tak981

April 19, 2024

Contents

Introduction **2**

PI Summary **2**

Sample Summary **3**

Sequencing and Integration Summary Table **3**

Abundance Frequency Plots **4**

AAV ITR Breakpoint Plots **8**

Gene distribution dot plots **13**

Methods **15**

Introduction

The attached report describes the results of integration site analysis for samples from gene therapy trials. Cellular DNA was processed as described in Sherman et al. 2017 (doi: 10.1016/j.omtm.2016.11.002.) and analyzed using the analysis software AAVengeR (https://github.com/helixscript/AAVengeR).

PI Summary

This is a sample text.

Sample Summary

No Summary Given

Sequencing and Integration Summary Table

Presented below are summary details of each sample. We estimate the number of cell clones sampled using the SonicLength method (Berry, 2012); this is summarized in the column "Inferred cells". Relative abundance was not measured from read counts, which are known to be inaccurate, but from marks introduced into DNA specimens prior to PCR amplification using the SonicLength method PMID:22238265.

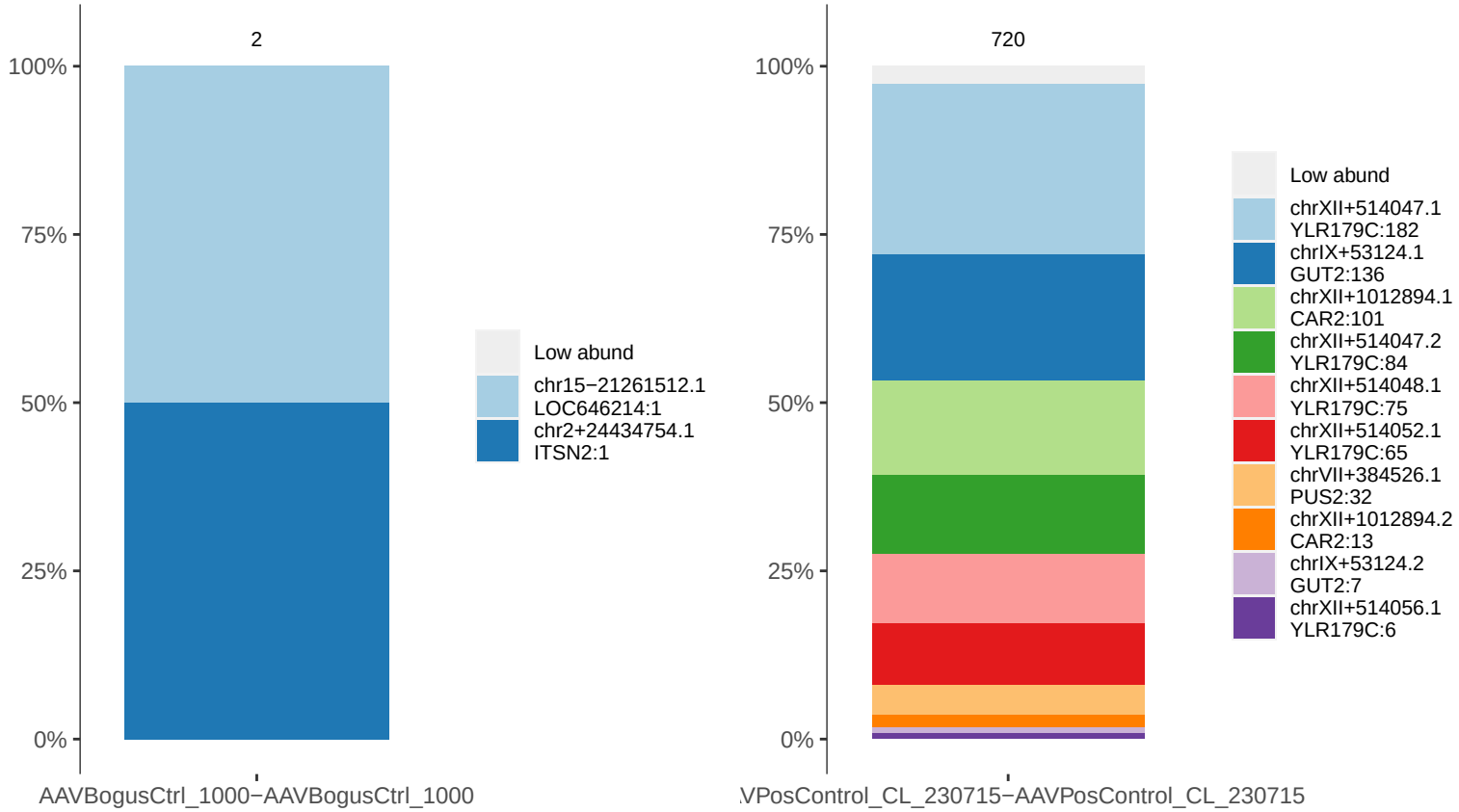
Under most circumstances, only a subset of sites will be sampled. We thus include an estimate of sample size based on the frequency of isolation information from the SonicLength method (Berry, 2012). The 'S.chao1' column denotes the estimated lower bound for population size derived using Chao estimate (Chao, 1987).

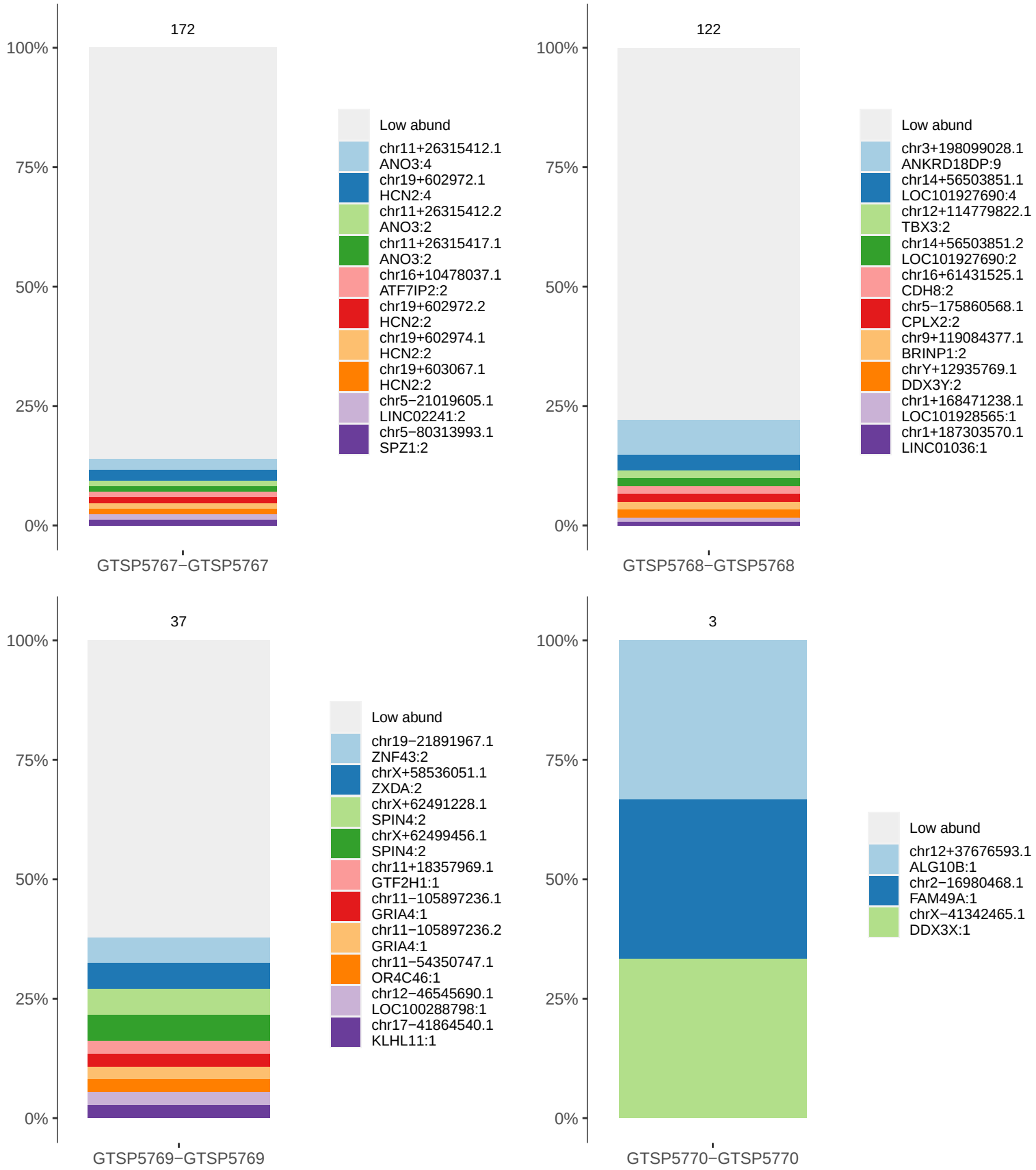
Table 1: integration summary table

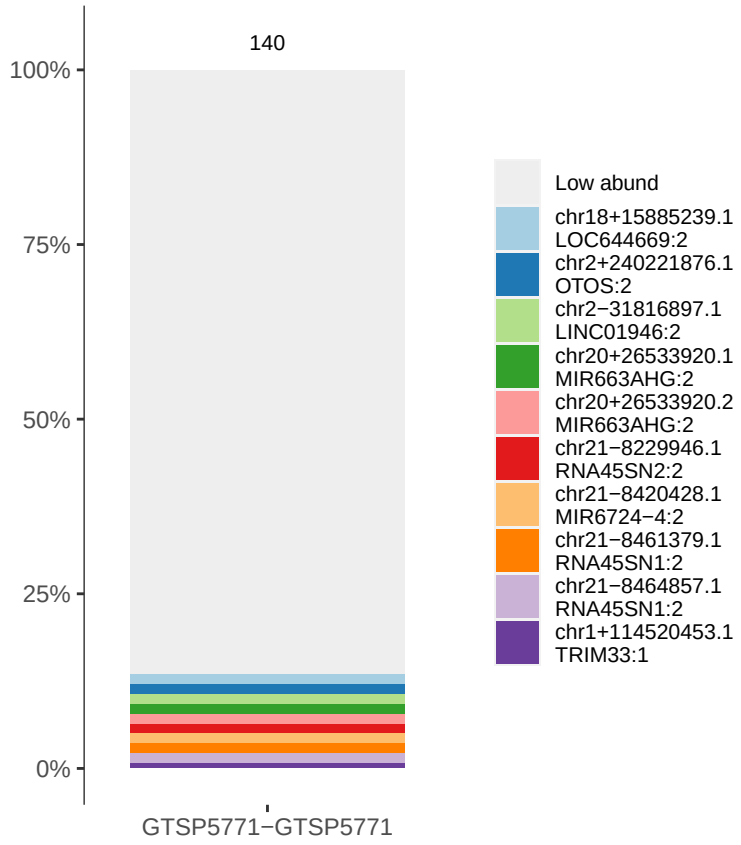
sample	patientID	total reads	Unique Sites	inferred cell	Chao1
AAVBogusCtrl_1000	AAVBogusCtrl_1000	2	2	2	3.00000
AAVPosControl_CL_230715	AAVPosControl_CL_230715	72648	19	720	22.33333
GTSP5767	GTSP5767	83174	153	172	828.21429
GTSP5768	GTSP5768	32442	105	122	770.14286
GTSP5769	GTSP5769	7855	33	37	114.20000
GTSP5770	GTSP5770	22	3	3	6.00000
GTSP5771	GTSP5771	41353	131	140	869.10000
GTSP5772	GTSP5772	2264	59	73	314.00000
GTSP5773	GTSP5773	20913	202	205	5127.25000
GTSP5774	GTSP5774	19639	68	68	2346.00000
GTSP5775	GTSP5775	16705	174	188	1374.27273
GTSP5776	GTSP5776	24948	151	156	1915.16667
GTSP5777	GTSP5777	13491	180	191	1616.50000

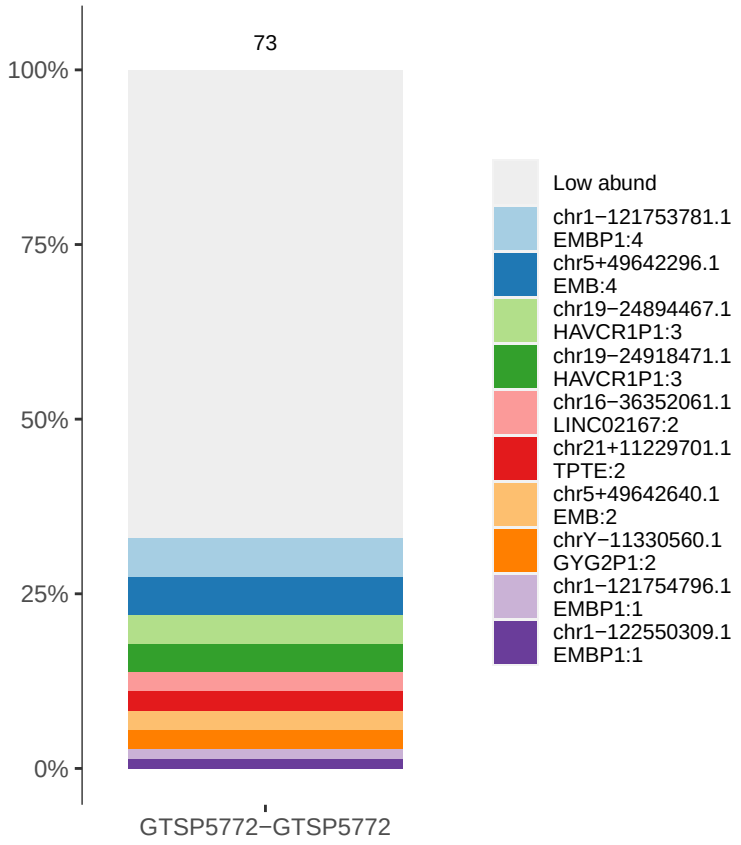
Abundance Frequency Plots

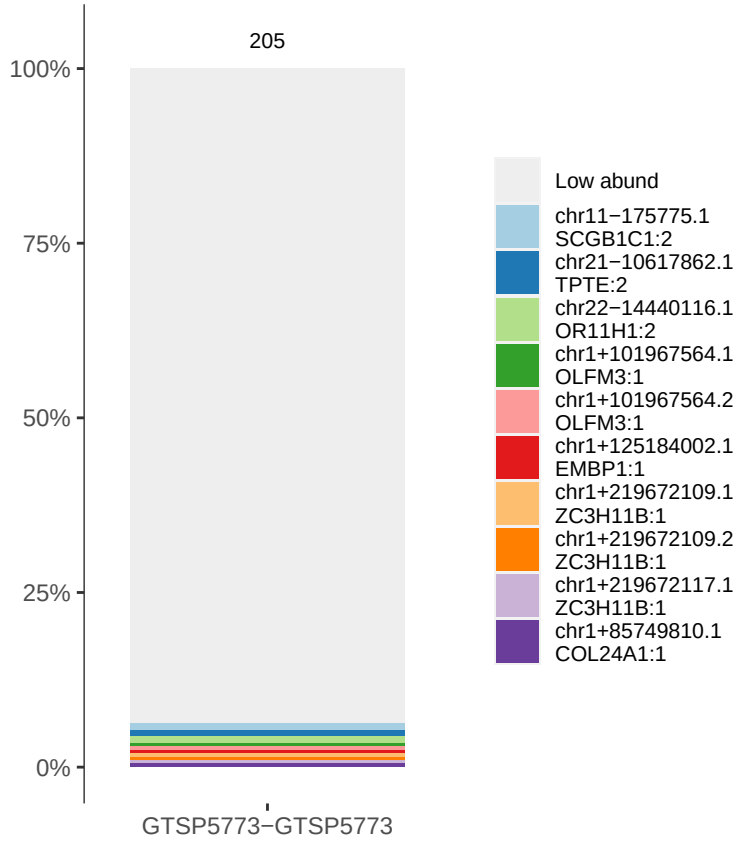
Stacked bar plots indicate clonal abundance frequency. Only the top 10 clonal types are being plotted, and the rest will be ploted in grey as "low abundance". The number above the stacked bar plots indicates the total clonal types. The legend in each plot correspond to the clonal type integration sites. The number after colon indicates the count for that specific clonal type.

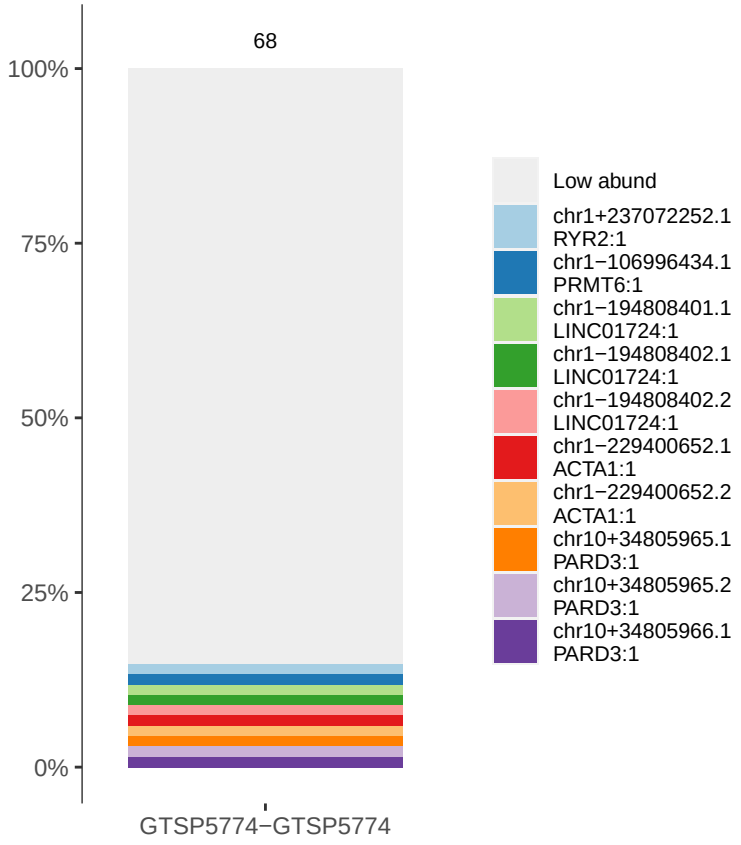


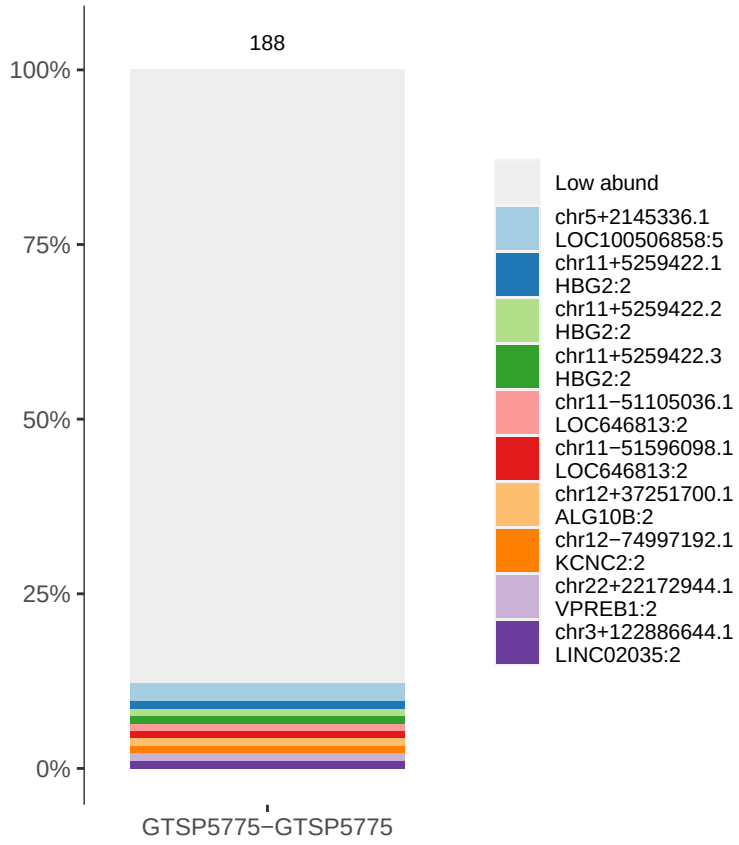

172
100%
75%
50%
25%
0%
GTSP5767−GTSP5767
Low abund
chr11+26315412.1
ANO3:4
chr19+602972.1
HCN2:4
chr11+26315412.2
ANO3:2
chr11+26315417.1
ANO3:2
chr16+10478037.1
ATF7IP2:2
chr19+602972.2
HCN2:2
chr19+602974.1
HCN2:2
chr19+603067.1
HCN2:2
chr5−21019605.1
LINC02241:2
chr5−80313993.1
SPZ1:2
122
GTSP5768−GTSP5768
Low abund
chr3+198099028.1
ANKRD18DP:9
chr14+56503851.1
LOC101927690:4
chr12+114779822.1
TBX3:2
chr14+56503851.2
LOC101927690:2
chr16+61431525.1
CDH8:2
chr5−175860568.1
CPLX2:2
chr9+119084377.1
BRINP1:2
chrY+12935769.1
DDX3Y:2
chr1+168471238.1
LOC101928565:1
chr1+187303570.1
LINC01036:1
37
GTSP5769−GTSP5769
Low abund
chr19−21891967.1
ZNF43:2
chrX+58536051.1
ZXDA:2
chrX+62491228.1
SPIN4:2
chrX+62499456.1
SPIN4:2
chr11+18357969.1
GTF2H1:1
chr11−105897236.1
GRIA4:1
chr11−105897236.2
GRIA4:1
chr11−54350747.1
OR4C46:1
chr12−46545690.1
LOC100288798:1
chr17−41864540.1
KLHL11:1
3
GTSP5770−GTSP5770
Low abund
chr12+37676593.1
ALG10B:1
chr2−16980468.1
FAM49A:1
chrX−41342465.1
DDX3X:1

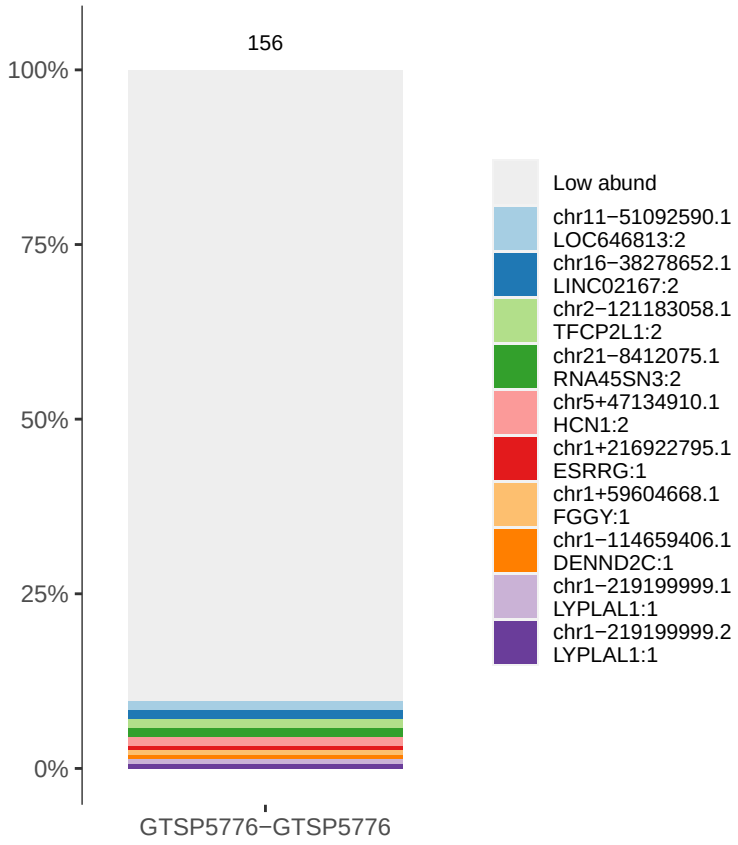

140
100%
75%
50%
25%
0%
GTSP5771−GTSP5771
Low abund
chr18+15885239.1
LOC644669:2
chr2+240221876.1
OTOS:2
chr2−31816897.1
LINC01946:2
chr20+26533920.1
MIR663AHG:2
chr20+26533920.2
MIR663AHG:2
chr21−8229946.1
RNA45SN2:2
chr21−8420428.1
MIR6724−4:2
chr21−8461379.1
RNA45SN1:2
chr21−8464857.1
RNA45SN1:2
chr1+114520453.1
TRIM33:1

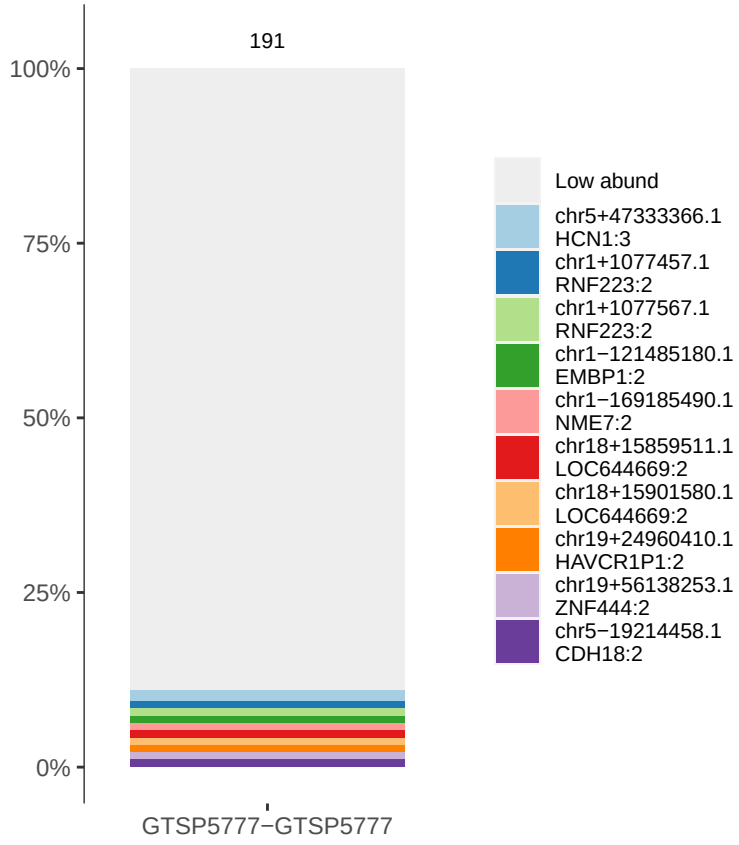

73
100%
75%
50%
25%
0%
GTSP5772−GTSP5772
Low abund
chr1−121753781.1
EMBP1:4
chr5+49642296.1
EMB:4
chr19−24894467.1
HAVCR1P1:3
chr19−24918471.1
HAVCR1P1:3
chr16−36352061.1
LINC02167:2
chr21+11229701.1
TPTE:2
chr5+49642640.1
EMB:2
chrY−11330560.1
GYG2P1:2
chr1−121754796.1
EMBP1:1
chr1−122550309.1
EMBP1:1


205
100%
75%
50%
25%
0%
GTSP5773−GTSP5773
Low abund
chr11−175775.1
SCGB1C1:2
chr21−10617862.1
TPTE:2
chr22−14440116.1
OR11H1:2
chr1+101967564.1
OLFM3:1
chr1+101967564.2
OLFM3:1
chr1+125184002.1
EMBP1:1
chr1+219672109.1
ZC3H11B:1
chr1+219672109.2
ZC3H11B:1
chr1+219672117.1
ZC3H11B:1
chr1+85749810.1
COL24A1:1


68
100%
75%
50%
25%
0%
GTSP5774−GTSP5774
Low abund
chr1+237072252.1
RYR2:1
chr1−106996434.1
PRMT6:1
chr1−194808401.1
LINC01724:1
chr1−194808402.1
LINC01724:1
chr1−194808402.2
LINC01724:1
chr1−229400652.1
ACTA1:1
chr1−229400652.2
ACTA1:1
chr10+34805965.1
PARD3:1
chr10+34805965.2
PARD3:1
chr10+34805966.1
PARD3:1

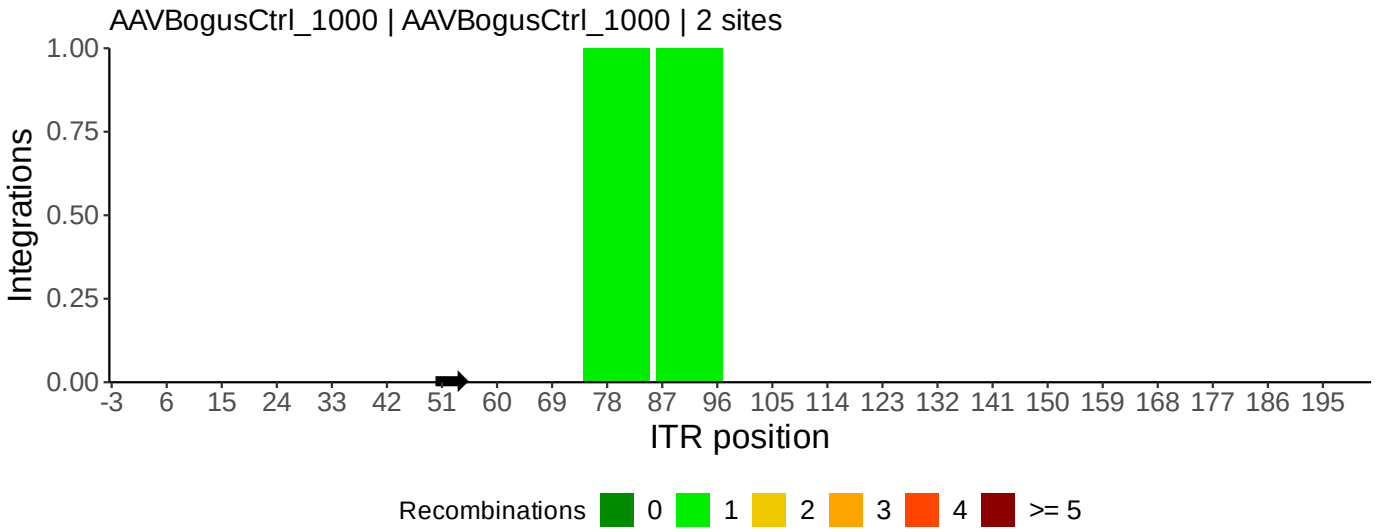

188
100%
75%
50%
25%
0%
GTSP5775−GTSP5775
Low abund
chr5+2145336.1
LOC100506858:5
chr11+5259422.1
HBG2:2
chr11+5259422.2
HBG2:2
chr11+5259422.3
HBG2:2
chr11−51105036.1
LOC646813:2
chr11−51596098.1
LOC646813:2
chr12+37251700.1
ALG10B:2
chr12−74997192.1
KCNC2:2
chr22+22172944.1
VPREB1:2
chr3+122886644.1
LINC02035:2

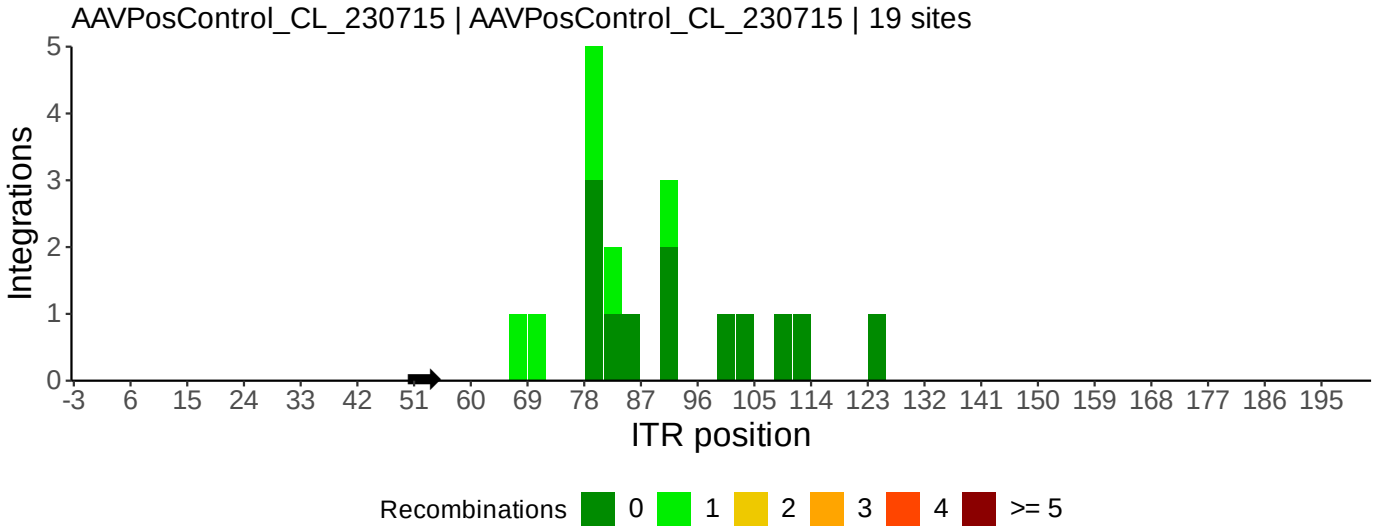

156
100%
75%
50%
25%
0%
GTSP5776−GTSP5776
Low abund
chr11−51092590.1
LOC646813:2
chr16−38278652.1
LINC02167:2
chr2−121183058.1
TFCP2L1:2
chr21−8412075.1
RNA45SN3:2
chr5+47134910.1
HCN1:2
chr1+216922795.1
ESRRG:1
chr1+59604668.1
FGGY:1
chr1−114659406.1
DENND2C:1
chr1−219199999.1
LYPLAL1:1
chr1−219199999.2
LYPLAL1:1

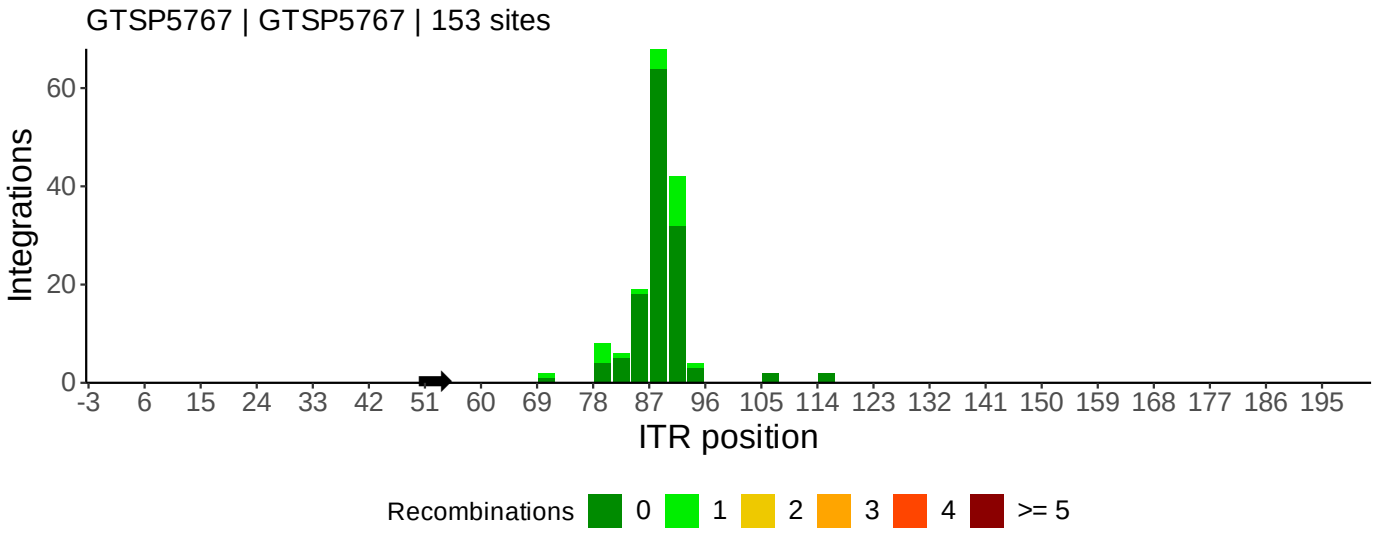

191
100%
75%
50%
25%
0%
GTSP5777−GTSP5777
Low abund
chr5+47333366.1
HCN1:3
chr1+1077457.1
RNF223:2
chr1+1077567.1
RNF223:2
chr1−121485180.1
EMBP1:2
chr1−169185490.1
NME7:2
chr18+15859511.1
LOC644669:2
chr18+15901580.1
LOC644669:2
chr19+24960410.1
HAVCR1P1:2
chr19+56138253.1
ZNF444:2
chr5−19214458.1
CDH18:2

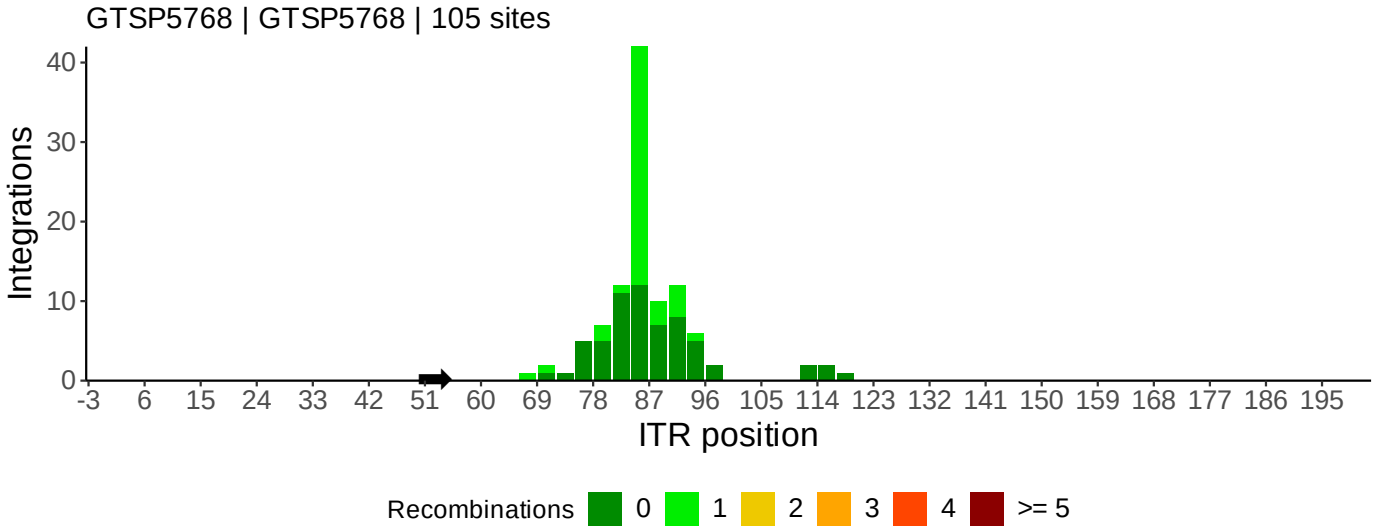
AAV ITR Breakpoint Plots

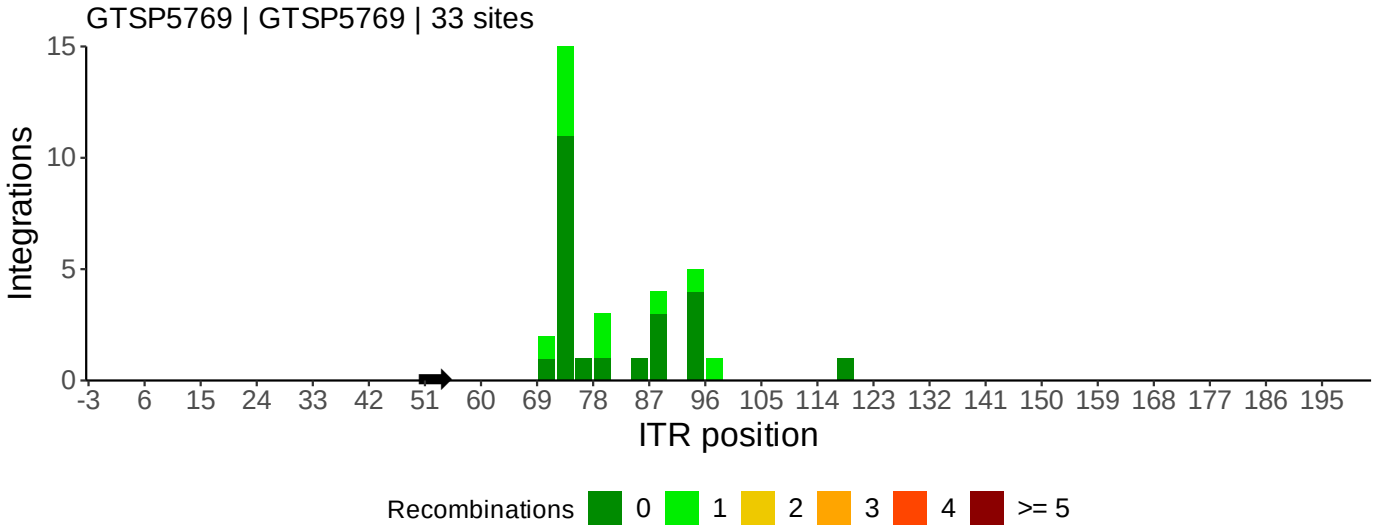
These plots are a visual representation of the ITR breakpoint at the site of genomic integration. The arrow indicates the priming location within the ITR and the x-coordinate of each bar corresponds to a departure from an expected ITR sequence. Each departure from expectation is demarcated as a rearrangement, and bars are colored to reflect the frequency of rearrangement at each integration site.

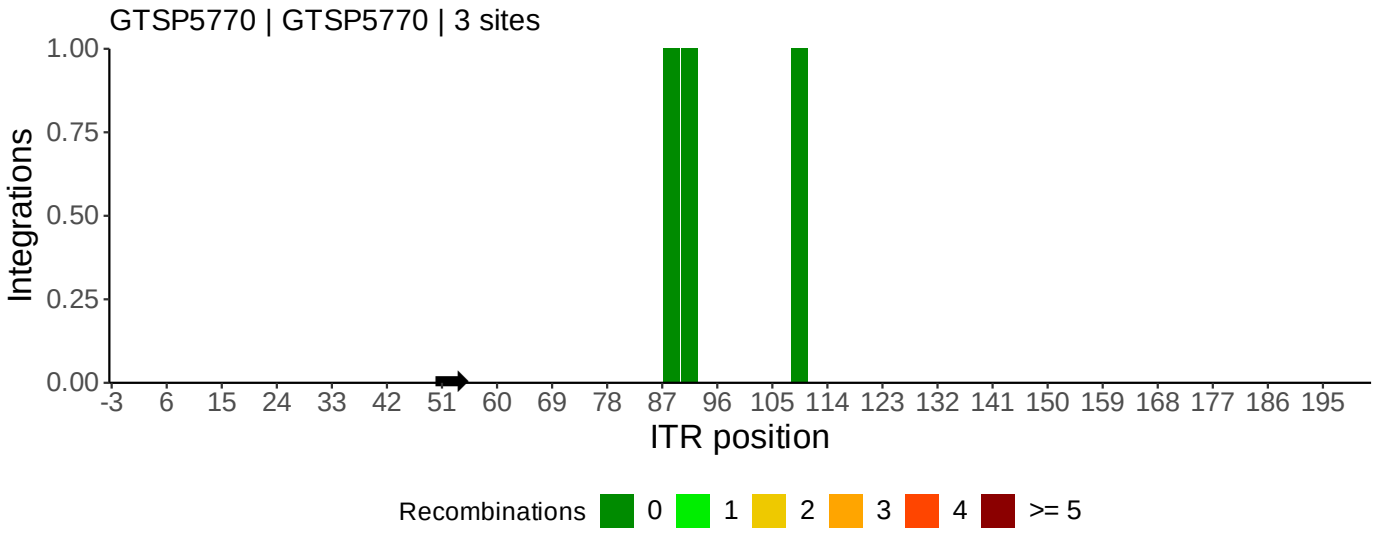


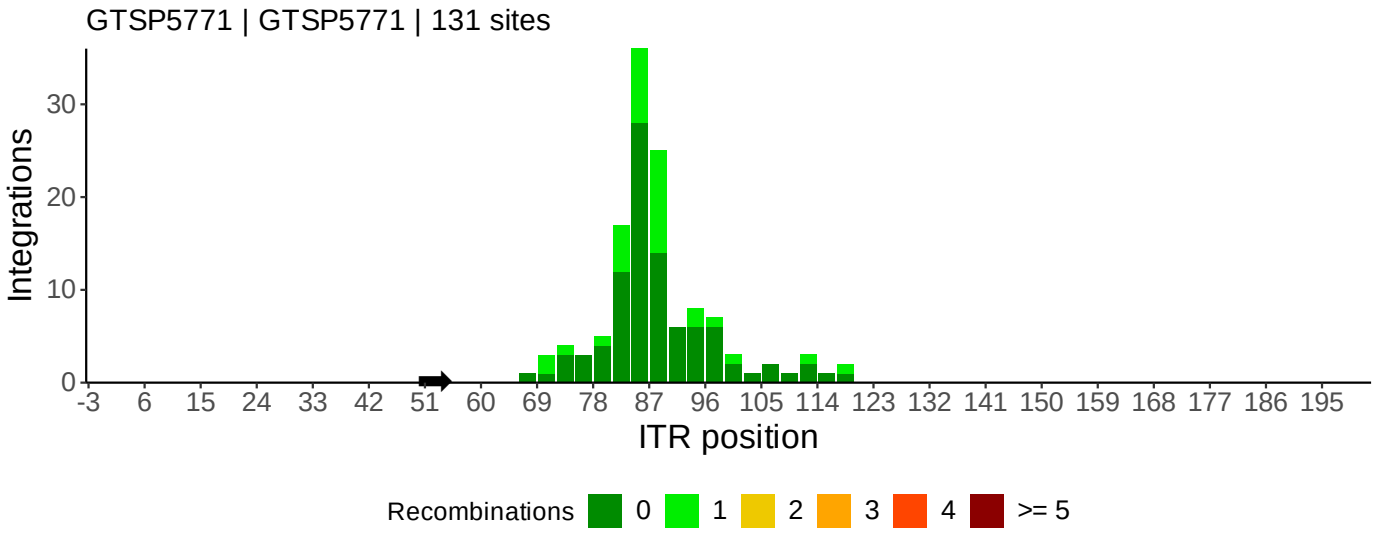


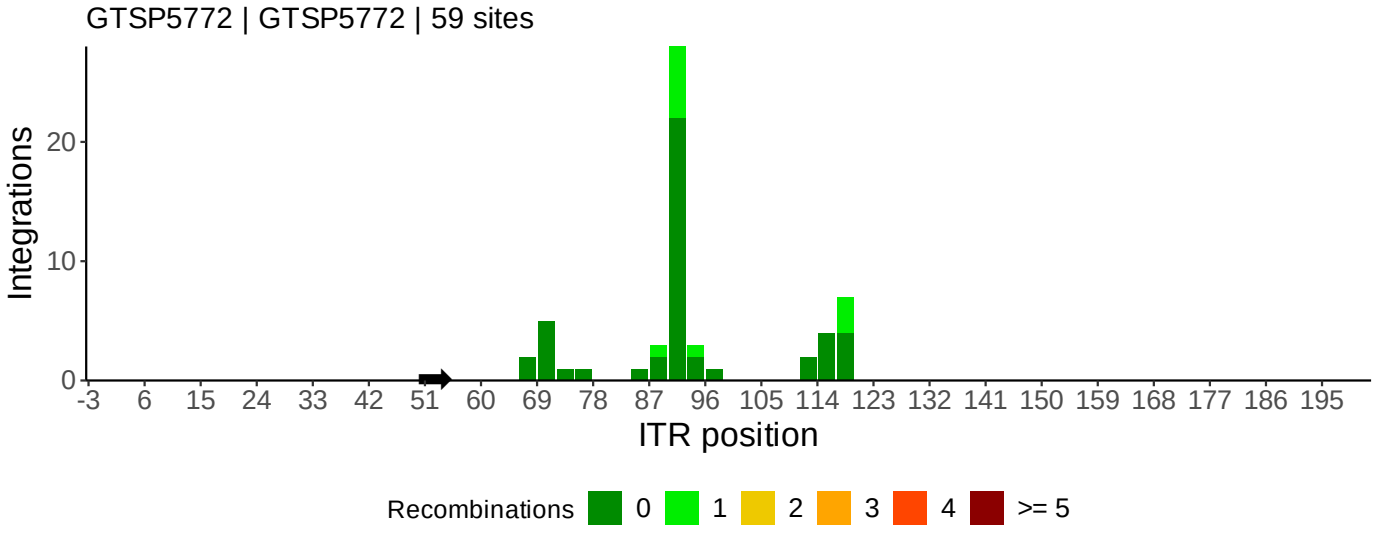

GTSP5767 | GTSP5767 | 153 sites
Integrations
0
20
40
60
-3 6 15 24 33 42 51 60 69 78 87 96 105 114 123 132 141 150 159 168 177 186 195
ITR position
Recombinations 0 1 2 3 4 >= 5

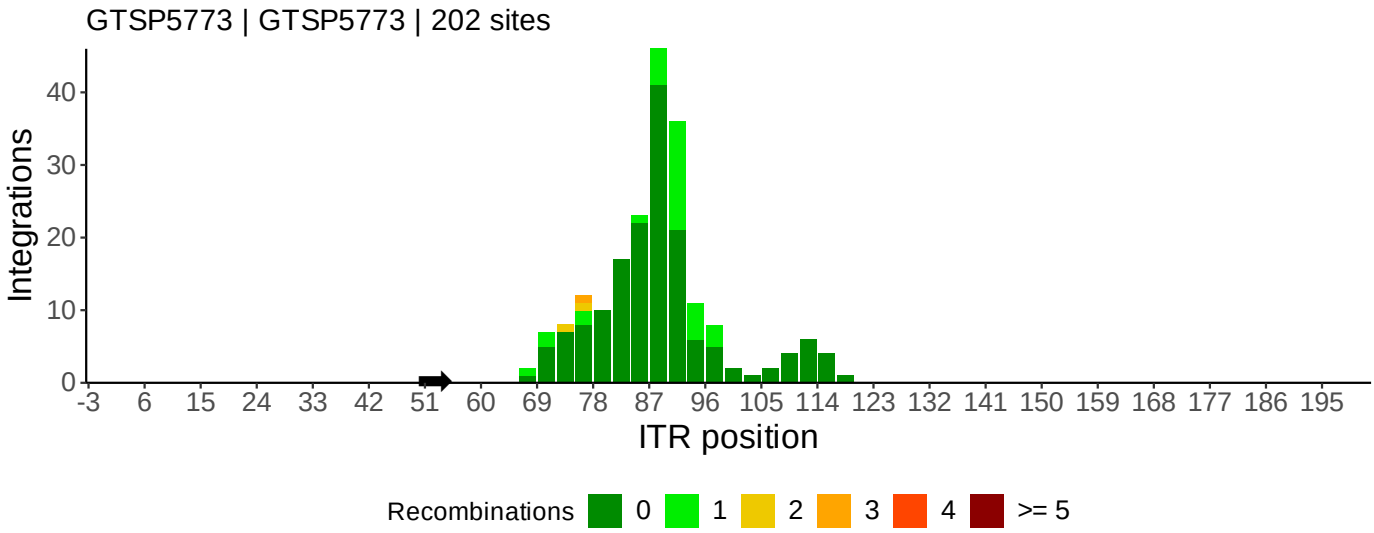

GTSP5768 | GTSP5768 | 105 sites
Integrations
0
10
20
30
40
-3 6 15 24 33 42 51 60 69 78 87 96 105 114 123 132 141 150 159 168 177 186 195
ITR position
Recombinations 0 1 2 3 4 >= 5

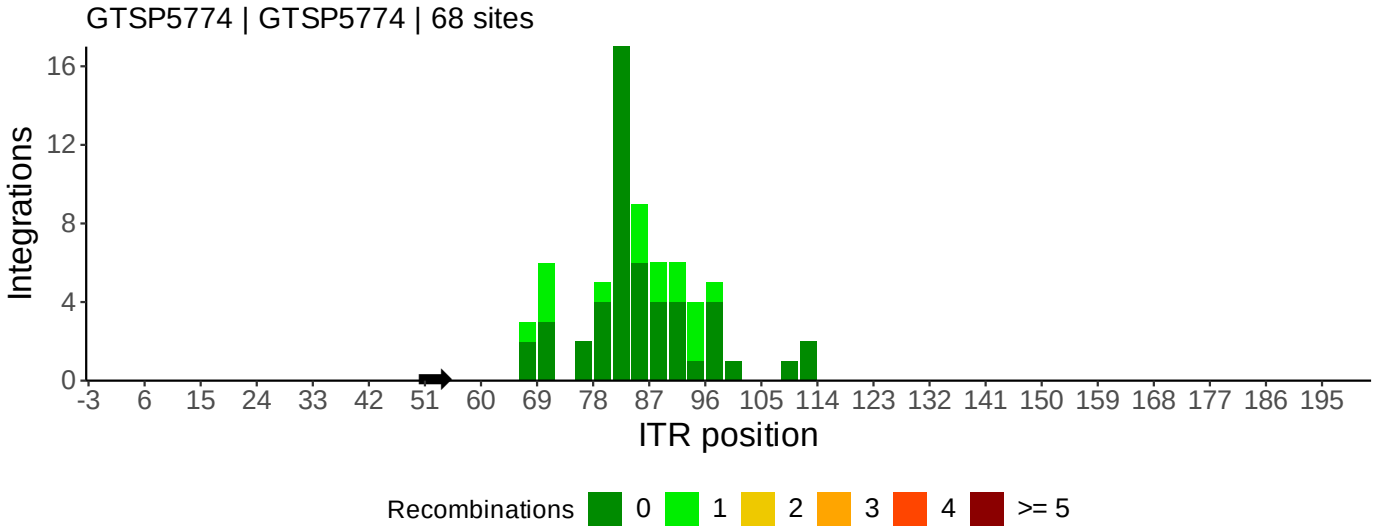

GTSP5769 | GTSP5769 | 33 sites
Integrations
0
5
10
15
-3 6 15 24 33 42 51 60 69 78 87 96 105 114 123 132 141 150 159 168 177 186 195
ITR position
Recombinations 0 1 2 3 4 >= 5

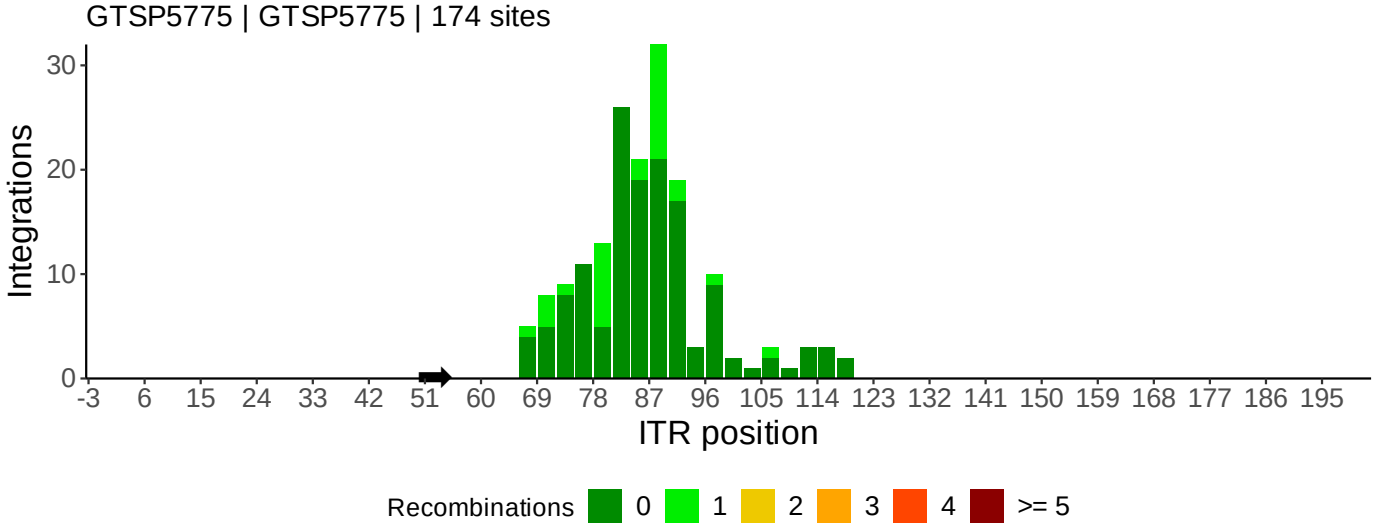

GTSP5770 | GTSP5770 | 3 sites
Integrations
1.00
0.75
0.50
0.25
0.00
-3 6 15 24 33 42 51 60 69 78 87 96 105 114 123 132 141 150 159 168 177 186 195
ITR position
Recombinations 0 1 2 3 4 >= 5

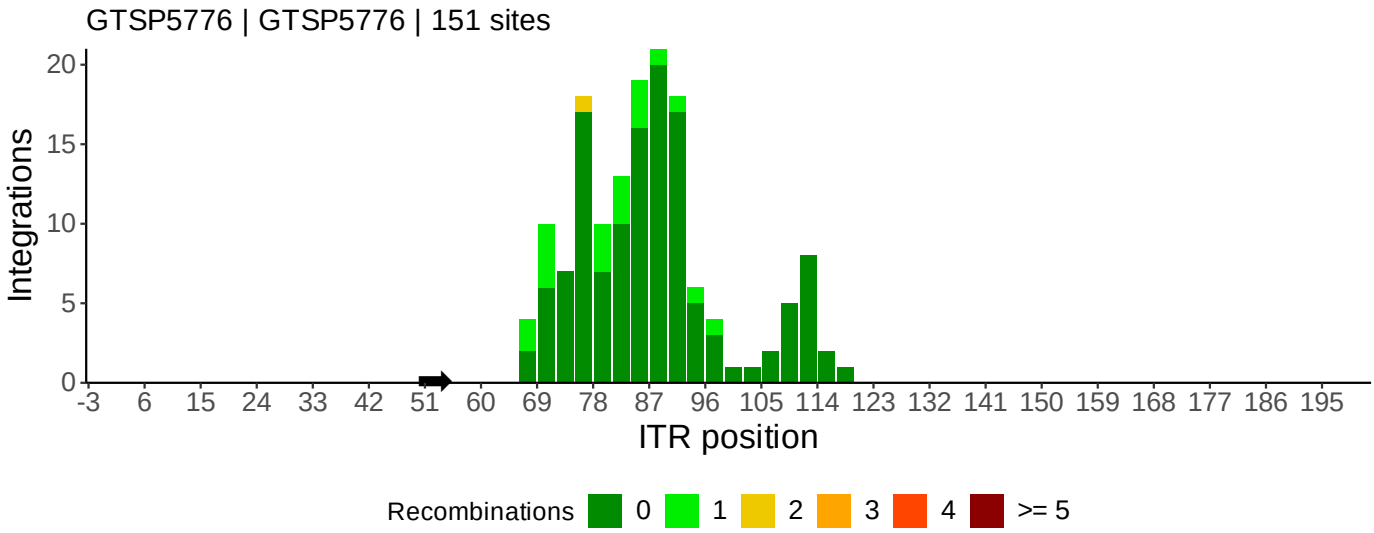

GTSP5771 | GTSP5771 | 131 sites
Integrations
30
20
10
0
-3 6 15 24 33 42 51 60 69 78 87 96 105 114 123 132 141 150 159 168 177 186 195
ITR position
Recombinations 0 1 2 3 4 >= 5

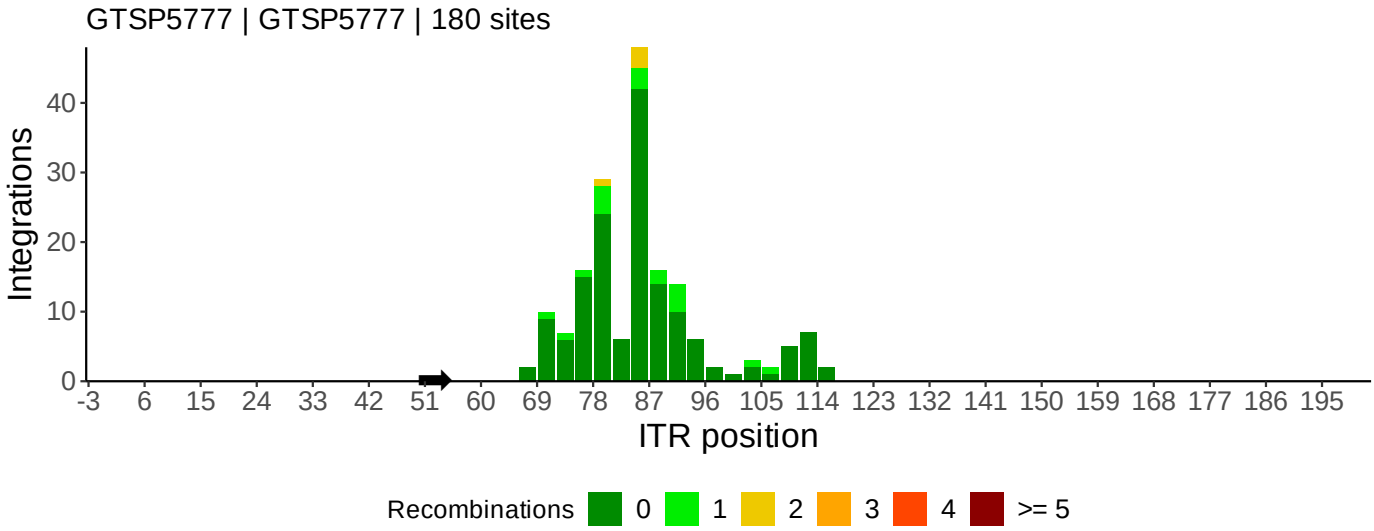

GTSP5772 | GTSP5772 | 59 sites
Integrations
20
10
0
-3 6 15 24 33 42 51 60 69 78 87 96 105 114 123 132 141 150 159 168 177 186 195
ITR position
Recombinations 0 1 2 3 4 >= 5


GTSP5773 | GTSP5773 | 202 sites
Integrations
0
10
20
30
40
-3 6 15 24 33 42 51 60 69 78 87 96 105 114 123 132 141 150 159 168 177 186 195
ITR position
Recombinations 0 1 2 3 4 >= 5


GTSP5774 | GTSP5774 | 68 sites
Integrations
0
4
8
12
16
-3 6 15 24 33 42 51 60 69 78 87 96 105 114 123 132 141 150 159 168 177 186 195
ITR position
Recombinations 0 1 2 3 4 >= 5


GTSP5775 | GTSP5775 | 174 sites
Integrations
0
10
20
30
-3 6 15 24 33 42 51 60 69 78 87 96 105 114 123 132 141 150 159 168 177 186 195
ITR position
Recombinations 0 1 2 3 4 >= 5


GTSP5776 | GTSP5776 | 151 sites
Integrations
0
5
10
15
20
-3 6 15 24 33 42 51 60 69 78 87 96 105 114 123 132 141 150 159 168 177 186 195
ITR position
Recombinations
0
1
2
3
4
>= 5


GTSP5777 | GTSP5777 | 180 sites
Integrations
0
10
20
30
40
-3 6 15 24 33 42 51 60 69 78 87 96 105 114 123 132 141 150 159 168 177 186 195
ITR position
Recombinations
0
1
2
3
4
>= 5

Gene distribution dot plots

The plot below details the % of integration sites within two features, 'in genes' and 'in Exons', summarized per sample. 'in Genes' refers to sites within transcription unit boundaries and 'in Exons' refers to sites within exon boundaries.

Table 2: gene distribution percentage table

sample	In Exon%	In Transcription Unit%
AAVBogusCtrl_1000	0.0	0.0
AAVPosControl_CL_230715	57.9	57.9
GTSP5767	0.7	36.6
GTSP5768	0.0	39.0
GTSP5769	3.0	15.2
GTSP5770	0.0	33.3
GTSP5771	10.7	29.8
GTSP5772	1.7	11.9
GTSP5773	3.5	32.2
GTSP5774	5.9	52.9
GTSP5775	2.9	35.6
GTSP5776	2.0	47.0
GTSP5777	0.6	26.1

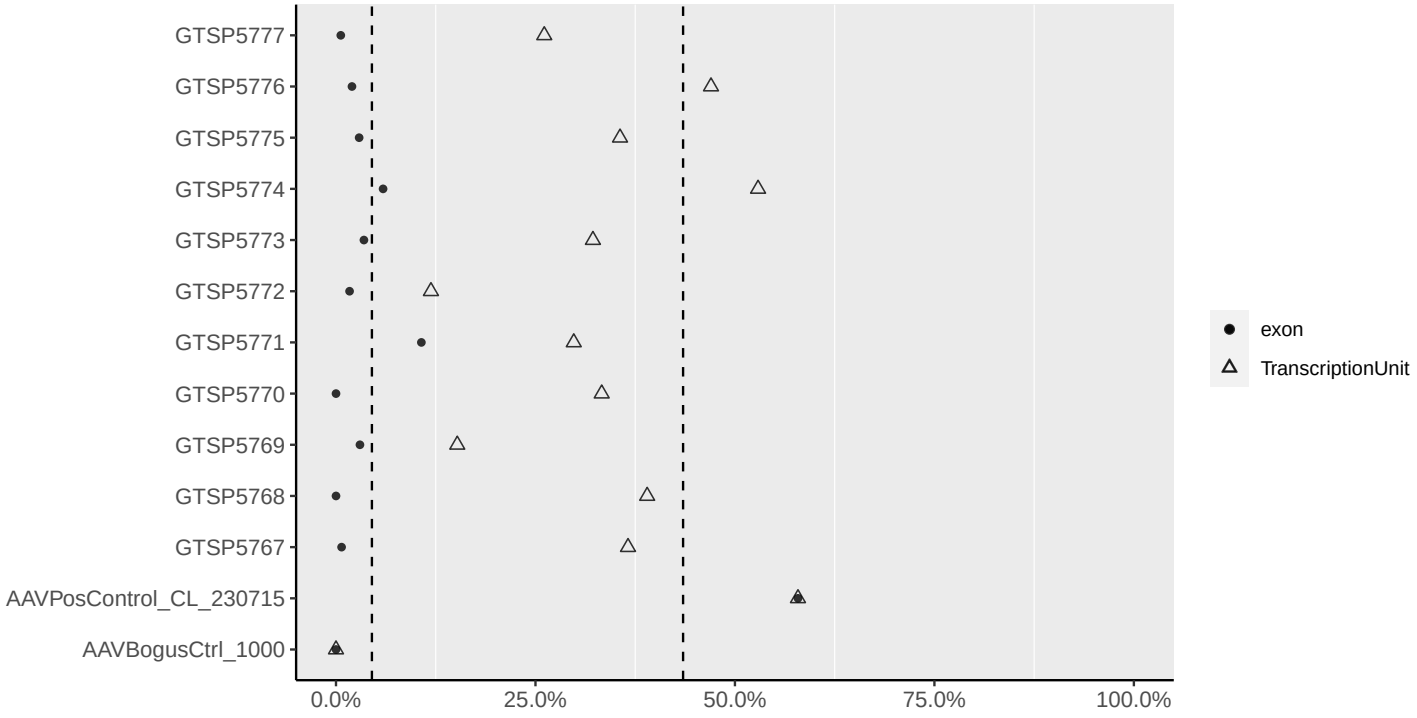


Figure 1: gene distribution scatter plot

Methods

Report Generated on: April 19, 2024
AAVenger Version: 1.1

Modules Called:

- core.R
- mapSiteLeaderSequences.R
- buildAAVremnantPlots.R
- anchorReadRearrangements.R