

# Analysis of integration site distributions and clonal abundances for gene therapy correction of cystinosis (Groups 1-21)

*John K. Everett, Ph.D. and Frederic Bushman, Ph.D.*

*June 2019*

## Contents

Summary of results	2
Human and mouse samples studied	2
Subject reports	2
UCSC browser exploration	2
Description of analysis techniques	3
Comparisons to previous trials	4
Integration events near oncogenes in human subjects . . . . .	4
Integration events near oncogenes in mouse subjects . . . . .	5
Mapping of integration site positions	6
Relative abundances of human subject samples	8
Expanded clones	10
Mouse transplant trials	11
References	26
Supplementary tables and figures	27
Numbers of inferred cells and integration sites identified in provided samples . . . . .	27
Analyzed samples in which no integration sites were identified . . . . .	29
Comparison of the number of integration sites and inferred cells . . . . .	30
Integration positions of the lentiviral therapeutic vector from a previous WAS correction trial <sup>1</sup> . .	31
Persistence of clones in mouse BM transplant trials . . . . .	32
Sequencing depth . . . . .	34

# Summary of results

The goal of this analysis is to investigate the integration profile of a gene therapy vector for the correction of cystinosis in both mouse and human subjects and assess potential clonal expansions. The list of human oncogenes used throughout this analysis was compiled from the literature ([link](#)) and the list of mouse oncogenes was compiled from the retroviral tagged cancer gene database (RTCGD)<sup>1</sup> using an inclusion threshold of three or more incidents. The human oncogene list comprises 7.46% of all human genes (NCBI RefSeq) whereas the mouse oncogene list comprises 2.52% of all mouse genes. The frequency of integration sites in human subjects near oncogenes was not significantly different than a published trial using a comparable vector to correct Wiskott-Aldrich syndrome (WAS) from which no adverse events have been reported<sup>2</sup>. The frequency of integration sites in bone marrow donor mice near oncogenes was generally less than that of mice in a previously published  $\beta$ -thalassemia mouse trial from which no adverse events have been reported<sup>3</sup>. The integration position profile for human subjects was very similar to that found in the WAS trial. Mouse bone marrow transplant trials yielded varying degrees of persistence in five of the nine trials where detection was limited due to both sequencing depth and vector copy number. No significant enrichment of integration events near oncogenes was identified between donor and recipient mice. The code base for this analysis is available on-line ([link](#)).

## Human and mouse samples studied

Integration sites were detected in 68 samples from both human and mouse subjects (Tables 1 & S1) while no integration sites were detected in 10 of the provided samples (Table S2).

Table 1. Overview of data collection.

Organism	Number of samples	Number of reads	Number of inferred cells	Number of integration sites
human	30	19,834,249	133,607	92,924
mouse	38	19,108,607	53,330	4,679

## Subject reports

Subject specific reports for all subjects are available via an online archive ([link](#)).

## UCSC browser exploration

UCSC browser sessions pre-loaded with the integration sites identified in this analysis are available via these links: ([human subjects](#)), ([mouse subjects](#)). Integration sites are shown as blue (positive orientation integration) and red (reverse orientation integration) tick marks. For each integration site, the intensity of the color reflects the maximum abundance observed.

# Description of analysis techniques

We investigate effects of integration on cell growth using the following criteria: Integration Frequency is the frequency at which unique integration sites are observed in or near a given gene. Clonal Abundance is determined by quantifying the number of sites of linker ligation associated with each unique integration site. This samples the number of DNA chains at the start of the experiment allowing clonal expansion to be quantified<sup>4</sup>.

Relative clonal Abundance is determined per sample and is the percentage of identified cells attributed to a given clone. Integration sites and the clones harboring them are sampled from a larger population. It would be rare for all integration sites in a sample to be represented in the sequence data.

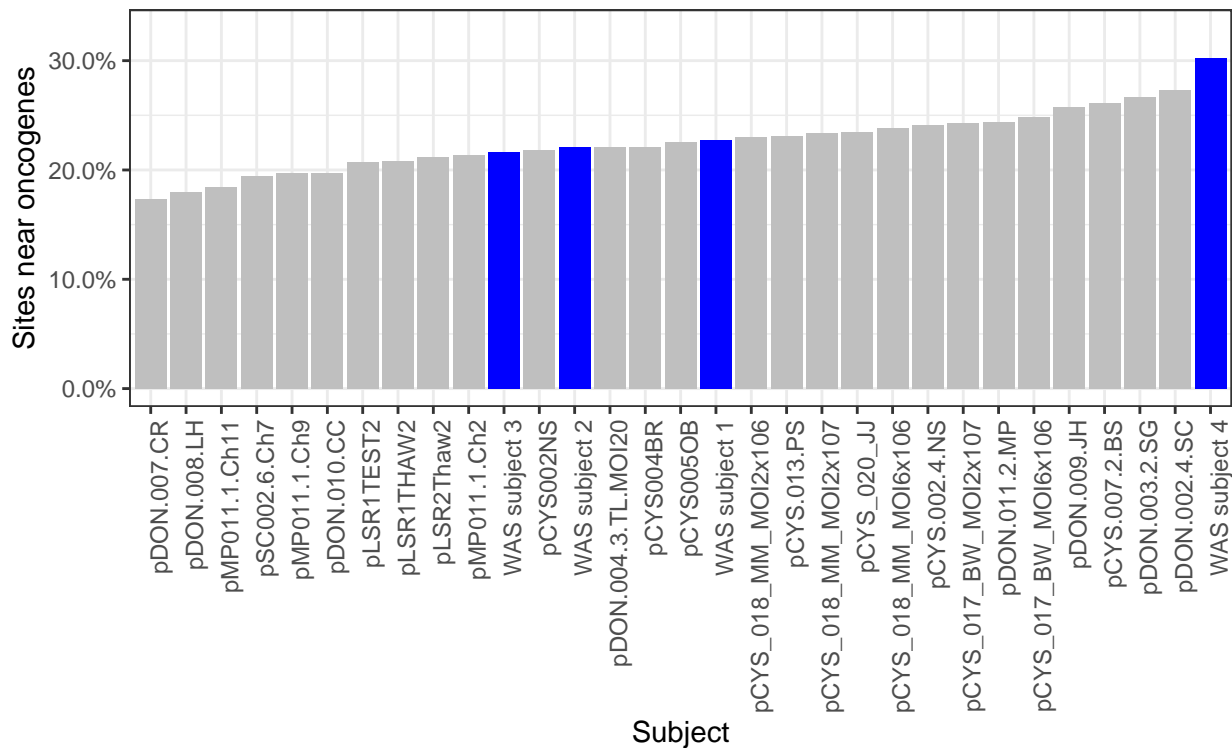
For this analysis, four technical replicates of each delivered sample were prepared, sequenced and analyzed with the INSPIRED integration site analysis pipeline (v1.2)<sup>5</sup>.

# Comparisons to previous trials

## Integration events near oncogenes in human subjects

In order to determine if the experimental vector has a higher propensity of integrating near suspected oncogene in humans than previously employed vectors, the frequency of integration near oncogenes was compared to a previously published human trial<sup>1</sup> which used a comparable lentiviral vector to correct Wiskott-Aldrich syndrome (WAS) where no adverse events have been reported to date. The frequency of integration near oncogenes in four WAS pre-transplant patient samples was compared to the frequency of integration in the experimental day 14 samples (Figure 1 [CYS: gray, WAS: blue]). The frequencies of integration near oncogenes of samples in this study and the previous WAS study were not significantly different (Mann-Whitney U-test p-value: 0.67).

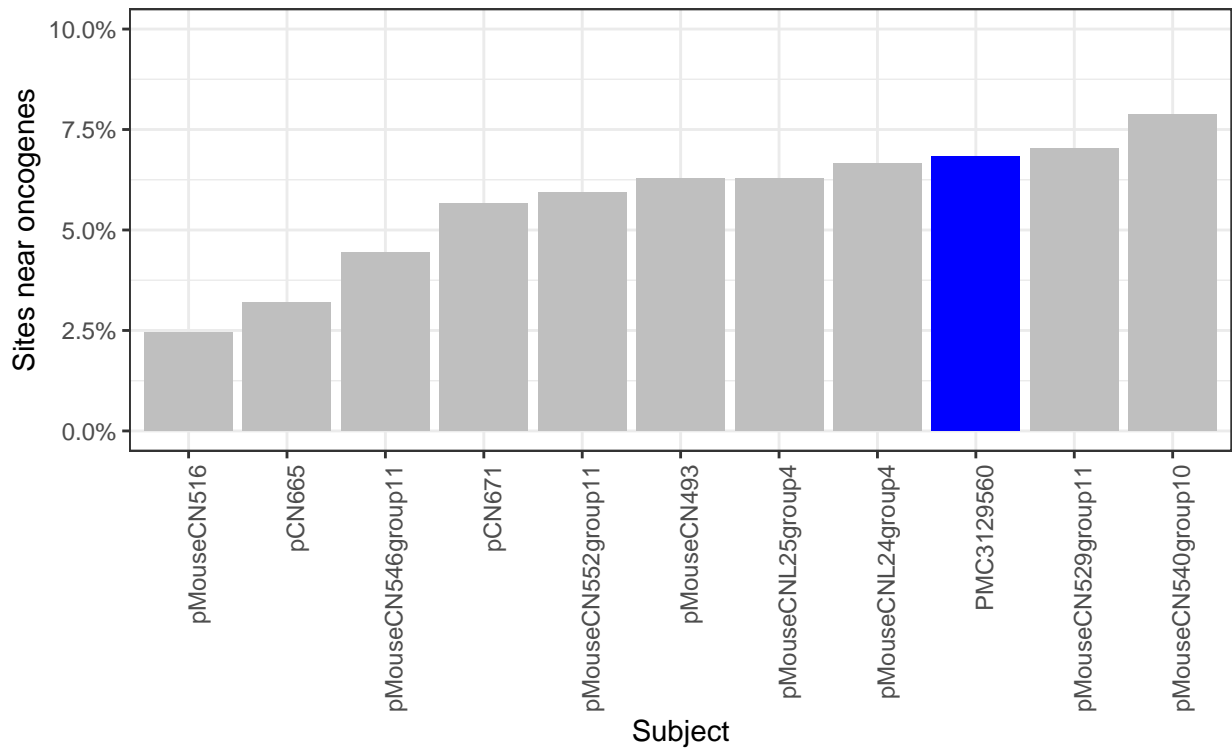
Figure 1. Comparison of frequencies of integration events near oncogenes.



# Integration events near oncogenes in mouse subjects

In order to determine if the experimental vector has a higher propensity of integrating near suspected oncogene in mice than previously employed vectors, the frequency of integration near oncogenes was compared to a previously published mouse trial<sup>3</sup> which used a comparable lentiviral vector to correct  $\beta$ -thalassemia. The frequency of integration events near onco genes in *bone marrow donor* mice was generally less than the mean frequency of integration events near oncogenes in the published trial (Figure 2 [CYS: gray,  $\beta$ -thalassemia trial: blue]).

Figure 2. Comparison of frequencies of integration events near oncogenes.



# Mapping of integration site positions

Heat maps of identified integration sites are shown below in Figure 3a (human subjects) & 3b (mouse subjects). The integration position profile from human subjects is very similar to the profile from the previous WAS study (Figure S2).

Figure 3a

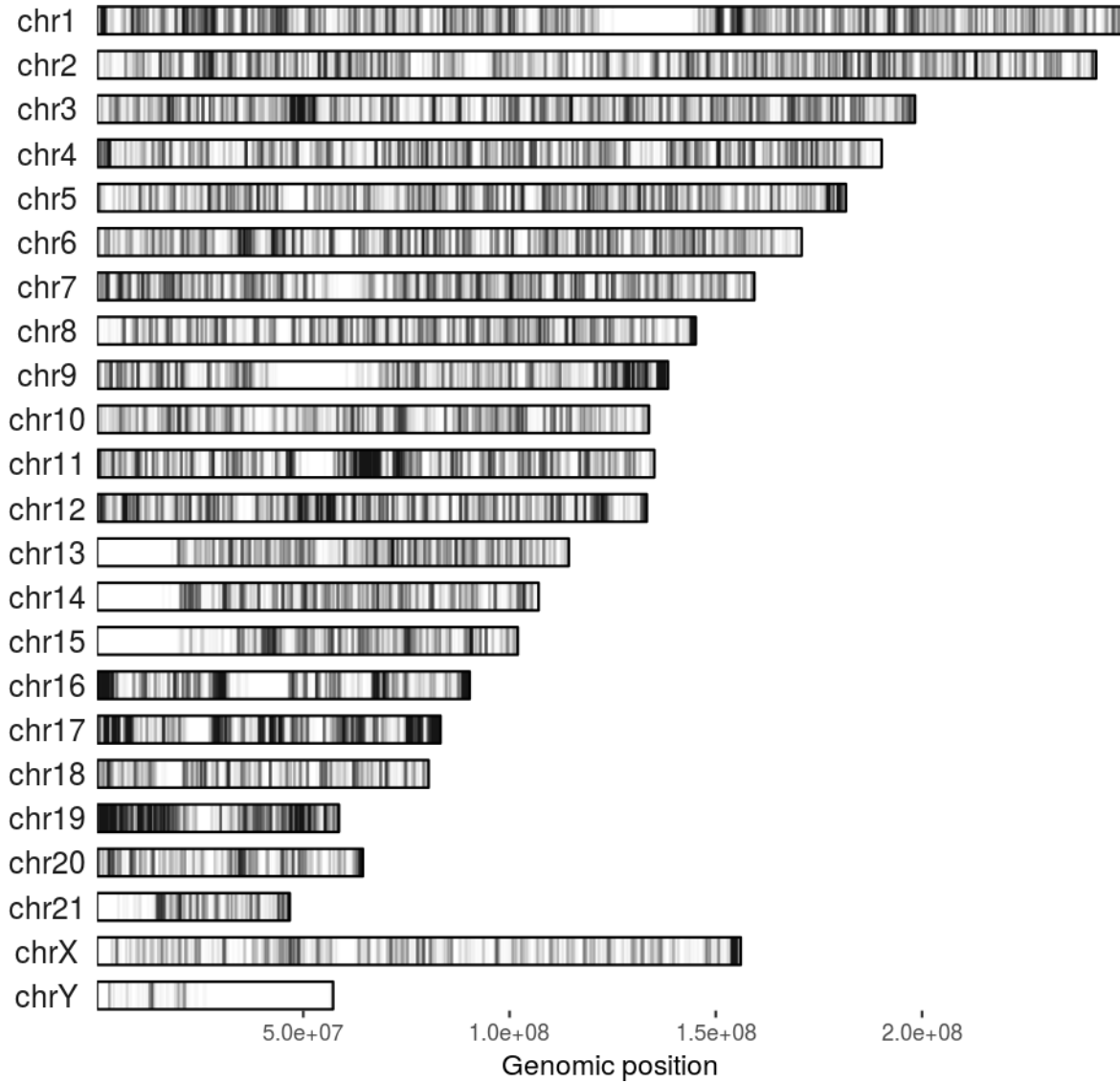
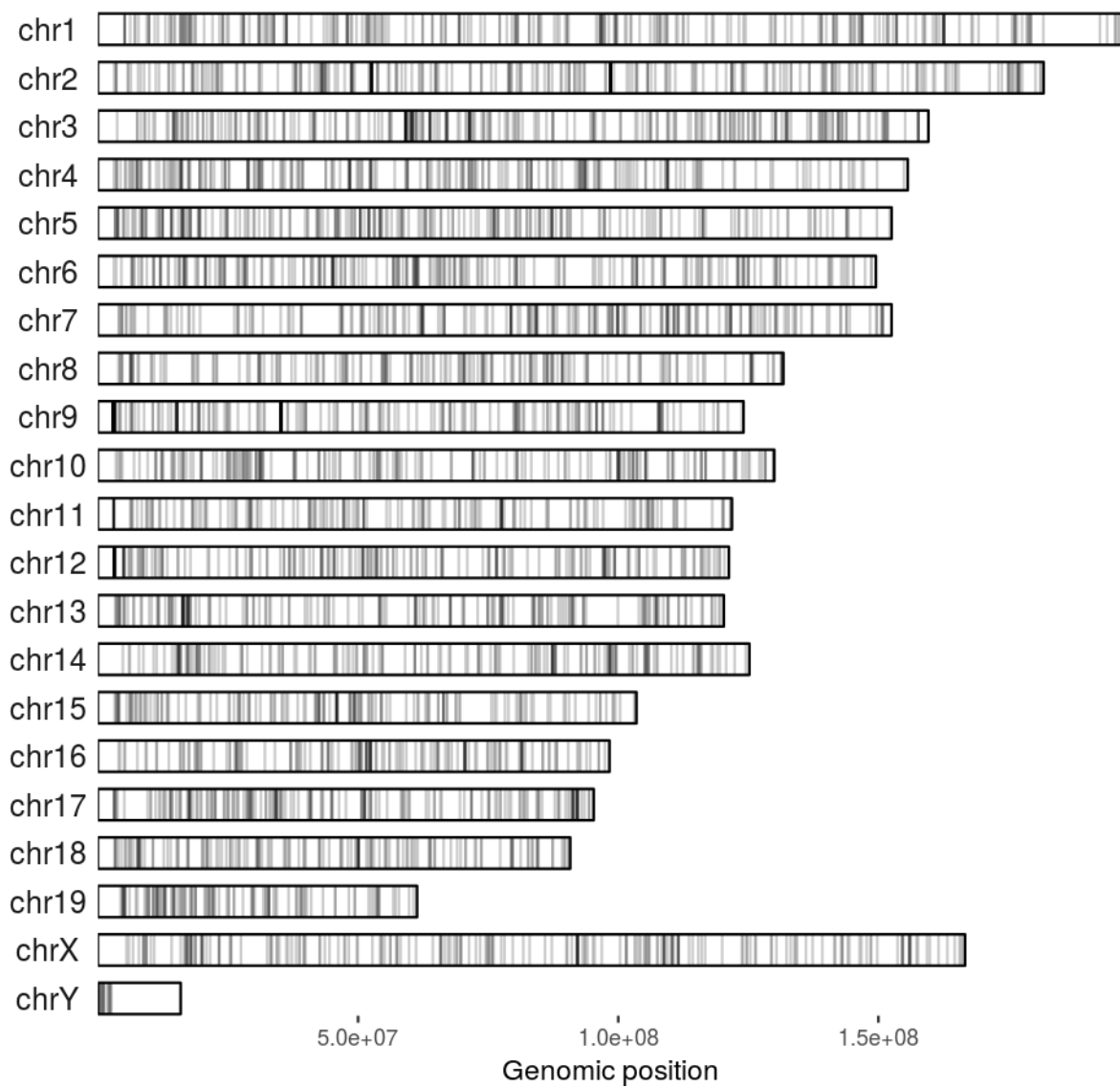


Figure 3b.



# Relative abundances of human subject samples

The sample relative abundance plots below (Figure 4) show the most abundant 25 clones in each human sample as colored bars while less abundant clones were relegated to a single low abundance bar shown in gray. Subject DON.002.4.SC showed an expanded clone (27% relative abundance) with an integration event down stream of the non-coding RNA gene LOC105374704. Subject pCYS.013.PS (PBCD34-MOI40) showed an expanded clone (27% relative abundance) with an integration event within intron 10 of ARFGEF1. These two expanded clones are not of immediate concern given that such expanded clones appeared in the WAS trial used for comparison (Figure S1) and both samples contained relatively few inferred cells which inflate relative abundance values.

Figure 4.

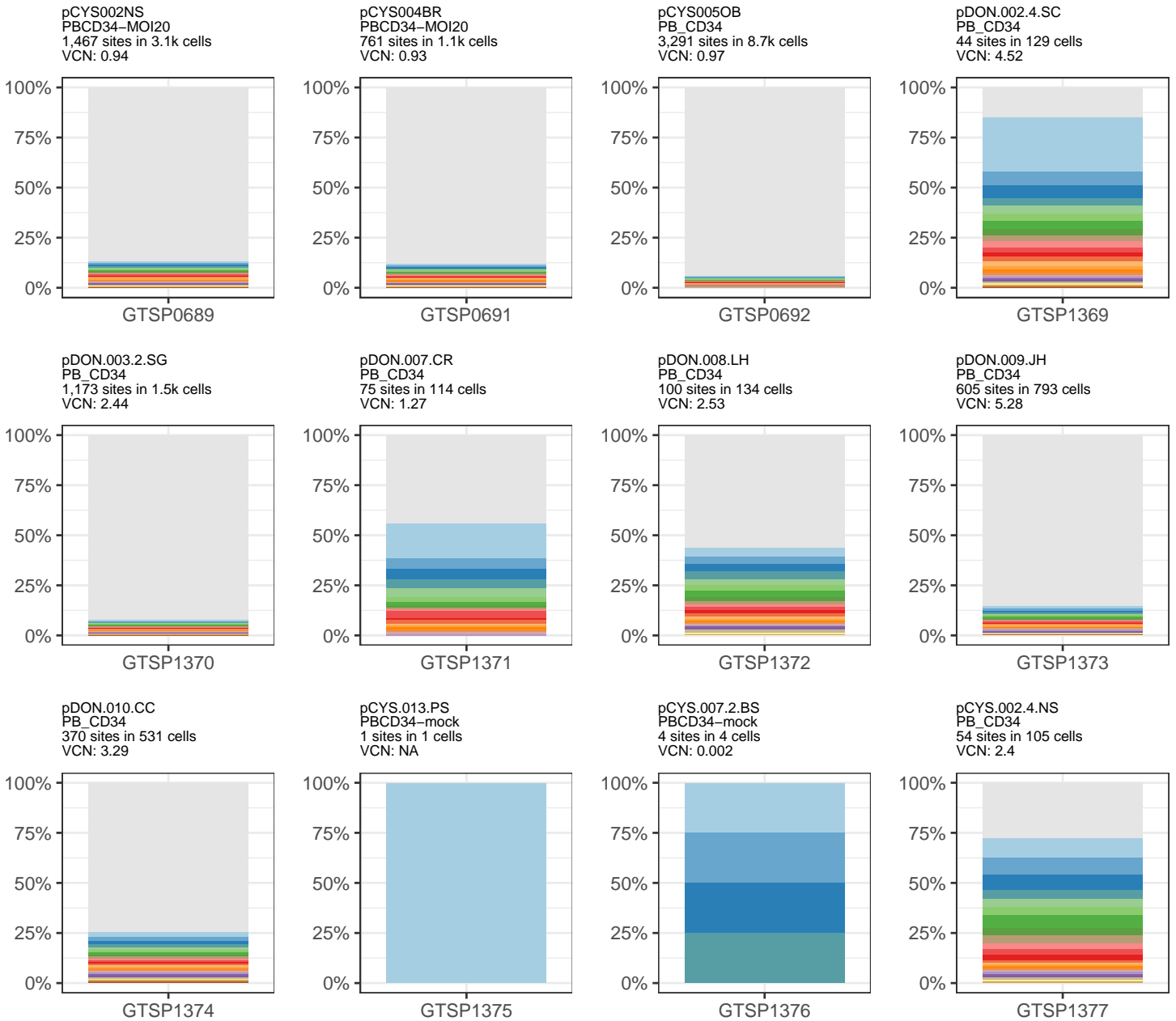




Figure 4. (continued)



# Expanded clones

Table 2 below lists clones with relative clonal abundances  $\geq 20\%$ . The estimated number of cells harboring each integration (Abundance) is shown for context.

Table 2.

Subject	Organism	Time point	Cell type	Position	Relative abundance	Abundance	Nearest Gene
pDON.002.4.SC	human	D14	PB_CD34	chr5+30409434	27.13%	35	LOC105374704
pCYS.013.PS	human	D14	PBCD34-mock	chr12+82136325	100.00%	1	LOC101928449
pCYS.007.2.BS	human	D14	PBCD34-mock	chr1-167962238	25.00%	1	DCAF6
pCYS.007.2.BS	human	D14	PBCD34-mock	chr1-218477558	25.00%	1	TGFB2
pCYS.007.2.BS	human	D14	PBCD34-mock	chr6+11842138	25.00%	1	ADTRP
pCYS.007.2.BS	human	D14	PBCD34-mock	chr8+133682373	25.00%	1	LOC105375773
pCYS.013.PS	human	D14	PBCD34-MOI40	chr8+67244348	26.53%	13	ARFGEF1
pMouseCNL23group4control	mouse	M6	Mouse_BM	chr3-59165408	33.33%	2	Igsf10
pMouseCNL24group4	mouse	M6	Mouse_BM	chr18-79750998	27.25%	785	Setbp1
pMouseCNL24group4	mouse	M6	Mouse_BM	chr2+141127890	32.45%	935	MacroD2
pMouseCNL25group4	mouse	M6	Mouse_BM	chr2+52586806	21.70%	1057	Stam2
pMouseCNL25group4	mouse	M6	Mouse_BM	chr3-59165408	22.22%	1082	Igsf10
pMouseCNL25group4	mouse	M6	Mouse_BM	chr3+71430357	20.97%	1021	Gm6634
pMouseCNL25group4	mouse	M6	Mouse_BM	chr9+15188460	22.77%	1109	Cep295
pMouseCNL38group1A	mouse	M6	Mouse_BM	chr14-102130900	50.00%	1	Lmo7
pMouseCNL38group1A	mouse	M6	Mouse_BM	chr7-115316991	50.00%	1	Olf486
pMouseCN493	mouse	M6	Mouse_BM	chr14+15556797	22.83%	989	Slc4a7
pMouseCN493	mouse	M6	Mouse_BM	chr6-10697710	22.25%	964	AA545190
pMouseCN493	mouse	M6	Mouse_BM	chrX-111592392	22.30%	966	Klh14
pMouseCN493	mouse	M6	Mouse_BM	chrX-155983905	23.50%	1018	Bclaf3
pMouseCNL53	mouse	M6	Mouse_BM	chr19+18123881	100.00%	2	Pcsk5
pMouseCNL58	mouse	M6	Mouse_BM	chr2+141127890	100.00%	3	MacroD2
pMouseCNL59	mouse	M6	Mouse_BM	chr2+52586806	25.11%	407	Stam2
pMouseCNL59	mouse	M6	Mouse_BM	chr3-59165408	22.52%	365	Igsf10
pMouseCNL59	mouse	M6	Mouse_BM	chr9+15188460	24.18%	392	Cep295
pMouseCN545group11	mouse	M6	Mouse_BM	chr18-15129653	36.11%	1006	Kctd1
pMouseCN545group11	mouse	M6	Mouse_BM	chr5-12432693	32.88%	916	Sema3d
pMouseCNL70group12	mouse	M6	Mouse_BM	chr11-77602820	22.05%	284	Myo18a
pCNL80	mouse	M6	Mouse_BM	chr3+123970190	34.43%	42	Tram111
pCNL82	mouse	M6	Mouse_BM	chr10-40246905	50.00%	1	Slc22a16
pCNL82	mouse	M6	Mouse_BM	chr10+94767778	50.00%	1	Cradd
pCN552	mouse	M6	Bone marrow CD11b	chr3+130245919	21.62%	16	Col25a1
pCNL59	mouse	M6	Bone marrow CD11b	chr2+52586806	30.38%	24	Stam2
pCNL59	mouse	M6	Bone marrow CD11b	chr3-59165408	26.58%	21	Igsf10
pCNL59	mouse	M6	Bone marrow CD11b	chr9+15188460	30.38%	24	Cep295
pCNL59	mouse	M6	B Cells	chr2+52586806	35.00%	7	Stam2
pCNL59	mouse	M6	B Cells	chr3+71430357	30.00%	6	Gm6634
pCNL59	mouse	M6	B Cells	chr9+15188460	30.00%	6	Cep295
pCN760	mouse	M6	Bone Marrow	chr9+45858945	29.41%	5	Sik3
pCN806	mouse	M6	Bone Marrow	chr13-16444618	22.22%	2	Inhba

# Mouse transplant trials

The positions of identified integration sites from cell transplant trials with nine pairs of mice are shown in Figure 5a (donor mice) and Figure 5b (recipient mice). The gRxCluster software package did not identify clusters of integration sites between donor and recipient mice with a false discovery rate of  $\leq 10\%$ . The relative clonal abundances of samples from the transplant trials are shown in Figure 6 where donor mice are shown on the left and recipient mice are shown on the right. Integration sites are denoted by both nearest gene and genomic coordinate and annotated with an asterisk (\*) if located within transcription units and with a tilde (~) if the integration site is within 50 KB of an oncogene. Below each abundance plot is a Fisher's exact test for the enrichment of oncogenes. None of the tests returned a significant result. The clonal abundances of clones found in both donor and recipient mice is shown in Table S3. The identification of relatively few persistent clones is likely due to sequencing experiments sampling only a subset of existing integration sites and a number of samples with low vector copy numbers (Figures 5B & S3).

Figure 5

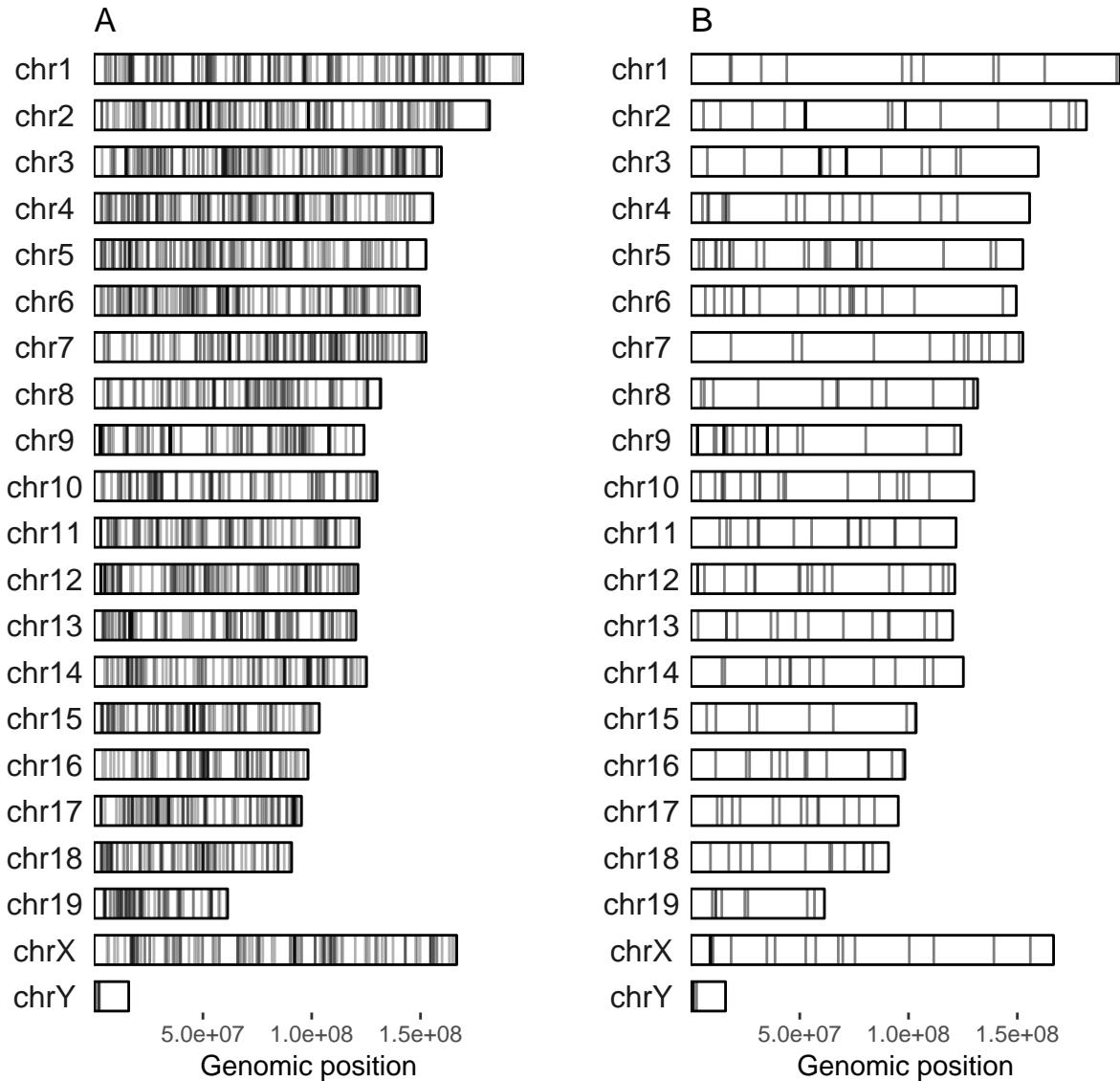
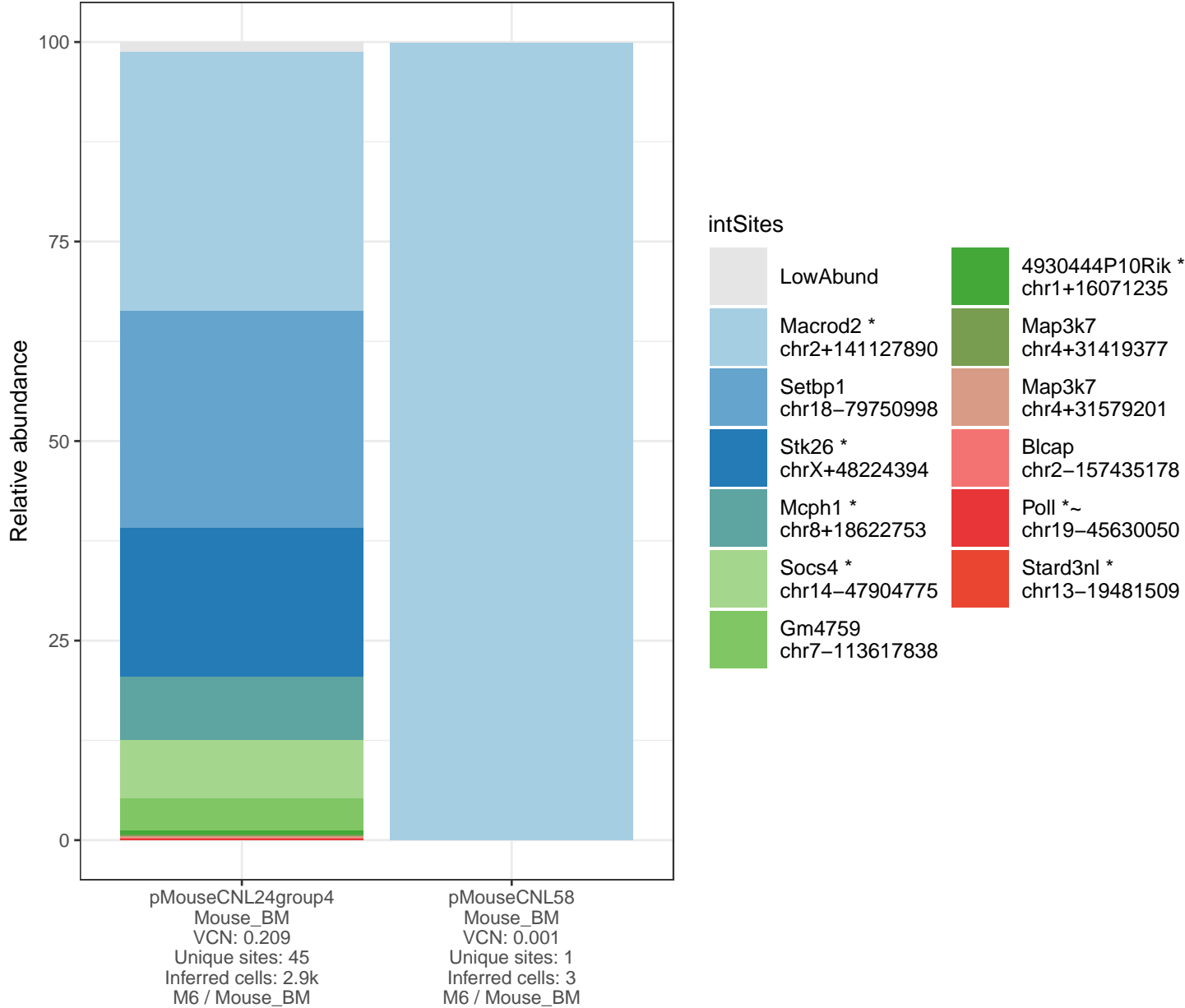


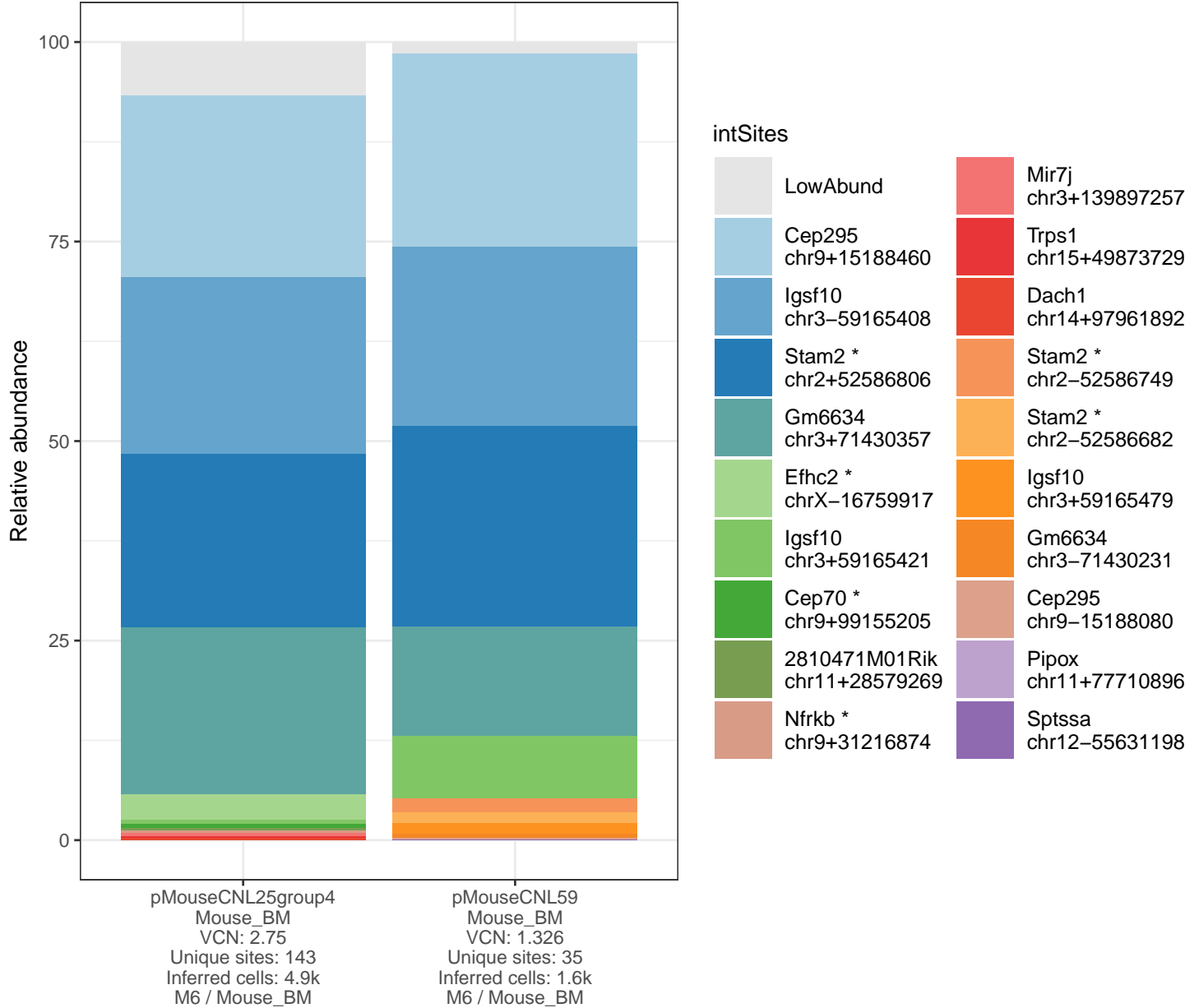
Figure 6a.



Fisher's exact p-value: 1.000

	Not near onco	Near onco
pMouseCNL24group4	42	3
pMouseCNL58	1	0

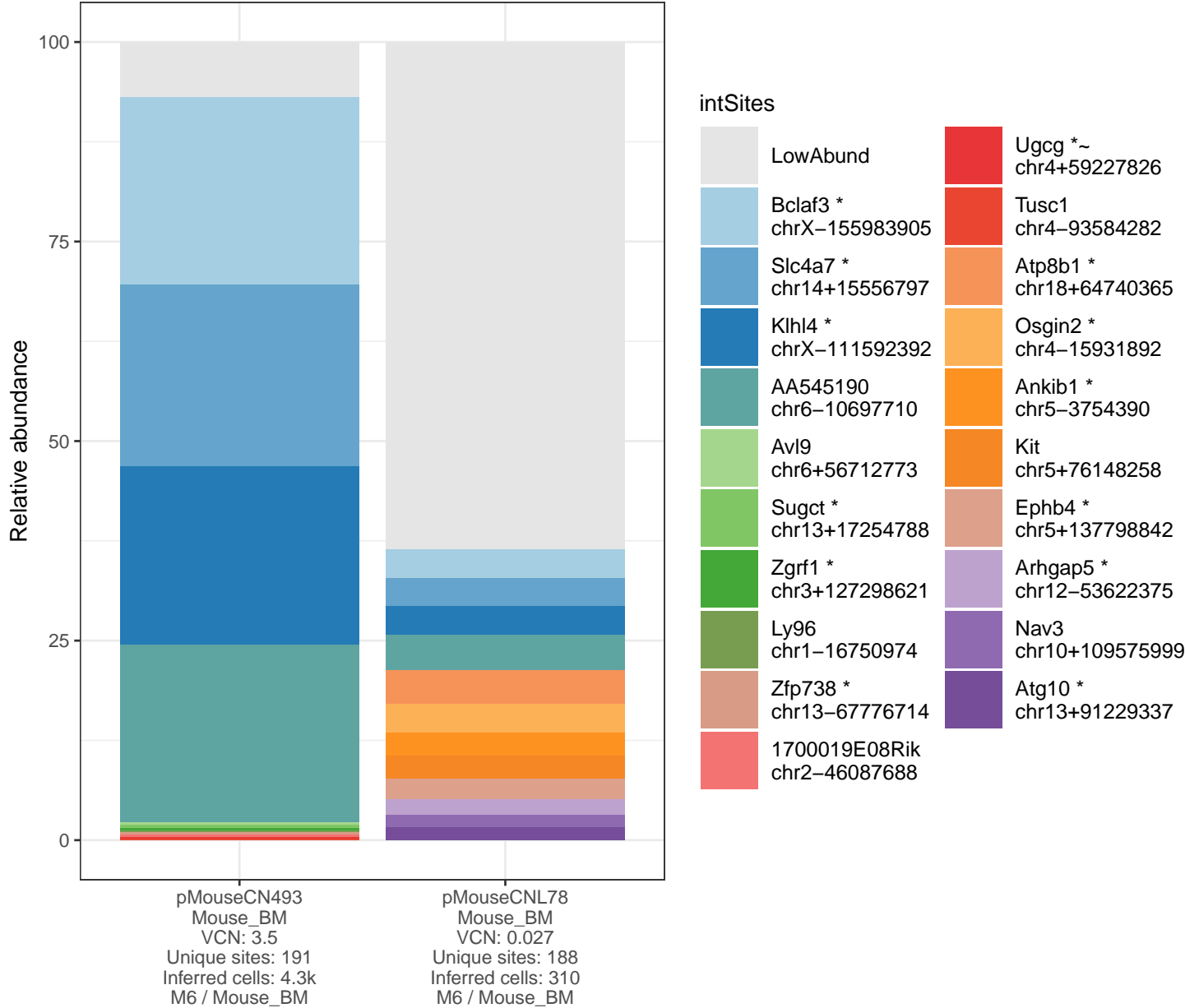
Figure 6b.



Fisher's exact p-value: 0.208

	Not near onco	Near onco
pMouseCNL25group4	134	9
pMouseCNL59	35	0

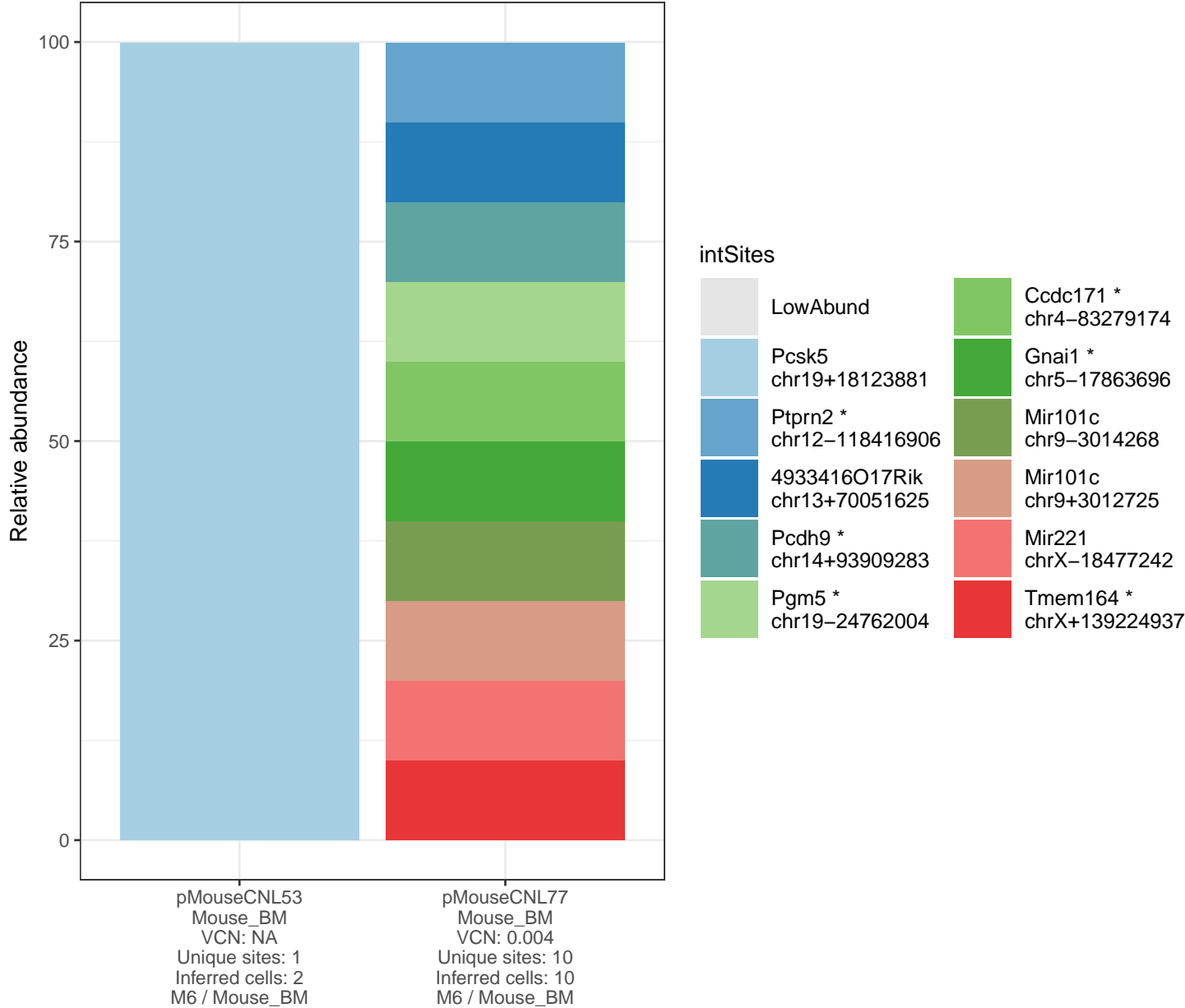
Figure 6c.



Fisher's exact p-value: 0.492

	Not near onco	Near onco
pMouseCN493	179	12
pMouseCNL78	180	8

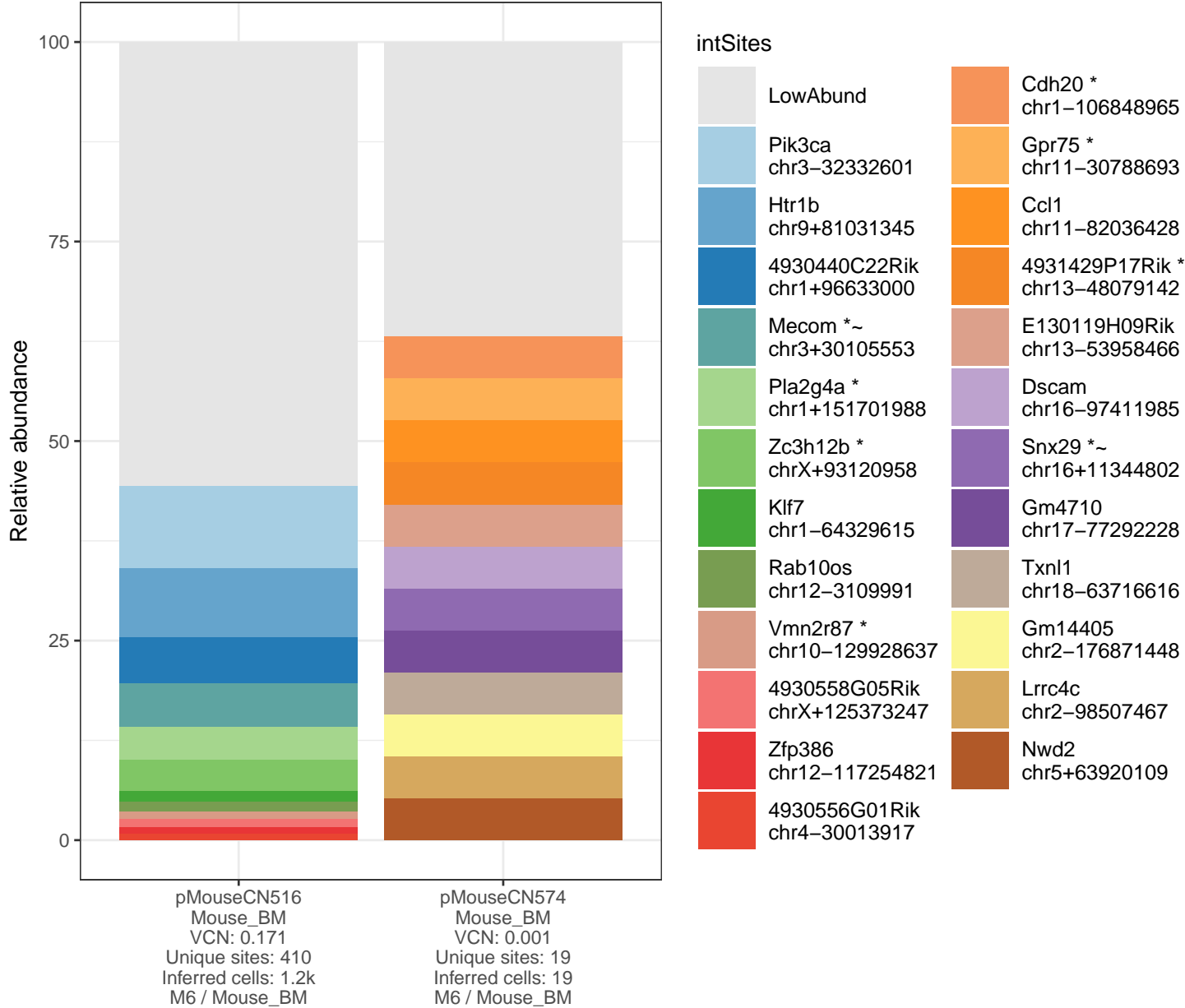
Figure 6d.



Fisher's exact p-value: 1.000

	Not near onco	Near onco
pMouseCNL53	1	0
pMouseCNL77	10	0

Figure 6e.

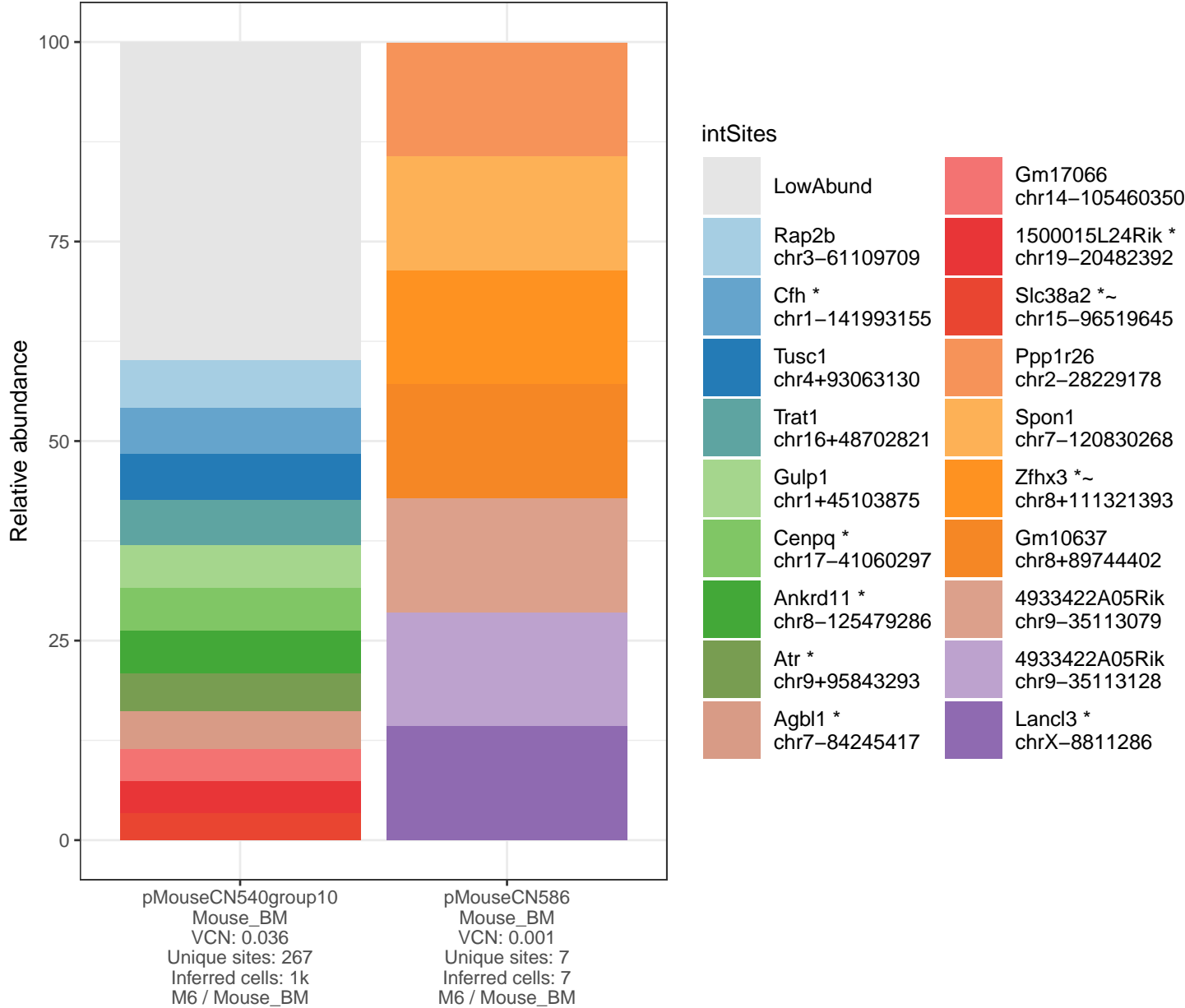


Fisher's exact p-value: 0.396

	Not near onco	Near onco
pMouseCN516	400	10
pMouseCN574	18	1



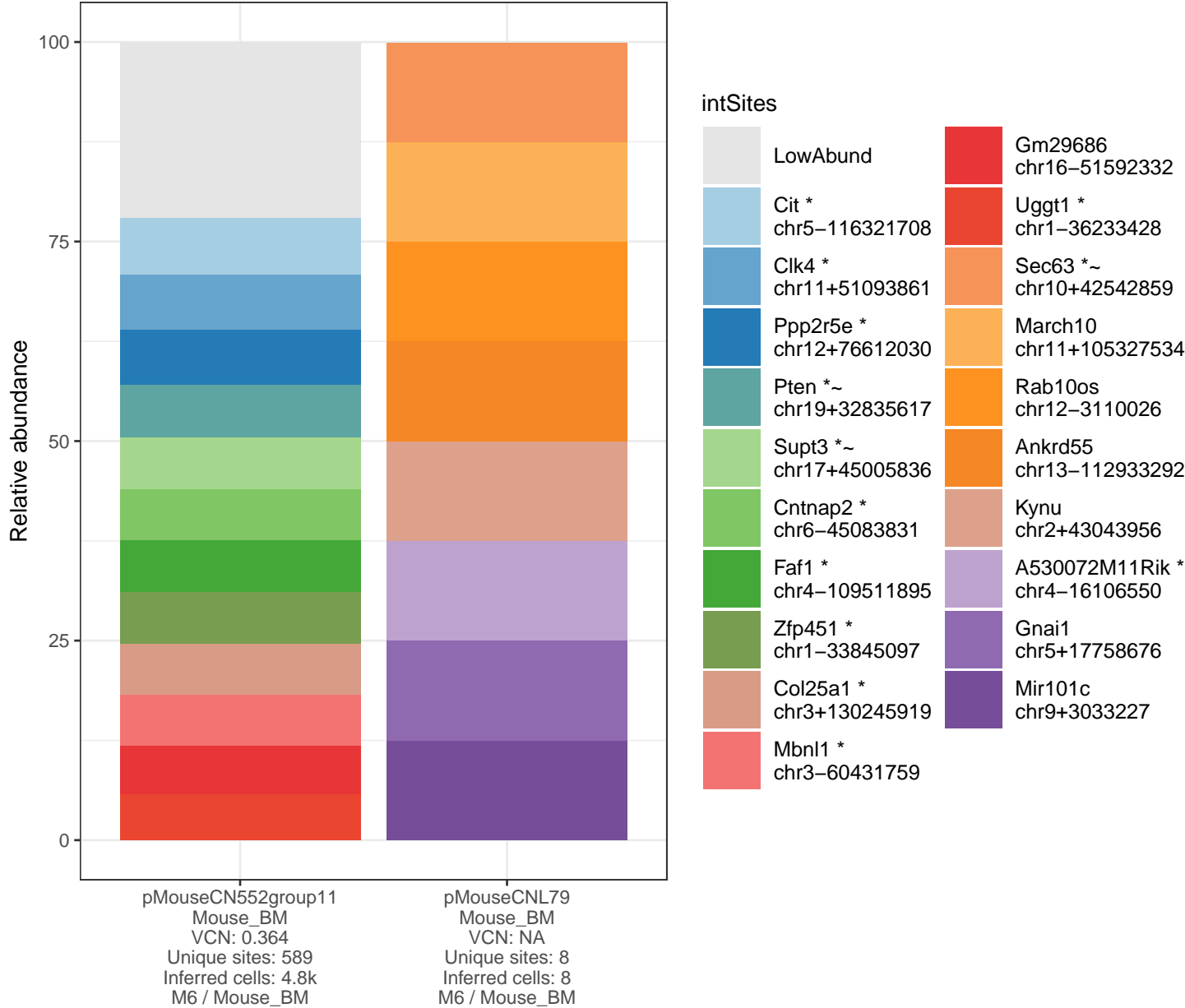
Figure 6f.



Fisher's exact p-value: 0.447

	Not near onco	Near onco
pMouseCN540group10	246	21
pMouseCN586	6	1

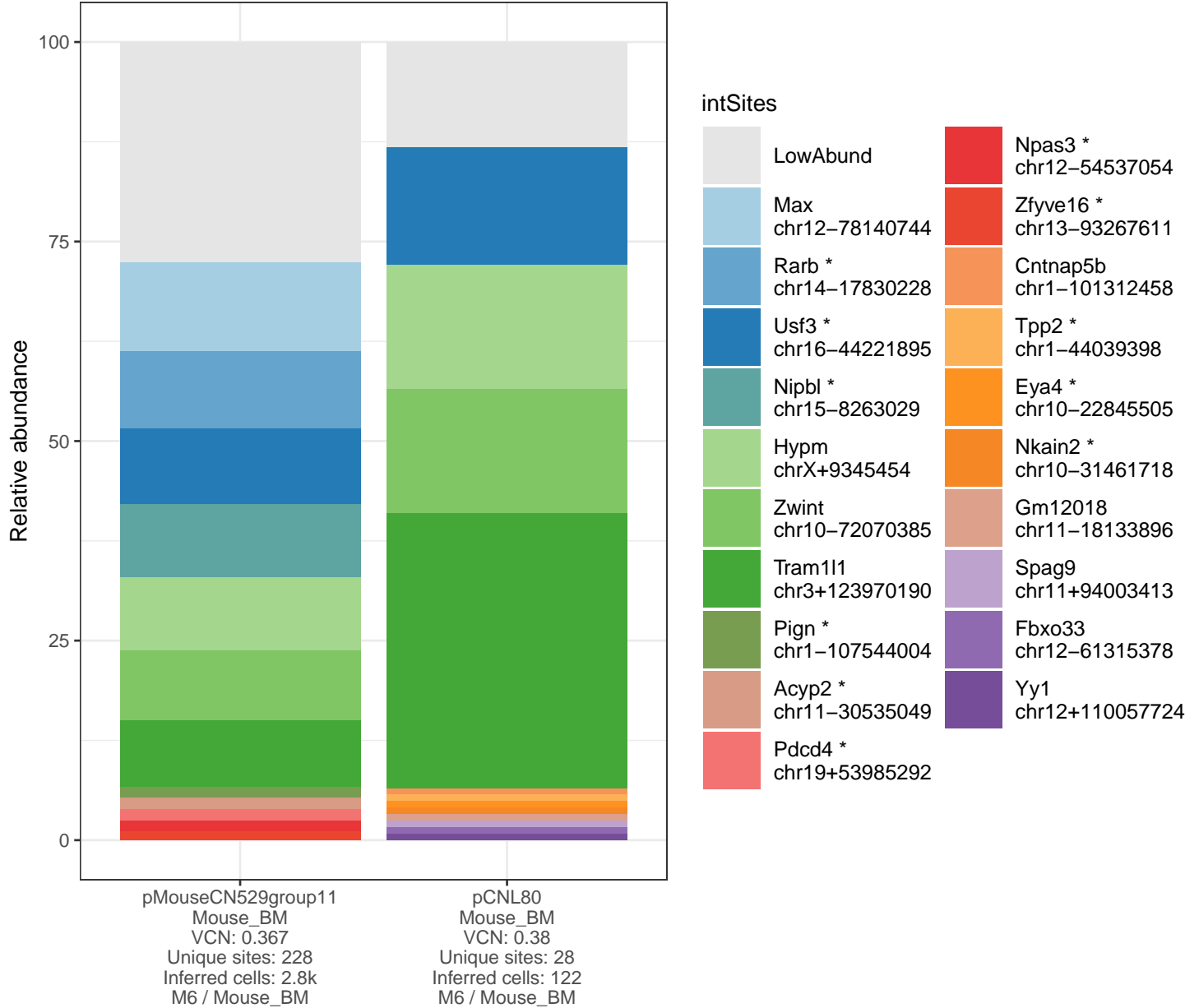
Figure 6g.



Fisher's exact p-value: 0.394

	Not near onco	Near onco
pMouseCN552group11	554	35
pMouseCNL79	7	1

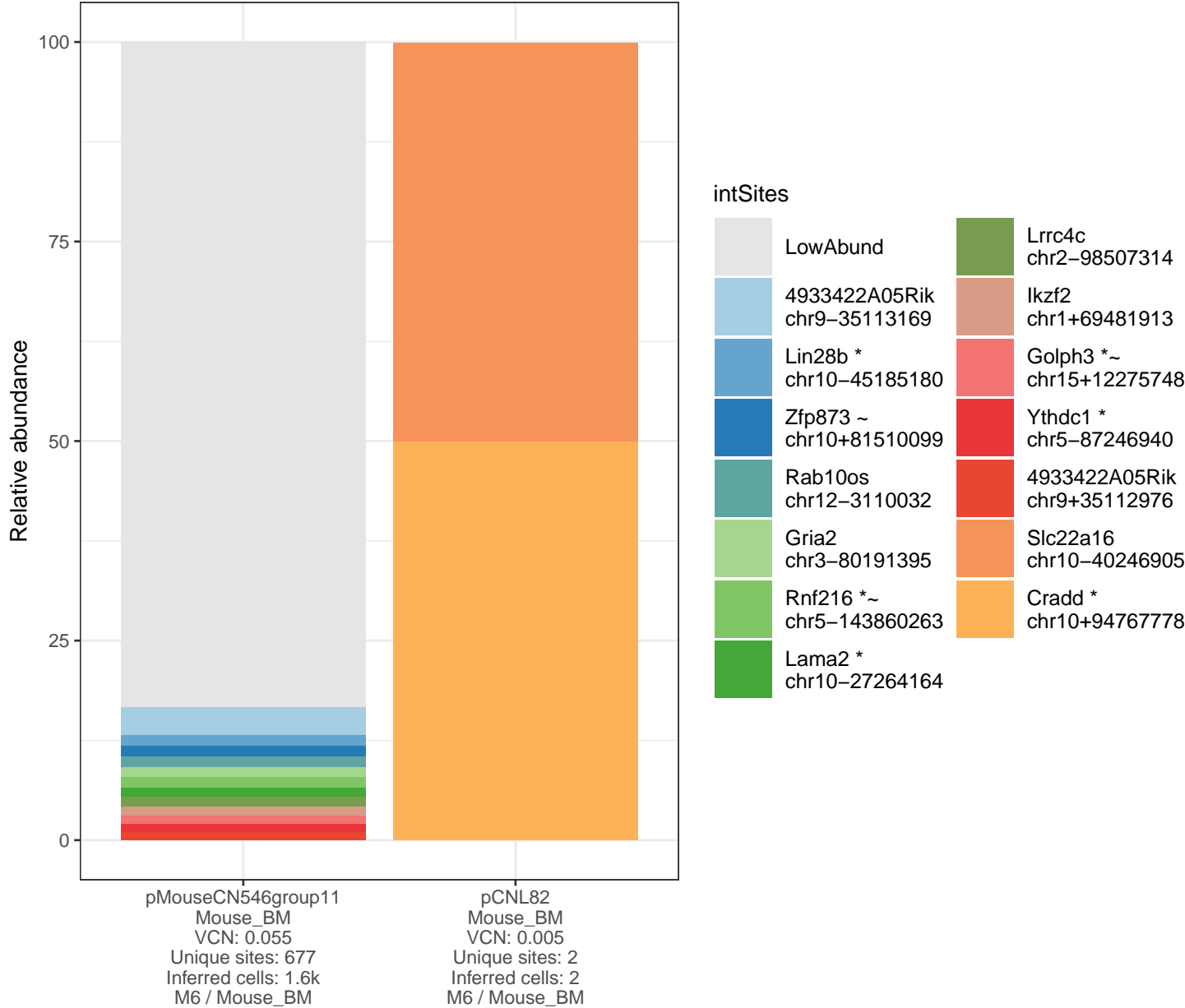
Figure 6h.



Fisher's exact p-value: 1.000

	Not near onco	Near onco
pMouseCN529group11	212	16
pCNL80	26	2

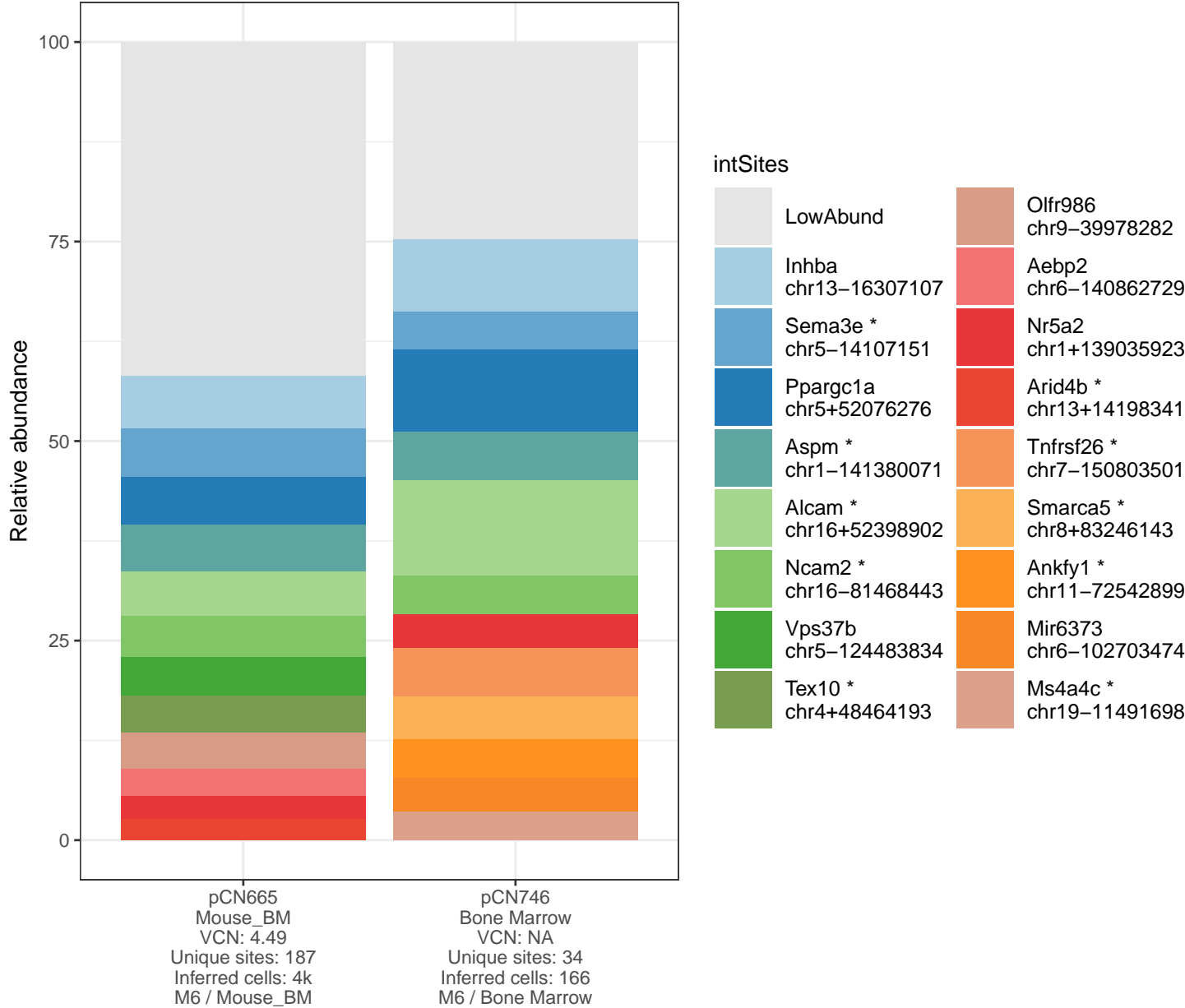
Figure 6i.



Fisher's exact p-value: 1.000

	Not near onco	Near onco
pMouseCN546group11	647	30
pCNL82	2	0

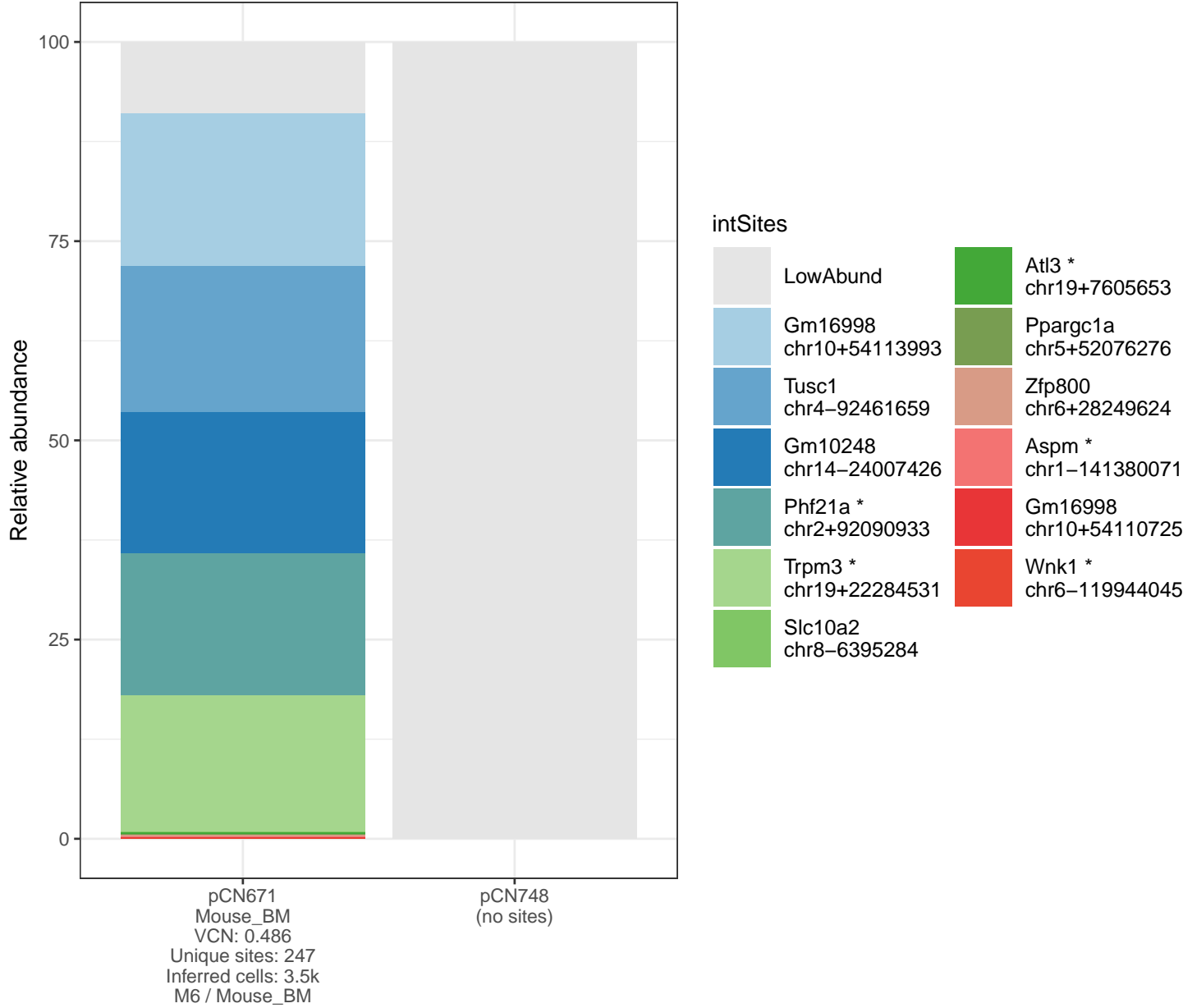
Figure 6j.



Fisher's exact p-value: 0.594

	Not near onco	Near onco
pCN665	181	6
pCN746	34	0

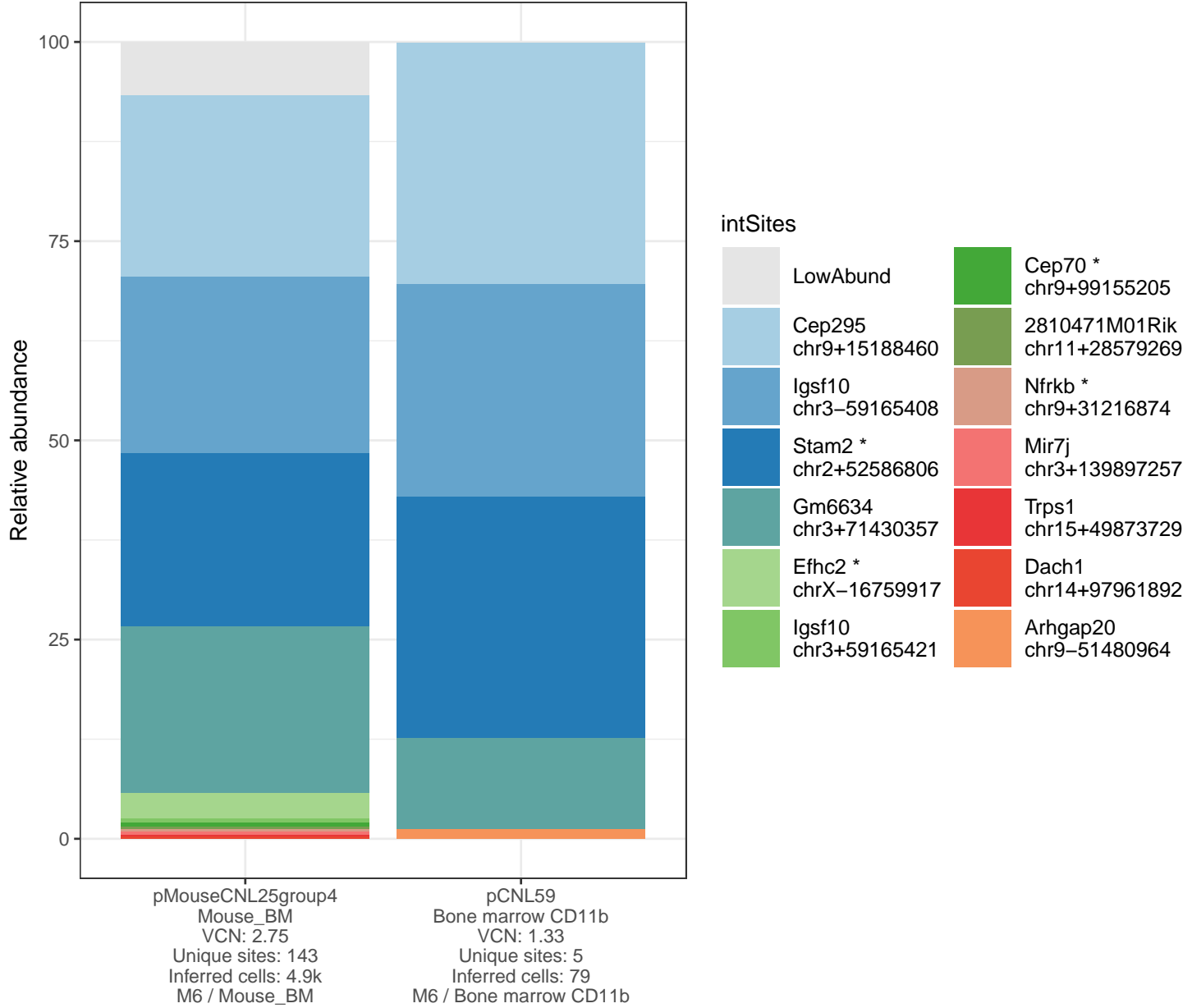
Figure 6k.



Fisher's exact p-value: 1.000

	Not near onco	Near onco
pCN671	233	14
NA	0	0

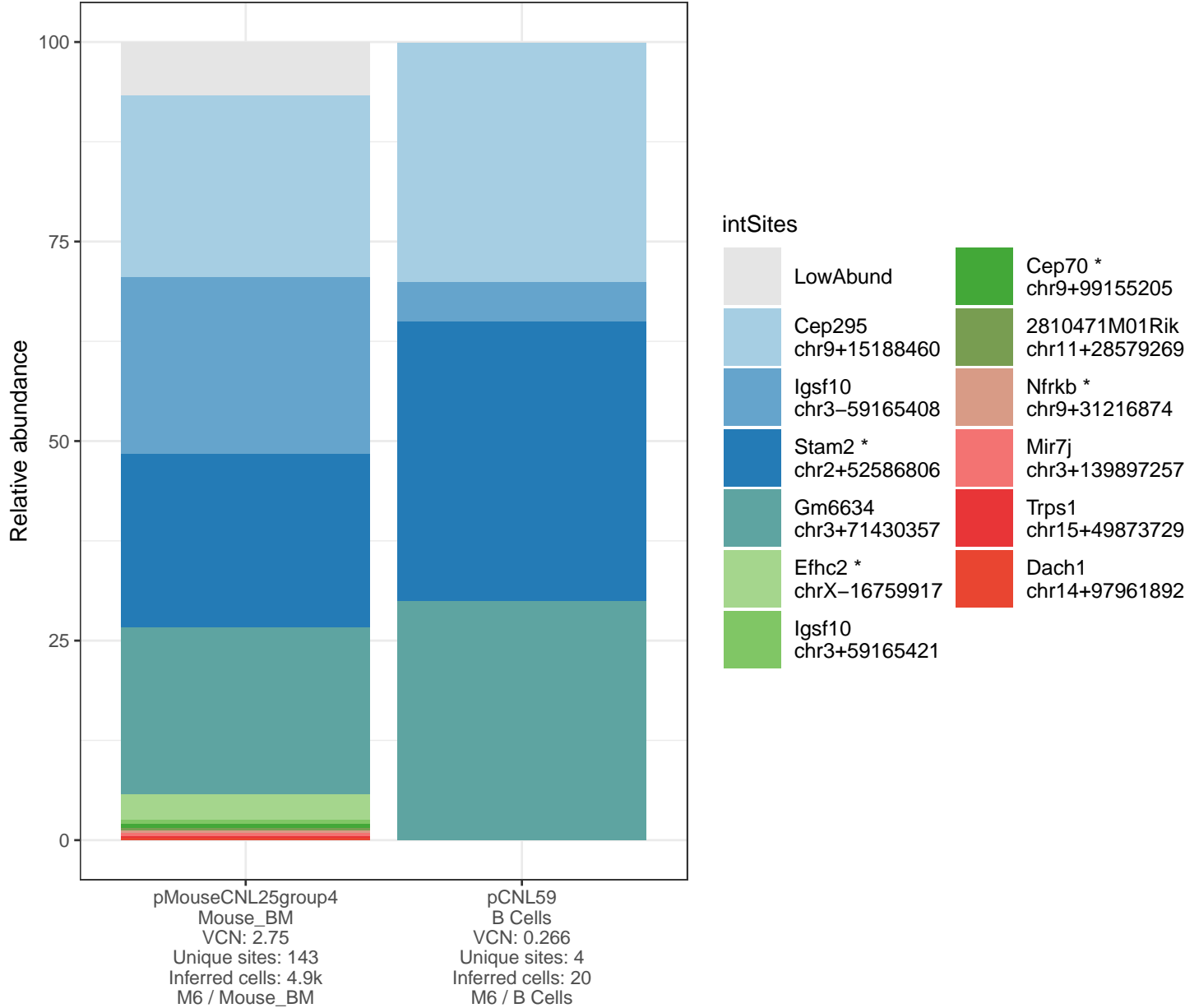
Figure 6l.



Fisher's exact p-value: 1.000

	Not near onco	Near onco
pMouseCNL25group4	134	9
pCNL59	5	0

Figure 6m.

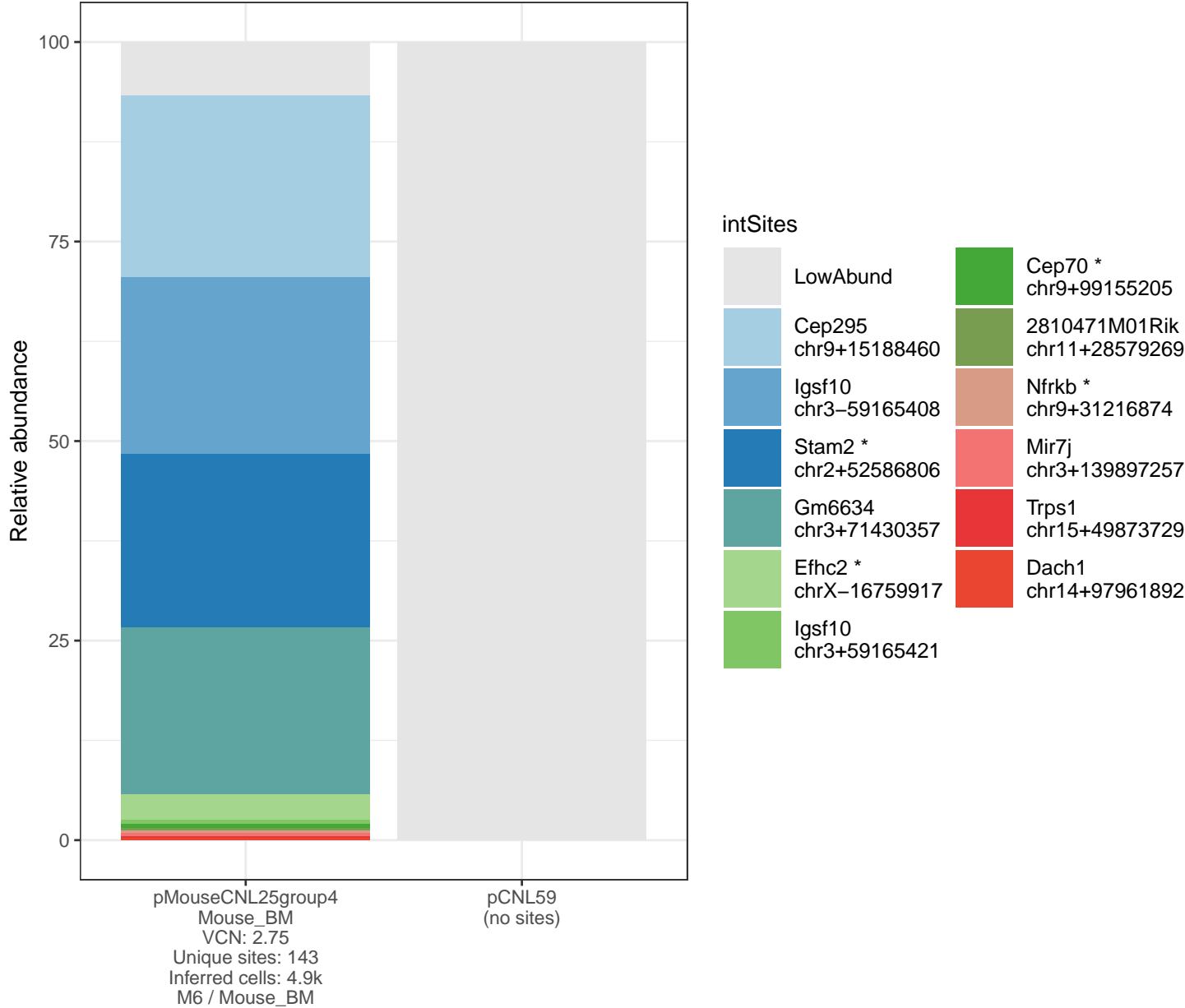


Fisher's exact p-value: 1.000

	Not near onco	Near onco
pMouseCNL25group4	134	9
pCNL59	4	0



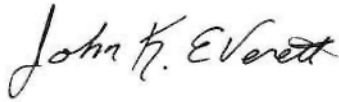
Figure 6n.



Fisher's exact p-value: 1.000

	Not near onco	Near onco
pMouseCNL25group4	134	9
NA	0	0

## Analyst



John K. Everett, Ph.D.

## Laboratory director



Frederic D. Bushman, Ph.D.

## References

1. RTCGD: retroviral tagged cancer gene database. Akagi K, Suzuki T, Stephens RM, Jenkins NA, Copeland NG. *Nucleic Acids Res.* 2004 Jan 1;32(Database issue):D523-7.
2. Outcomes following gene therapy in patients with severe Wiskott-Aldrich syndrome. Hacein-Bey Abina S, Gaspar HB, Blondeau J, Caccavelli L, Charrier S, Buckland K, Picard C, Six E, Himoudi N, Gilmour K, McNicol AM, Hara H, Xu-Bayford J, Rivat C, Touzot F, Mavilio F, Lim A, Treluyer JM, Héritier S, Lefrère F, Magalon J, Pengue-Koyi I, Honnet G, Blanche S, Sherman EA, Male F, Berry C, Malani N, Bushman FD, Fischer A, Thrasher AJ, Galy A, Cavazzana M. *JAMA.* 2015 Apr 21;313(15):1550-63.
3. Distribution of Lentiviral Vector Integration Sites in Mice Following Therapeutic Gene Transfer to Treat  $\beta$ -thalassemia. Ronen K, Negre O, Roth S, Colomb C, Malani N, Denaro M, Brady T, Fusil F, Gillet-Legrand B, Hehir K, Beuzard Y, Leboulch P, Down JD, Payen E, Bushman FD. *Mol Ther.* 2011 Jul;19(7):1273-86.
4. Estimating abundances of retroviral insertion sites from DNA fragment length data. Berry CC, Gillet NA, Melamed A, Gormley N, Bangham CR, Bushman FD. *Bioinformatics.* 2012 Mar 15;28(6):755-62.
5. INSPIRED: A Pipeline for Quantitative Analysis of Sites of New DNA Integration in Cellular Genomes. Sherman E, Nobles C, Berry CC, Six E, Wu Y, Dryga A, Malani N, Male F, Reddy S, Bailey A, Bittinger K, Everett JK, Caccavelli L, Drake MJ, Bates P, Hacein-Bey-Abina S, Cavazzana M, Bushman FD. *Mol Ther Methods Clin Dev.* 2016 Dec 18;4:39-49.

# Supplementary tables and figures

## Numbers of inferred cells and integration sites identified in provided samples

Table S1.

Organism	GTSP	Subject	Cell type	VCN	Time point	Number inferred cells	Number of intSites
human	GTSP0689	pCYS002NS	PBCD34-MOI20	0.940	D14	3,146	1,467
human	GTSP0691	pCYS004BR	PBCD34-MOI20	0.930	D14	1,091	761
human	GTSP0692	pCYS005OB	PB_CD34	0.970	D14	8,698	3,291
human	GTSP1369	pDON.002.4.SC	PB_CD34	4.520	D14	129	44
human	GTSP1370	pDON.003.2.SG	PB_CD34	2.440	D14	1,489	1,173
human	GTSP1371	pDON.007.CR	PB_CD34	1.270	D14	114	75
human	GTSP1372	pDON.008.LH	PB_CD34	2.530	D14	134	100
human	GTSP1373	pDON.009.JH	PB_CD34	5.280	D14	793	605
human	GTSP1374	pDON.010.CC	PB_CD34	3.290	D14	531	370
human	GTSP1375	pCYS.013.PS	PBCD34-mock	NA	D14	1	1
human	GTSP1376	pCYS.007.2.BS	PBCD34-mock	0.002	D14	4	4
human	GTSP1377	pCYS.002.4.NS	PB_CD34	2.400	D14	105	54
human	GTSP1378	pCYS.007.2.BS	PBCD34-MOI20	3.640	D14	568	460
human	GTSP1379	pCYS.007.2.BS	PBCD34-MOI40	6.100	D14	1,225	1,057
human	GTSP1380	pCYS.013.PS	PBCD34-MOI40	1.000	D14	49	25
human	GTSP1682	pCYS_017_BW_MOI2x107	PB_CD34	1.660	D14	16,712	11,675
human	GTSP1683	pCYS_017_BW_MOI6x106	PB_CD34	1.415	D14	16,717	11,074
human	GTSP1685	pCYS_018_MM_MOI2x107	PB_CD34	0.958	D14	11,700	8,430
human	GTSP1686	pCYS_018_MM_MOI6x106	PB_CD34	0.530	D14	7,126	4,541
human	GTSP1687	pCYS_018_MM_MOI2x106	PB_CD34	0.331	D14	3,589	2,502
human	GTSP1688	pCYS_020_JJ	PB_CD34	2.590	D14	21,755	14,752
human	GTSP1689	pMP011.1.Ch2	PB_CD34	1.607	D14	1,517	938
human	GTSP1690	pSC002.6.Ch7	PB_CD34	2.350	D14	427	221
human	GTSP1691	pMP011.1.Ch9	PB_CD34	2.420	D14	726	426
human	GTSP1692	pMP011.1.Ch11	PB_CD34	1.074	D14	626	352
human	GTSP1694	pDON.004.3.TL.MOI20	PB_CD34	3.250	D14	5,270	4,562
human	GTSP1695	pDON.011.2.MP	PB_CD34	3.950	D14	11,241	7,060
human	GTSP1969	pLSR1TEST2	BM CD34+ Cells	0.695	D0	2,584	2,215
human	GTSP1970	pLSR1THAW2	BM CD34+ Cells	0.709	D0	6,344	6,003
human	GTSP1971	pLSR2Thaw2	BM CD34+ Cells	1.020	D0	9,196	8,686

Table S1 (continued).

Organism	GTSP	Subject	Cell type	VCN	Time point	Number inferred cells	Number of intSites
mouse	GTSP0829	pMouseCNL23group4control	Mouse_BM	NA	M6	6	5
mouse	GTSP0830	pMouseCNL24group4	Mouse_BM	0.209	M6	2,881	45
mouse	GTSP0831	pMouseCNL25group4	Mouse_BM	2.750	M6	4,870	143
mouse	GTSP0832	pMouseCNL38group1A	Mouse_BM	0.001	M6	2	2
mouse	GTSP0833	pMouseCulture	Mouse_Sca1pos	2.760	D14	89	28
mouse	GTSP0891	pMouseCN493	Mouse_BM	3.500	M6	4,332	191
mouse	GTSP0892	pMouseCNL53	Mouse_BM	NA	M6	2	1
mouse	GTSP0947	pMouseCNL58	Mouse_BM	0.001	M6	3	1
mouse	GTSP0948	pMouseCNL59	Mouse_BM	1.326	M6	1,621	35
mouse	GTSP0949	pMouseCN516	Mouse_BM	0.171	M6	1,174	410
mouse	GTSP0950	pMouseCN518	Mouse_BM	2.300	M6	3,379	101
mouse	GTSP1079	pMouseCN540group10	Mouse_BM	0.036	M6	1,009	267
mouse	GTSP1080	pMouseCN552group11	Mouse_BM	0.364	M6	4,849	589
mouse	GTSP1081	pMouseCN529group11	Mouse_BM	0.367	M6	2,786	228
mouse	GTSP1082	pMouseCN545group11	Mouse_BM	1.339	M6	2,786	35
mouse	GTSP1083	pMouseCN546group11	Mouse_BM	0.055	M6	1,644	677
mouse	GTSP1084	pMouseCNL70group12	Mouse_BM	0.610	M6	1,288	92
mouse	GTSP1146	pMouseCN557	Mouse_BM	2.600	M6	10,839	745
mouse	GTSP1147	pMouseCNL77	Mouse_BM	0.004	M6	10	10
mouse	GTSP1148	pMouseCNL78	Mouse_BM	0.027	M6	310	188
mouse	GTSP1149	pMouseCN574	Mouse_BM	0.001	M6	19	19
mouse	GTSP1150	pMouseCN586	Mouse_BM	0.001	M6	7	7
mouse	GTSP1151	pMouseCNL79	Mouse_BM	NA	M6	8	8
mouse	GTSP1481	pCNL80	Mouse_BM	0.380	M6	122	28
mouse	GTSP1482	pCNL82	Mouse_BM	0.005	M6	2	2
mouse	GTSP1696	pCN665	Mouse_BM	4.490	M6	4,012	187
mouse	GTSP1697	pCN671	Mouse_BM	0.486	M6	3,453	247
mouse	GTSP1705	pCN552	Bone marrow CD11b	0.101	M6	74	15
mouse	GTSP1706	pCN552	T-Cells	0.514	M6	192	49
mouse	GTSP1707	pCN552	B Cells	4.230	M6	922	161
mouse	GTSP1710	pCNL59	Bone marrow CD11b	1.330	M6	79	5
mouse	GTSP1712	pCNL59	B Cells	0.266	M6	20	4
mouse	GTSP1867	pCN746	Bone Marrow	NA	M6	166	34
mouse	GTSP1972	pCN752	Bone Marrow	0.004	M6	238	71
mouse	GTSP1973	pCN755	Bone Marrow	NA	M6	28	27
mouse	GTSP1974	pCN760	Bone Marrow	0.001	M6	17	10
mouse	GTSP1980	pCN804	Bone Marrow	0.003	M6	82	28
mouse	GTSP1981	pCN806	Bone Marrow	0.001	M6	9	8

## Analyzed samples in which no integration sites were identified

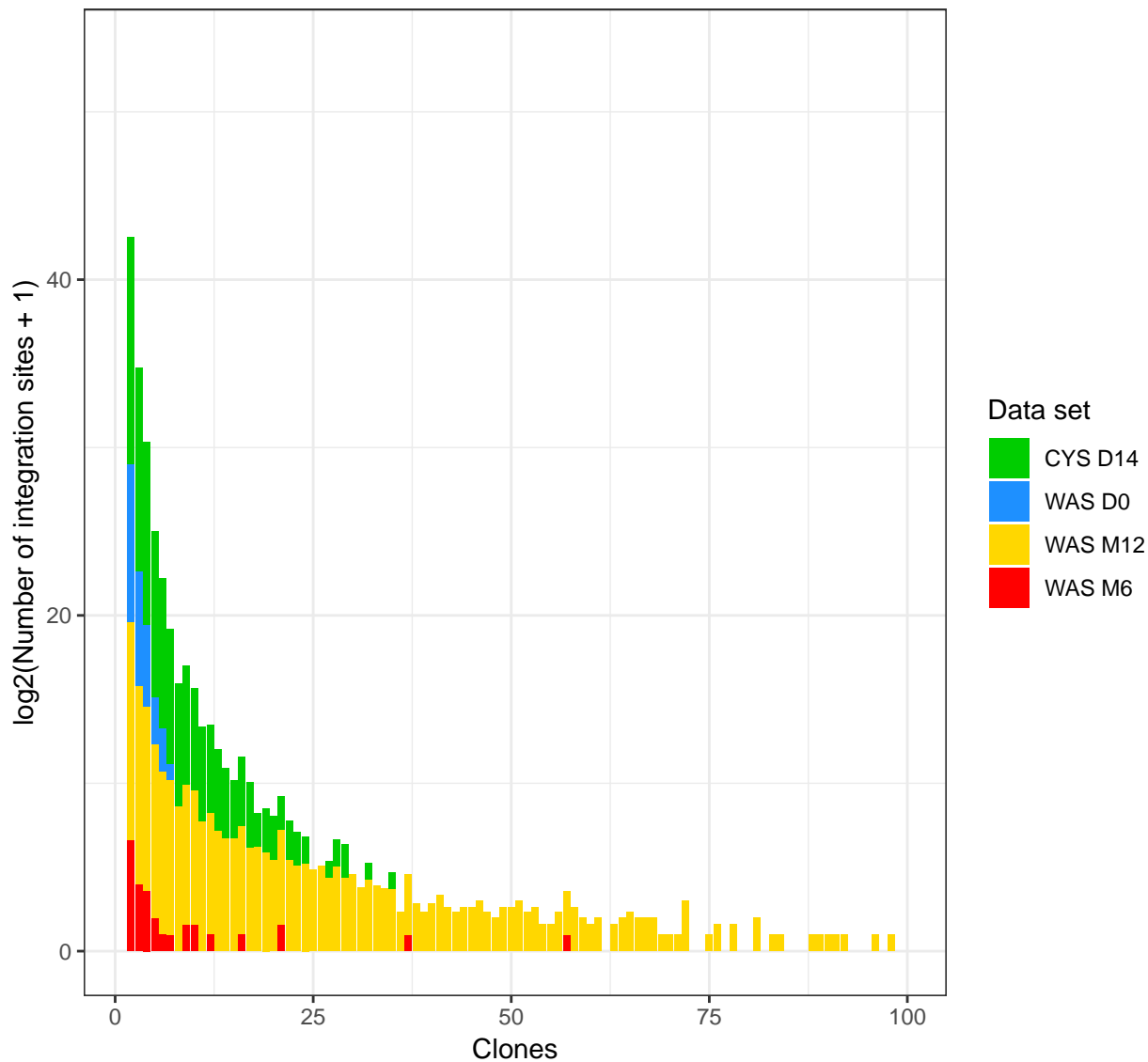
Table S2.

GTSP	Cell type	Subject	Time point	Comment
GTSP0688	PBCD34-mock	pCYS002NS	D14	Mock transduced control; Stage: Mock
GTSP0690	PBCD34-mock	pCYS004BR	D14	Mock transduced control; Stage: Mock
GTSP0828	Mouse_BM	pMouseCNL1group1	M6	pCCL-CTNS; MOI 10; Stage: Primary
GTSP0834	Mouse_Sca1pos	pMouseCultureControl	D14	Control, DNA was extracted from mouse Sca1+ cells and cultured for 2 weeks (Mock)
GTSP1684	PB_CD34	pCYS_017_BW MOCK	D14	Mock transduced control; Stage: Mock
GTSP1693	PB_CD34	pDON.004.3.TL.MOCK	D14	Mock transduced control; Stage: Mock
GTSP1711	T-Cells	pCNL59	M6	Mouse Thymus, T cells
GTSP1868	Bone Marrow	pCN748	M6	pCCL-CTNS transduced, Secondary graft, Primary graft mouse: CN671, Primary graft
GTSP1975	Thorax	pCN752P1	6M	Pathology sample 1 from CN752 (found in thorax)
GTSP1976	Thymus	pCN752P2	6M	Pathology sample 2 from CN752 (thymus)

## Comparison of the number of integration sites and inferred cells

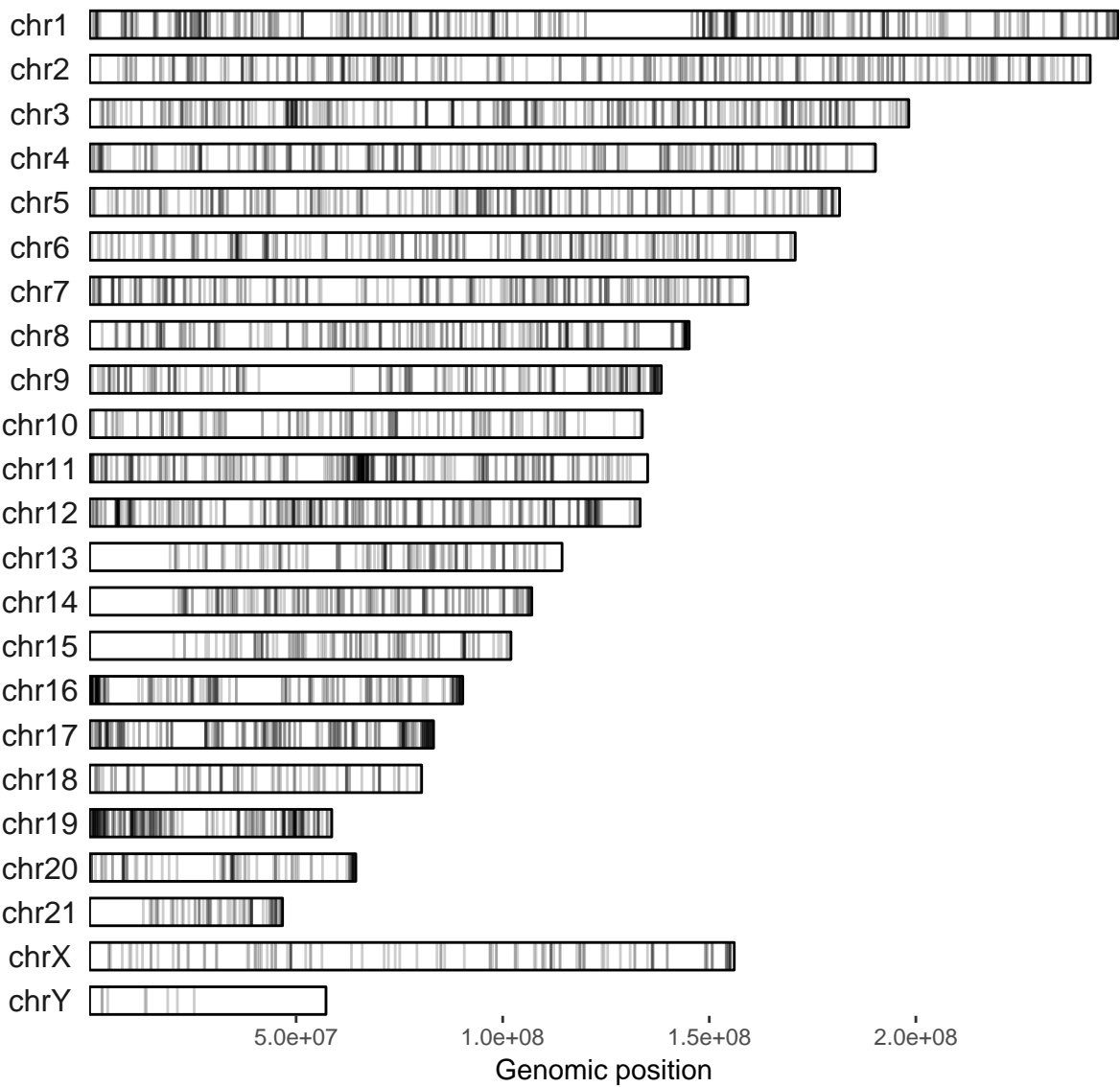
Kolmogorov–Smirnov tests show no significant differences between the distributions of estimated abundances between trials. The plot below was truncated at 100 inferred cells for clarity which eliminated 16 data points from the ‘WAS m12’ distribution.

Figure S1.



Integration positions of the lentiviral therapeutic vector from a previous WAS correction trial<sup>1</sup>

Figure S2.



## Persistence of clones in mouse BM transplant trials

Table S3.

Donor	Recipient	Position	Donor cells	Recipient cells
pMouseCNL24group4	pMouseCNL58	chr2+141127890	935	3
pMouseCNL25group4	pMouseCNL59	chr12-55631198	3	1
pMouseCNL25group4	pMouseCNL59	chr2-52586682	1	22
pMouseCNL25group4	pMouseCNL59	chr2-52586749	7	27
pMouseCNL25group4	pMouseCNL59	chr2+52586806	1,057	407
pMouseCNL25group4	pMouseCNL59	chr3-59165408	1,082	365
pMouseCNL25group4	pMouseCNL59	chr3-71430231	2	9
pMouseCNL25group4	pMouseCNL59	chr3+59165404	1	1
pMouseCNL25group4	pMouseCNL59	chr3+59165421	27	127
pMouseCNL25group4	pMouseCNL59	chr3+59165479	6	22
pMouseCNL25group4	pMouseCNL59	chr3+71430357	1,021	222
pMouseCNL25group4	pMouseCNL59	chr9+15188460	1,109	392
pMouseCN493	pMouseCNL78	chr14+15556797	989	11
pMouseCN493	pMouseCNL78	chr6-10697710	964	14
pMouseCN493	pMouseCNL78	chrX-111592392	966	11
pMouseCN493	pMouseCNL78	chrX-155983905	1,018	11
pMouseCN540group10	pMouseCN586	chr9-35113079	1	1
pMouseCN552group11	pMouseCNL79	chr12-3110026	10	1
pMouseCN529group11	pCNL80	chr10-72070385	242	19
pMouseCN529group11	pCNL80	chr16-44221895	263	18
pMouseCN529group11	pCNL80	chr3+123970190	233	42
pMouseCN529group11	pCNL80	chrX+9345454	255	19
pCN665	pCN746	chr1-141380071	235	10
pCN665	pCN746	chr1+139035923	115	7
pCN665	pCN746	chr11-72542899	6	8
pCN665	pCN746	chr13-16307093	1	2
pCN665	pCN746	chr13-16307107	265	15
pCN665	pCN746	chr13+90648207	3	5
pCN665	pCN746	chr16-81468443	208	8
pCN665	pCN746	chr16+52398902	224	20



Table S3. (continued)

	Donor	Recipient	Position	Donor cells	Recipient cells
31	pCN665	pCN746	chr19-11491698	7	6
32	pCN665	pCN746	chr2+5793957	4	3
33	pCN665	pCN746	chr4+48464193	189	4
34	pCN665	pCN746	chr5-14107151	245	8
35	pCN665	pCN746	chr5+52076276	240	17
36	pCN665	pCN746	chr6-102703474	4	7
37	pCN665	pCN746	chr7-150803501	105	10
38	pCN665	pCN746	chr7+46794970	5	2
39	pCN665	pCN746	chr8+83246143	6	9
40	pCN665	pCN746	chr9-39978282	181	6
41	pCN665	pCN746	chrUn_random+46156	4	4
42	pMouseCNL25group4	pCNL59	chr2+52586806	1,057	24
43	pMouseCNL25group4	pCNL59	chr3-59165408	1,082	21
44	pMouseCNL25group4	pCNL59	chr3+71430357	1,021	9
45	pMouseCNL25group4	pCNL59	chr9+15188460	1,109	24
46	pMouseCNL25group4	pCNL59	chr2+52586806	1,057	7
47	pMouseCNL25group4	pCNL59	chr3-59165408	1,082	1
48	pMouseCNL25group4	pCNL59	chr3+71430357	1,021	6
49	pMouseCNL25group4	pCNL59	chr9+15188460	1,109	6

# Sequencing depth

Identified integration site are shown as colored squares that are positioned by the number of reads leading to their identification.

Figure S3.

