

# Analysis of integration site distributions and clonal abundances for gene therapy correction of cystinosis (Groups 29-30)

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

*November 2019*

## Contents

Summary of results	2
Mouse samples studied	3
Subject reports	3
UCSC browser exploration	3
Description of analysis techniques	4
Comparisons to previous trials	5
Integration events near oncogenes in mouse subjects . . . . .	5
Relative abundances of mouse subject samples	6
Expanded clones	8
Mapping of integration site positions	9
Mouse transplant trials	10
References	18
Supplementary tables and figures	19
Numbers of inferred cells and integration sites identified in provided samples . . . . .	19

## 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 mouse subjects and assess potential clonal expansions. 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 where the mouse oncogene list comprises 2.5% of all mouse genes. The frequency of integration 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>2</sup>. The code base for this analysis is available online ([link](#)).

# Mouse samples studied

Integration sites were detected in 34 samples from mouse subjects (Tables 1 & S1).

Table 1. Overview of data collection.

Organism	Number of samples	Number of reads	Number of inferred cells	Number of integration sites
mouse	34	14,339,078	256,987	8,391

## 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 this ([link](#)). 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>4</sup>.

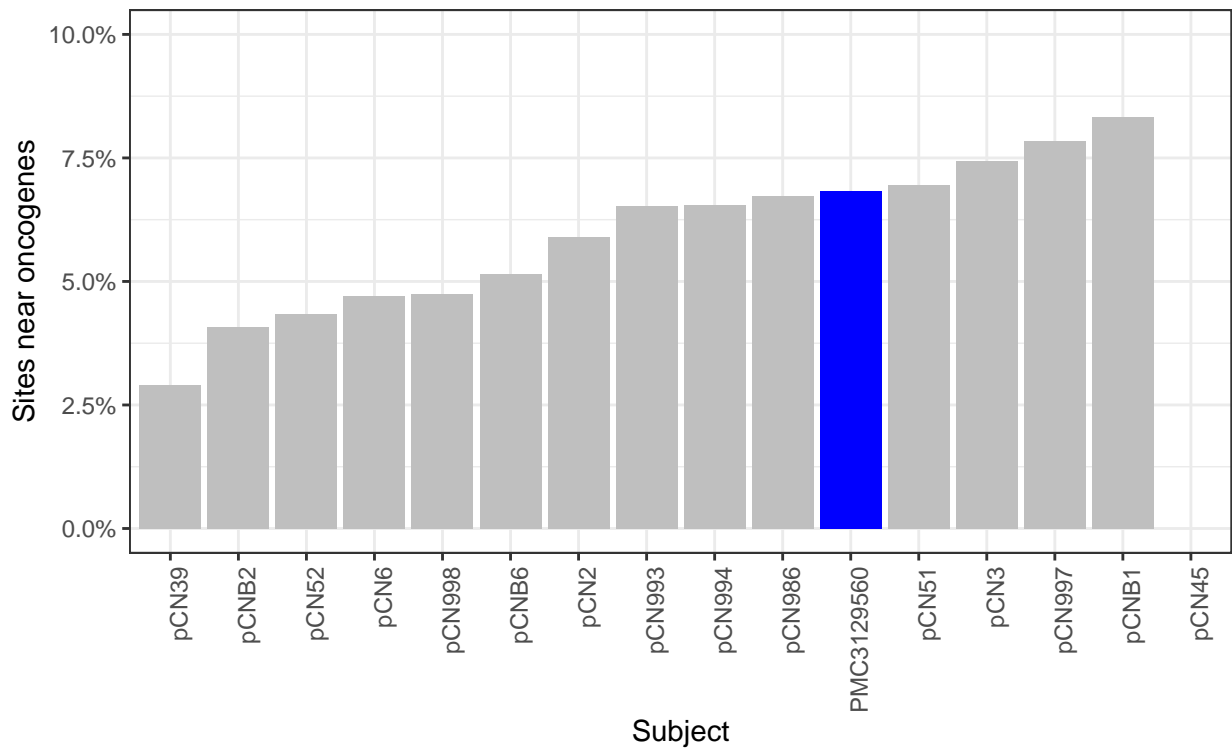
# Comparisons to previous trials

## 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>2</sup> which used a comparable lentiviral vector to correct  $\beta$ -thalassemia. The frequency of integration events near onco genes in *all transfected mice* was less than the mean frequency of integration events near oncogenes in the published trial (Figure 1 [experimental subjects: gray, previous  $\beta$ -thalassemia trial: blue]).

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

## Warning: Removed 1 rows containing missing values (position\_stack).



# Relative abundances of mouse subject samples

The sample relative abundance plots below (Figure 2) show the most abundant 25 clones in each sample as colored bars while less abundant clones were relegated to a single low abundance bar shown in gray.

Figure 2.

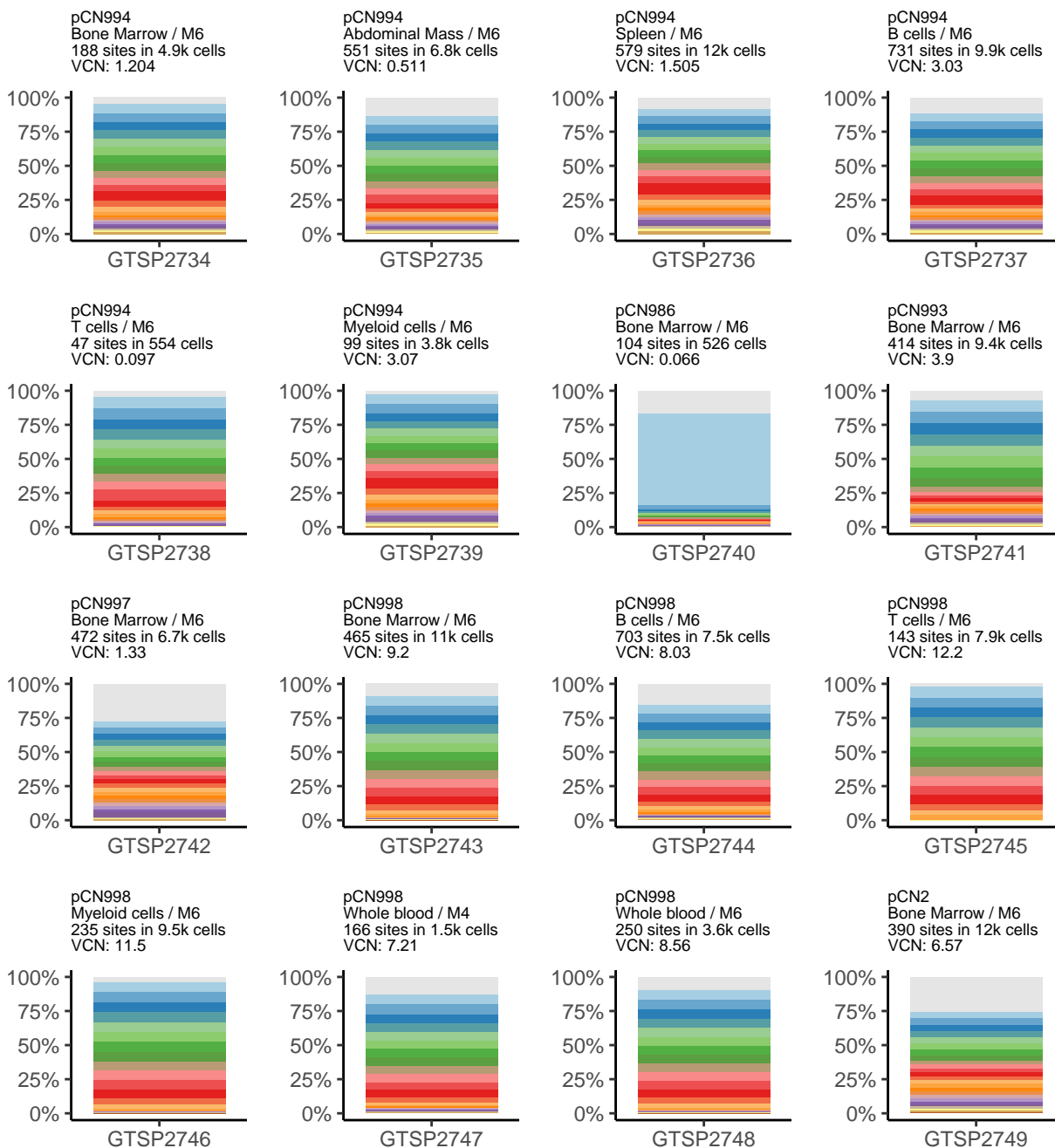
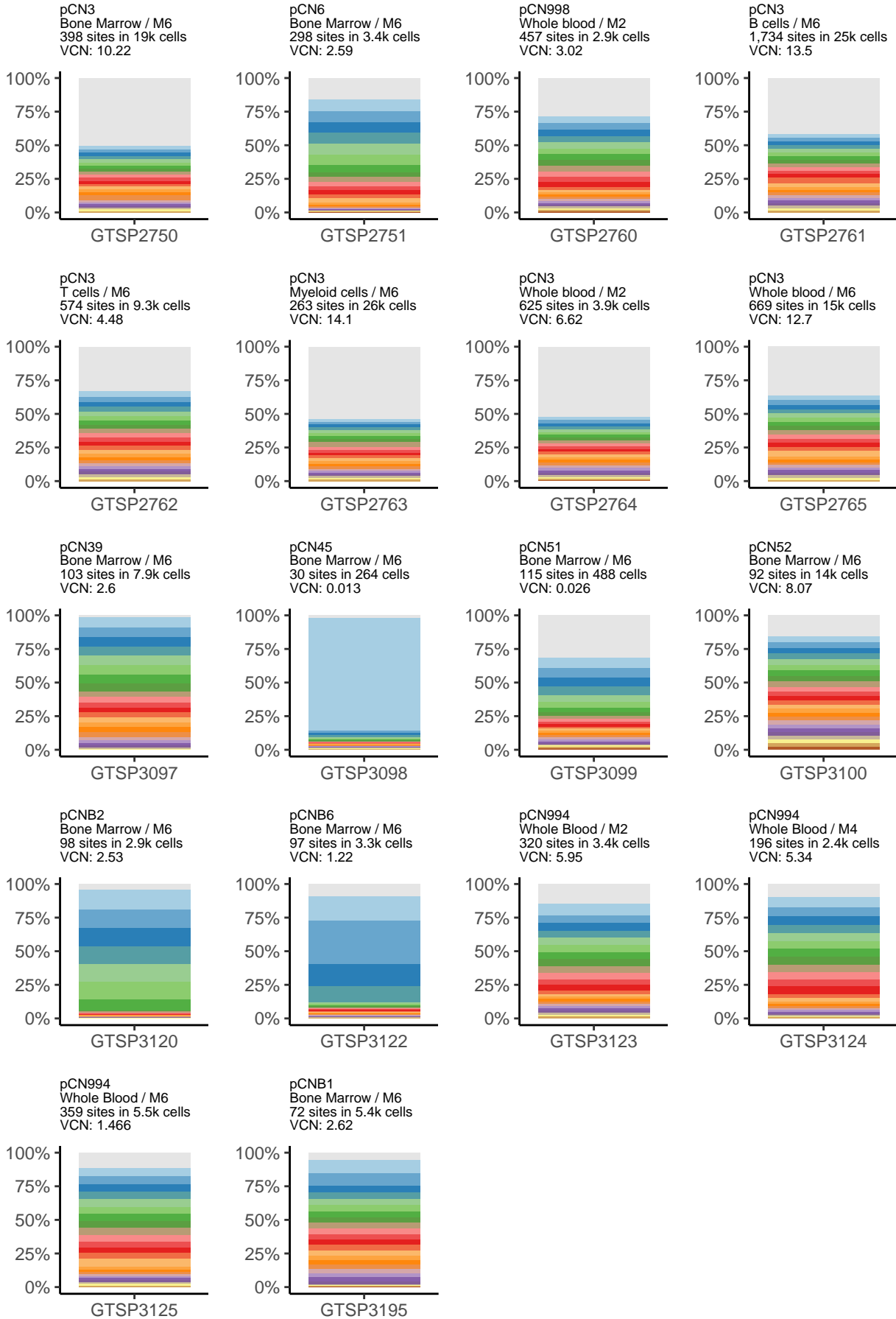


Figure 2 (continued).



# 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.

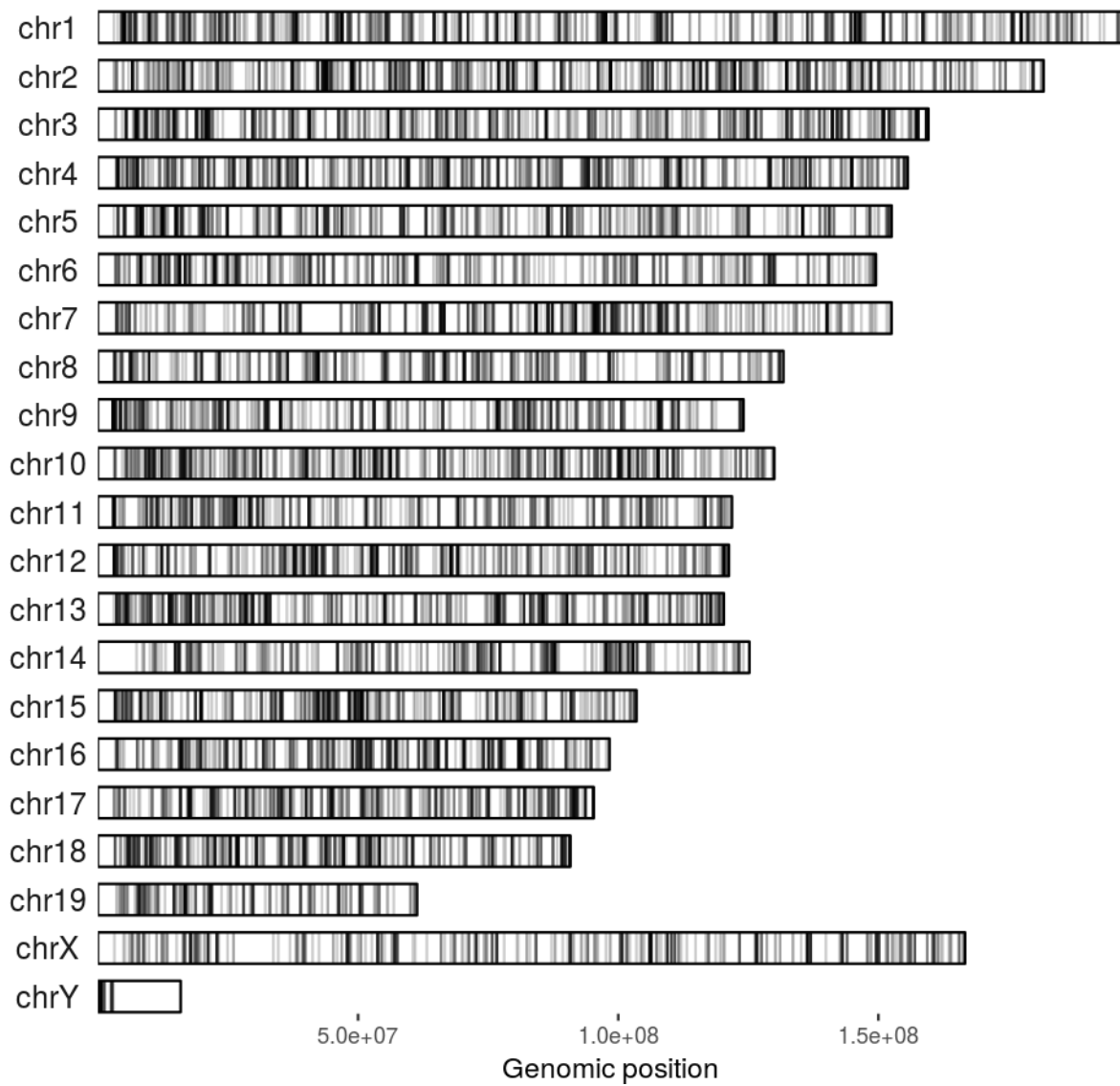
Subject	Organism	Time point	Cell type	Position	Relative abundance	Abundance	Nearest gene
pCN45	mouse	M6	Bone Marrow	chr9+66207164	83.71%	221	Herc1
pCN986	mouse	M6	Bone Marrow	chr9+66207164	66.92%	352	Herc1



# Mapping of integration site positions

Integration events were observed across all mouse subject chromosomes (Figure 3).

Figure 3.



# Mouse transplant trials

The positions of identified integration sites from cell transplant trials with nine pairs of mice are shown in Figure 4a (donor mice) and Figure 4b (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 5 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. Instances where no integration sites were identified in the recipient mouse are listed as 'NA'. The clonal abundances of clones found in both donor and recipient mice is shown in Table S2. 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 4B & S3).

Figure 4.

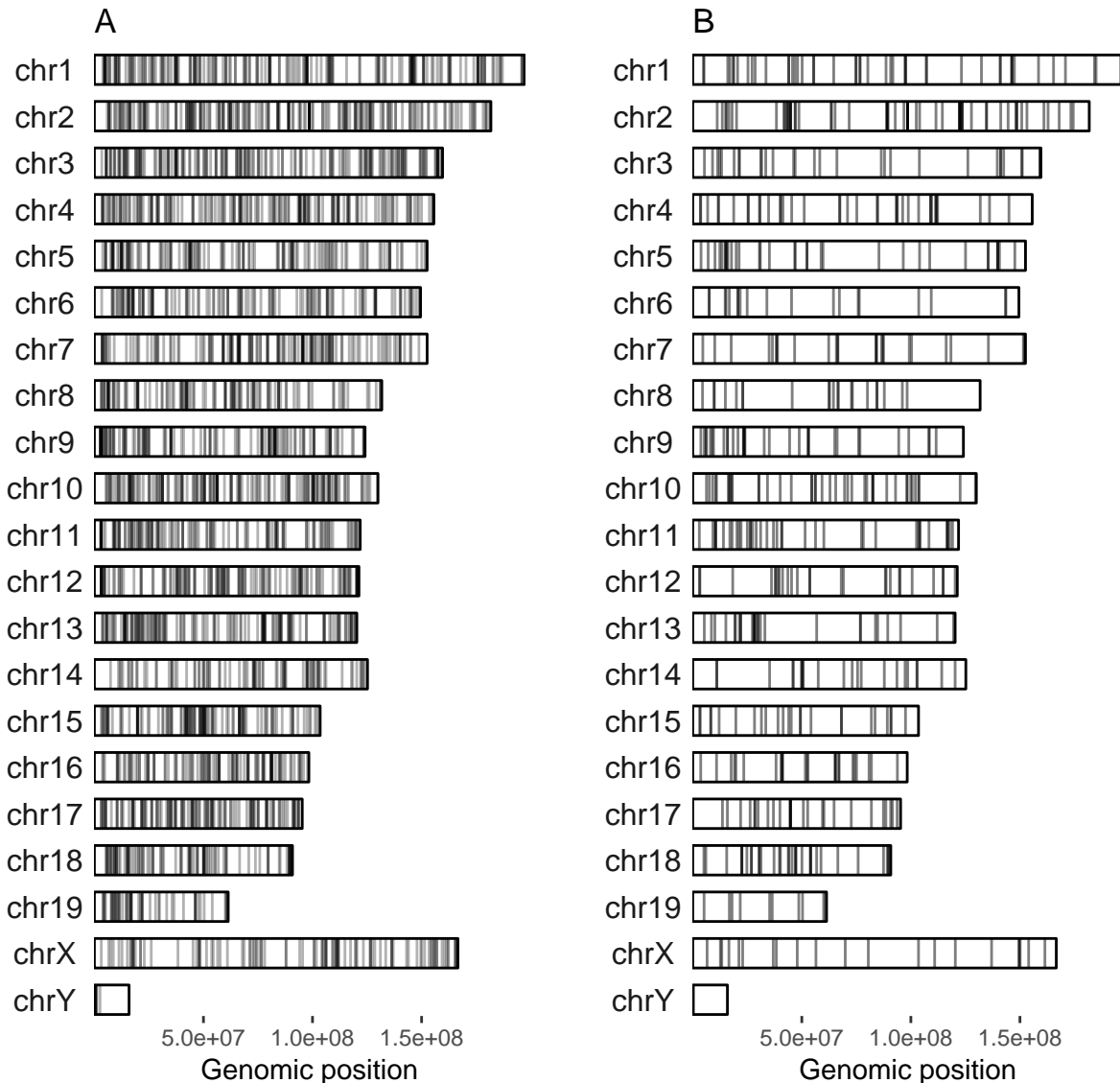
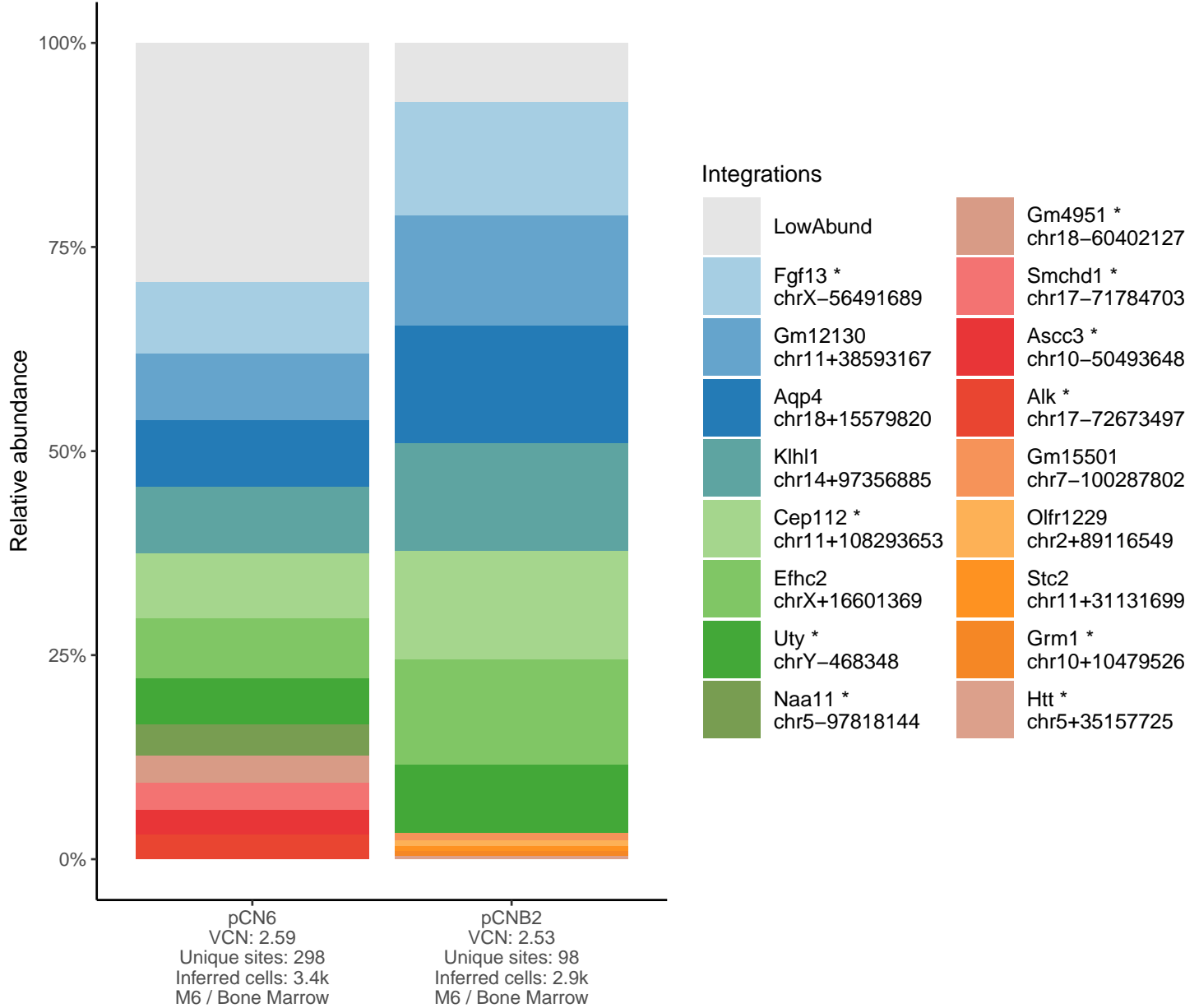


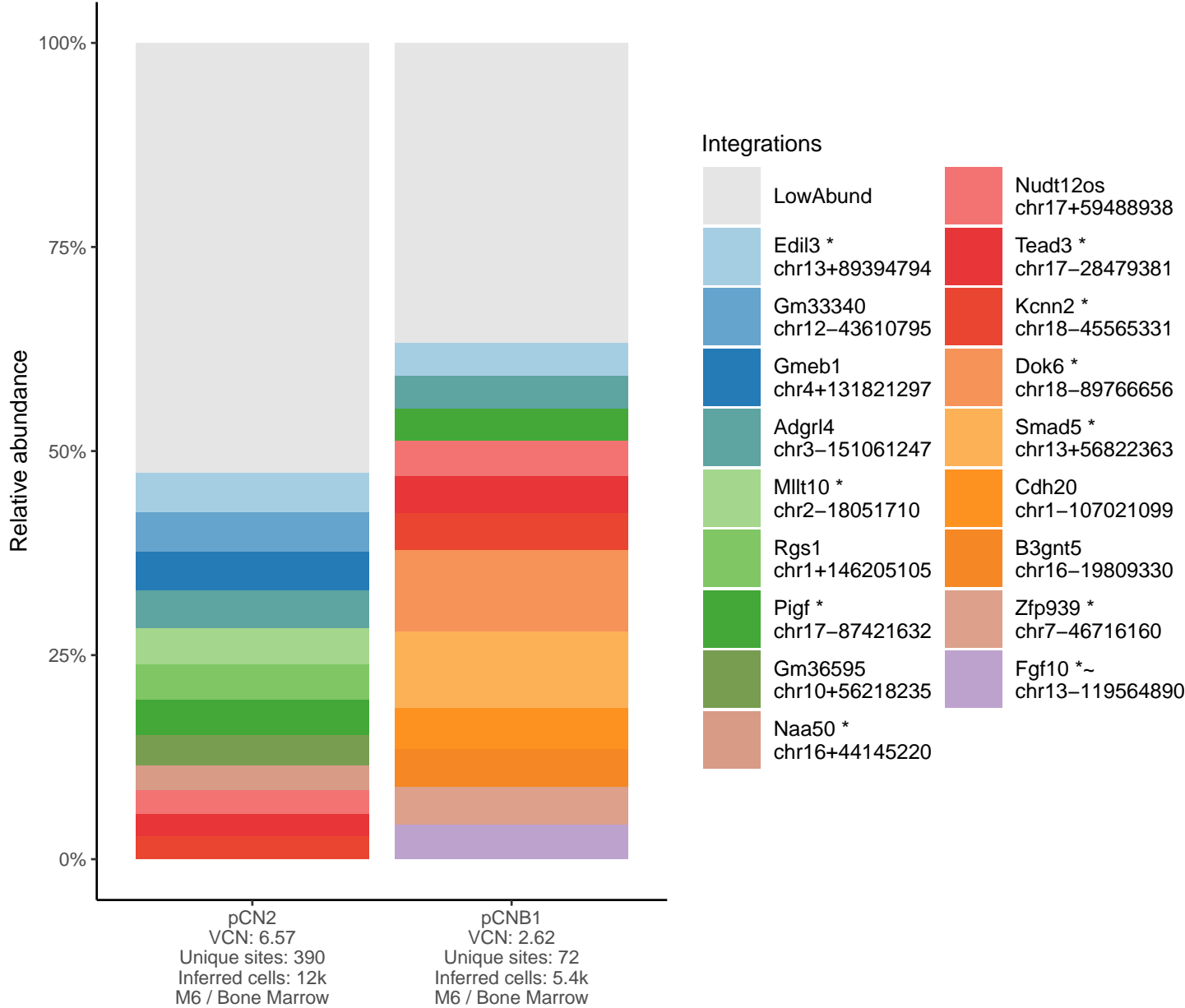
Figure 5a.



Fisher's exact p-value: 1.000

	Not near onco	Near onco
pCN6	284	14
pCNB2	94	4

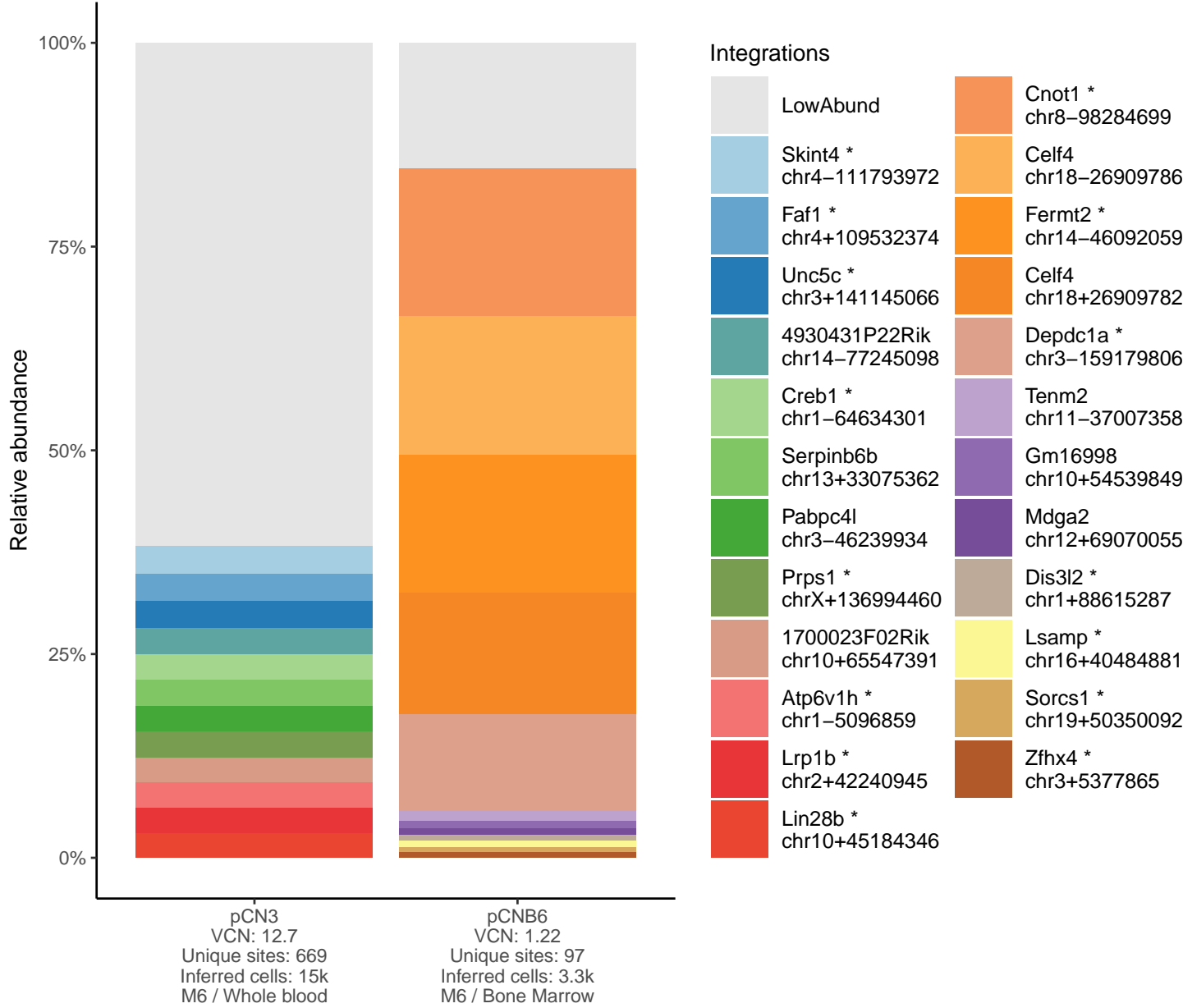
Figure 5b.



Fisher's exact p-value: 0.429

	Not near onco	Near onco
pCN2	367	23
pCNB1	66	6

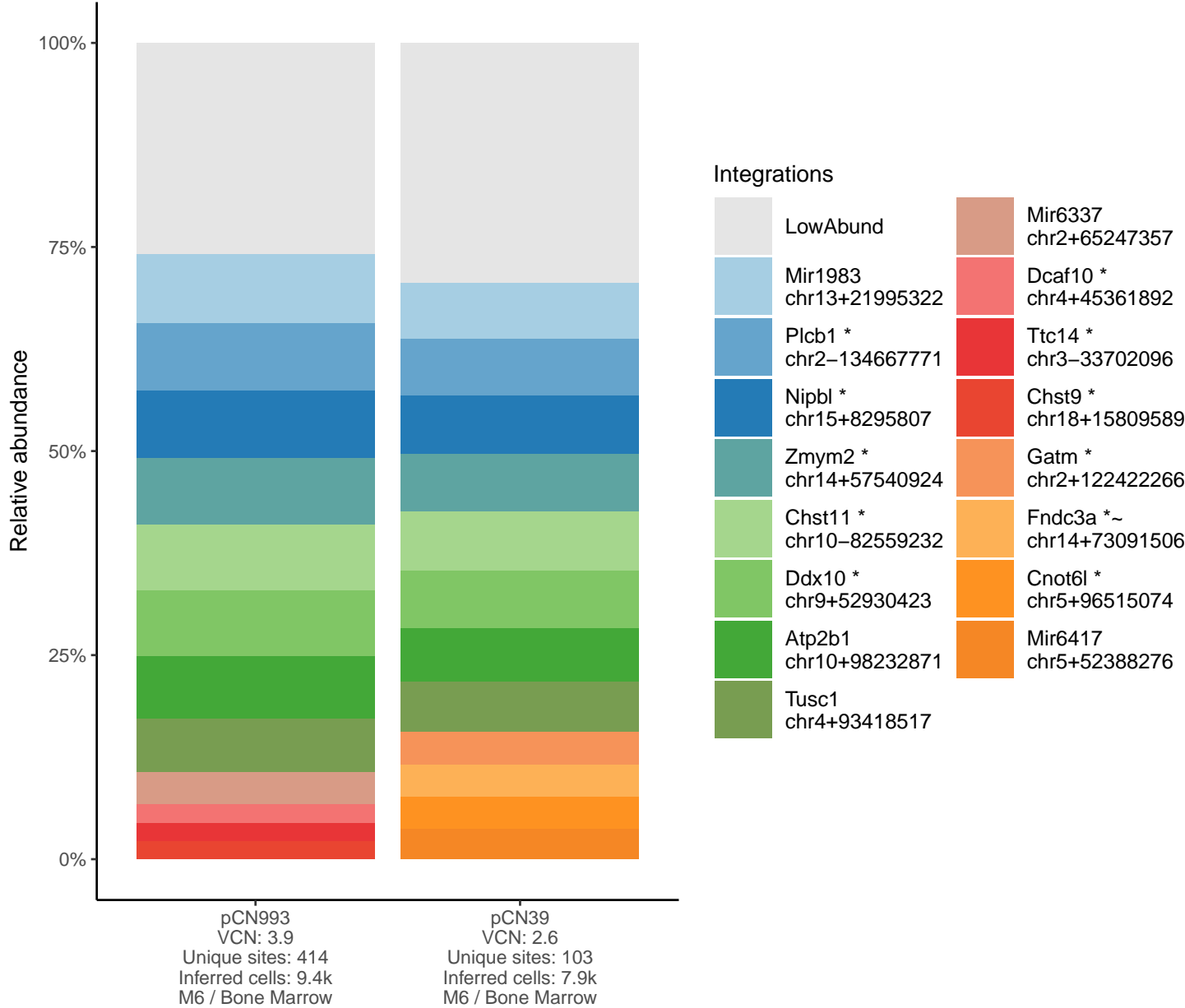
Figure 5c.



Fisher's exact p-value: 0.249

	Not near onco	Near onco
pCN3	606	63
pCNB6	92	5

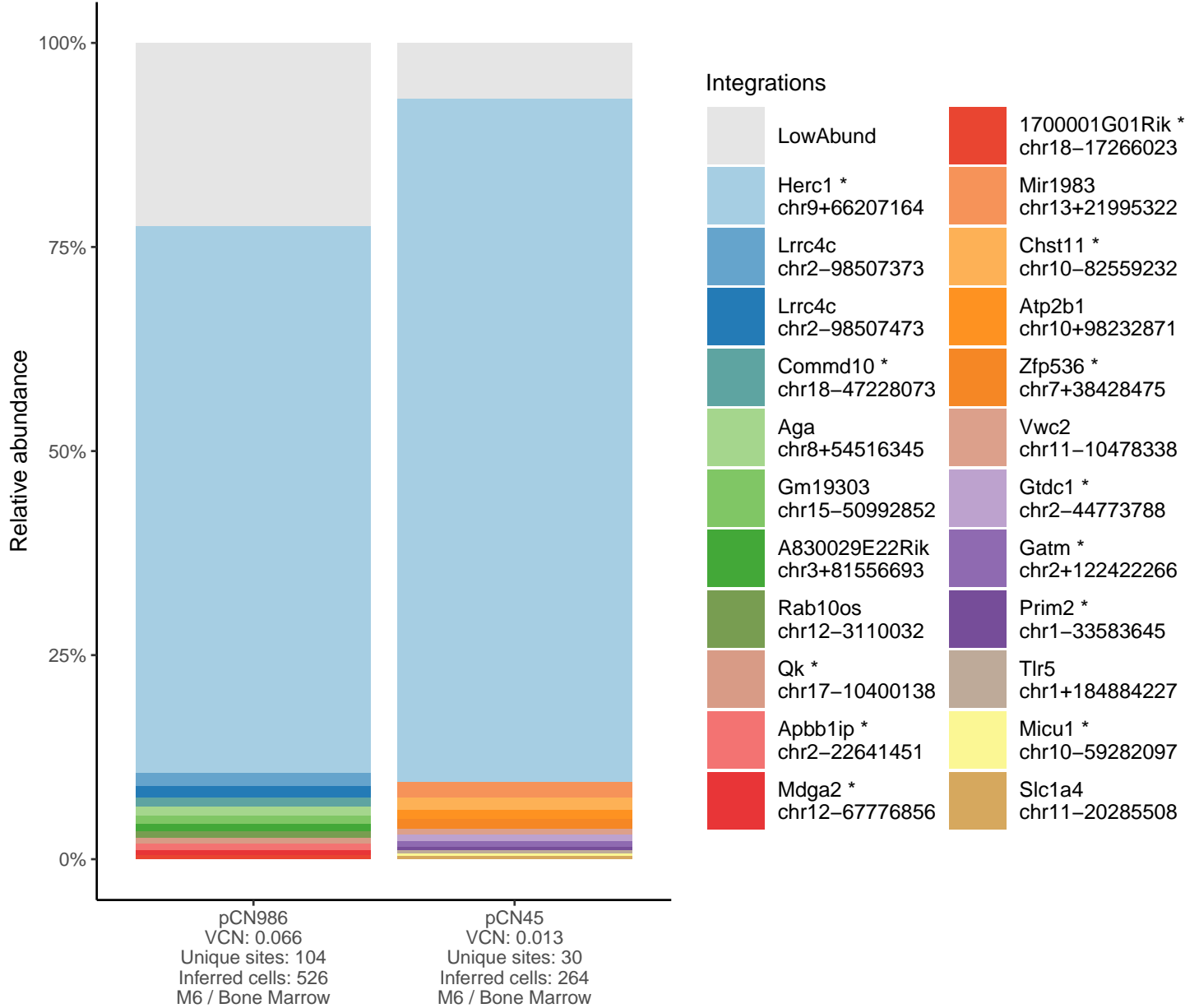
Figure 5d.



Fisher's exact p-value: 0.237

	Not near onco	Near onco
pCN993	387	27
pCN39	100	3

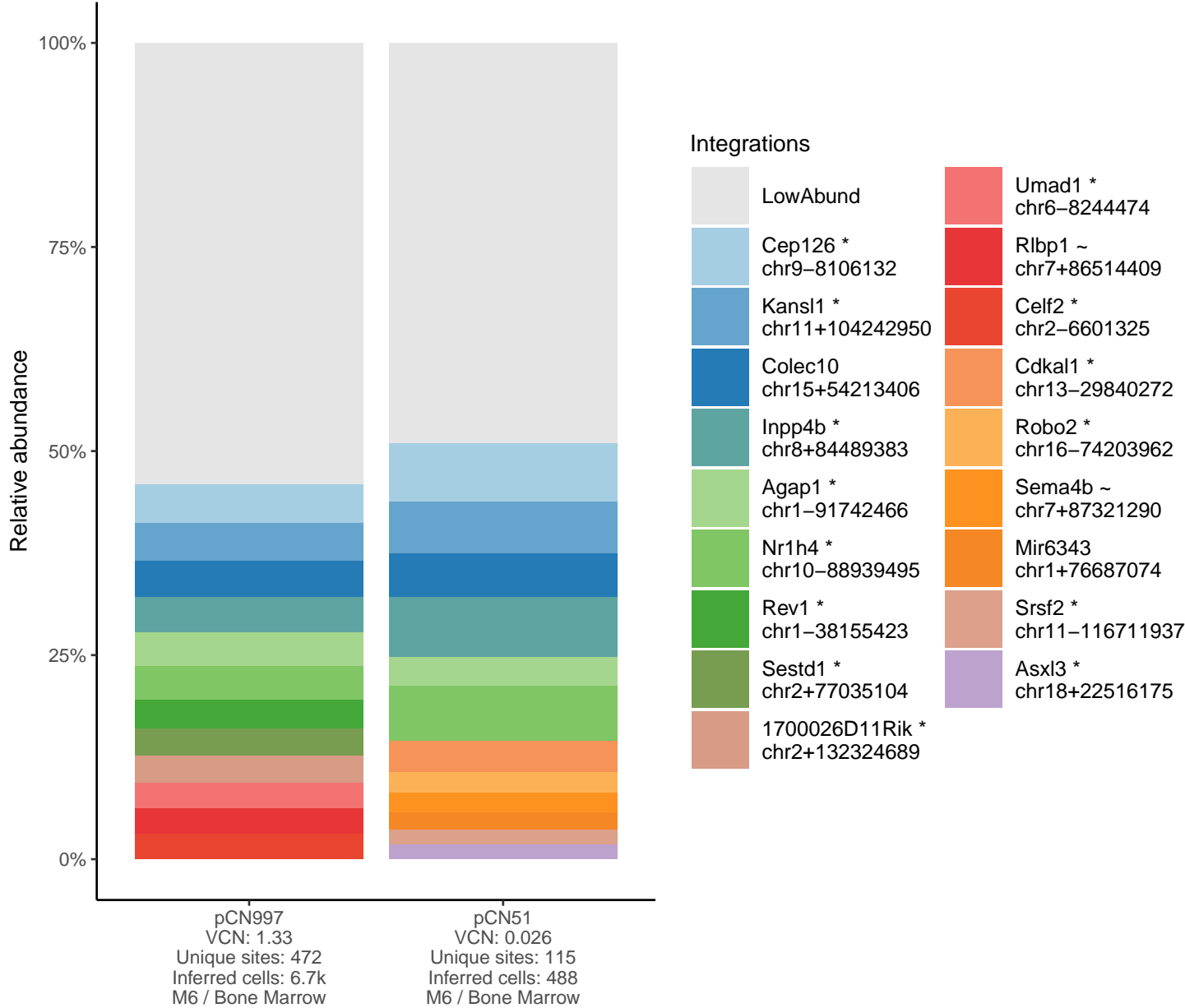
Figure 5e.



Fisher's exact p-value: 0.264

	Not near onco	Near onco
pCN986	97	7
pCN45	26	4

Figure 5f.

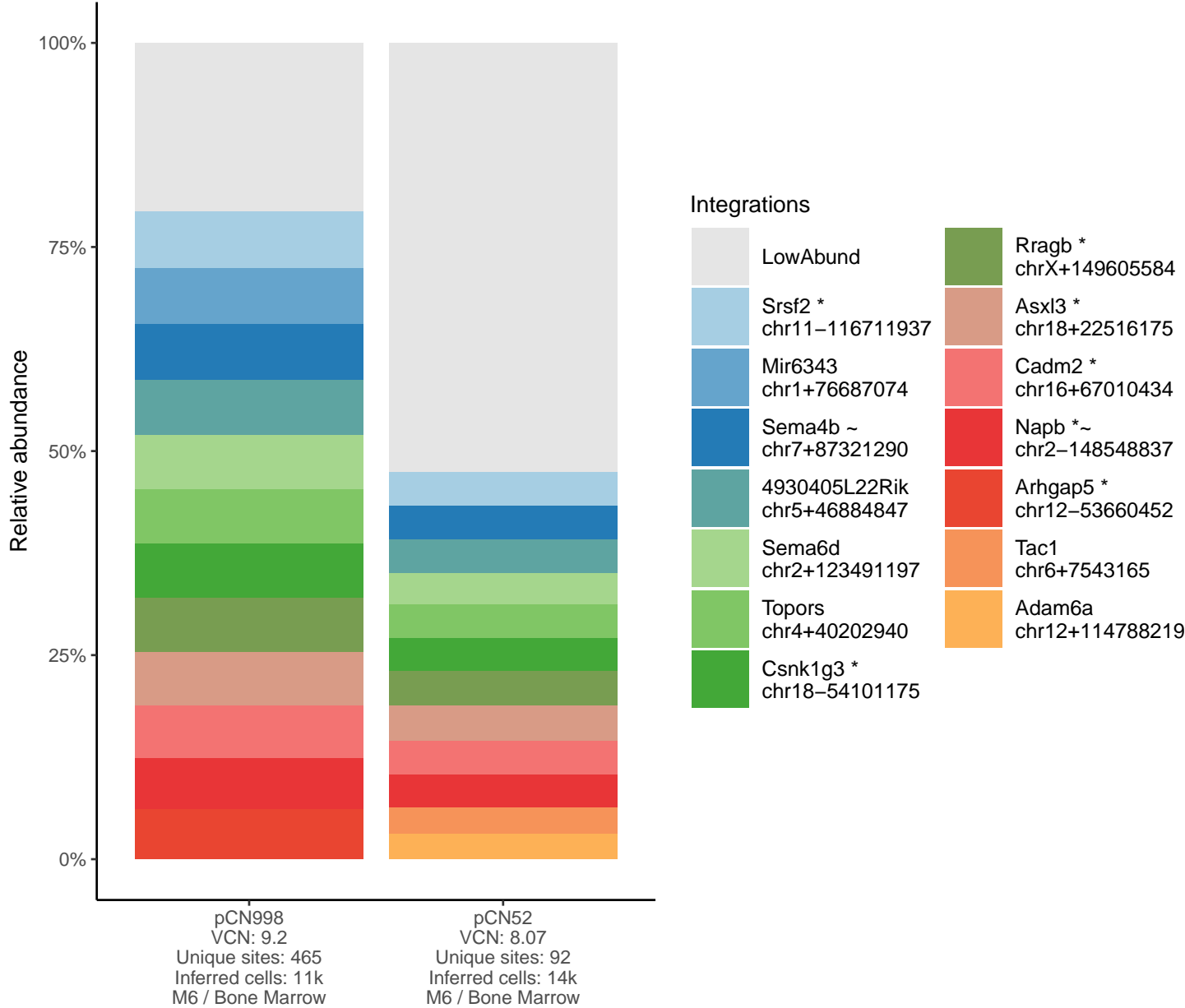


Fisher's exact p-value: 0.847

	Not near onco	Near onco
pCN997	435	37
pCN51	107	8



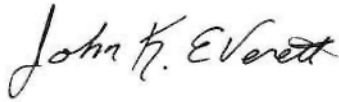
Figure 5g.



Fisher's exact p-value: 0.535

	Not near onco	Near onco
pCN998	450	15
pCN52	88	4

**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. 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.
3. 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.
4. 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
mouse	GTSP2734	pCN994	Bone Marrow	1.204	M6	4,881	188
mouse	GTSP2735	pCN994	Abdominal Mass	0.511	M6	6,838	551
mouse	GTSP2736	pCN994	Spleen	1.505	M6	12,114	579
mouse	GTSP2737	pCN994	B cells	3.030	M6	9,936	731
mouse	GTSP2738	pCN994	T cells	0.097	M6	554	47
mouse	GTSP2739	pCN994	Myeloid cells	3.070	M6	3,848	99
mouse	GTSP2740	pCN986	Bone Marrow	0.066	M6	526	104
mouse	GTSP2741	pCN993	Bone Marrow	3.900	M6	9,398	414
mouse	GTSP2742	pCN997	Bone Marrow	1.330	M6	6,697	472
mouse	GTSP2743	pCN998	Bone Marrow	9.200	M6	11,044	465
mouse	GTSP2744	pCN998	B cells	8.030	M6	7,501	703
mouse	GTSP2745	pCN998	T cells	12.200	M6	7,919	143
mouse	GTSP2746	pCN998	Myeloid cells	11.500	M6	9,486	235
mouse	GTSP2747	pCN998	Whole blood	7.210	M4	1,458	166
mouse	GTSP2748	pCN998	Whole blood	8.560	M6	3,558	250
mouse	GTSP2749	pCN2	Bone Marrow	6.570	M6	12,164	390
mouse	GTSP2750	pCN3	Bone Marrow	10.220	M6	18,701	398
mouse	GTSP2751	pCN6	Bone Marrow	2.590	M6	3,436	298
mouse	GTSP2760	pCN998	Whole blood	3.020	M2	2,931	457
mouse	GTSP2761	pCN3	B cells	13.500	M6	24,687	1,734
mouse	GTSP2762	pCN3	T cells	4.480	M6	9,251	574
mouse	GTSP2763	pCN3	Myeloid cells	14.100	M6	25,816	263
mouse	GTSP2764	pCN3	Whole blood	6.620	M2	3,863	625
mouse	GTSP2765	pCN3	Whole blood	12.700	M6	15,147	669
mouse	GTSP3097	pCN39	Bone Marrow	2.600	M6	7,867	103
mouse	GTSP3098	pCN45	Bone Marrow	0.013	M6	264	30
mouse	GTSP3099	pCN51	Bone Marrow	0.026	M6	488	115
mouse	GTSP3100	pCN52	Bone Marrow	8.070	M6	13,839	92
mouse	GTSP3120	pCNB2	Bone Marrow	2.530	M6	2,876	98
mouse	GTSP3122	pCNB6	Bone Marrow	1.220	M6	3,291	97
mouse	GTSP3123	pCN994	Whole Blood	5.950	M2	3,394	320
mouse	GTSP3124	pCN994	Whole Blood	5.340	M4	2,364	196
mouse	GTSP3125	pCN994	Whole Blood	1.466	M6	5,493	359
mouse	GTSP3195	pCNB1	Bone Marrow	2.620	M6	5,357	72