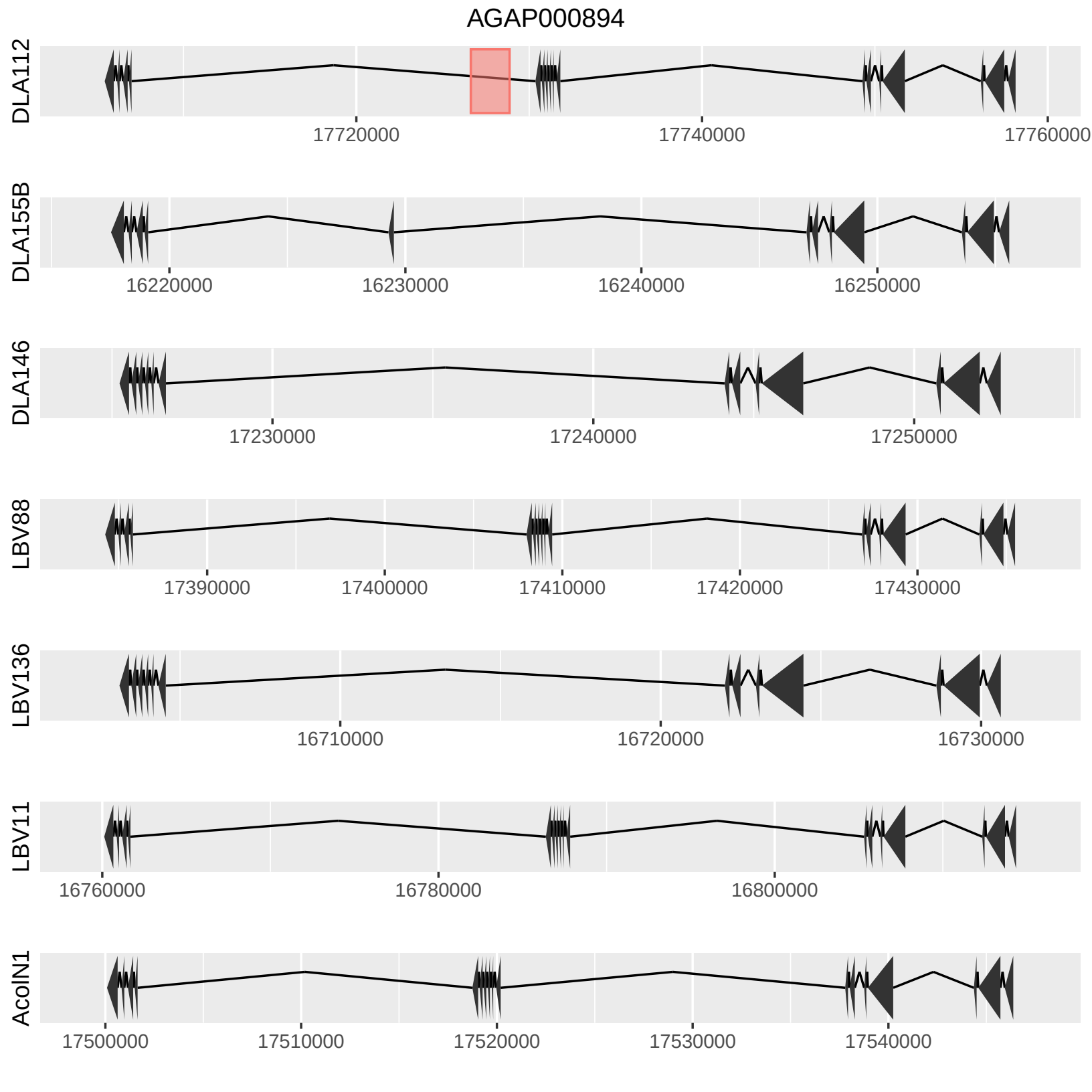


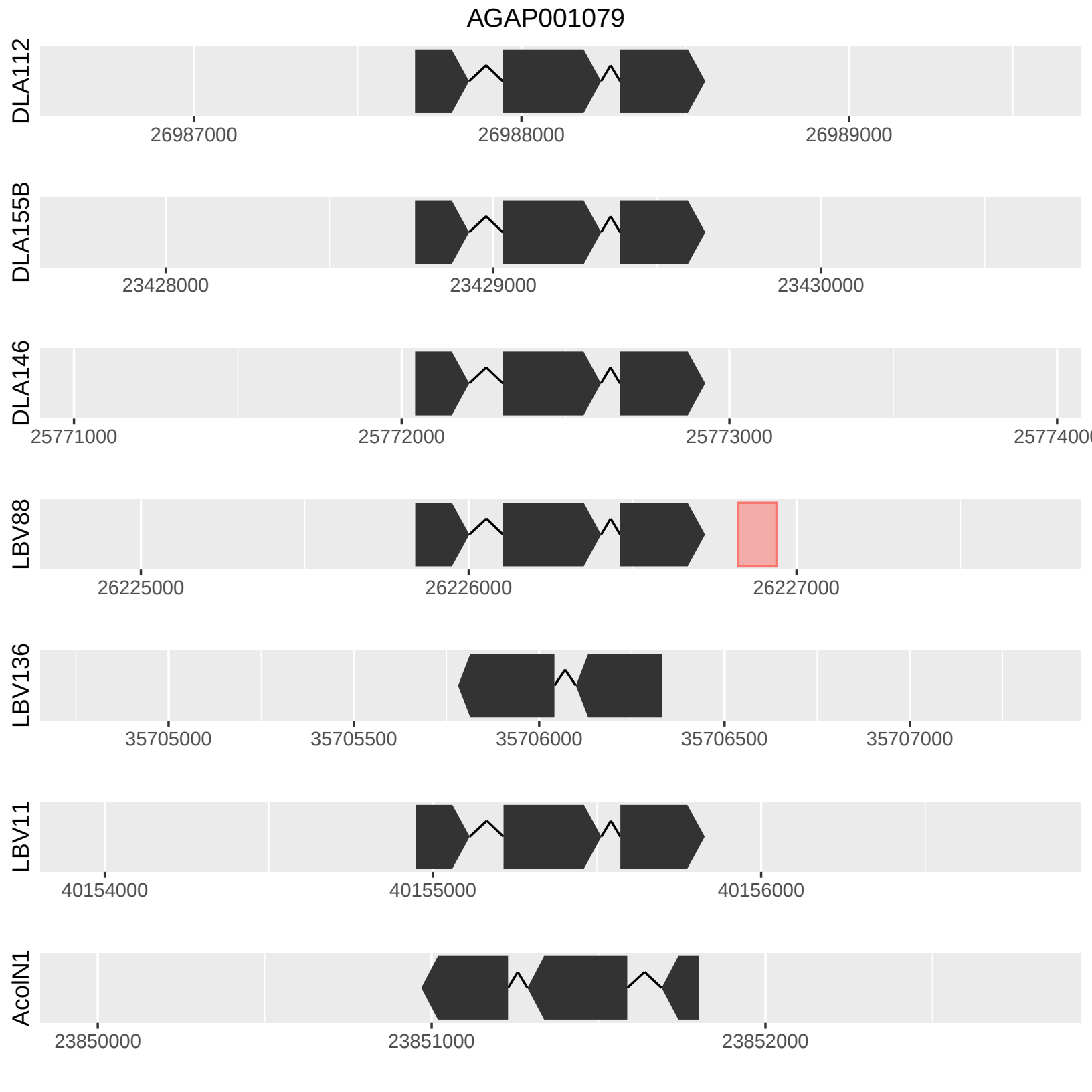

AGAP000368
DLA112
6770000
6780000
6790000
DLA155B
6355000
6360000
6365000
6370000
6375000
DLA146
6420000
6430000
6440000
LBV88
6970000
6980000
6990000
LBV136
6485000
6490000
6495000
6500000
LBV11
6470000
6480000
6490000
AcolN1
6880000
6890000
6900000

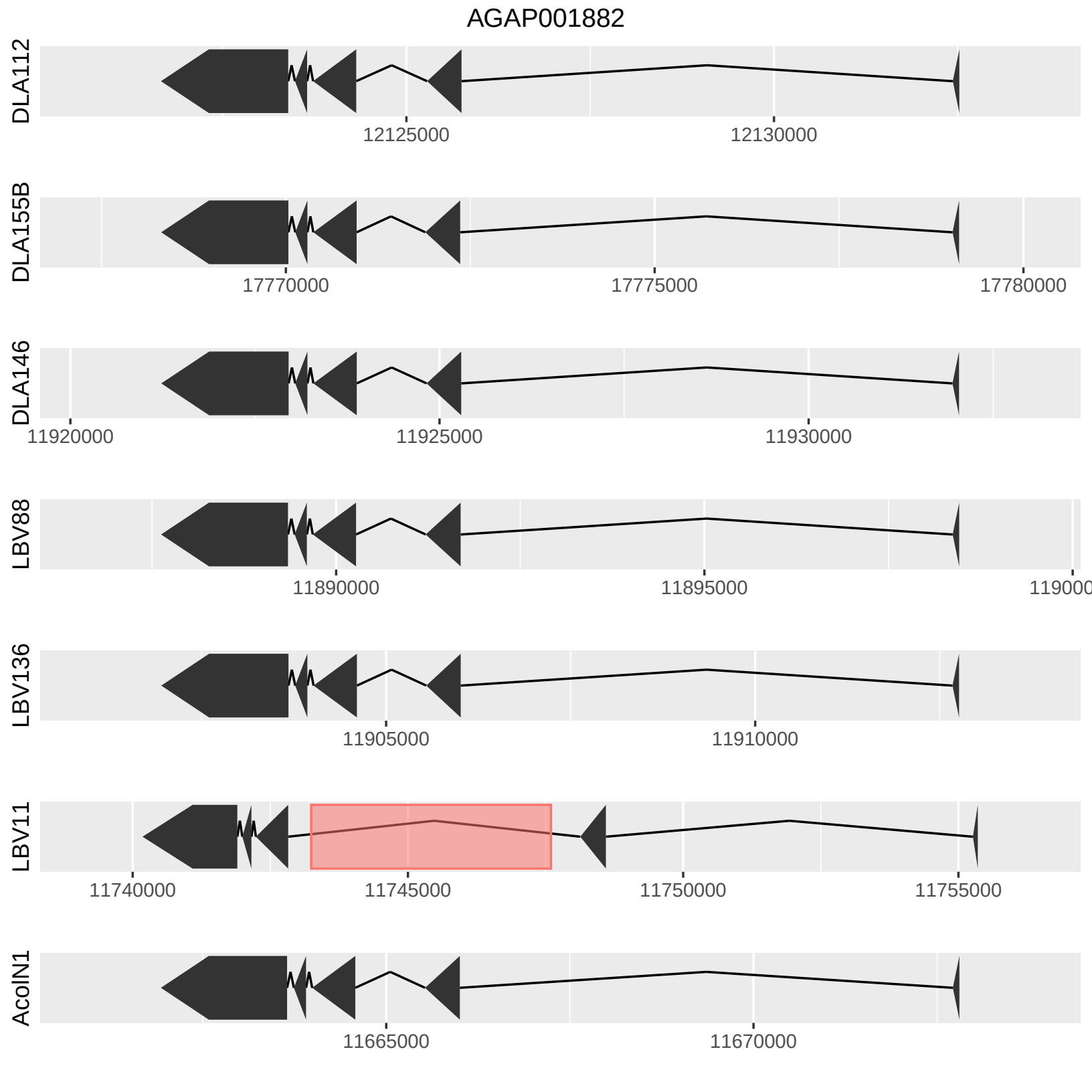

AGAP000531
DLA112
9513000
9514000
9515000
9516000
DLA155B
9077000
9078000
9079000
9080000
DLA146
9111000
9112000
9113000
911400
LBV88
9620000
9621000
9622000
9623000
LBV136
8742000
8743000
8744000
8745000
LBV11
9095000
9096000
9097000
AcoIN1
9657000
9658000
9659000
9660000

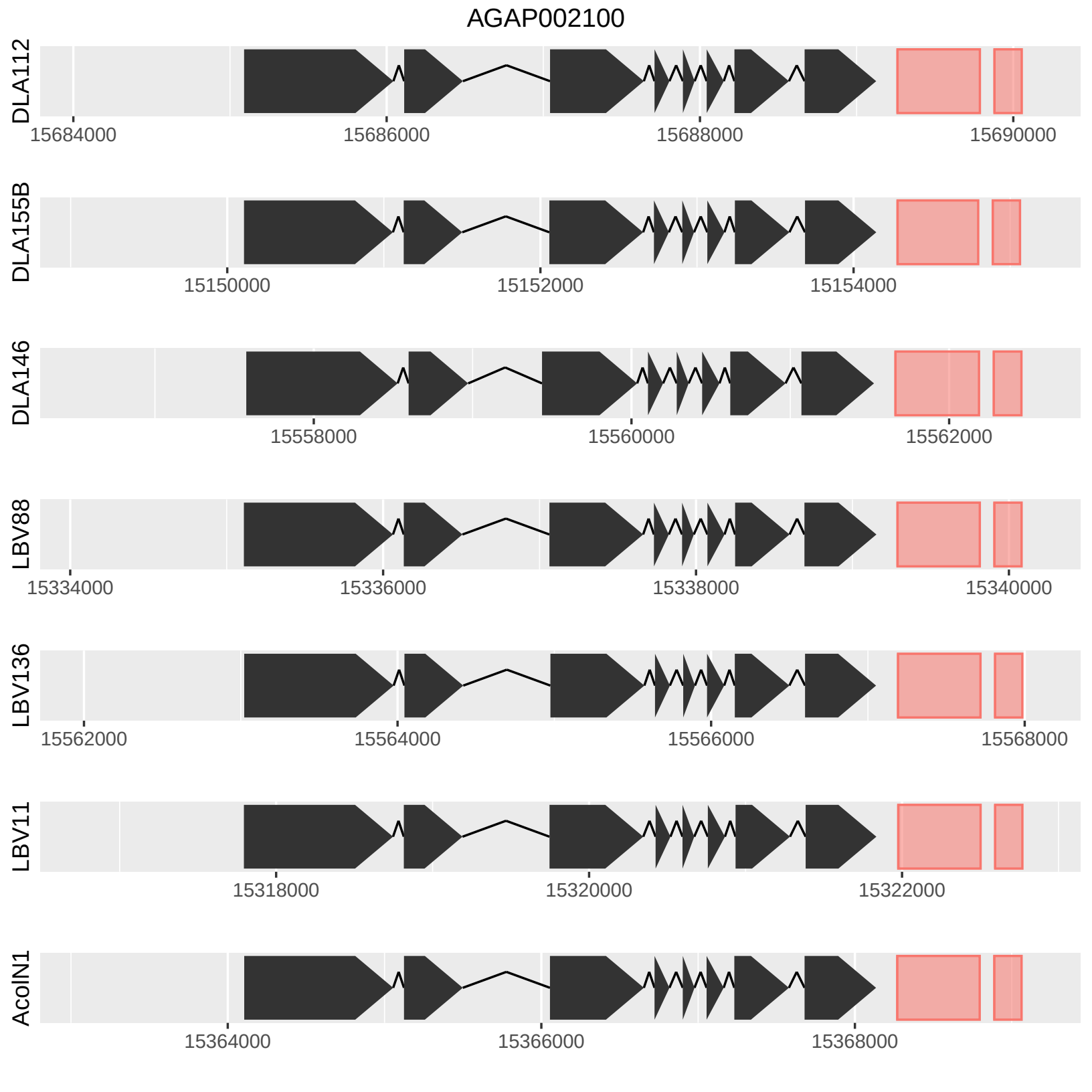

AGAP000707
DLA112
12900000
12910000
12920000
DLA155B
12080000
12090000
12100000
DLA146
12500000
12510000
12520000
LBV88
13030000
13040000
13050000
LBV136
12100000
12105000
12110000
LBV11
12415000
12420000
12425000
12430000
AcolN1
12900000
12910000

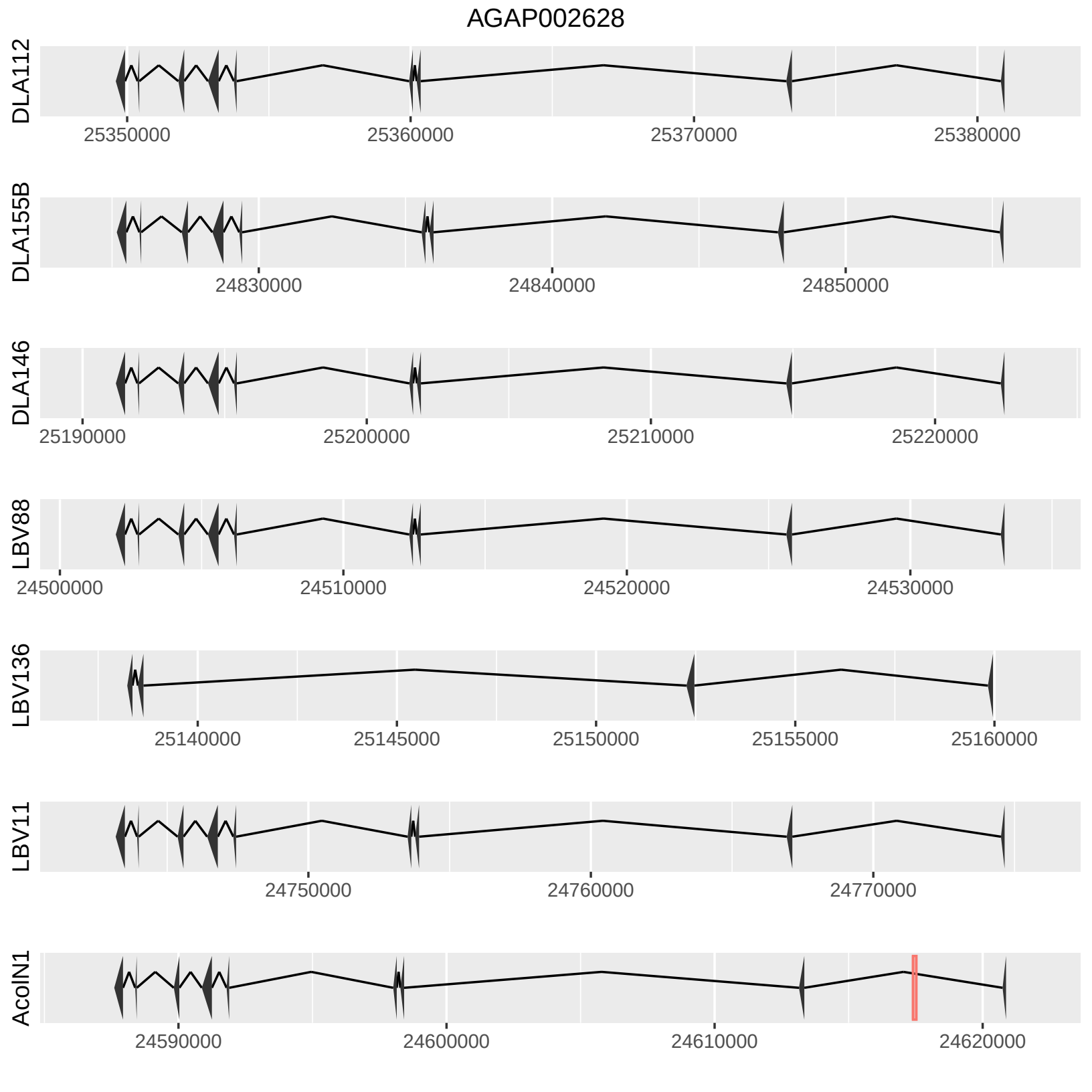

AGAP000741
DLA112
13810000
13812500
13815000
DLA155B
12967500
12970000
12972500
DLA146
13442000
13444000
13446000
13448000
13450000
LBV88
13830000
13832500
13835000
13837500
LBV136
13000000
13002500
13005000
13007500
LBV11
13325000
13327500
13330000
13332500
AcoIN1
14002000
14004000
14006000
14008000
14010000

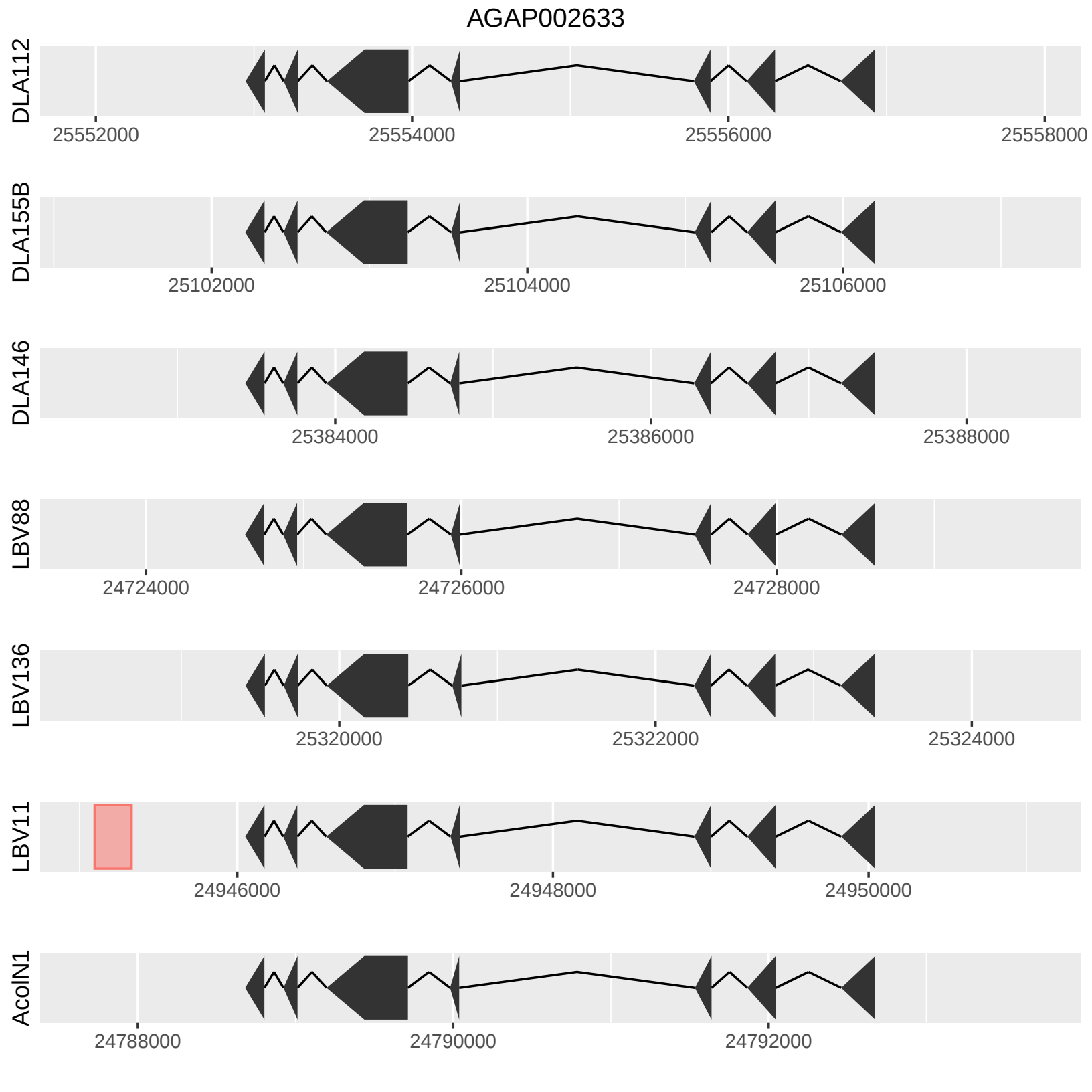

AGAP000894
DLA112
17720000
17740000
17760000
DLA155B
16220000
16230000
16240000
16250000
DLA146
17230000
17240000
17250000
LBV88
17390000
17400000
17410000
17420000
17430000
LBV136
16710000
16720000
16730000
LBV11
16760000
16780000
16800000
AcolN1
17500000
17510000
17520000
17530000
17540000

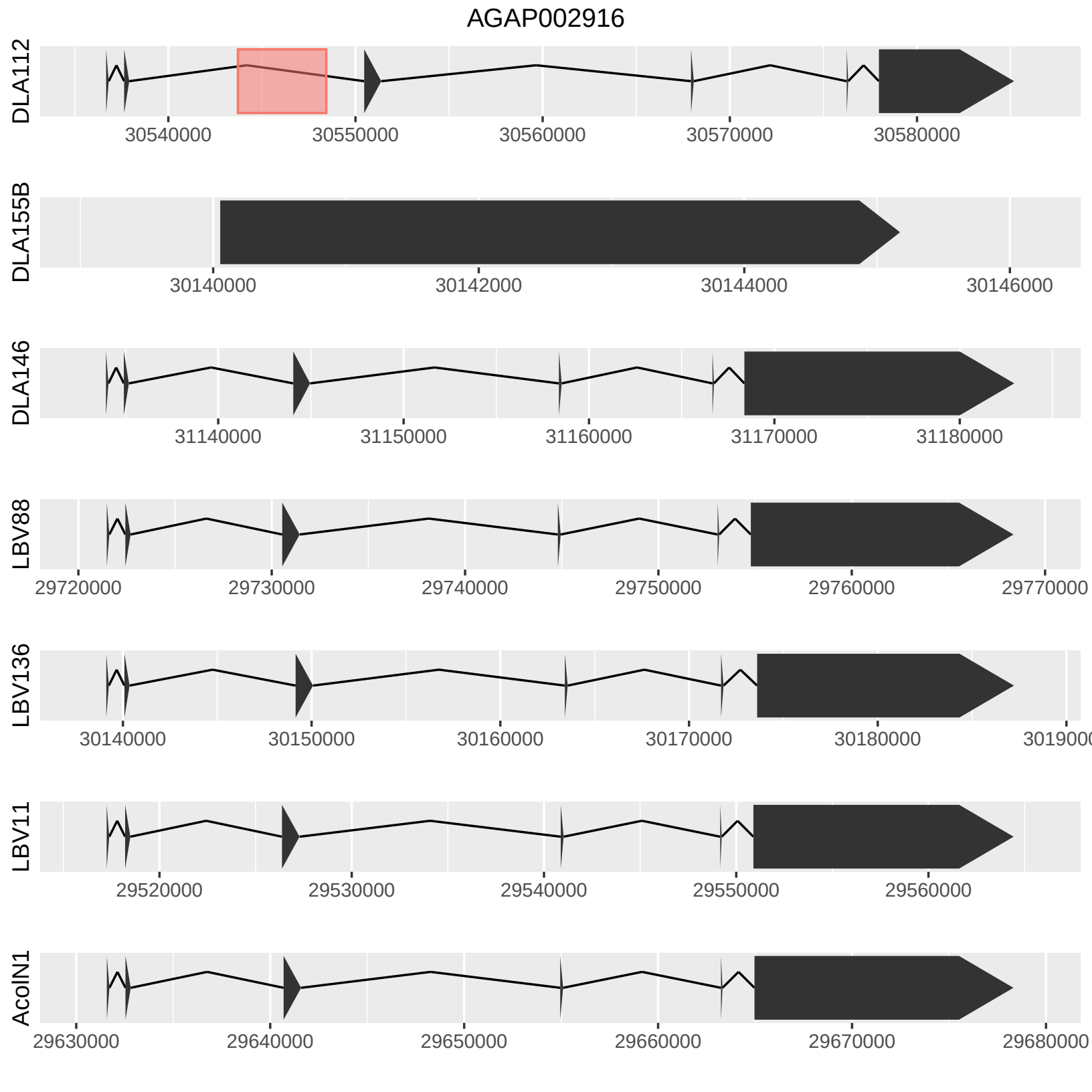

AGAP001079
DLA112
26987000
26988000
26989000
DLA155B
23428000
23429000
23430000
DLA146
25771000
25772000
25773000
LBV88
26225000
26226000
26227000
LBV136
35705000
35705500
35706000
35706500
35707000
LBV11
40154000
40155000
40156000
AcolN1
23850000
23851000
23852000

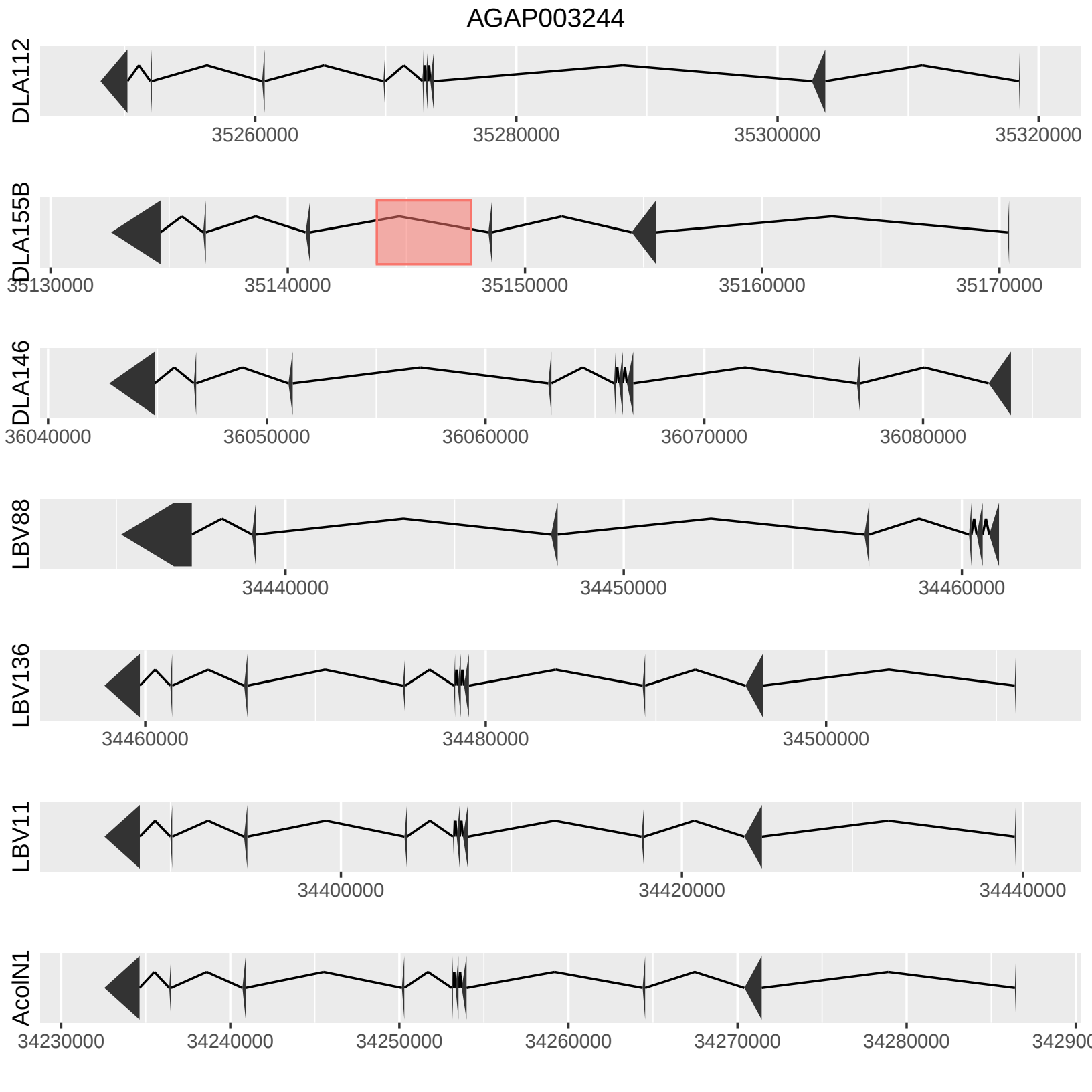

AGAP001882
DLA112
12125000
12130000
DLA155B
17770000
17775000
17780000
DLA146
11920000
11925000
11930000
LBV88
11890000
11895000
119000
LBV136
11905000
11910000
LBV11
11740000
11745000
11750000
11755000
AcolN1
11665000
11670000

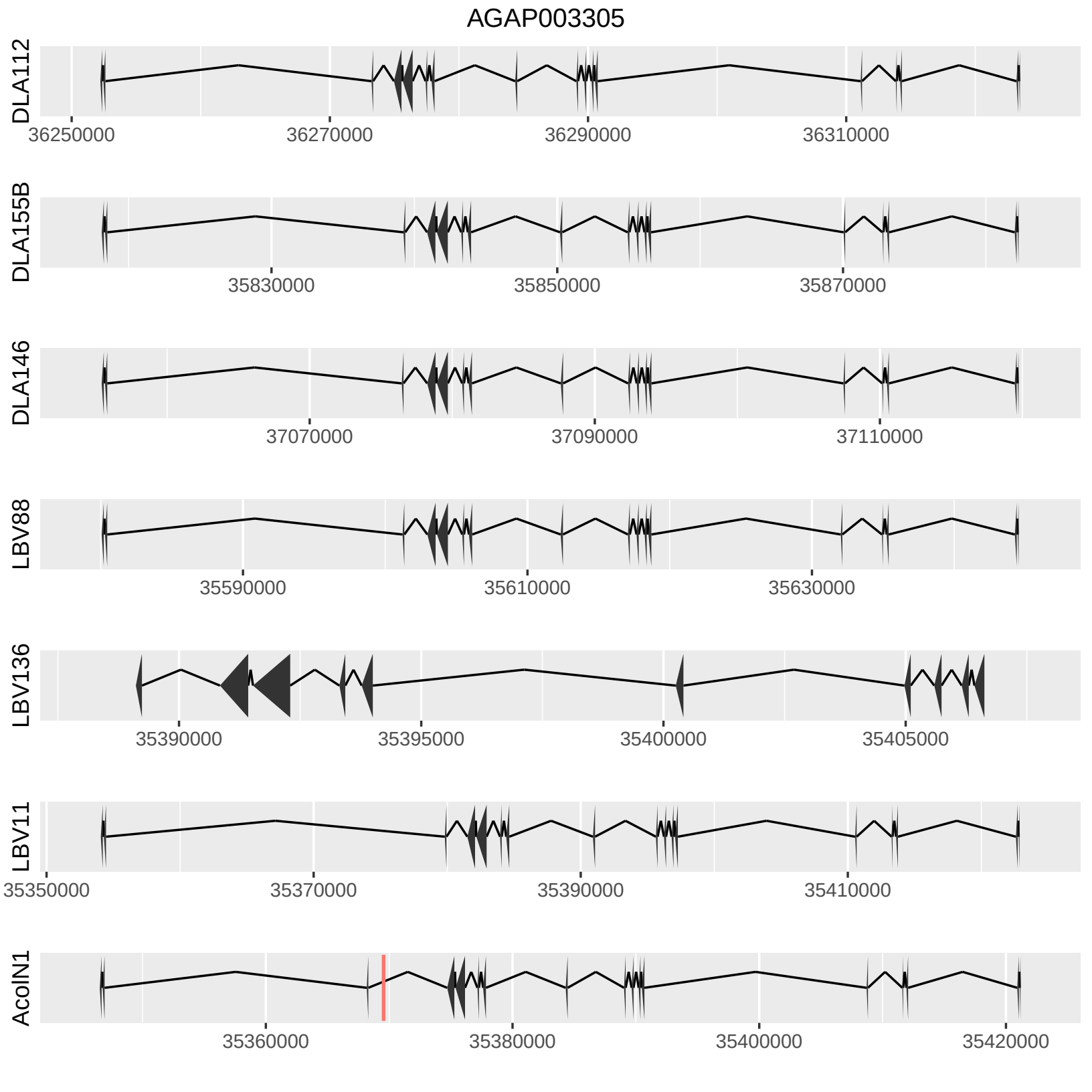

AGAP002100
DLA112
15684000
15686000
15688000
15690000
DLA155B
15150000
15152000
15154000
DLA146
15558000
15560000
15562000
LBV88
15334000
15336000
15338000
15340000
LBV136
15562000
15564000
15566000
15568000
LBV11
15318000
15320000
15322000
AcolN1
15364000
15366000
15368000

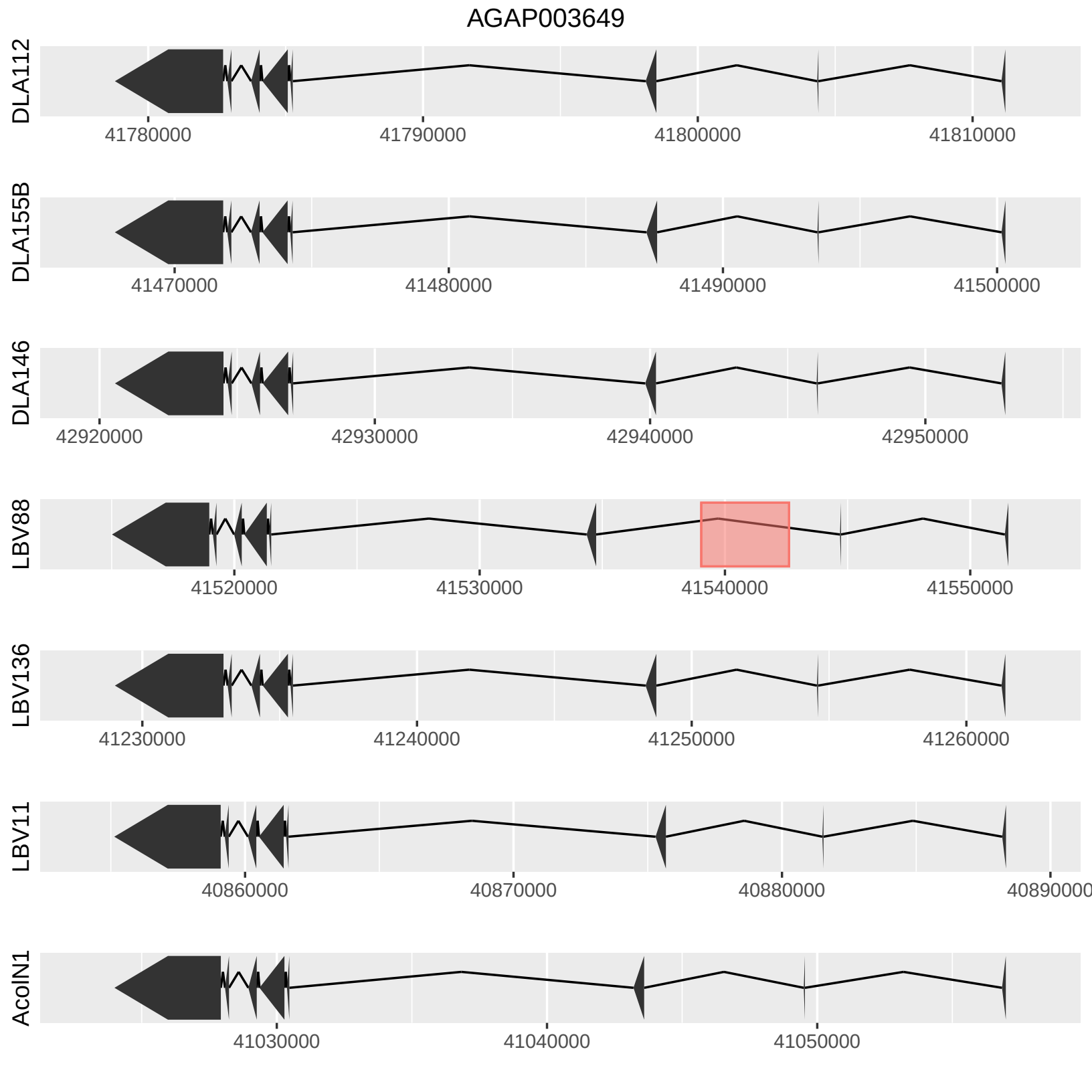

AGAP002628
DLA112
25350000
25360000
25370000
25380000
DLA155B
24830000
24840000
24850000
DLA146
25190000
25200000
25210000
25220000
LBV88
24500000
24510000
24520000
24530000
LBV136
25140000
25145000
25150000
25155000
25160000
LBV11
24750000
24760000
24770000
AcolN1
24590000
24600000
24610000
24620000

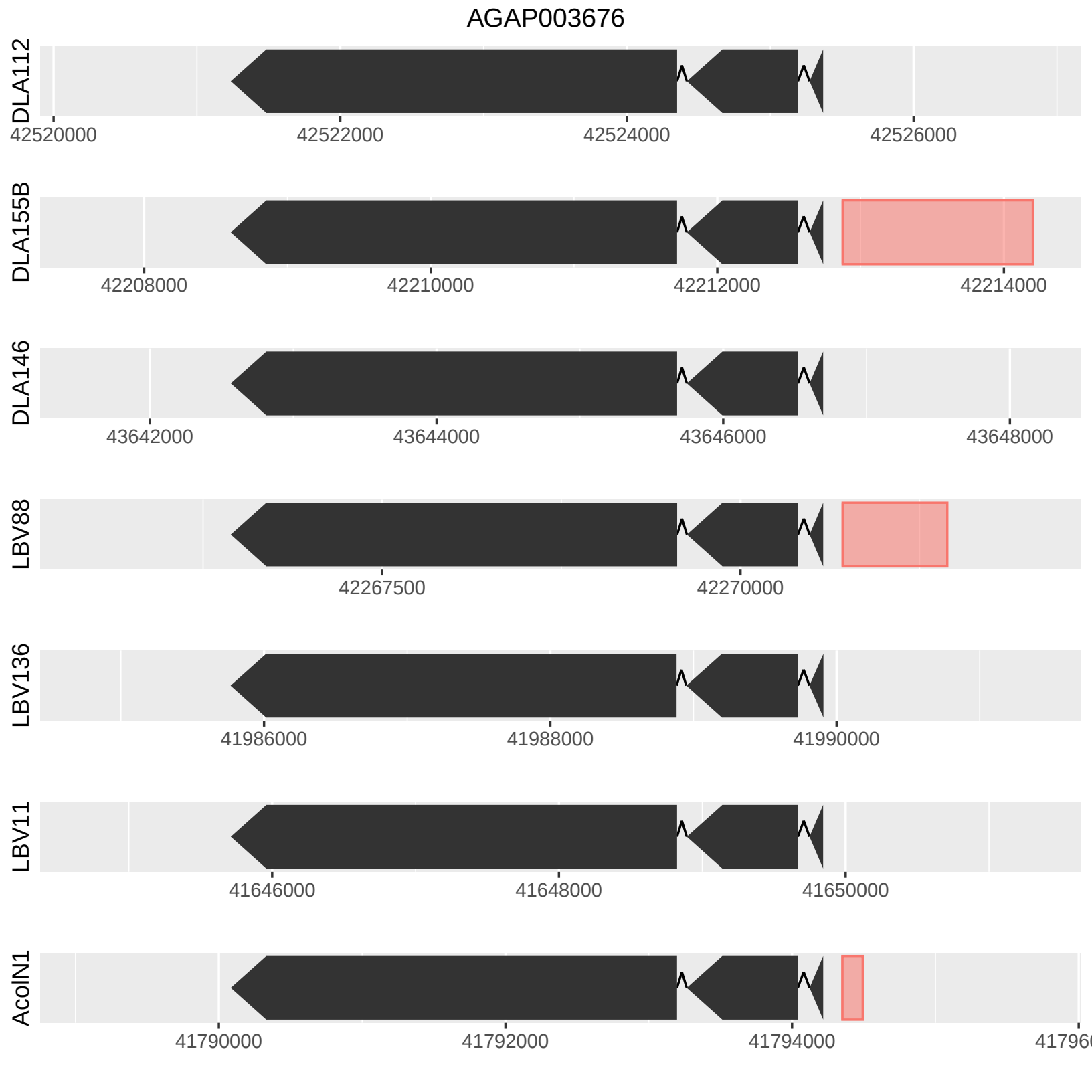

AGAP002633
DLA112
25552000
25554000
25556000
25558000
DLA155B
25102000
25104000
25106000
DLA146
25384000
25386000
25388000
LBV88
24724000
24726000
24728000
LBV136
25320000
25322000
25324000
LBV11
24946000
24948000
24950000
AcoIN1
24788000
24790000
24792000

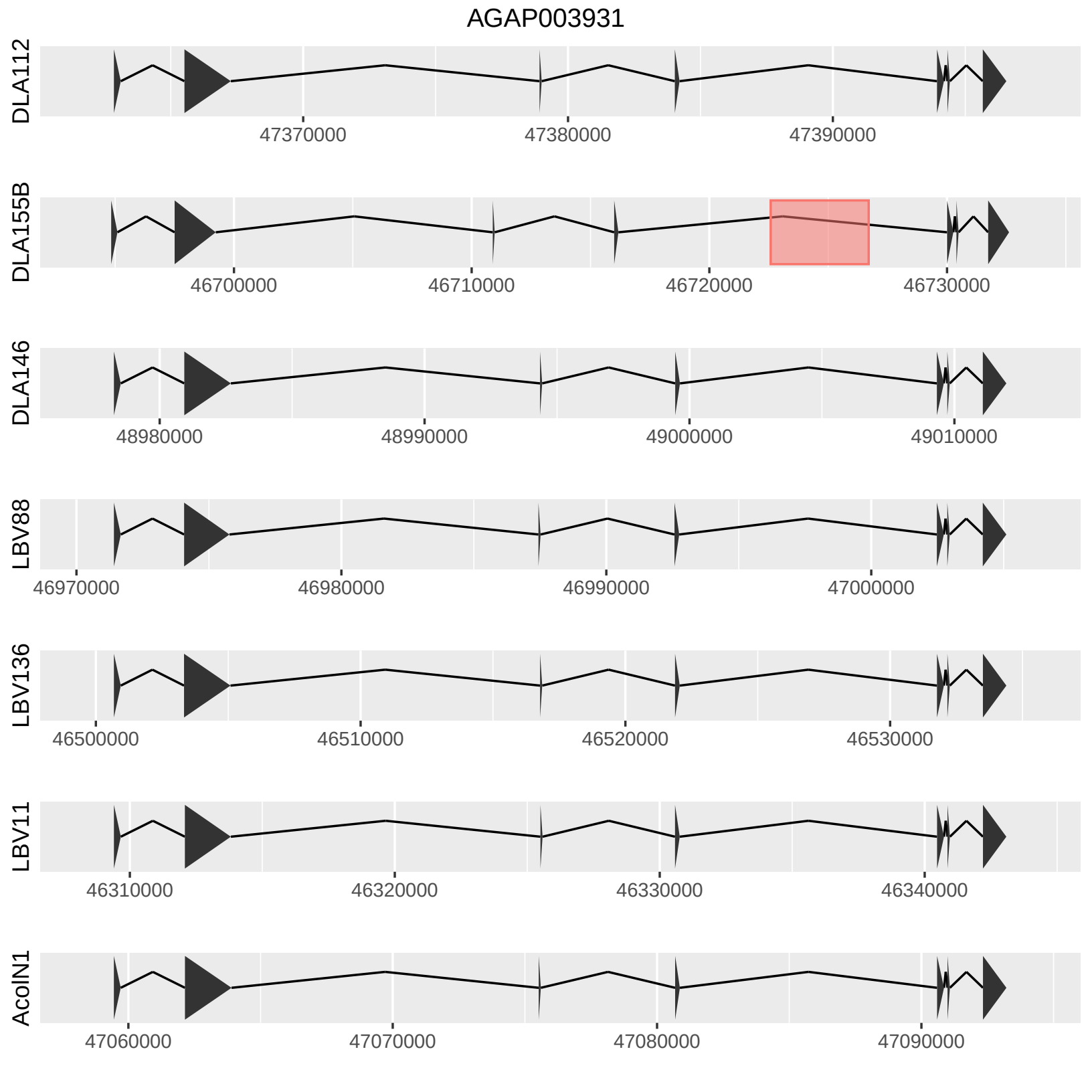

AGAP002916
DLA112
30540000
30550000
30560000
30570000
30580000
DLA155B
30140000
30142000
30144000
30146000
DLA146
31140000
31150000
31160000
31170000
31180000
LBV88
29720000
29730000
29740000
29750000
29760000
29770000
LBV136
30140000
30150000
30160000
30170000
30180000
LBV11
29520000
29530000
29540000
29550000
29560000
AcolN1
29630000
29640000
29650000
29660000
29670000
29680000

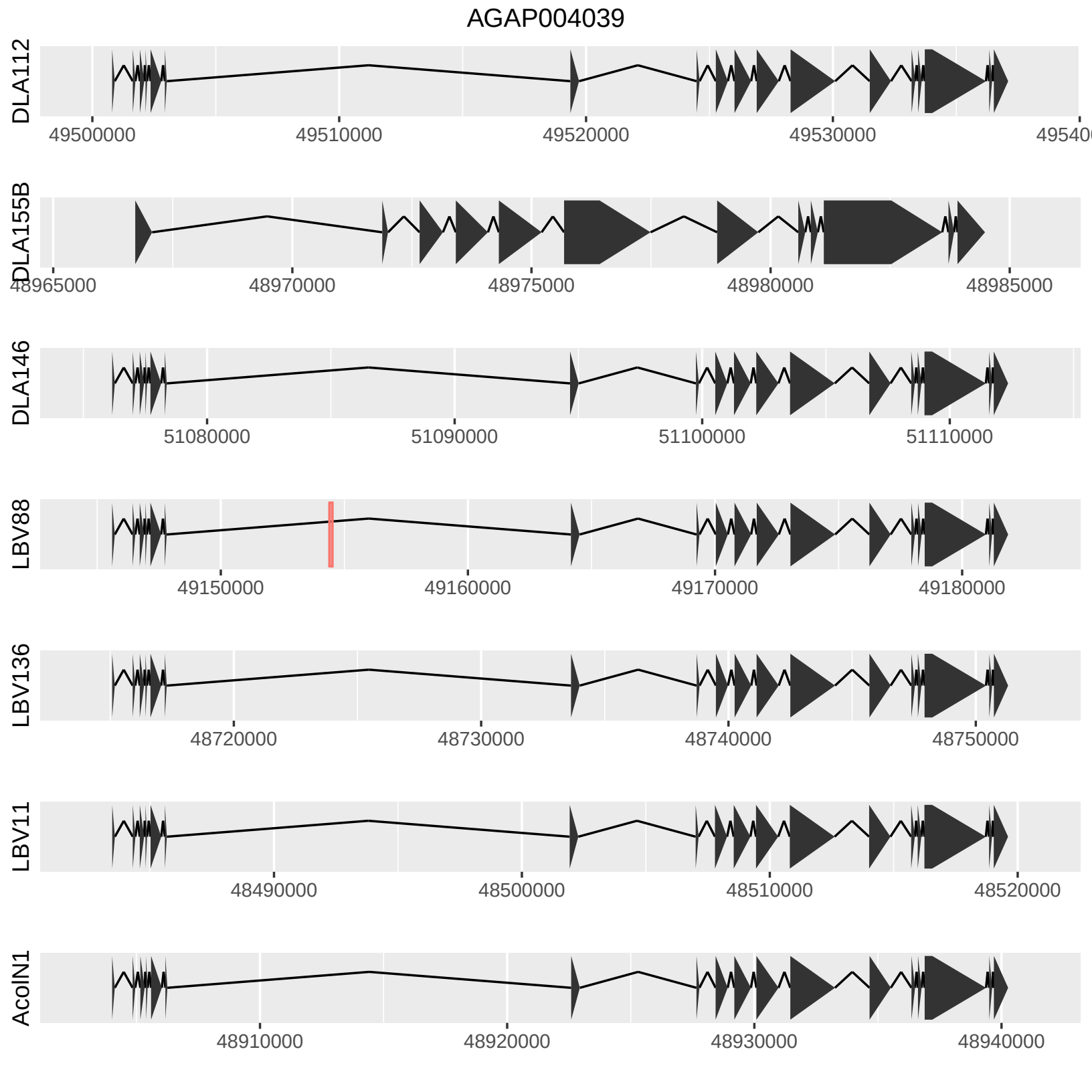

AGAP003244
DLA112
35260000
35280000
35300000
35320000
DLA155B
35130000
35140000
35150000
35160000
35170000
DLA146
36040000
36050000
36060000
36070000
36080000
LBV88
34440000
34450000
34460000
LBV136
34460000
34480000
34500000
LBV11
34400000
34420000
34440000
AcolN1
34230000
34240000
34250000
34260000
34270000
34280000
34290

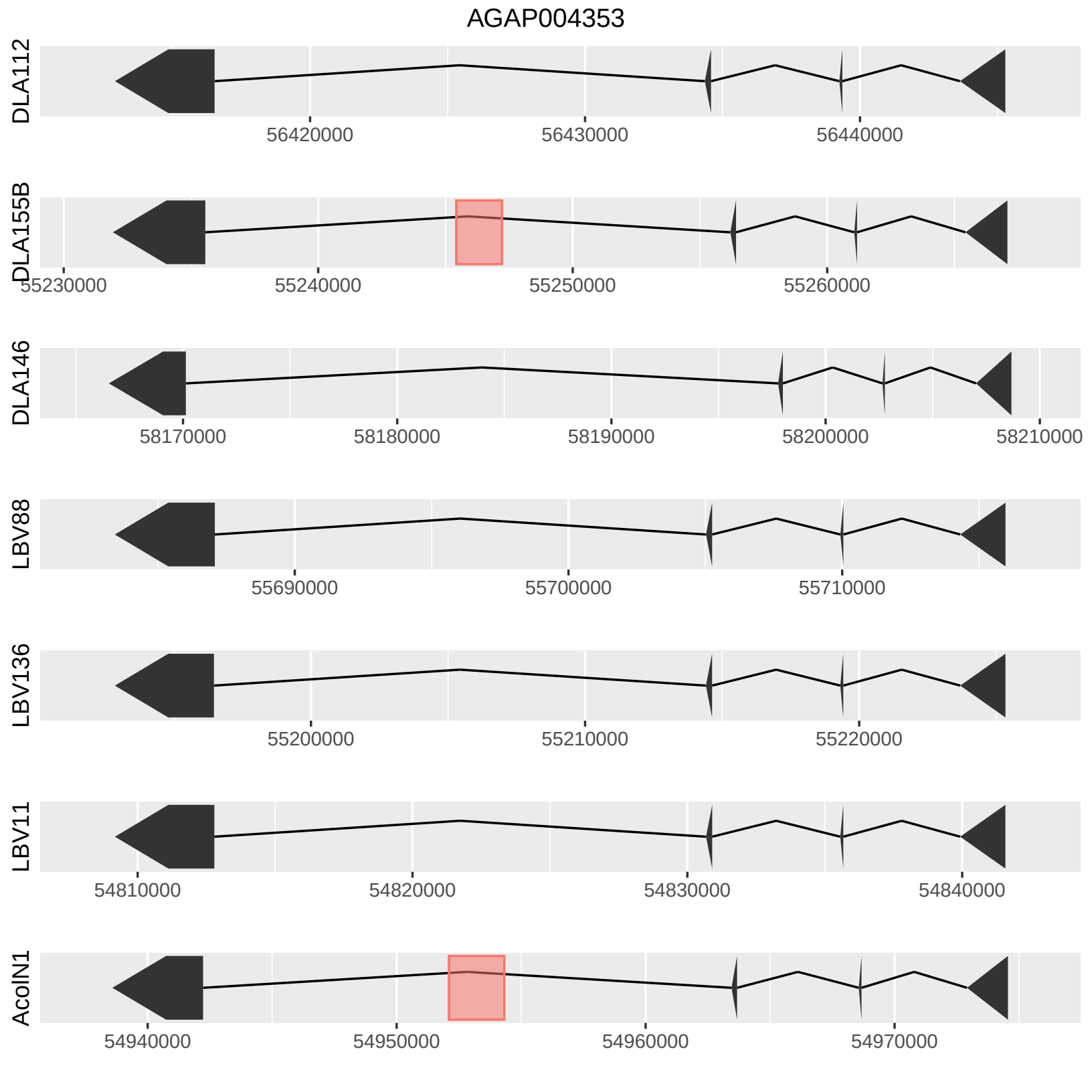

AGAP003305
DLA112
36250000
36270000
36290000
36310000
DLA155B
35830000
35850000
35870000
DLA146
37070000
37090000
37110000
LBV88
35590000
35610000
35630000
LBV136
35390000
35395000
35400000
35405000
LBV11
35350000
35370000
35390000
35410000
AcolN1
35360000
35380000
35400000
35420000

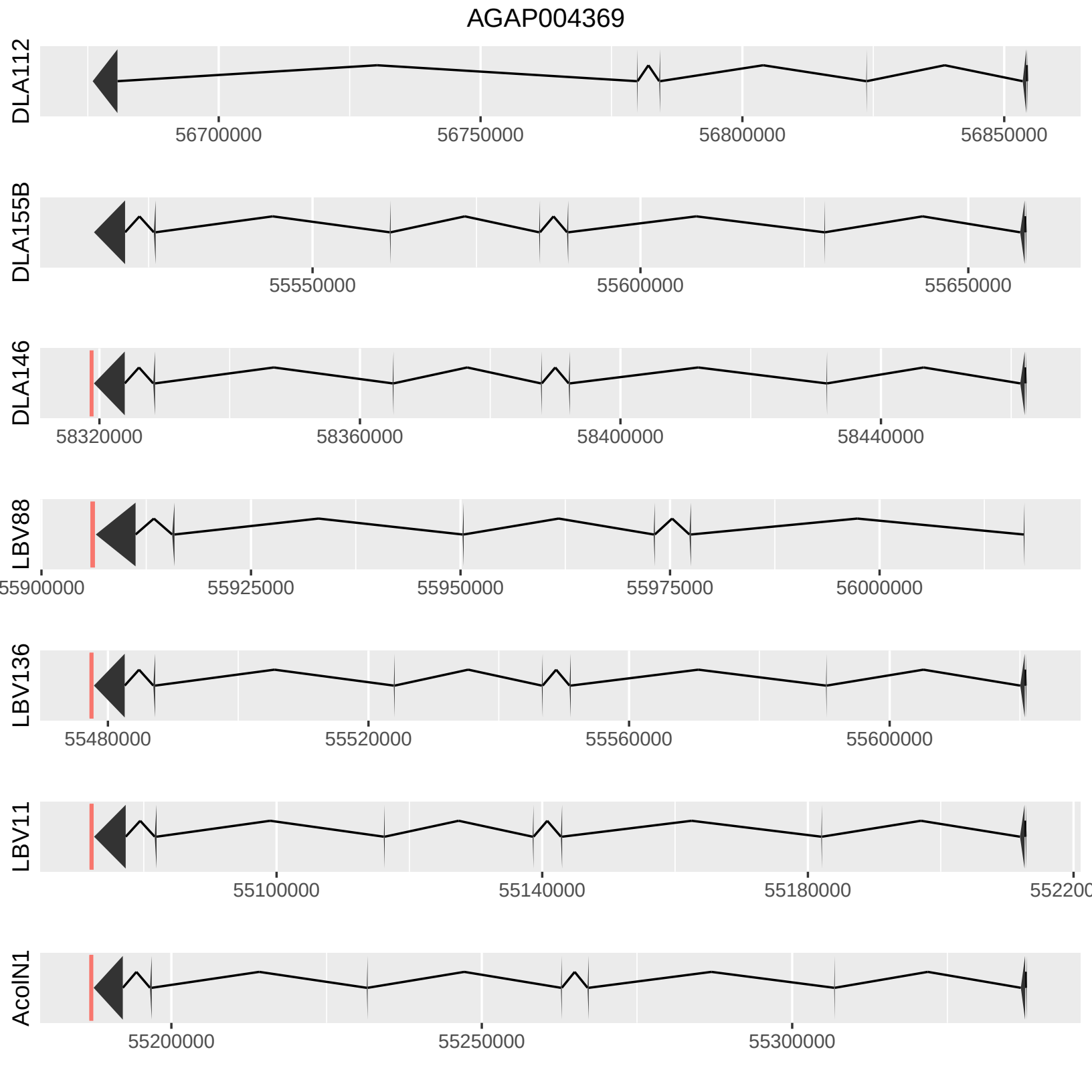

AGAP003649
DLA112
41780000
41790000
41800000
41810000
DLA155B
41470000
41480000
41490000
41500000
DLA146
42920000
42930000
42940000
42950000
LBV88
41520000
41530000
41540000
41550000
LBV136
41230000
41240000
41250000
41260000
LBV11
40860000
40870000
40880000
40890000
AcolN1
41030000
41040000
41050000

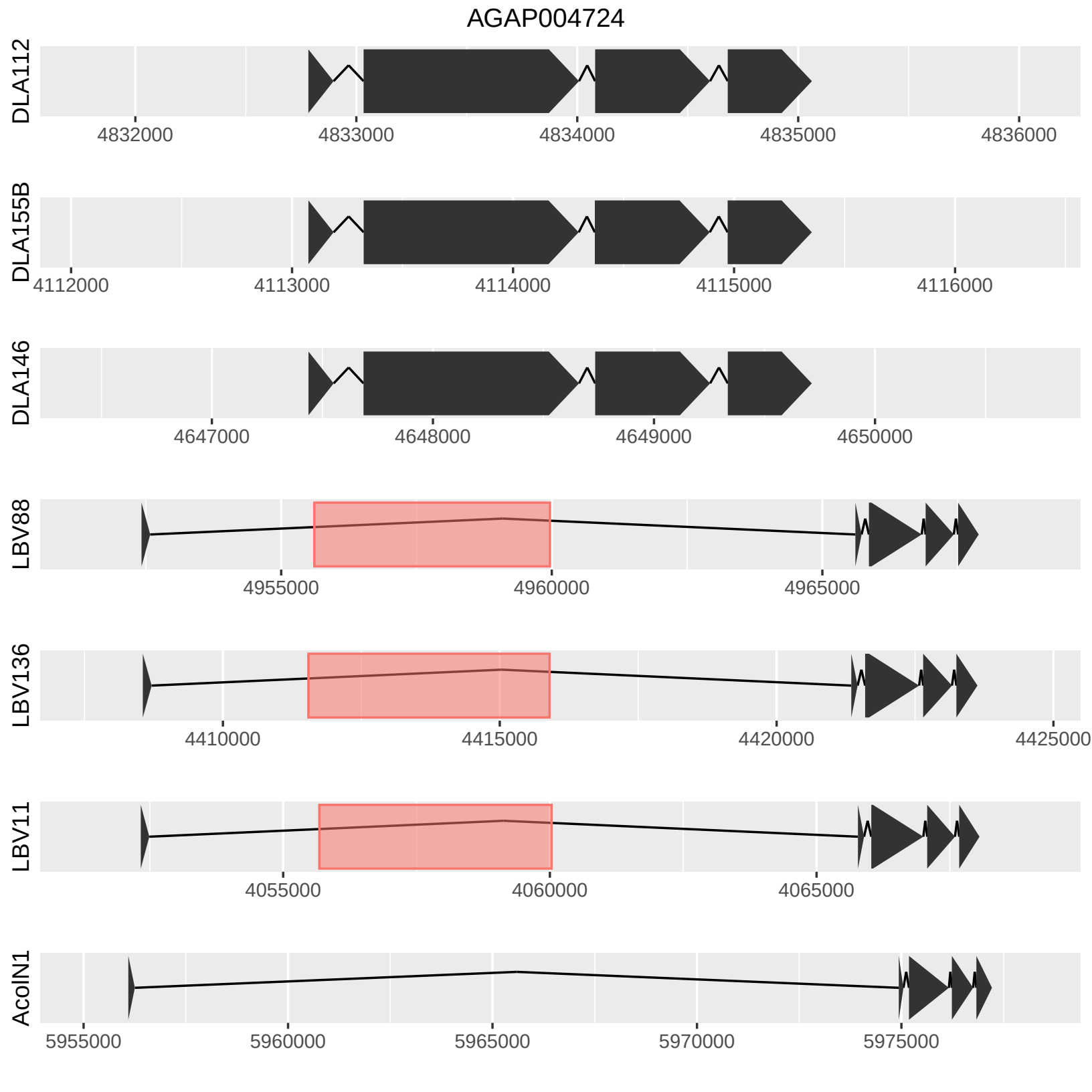

AGAP003676
DLA112
42520000
42522000
42524000
42526000
DLA155B
42208000
42210000
42212000
42214000
DLA146
43642000
43644000
43646000
43648000
LBV88
42267500
42270000
LBV136
41986000
41988000
41990000
LBV11
41646000
41648000
41650000
AcolN1
41790000
41792000
41794000

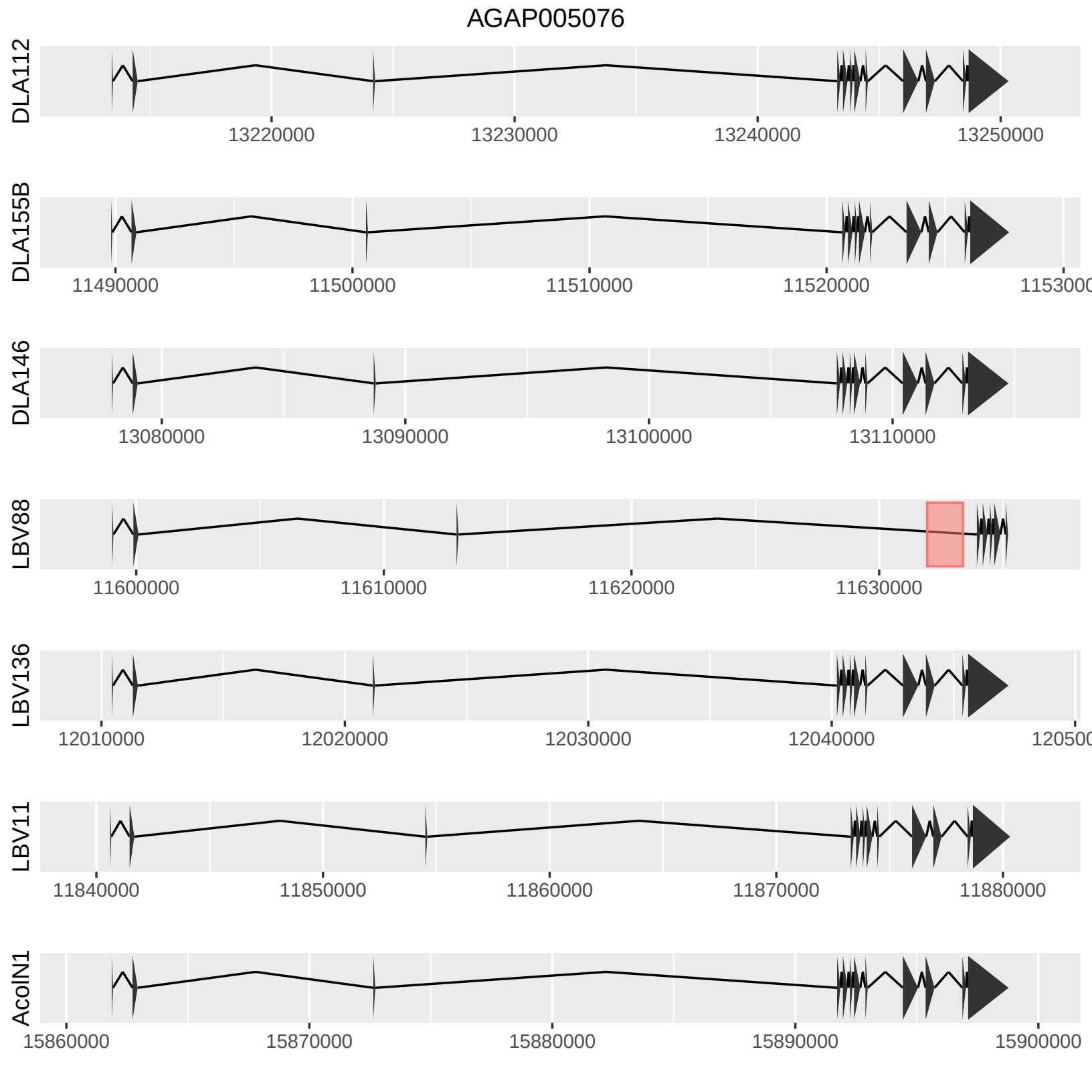

AGAP003931
DLA112
47370000
47380000
47390000
DLA155B
46700000
46710000
46720000
46730000
DLA146
48980000
48990000
49000000
49010000
LBV88
46970000
46980000
46990000
47000000
LBV136
46500000
46510000
46520000
46530000
LBV11
46310000
46320000
46330000
46340000
AcoIN1
47060000
47070000
47080000
47090000

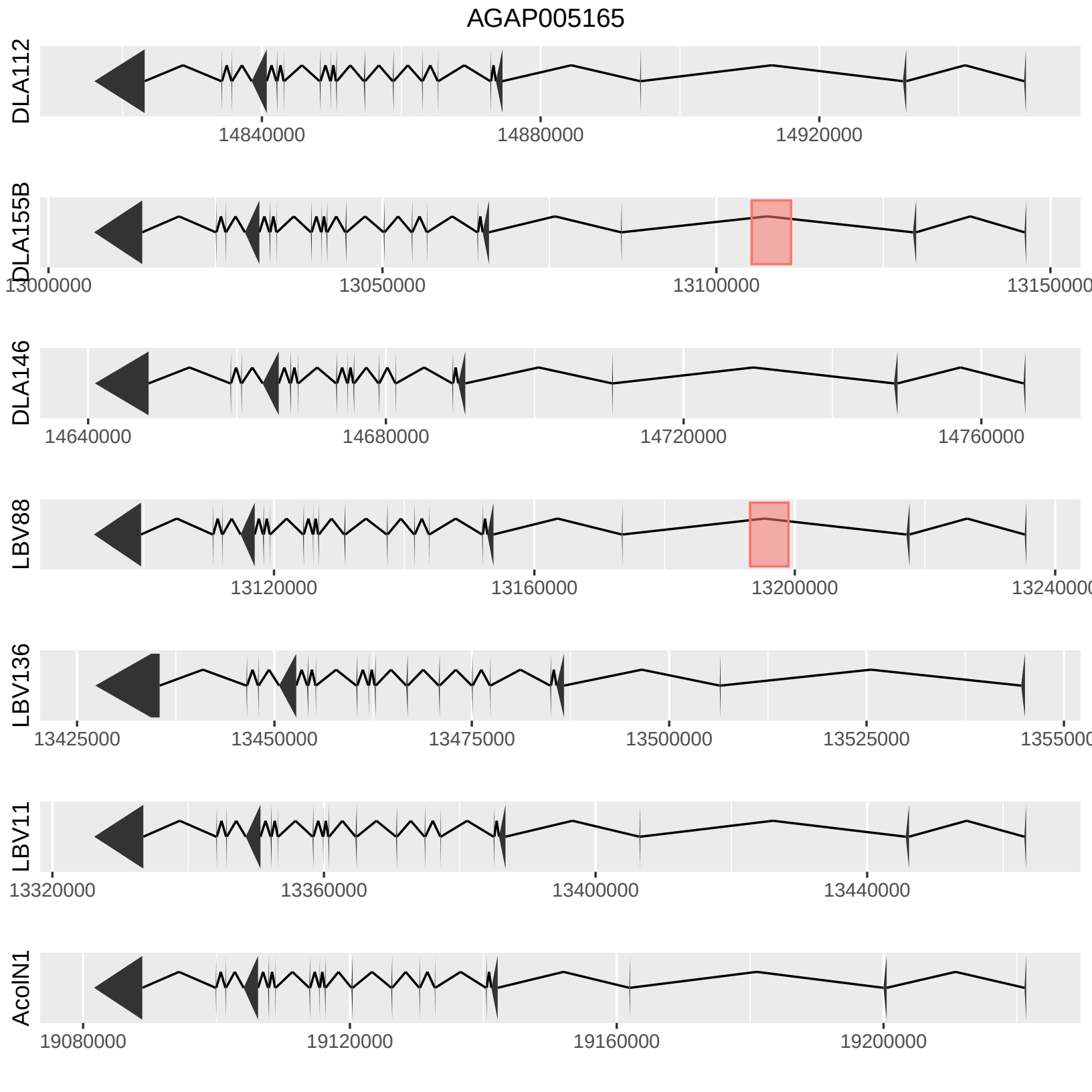

AGAP004039
DLA112
49500000
49510000
49520000
49530000
49540
DLA155B
48965000
48970000
48975000
48980000
48985000
DLA146
51080000
51090000
51100000
51110000
LBV88
49150000
49160000
49170000
49180000
LBV136
48720000
48730000
48740000
48750000
LBV11
48490000
48500000
48510000
48520000
AcolN1
48910000
48920000
48930000
48940000

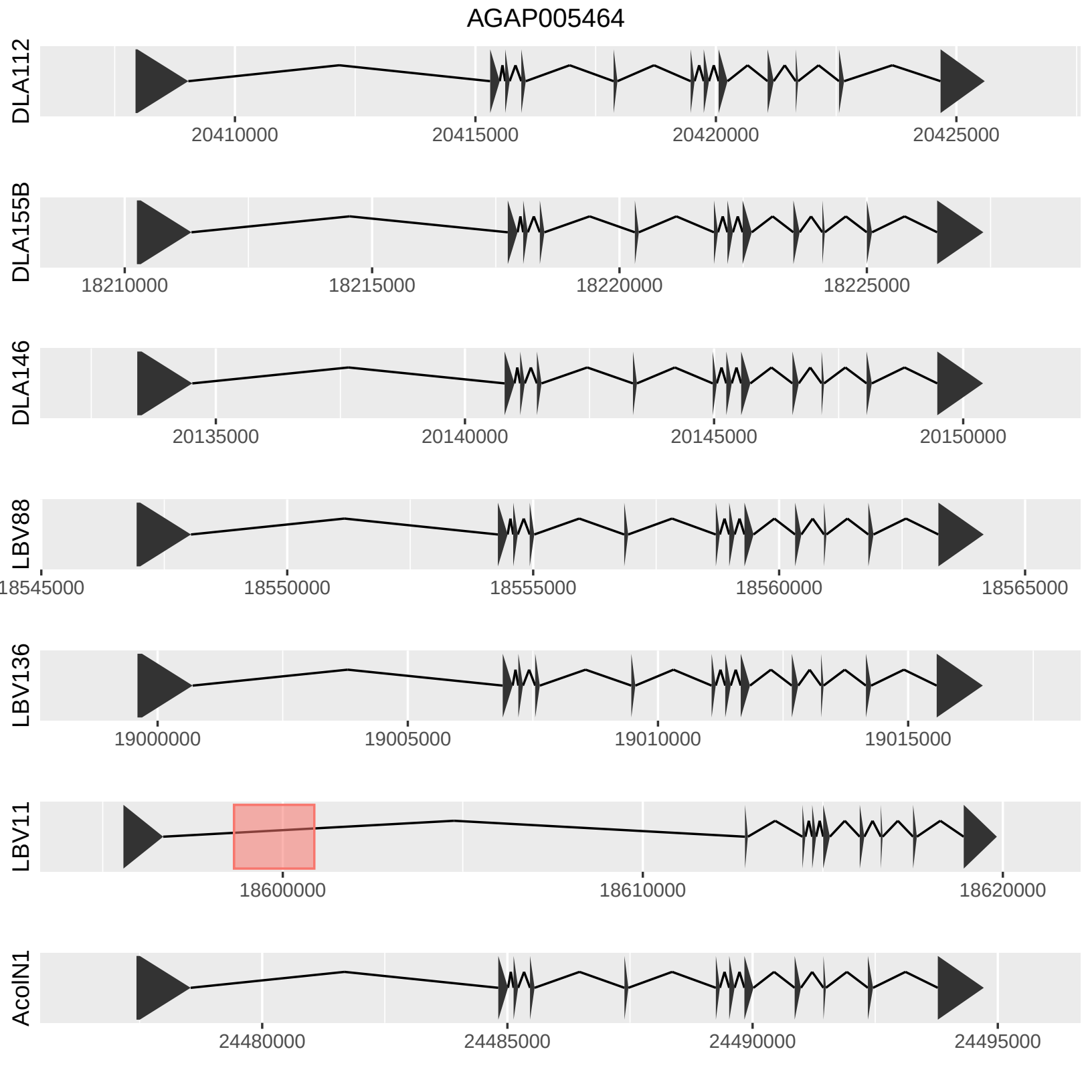

AGAP004353
DLA112
56420000
56430000
56440000
DLA155B
55230000
55240000
55250000
55260000
DLA146
58170000
58180000
58190000
58200000
58210000
LBV88
55690000
55700000
55710000
LBV136
55200000
55210000
55220000
LBV11
54810000
54820000
54830000
54840000
AcoIN1
54940000
54950000
54960000
54970000

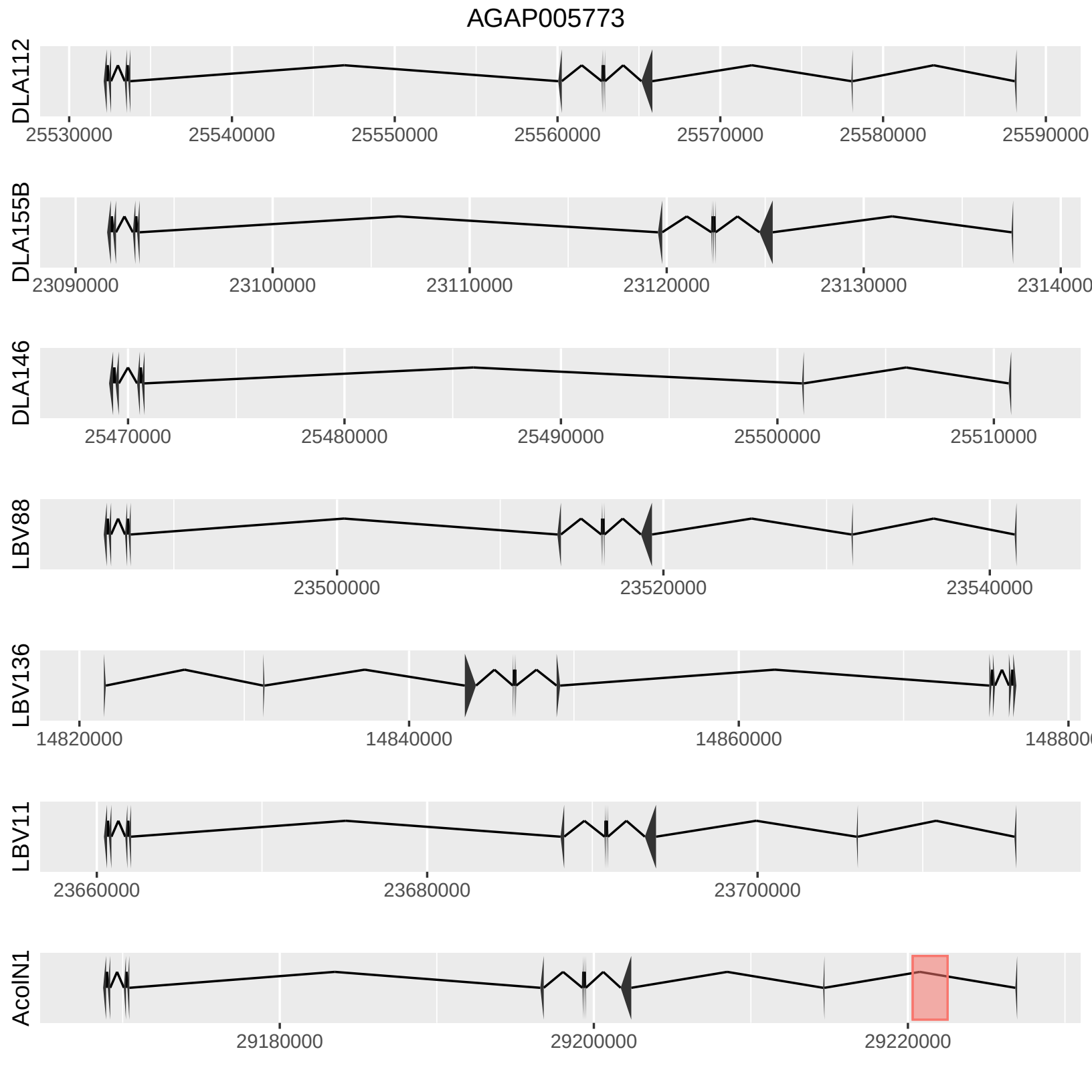

AGAP004369
DLA112
56700000
56750000
56800000
56850000
DLA155B
55550000
55600000
55650000
DLA146
58320000
58360000
58400000
58440000
LBV88
55900000
55925000
55950000
55975000
56000000
LBV136
55480000
55520000
55560000
55600000
LBV11
55100000
55140000
55180000
AcolN1
55200000
55250000
55300000

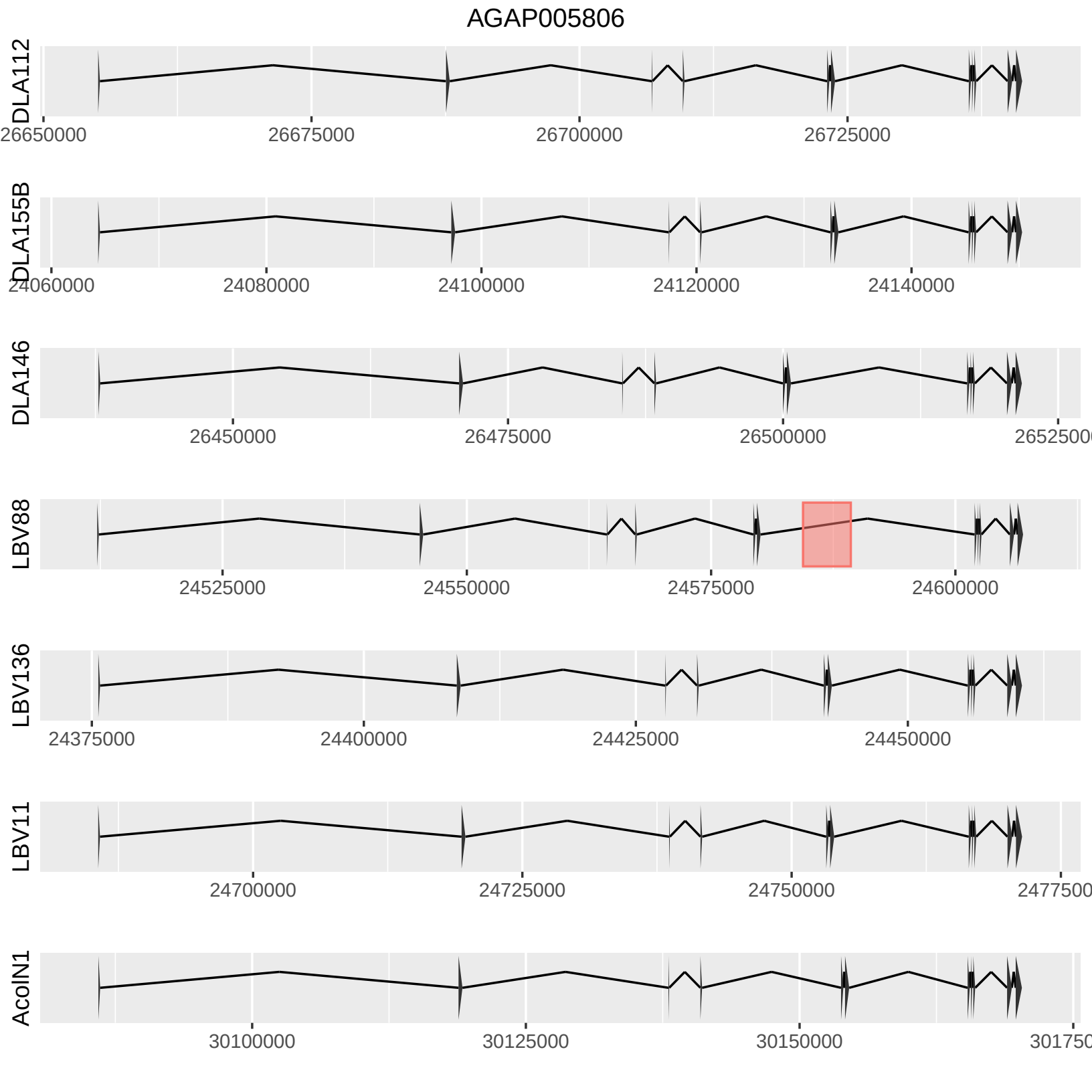

AGAP004724
DLA112
4832000
4833000
4834000
4835000
4836000
DLA155B
4112000
4113000
4114000
4115000
4116000
DLA146
4647000
4648000
4649000
4650000
LBV88
4955000
4960000
4965000
LBV136
4410000
4415000
4420000
4425000
LBV11
4055000
4060000
4065000
AcolN1
5955000
5960000
5965000
5970000
5975000

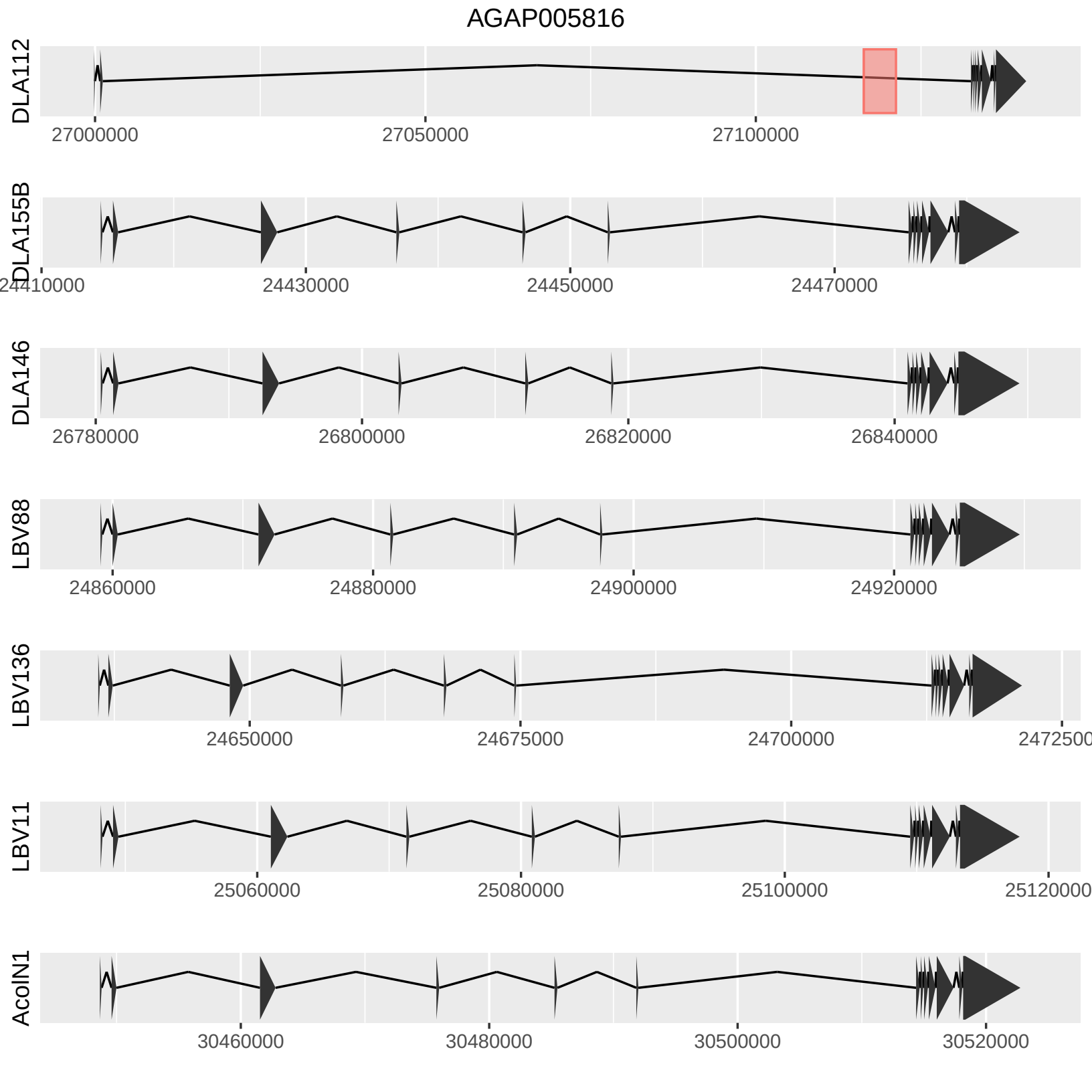

AGAP005076
DLA112
13220000
13230000
13240000
13250000
DLA155B
11490000
11500000
11510000
11520000
1153000
DLA146
13080000
13090000
13100000
13110000
LBV88
11600000
11610000
11620000
11630000
LBV136
12010000
12020000
12030000
12040000
120500
LBV11
11840000
11850000
11860000
11870000
11880000
AcolN1
15860000
15870000
15880000
15890000
15900000

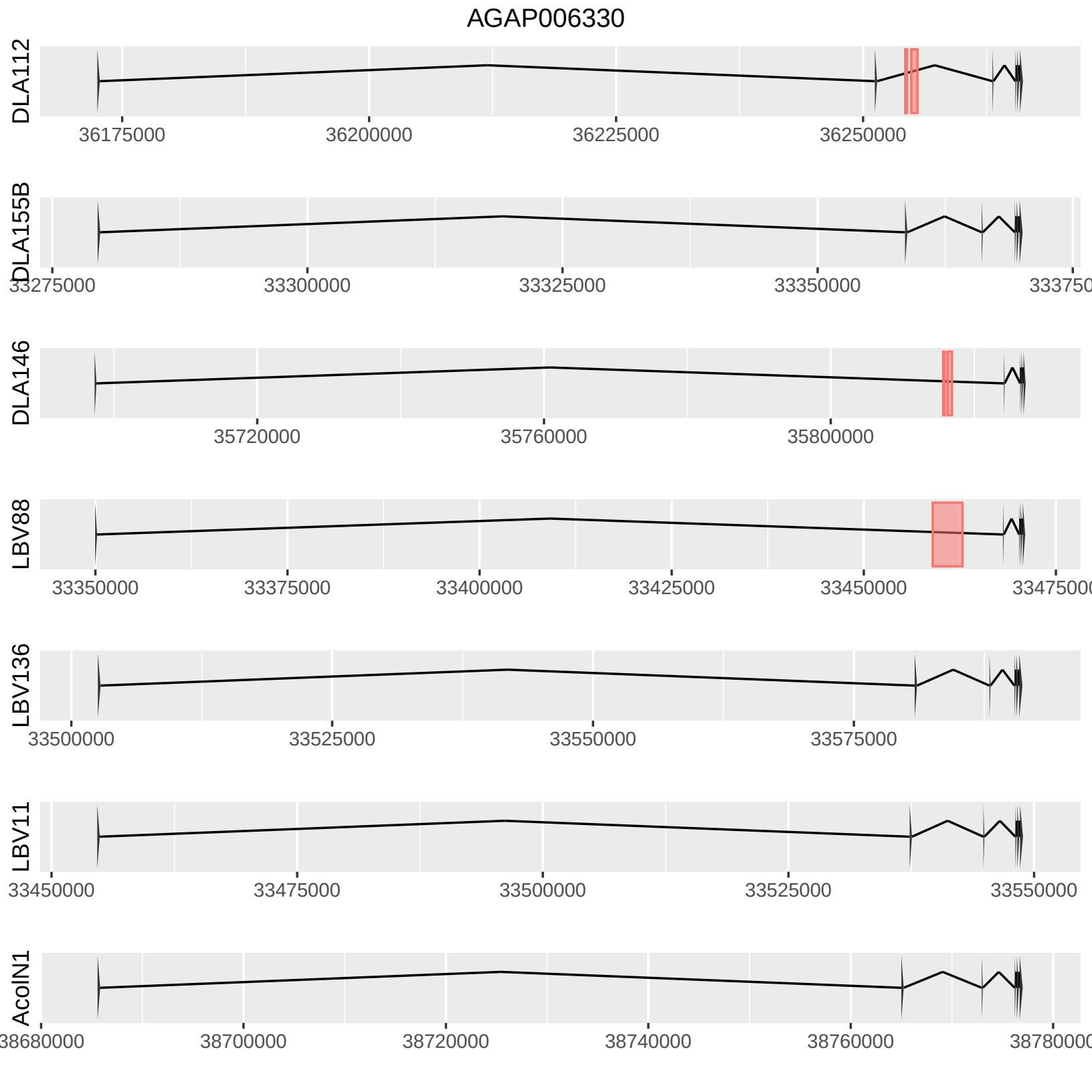

AGAP005165
DLA112
14840000
14880000
14920000
DLA155B
13000000
13050000
13100000
13150000
DLA146
14640000
14680000
14720000
14760000
LBV88
13120000
13160000
13200000
13240000
LBV136
13425000
13450000
13475000
13500000
13525000
13550000
LBV11
13320000
13360000
13400000
13440000
AcolN1
19080000
19120000
19160000
19200000

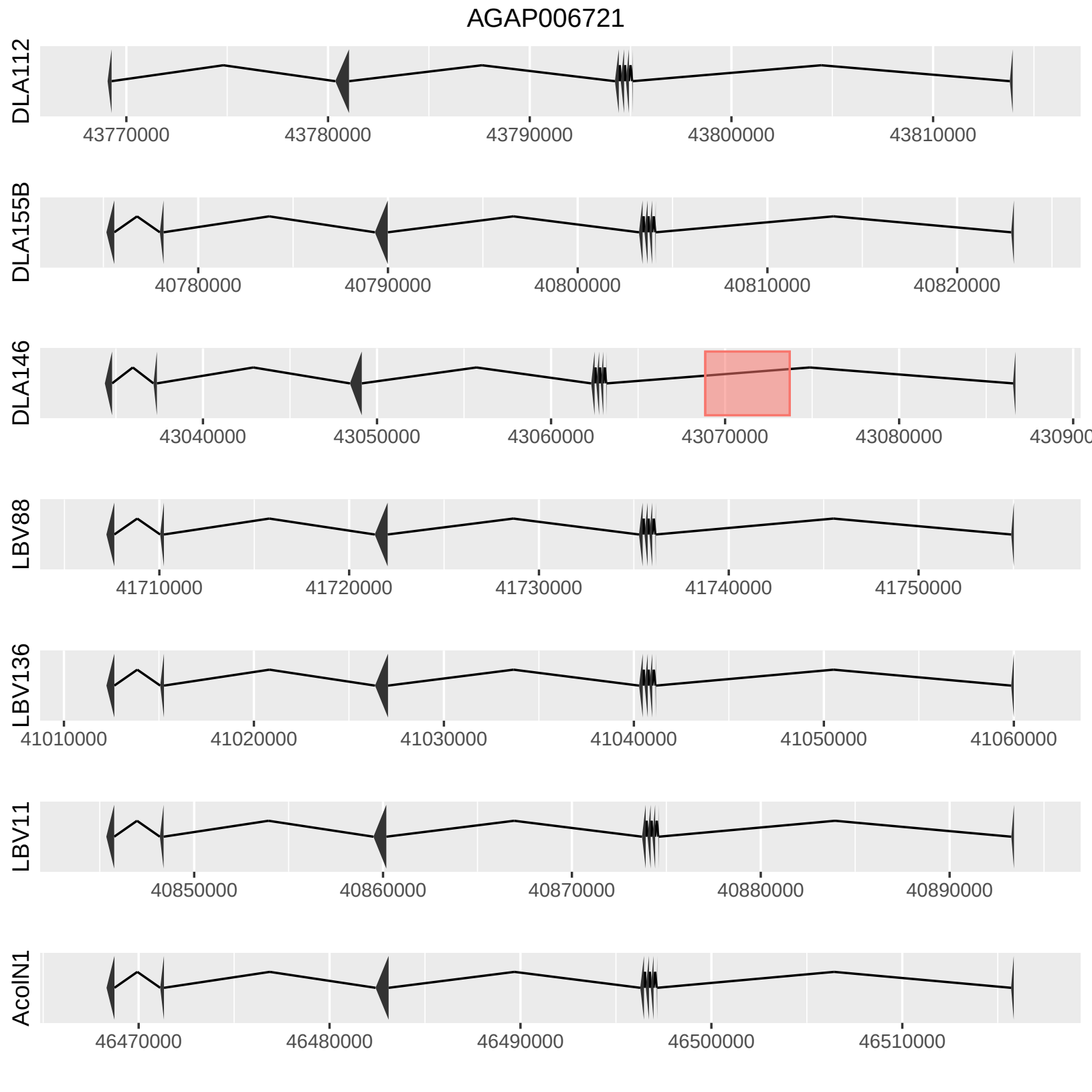

AGAP005464
DLA112
20410000
20415000
20420000
20425000
DLA155B
18210000
18215000
18220000
18225000
DLA146
20135000
20140000
20145000
20150000
LBV88
18545000
18550000
18555000
18560000
18565000
LBV136
19000000
19005000
19010000
19015000
LBV11
18600000
18610000
18620000
AcolN1
24480000
24485000
24490000
24495000

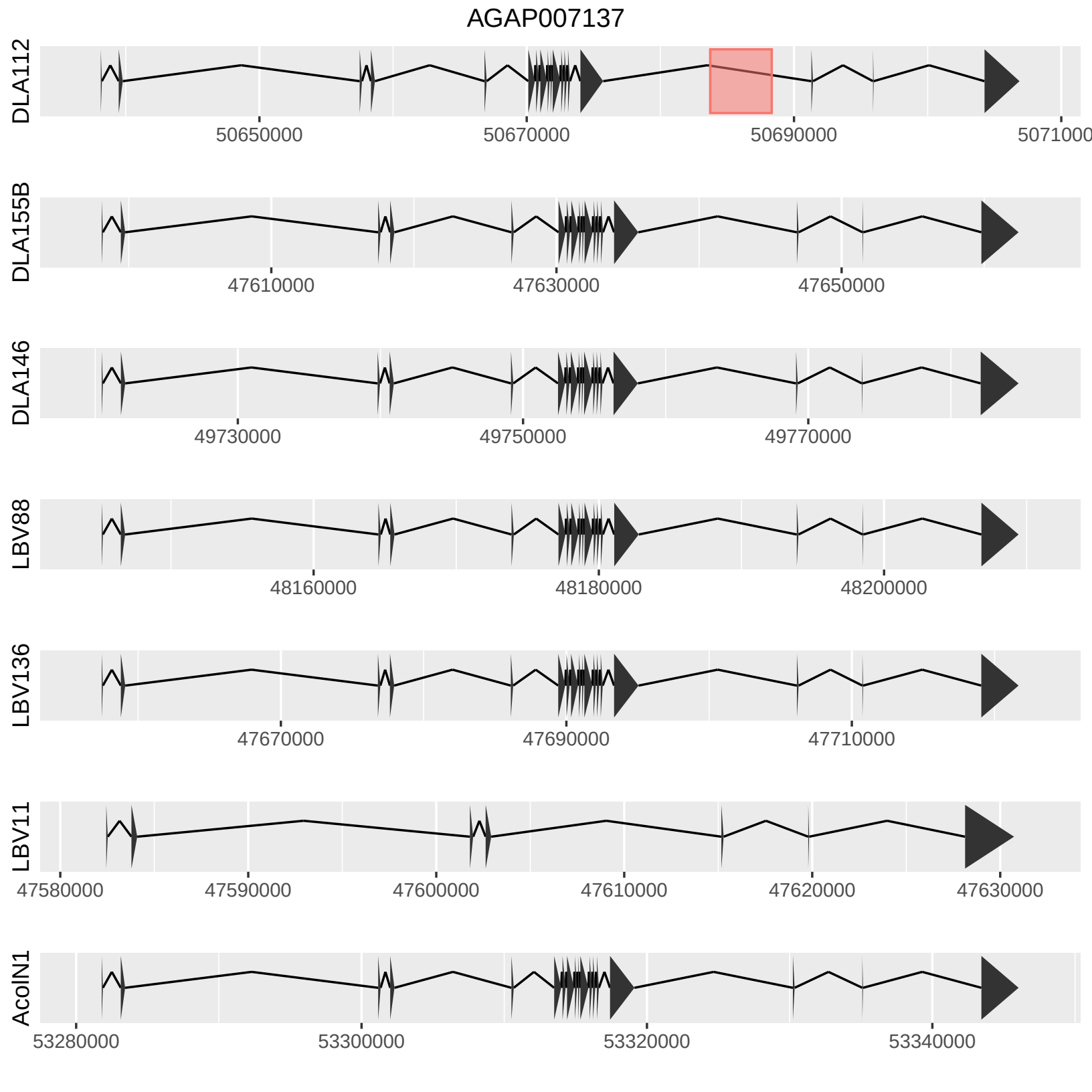

AGAP005773
DLA112
25530000
25540000
25550000
25560000
25570000
25580000
25590000
DLA155B
23090000
23100000
23110000
23120000
23130000
2314000
DLA146
25470000
25480000
25490000
25500000
25510000
LBV88
23500000
23520000
23540000
LBV136
14820000
14840000
14860000
148800
LBV11
23660000
23680000
23700000
AcolN1
29180000
29200000
29220000

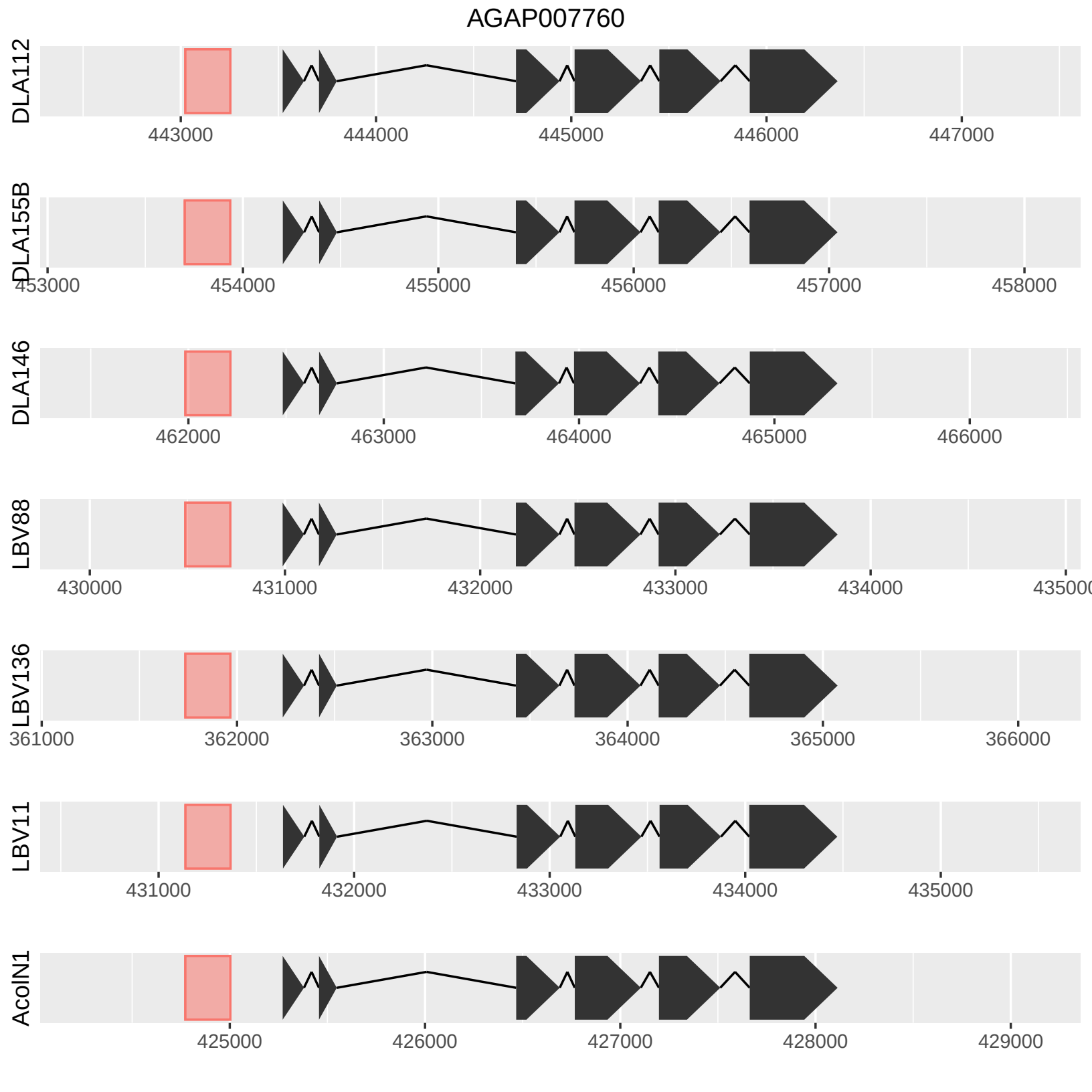

AGAP005806
DLA112
26650000
26675000
26700000
26725000
DLA155B
24060000
24080000
24100000
24120000
24140000
DLA146
26450000
26475000
26500000
LBV88
24525000
24550000
24575000
24600000
LBV136
24375000
24400000
24425000
24450000
LBV11
24700000
24725000
24750000
AcolN1
30100000
30125000
30150000

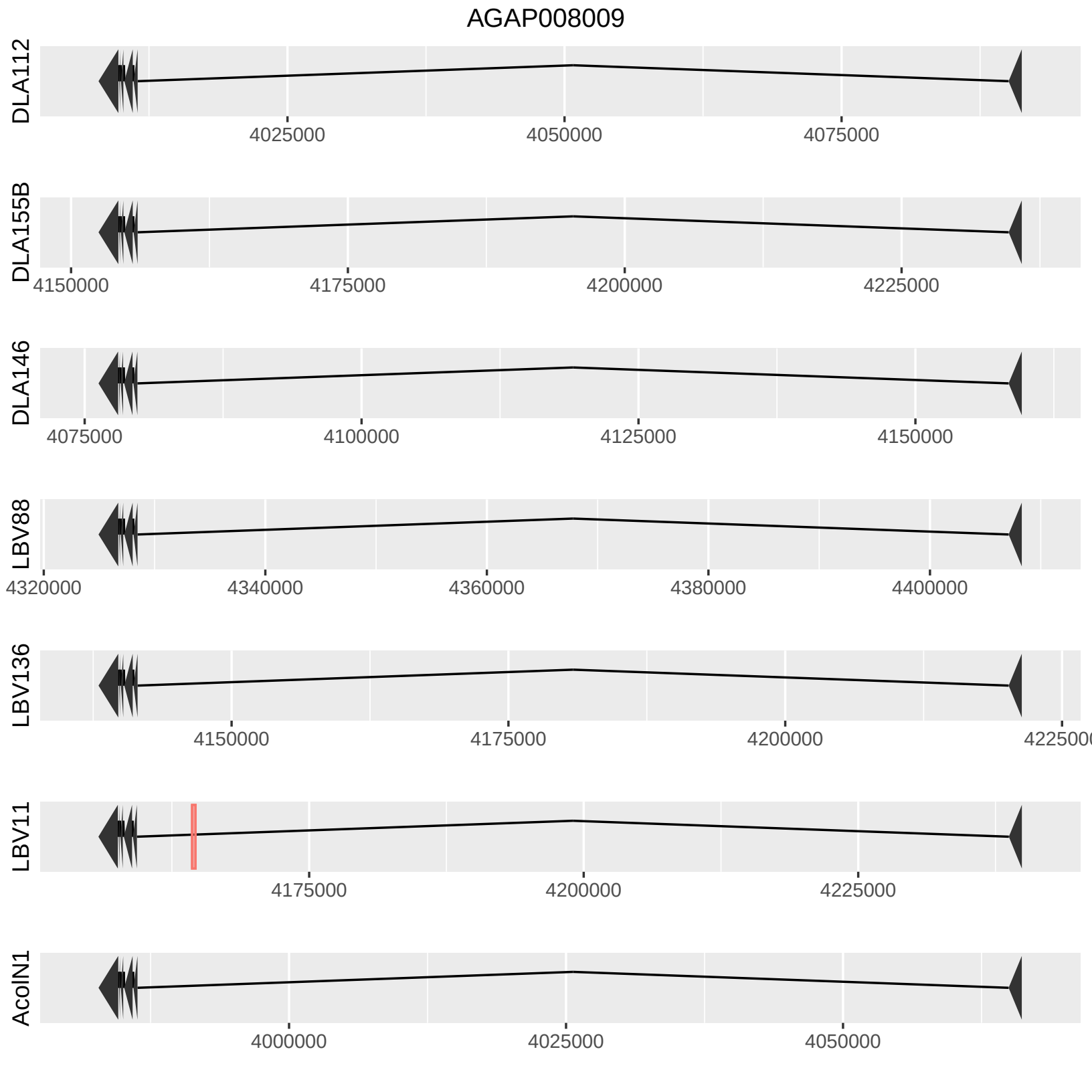

AGAP005816
DLA112
27000000
27050000
27100000
DLA155B
24410000
24430000
24450000
24470000
DLA146
26780000
26800000
26820000
26840000
LBV88
24860000
24880000
24900000
24920000
LBV136
24650000
24675000
24700000
2472500
LBV11
25060000
25080000
25100000
25120000
AcolN1
30460000
30480000
30500000
30520000

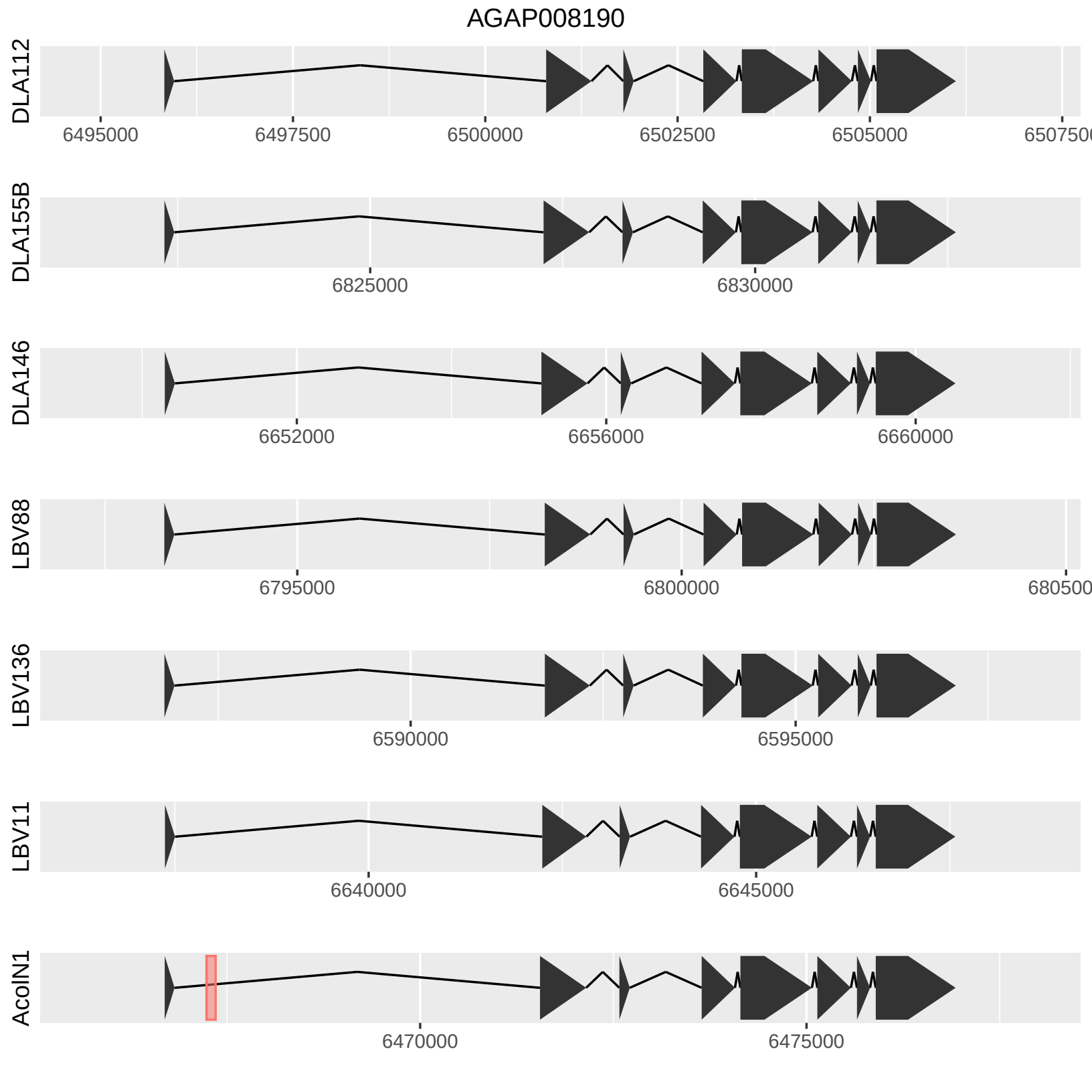

AGAP006330
DLA112
36175000
36200000
36225000
36250000
DLA155B
33275000
33300000
33325000
33350000
DLA146
35720000
35760000
35800000
LBV88
33350000
33375000
33400000
33425000
33450000
LBV136
33500000
33525000
33550000
33575000
LBV11
33450000
33475000
33500000
33525000
33550000
AcoIN1
38680000
38700000
38720000
38740000
38760000

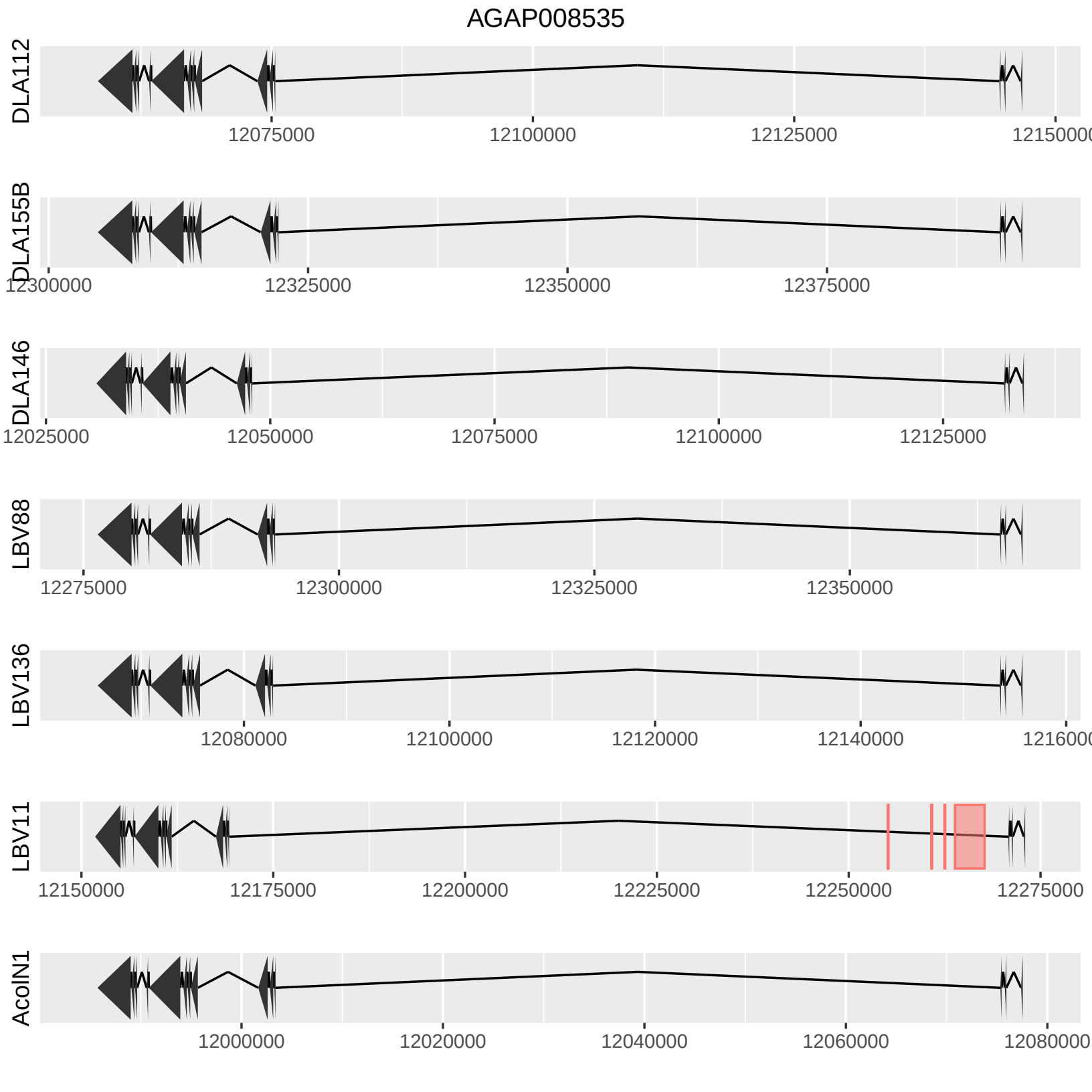

AGAP006721
DLA112
43770000
43780000
43790000
43800000
43810000
DLA155B
40780000
40790000
40800000
40810000
40820000
DLA146
43040000
43050000
43060000
43070000
43080000
LBV88
41710000
41720000
41730000
41740000
41750000
LBV136
41010000
41020000
41030000
41040000
41050000
41060000
LBV11
40850000
40860000
40870000
40880000
40890000
AcolN1
46470000
46480000
46490000
46500000
46510000

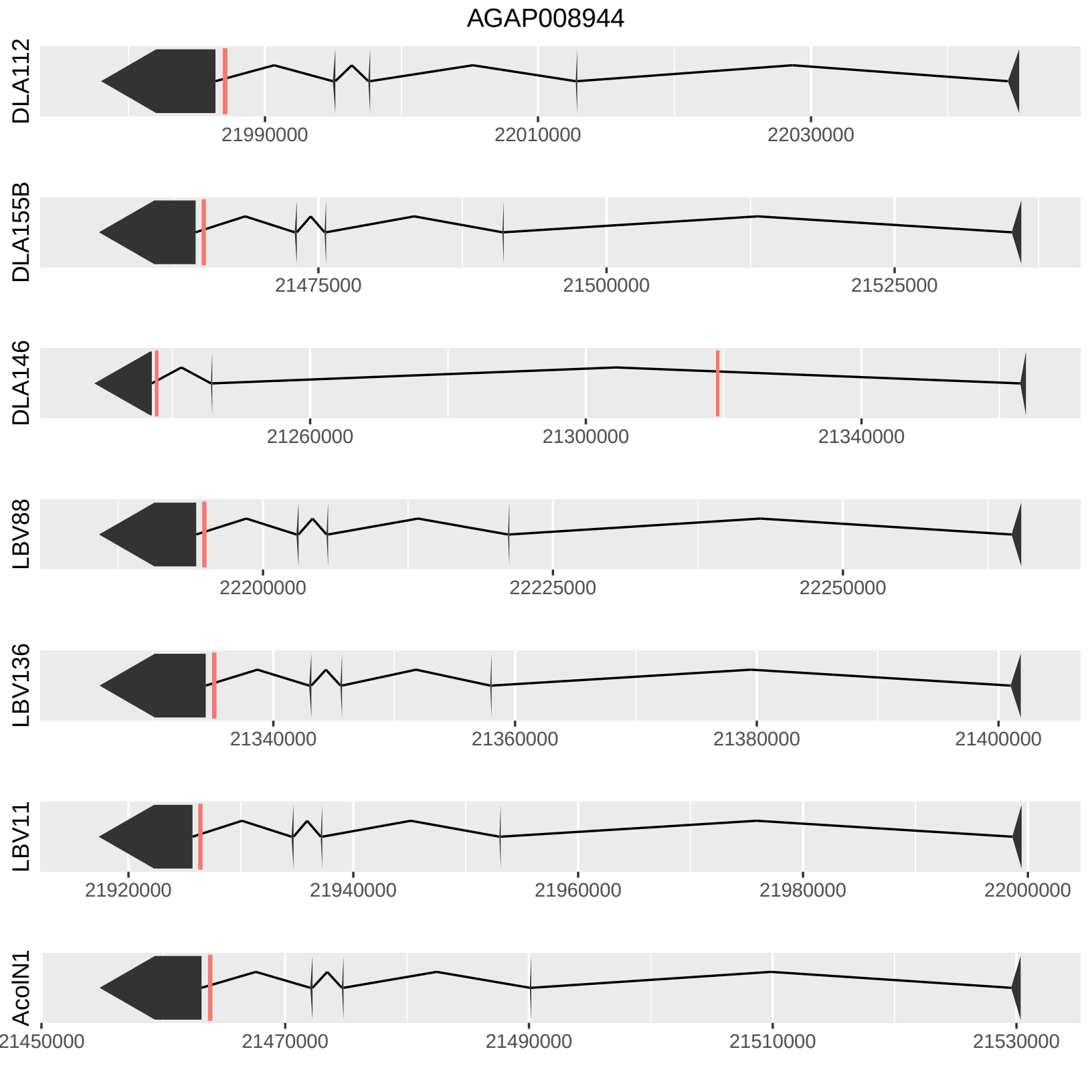

AGAP007137
DLA112
50650000
50670000
50690000
5071000
DLA155B
47610000
47630000
47650000
DLA146
49730000
49750000
49770000
LBV88
48160000
48180000
48200000
LBV136
47670000
47690000
47710000
LBV11
47580000
47590000
47600000
47610000
47620000
47630000
AcolN1
53280000
53300000
53320000
53340000

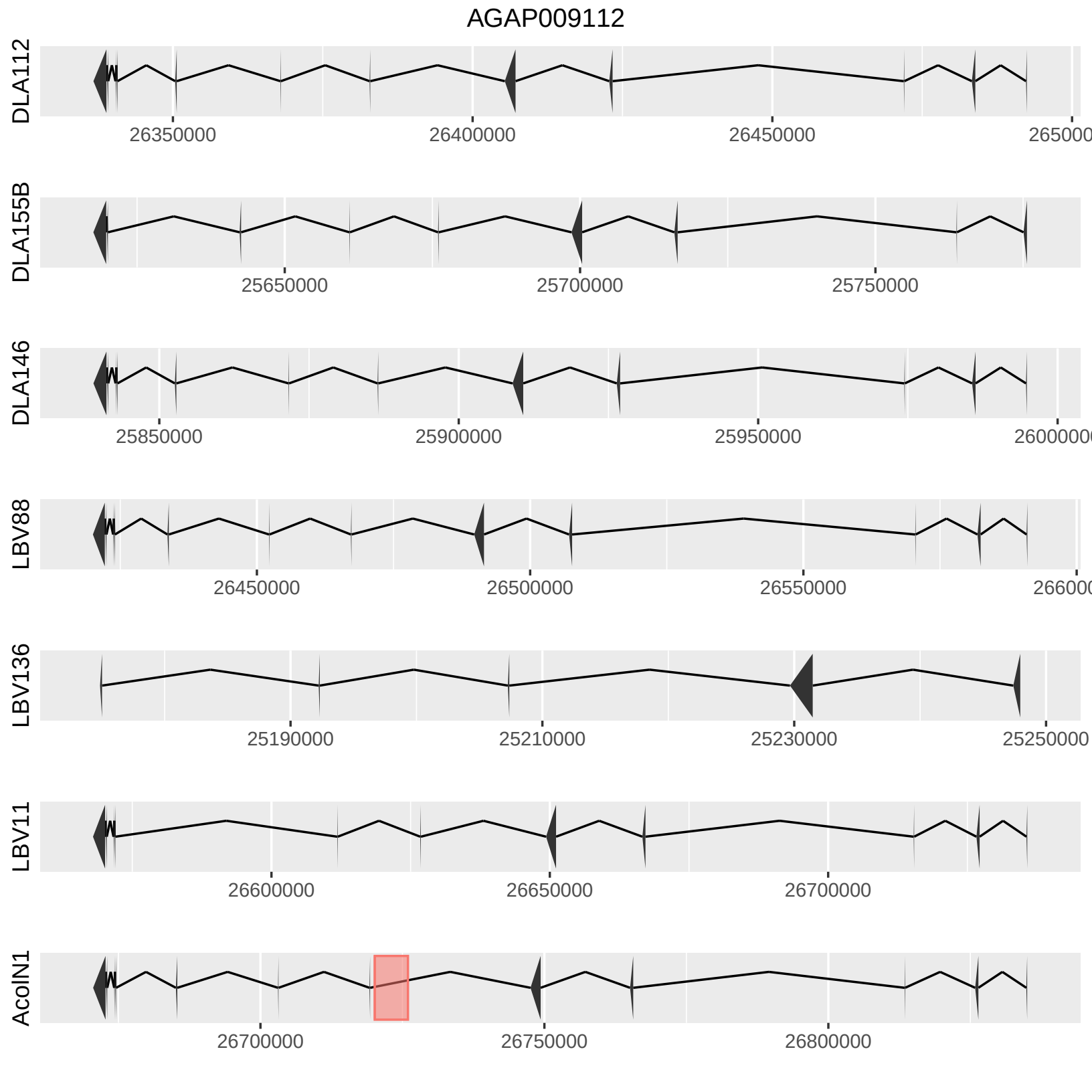

AGAP007760
DLA112
443000
444000
445000
446000
447000
DLA155B
453000
454000
455000
456000
457000
458000
DLA146
462000
463000
464000
465000
466000
LBV88
430000
431000
432000
433000
434000
435000
LBV136
361000
362000
363000
364000
365000
366000
LBV11
431000
432000
433000
434000
435000
AcolN1
425000
426000
427000
428000
429000

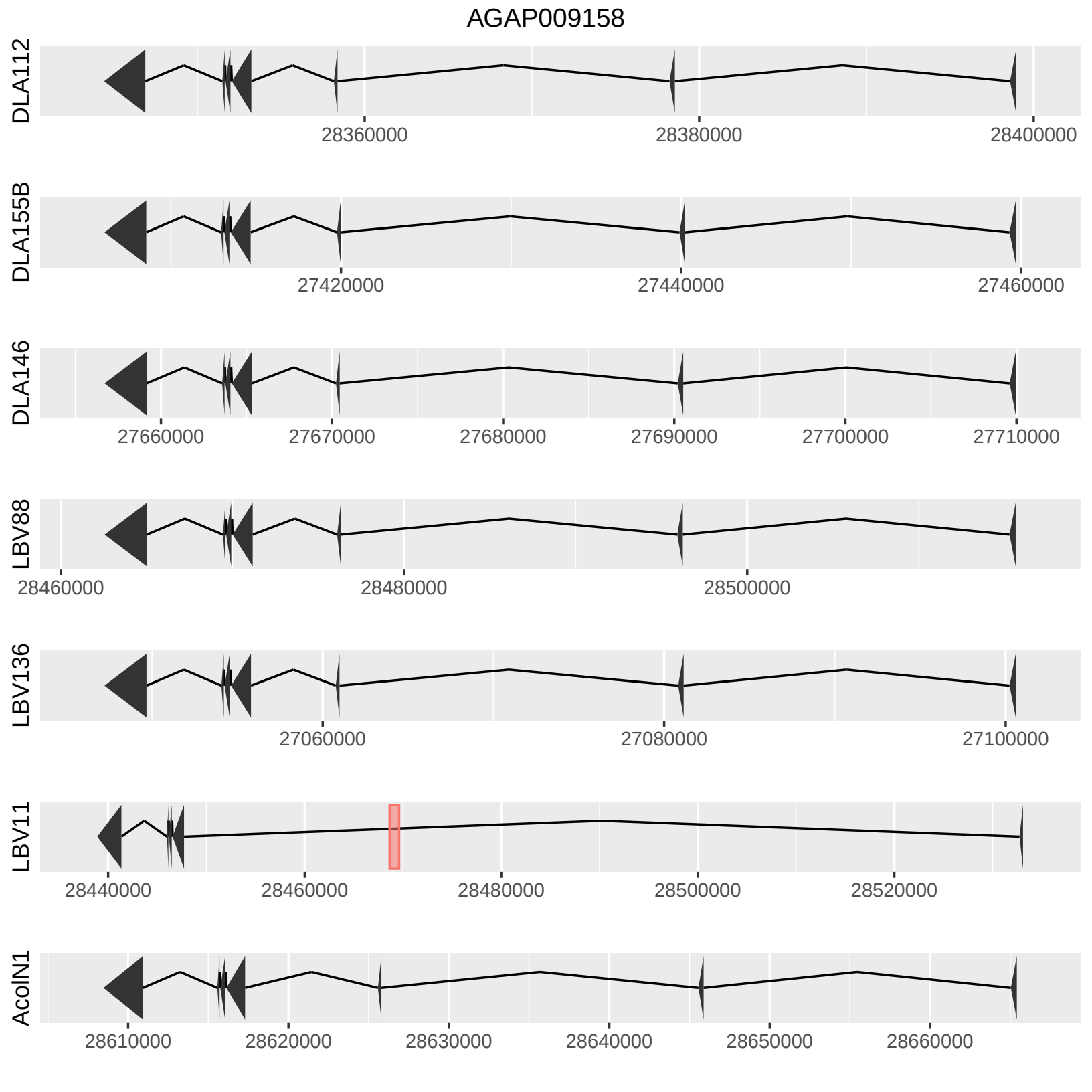

AGAP008009
DLA112
4025000
4050000
4075000
DLA155B
4150000
4175000
4200000
4225000
DLA146
4075000
4100000
4125000
4150000
LBV88
4320000
4340000
4360000
4380000
4400000
LBV136
4150000
4175000
4200000
LBV11
4175000
4200000
4225000
AcoIN1
4000000
4025000
4050000

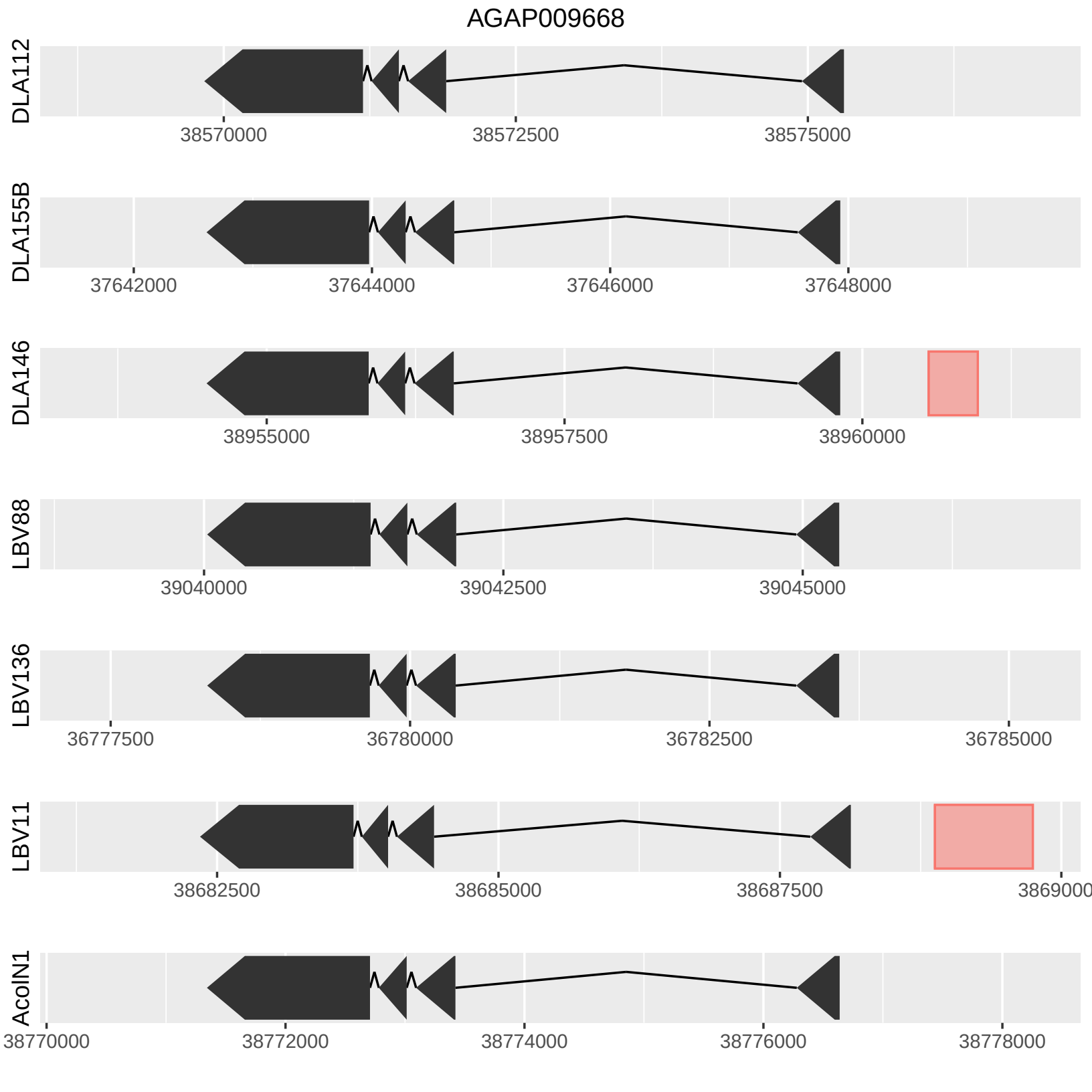

AGAP008190
DLA112
6495000
6497500
6500000
6502500
6505000
DLA155B
6825000
6830000
DLA146
6652000
6656000
6660000
LBV88
6795000
6800000
LBV136
6590000
6595000
LBV11
6640000
6645000
AcolN1
6470000
6475000

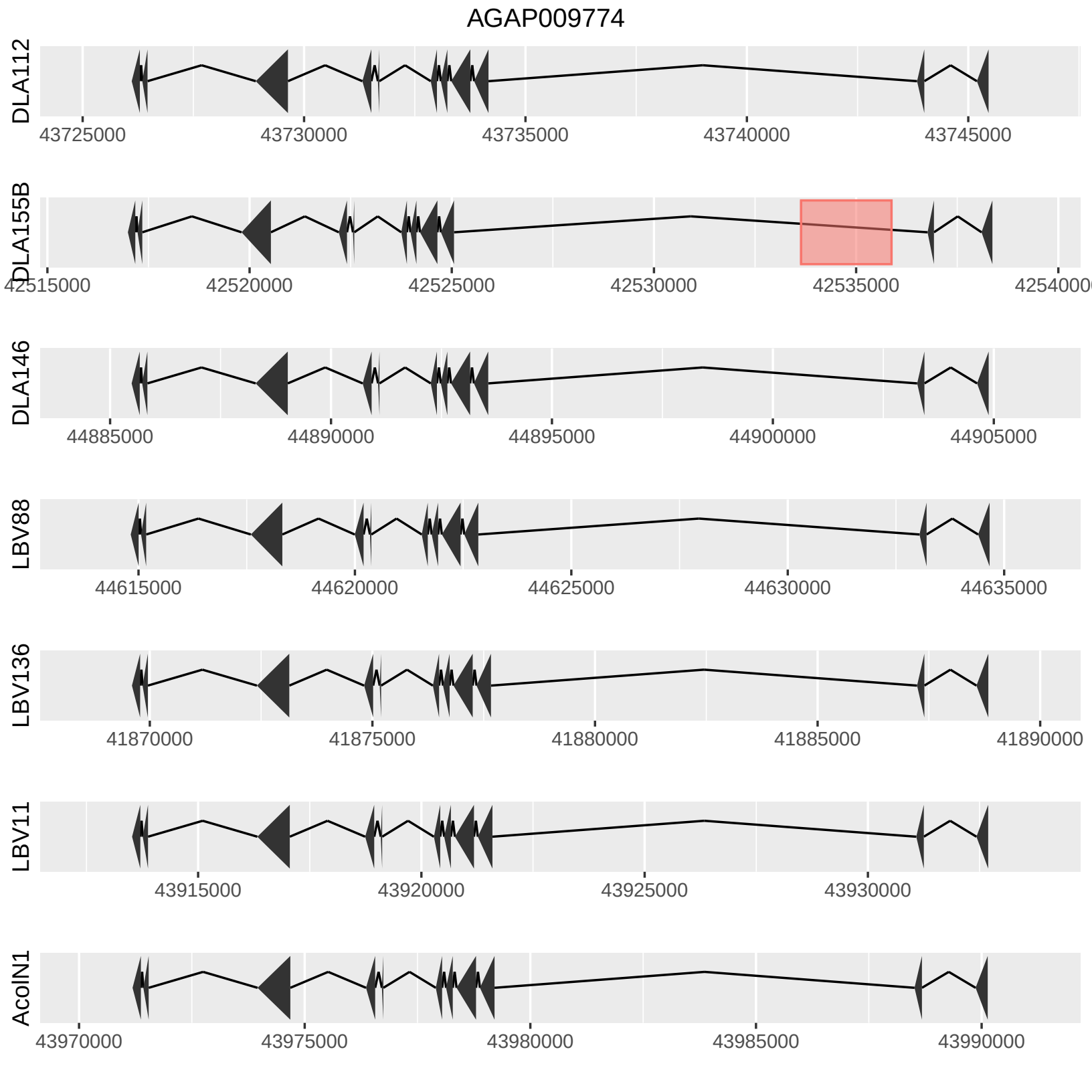

AGAP008535
DLA112
12075000
12100000
12125000
12150000
DLA155B
12300000
12325000
12350000
12375000
DLA146
12025000
12050000
12075000
12100000
12125000
LBV88
12275000
12300000
12325000
12350000
LBV136
12080000
12100000
12120000
12140000
12160000
LBV11
12150000
12175000
12200000
12225000
12250000
12275000
AcolN1
12000000
12020000
12040000
12060000
12080000

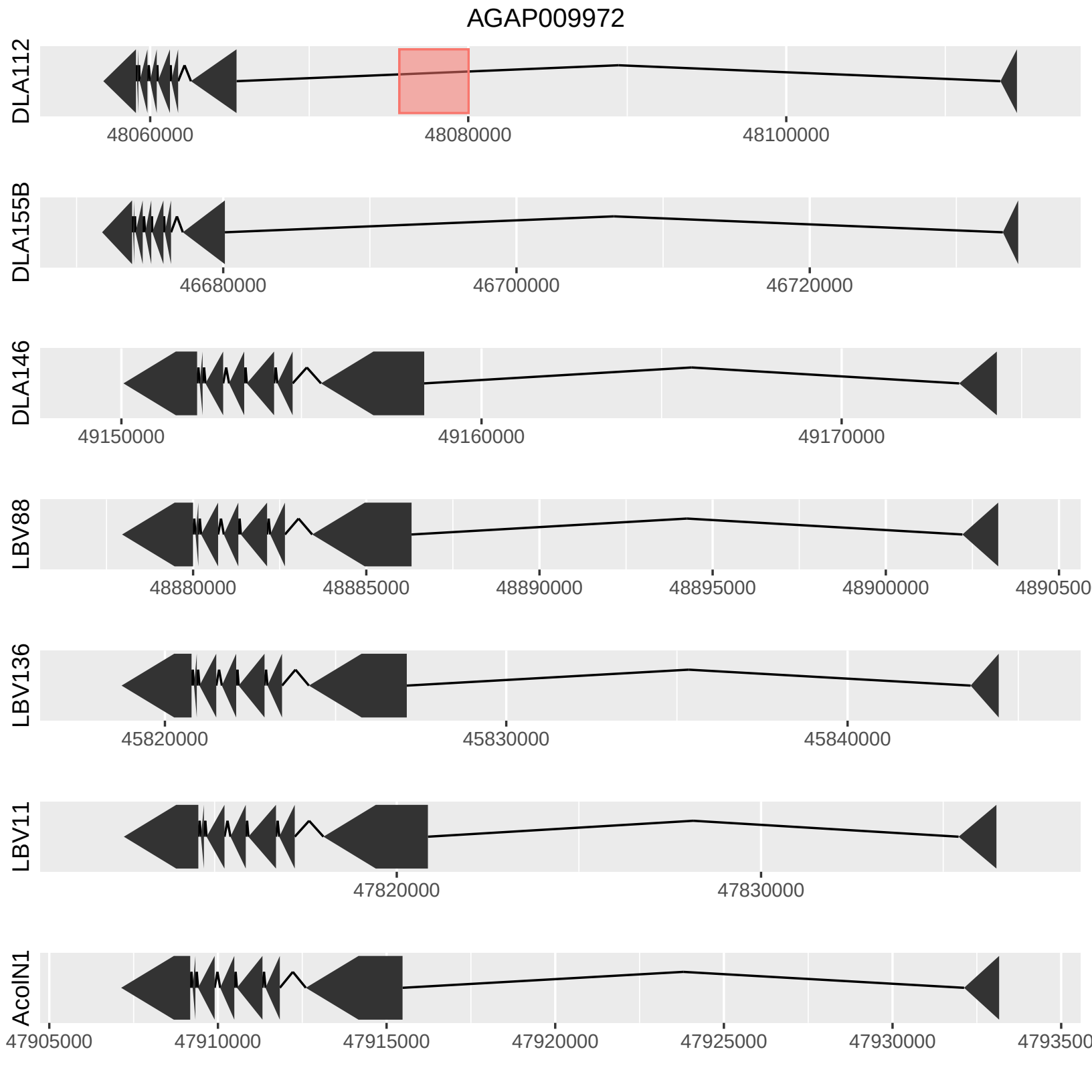

AGAP008944
DLA112
21990000
22010000
22030000
DLA155B
21475000
21500000
21525000
DLA146
21260000
21300000
21340000
LBV88
22200000
22225000
22250000
LBV136
21340000
21360000
21380000
21400000
LBV11
21920000
21940000
21960000
21980000
22000000
AcolN1
21450000
21470000
21490000
21510000
21530000

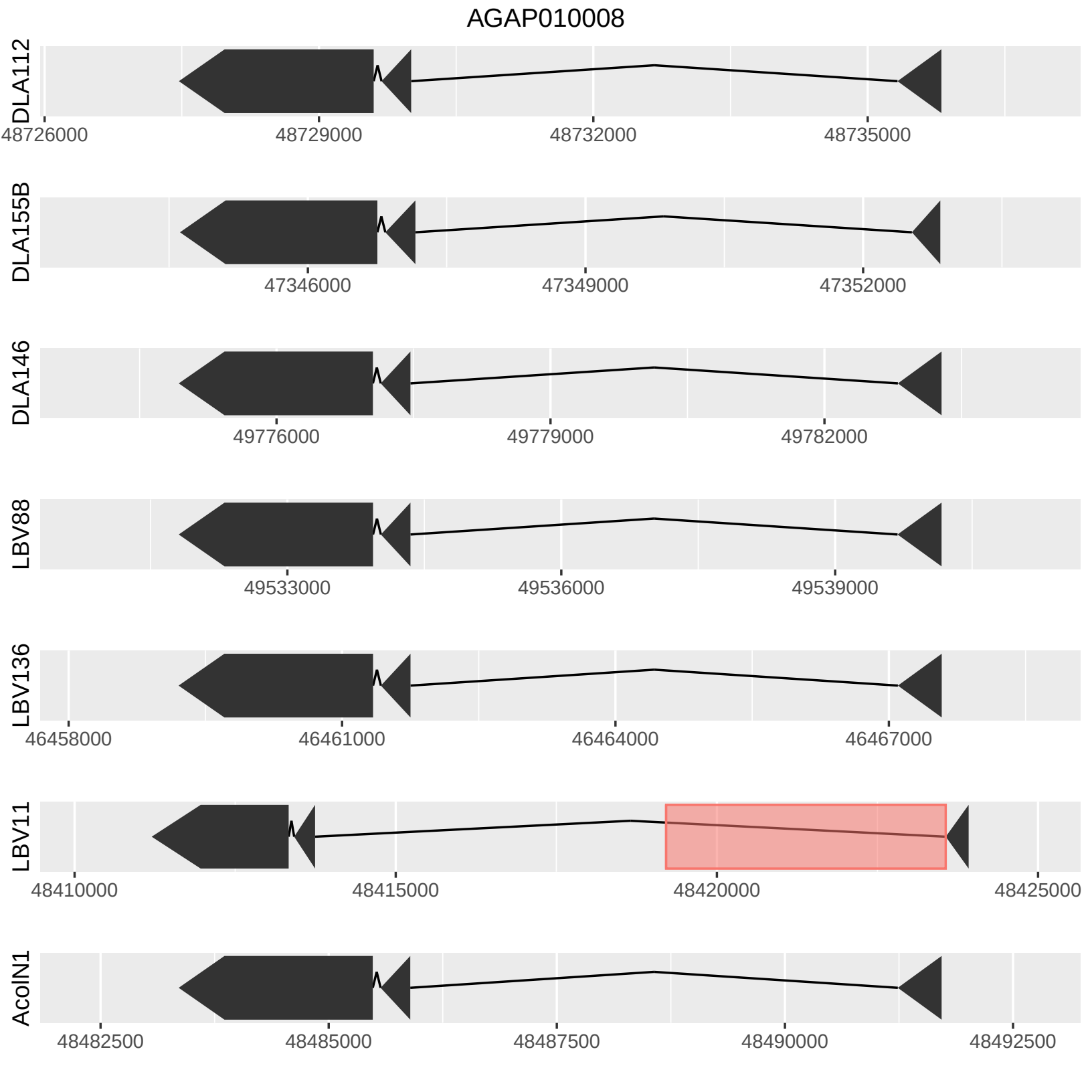

AGAP009112
DLA112
26350000
26400000
26450000
265000
DLA155B
25650000
25700000
25750000
DLA146
25850000
25900000
25950000
2600000
LBV88
26450000
26500000
26550000
26600
LBV136
25190000
25210000
25230000
25250000
LBV11
26600000
26650000
26700000
AcolN1
26700000
26750000
26800000

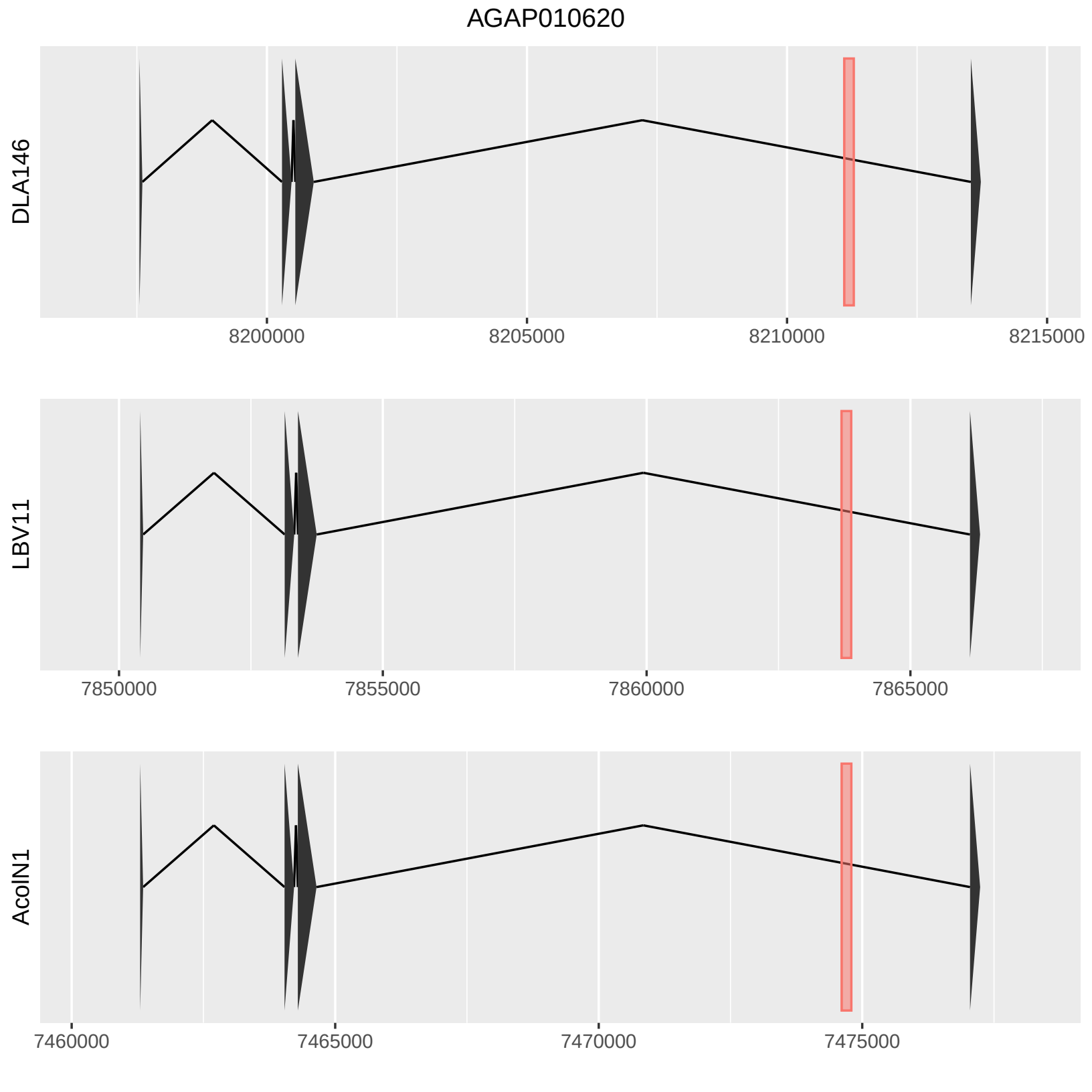

AGAP009158
DLA112
28360000
28380000
28400000
DLA155B
27420000
27440000
27460000
DLA146
27660000
27670000
27680000
27690000
27700000
27710000
LBV88
28460000
28480000
28500000
LBV136
27060000
27080000
27100000
LBV11
28440000
28460000
28480000
28500000
28520000
AcolN1
28610000
28620000
28630000
28640000
28650000
28660000

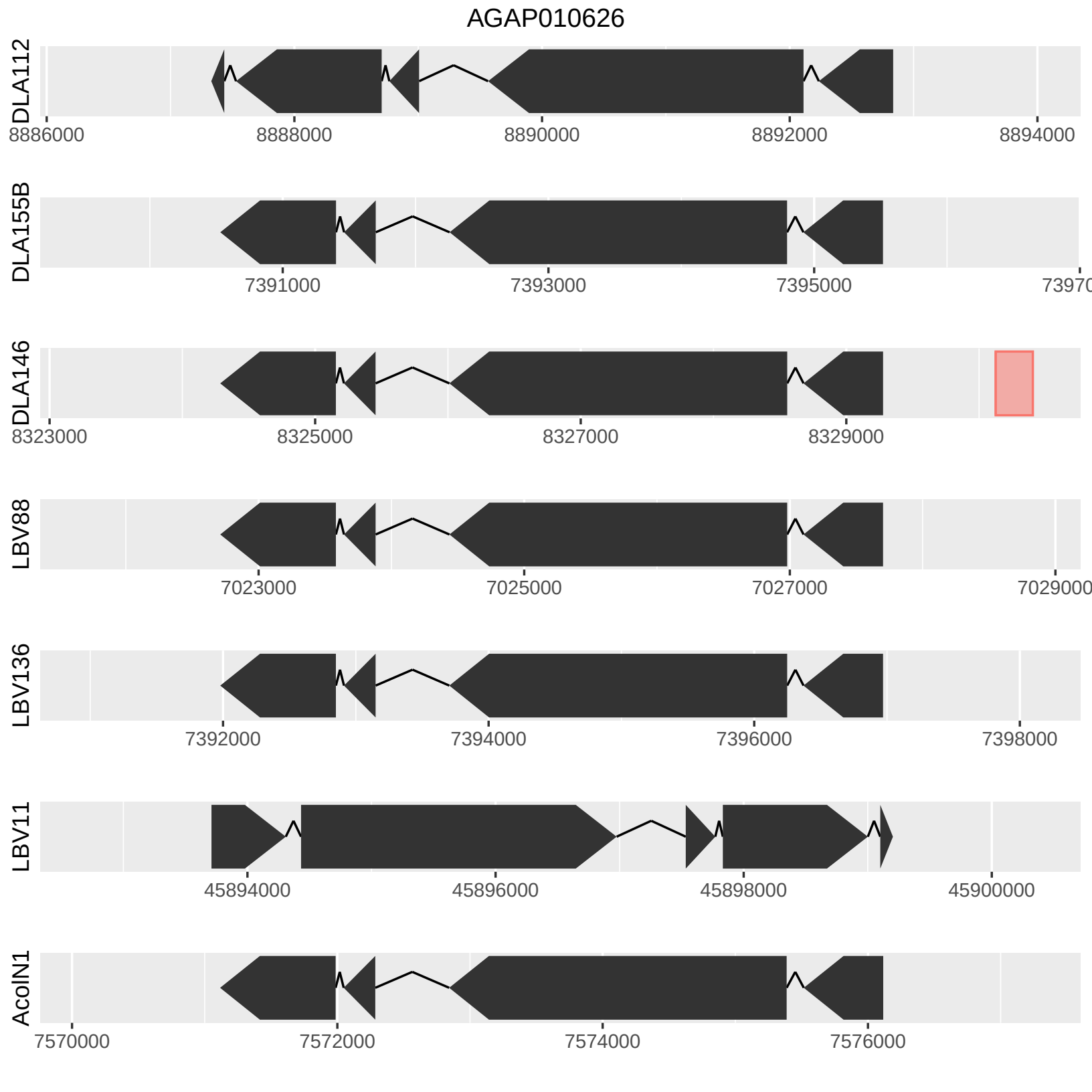

AGAP009668
DLA112
38570000
38572500
38575000
DLA155B
37642000
37644000
37646000
37648000
DLA146
38955000
38957500
38960000
LBV88
39040000
39042500
39045000
LBV136
36777500
36780000
36782500
36785000
LBV11
38682500
38685000
38687500
3869000
AcoIN1
38770000
38772000
38774000
38776000
38778000

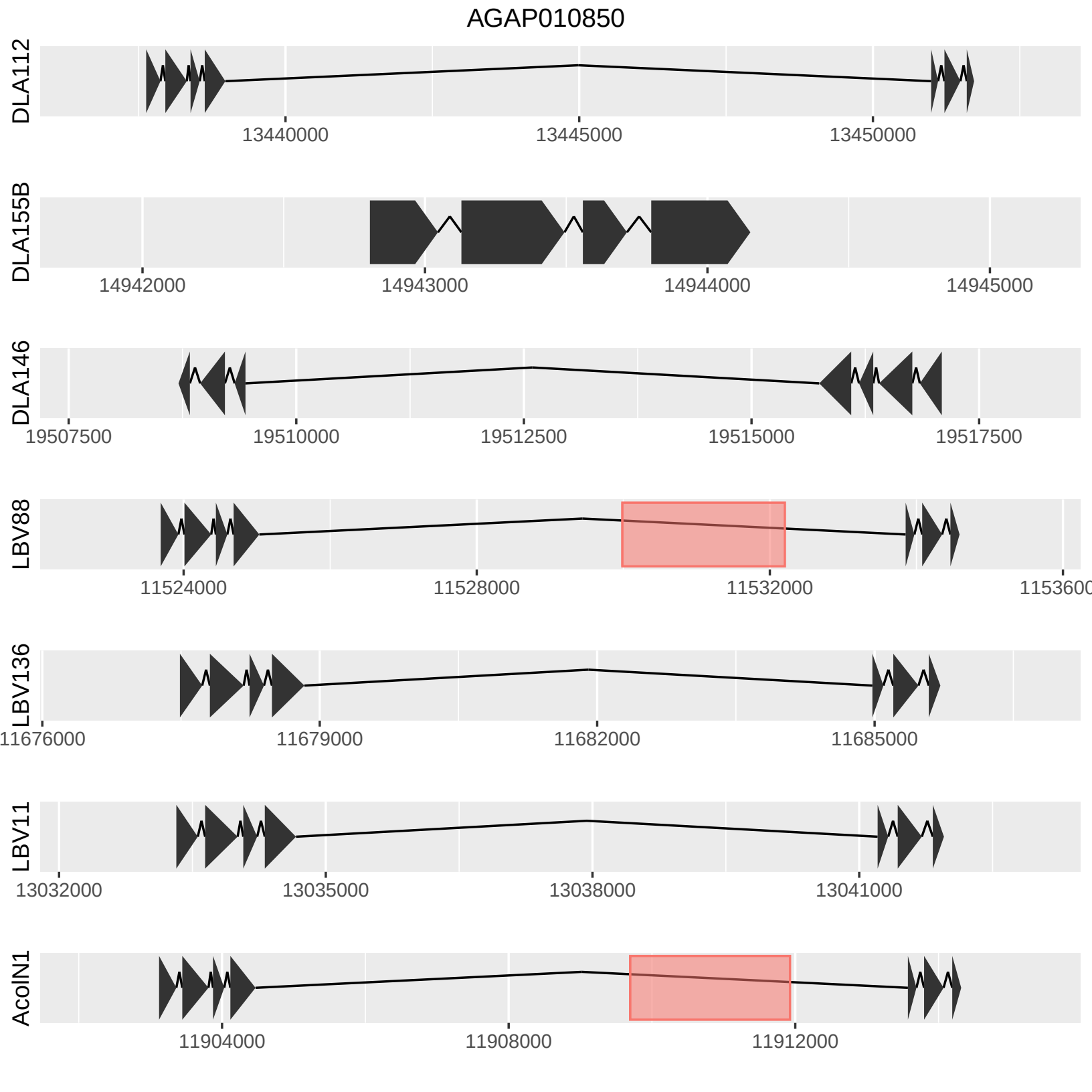

AGAP009774
DLA112
43725000
43730000
43735000
43740000
43745000
DLA155B
42515000
42520000
42525000
42530000
42535000
4254000
DLA146
44885000
44890000
44895000
44900000
44905000
LBV88
44615000
44620000
44625000
44630000
44635000
LBV136
41870000
41875000
41880000
41885000
41890000
LBV11
43915000
43920000
43925000
43930000
AcolN1
43970000
43975000
43980000
43985000
43990000

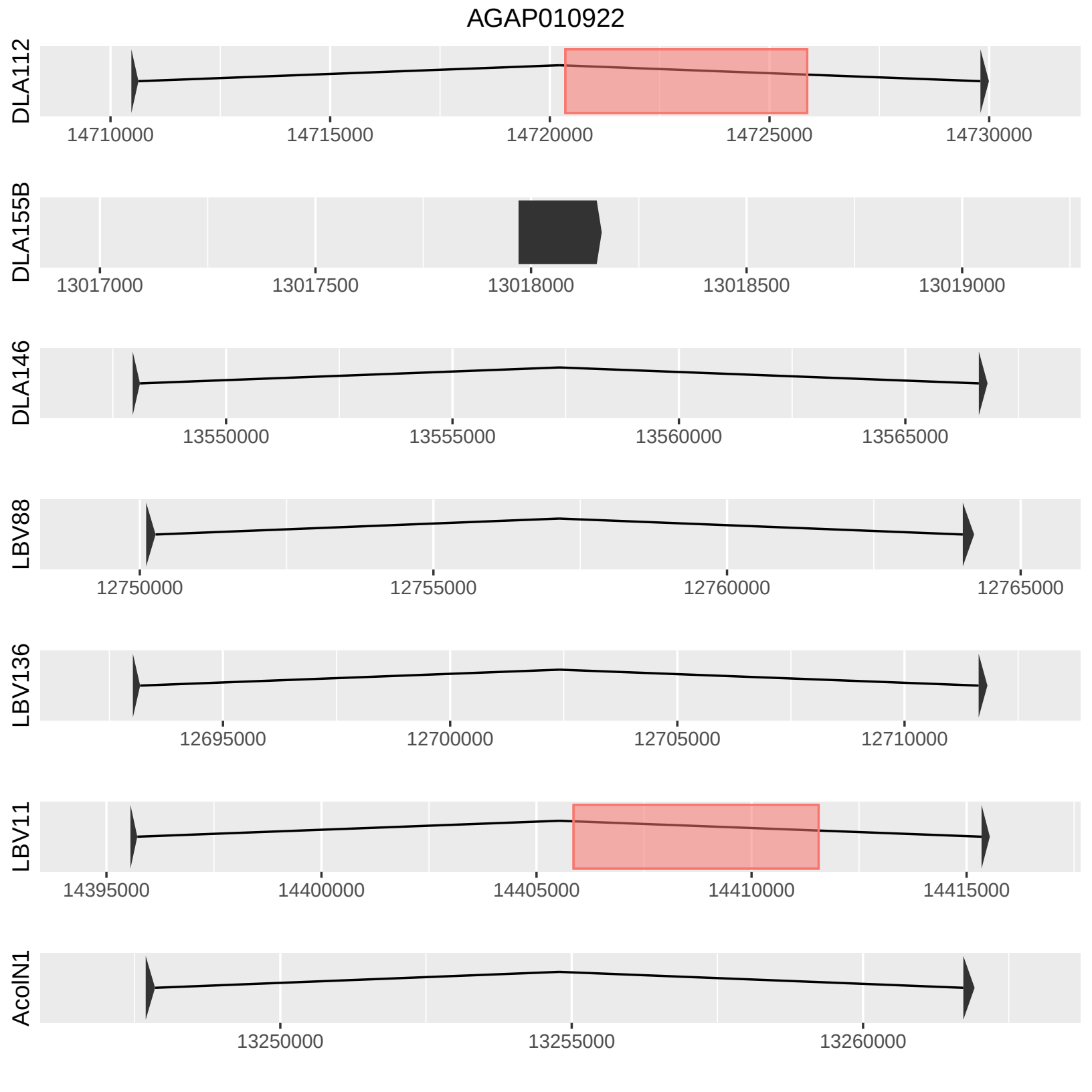

AGAP009972
DLA112
48060000
48080000
48100000
DLA155B
46680000
46700000
46720000
DLA146
49150000
49160000
49170000
LBV88
48880000
48885000
48890000
48895000
48900000
4890500
LBV136
45820000
45830000
45840000
LBV11
47820000
47830000
AcolN1
47905000
47910000
47915000
47920000
47925000
47930000
4793500

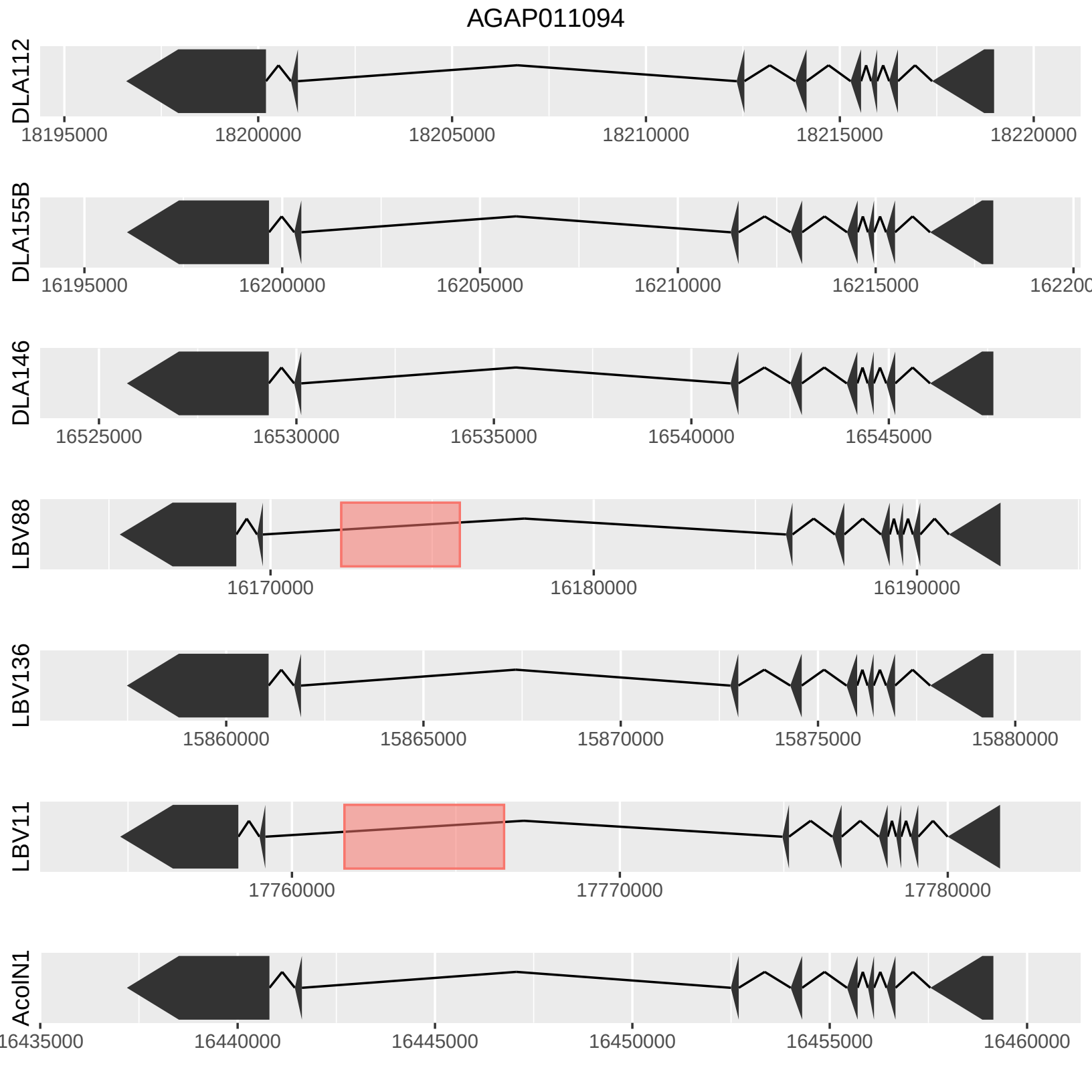

AGAP010008
DLA112
48726000
48729000
48732000
48735000
DLA155B
47346000
47349000
47352000
DLA146
49776000
49779000
49782000
LBV88
49533000
49536000
49539000
LBV136
46458000
46461000
46464000
46467000
LBV11
48410000
48415000
48420000
48425000
AcoIN1
48482500
48485000
48487500
48490000
48492500

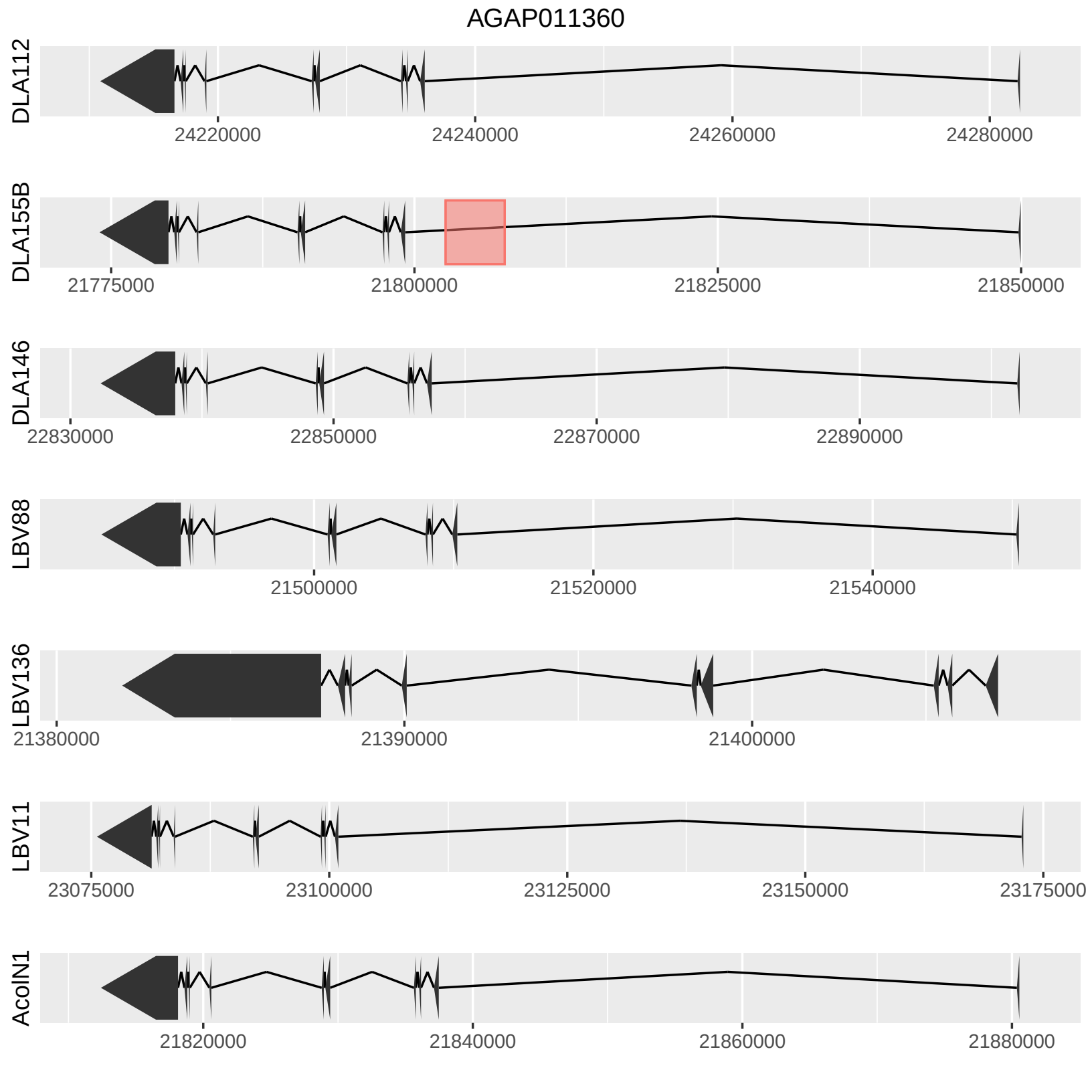

AGAP010620
DLA146
8200000
8205000
8210000
8215000
LBV11
7850000
7855000
7860000
7865000
AcolN1
7460000
7465000
7470000
7475000

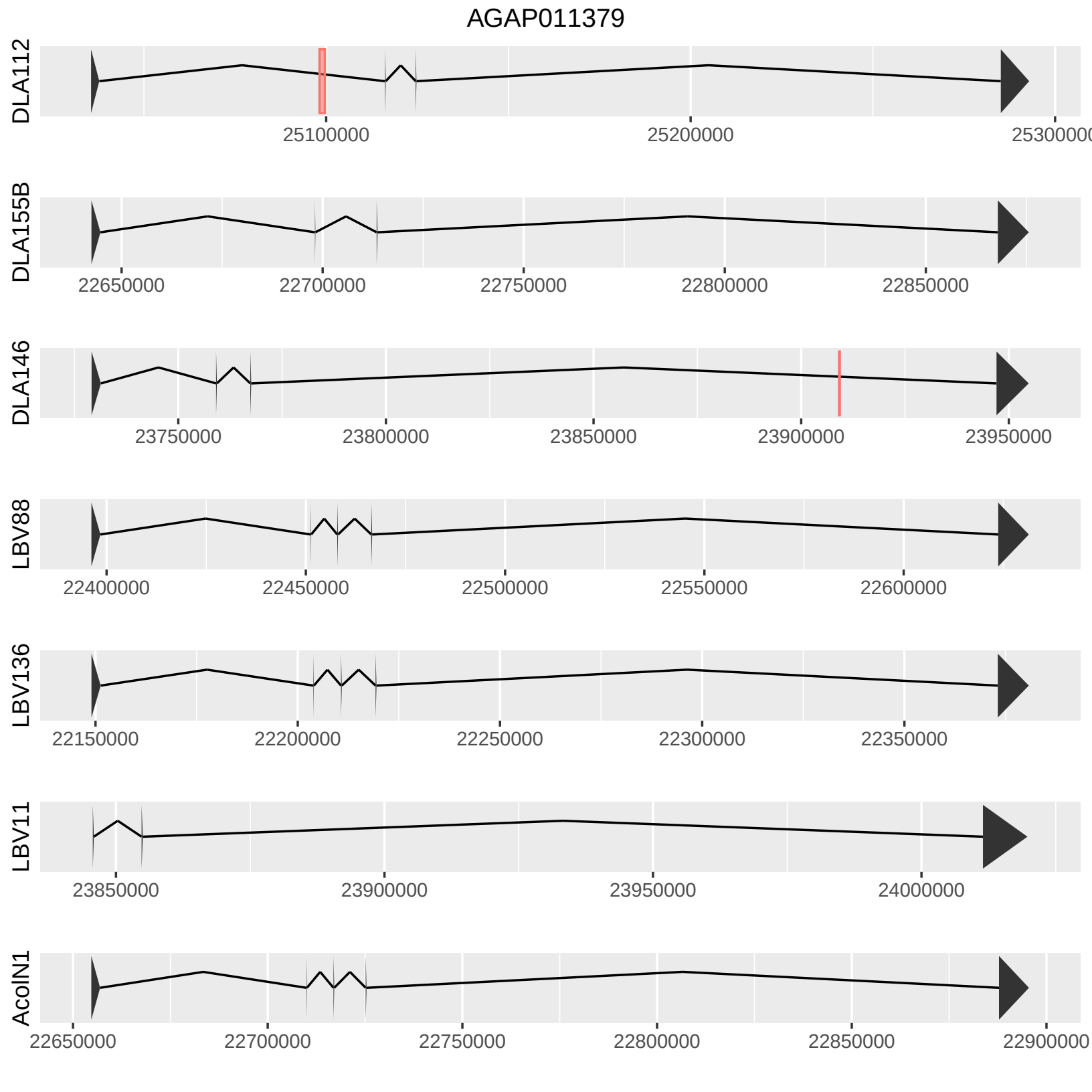

AGAP010626
DLA112
8886000
8888000
8890000
8892000
8894000
DLA155B
7391000
7393000
7395000
DLA146
8323000
8325000
8327000
8329000
LBV88
7023000
7025000
7027000
7029000
LBV136
7392000
7394000
7396000
7398000
LBV11
45894000
45896000
45898000
45900000
AcoIN1
7570000
7572000
7574000
7576000

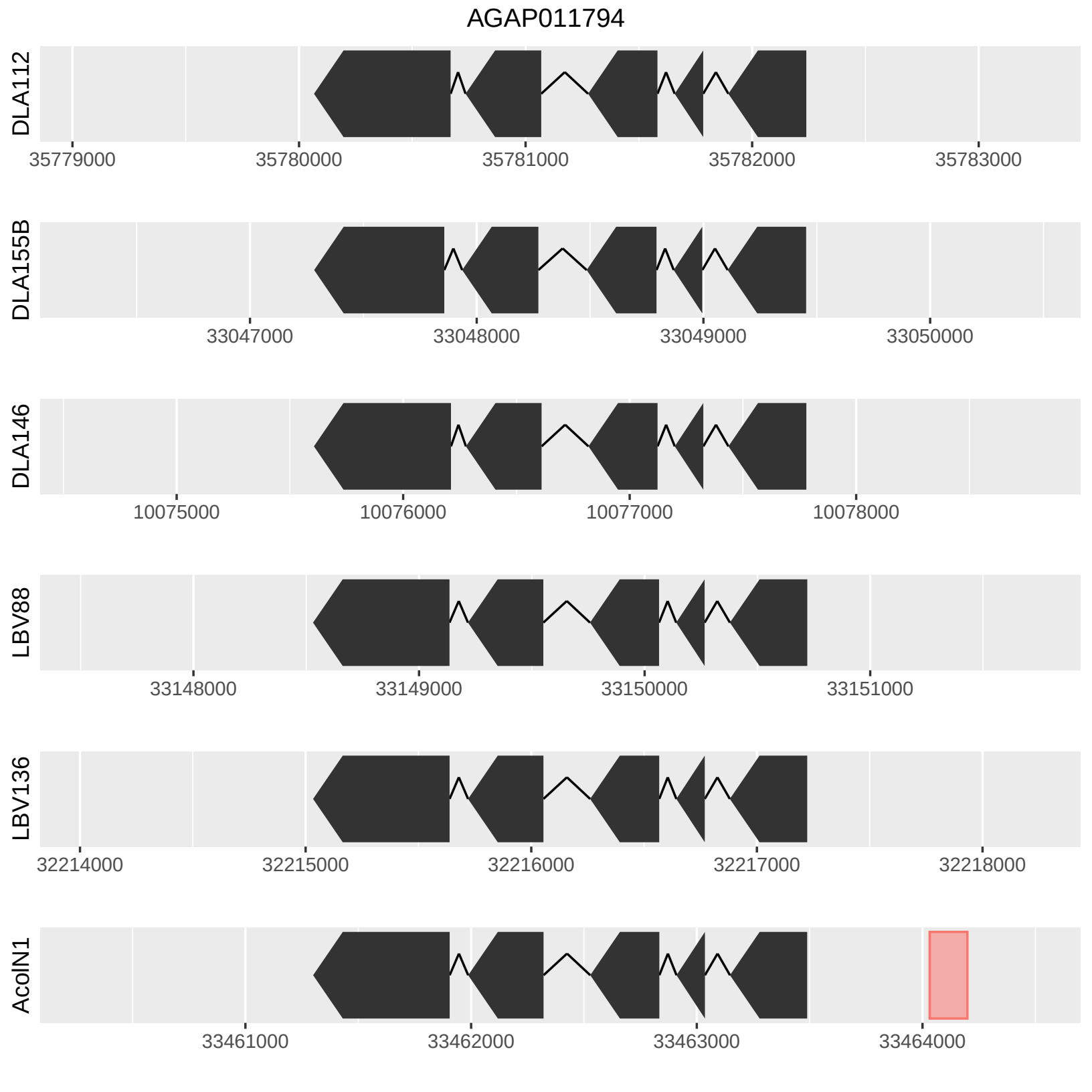

AGAP010850
DLA112
13440000
13445000
13450000
DLA155B
14942000
14943000
14944000
14945000
DLA146
19507500
19510000
19512500
19515000
19517500
LBV88
11524000
11528000
11532000
LBV136
11676000
11679000
11682000
11685000
LBV11
13032000
13035000
13038000
13041000
AcolN1
11904000
11908000
11912000

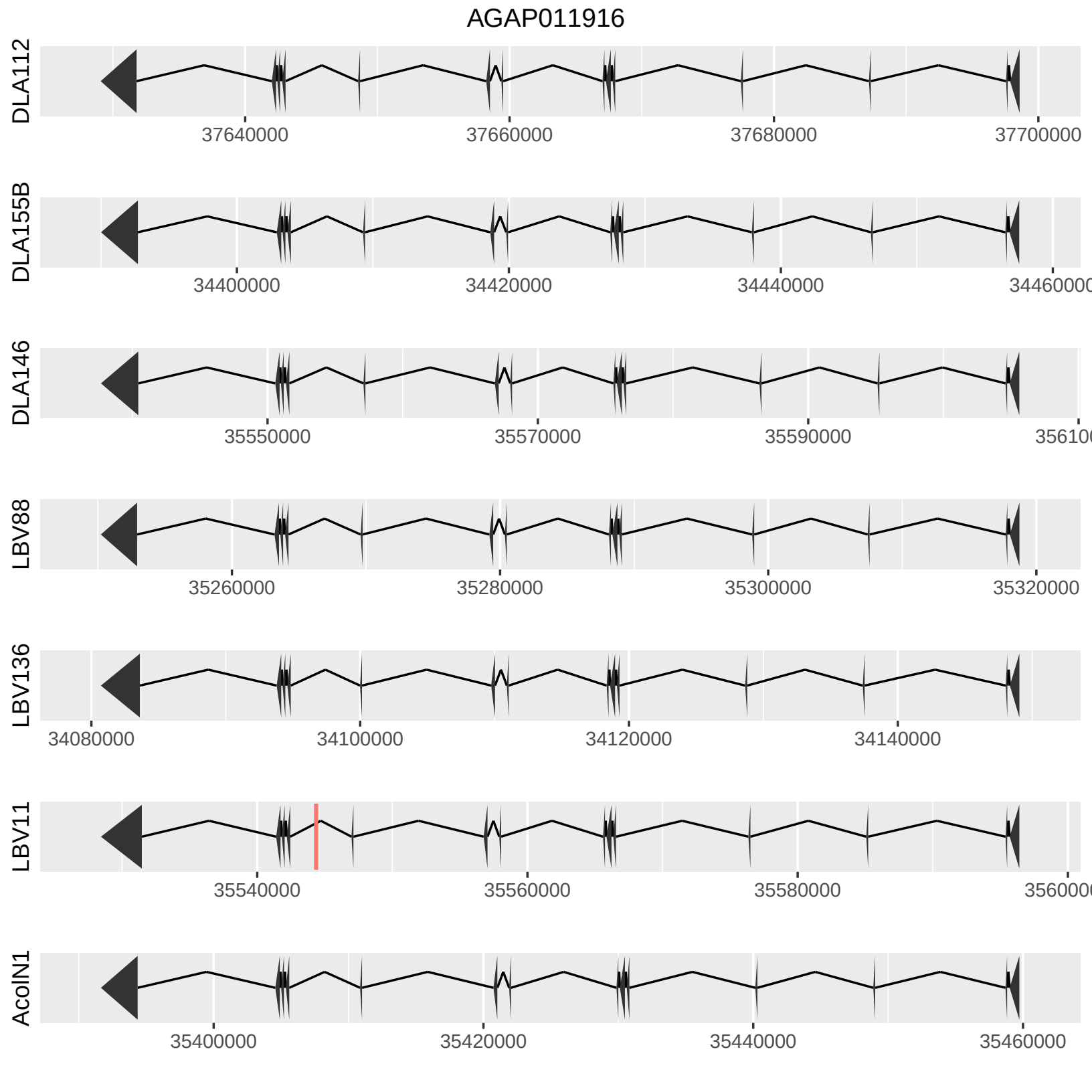

AGAP010922
DLA112
14710000
14715000
14720000
14725000
14730000
DLA155B
13017000
13017500
13018000
13018500
13019000
DLA146
13550000
13555000
13560000
13565000
LBV88
12750000
12755000
12760000
12765000
LBV136
12695000
12700000
12705000
12710000
LBV11
14395000
14400000
14405000
14410000
14415000
AcolN1
13250000
13255000
13260000

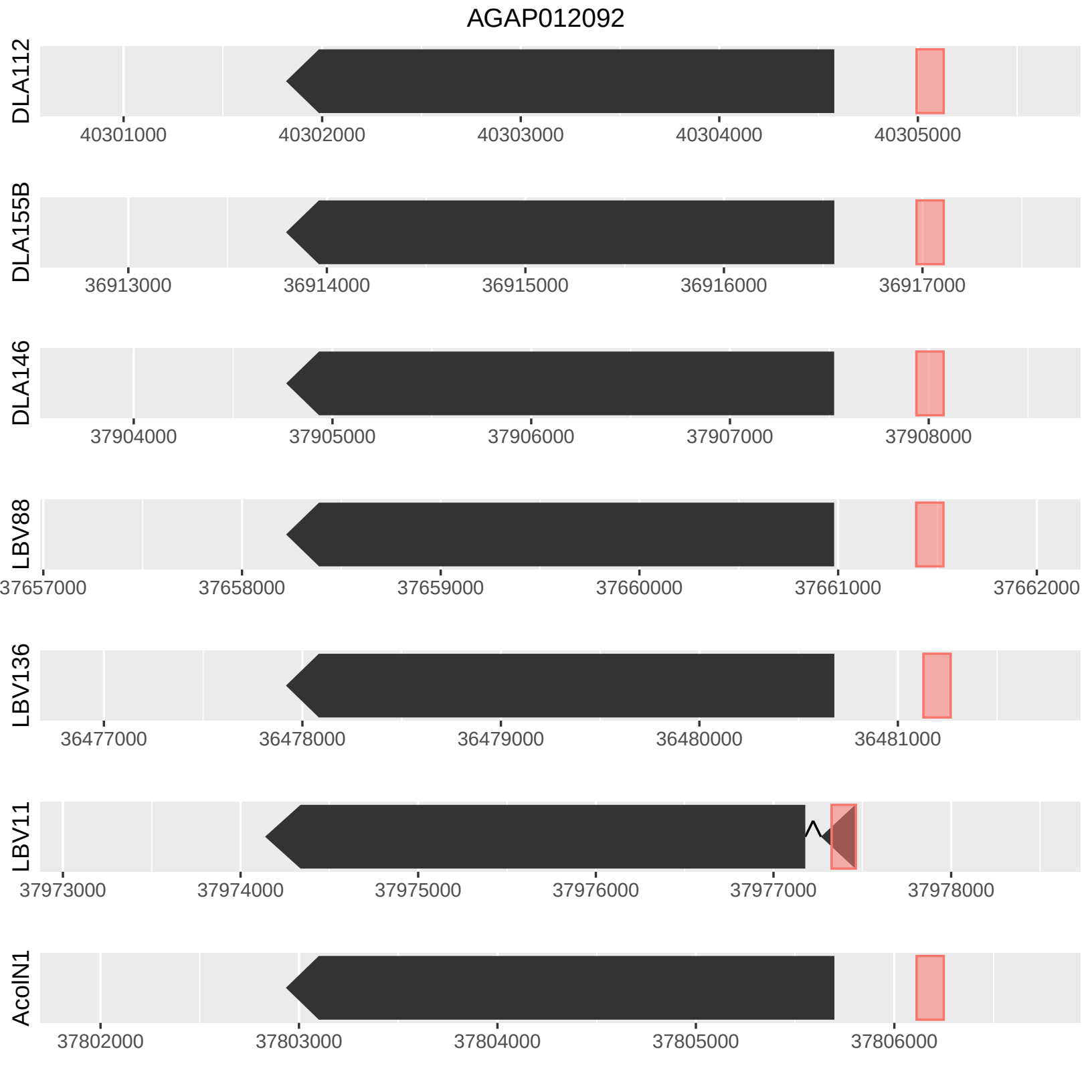

AGAP011094
DLA112
18195000
18200000
18205000
18210000
18215000
18220000
DLA155B
16195000
16200000
16205000
16210000
16215000
162200
DLA146
16525000
16530000
16535000
16540000
16545000
LBV88
16170000
16180000
16190000
LBV136
15860000
15865000
15870000
15875000
15880000
LBV11
17760000
17770000
17780000
AcolN1
16435000
16440000
16445000
16450000
16455000
16460000

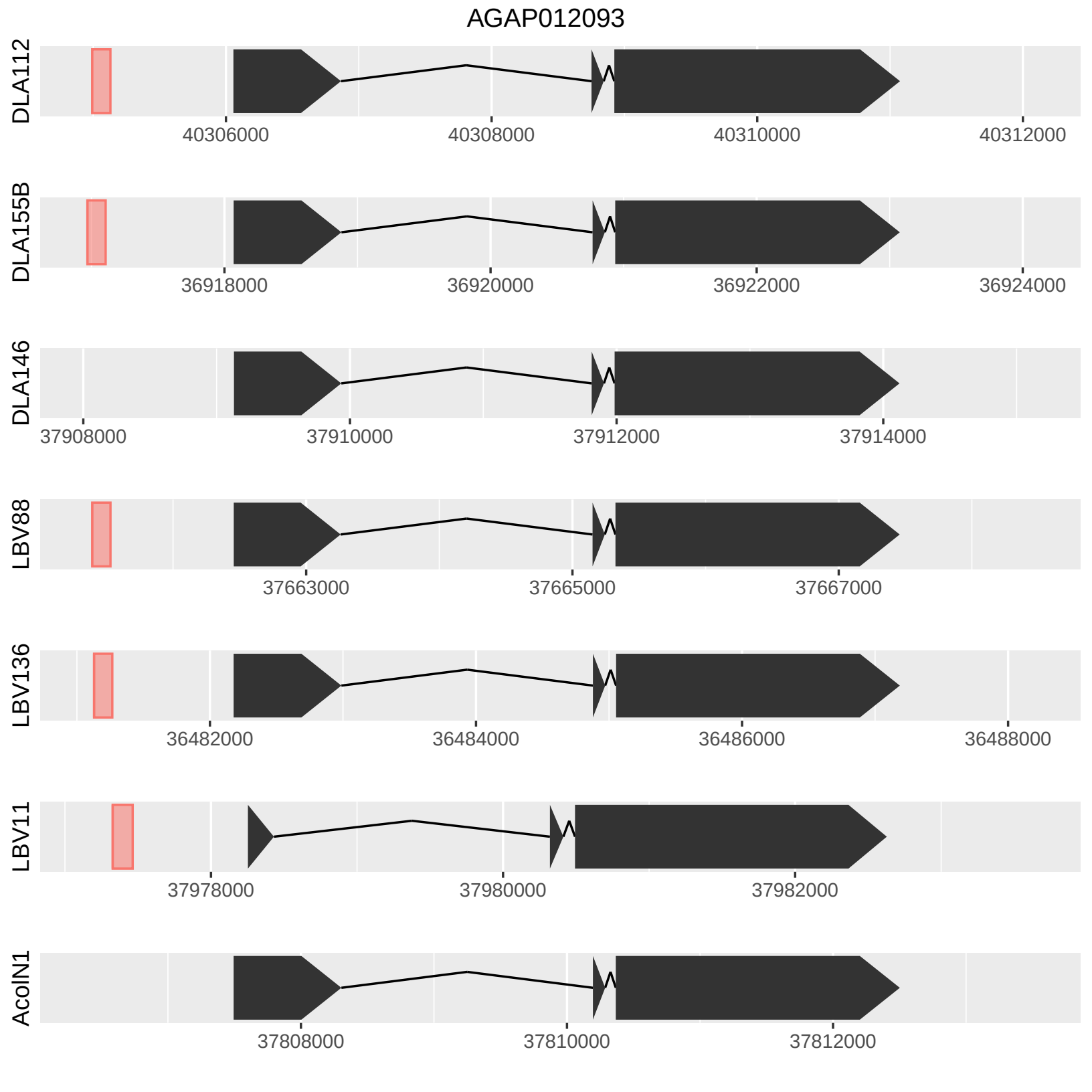

AGAP011360
DLA112
24220000
24240000
24260000
24280000
DLA155B
21775000
21800000
21825000
21850000
DLA146
22830000
22850000
22870000
22890000
LBV88
21500000
21520000
21540000
LBV136
21380000
21390000
21400000
LBV11
23075000
23100000
23125000
23150000
23175000
AcolN1
21820000
21840000
21860000
21880000

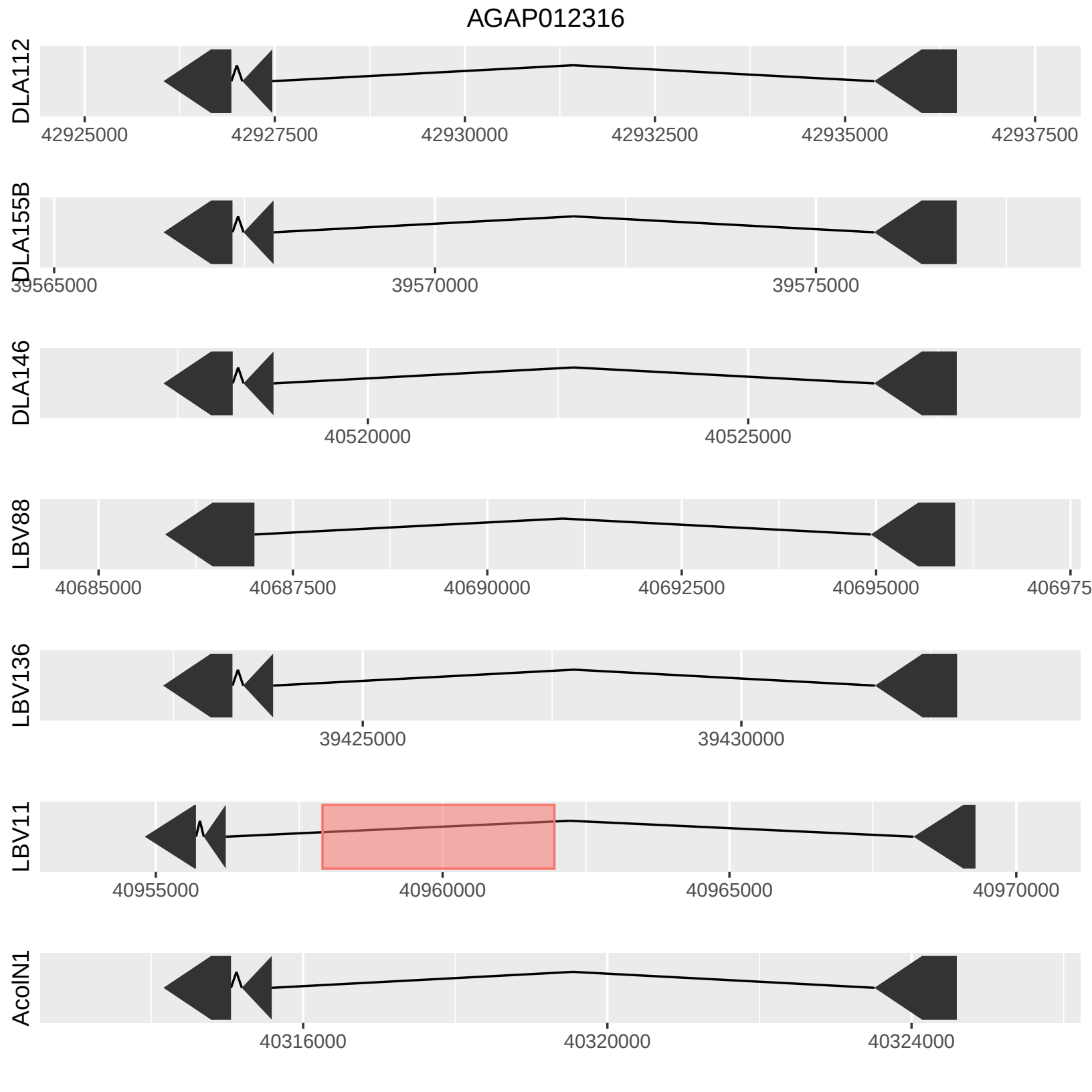

AGAP011379
DLA112
25100000
25200000
25300000
DLA155B
22650000
22700000
22750000
22800000
22850000
DLA146
23750000
23800000
23850000
23900000
23950000
LBV88
22400000
22450000
22500000
22550000
22600000
LBV136
22150000
22200000
22250000
22300000
22350000
LBV11
23850000
23900000
23950000
24000000
AcolN1
22650000
22700000
22750000
22800000
22850000
22900000

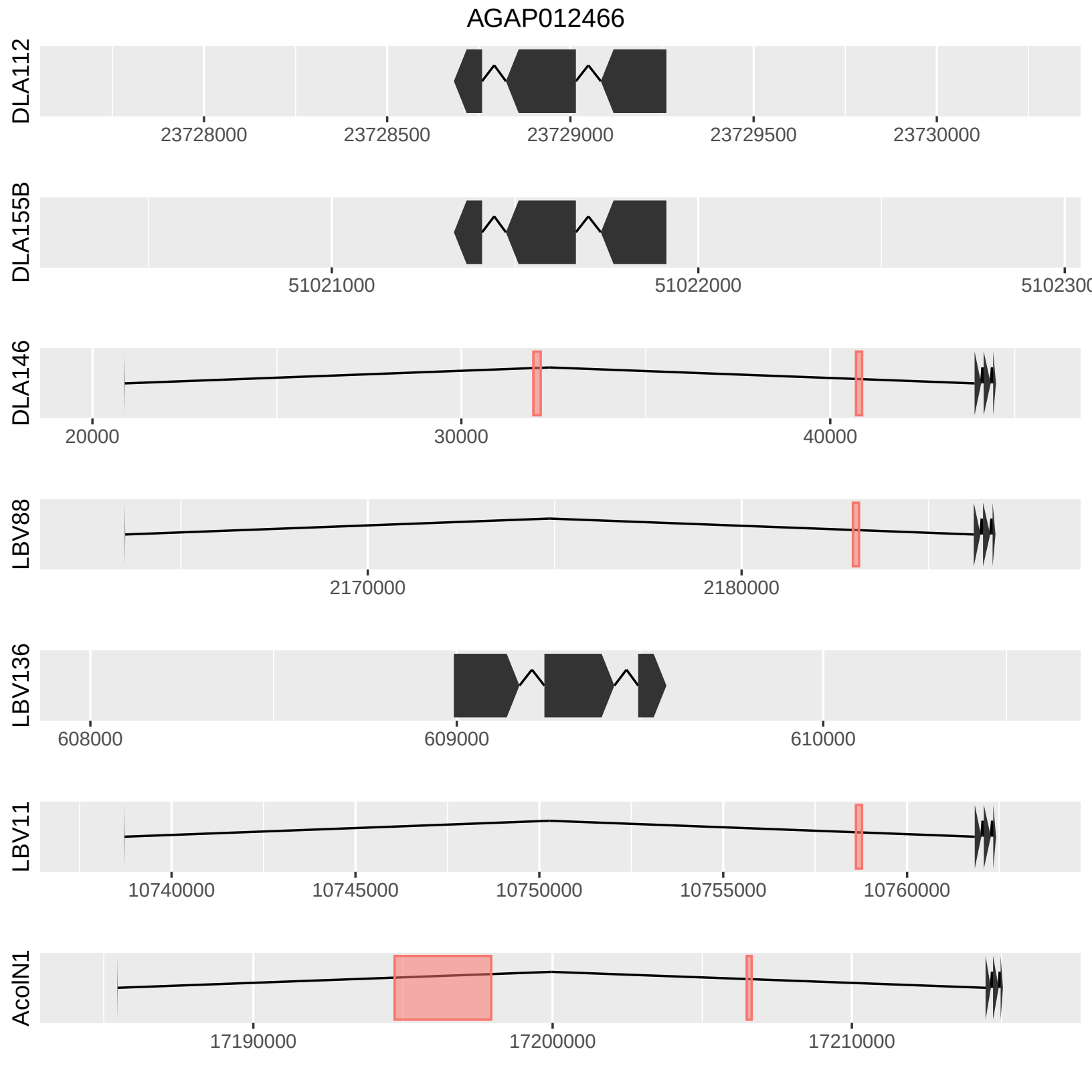

AGAP011794
DLA112
35779000
35780000
35781000
35782000
35783000
DLA155B
33047000
33048000
33049000
33050000
DLA146
10075000
10076000
10077000
10078000
LBV88
33148000
33149000
33150000
33151000
LBV136
32214000
32215000
32216000
32217000
32218000
AcolN1
33461000
33462000
33463000
33464000

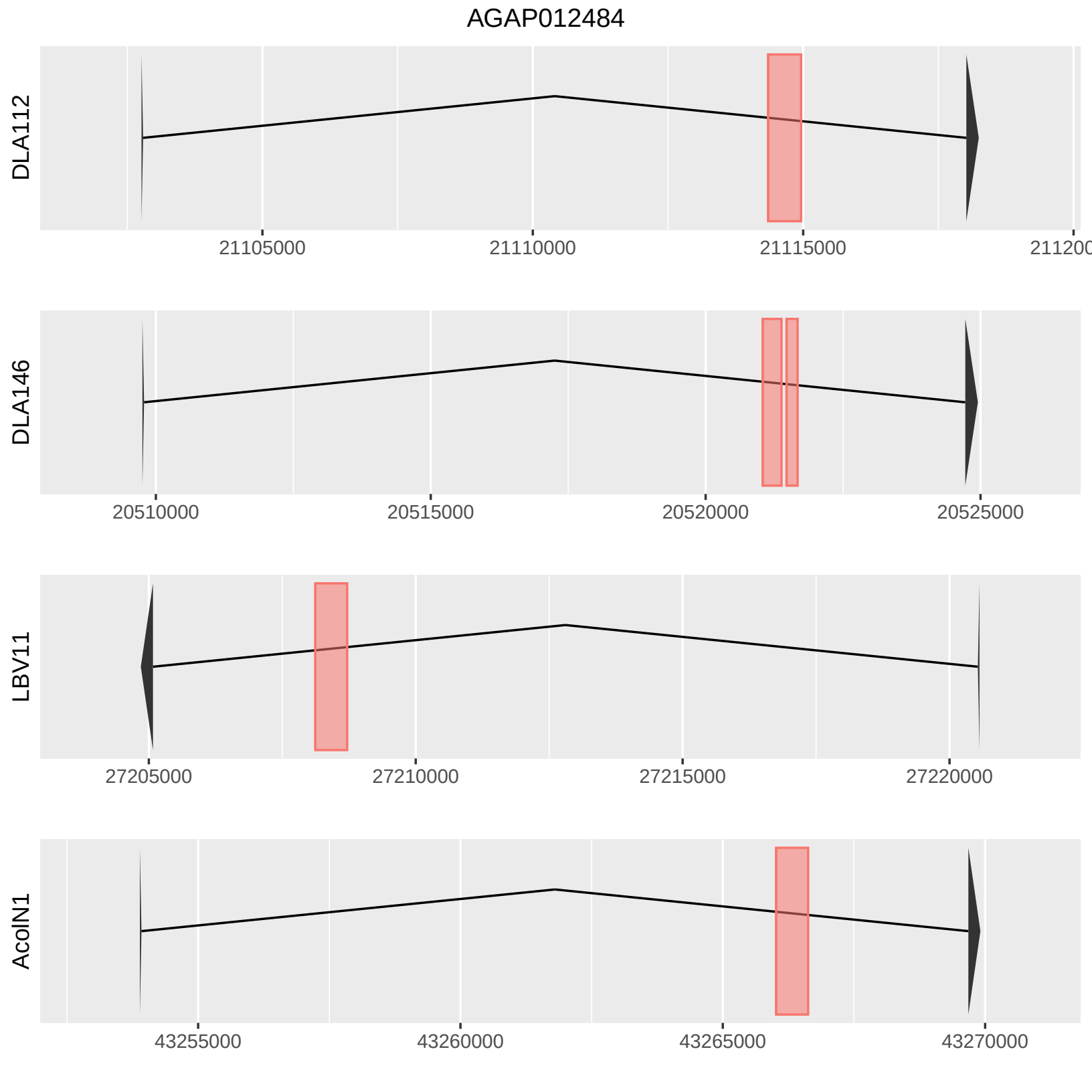

AGAP011916
DLA112
37640000
37660000
37680000
37700000
DLA155B
34400000
34420000
34440000
34460000
DLA146
35550000
35570000
35590000
35610
LBV88
35260000
35280000
35300000
35320000
LBV136
34080000
34100000
34120000
34140000
LBV11
35540000
35560000
35580000
356000
AcolN1
35400000
35420000
35440000
35460000

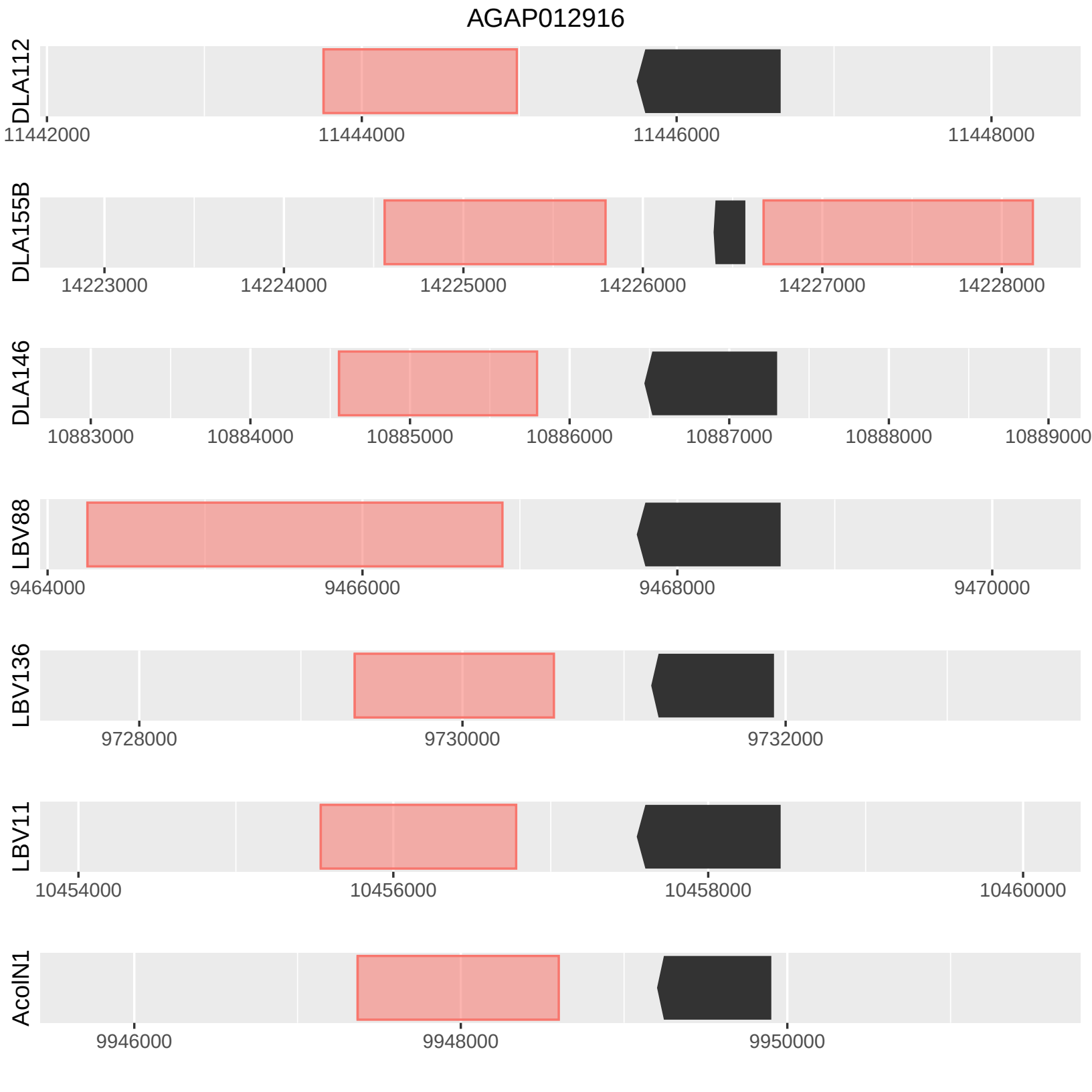

AGAP012092
DLA112
40301000
40302000
40303000
40304000
40305000
DLA155B
36913000
36914000
36915000
36916000
36917000
DLA146
37904000
37905000
37906000
37907000
37908000
LBV88
37657000
37658000
37659000
37660000
37661000
37662000
LBV136
36477000
36478000
36479000
36480000
36481000
LBV11
37973000
37974000
37975000
37976000
37977000
37978000
AcoIN1
37802000
37803000
37804000
37805000
37806000

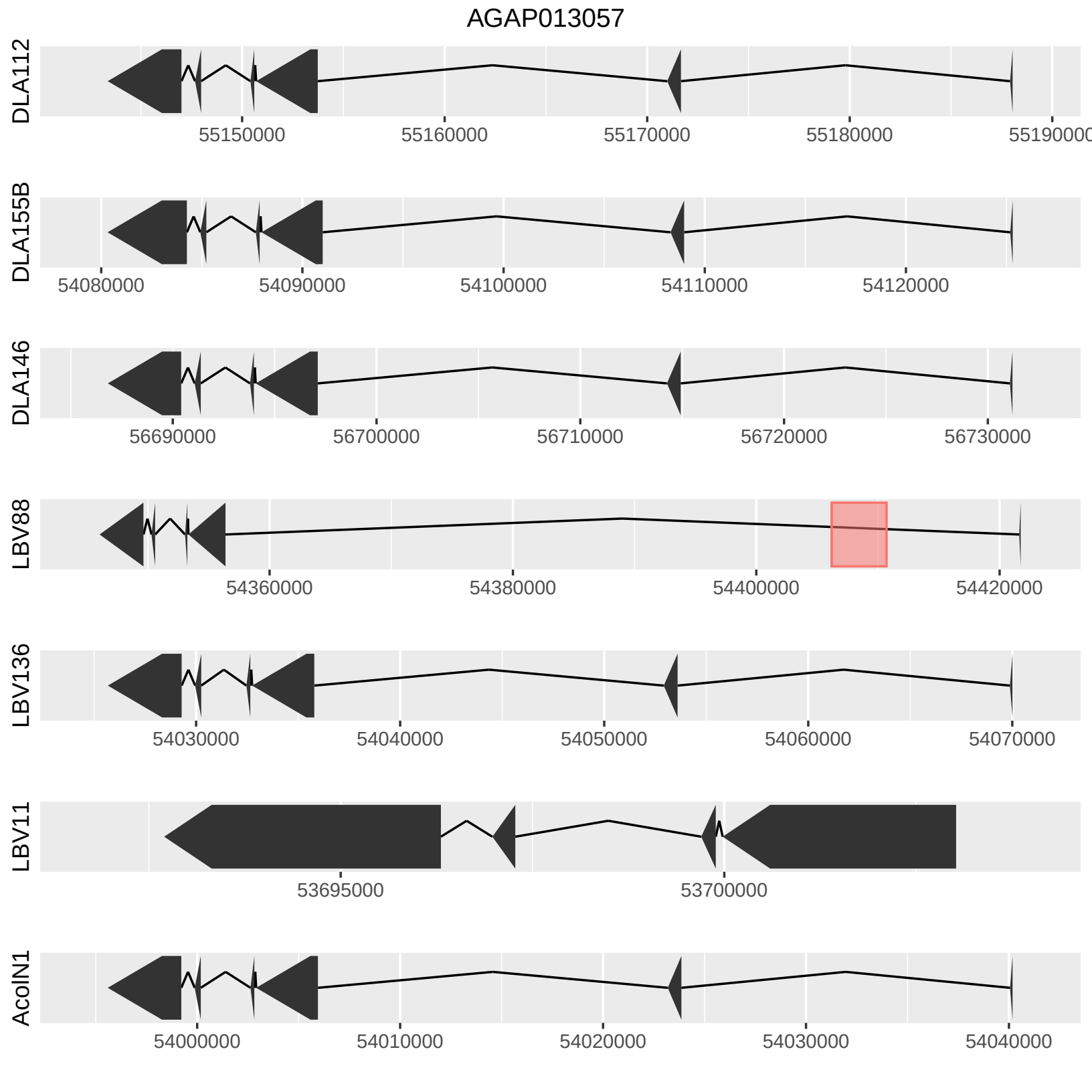

AGAP012093
DLA112
40306000
40308000
40310000
40312000
DLA155B
36918000
36920000
36922000
36924000
DLA146
37908000
37910000
37912000
37914000
LBV88
37663000
37665000
37667000
LBV136
36482000
36484000
36486000
36488000
LBV11
37978000
37980000
37982000
AcoIN1
37808000
37810000
37812000

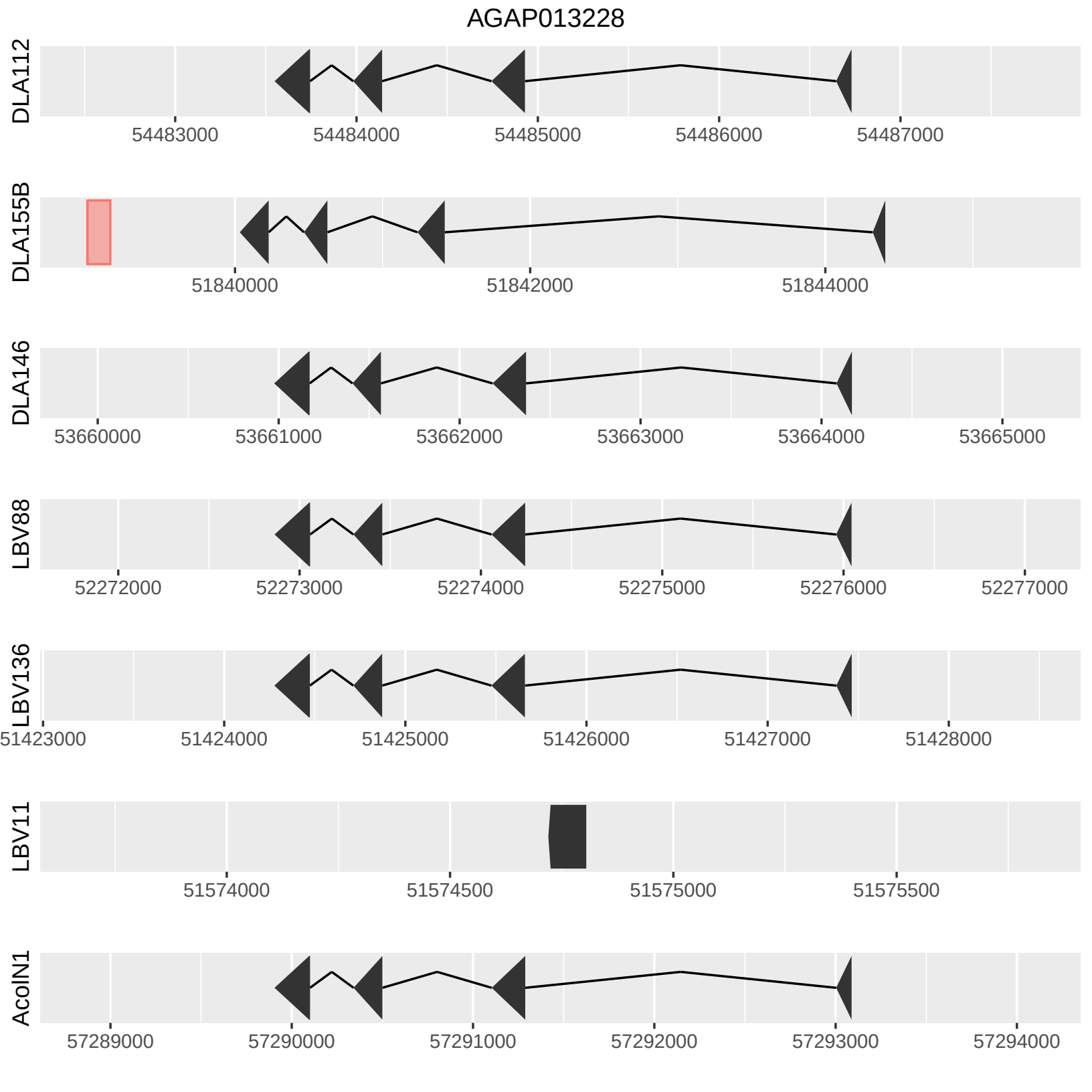

AGAP012316
DLA112
42925000
42927500
42930000
42932500
42935000
42937500
DLA155B
39565000
39570000
39575000
DLA146
40520000
40525000
LBV88
40685000
40687500
40690000
40692500
40695000
406975
LBV136
39425000
39430000
LBV11
40955000
40960000
40965000
40970000
AcoIN1
40316000
40320000
40324000

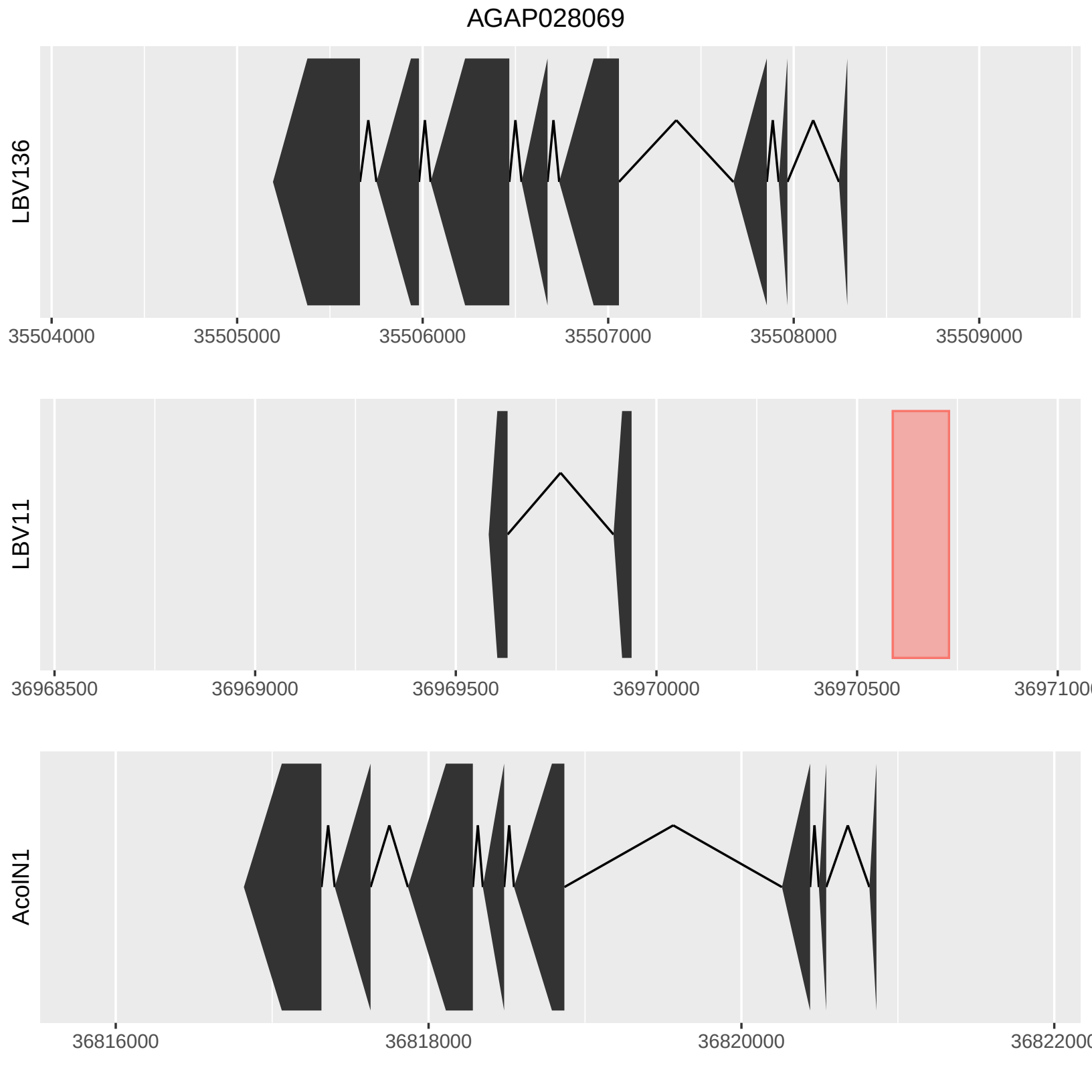

AGAP012466
DLA112
23728000
23728500
23729000
23729500
23730000
DLA155B
51021000
51022000
5102300
DLA146
20000
30000
40000
LBV88
2170000
2180000
LBV136
608000
609000
610000
LBV11
10740000
10745000
10750000
10755000
10760000
AcolN1
17190000
17200000
17210000

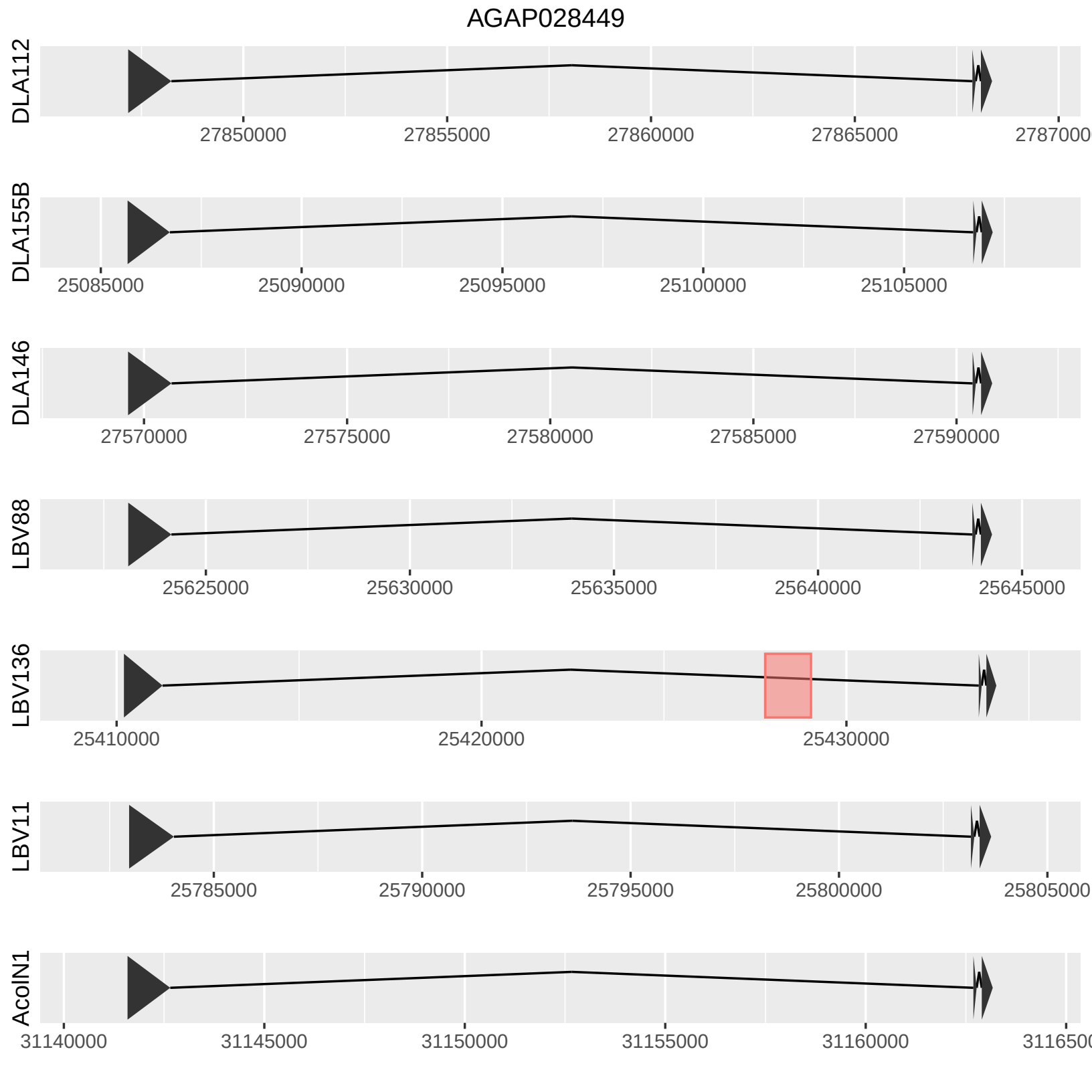

AGAP012484
DLA112
21105000
21110000
21115000
211200
DLA146
20510000
20515000
20520000
20525000
LBV11
27205000
27210000
27215000
27220000
AcoIN1
43255000
43260000
43265000
43270000

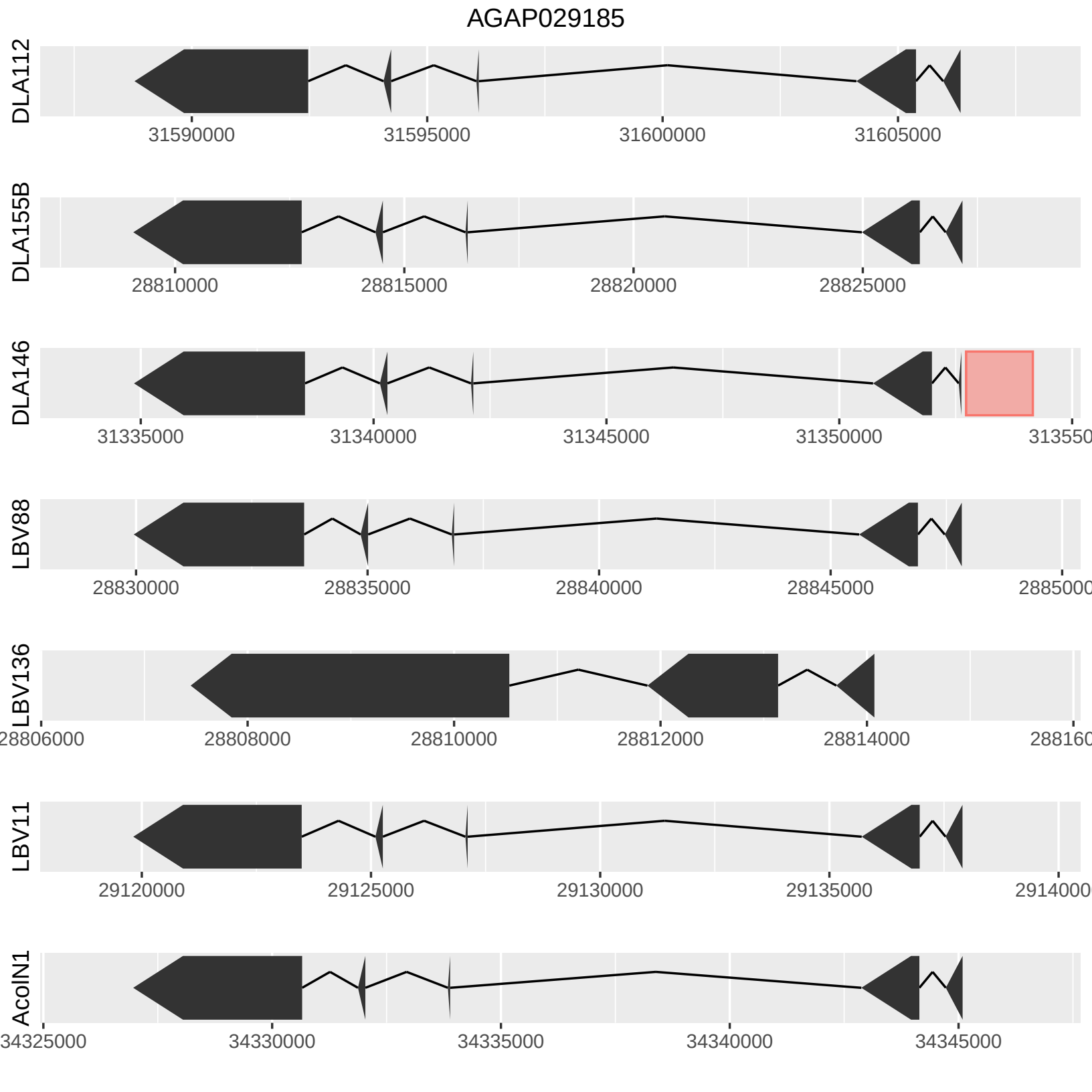

AGAP012916
DLA112
11442000
11444000
11446000
11448000
DLA155B
14223000
14224000
14225000
14226000
14227000
14228000
DLA146
10883000
10884000
10885000
10886000
10887000
10888000
10889000
LBV88
9464000
9466000
9468000
9470000
LBV136
9728000
9730000
9732000
LBV11
10454000
10456000
10458000
10460000
AcoIN1
9946000
9948000
9950000

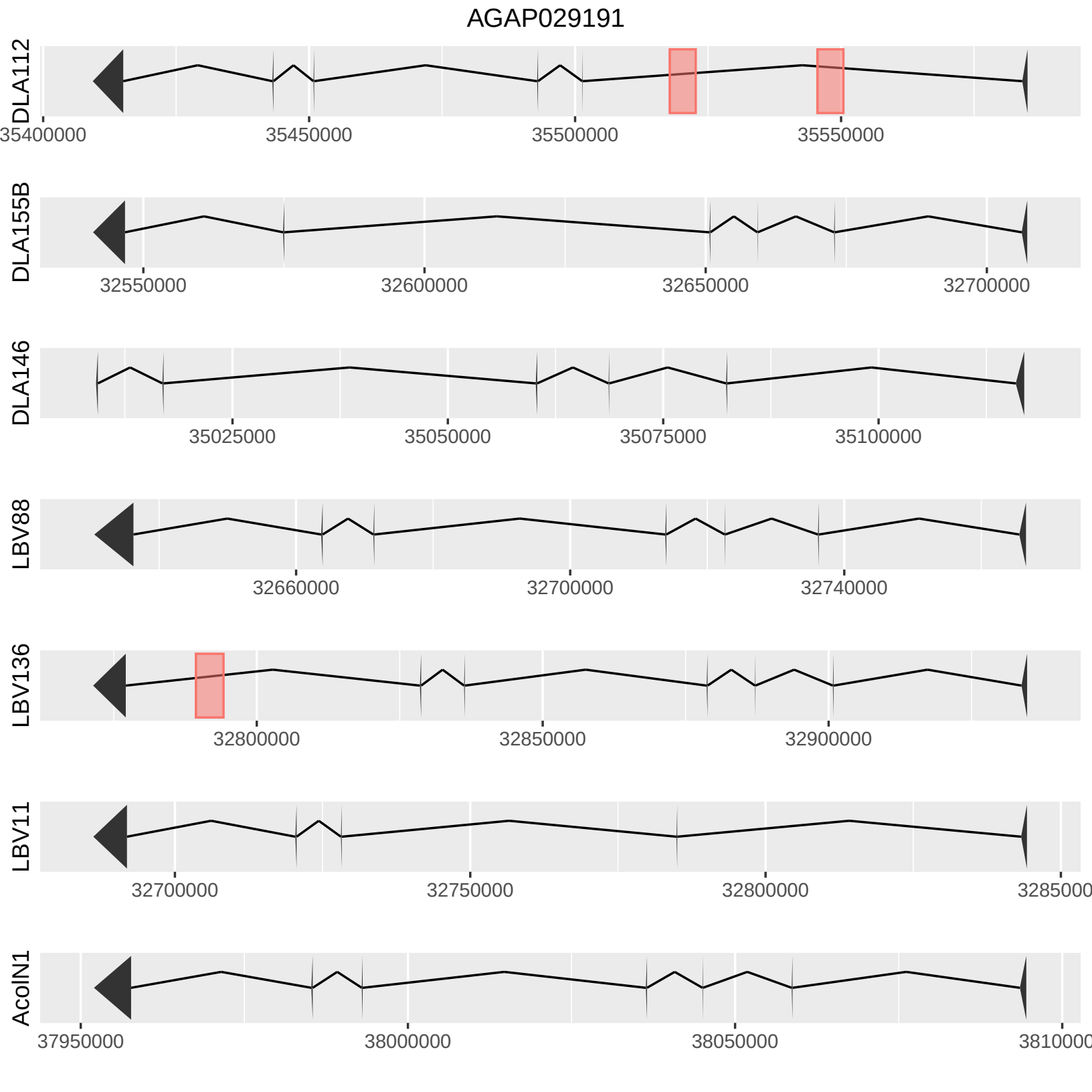

AGAP013057
DLA112
55150000
55160000
55170000
55180000
55190000
DLA155B
54080000
54090000
54100000
54110000
54120000
DLA146
56690000
56700000
56710000
56720000
56730000
LBV88
54360000
54380000
54400000
54420000
LBV136
54030000
54040000
54050000
54060000
54070000
LBV11
53695000
53700000
AcolN1
54000000
54010000
54020000
54030000
54040000

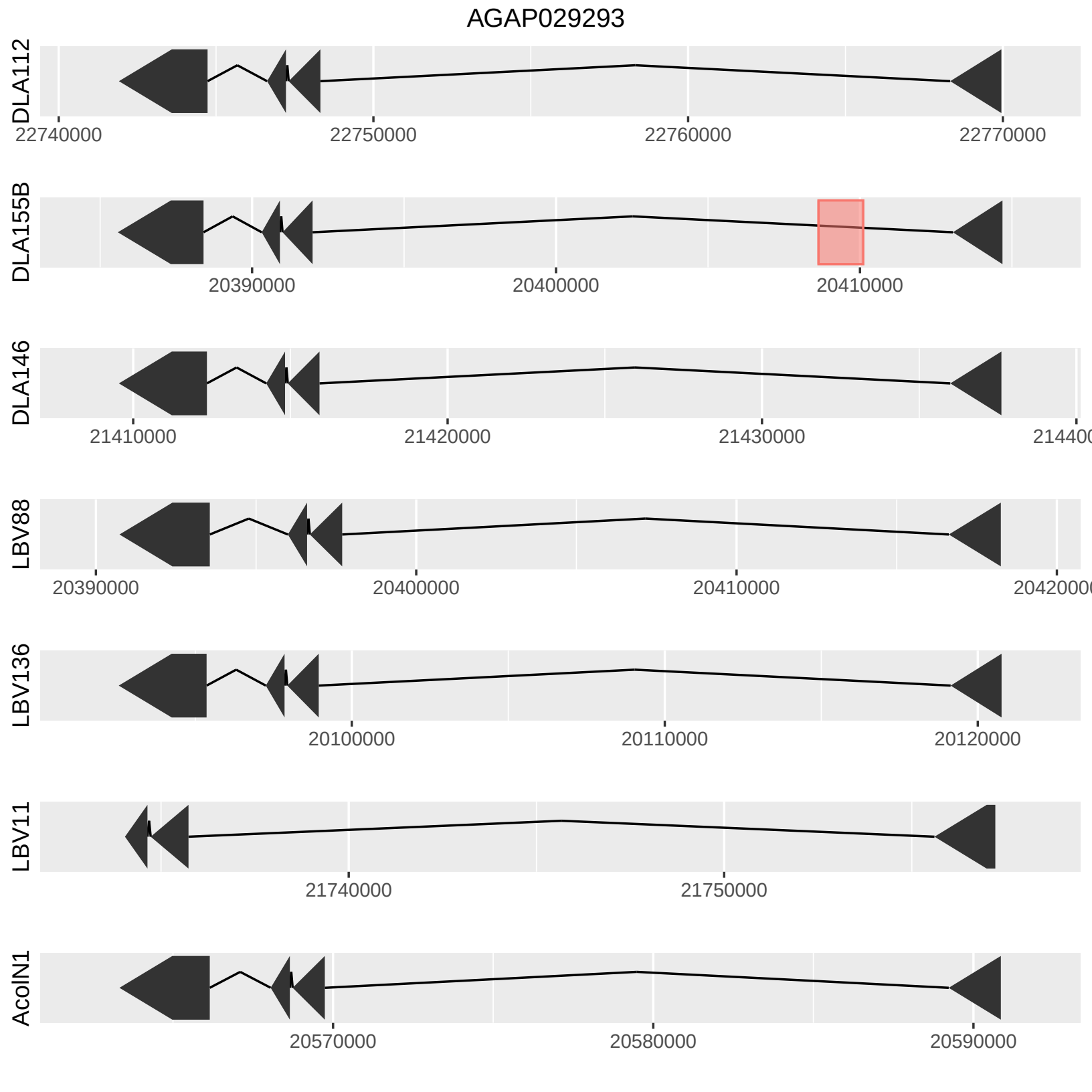

AGAP013228
DLA112
54483000
54484000
54485000
54486000
54487000
DLA155B
51840000
51842000
51844000
DLA146
53660000
53661000
53662000
53663000
53664000
53665000
LBV88
52272000
52273000
52274000
52275000
52276000
52277000
LBV136
51423000
51424000
51425000
51426000
51427000
51428000
LBV11
51574000
51574500
51575000
51575500
AcoIN1
57289000
57290000
57291000
57292000
57293000
57294000

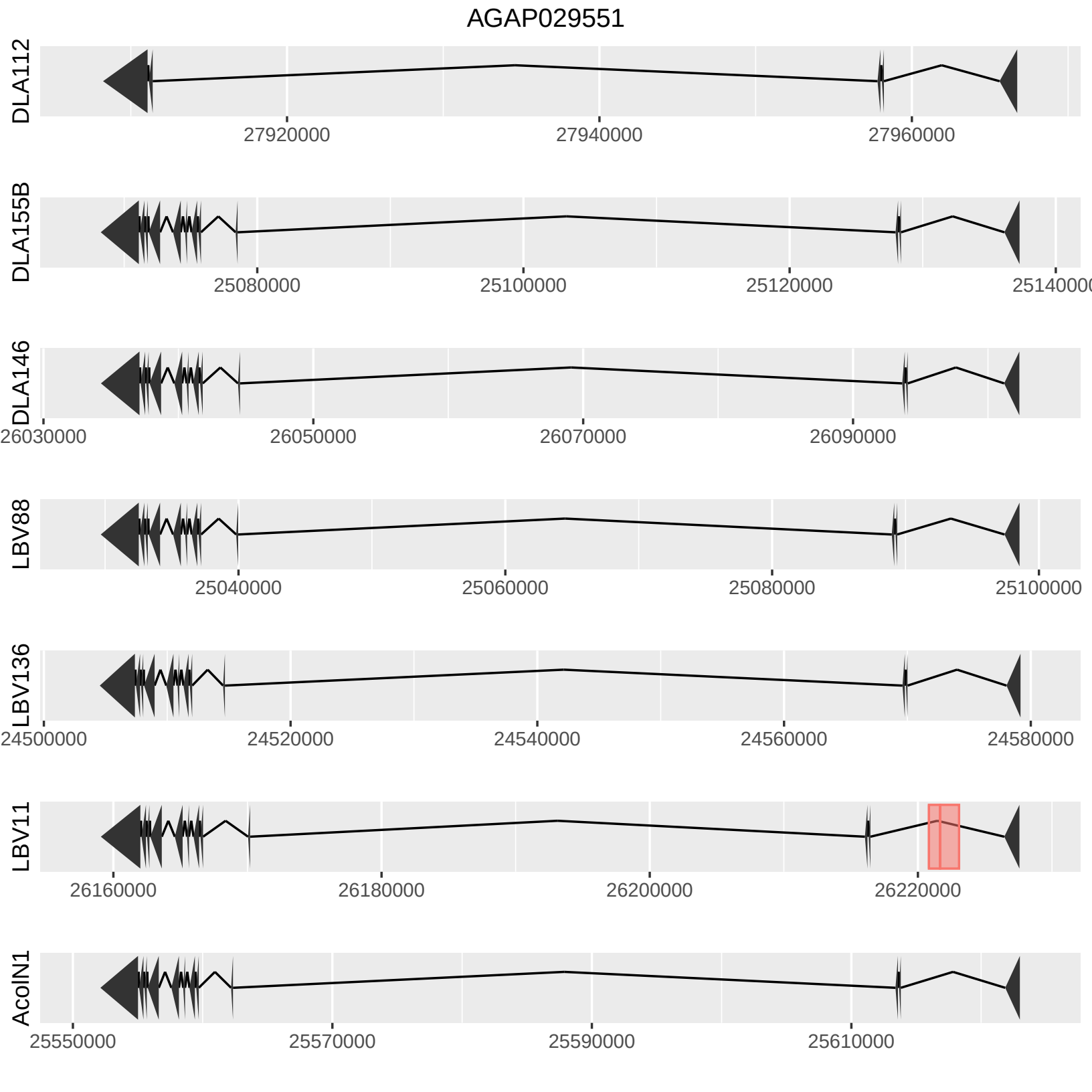

AGAP028069
LBV136
35504000
35505000
35506000
35507000
35508000
35509000
LBV11
36968500
36969000
36969500
36970000
36970500
AcolN1
36816000
36818000
36820000

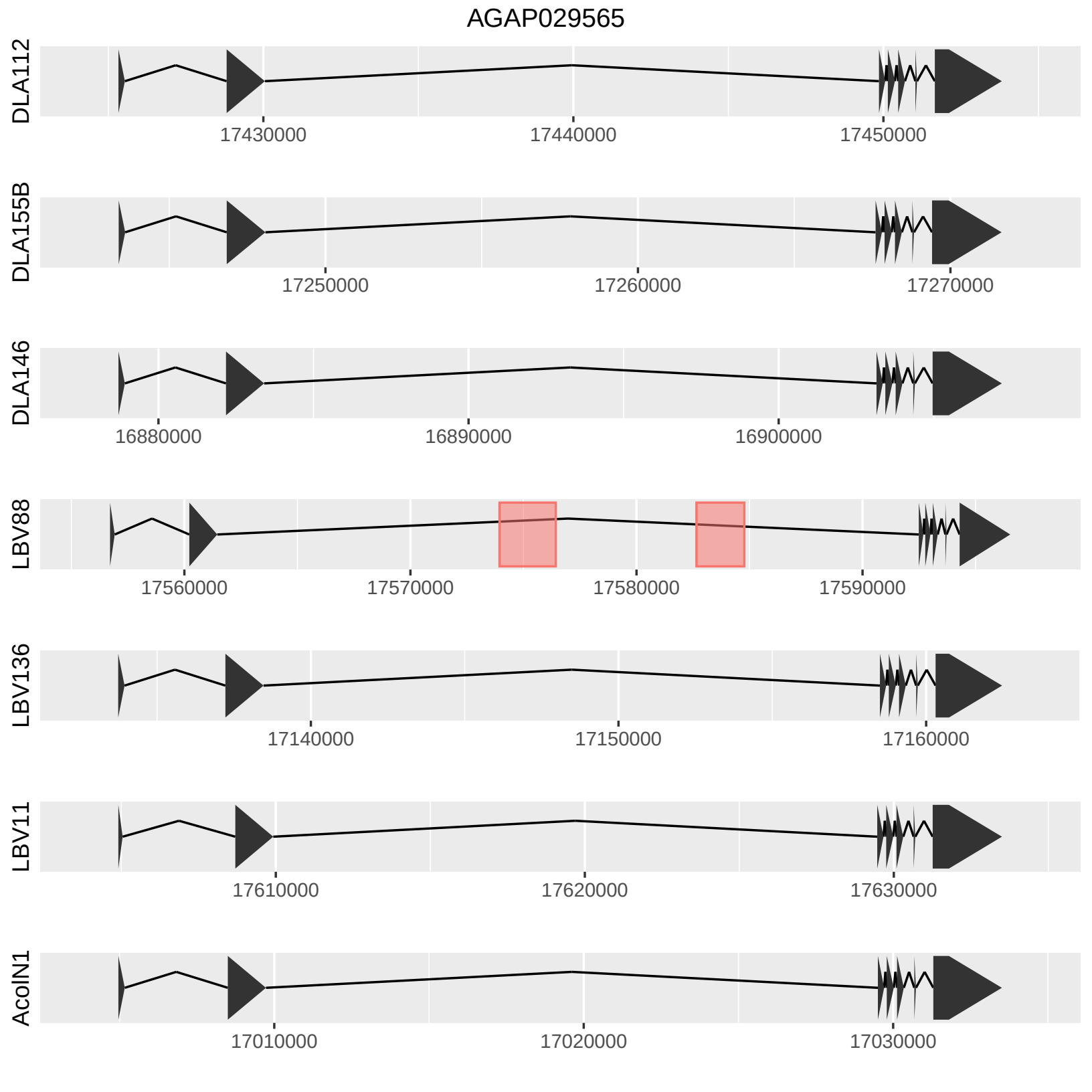

AGAP028449
DLA112
27850000
27855000
27860000
27865000
2787000
DLA155B
25085000
25090000
25095000
25100000
25105000
DLA146
27570000
27575000
27580000
27585000
27590000
LBV88
25625000
25630000
25635000
25640000
25645000
LBV136
25410000
25420000
25430000
LBV11
25785000
25790000
25795000
25800000
25805000
AcolN1
31140000
31145000
31150000
31155000
31160000
311650


AGAP029185
DLA112
31590000
31595000
31600000
31605000
DLA155B
28810000
28815000
28820000
28825000
DLA146
31335000
31340000
31345000
31350000
313550
LBV88
28830000
28835000
28840000
28845000
2885000
LBV136
28806000
28808000
28810000
28812000
28814000
288160
LBV11
29120000
29125000
29130000
29135000
2914000
AcoIN1
34325000
34330000
34335000
34340000
34345000


AGAP029191
DLA112
35400000
35450000
35500000
35550000
DLA155B
32550000
32600000
32650000
32700000
DLA146
35025000
35050000
35075000
35100000
LBV88
32660000
32700000
32740000
LBV136
32800000
32850000
32900000
LBV11
32700000
32750000
32800000
3285000
AcolN1
37950000
38000000
38050000
3810000


AGAP029293
DLA112
22740000
22750000
22760000
22770000
DLA155B
20390000
20400000
20410000
DLA146
21410000
21420000
21430000
LBV88
20390000
20400000
20410000
LBV136
20100000
20110000
20120000
LBV11
21740000
21750000
AcolN1
20570000
20580000
20590000


AGAP029551
DLA112
27920000
27940000
27960000
DLA155B
25080000
25100000
25120000
25140000
DLA146
26030000
26050000
26070000
26090000
LBV88
25040000
25060000
25080000
25100000
LBV136
24500000
24520000
24540000
24560000
24580000
LBV11
26160000
26180000
26200000
26220000
AcolN1
25550000
25570000
25590000
25610000


AGAP029565
DLA112
17430000
17440000
17450000
DLA155B
17250000
17260000
17270000
DLA146
16880000
16890000
16900000
LBV88
17560000
17570000
17580000
17590000
LBV136
17140000
17150000
17160000
LBV11
17610000
17620000
17630000
AcoIN1
17010000
17020000
17030000

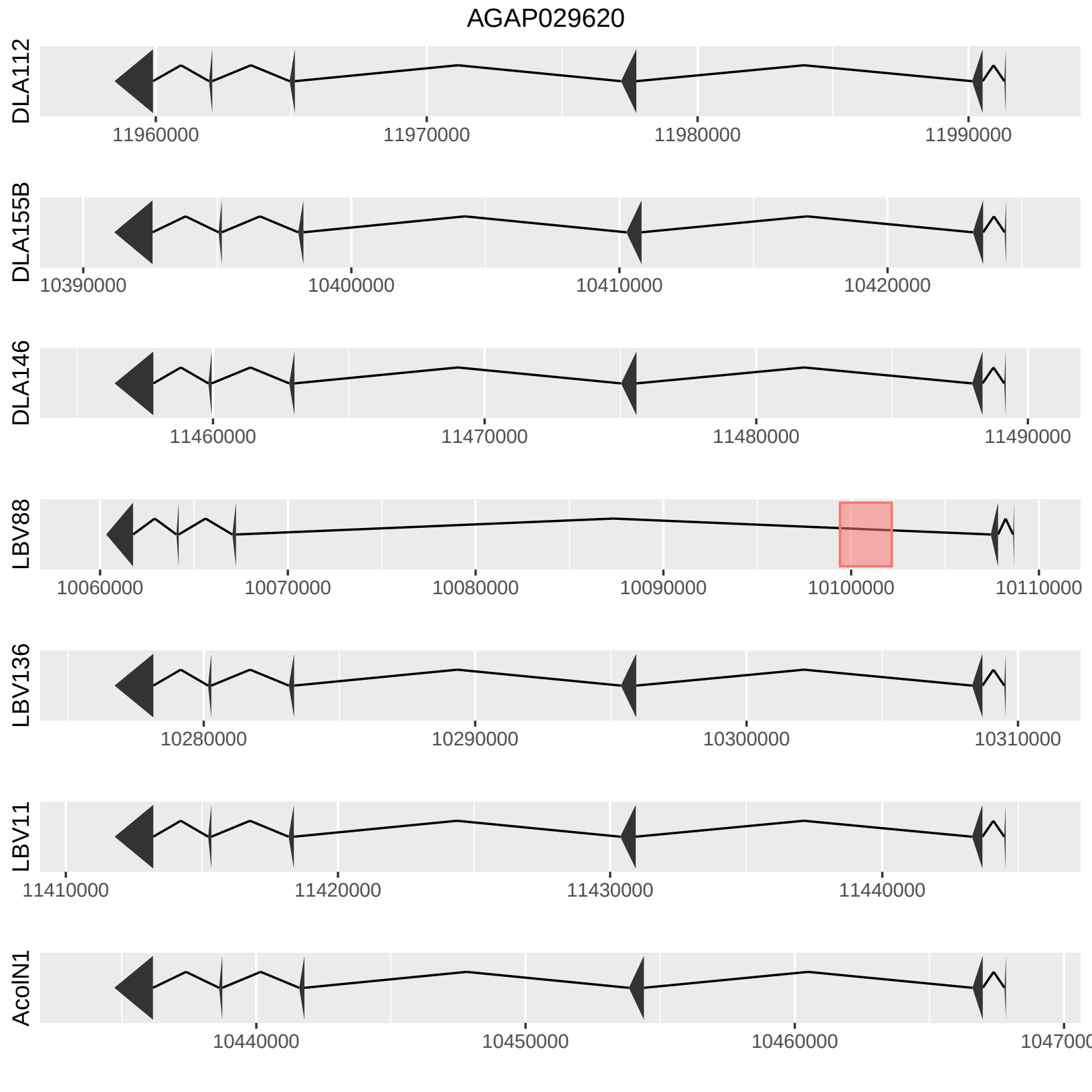
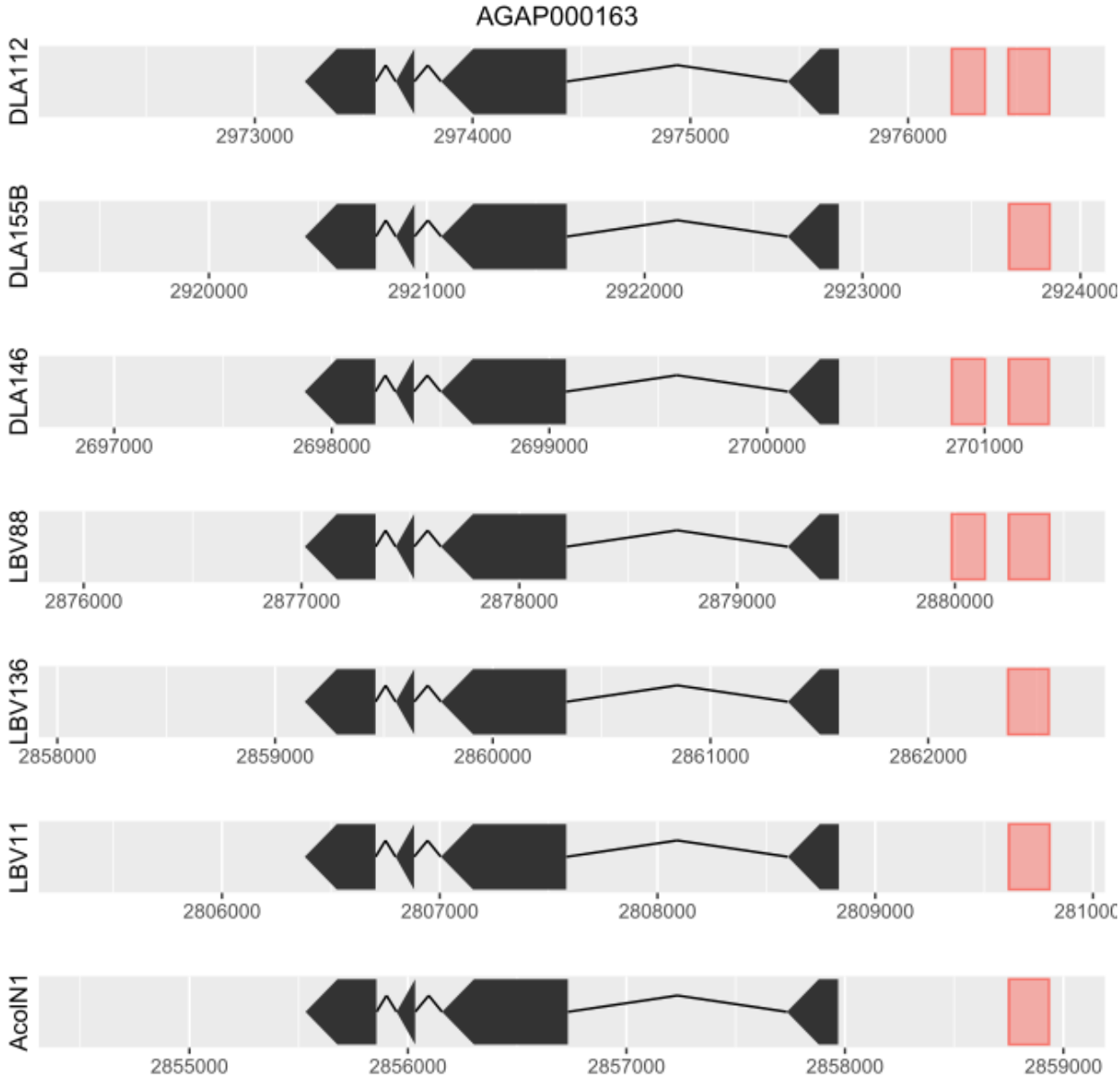
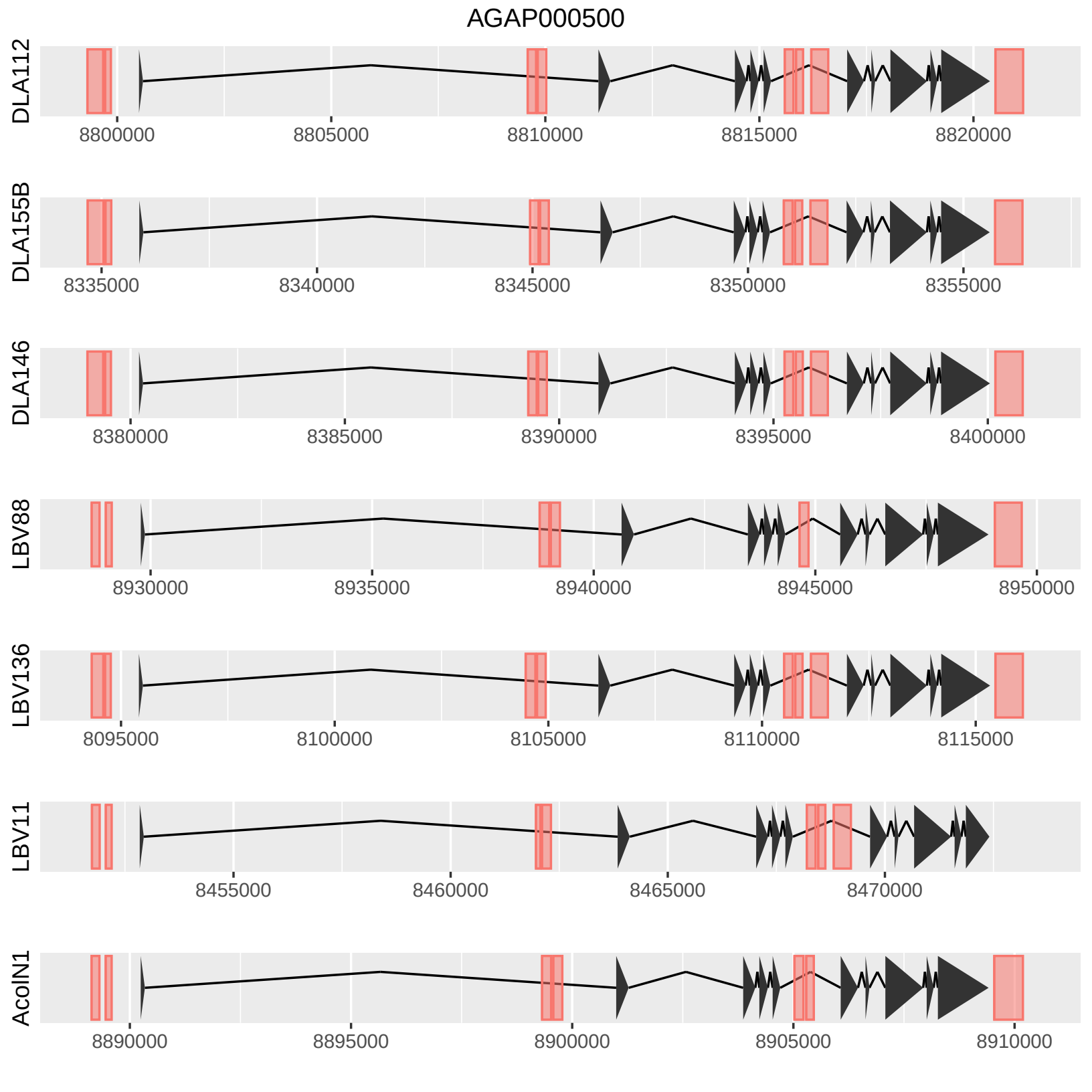
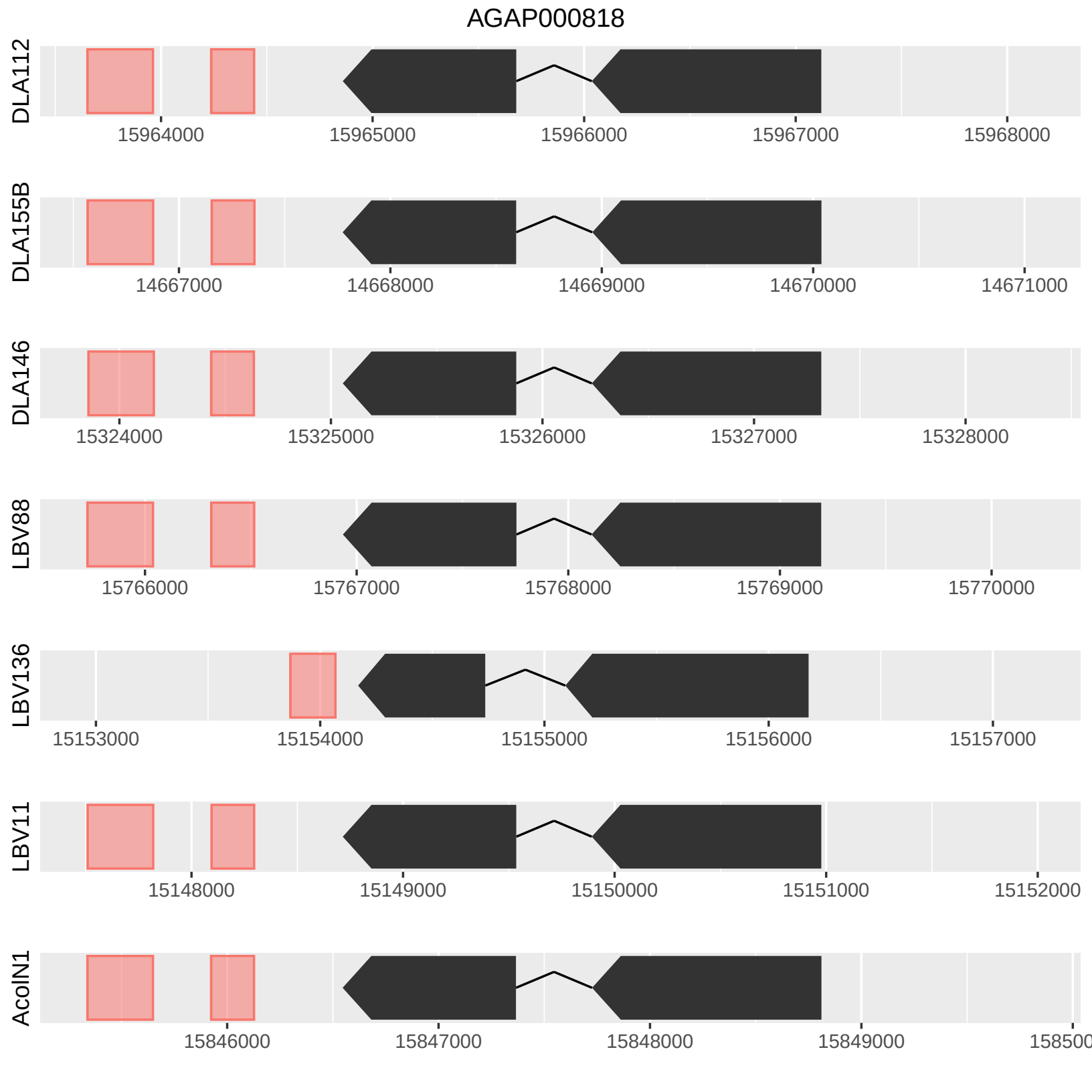
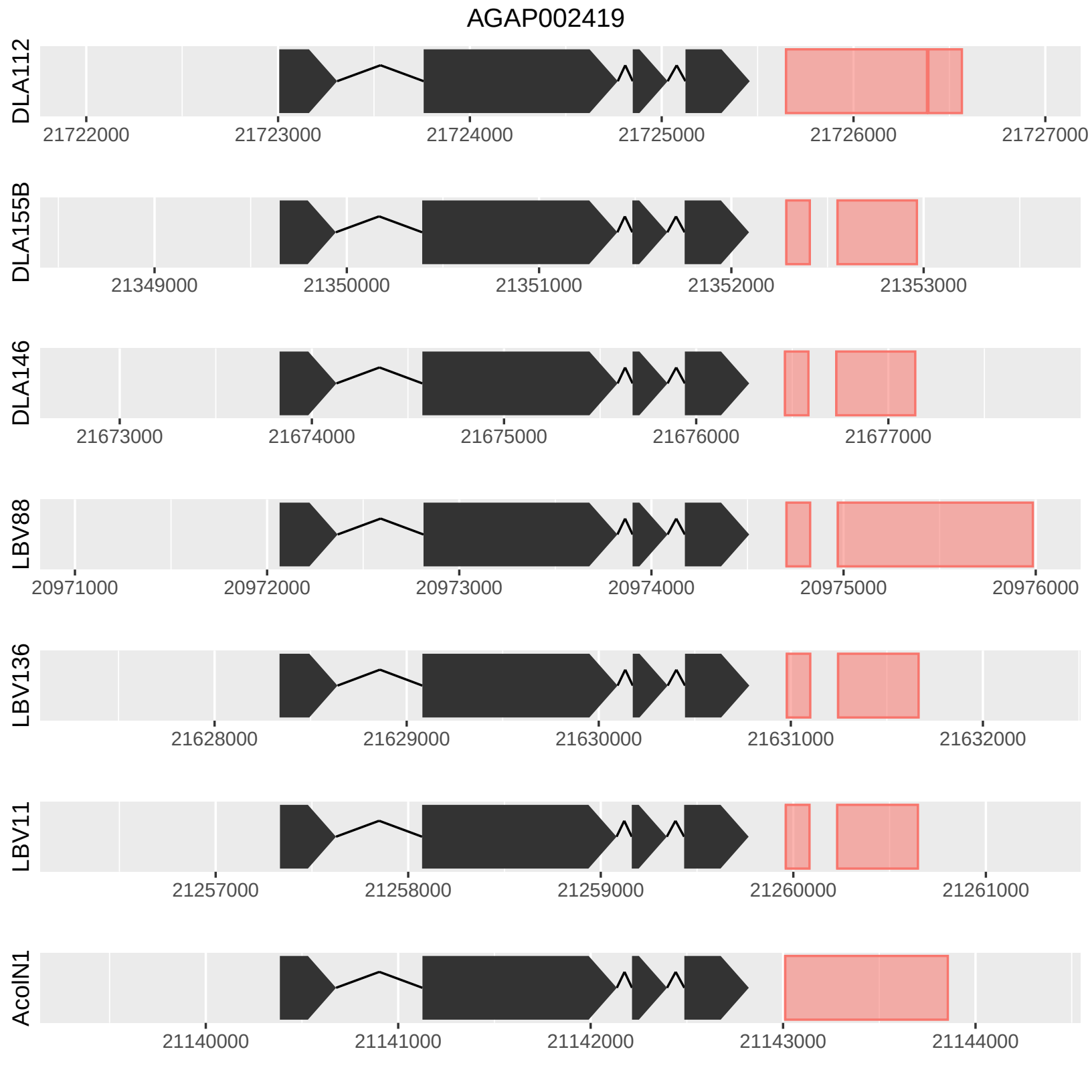

AGAP029620
DLA112
11960000
11970000
11980000
11990000
DLA155B
10390000
10400000
10410000
10420000
DLA146
11460000
11470000
11480000
11490000
LBV88
10060000
10070000
10080000
10090000
10100000
10110000
LBV136
10280000
10290000
10300000
10310000
LBV11
11410000
11420000
11430000
11440000
AcolN1
10440000
10450000
10460000
1047000

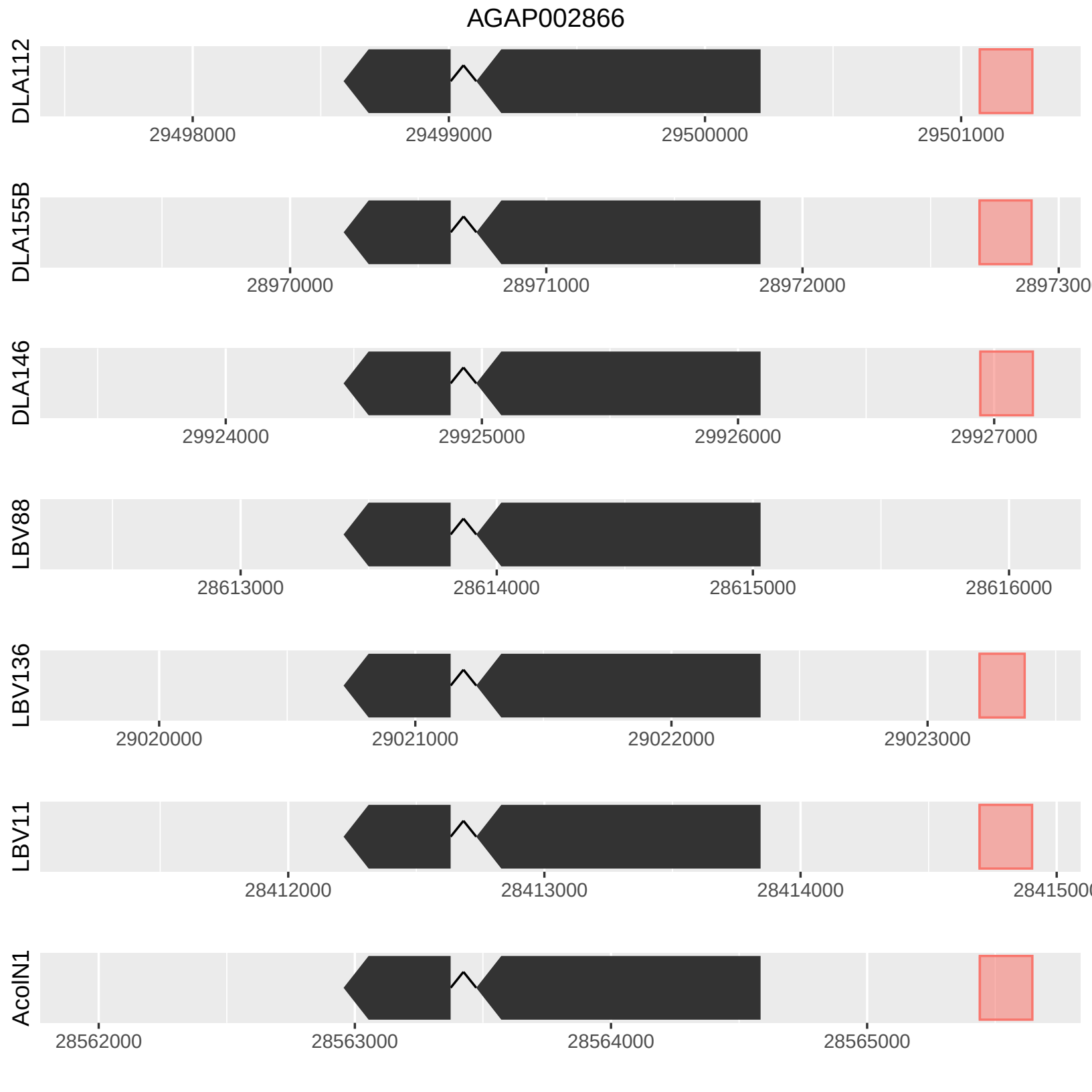
Figure S7. Genes associated with insecticide resistance with TE insertions. Diagrams of genes associated with insecticide resistance showing the gene structure and the TE insertions (in red) inside or closer than 1 kb to gene.

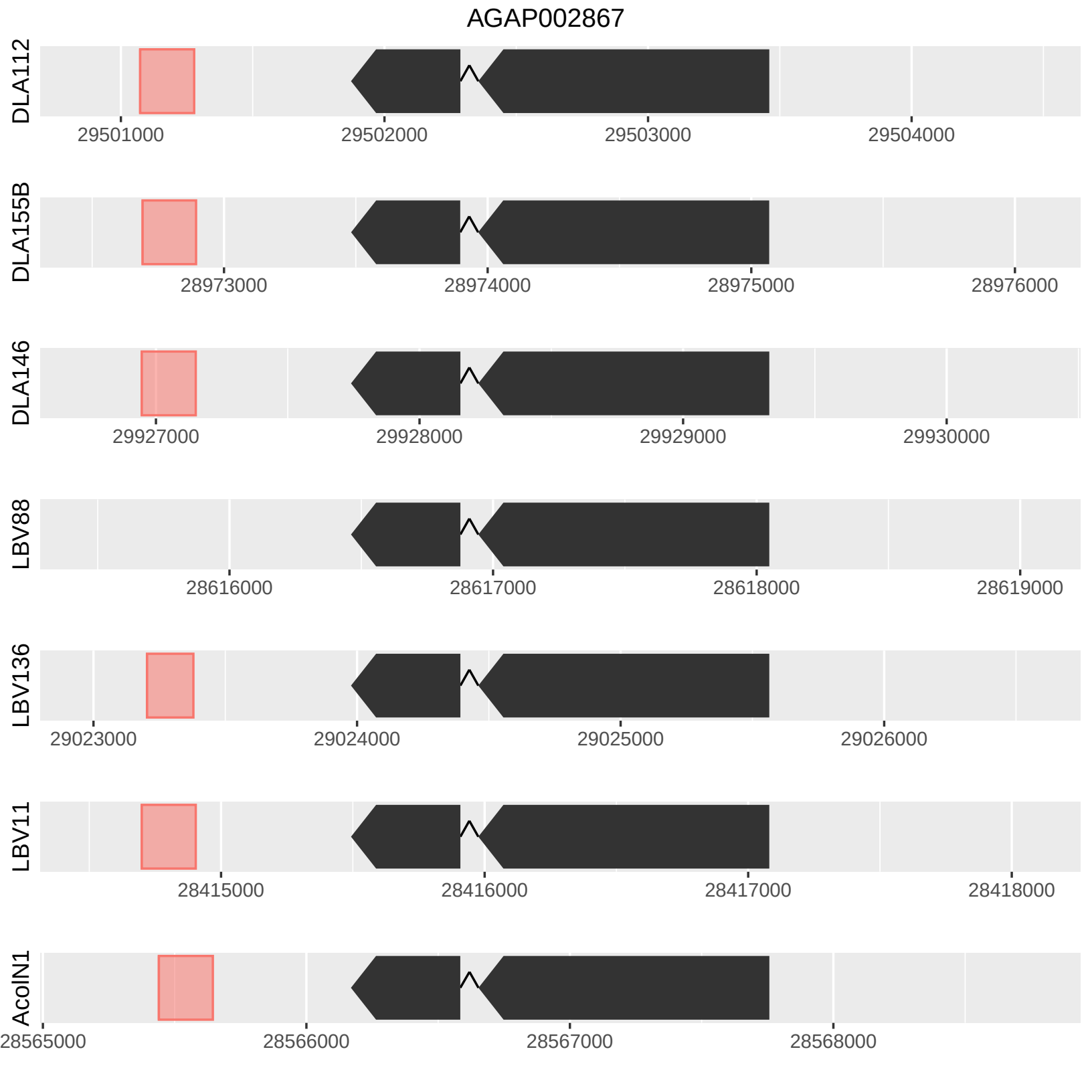


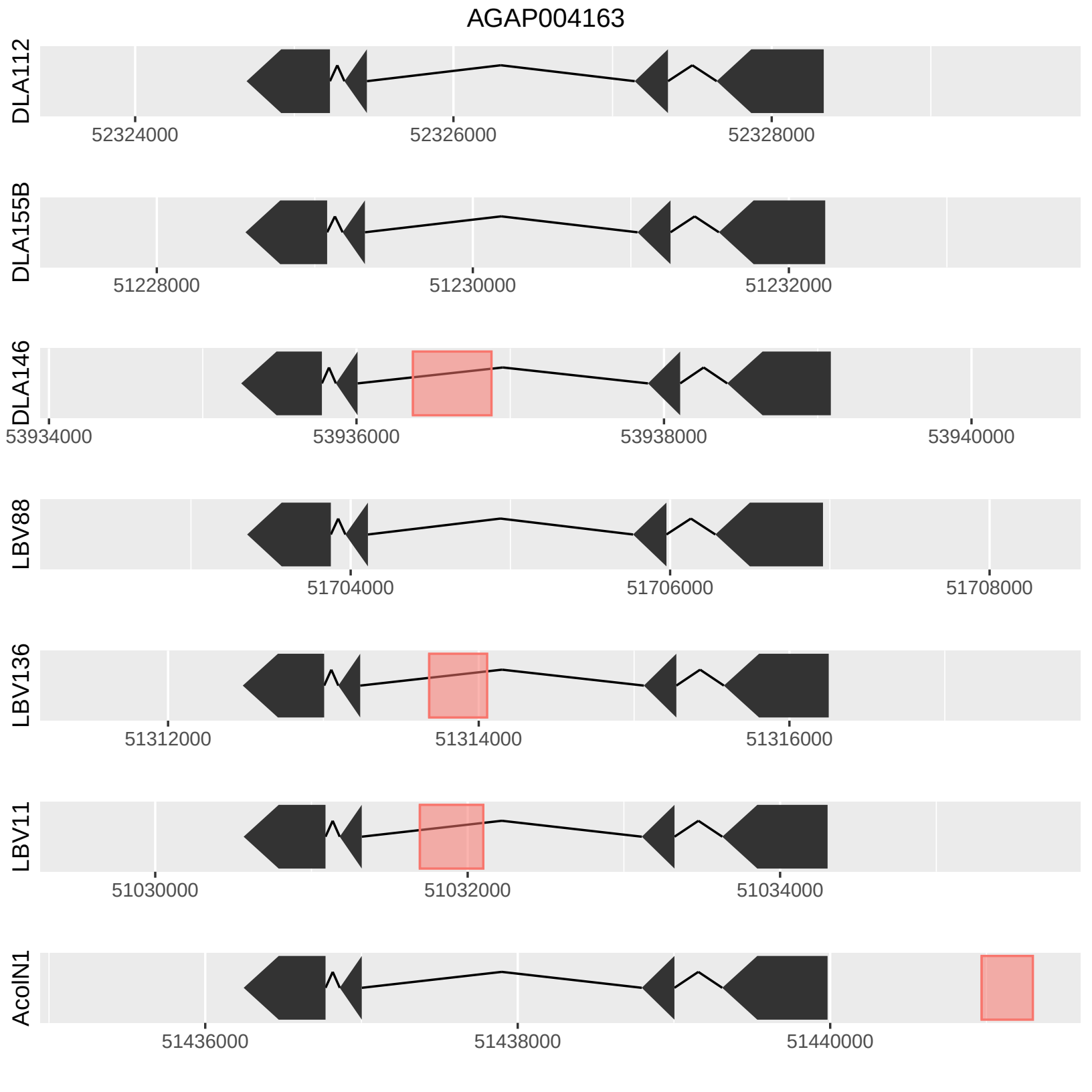

AGAP000500
DLA112
8800000
8805000
8810000
8815000
8820000
DLA155B
8335000
8340000
8345000
8350000
8355000
DLA146
8380000
8385000
8390000
8395000
8400000
LBV88
8930000
8935000
8940000
8945000
8950000
LBV136
8095000
8100000
8105000
8110000
8115000
LBV11
8455000
8460000
8465000
8470000
AcolN1
8890000
8895000
8900000
8905000
8910000

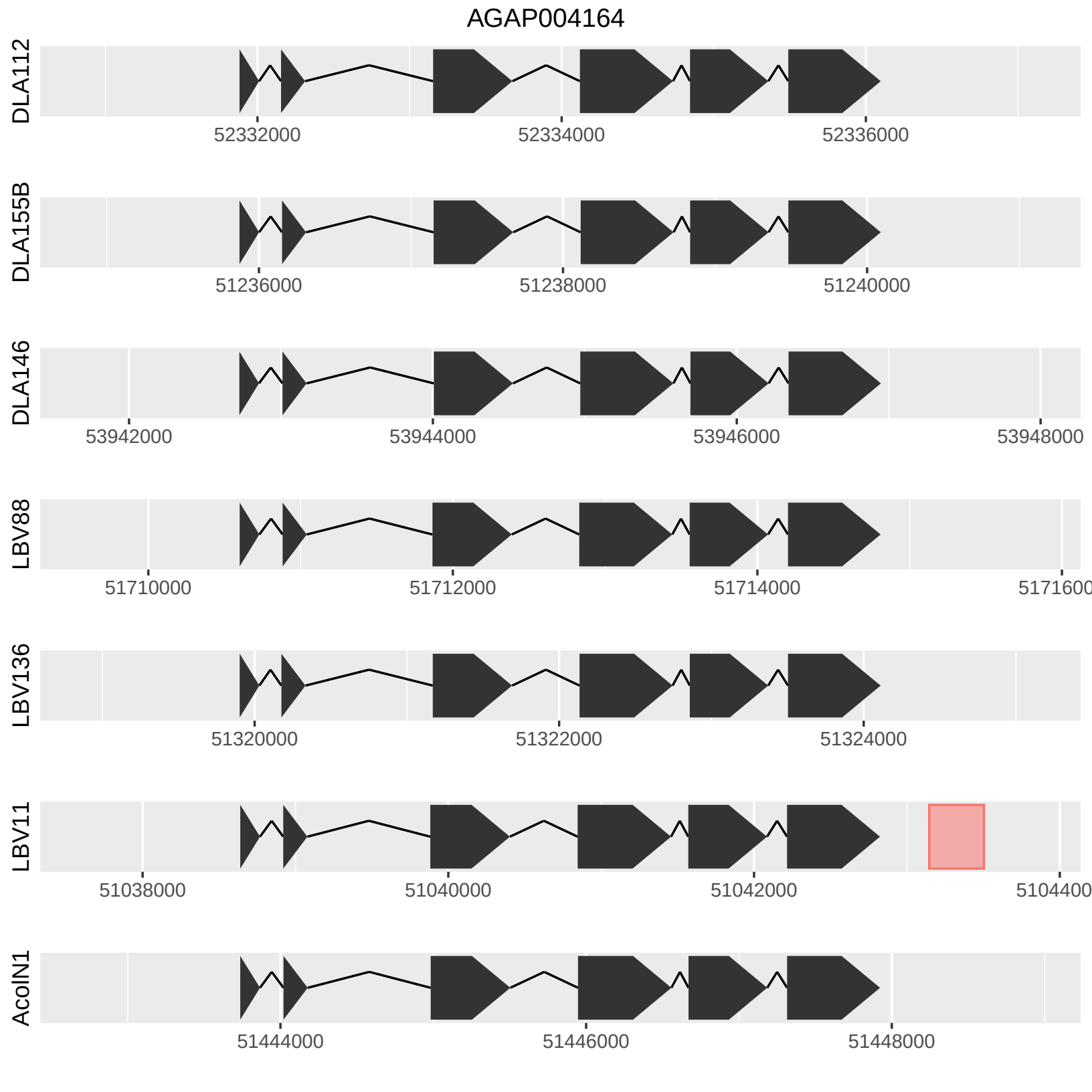

AGAP000818
DLA112
15964000
15965000
15966000
15967000
15968000
DLA155B
14667000
14668000
14669000
14670000
14671000
DLA146
15324000
15325000
15326000
15327000
15328000
LBV88
15766000
15767000
15768000
15769000
15770000
LBV136
15153000
15154000
15155000
15156000
15157000
LBV11
15148000
15149000
15150000
15151000
15152000
AcolN1
15846000
15847000
15848000
15849000
158500

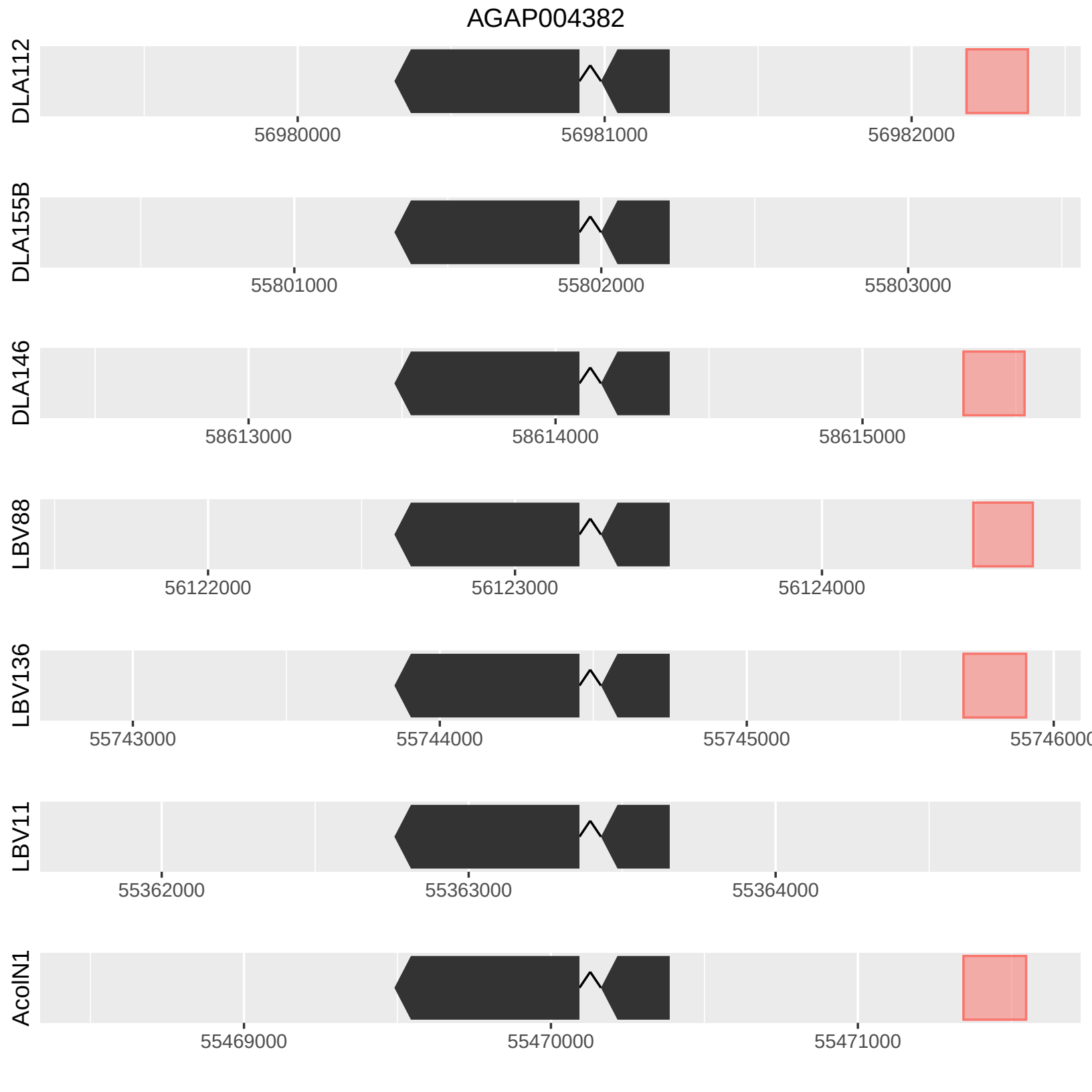

AGAP002419
DLA112
21722000
21723000
21724000
21725000
21726000
21727000
DLA155B
21349000
21350000
21351000
21352000
21353000
DLA146
21673000
21674000
21675000
21676000
21677000
LBV88
20971000
20972000
20973000
20974000
20975000
20976000
LBV136
21628000
21629000
21630000
21631000
21632000
LBV11
21257000
21258000
21259000
21260000
21261000
AcolN1
21140000
21141000
21142000
21143000
21144000

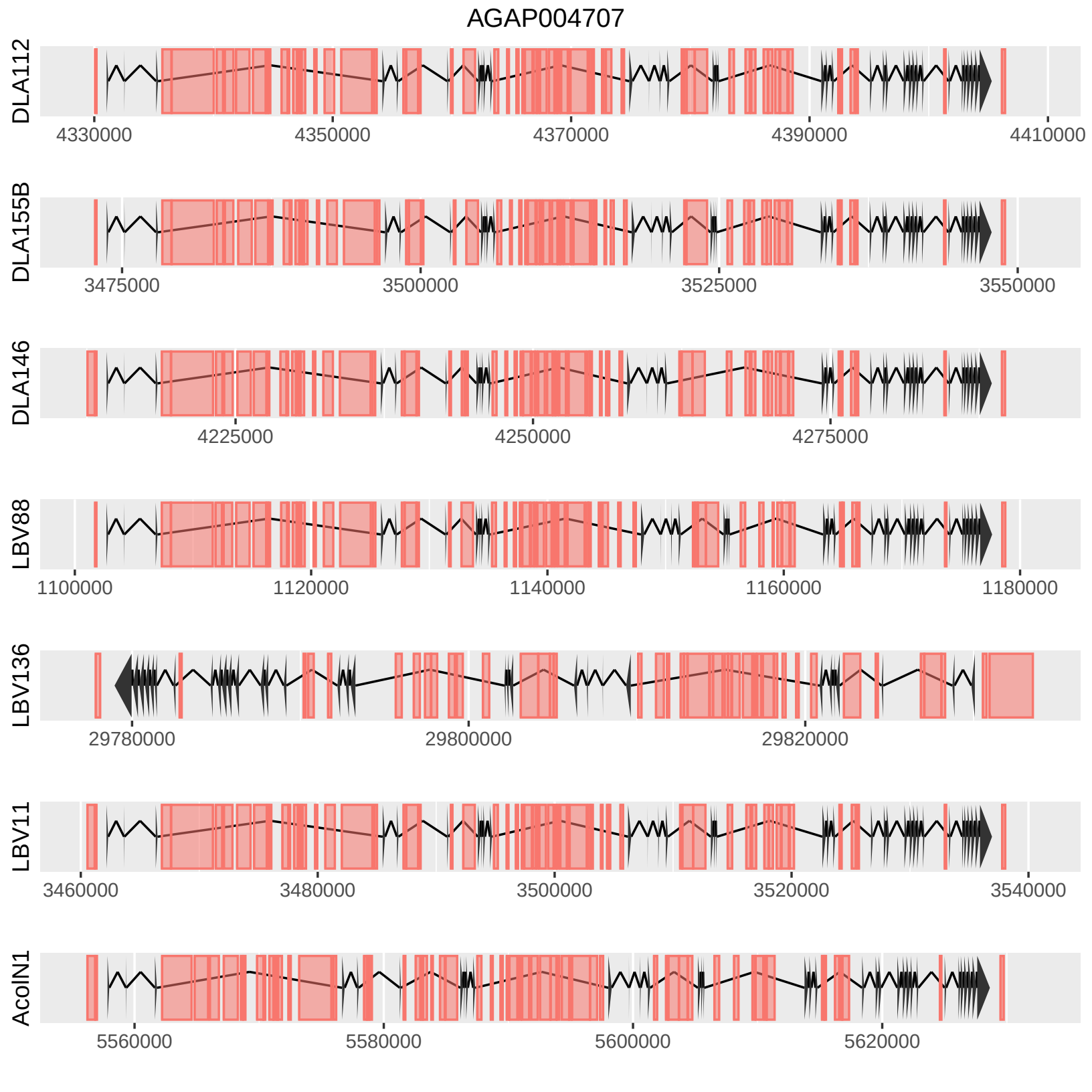

AGAP002866
DLA112
29498000
29499000
29500000
29501000
DLA155B
28970000
28971000
28972000
2897300
DLA146
29924000
29925000
29926000
29927000
LBV88
28613000
28614000
28615000
28616000
LBV136
29020000
29021000
29022000
29023000
LBV11
28412000
28413000
28414000
2841500
AcoIN1
28562000
28563000
28564000
28565000

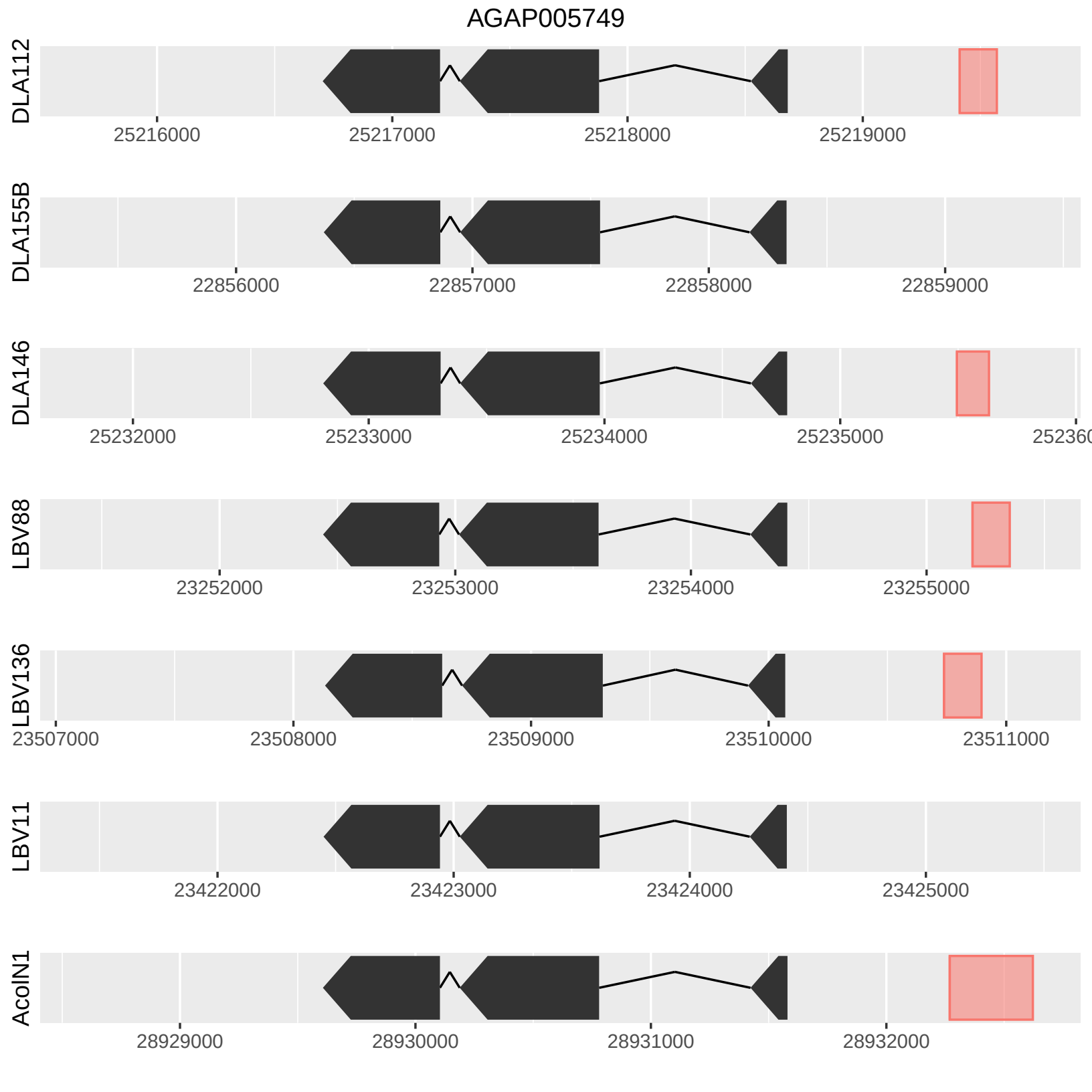

AGAP002867
DLA112
29501000
29502000
29503000
29504000
DLA155B
28973000
28974000
28975000
28976000
DLA146
29927000
29928000
29929000
29930000
LBV88
28616000
28617000
28618000
28619000
LBV136
29023000
29024000
29025000
29026000
LBV11
28415000
28416000
28417000
28418000
AcoIN1
28565000
28566000
28567000
28568000

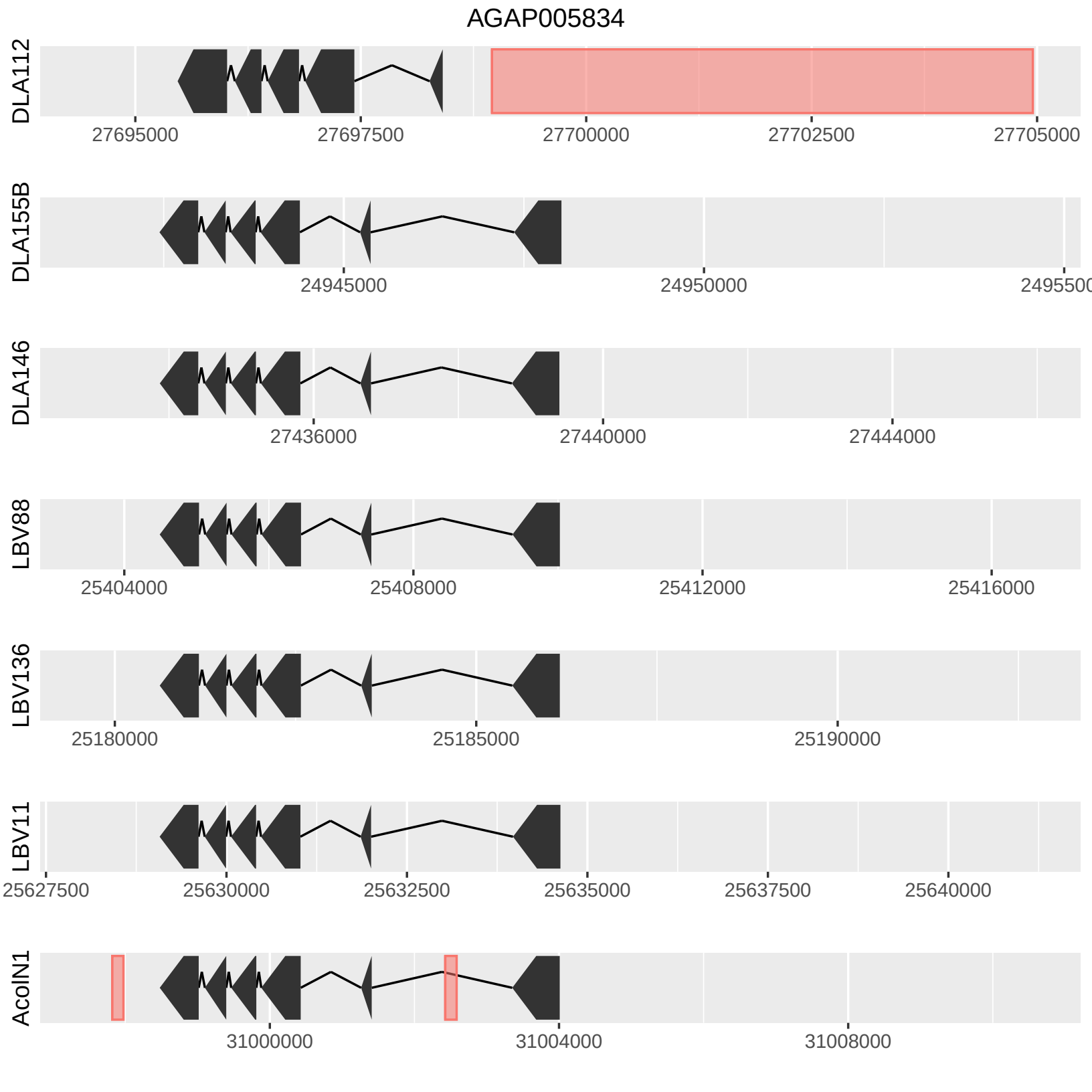

AGAP004163
DLA112
52324000
52326000
52328000
DLA155B
51228000
51230000
51232000
DLA146
53934000
53936000
53938000
53940000
LBV88
51704000
51706000
51708000
LBV136
51312000
51314000
51316000
LBV11
51030000
51032000
51034000
AcoIN1
51436000
51438000
51440000

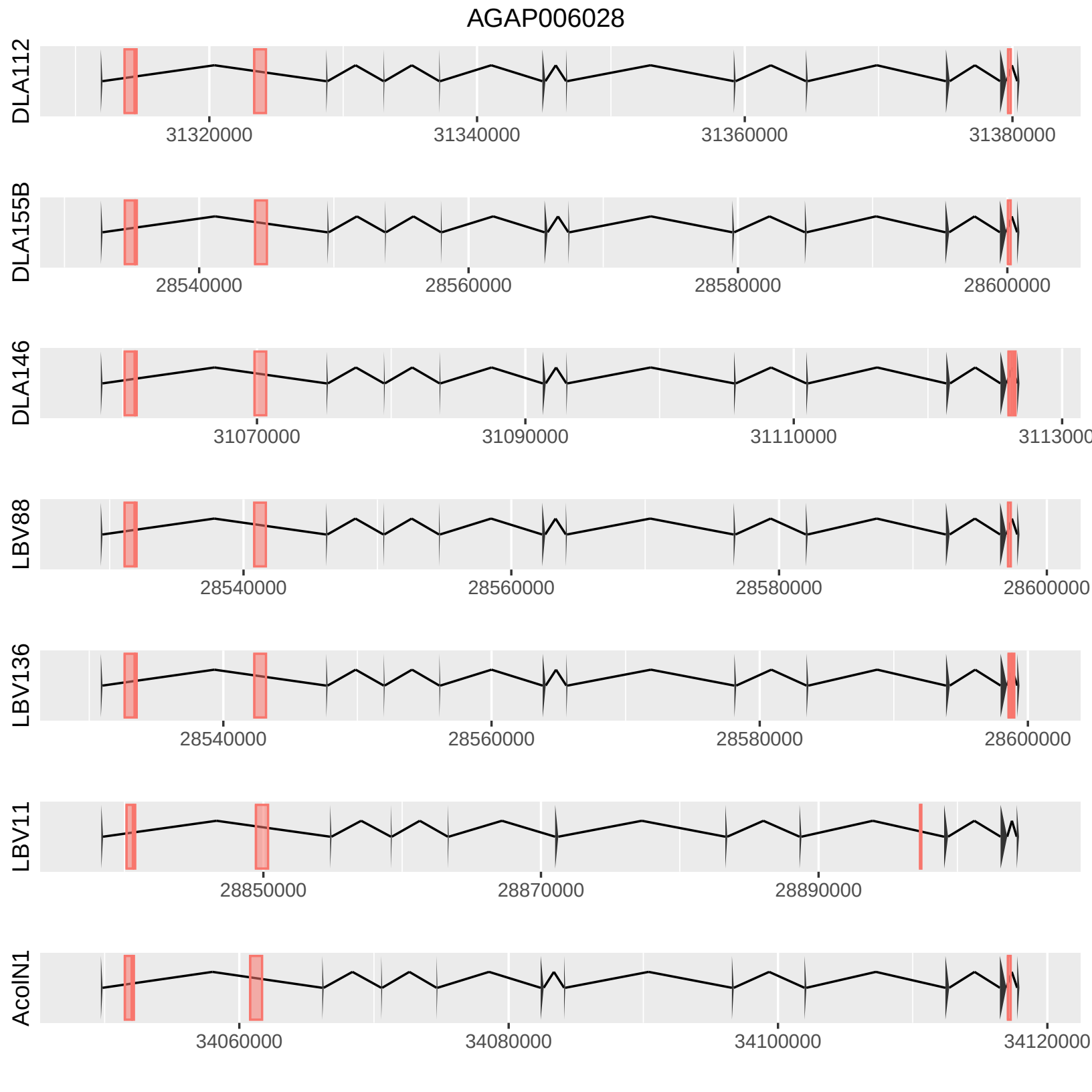

AGAP004164
DLA112
52332000
52334000
52336000
DLA155B
51236000
51238000
51240000
DLA146
53942000
53944000
53946000
53948000
LBV88
51710000
51712000
51714000
5171600
LBV136
51320000
51322000
51324000
LBV11
51038000
51040000
51042000
5104400
AcoIN1
51444000
51446000
51448000

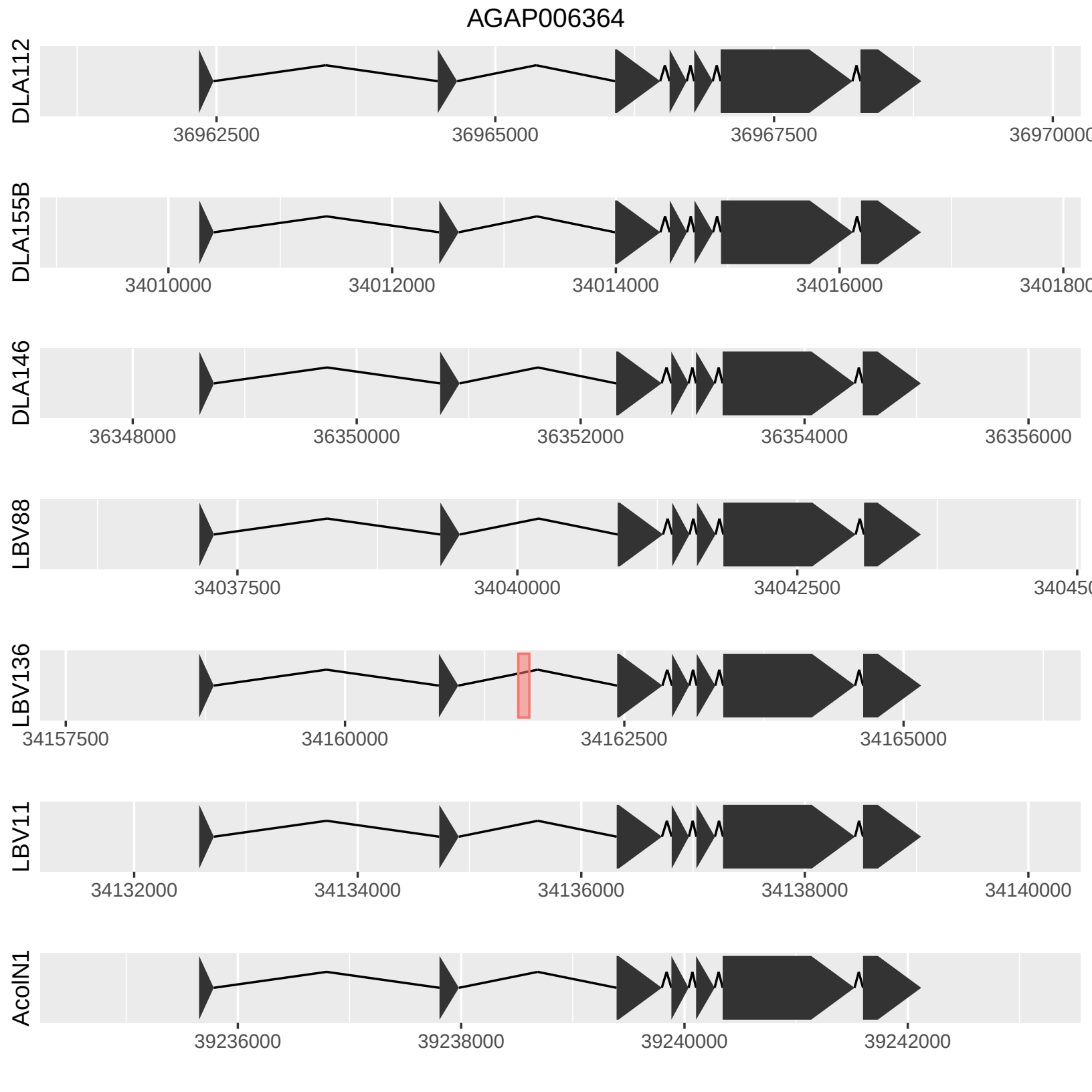

AGAP004382
DLA112
56980000
56981000
56982000
DLA155B
55801000
55802000
55803000
DLA146
58613000
58614000
58615000
LBV88
56122000
56123000
56124000
LBV136
55743000
55744000
55745000
55746000
LBV11
55362000
55363000
55364000
AcolN1
55469000
55470000
55471000

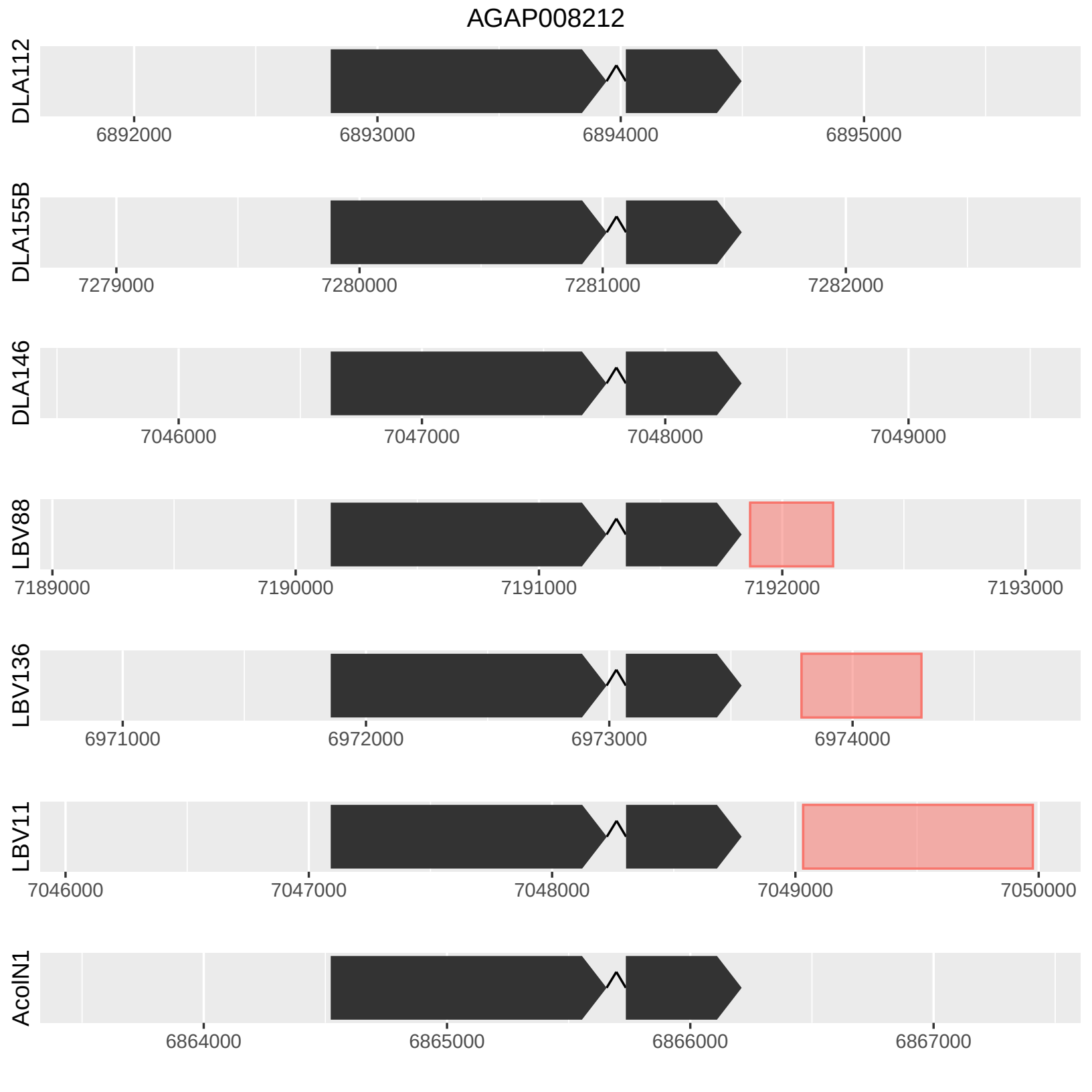

AGAP004707
DLA112
4330000
4350000
4370000
4390000
4410000
DLA155B
3475000
3500000
3525000
3550000
DLA146
4225000
4250000
4275000
LBV88
1100000
1120000
1140000
1160000
1180000
LBV136
29780000
29800000
29820000
LBV11
3460000
3480000
3500000
3520000
3540000
AcolN1
5560000
5580000
5600000
5620000

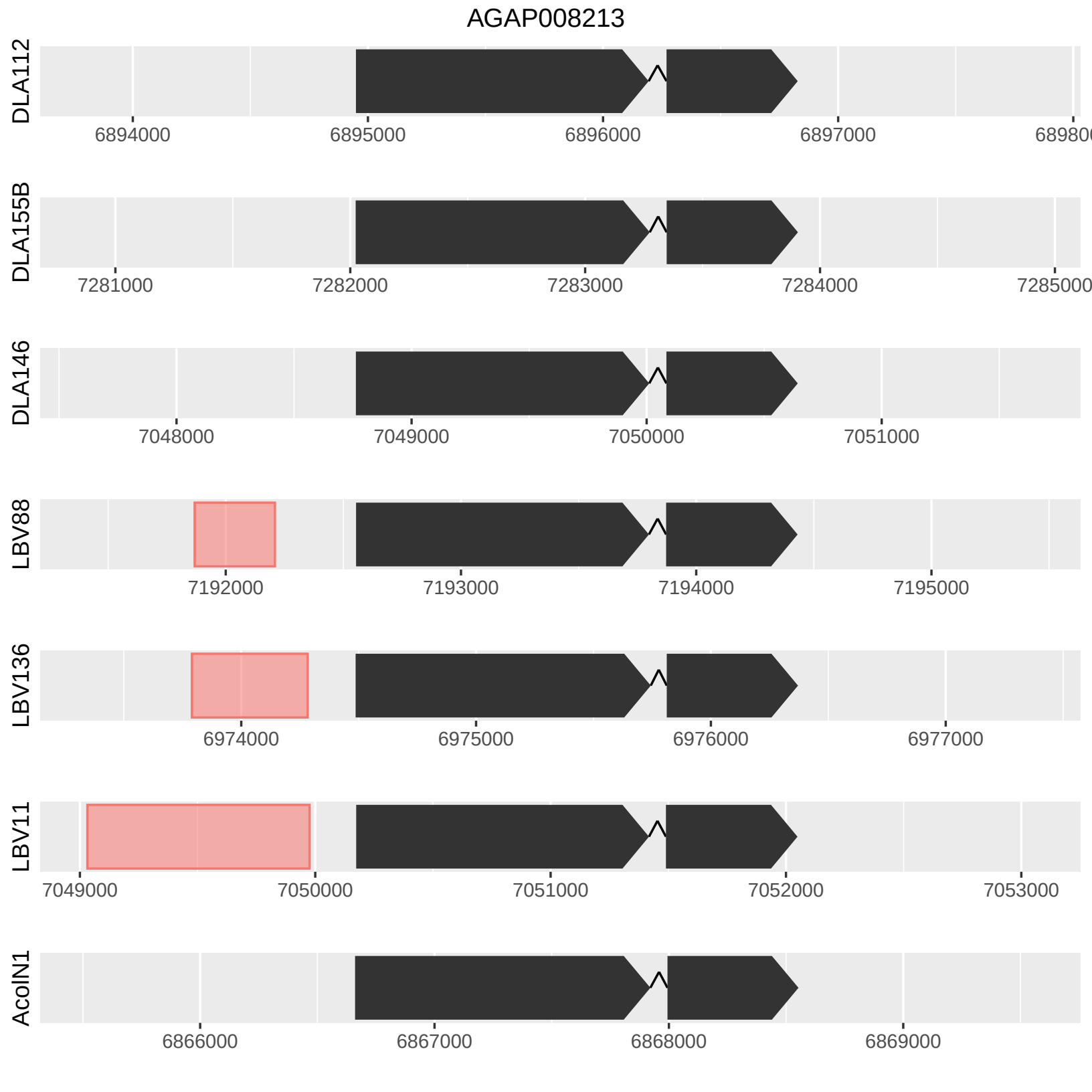

AGAP005749
DLA112
25216000
25217000
25218000
25219000
DLA155B
22856000
22857000
22858000
22859000
DLA146
25232000
25233000
25234000
25235000
LBV88
23252000
23253000
23254000
23255000
LBV136
23507000
23508000
23509000
23510000
23511000
LBV11
23422000
23423000
23424000
23425000
AcolN1
28929000
28930000
28931000
28932000

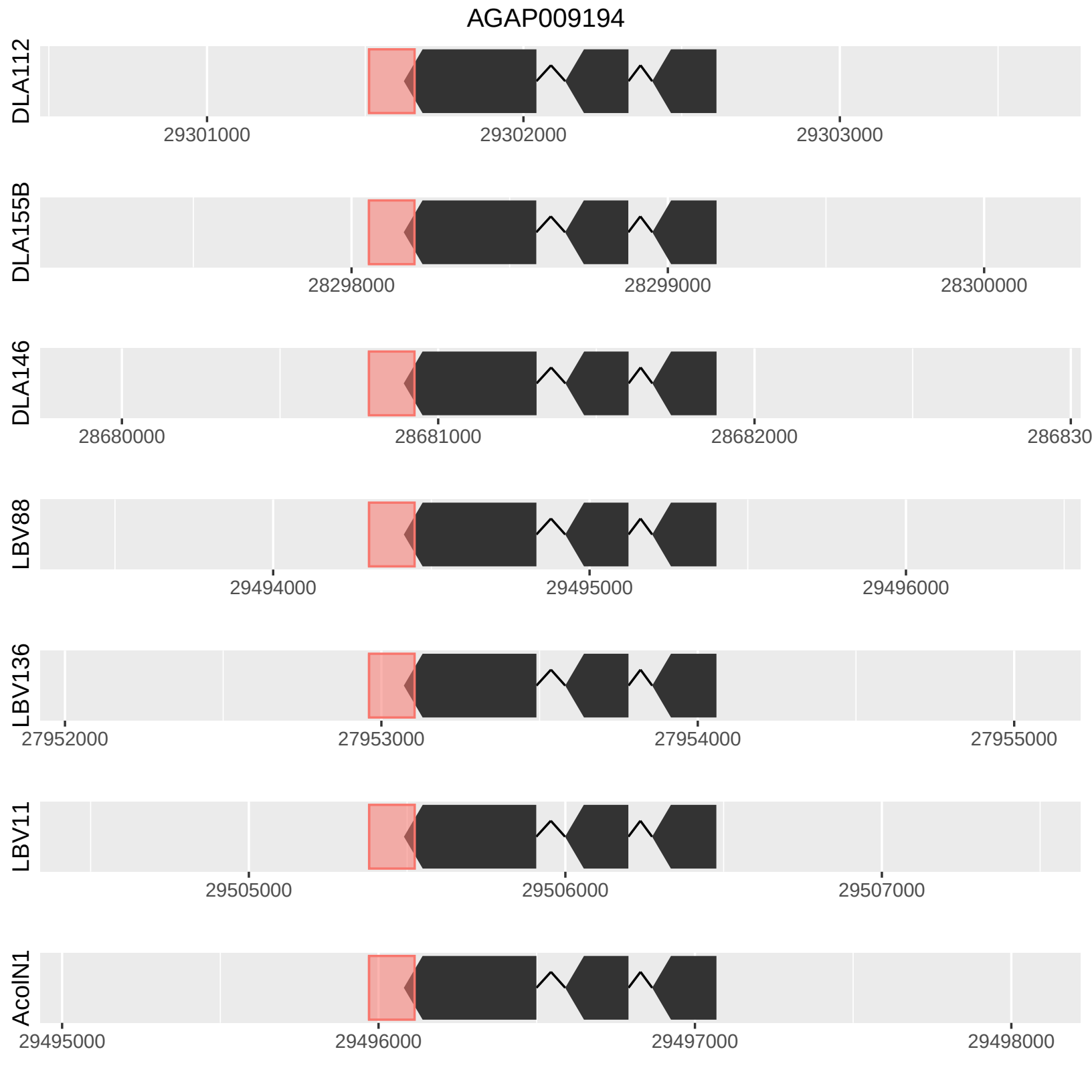

AGAP005834
DLA112
27695000
27697500
27700000
27702500
27705000
DLA155B
24945000
24950000
DLA146
27436000
27440000
27444000
LBV88
25404000
25408000
25412000
25416000
LBV136
25180000
25185000
25190000
LBV11
25627500
25630000
25632500
25635000
25637500
25640000
AcolN1
31000000
31004000
31008000

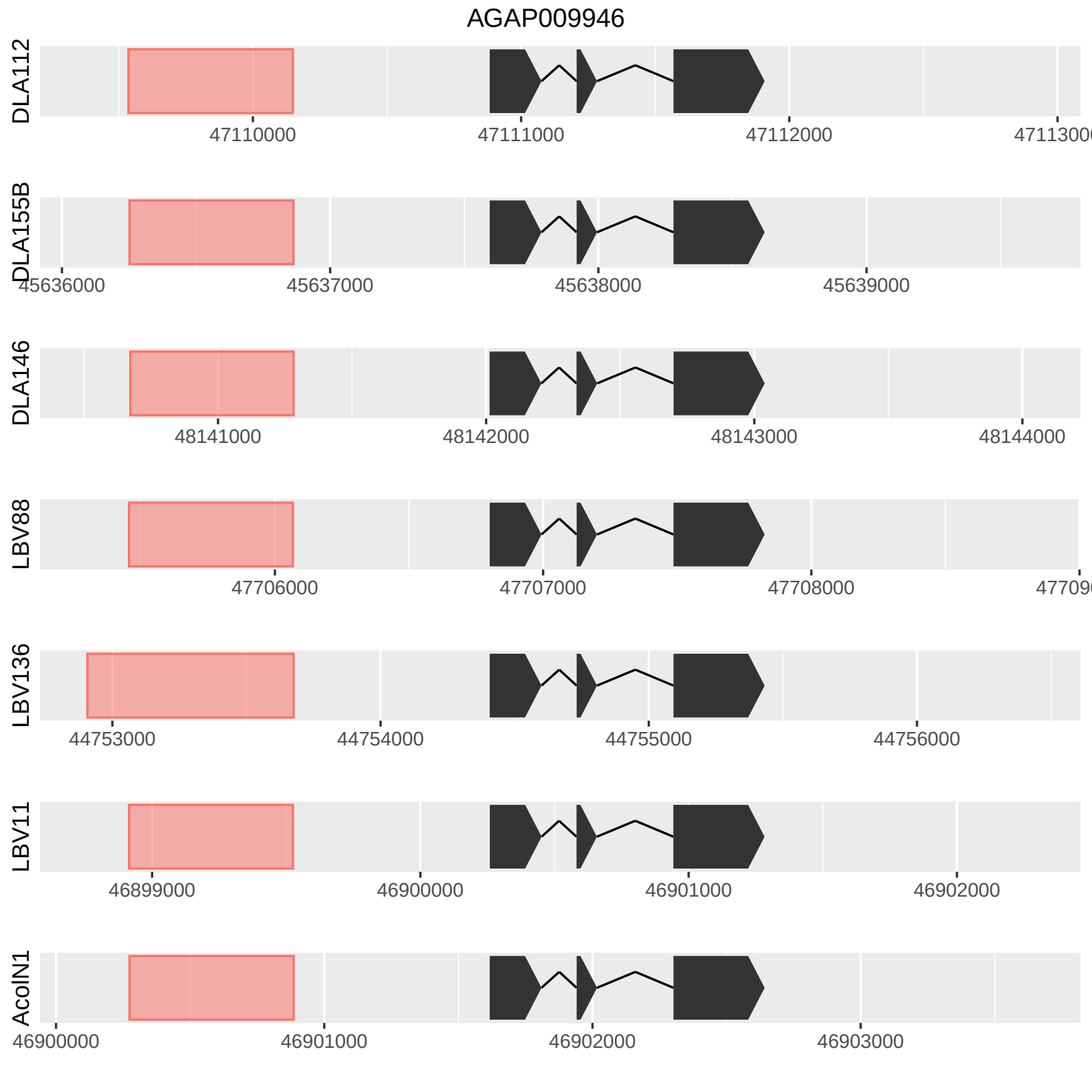

AGAP006028
DLA112
31320000
31340000
31360000
31380000
DLA155B
28540000
28560000
28580000
28600000
DLA146
31070000
31090000
31110000
3113000
LBV88
28540000
28560000
28580000
28600000
LBV136
28540000
28560000
28580000
28600000
LBV11
28850000
28870000
28890000
AcolN1
34060000
34080000
34100000
34120000

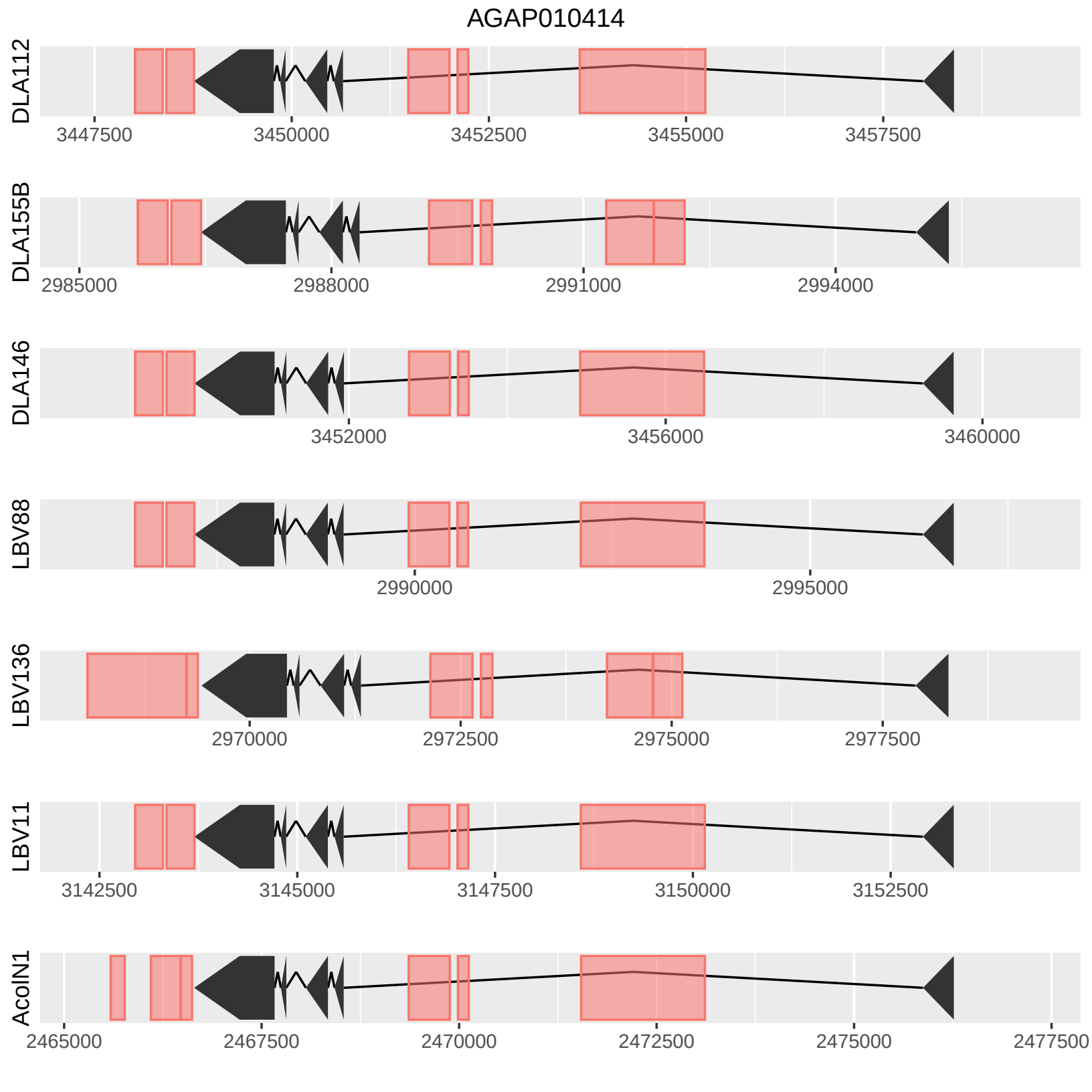

AGAP006364
DLA112
36962500
36965000
36967500
36970000
DLA155B
34010000
34012000
34014000
34016000
34018000
DLA146
36348000
36350000
36352000
36354000
36356000
LBV88
34037500
34040000
34042500
34045
LBV136
34157500
34160000
34162500
34165000
LBV11
34132000
34134000
34136000
34138000
34140000
AcolN1
39236000
39238000
39240000
39242000

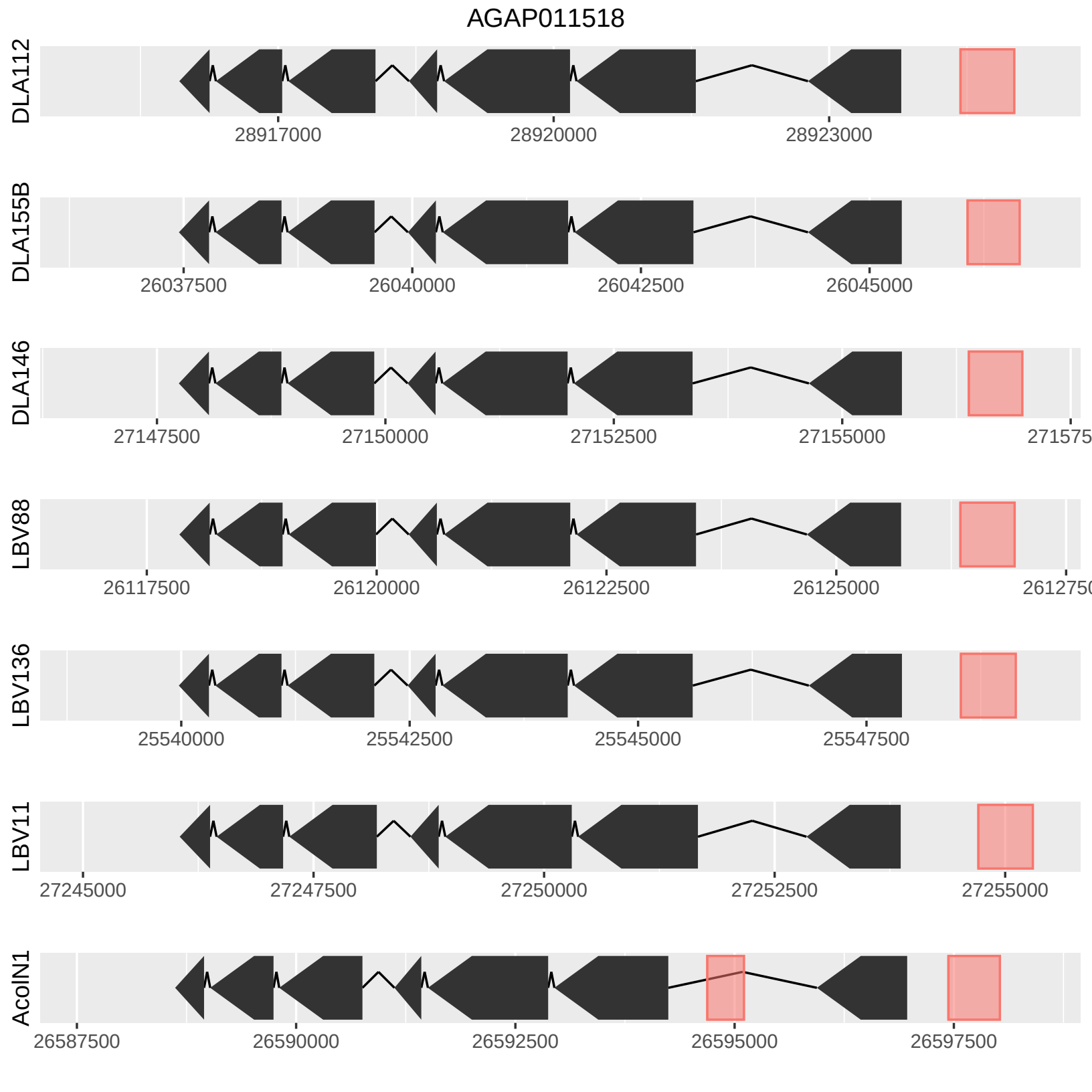

AGAP008212
DLA112
6892000
6893000
6894000
6895000
DLA155B
7279000
7280000
7281000
7282000
DLA146
7046000
7047000
7048000
7049000
LBV88
7189000
7190000
7191000
7192000
7193000
LBV136
6971000
6972000
6973000
6974000
LBV11
7046000
7047000
7048000
7049000
7050000
AcoIN1
6864000
6865000
6866000
6867000


AGAP008213
DLA112
6894000
6895000
6896000
6897000
DLA155B
7281000
7282000
7283000
7284000
7285000
DLA146
7048000
7049000
7050000
7051000
LBV88
7192000
7193000
7194000
7195000
LBV136
6974000
6975000
6976000
6977000
LBV11
7049000
7050000
7051000
7052000
7053000
AcoIN1
6866000
6867000
6868000
6869000


AGAP009194
DLA112
29301000
29302000
29303000
DLA155B
28298000
28299000
28300000
DLA146
28680000
28681000
28682000
286830
LBV88
29494000
29495000
29496000
LBV136
27952000
27953000
27954000
27955000
LBV11
29505000
29506000
29507000
AcoIN1
29495000
29496000
29497000
29498000


AGAP009946
DLA112
47110000
47111000
47112000
DLA155B
45636000
45637000
45638000
45639000
DLA146
48141000
48142000
48143000
48144000
LBV88
47706000
47707000
47708000
LBV136
44753000
44754000
44755000
44756000
LBV11
46899000
46900000
46901000
46902000
AcoIN1
46900000
46901000
46902000
46903000


AGAP010414
DLA112
3447500
3450000
3452500
3455000
3457500
DLA155B
2985000
2988000
2991000
2994000
DLA146
3452000
3456000
3460000
LBV88
2990000
2995000
LBV136
2970000
2972500
2975000
2977500
LBV11
3142500
3145000
3147500
3150000
3152500
AcoIN1
2465000
2467500
2470000
2472500
2475000
2477500


AGAP011518
DLA112
28917000
28920000
28923000
DLA155B
26037500
26040000
26042500
26045000
DLA146
27147500
27150000
27152500
27155000
271575
LBV88
26117500
26120000
26122500
26125000
261275
LBV136
25540000
25542500
25545000
25547500
LBV11
27245000
27247500
27250000
27252500
27255000
AcoIN1
26587500
26590000
26592500
26595000
26597500

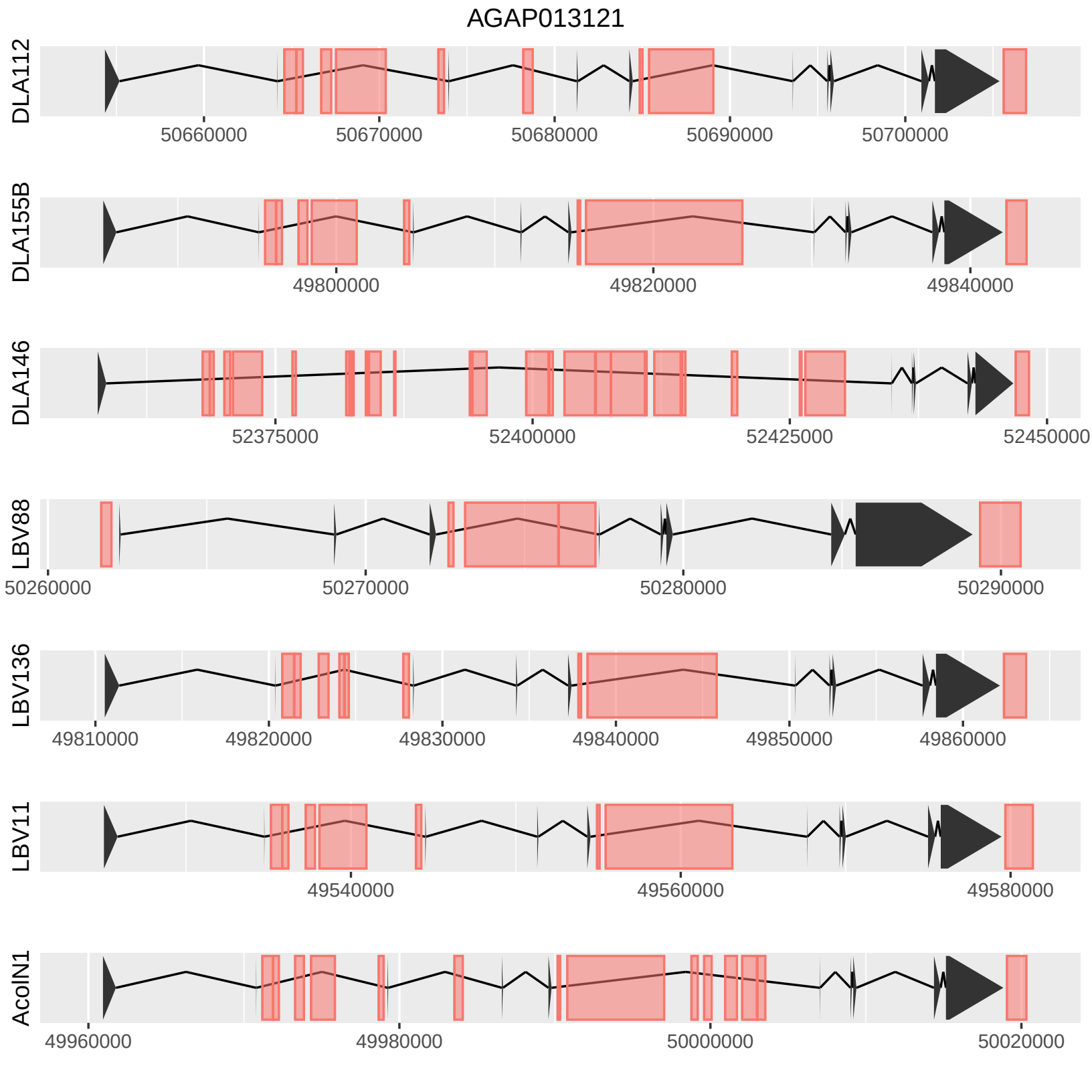
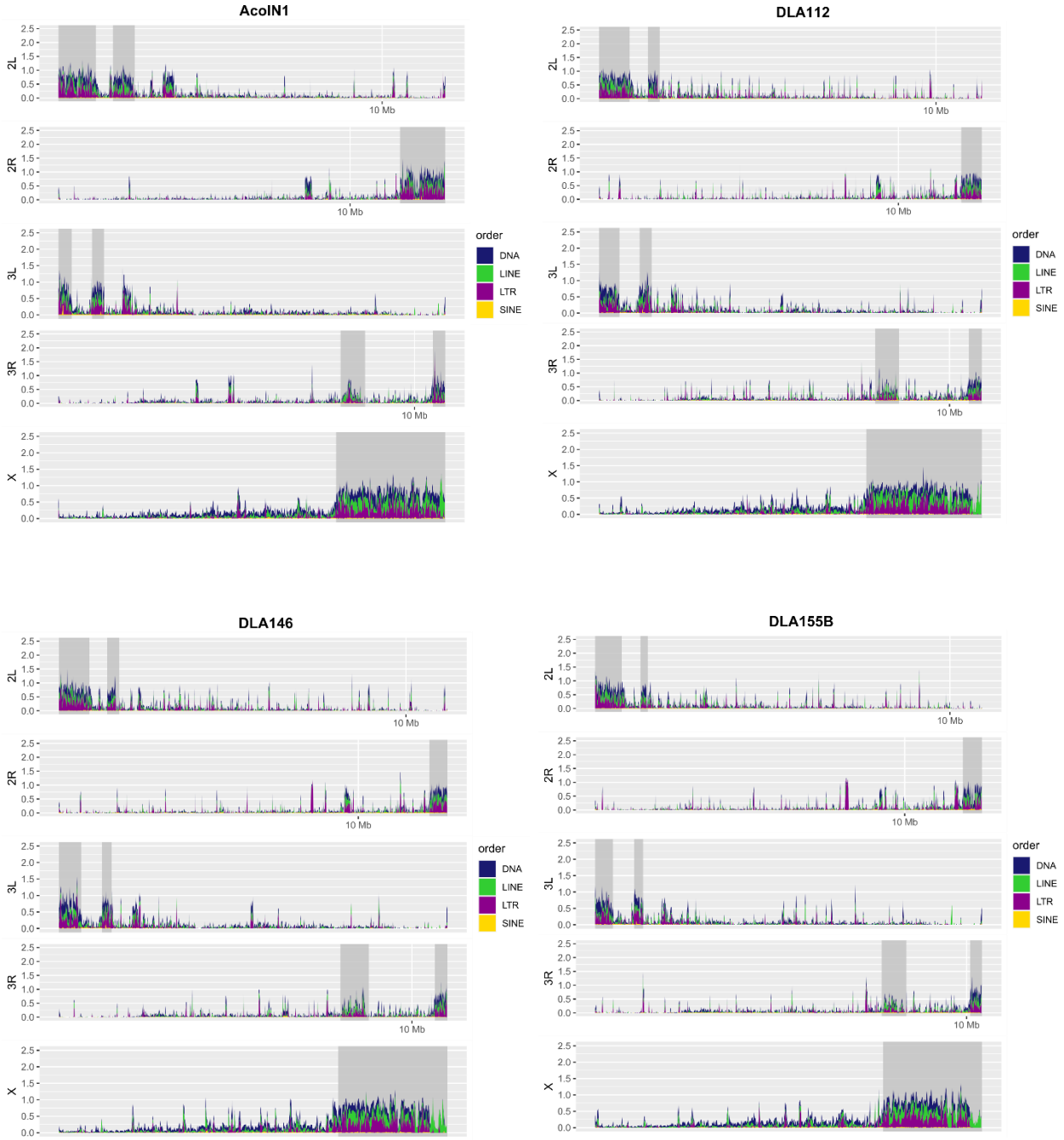
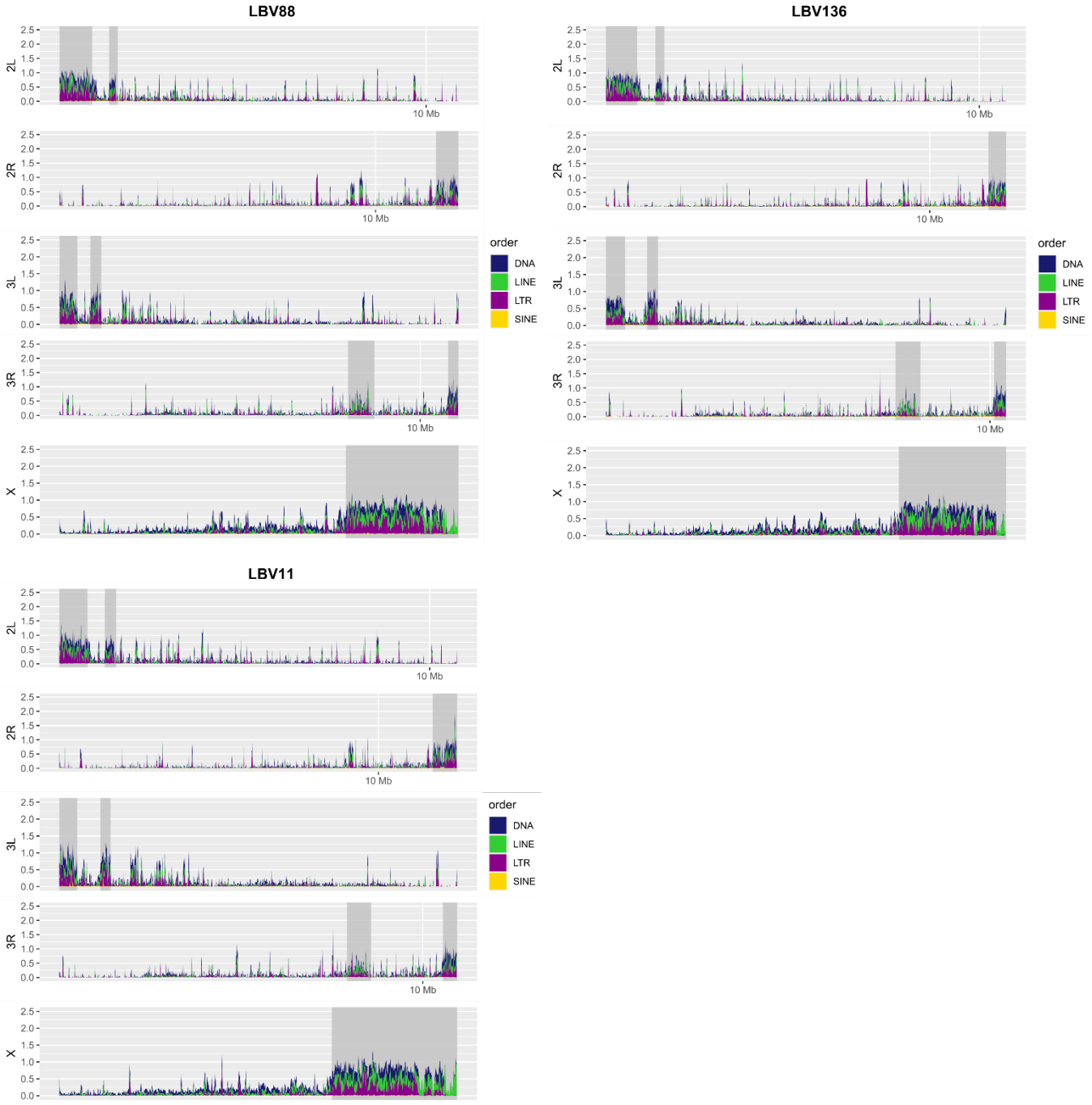

AGAP013121
DLA112
50660000
50670000
50680000
50690000
50700000
DLA155B
49800000
49820000
49840000
DLA146
52375000
52400000
52425000
52450000
LBV88
50260000
50270000
50280000
50290000
LBV136
49810000
49820000
49830000
49840000
49850000
49860000
LBV11
49540000
49560000
49580000
AcolN1
49960000
49980000
50000000
50020000

Figure S8. TE abundance plots by chromosome. The TE coverage by position is plotted. Grey boxes delimit the regions where the heterochromatin coordinates were transferred from the AgamP3 genome to each of the genomes used in this work.




LBV88
LBV136
LBV11
2L
2R
3L
3R
X
2.5
2.0
1.5
1.0
0.5
0.0
10 Mb
order
DNA
LINE
LTR
SINE

Supplemental Table Descriptions

Supplemental Table S1. General statistics, linear models, and euchromatin coordinates for all the genomes analyzed in this work. A) General statistics for the assembled genomes from the six sampled breeding sites. Whenever available, these statistics are also shown for the reference genome *AcolN1*. Scaffolding was performed using the AgamP4 chromosome level genome assembly. B) Linear models testing for significant associations between long-reads coverage, long-reads N50, number of contigs, N50 of contigs, complete BUSCO and number of transferred genes with different sequencing, assembly, and scaffolding statistics as well as different TE content and TE diversity values. FDR values smaller than 0.05 are highlighted in red. C) Coordinates for the euchromatin transferred using the reference from Sharakhova *et al*., (2010).

Supplemental Table S2. TEs identified using REPET. A) Number of TEs identified in each stage of the annotation process. TEs with MAP > 1 correspond to TEs annotated using more than one sequence. TEs with FC >3 correspond to TEs with more than three complete copies in the genome in a single or multiple blocks. TEs with FF >1 correspond to TEs with more than one complete copy in the genome in a single block. Reliable TEs correspond to TEs that upon observation of the coverage plots and the presence of features typical of their orders were considered to be accurate. TE consensus manually corrected correspond to TEs that although reliable required removal of features or extension, this category comprised mostly chimeric annotations. B) Families identified in each genome.

Supplemental Table S3. TE families annotated in *An. coluzzii*. A) "Known" describes whether the family had previously been described in another species (Yes) or first described in the present study (New)."Classification" describes whether families are degenerate (DG), putatively autonomous (PA), putatively autonomous but lacking TIRs or LTRs (PA*) or putatively non autonomous (PNA)." BBH in TEfam" refers to the ID of the bidirectional best hit in TEfam. This applies as well for BBH in AnnoTExcel and BBH in Repbase. B) Summary of the number of insertions or bases from each order or superfamily in each of the seven genomes.

Supplemental Table S4. Newly identified TE families. The 64 families were classified as follows: 32 DNA, 9 LINEs and 23 LTRs. Moreover, the majority of the new families (43/64) had partial matches to other known TEs, thus allowing us to classify them at the superfamily level. For each genome, the total number of insertions, the number of NFL (near full-length fragments, covering at least 90% of the consensus) and FL (full-length fragments, with differences in size with the consensus of less than 30 bp) insertions are given. The median number of insertions considering all strains and the fraction of insertions that are NFL are also given. The number of insertions with greater than 80% identity and covering at least 80% of the consensus in the reference genomes of several species of the Anopheles genus are shown and their columns colored differently according to the series they belong to. Aalb - *An. albimanus* (Nyssorhynchus), Aatr - *An. atroparvus* (Anopheles), Afar - *An. farauti* (Cellia [Neomyzomyia]), Afun - *An. funestus* (Cellia [Myzomyia]), Aste - *An. stephensi* (Cellia [Neocellia]), Aepi - *An. epiroticus*, Achr - *An. christyi*, Amer - *An. merus*, Agam - *An. gambiae*, Acol - *An. coluzzii*, Amel - *An. melas*, Aara - An. arabiensis, Aqua - *An. quadriannulatus*, Abwa - *An. bwambae* and

Afon *An. fontenillei* (Cellia [Pyretophorus]), Cqui - *Cx. quinquefasciatus*, Aaeg - *Ae. aegypti*, Dmel - *D. melanogaster*. The five anopheline species that do not belong to the Pyretophorus series were selected given that they had the most complete genomes in their taxonomic group.

Supplemental Table S5. Distribution of TE insertions regarding genes. A) χ^2 tests for the total number of insertions over the size of each compartment. The expected values are calculated determining the number of insertions that would be expected given the size of the compartment. The residuals for each category are shown. **B)** χ^2 tests for the number of insertions group by order in each compartment by strain. The residuals are shown, in yellow positions with absolute values larger than 5 and in red positions with absolute values larger than 10, and thus with more weight. **C)** Shows the TEs grouped by superfamily.

Supplemental Table S6. Inversion breakpoints in *An. coluzzii*. The known coordinates or the breakpoints of five inversions were transferred from the An. gambiae genome using the Corbett-Detig *et al*., 2019 reference. Given that the distal 2Ru breakpoint is shared with the 2Rc proximal and 2Rd distal we were able to include it as well but we did not take into consideration the proximal 2Ru given that the exact coordinates for this breakpoint are not known. The reference column includes information of the experimental characterization of the inversion.

Supplemental Table S7. Identification and analysis of active TE families. **A)** Strains by dnaPipeTE refers to the number of strains where more than 50% of the insertions are >99% identical to the consensus according to dnaPipeTE. Identical using BLAST is the median percentage of insertions >99% identical to the consensus in all seven strains. Strains with >2 identical FL is the number of strains with more than two identical full-length insertion. Left and right side LTR are the coordinates of the identified LTRs on the consensus sequence, ORF1 and ORF2 are the identified open reading frames in each family. TSD describes the identified length of the target site duplication. **B)** Identities between both LTRs in full-length insertions of the six putatively active LTRs.

Supplemental Table S8. TE insertions from active families in proximity of genes. A) Presence or absence of the active insertions by genome is shown. **B)** Genes with TE insertions from active families. In yellow the six genes selected for further analysis. Status refers to whether the insertion(s) found on each of the genes is (are) fixed, polymorphic or found on a single genome.

Supplemental Table S9. Transcription factor binding sites (TFBS) and promoter prediction for TEs belonging to the active families and insecticide resistance genes. A) Position and score for each identified binding site of the transcription factors *cnc:mafS*, dl and *Stat92E*. **B)** Summary of the TFBS identified per genome. **C)** Promoter motifs in the TE insertions in genes involved in insecticide resistance and in insertions from active families located in gene upstream regions. In orange insertions that contained one of the promoter motifs (TATA box, Drosophila Initiator or Drosophila TCT), in blue insertions that contained a pair of the promoter motifs (TATA and either Initiator or TCT) and the distance between these two motifs was the expected one (Sloutskin *et al* 2015).

Supplemental Table S10. TEs in genes which are differentially expressed upon exposure to insecticides. In yellow eight genes well-known to be involved in insecticide response.

Supplemental Table S11. TE frequency estimation in rural populations using PoPoolationTE2. A) Accession number of the 37 genomes from Burkina Faso and the 30 genomes from Cote d'Ivoire analyzed. **B)** Frequency of the 59 TE insertions located inside or nearby insecticide-related genes in the Burkina Faso (BF) and Cote d'Ivoire (CI) populations and in our genomes. In red frequencies below 10%. **C)** Frequency of the 438 TE insertions located inside or nearby immune-related genes in the BF and CI populations and in our genomes. In red frequencies below 10%.

Supplemental Table S12. TE insertions in immune-related genes. Genes were obtained from ImmunoDB (Waterhouse *et al*., 2007). "Promoter" can be YES or NO if found or not in the TE, "*dorsal*" and "*STAT*" are the number of binding sites, "# of genomes" is the number of genomes where the TE is found and columns with the genome names can be YES or NO if the insertion is present or not in that particular genome. "Plasmodium responsive" indicates if the gene has been identified as a plasmodium responsive gene according to Ruiz *et al*., (2019).

Supplemental Table S13. Comparison of the TE copies age distribution (TE landscape) obtained with dnaPipeTE and estimating the identity of the insertions identified by REPET using BLAST. The distributions were compared using a Kolmogorov-Smirnov test and the p-values are reported. Significant differences between superfamilies in each strain are highlighted in yellow.

Supplemental Files

Supplemental File 1. GFF file with the TE annotation for the *AcolN1* genome.

Supplemental File 2. GFF file with the TE annotation for the *DLA112* genome.

Supplemental File 3. GFF file with the TE annotation for the *DLA146* genome.

Supplemental File 4. GFF file with the TE annotation for the *DLA155B* genome.

Supplemental File 5. GFF file with the TE annotation for the *LBV11* genome.

Supplemental File 6. GFF file with the TE annotation for the *LBV88* genome.

Supplemental File 7. GFF file with the TE annotation for the *LBV136* genome.

Supplemental File 8. GFF file with the gene annotation for the *AcolN1* genome.

Supplemental File 9. GFF file with the gene annotation for the *DLA112* genome.

Supplemental File 10. GFF file with the gene annotation for the *DLA146* genome.

Supplemental File 11. GFF file with the gene annotation for the *DLA155B* genome.

Supplemental File 12. GFF file with the gene annotation for the *LBV11* genome.

Supplemental File 13. GFF file with the gene annotation for the *LBV88* genome.

Supplemental File 14. GFF file with the gene annotation for the *LBV136* genome.

Supplemental File 15. TE library for A. coluzzii comprising all 520 families in fasta format.

Supplemental File 16. TEs transferred across all genomes in txt format. The first three columns include the chromosome, start and end positions for each insertion are reported. The fourth column the TE family name with a unique number appended after an underscore. Columns five to 11 describe if that insertion was found (PRESENT) or not (ABSENT) or if no information was available (NA).

Supplemental Methods

Sample collection and DNA isolation

We sampled *An. coluzzii* larvae in two cities of Central Africa: Libreville, Gabon, in January 2016 and Douala, Cameroon, in April 2018 (Additional File 1: Table S1). A systematic inspection of potential breeding sites was conducted to determine the presence of *Anopheles* larvae. We manually separated the anopheline from the culicine larvae based on morphological recognition and positioning of their bodies on or under the water surface (Robert, 2017). We collected immature 3rd and 4th stage larvae of *Anopheles* from water bodies using the standard dipping method (Service, 1993). Larvae were stored in 1.5 ml of absolute ethanol. After each daily sampling session, the samples were stored at -20 °C. All the samples were PCR tested to differentiate *An. coluzzii* larvae from *An. gambiae* larvae before library preparation, using primers SINE200_F (TCGCCTTAGACCTTGCGTTA) and SINE200_R (CGCTTCAAGAATTCGAGATAC) (Santolamazza *et al.*, 2008). These primers target a single copy *SINE200* transposable element insertion that is fixed in *An. coluzzii* and absent in *An. gambiae*.

For PacBio sequencing, DNA from a single *An. coluzzii* larva from the *LBV11* site was extracted using the MagAttract HMW DNA extraction kit (Qiagen) following manufacturer's instructions. Briefly, the larva was air-dried and lysed in 240 µl of buffer ATL (proteinase K added) shaking overnight at 56 ºC. Next, the DNA was isolated using the MagAttract magnetic beads and eluted twice in 50 µl of buffer AE. The DNA concentration was measured using a Qubit fluorometer.

For Nanopore sequencing, DNA from six larvae from each of the five breeding sites was extracted either with the QiaAMP UCP DNA kit (Qiagen) or MagAttract HMW DNA extraction kit (Qiagen). We performed individual larvae extractions as our objective was to use the minimum number of larvae possible to avoid the presence of excess polymorphisms that could affect the genome assembly. For the QiaAMP UCP DNA kit, we followed the manufacturer's instructions. Each larva was air-dried and lysed in 200 µl of buffer AUT (proteinase K added) shaking overnight at 56 ºC, then DNA was isolated using a QIAamp UCP MinElute column and eluted twice in 25 µl of buffer AUE. For the MagAttract HMW DNA extraction kit, we followed manufacturer's instructions but using lower buffer amounts to increase DNA concentration. Briefly, each larva was lysed in 120 µl of buffer ATL (proteinase K added) shaking overnight at 56 ºC, then DNA was isolated using the MagAttract magnetic beads and eluted twice in 25 µl of

buffer AE. The DNA concentration was measured using a Qubit fluorometer. Both elutions of the same sample were mixed before library preparation. For Illumina sequencing, DNA from one larva from each of the six different breeding sites was extracted following the same extraction protocol as for Nanopore sequencing.

Genome Assemblies

The PacBio sequenced genome was assembled using Canu version 1.8 (Koren *et al.*, 2017) with an estimated genome size of 250Mb and parameters: '*stopOnLowCoverage*=5, *corMinCoverage*=0, *correctedErrorRate*=0.105, *CorMhapFilterThreshold*=0.0000000002, *corMhapOptions="--threshold 0.80 --num-hashes 512 --num-min-matches 3 --ordered-sketch-size 1000 --ordered-kmer-size 14 --min-olap-length 2000 --repeat-idf-scale 50*" *mhapMemory*=60g, *mhapBlockSize*=500, *ovlMerDistinct*=0.975'. The parameter stopOnLowCoverage was set to 5 to prevent fragmentation given that some of our samples had medium coverage. corMinCoverage was set to 0 to conserve the full length of the reads during the correction stage. correctedErrorRate was set to 0.105 following the recommendations in Canu's manual for low coverage genome assemblies. All of the remaining parameters were set to reduce disk space and run time following the recommendations for repetitive genomes. Next, we identified and removed allelic variants using Purge Haplotigs version 1.0.4 (Roach, Schmidt and Borneman, 2018) with the "*-l 15 -m 100 -h 195*" parameters.

The Nanopore genomes were assembled using Canu version 1.8 using the same parameters as previously described, except for *correctedErrorRate* which was set to 0.16, followed by a round of polishing using *racon* version 1.3.3 (Vaser *et al.*, 2017), followed by nanopolish version 0.11.1 (Loman, Quick and Simpson, 2015) and Pilon version 1.23-0 (Walker *et al.*, 2014) with the fix parameter set on '*bases*'. Pilon requires high coverage short-read data to perform the polishing and these data came from the aforementioned single larvae sequenced from each of the sites. Polishing was not performed for the PacBio sequenced genome given that its assembly had been performed using a single individual and we did not want to incorporate additional variants from the Illumina sequenced individual. In the case of the Nanopore genomes, and although we cannot discard that using Illumina data from an additional individual could introduce novel variants, we did perform the polishing given that those genomes had been obtained using several individuals and polishing increased the BUSCO values. Next, we identified and removed allelic

variants using Purge Haplotigs version 1.0.4 (Roach, Schmidt and Borneman, 2018) with the "*-l 15 -m 100 -h 195*" parameters. Finally, BlobTools version 1.1.1 (Laetsch and Blaxter, 2017) was used to remove contamination from all six genome assemblies taking into consideration fragment sizes, their taxonomic assignation and the coverage using the Illumina reads.

Construction of the curated TE library and *de novo* TE annotation

We ran the TEdenovo pipeline (Flutre *et al.*, 2011) independently on each of the seven genomes with default parameters. The obtained consensus in each genome were further filtered by discarding those generated with (i) only one sequence; (ii) with less than one full-length fragment mapping to the genome; (iii) with less than three full-length copies; and (iv) shorter than 100bp (Additional file 1: Table S2). The remaining consensuses were manually curated to remove redundant sequences and artifacts by manual inspection of coverage plots generated using the plotCoverage tool from REPET and visualization of the structural features on the genome browser IGV version 2.4.19 (Robinson *et al.*, 2011).

To ensure that we identified as much of the TE diversity as possible, the TEfam (tefam.biochem.vt.edu) database, which contains the TE libraries for several species of mosquitoes, was used to annotate the seven genomes using RepeatMasker version open-4.0.9 (Smit *et al*. 2015). Because TEfam contains sequences from different species, we used a conservative approach to minimize false positives. We required candidate families to have more than three matches longer than 90% in any genome. Then, the hit with the highest identity from each genome was extracted. These sequences were added to the REPET library and all the consensuses were clustered using CD-HIT version 4.8.1 (Fu *et al.*, 2012) with the *-c* and *-s* parameters set to 0.8. These filters ensured that all TEs with an identity greater than 80% throughout more than 80% of their sequence were grouped in the same family. 85 clusters contained sequences only identified by TEfam. The sequences belonging to the same cluster were used to perform a multiple sequence alignment and the consensuses were obtained.

The consensuses were classified using PASTEC (Hoede *et al.*, 2014) with default parameters. Next their bidirectional best-hits were calculated using BLAST (Camacho *et al.*, 2009) against the TEfam (tefam.biochem.vt.edu), AnoTExcel (Fernández-Medina, Struchiner and Ribeiro,

2011) and Repbase (Bao, Kojima and Kohany, 2015) databases. When more than 80% of a consensus matched to a feature from the databases with an identity higher than 80%, the classification was transferred to the consensus. While not an order *per se*, MITEs were grouped together for subsequent analysis. Additionally, we classified the families based on the conservation of features characteristic of their orders into putative autonomous, putative autonomous lacking terminal inverted repeats (TIRs) or long terminal repeats (LTRs), putative non-autonomous, such as MITEs and TRIMs, and degenerated (Additional file 1: Table S4) (Fonseca *et al* 2019). These classified consensuses were used to re-annotate the assembled genomes with the TEannot pipeline using default parameters and we discarded copies whose length overlapped >80% with satellite annotations (Quesneville *et al.*, 2005).

Transfer of TE annotations to the *AcolN1* reference genome

We transferred the euchromatic TE annotations from the six genomes we sequenced to the *AcolN1* genome. First, we built a GFF file composed by the coordinates of two 500 bp long "anchors" adjacent to each TE. We transferred these features considering each pair of anchors as exons from a single gene using the Liftoff tool with the -exclude_partial -overlap 1 -s 0.8 parameters (Shumate and Salzberg, 2021). We conserved only transfers where both anchors were transferred to the *AcolN1* genome. When both anchors were separated by less than 10 bp we considered the TE to be absent. This allowed us to identify most TE insertions, both present and absent in the *AcolN1* genome. Next, following the same strategy we transferred these regions from the *AcolN1* genome to the other six genomes. This was performed to allow the identification of TEs that were indeed present in these genomes but that had not been initially annotated by REPET. To summarize this information, we built a matrix containing the status for all of the TE insertions transferred to the *AcolN1* genome in every genome. When the anchors were found more than 10 bp away the TE was considered to be present and this was represented with a 1, when the distance was less than 10 bp the TE was considered to be absent, and it was represented with a -1. When any of the anchors was not transferred the TE was considered to be not transferred and represented with a 0. For a graphical description see Figure 1.

Overall, we were able to transfer to the *AcolN1* genome 71.93% to 75.29% of the TEs present in each of the six genomes leading to a total of 67,548 TEs transferred. We then attempt to transferred back these 67,548 from the *AcolN1* reference to the remaining six genomes and we

were able to confidently transfer 53,893 (79.78%) of these regions to at least three of the genomes, with 32,185 (47.65%) transferred to all six genomes. We checked whether the TEs that failed to be transferred were enriched for nested TEs and we found that this was the case: while 68.58% and 51.18% of the TEs located more and less than 500 bp away from other TEs, respectively, were transferred, only 38.26% of the TEs overlapping other TEs were transferred to the six genomes.

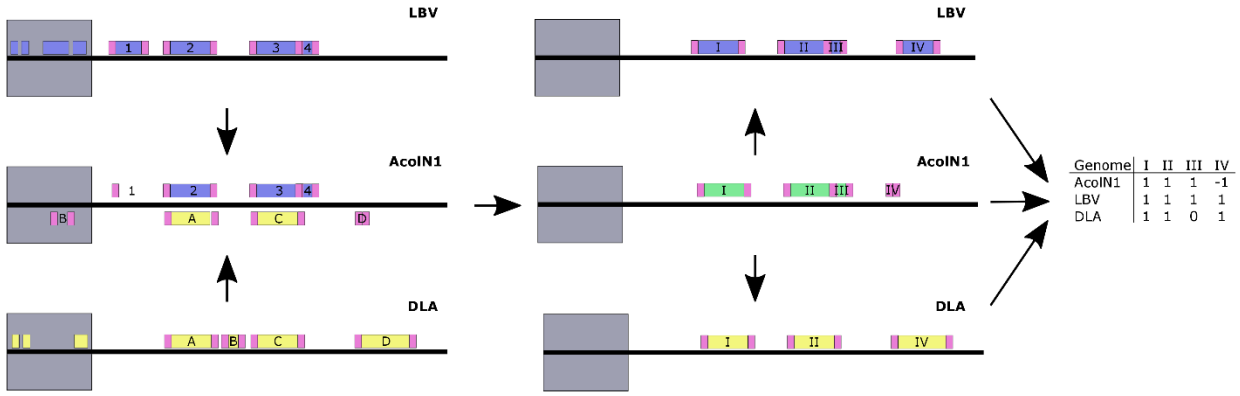


Figure 1. Transfer of TE annotations. Three genomes are shown (LBV, DLA and the reference *AcolN1*) as black lines. The heterochromatin in each genome is displayed using a gray box. The euchromatic TEs from genomes LBV (shown as blue boxes) and DLA (shown as yellow boxes) are transferred to the *AcolN1* genome. This is performed by obtaining 500 bp long anchors from each side of the TE (shown as pink boxes). Then using Liftoff these anchors are transferred to the *AcolN1* genome.

Note that the anchors of TE B were transferred to the heterochromatin and thus this TE is discarded. Also, only one anchor of TE 1 was transferred, therefore it was also discarded. If the distance between both anchors was smaller than 10 (TE D), the TE was considered to be absent in the *AcolN1* genome, however these regions were conserved to assess whether these TEs were present in other genomes. The threshold of 10, instead of a fraction of the size of the TE, was selected to detect solo-LTR insertions given that LTRs represent a small fraction of the total size of the full element. In the case of regions where TEs from two genomes were transferred (for example TE 2 and TE A) if they belonged to the same family the coordinates of the TE with a size closer to the consensus of that family were conserved.

Next the anchors for all TEs that had been transferred to the *AcolN1* genome were transferred to every other genome. Notice that TE IV had not been discovered during the LBV annotation, but it was identified during the transfer. Also, the anchors for TE III were not transferred to the DLA genome. Finally, all this information was summarized in a matrix where 1 represents that the TE was identified, -1 that the TE was absent and 0 when the anchors were not transferred and thus no information of the presence of the TE is available.

Identification of newly described families in other species

We analyzed all 10 available fully sequenced species from the Pyretophorus series, which belongs to the Cellia subgenus. We also included an additional five *Anopheles* species, three from each of the other series from the Cellia subgenus and two from the other subgenera with available fully sequenced species. As outgroups we included the genomes of *Cx. quinquefasciatus*, *Ae. aegypti* and *D. melanogaster*. RepeatMasker version open-4.0.9 (Smit *et al*. 2015) was run with default parameters using the 64 newly described families as the library on the following genomes: *An. albimanus* (AalbS2), *An. atroparvus* (AatrE3), *An. farauti* (AfarF2), *An. funestus* (AfunF3), *An. stephensi* (AsteS1), *An. epiroticus* (AepiE1), *An. christyi* (AchrA1), *An. merus* (AmerM2), *An. gambiae* (AgamP4), *An. coluzzii* (*AcolN1*), *An. melas* (AmelC2), *An. arabiensis* (AaraD1), *An. quadriannulatus* (AquaS1), *An. bwambae* (Abwa2) and *An. fontenillei* (ASM881789v1), *Cx. quinquefasciatus* (CulPip1.0), *Ae. aegypti* (AaegL5.0) and *D. melanogaster* (ISO1 release 6).

Insertion frequency estimation in rural populations

We used the PoPoolationTE2 v-1.10.03 pipeline (Kofler, Gómez-Sánchez and Schlötterer, 2016) to compute the TE insertion frequencies in the Ivory Coast and Burkina Faso populations from the Ag1000G project (Anopheles gambiae 1000 Genomes Consortium, 2020).

The TE hierarchy and the TE-merged reference files were created using the TE library previously created and the *AcolN1* reference genome. We used RepeatMasker version open-4.0.9 (Smit, Hubley and Green, 2013-2015) to mask the TE sequences in the reference and concatenated the masked reference file with the full TE library fasta. Next, we mapped the sequencing data from each sample in fastq format against the concatenated reference using bwa version 0.7.12-r1039 (Li and Durbin, 2009) with default parameters. The obtained files were

converted to bam format, sorted and indexed using samtools version 1.9-55-g8a4d196 (Li *et al.*, 2009).

Using the ppileup tool from PopoolationTE2 we generated a pileup from all the bam files. Next, the signatures were identified with the identifySignatures tool using the separate mode and requiring a minimum count of 3. The output file was processed with the frequency tool and the pairupSignatures tool using a minimum distance of 200 and a maximum distance of 300 as recommended in the manual. Once we obtained the insertion locations we converted them to GenomicRanges objects ignoring strand information and used the findOverlaps function from the GenomicAlignments R package (Lawrence *et al.*, 2013) to identify overlaps with the previously identified TEs of interest. We only considered matches where the overlapping TEs belonged to the same superfamily.

Supplemental References

Bao, W., Kojima, K. K. and Kohany, O. (2015) 'Repbase Update, a database of repetitive elements in eukaryotic genomes', *Mobile DNA,* 6(1), pp. 4-9.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T. L. (2009) 'BLAST+: architecture and applications', *BMC bioinformatics,* 10, pp. 421-421.

Anopheles gambiae 1000 Genomes Consortium (2020) 'Genome variation and population structure among 1,142 mosquitoes of the African malaria vector species Anopheles gambiae and Anopheles coluzzii', *Genome Research*, 30(10):1533-1546.

Fernández-Medina, R. D., Struchiner, C. J. and Ribeiro, J. M. C. (2011) 'Novel transposable elements from Anopheles gambiae', *BMC Genomics,* 12(1), pp. 260-260.

Flutre, T., Duprat, E., Feuillet, C. and Quesneville, H. (2011) 'Considering Transposable Element Diversification in De Novo Annotation Approaches', *PLOS ONE,* 6(1), pp. e16526.

Fu, L., Niu, B., Zhu, Z., Wu, S. and Li, W. (2012) 'CD-HIT: accelerated for clustering the next-generation sequencing data', *Bioinformatics (Oxford, England),* 28(23), pp. 3150-3152.

Hoede, C., Arnoux, S., Moisset, M., Chaumier, T., Inizan, O., Jamilloux, V. and Quesneville, H. (2014) 'PASTEC: An Automatic Transposable Element Classification Tool', *PLOS ONE,* 9(5), pp. e91929.

Kofler, R., Gómez-Sánchez, D. and Schlötterer, C. (2016) 'PoPoolationTE2: Comparative Population Genomics of Transposable Elements Using Pool-Seq', *Molecular biology and evolution,* 33(10), pp. 2759-2764.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H. and Phillippy, A. M. (2017) 'Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation', *Genome research,* 27(5), pp. 722-736.

Laetsch, D. R. and Blaxter, M. L. (2017) 'BlobTools: Interrogation of genome assemblies', *F1000Research,* 6(1287).

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T. and Carey, V. J. (2013) 'Software for Computing and Annotating Genomic Ranges', *PLOS Computational Biology,* 9(8), pp. e1003118.

Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows–Wheeler transform', *Bioinformatics,* 25(14), pp. 1754-1760.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics,* 25(16), pp. 2078-2079.
Loman, N. J., Quick, J. and Simpson, J. T. (2015) 'A complete bacterial genome assembled de novo using only nanopore sequencing data', *Nature Methods,* 12(8), pp. 733-735.
Quesneville, H., Bergman, C. M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M. and Anxolabehere, D. (2005) 'Combined evidence annotation of transposable elements in genome sequences', *PLoS computational biology,* 1(2), pp. 166-175.
Roach, M. J., Schmidt, S. A. and Borneman, A. R. (2018) 'Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies', *BMC Bioinformatics,* 19(1), pp. 460.
Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G. and Mesirov, J. P. (2011) 'Integrative genomics viewer', *Nature biotechnology,* 29(1), pp. 24-26.
Santolamazza, F., Mancini, E., Simard, F., Qi, Y., Tu, Z. and Della Torre, A. (2008) 'Insertion polymorphisms of SINE200 retrotransposons within speciation islands of Anopheles gambiae molecular forms', *Malaria Journal,* 7, pp. 1-10.
Shumate, A. and Salzberg, S. L. (2021) 'Liftoff: accurate mapping of gene annotations', *Bioinformatics*, 37(12), pp. 1639–1643.
Smit, A., Hubley, R. and Green, P. (2013-2015) *RepeatMasker Open-4.0*. Available at: http://www.repeatmasker.org.
Vaser, R., Sović, I., Nagarajan, N. and Šikić, M. (2017) 'Fast and accurate de novo genome assembly from long uncorrected reads', *Genome research,* 27(5), pp. 737-746.
Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K. and Earl, A. M. (2014) 'Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement', *PLOS ONE,* 9(11), pp. e112963.