Analysis of moderately abundant clones from a beta-thalassemia gene therapy trial (M. Sadelain)

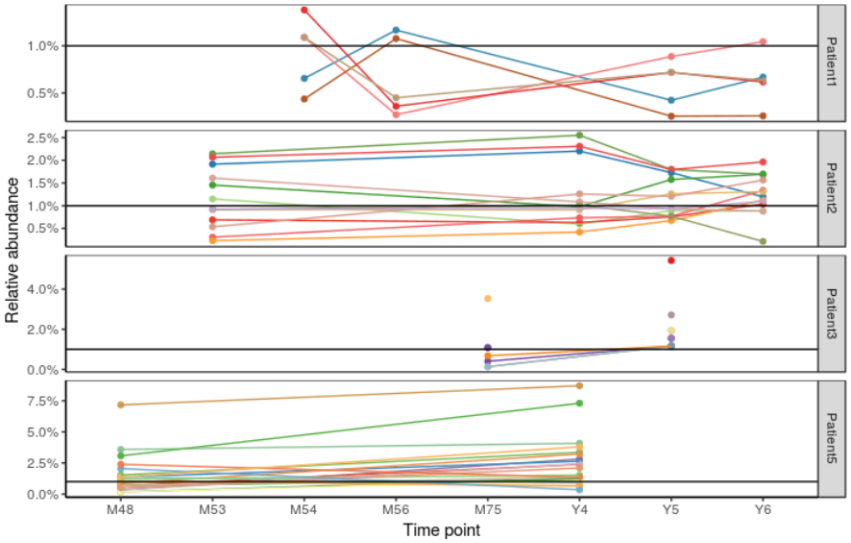
John K. Everett, Ph.D.

October 8, 2020

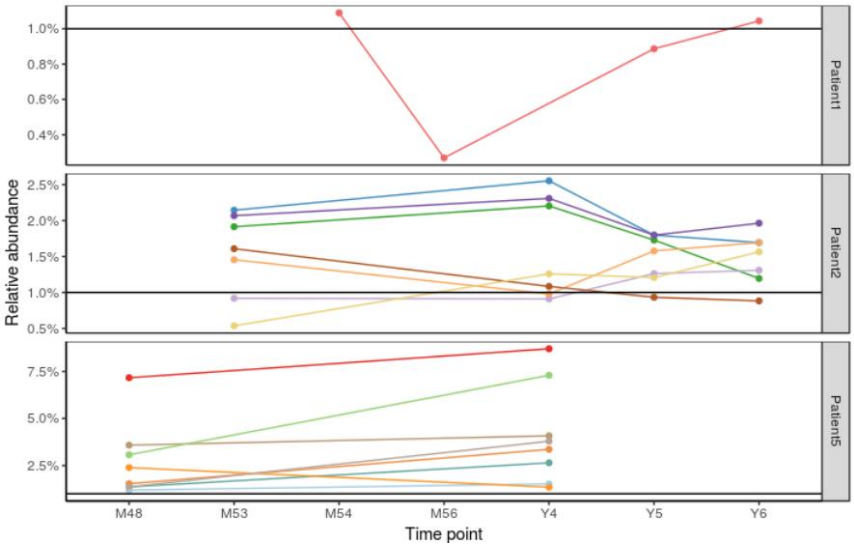
Data and spreadsheets can be found in the analysis code base:
http://github.com/everettJK/sadelaineAbundantClones

Four (4) patients had final whole blood time points ranging between 5 and 6 years.
We focused our analyses on samples with time points ≥ 4 years and at least 100 inferred cells in order not to be misled by the relative abundances.

58 clones exceeded 1% sample relative abundances at least 1 time point.



16 clones exceeded 1% sample relative abundances at 2 or more time points.
The following analyses did not yield significant results when focused solely on these clones.



Distances between integrations and the nearest oncogene boundaries were determined using four different methods.

1. Using nearest gene boundaries and the Bushman oncogene list [a] (composite of several resources).
2. Using gene start positions and the Bushman oncogene list.
3. Using nearest gene boundaries and the COSMIC oncogene list [b].
4. Using gene start positions and the COSMIC oncogene list.

a. 2009 genes with coordinates in human hg38 reference genome (http://www.bushmanlab.org/links/genelists).
b. 1166 genes with coordinates in human hg38 reference genome (https://cancer.sanger.ac.uk/census).

The table below shows the results of Fisher exact tests gauging the enrichment of clones which met or exceeded 1% sample relative abundances with integrations within a distance window.

pNearOncoGreater1:
Percent of clones with integrations proximal to oncogenes (within window) and relative abundances ≥ 1%.

pNotNearOncoGreater1:
Percent of clones with integrations distant to oncogenes (not within window) and relative abundances ≥ 1%.

method	window	pNearOncoGreater1	pNotNearOncoGreater1	pVal	pVal_1sided
1	10000	1.09%	0.61%	0.0719	0.0431
1	25000	1.07%	0.60%	0.0470	0.0318
1	50000	1.11%	0.55%	0.00988	0.00689
1	100000	0.98%	0.53%	0.0195	0.0124
2	10000	0.94%	0.68%	0.487	0.383
2	25000	0.98%	0.66%	0.271	0.203
2	50000	1.21%	0.58%	0.0145	0.00993
2	100000	0.92%	0.59%	0.116	0.0619
3	10000	1.18%	0.66%	0.190	0.115
3	25000	1.17%	0.65%	0.142	0.0986
3	50000	1.06%	0.65%	0.188	0.132
3	100000	0.91%	0.66%	0.345	0.205
4	10000	0.00%	0.70%	1	1
4	25000	1.35%	0.67%	0.150	0.150
4	50000	0.89%	0.68%	0.592	0.348
4	100000	0.80%	0.68%	0.682	0.387

The table below shows the results of Fisher Exact tests gauging the enrichment of clones which met or exceeded 1% sample relative abundances with integrations within oncogenes transcription units.

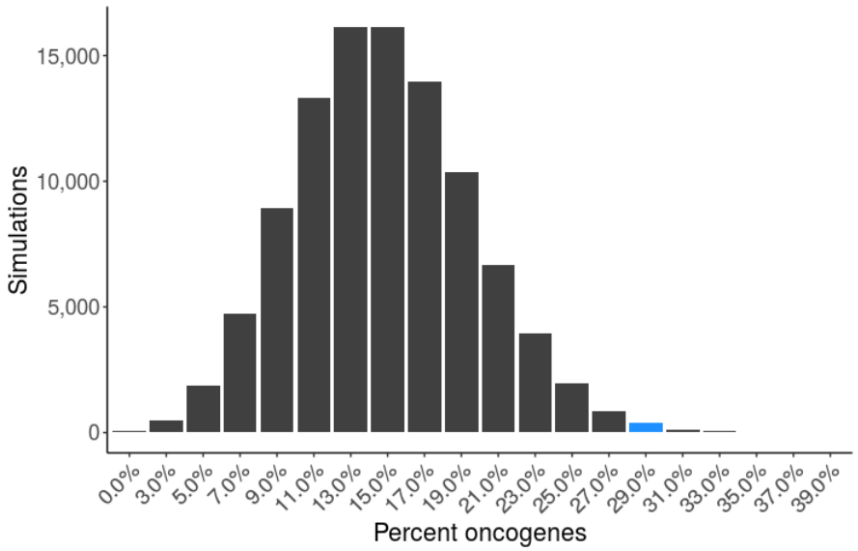
method	pInOncoGreater1	pNotInOncoGreater1	pVal	pVal_1sided
1	1.19%	0.612%	0.0350	0.0265
3	1.30%	0.651%	0.0974	0.0763

49 / 58 (84%) of clones with relative abundances meeting or exceeding 1% had integrations within gene transcription units.

We created a list of genes with one or more integrations within their boundaries from the ≥ 4 year time point data set (3,243 genes).

Using both the Bushman composite and COSMIC oncogene lists, we randomly drew 49 genes from this list of genes with one or more integrations 100,000 times and the percentage of drawn genes from the two oncogene lists are plotted below where the experimental percentages are highlighted blue.

Bushman composite oncogene list, p-value: 0.0061



COSMIC oncogene list, p-value: 0.0205