Analysis of integration site distributions and clonal abundances for gene therapy correction of cystinosis

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

October 2018

Contents

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)[1] using an inclusion threshold of three or more incidents where the mouse oncogene list comprises 2.01% of all mouse genes. The frequency of integration near oncogenes was generally less than that of mice in a previously published β-thalassemia mouse trial from which no adverse events have been reported [2]. The code base for this analysis is available online (link).

Twenty-nine mice were tested by serial transfer of vector-treated bone marrow. From these, 64,413 cells were sampled, yielding 3,647 total unique integration sites. In both donor and recipient samples, there was no overall increase in integration sites near annotated murine cancer-associated genes, thus providing no evidence for genotoxicity. Some cells expanded more than others, but analysis of cells showing the most pronounced expansion showed no enrichment in integration sites in or near cancer-associated genes. In one case, decendants of a single cell appears to have mainly repopulated a test mouse (pCN801). This cell appears to have harbored ~12 independent integration events, and proliferated robustly, so that the final specimen showed a VCN of 9.58. Upon transfer to a recipient mouse (pCN951), the cell clone proliferate sluggishly (VCN 0.083), suggesting that despite the number of integration events, no major change in growth properties took place. None of the 12 genes are annotated as cancer associated genes. Specimens of integration into human cells in culture were also compared (VIS UCSD CYS Report.group1). Results showed no favored integration near cancer associated genes compared to previous studies of integration site distributions. Thus the tests of integration site distributions here provide no strong evidence for genotoxicity of the vector analyzed greater than for vectors used safely to date in ongoing human trials.

Mouse samples studied

Integration sites were detected in 29 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	29	16,317,429	64,413	3,647

Subject reports

Subject specific reports for all subjects are available via a protected web archive (link).
user: cherqui
pass: geneTherapy@!#

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, a second track provides the maximum clonal abundance. Entering gene names into the search bar will direct the browser to specific genes.

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[4].

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 INSPIIRED integration site analysis pipeline (v1.2)[4].

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[2] which used a comparable lentiviral vector to correct β-thalassemia. The frequency of integration events near oncogenes in the subjects was less than the mean frequency of integration events near oncogenes in the published trial (Figure 1 [experimental subjects: gray, previous β-thalassemia trial: blue]).

Figure 1. Comparison of frequencies of integration events near oncogenes.
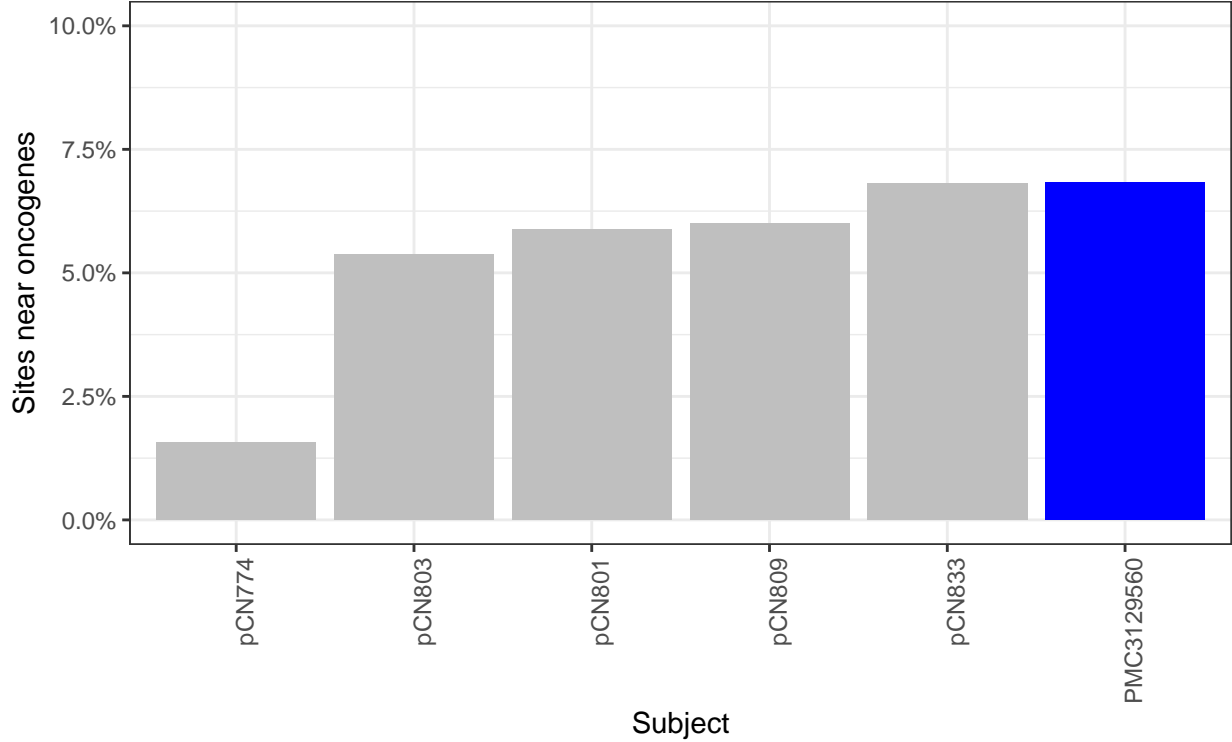
Relative abundances of mouse subject samples

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.
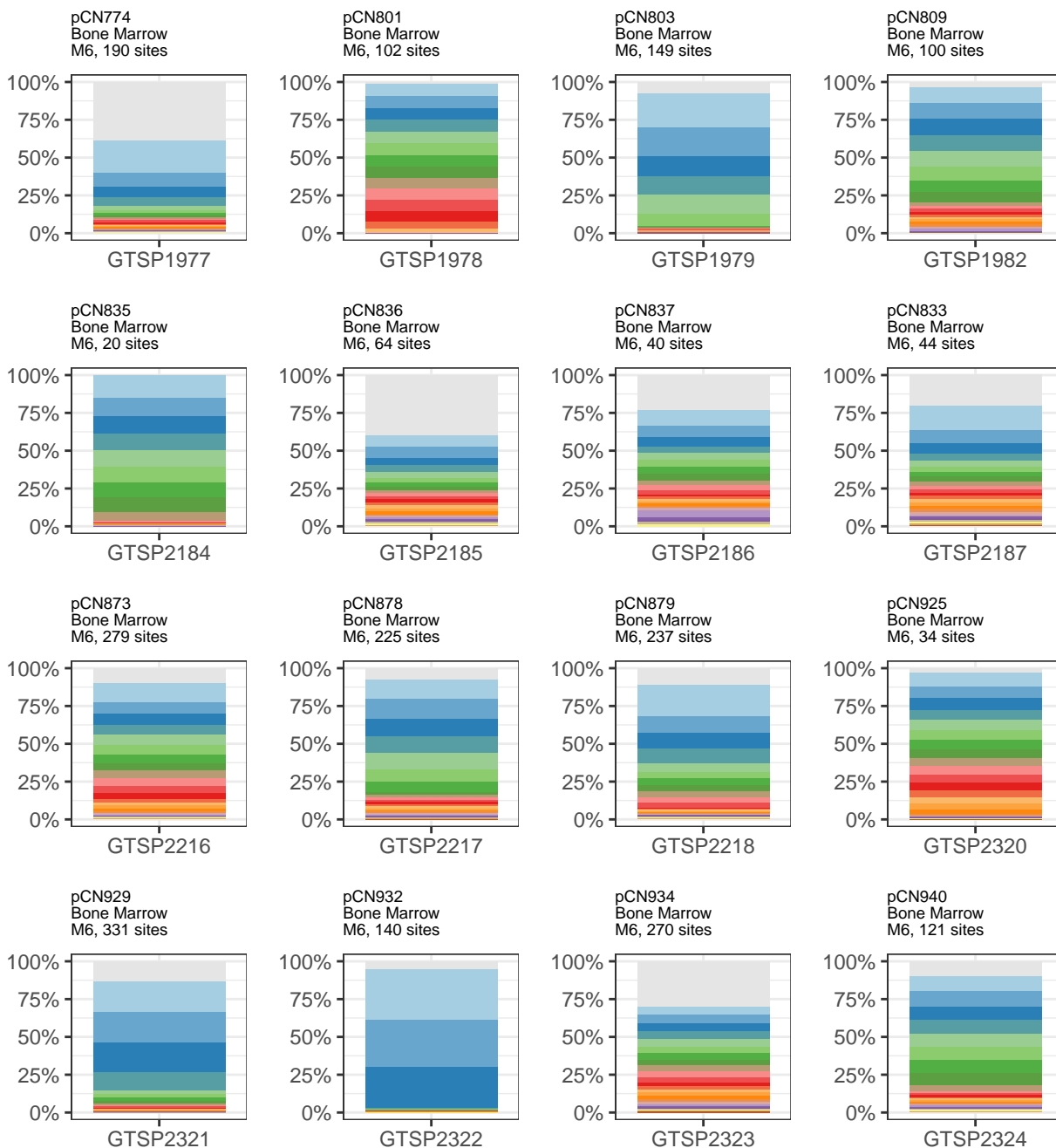
Figure 2.

Figure 2 (continued).

pCN951
Bone Marrow
M6, 32 sites
GTSP2326

pCN956
Bone Marrow
M6, 122 sites
GTSP2328

pCN801
Whole Blood
M2, 145 sites
GTSP2341

pCN801
Whole Blood
M6, 337 sites
GTSP2343

pCN801
BM Myeloid cells
M6, 45 sites
GTSP2344

pCN801
B–cells
M6, 496 sites
GTSP2345

pCN801
T–Cells
M6, 38 sites
GTSP2346

pCN948
Hematoma
M6, 207 sites
GTSP2347

pCN948
B–cells
M6, 258 sites
GTSP2349

pCN951
BM Myeloid cells
M6, 35 sites
GTSP2350

pCN951
B–cells
M6, 24 sites
GTSP2351

pCN951
T–Cells
M6, 35 sites
GTSP2352

pCN967
Bone Marrow
M6, 24 sites
GTSP2353
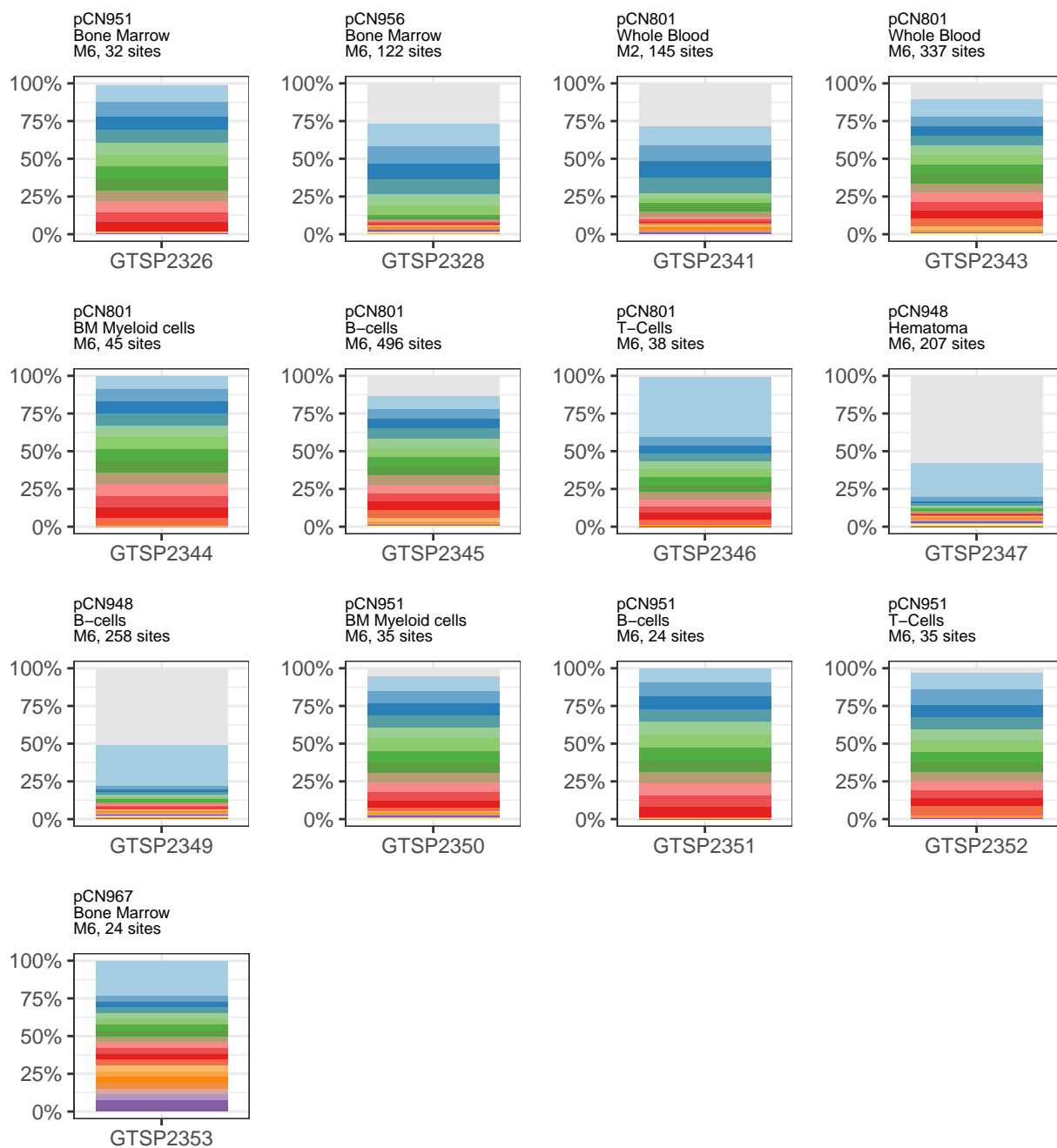
Expanded clones

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
pCN774	mouse	M6	Bone Marrow	chr1-21011655	21.02%	99	Tram2
pCN803	mouse	M6	Bone Marrow	chr3+60155698	22.64%	461	Mbnl1
pCN879	mouse	M6	Bone Marrow	chr15+44227316	20.85%	829	Nudcd1
pCN929	mouse	M6	Bone Marrow	chr9+100734703	20.67%	669	Stag1
pCN929	mouse	M6	Bone Marrow	chrX-126866229	20.02%	648	Diaph2
pCN932	mouse	M6	Bone Marrow	chr12+80277915	33.13%	807	Rdh11
pCN932	mouse	M6	Bone Marrow	chr8-66234933	31.53%	768	Sgo2b
pCN932	mouse	M6	Bone Marrow	chr8+9405148	26.56%	647	Fam155a
pCN801	mouse	M6	T-Cells	chr5-90109547	39.80%	780	Adamts3
pCN948	mouse	M6	Hematoma	chr1-21011655	22.75%	76	Tram2
pCN948	mouse	M6	B-cells	chr1-21011655	27.19%	143	Tram2

Mapping of integration site positions

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

Figure 3
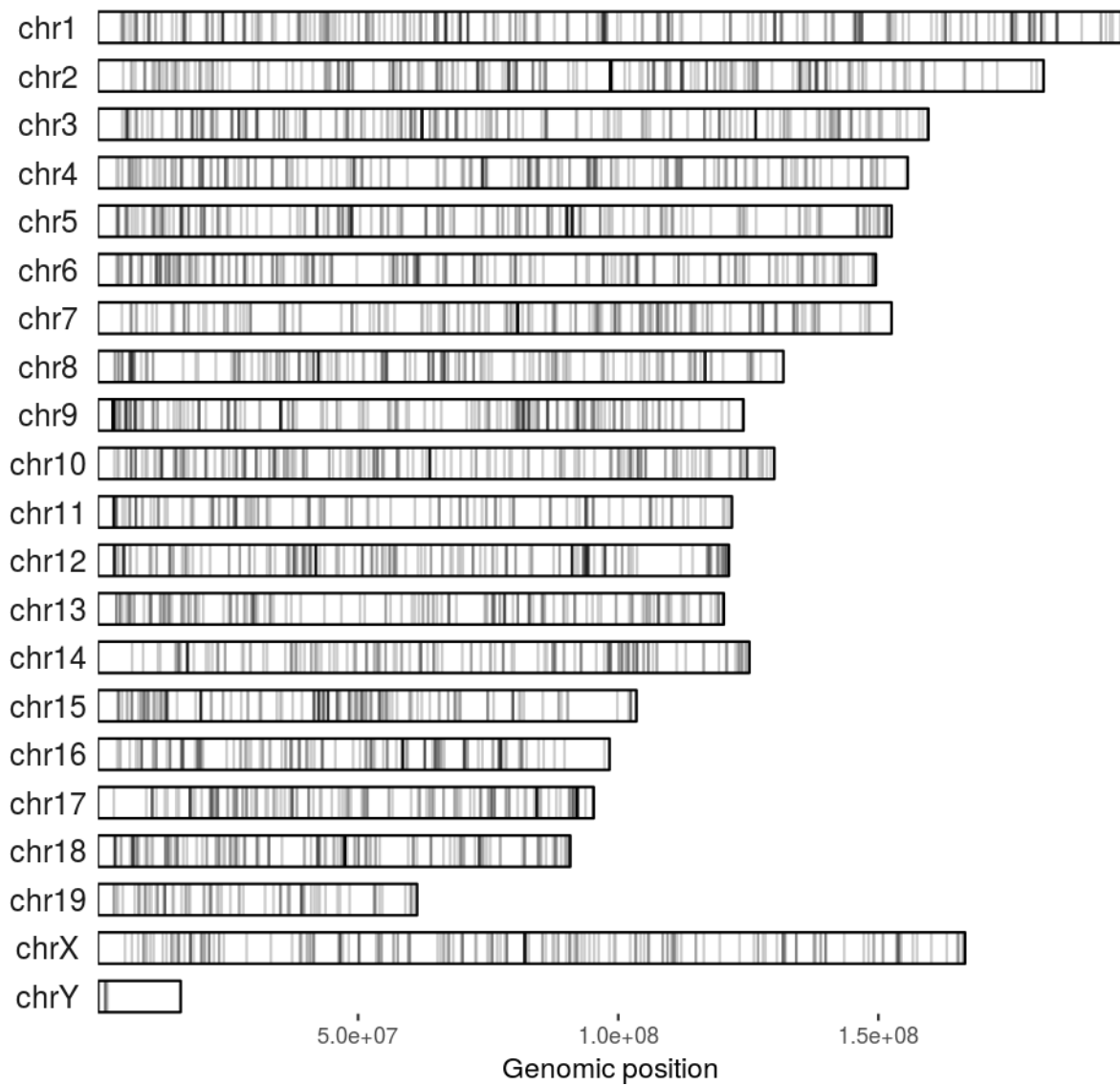
Mouse transplant trials

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 tilda (~) 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).
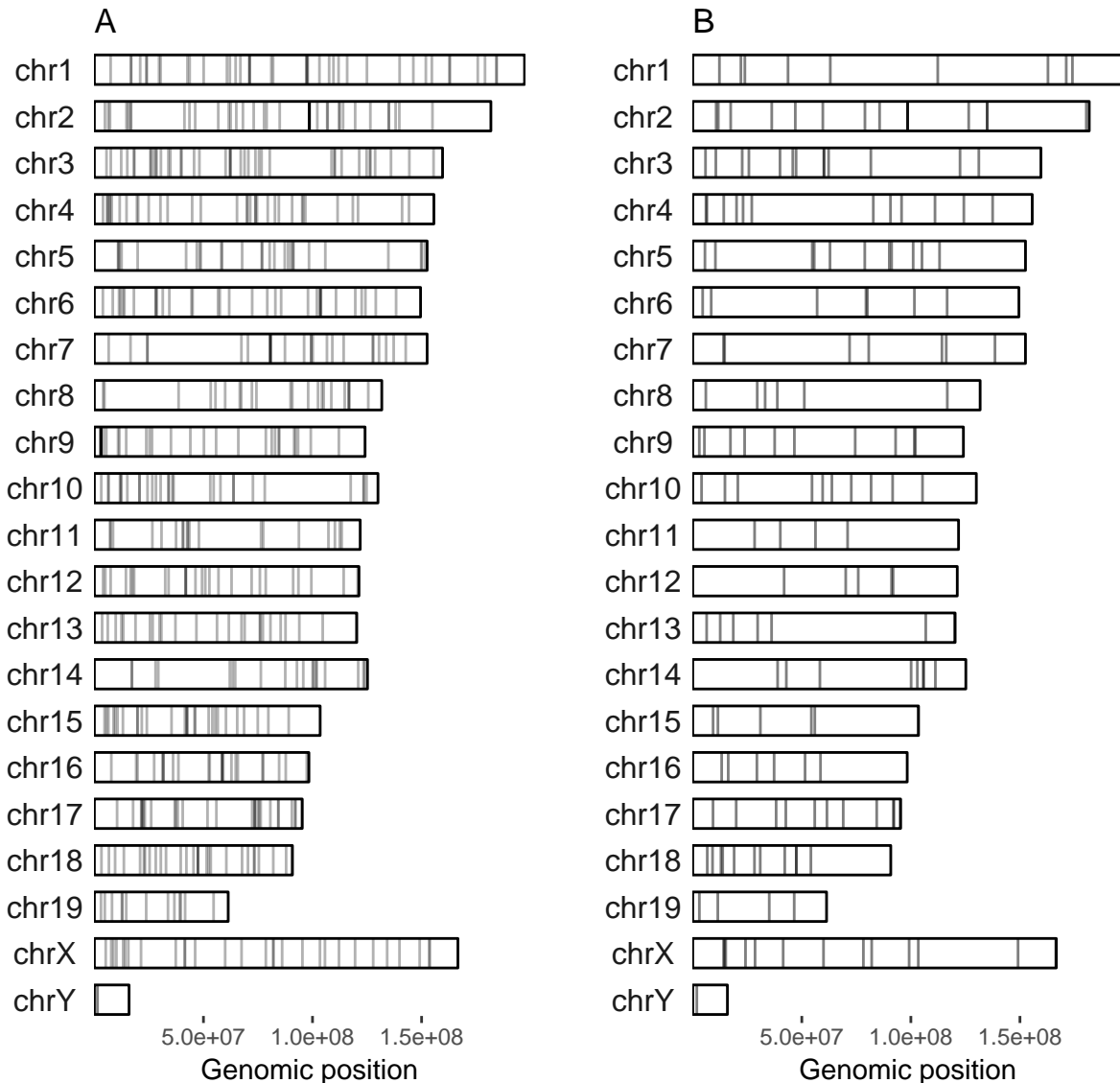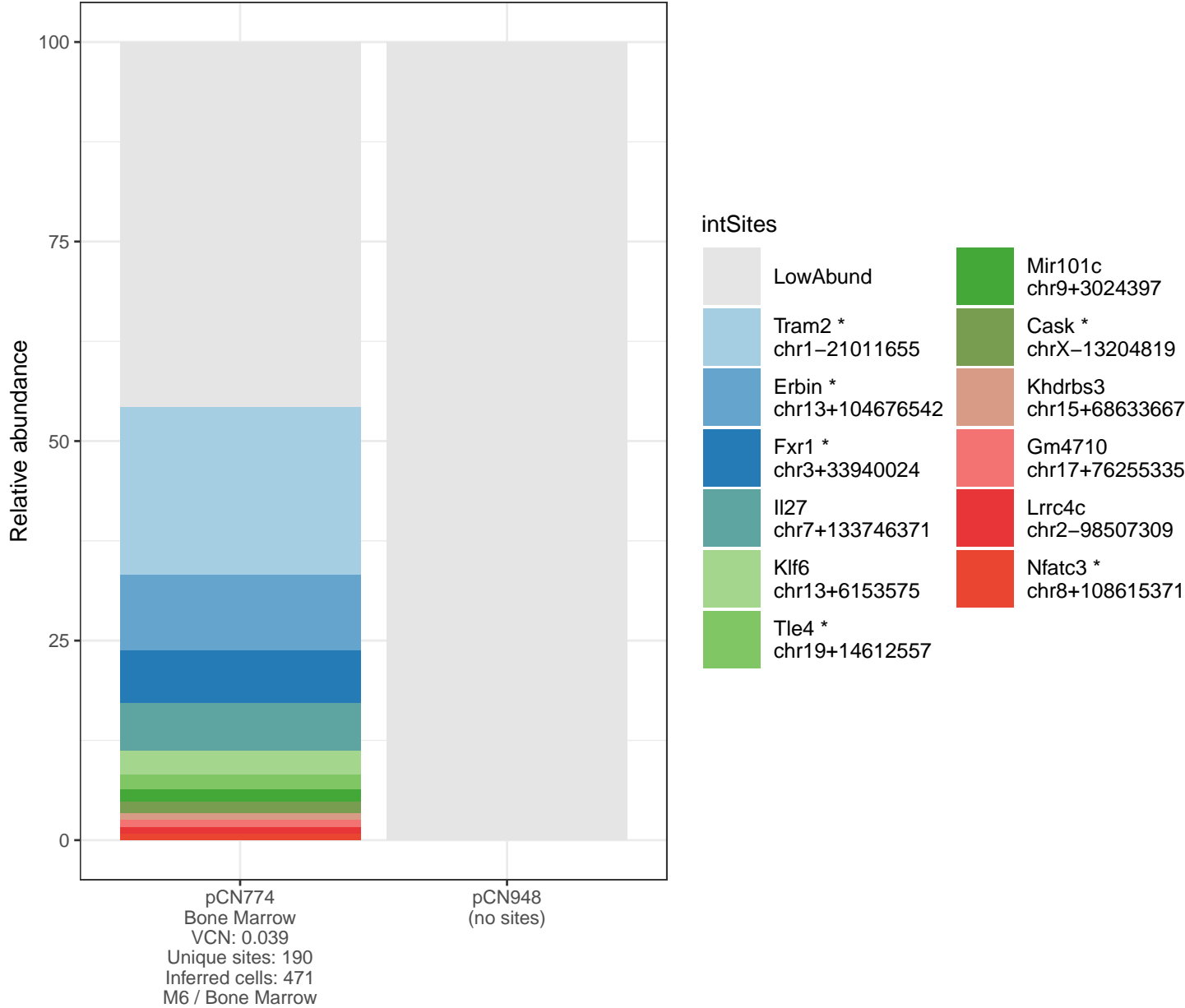
Figure 4

Figure 5a.

Fisher's exact p-value: 1.000

	Not near onco	Near onco
pCN774	187	3
NA	0	0

Figure 5b.
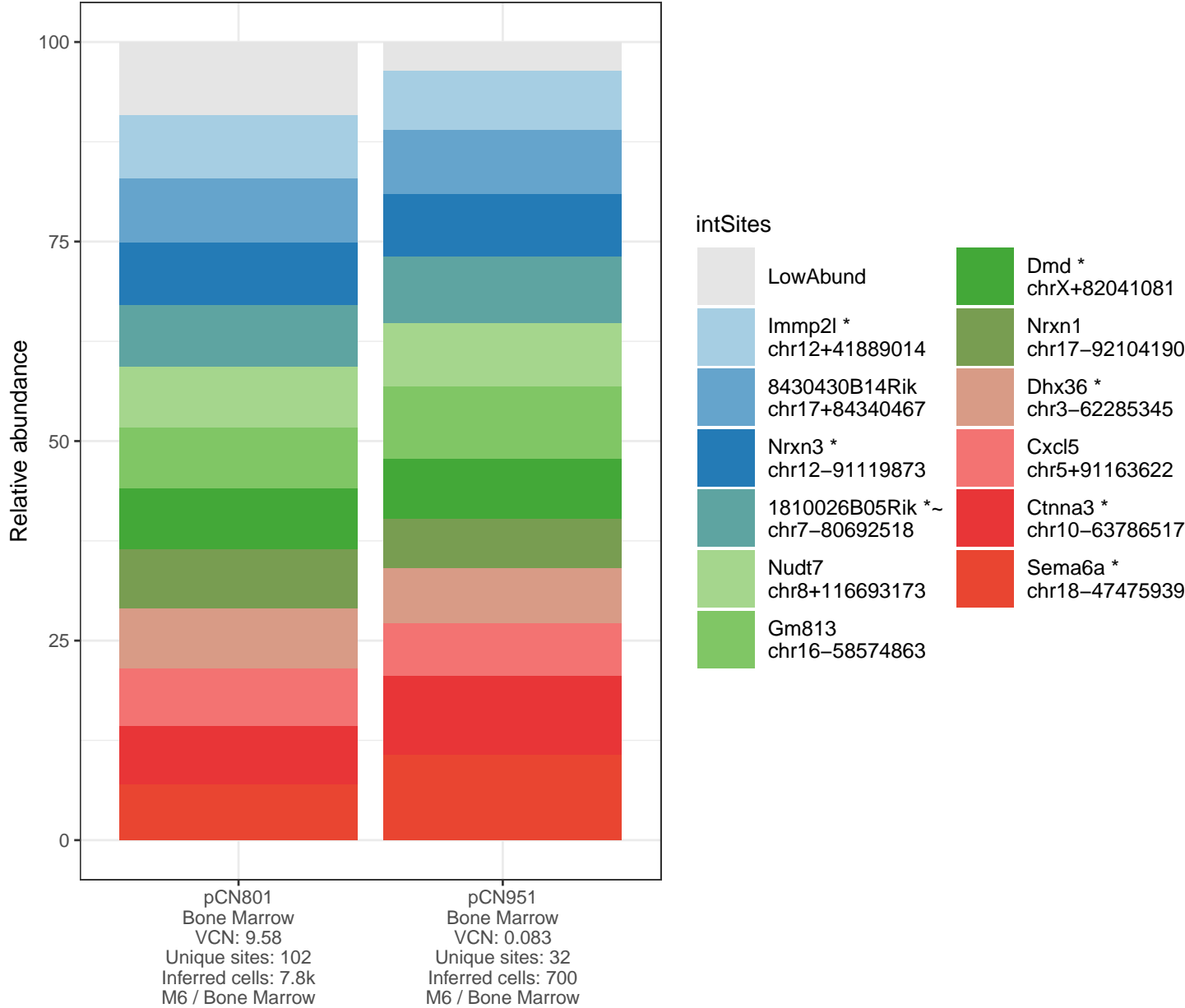
Fisher's exact p-value: 1.000

	Not near onco	Near onco
pCN801	96	6
pCN951	31	1

• Integration at chr7-80692518, near 1810026B05Rik, is 6 KB upstream of oncogene Chd2.

Figure 5c.
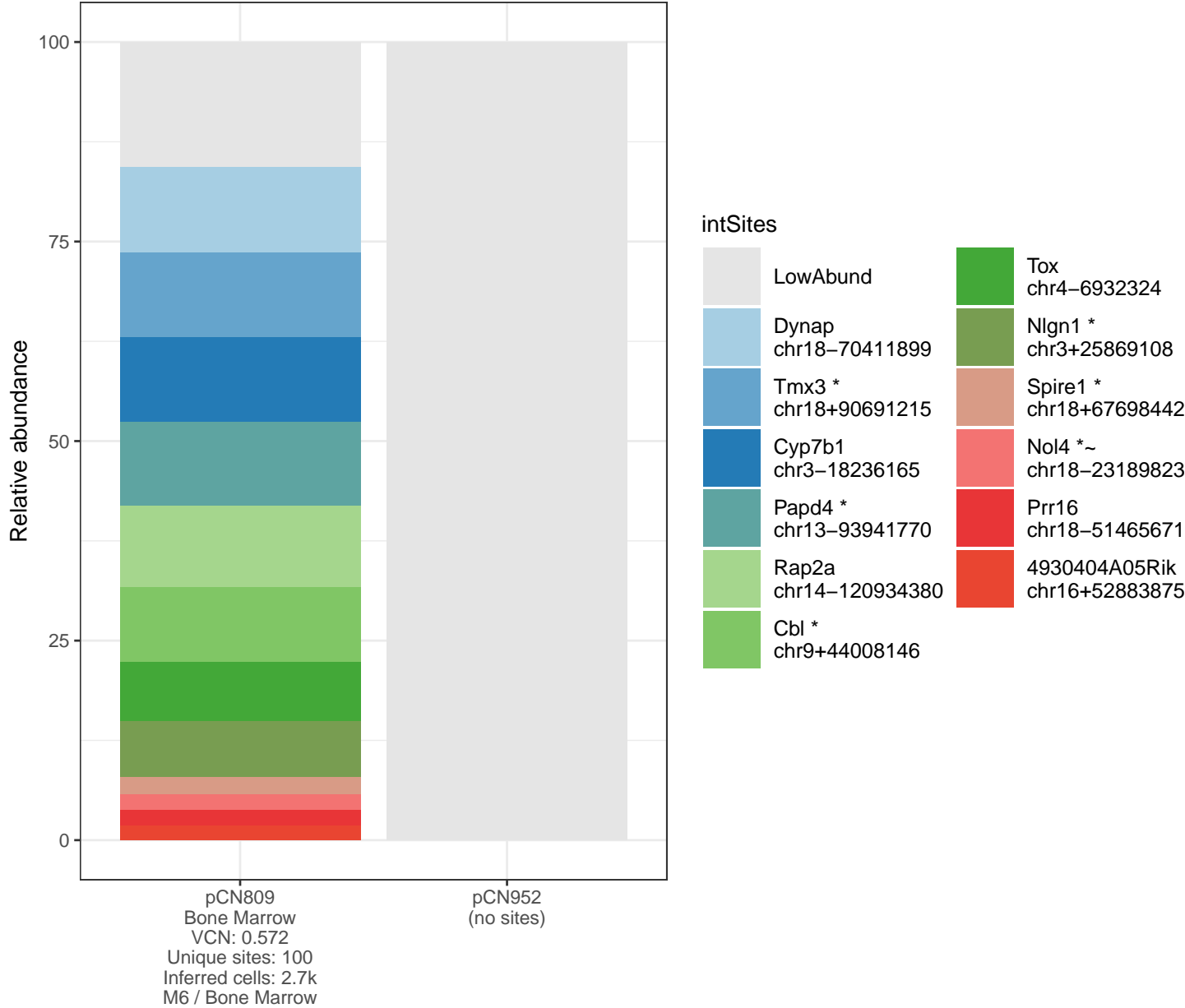
Fisher's exact p-value: 1.000

	Not near onco	Near onco
pCN809	94	6
NA	0	0

Figure 5d.
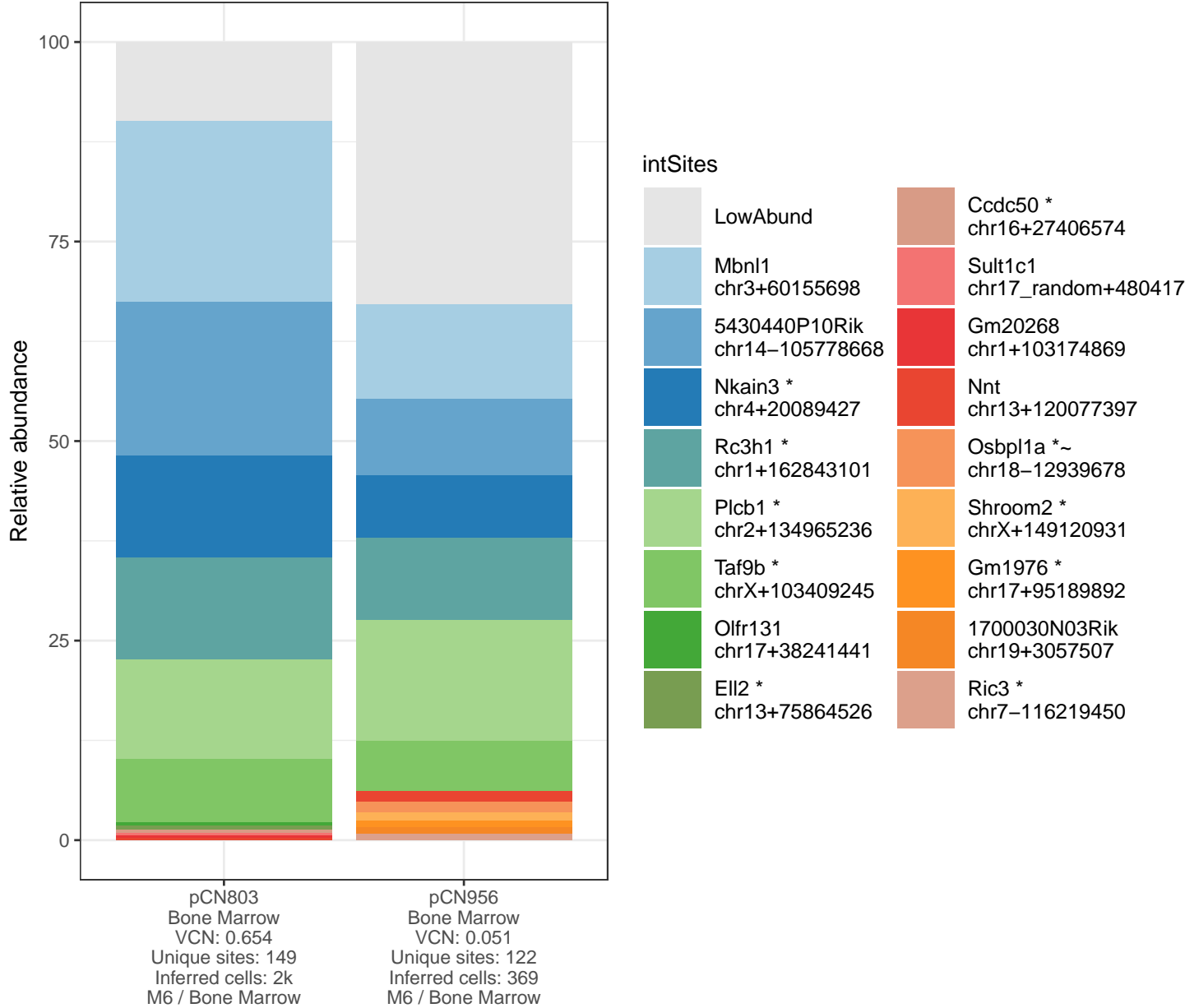
Fisher's exact p-value: 0.355

	Not near onco	Near onco
pCN803	141	8
pCN956	119	3
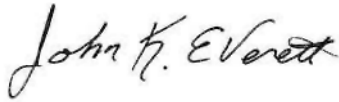
Figure 5e.
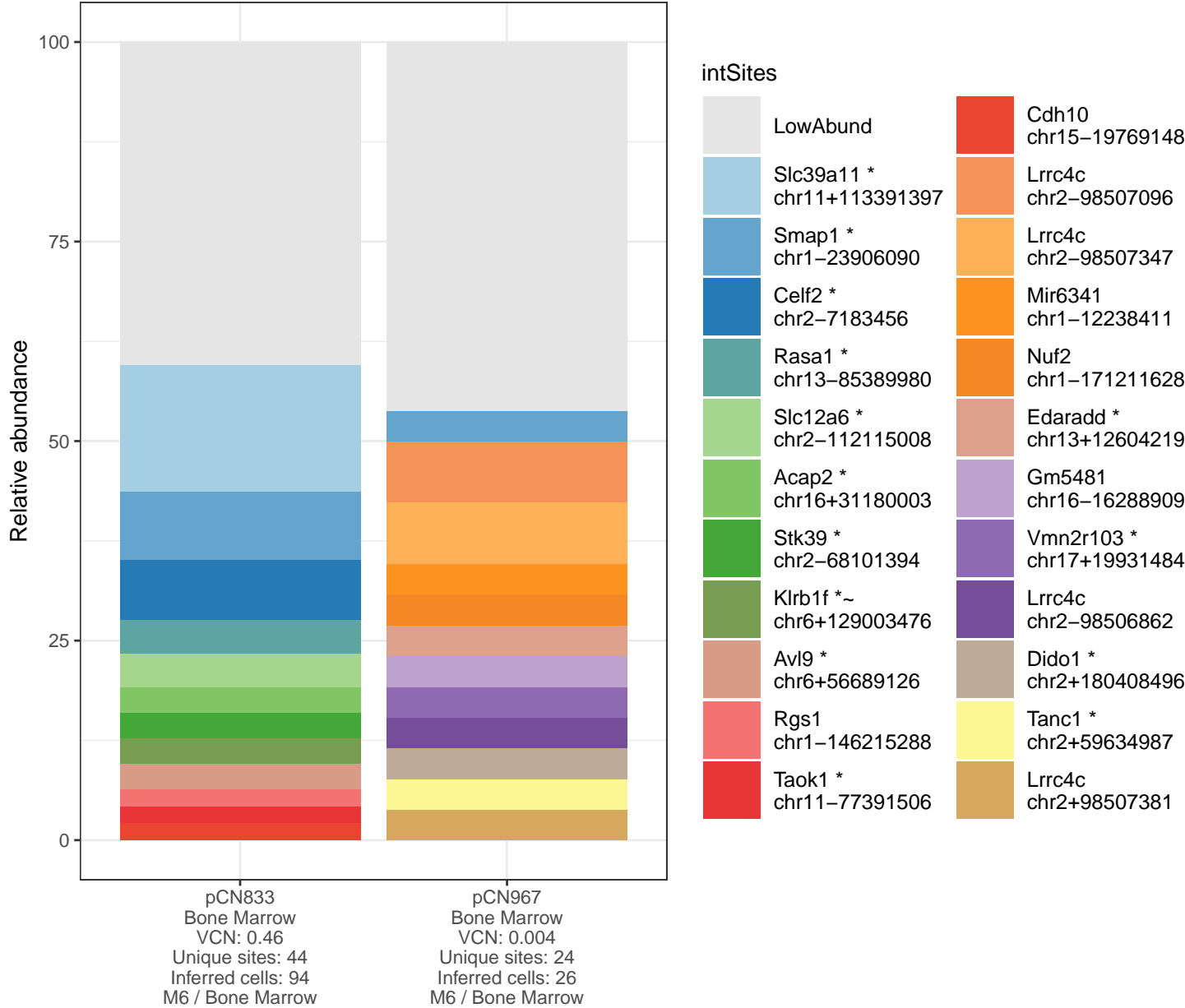
	Not near onco	Near onco
pCN833	41	3
pCN967	24	0

Fisher's exact p-value: 0.547

Analyst

John K. Everett

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 β-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. INSPIIRED: 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	GTSP1977	pCN774	Bone Marrow	0.039	M6	471	190
mouse	GTSP1978	pCN801	Bone Marrow	9.580	M6	7,842	102
mouse	GTSP1979	pCN803	Bone Marrow	0.654	M6	2,036	149
mouse	GTSP1982	pCN809	Bone Marrow	0.572	M6	2,675	100
mouse	GTSP2184	pCN835	Bone Marrow	1.481	M6	393	20
mouse	GTSP2185	pCN836	Bone Marrow	0.256	M6	99	64
mouse	GTSP2186	pCN837	Bone Marrow	0.138	M6	66	40
mouse	GTSP2187	pCN833	Bone Marrow	0.460	M6	94	44
mouse	GTSP2216	pCN873	Bone Marrow	0.786	M6	4,681	279
mouse	GTSP2217	pCN878	Bone Marrow	1.260	M6	4,125	225
mouse	GTSP2218	pCN879	Bone Marrow	1.061	M6	3,976	237
mouse	GTSP2320	pCN925	Bone Marrow	0.038	M6	363	34
mouse	GTSP2321	pCN929	Bone Marrow	2.000	M6	3,237	331
mouse	GTSP2322	pCN932	Bone Marrow	1.750	M6	2,436	140
mouse	GTSP2323	pCN934	Bone Marrow	0.506	M6	2,720	270
mouse	GTSP2324	pCN940	Bone Marrow	0.137	M6	1,431	121
mouse	GTSP2326	pCN951	Bone Marrow	0.083	M6	700	32
mouse	GTSP2328	pCN956	Bone Marrow	0.051	M6	369	122
mouse	GTSP2341	pCN801	Whole Blood	1.300	M2	533	145
mouse	GTSP2343	pCN801	Whole Blood	3.020	M6	4,356	337
mouse	GTSP2344	pCN801	BM Myeloid cells	7.950	M6	8,411	45
mouse	GTSP2345	pCN801	B-cells	2.690	M6	7,174	496
mouse	GTSP2346	pCN801	T-Cells	1.100	M6	1,960	38
mouse	GTSP2347	pCN948	Hematoma	0.018	M6	334	207
mouse	GTSP2349	pCN948	B-cells	0.019	M6	526	258
mouse	GTSP2350	pCN951	BM Myeloid cells	0.021	M6	184	35
mouse	GTSP2351	pCN951	B-cells	1.269	M6	2,869	24
mouse	GTSP2352	pCN951	T-Cells	0.026	M6	326	35
mouse	GTSP2353	pCN967	Bone Marrow	0.004	M6	26	24

Persistence of clones in mouse BM transplant trials

Table S2.

Donor	Recipient	Position	Donor cells	Recipient cells
pCN801	pCN951	chr10-63786517	568	69
pCN801	pCN951	chr12-91119873	615	55
pCN801	pCN951	chr12+41889014	629	52
pCN801	pCN951	chr16-58574863	596	63
pCN801	pCN951	chr17-92104190	589	43
pCN801	pCN951	chr17+84340467	626	56
pCN801	pCN951	chr18-47475939	552	75
pCN801	pCN951	chr18-6677426	5	4
pCN801	pCN951	chr3-62285345	585	49
pCN801	pCN951	chr5-90109547	195	2
pCN801	pCN951	chr5+91163622	569	46
pCN801	pCN951	chr7-80692518	604	58
pCN801	pCN951	chr8+116693173	599	56
pCN801	pCN951	chrX+82041081	596	53
pCN803	pCN956	chr1+162843101	258	38
pCN803	pCN956	chr1+43686997	1	1
pCN803	pCN956	chr10-20694821	2	1
pCN803	pCN956	chr10+54685831	1	1
pCN803	pCN956	chr10+72683496	1	2
pCN803	pCN956	chr12+75878946	4	1
pCN803	pCN956	chr13+120077397	6	5
pCN803	pCN956	chr14-105778668	392	35
pCN803	pCN956	chr17+38241441	10	2
pCN803	pCN956	chr17+55932173	1	1
pCN803	pCN956	chr17+95189892	2	3
pCN803	pCN956	chr18-42192342	1	1
pCN803	pCN956	chr19+3057507	2	3
pCN803	pCN956	chr2+126512656	6	2
pCN803	pCN956	chr2+134965236	254	56
pCN803	pCN956	chr3-25751852	1	1

Sequencing depth

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

Figure S3.