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

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

August 2018

Contents

Summary of results **2**

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 mouse subjects 4

Mapping of integration site positions **5**

Relative abundances of mouse subject samples **6**

Expanded clones **7**

References **8**

Supplementary tables and figures **9**

Numbers of inferred cells and integration sites identified in provided samples 9

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).

Mouse samples studied

Integration sites were detected in 16 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	16	9,464,092	29,630	2,346

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.

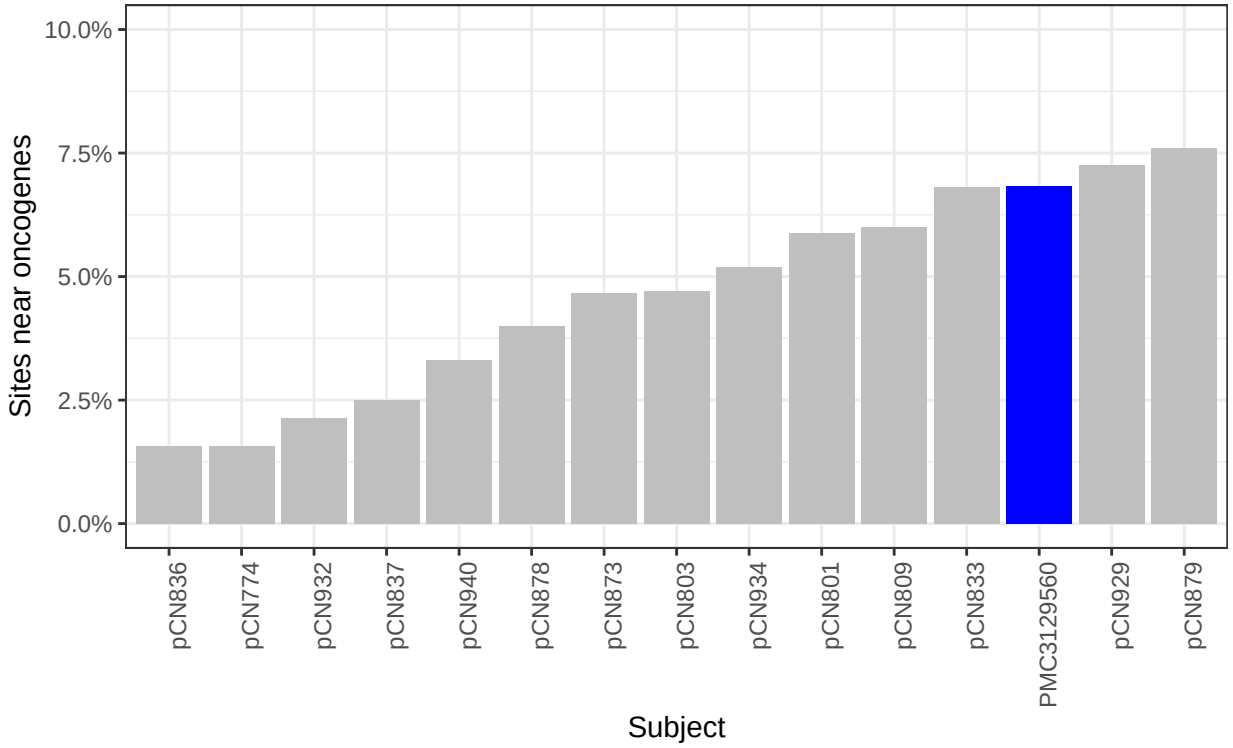
For this analysis, four technical replicates of each delivered sample were prepared, sequenced and analyzed with the INSPIIRED integration site analysis pipeline (v1.2)[4].

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 was generally less than the mean frequency of integration events near oncogenes in the published trial (Figure 1 [CYS: gray, β-thalassemia trial: blue]).

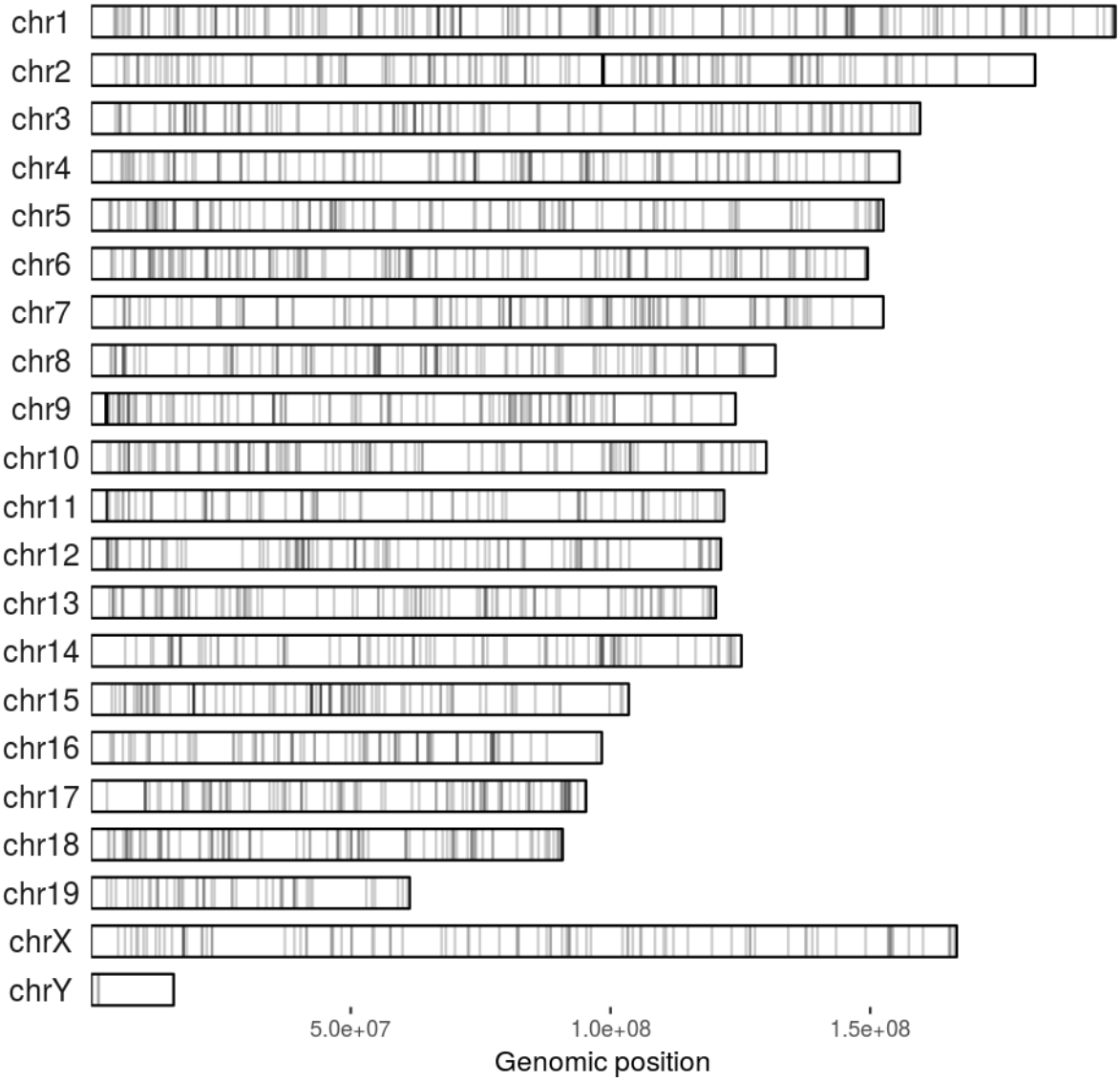
Figure 1. Comparison of frequencies of integration events near oncogenes.



Mapping of integration site positions

Heat map of identified integration positions is shown below in Figure 2.

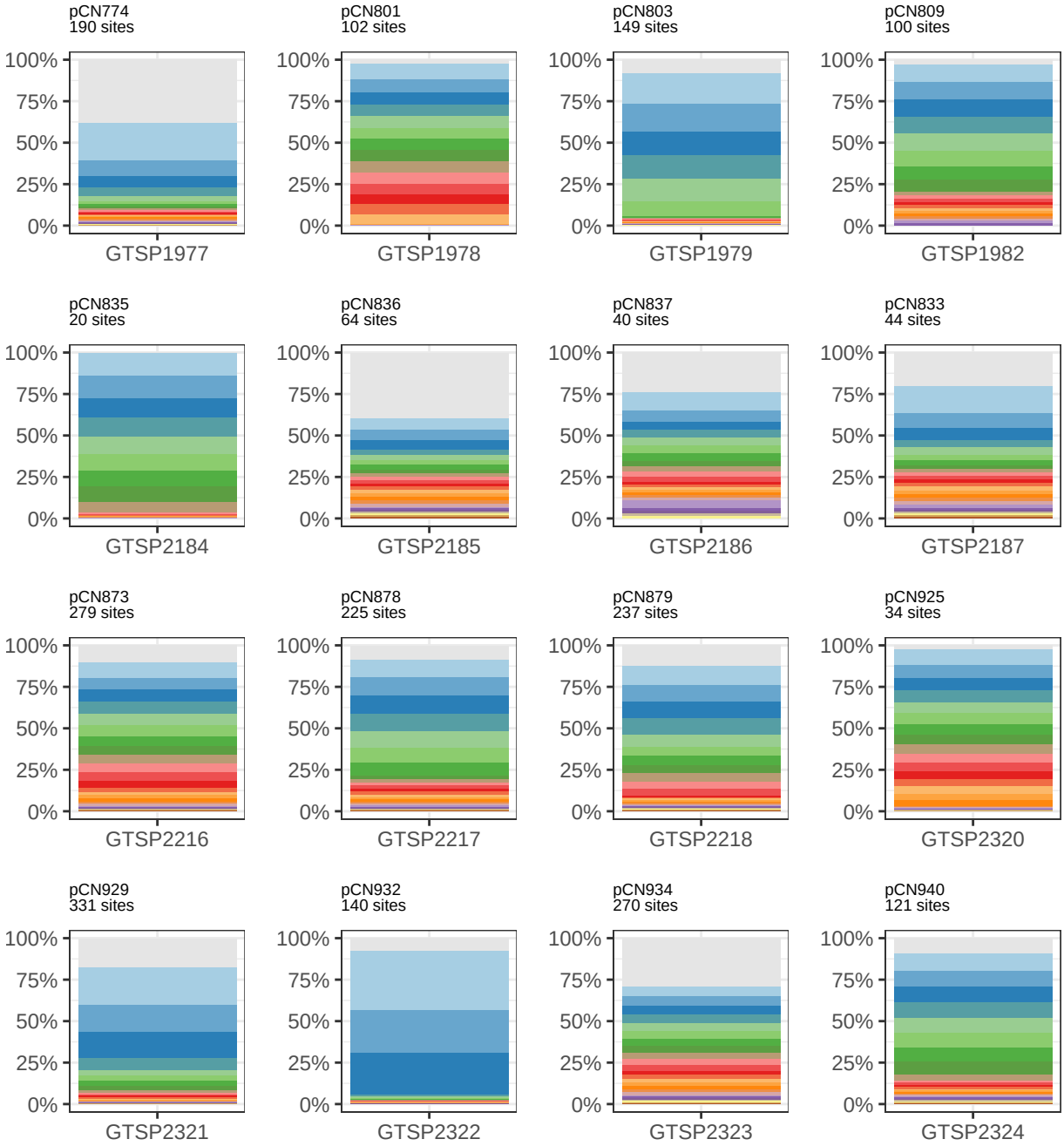
Figure 2



Relative abundances of mouse subject samples

The sample relative abundance plots below (Figure 3) 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 3.



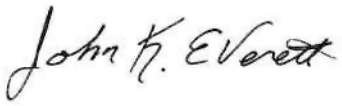
Expanded clones

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	22.15%	107	Tram2
pCN929	mouse	M6	Bone Marrow	chr12-38193729	22.61%	548	Agmo
pCN932	mouse	M6	Bone Marrow	chr12+80277915	35.25%	553	Rdh11
pCN932	mouse	M6	Bone Marrow	chr8-66234933	25.56%	401	Sgo2b
pCN932	mouse	M6	Bone Marrow	chr8+9405148	25.30%	397	Fam155a

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 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	483	190
mouse	GTSP1978	pCN801	Bone Marrow	10.490	M6	3,397	102
mouse	GTSP1979	pCN803	Bone Marrow	0.654	M6	1,827	149
mouse	GTSP1982	pCN809	Bone Marrow	0.572	M6	2,765	100
mouse	GTSP2184	pCN835	Bone Marrow	1.270	M6	404	20
mouse	GTSP2185	pCN836	Bone Marrow	2.970	M6	99	64
mouse	GTSP2186	pCN837	Bone Marrow	1.720	M6	63	40
mouse	GTSP2187	pCN833	Bone Marrow	1.070	M6	93	44
mouse	GTSP2216	pCN873	Bone Marrow	0.790	M6	4,716	279
mouse	GTSP2217	pCN878	Bone Marrow	1.260	M6	3,657	225
mouse	GTSP2218	pCN879	Bone Marrow	1.060	M6	3,469	237
mouse	GTSP2320	pCN925	Bone Marrow	0.040	M6	375	34
mouse	GTSP2321	pCN929	Bone Marrow	2.000	M6	2,424	331
mouse	GTSP2322	pCN932	Bone Marrow	1.750	M6	1,569	140
mouse	GTSP2323	pCN934	Bone Marrow	0.510	M6	2,820	270
mouse	GTSP2324	pCN940	Bone Marrow	0.140	M6	1,469	121