

Gene of Interest Report (GOI)

CART research group

April 01, 2024

<https://github.com/helixscript/myGOI>

Introduction

This report presents an analysis of genes targeted during T cell therapy over 783 treated subjects (Described in Jadowski et al.)

Lentiviral and gammaretroviral vectors integrate into genomes of targeted host cells in these studies T-cells. These genomic locations of vector integrations are identifiable through integration site sequencing. Abundances of individual cell clones can be inferred by the sonicLength method (Berry et al. 2012). In this report, we analyzed the data collected from integration site sequencing for subjects treated with engineered T cells.

We constructed 4 gene lists based on: 1 & 2) decreased (D) or enriched (E) integration site occurrence in patient samples relative to the initial transduction product, 3) peak clonal abundance (A), and 4) longitudinal clonal persistence (L).

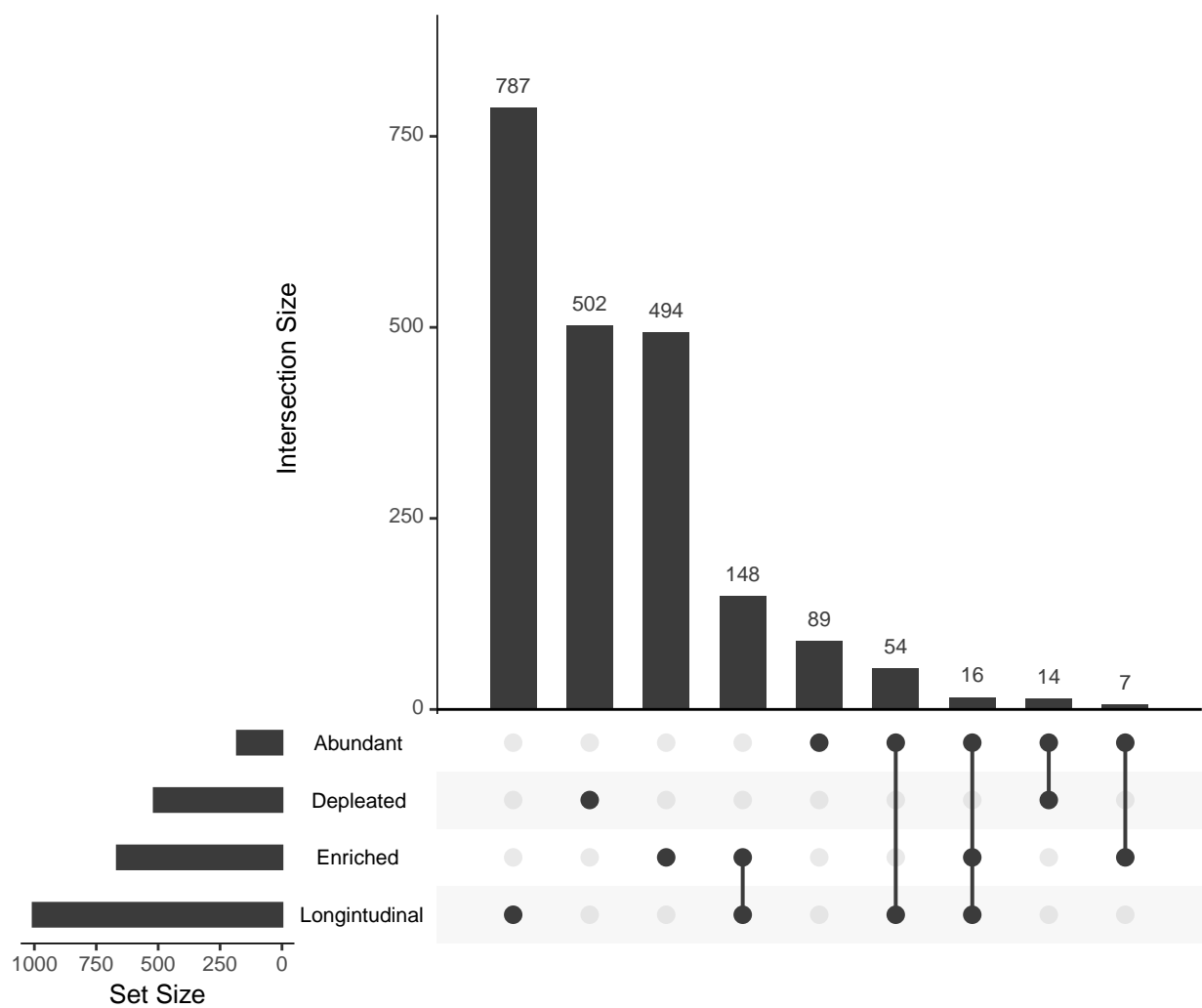
More about each of these criteria is below:

| | |
|------------------|---|
| Depleted (D) | Depleted genes show a significant decrease ($p \leq 0.05$) in integration frequency at later time points. |
| Enriched (E) | Enriched genes show a significant increase ($p \leq 0.05$) in integration frequency at later time points. |
| Abundant (A) | Abundant genes are genes associated with the top 1% of clonal abundance estimates at later time points. |
| Longitudinal (L) | Longitudinal genes are genes with ≥ 3 unique integrations from ≥ 3 patients recovered ≥ 90 days post-transduction. |

Note that integration sites are sampled from a larger population, so that that sampling of each specimen is incomplete.

The assignment of genes to the Enriched and Depleted categories is dependent on gene specific Fisher's Exact tests (uncorrected p-values ≤ 0.05). Figure 1 below depicts the number of genes associated with more than one category.

Figure 1.



The table below contains gene that have been associated with the Enriched, Abundant, and Longintudinal categories. These genes are of particular interest since three separate metrics suggests that integration near these genes may bolster persistence and/or proliferation over time.

Table 1. Genes annotated as EAL.

| gene | subjects | totalSites | percentChange | maxAbund | longitudinalSites | oncoGene | categories |
|----------|----------|------------|---------------|----------|-------------------|----------|------------|
| TRIO | 35 | 52 | 121.4% | 66 | 4 | yes | EAL |
| PPP6R3 | 124 | 642 | 58.4% | 154 | 13 | | EAL |
| CLK4 | 102 | 324 | 52.4% | 53 | 6 | | EAL |
| SRRM2 | 101 | 231 | 44.8% | 33 | 6 | | EAL |
| NELL2 | 103 | 316 | 43.6% | 441 | 6 | | EAL |
| SUPT3H | 119 | 589 | 38.6% | 31 | 4 | | EAL |
| MED13L | 112 | 358 | 36.9% | 38 | 4 | | EAL |
| FOXP1 | 120 | 431 | 35.6% | 49 | 5 | yes | EAL |
| AKAP13 | 109 | 344 | 33.7% | 27 | 4 | yes | EAL |
| VAV1 | 133 | 981 | 30.5% | 37 | 9 | yes | EAL |
| ANKRD11 | 136 | 807 | 26.8% | 25 | 14 | | EAL |
| ADD1 | 121 | 416 | 25.5% | 43 | 5 | | EAL |
| RNF213 | 126 | 603 | 24.6% | 76 | 7 | yes | EAL |
| PAFAH1B1 | 121 | 579 | 24.4% | 33 | 11 | | EAL |
| SAE1 | 121 | 520 | 20.7% | 24 | 5 | | EAL |
| CRAMP1 | 127 | 634 | 20.3% | 31 | 7 | | EAL |

The degree of integration near suspected oncogenes is important given that genotoxicity through disruption of oncogene transcription units or over expression via promoter insertion can potentially lead to adverse events. Here we test for significance of overlap between select oncogene lists and genes near recovered integration events. Test data sets include:

allOnco [2,579 genes (link)]: A comprehensive list of oncogenes compiled from multiple projects and consortia.

cosmic [581 genes (link)]: “Catalogue Of Somatic Mutations In Cancer” is an expert-curated database encompassing the wide variety of somatic mutation mechanisms causing human cancer.

cosmic_tsg [273 genes]: This gene list is a subset of the cosmic gene list annotated as tumor suppressors.

Table 2 summarizes the size of each criteria gene list identified by the various methods. Significance of overlap between gene lists are displayed by asterisks before the percent of genes identified from the criteria list which overlap with the column specified group. The asterisk to the left of the “/” indicates a p-value below 0.05 before multiple comparison corrections, while an asterisk to the right of the “/” indicates a p-value below 0.05 after multiple comparison corrections. Significance was tested using Fisher’s Exact test and multiple comparison corrections were made using a Benjamini-Hochberg (FDR) method for each criteria based list.

Table 2.

| Criteria | Genes | allOnco | cosmic | cosmic_tsg |
|--------------|-------|------------|-----------|------------|
| Depleted | 516 | */- 19.6% | */- 7.2% | */- 3.7% |
| Enriched | 665 | */ * 16.1% | */ * 5.1% | */ * 3.5% |
| Abundant | 180 | */ * 25.6% | */ * 9.4% | */ * 5.6% |
| Longitudinal | 1005 | */ * 22.7% | */ * 7.9% | */ * 4.4% |

Genes associated with the Enriched category have greater integration frequencies at later time points compared to earlier time points which suggests that integration near these genes bolsters cell survival.

| | |
|---------------|--|
| gene | Gene symbol. |
| subjects | Total number of subjects with an integration near gene. |
| earlyCount | Number of integration sites recovered from earlier time points (≤ 0 days). |
| lateCount | Number of integration sites recovered from later time points (> 0 days). |
| percentChange | Percent increase in integration frequency compared to earlier time period (> 0 days). |
| pVal | p-value from Fisher's Exact test. |
| pVal.adj | BH corrected p-value from Fisher's Exact test. |
| oncoGene | Gene is found in a broad lists of oncogenes. |
| categories | DEAL categories associated with gene. |

Table 3. Top 100 enriched genes. p-Values are marked with an * if ≤ 0.05 .

| gene | subjects | earlyCount | lateCount | percentChange | pVal | pVal.adj | oncoGene | categories |
|--------------|----------|------------|-----------|---------------|-----------|-----------|----------|------------|
| EP300 | 115 | 249 | 243 | 85.21% | 1.5e-11 * | 8.4e-08 * | yes | EL |
| KDM6A | 107 | 178 | 185 | 97.25% | 1.4e-10 * | 4.7e-07 * | yes | EL |
| CBFB | 130 | 389 | 327 | 59.54% | 7.2e-10 * | 1.7e-06 * | yes | EL |
| PPP6R3 | 124 | 357 | 298 | 58.42% | 7.4e-09 * | 1.2e-05 * | | EAL |
| KMT2D | 74 | 99 | 110 | 110.87% | 8.1e-08 * | 8.3e-05 * | yes | EA |
| LRPPRC | 68 | 52 | 69 | 151.83% | 4.4e-07 * | 3.8e-04 * | | EL |
| NUCB2 | 47 | 28 | 44 | 198.23% | 4.9e-06 * | 2.7e-03 * | | EL |
| CSMD1 | 45 | 12 | 28 | 342.83% | 6.1e-06 * | 3.2e-03 * | | EL |
| CREBBP | 123 | 300 | 233 | 47.40% | 1.1e-05 * | 5.3e-03 * | yes | EL |
| LPP | 116 | 184 | 155 | 59.87% | 2.2e-05 * | 9.1e-03 * | yes | EL |
| PPP4R2 | 47 | 27 | 40 | 181.16% | 2.8e-05 * | 1.1e-02 * | | EL |
| SMG1P7 | 68 | 56 | 64 | 116.90% | 2.9e-05 * | 1.1e-02 * | | E |
| VAV1 | 133 | 605 | 416 | 30.50% | 3.7e-05 * | 1.3e-02 * | yes | EAL |
| SUPT3H | 119 | 345 | 252 | 38.62% | 1.0e-04 * | 3.1e-02 * | | EAL |
| MICAL3 | 20 | 5 | 16 | 507.31% | 1.2e-04 * | 3.5e-02 * | | EL |
| ATG5 | 77 | 78 | 77 | 87.35% | 1.3e-04 * | 3.7e-02 * | | E |
| HELQ | 40 | 22 | 33 | 184.68% | 1.6e-04 * | 4.3e-02 * | | E |
| SMG1P1 | 110 | 127 | 110 | 64.38% | 1.6e-04 * | 4.3e-02 * | | EL |
| CLK4 | 102 | 183 | 147 | 52.45% | 1.6e-04 * | 4.3e-02 * | | EAL |
| SLC1A1 | 10 | 1 | 10 | 1797.84% | 1.8e-04 * | 4.5e-02 * | | E |
| MPP2 | 30 | 14 | 25 | 238.90% | 2.6e-04 * | 5.5e-02 | | E |
| LCOR | 81 | 102 | 92 | 71.18% | 2.7e-04 * | 5.6e-02 | | E |
| FANCL | 46 | 27 | 36 | 153.04% | 2.8e-04 * | 5.6e-02 | | E |
| CSDE1 | 37 | 18 | 28 | 195.22% | 2.9e-04 * | 5.7e-02 | | E |
| A4GALT | 12 | 2 | 11 | 943.81% | 3.0e-04 * | 5.9e-02 | | E |
| USP16 | 39 | 19 | 29 | 189.67% | 3.5e-04 * | 6.4e-02 | | E |
| MIR4435-2HG | 19 | 6 | 16 | 406.09% | 3.8e-04 * | 6.7e-02 | | EL |
| CALN1 | 25 | 10 | 20 | 279.57% | 3.8e-04 * | 6.7e-02 | | EL |
| PTPRT | 10 | 3 | 12 | 659.13% | 4.1e-04 * | 7.1e-02 | yes | E |
| NRXN3 | 24 | 9 | 19 | 300.65% | 4.4e-04 * | 7.5e-02 | | EL |
| TLL1 | 8 | 1 | 9 | 1608.05% | 4.8e-04 * | 7.8e-02 | | E |
| RHEB | 25 | 11 | 21 | 262.31% | 5.3e-04 * | 8.2e-02 | | E |
| AGAP1 | 21 | 8 | 18 | 327.01% | 5.7e-04 * | 8.6e-02 | | EL |
| FRG1CP | 57 | 48 | 51 | 101.64% | 6.3e-04 * | 9.3e-02 | | E |
| TCF12 | 84 | 110 | 94 | 62.18% | 6.7e-04 * | 9.6e-02 | yes | EL |
| FDX1 | 21 | 7 | 16 | 333.79% | 7.1e-04 * | 9.9e-02 | | E |
| LOC101927550 | 10 | 2 | 10 | 848.92% | 7.5e-04 * | 1.0e-01 | | E |
| RLIM | 19 | 6 | 15 | 374.46% | 7.6e-04 * | 1.0e-01 | | E |
| EIF4H | 29 | 16 | 25 | 196.54% | 7.6e-04 * | 1.0e-01 | | E |
| TAF2 | 62 | 54 | 55 | 93.30% | 8.0e-04 * | 1.0e-01 | | EL |
| PHF3 | 97 | 121 | 101 | 58.41% | 8.6e-04 * | 1.1e-01 | | EL |
| TIAM1 | 76 | 59 | 58 | 86.57% | 8.8e-04 * | 1.1e-01 | yes | EL |

Table 3. Top 100 enriched genes. p-Values are marked with an * if ≤ 0.05 . (continued)

| gene | subjects | earlyCount | lateCount | percentChange | pVal | pVal.adj | oncoGene | categories |
|--------------|----------|------------|-----------|---------------|-----------|----------|----------|------------|
| UTY | 72 | 95 | 83 | 65.81% | 8.9e-04 * | 1.1e-01 | | E |
| GOLT1B | 26 | 11 | 20 | 245.06% | 9.2e-04 * | 1.1e-01 | | E |
| FAM81A | 13 | 3 | 11 | 595.87% | 9.7e-04 * | 1.1e-01 | | E |
| GFRA1 | 12 | 3 | 11 | 595.87% | 9.7e-04 * | 1.1e-01 | | E |
| ANKRD11 | 136 | 488 | 326 | 26.78% | 1.0e-03 * | 1.2e-01 | | EAL |
| DNAAF4-CCPG1 | 25 | 8 | 17 | 303.29% | 1.0e-03 * | 1.2e-01 | | EL |
| FMR1 | 44 | 25 | 32 | 142.92% | 1.1e-03 * | 1.3e-01 | | E |
| WAC | 72 | 67 | 63 | 78.45% | 1.2e-03 * | 1.3e-01 | | EL |
| ESRRG | 8 | 1 | 8 | 1418.27% | 1.3e-03 * | 1.4e-01 | yes | EL |
| PPP3CA | 108 | 191 | 144 | 43.08% | 1.3e-03 * | 1.4e-01 | | EL |
| CHD8 | 65 | 75 | 68 | 72.07% | 1.5e-03 * | 1.5e-01 | | E |
| CNTN5 | 16 | 6 | 14 | 342.83% | 1.5e-03 * | 1.5e-01 | | E |
| CRAT37 | 17 | 6 | 14 | 342.83% | 1.5e-03 * | 1.5e-01 | | E |
| HHAT | 16 | 6 | 14 | 342.83% | 1.5e-03 * | 1.5e-01 | | EL |
| TTC7B | 19 | 6 | 14 | 342.83% | 1.5e-03 * | 1.5e-01 | | E |
| NELL2 | 103 | 181 | 137 | 43.65% | 1.7e-03 * | 1.6e-01 | | EAL |
| KIAA1217 | 14 | 5 | 13 | 393.44% | 1.7e-03 * | 1.6e-01 | | E |
| KIF13A | 15 | 5 | 13 | 393.44% | 1.7e-03 * | 1.6e-01 | | EL |
| PDCD6 | 14 | 5 | 13 | 393.44% | 1.7e-03 * | 1.6e-01 | yes | E |
| IFNGR2 | 9 | 2 | 9 | 754.03% | 1.8e-03 * | 1.7e-01 | yes | EA |
| PTPRN2 | 9 | 2 | 9 | 754.03% | 1.8e-03 * | 1.7e-01 | | E |
| ORC4 | 48 | 38 | 41 | 104.77% | 1.9e-03 * | 1.7e-01 | | E |
| FOXP1 | 120 | 252 | 180 | 35.56% | 2.0e-03 * | 1.8e-01 | yes | EAL |
| ATXN1 | 84 | 91 | 78 | 62.67% | 2.0e-03 * | 1.8e-01 | | EL |
| PPWD1 | 25 | 10 | 18 | 241.61% | 2.1e-03 * | 1.8e-01 | | E |
| CDK17 | 82 | 103 | 86 | 58.46% | 2.1e-03 * | 1.8e-01 | | E |
| ZNF148 | 78 | 103 | 86 | 58.46% | 2.1e-03 * | 1.8e-01 | | E |
| USP11 | 54 | 43 | 44 | 94.20% | 2.2e-03 * | 1.8e-01 | | E |
| LINC01687 | 12 | 3 | 10 | 532.61% | 2.2e-03 * | 1.8e-01 | | E |
| LTBP1 | 13 | 3 | 10 | 532.61% | 2.2e-03 * | 1.8e-01 | | E |
| NECAB1 | 13 | 3 | 10 | 532.61% | 2.2e-03 * | 1.8e-01 | | E |
| VAMP7 | 8 | 3 | 10 | 532.61% | 2.2e-03 * | 1.8e-01 | | E |
| RBPJ | 50 | 33 | 37 | 112.79% | 2.3e-03 * | 1.8e-01 | | E |
| RASA2 | 114 | 185 | 137 | 40.54% | 2.8e-03 * | 2.1e-01 | | E |
| FRG1DP | 35 | 23 | 28 | 131.04% | 3.0e-03 * | 2.2e-01 | | EL |
| ATP9A | 16 | 6 | 13 | 311.20% | 3.0e-03 * | 2.2e-01 | | E |
| LINGO2 | 18 | 6 | 13 | 311.20% | 3.0e-03 * | 2.2e-01 | | E |
| TENM2 | 18 | 6 | 13 | 311.20% | 3.0e-03 * | 2.2e-01 | | EL |
| TPST1 | 15 | 6 | 13 | 311.20% | 3.0e-03 * | 2.2e-01 | | E |
| CDK8 | 37 | 26 | 31 | 126.28% | 3.0e-03 * | 2.2e-01 | | E |
| CEP128 | 71 | 74 | 65 | 66.70% | 3.1e-03 * | 2.3e-01 | | EL |
| PLEKHA1 | 65 | 65 | 59 | 72.27% | 3.2e-03 * | 2.3e-01 | | EL |
| ADCYAP1 | 7 | 1 | 7 | 1228.48% | 3.3e-03 * | 2.3e-01 | yes | E |
| DOK5 | 7 | 1 | 7 | 1228.48% | 3.3e-03 * | 2.3e-01 | | E |
| FAM83A | 8 | 1 | 7 | 1228.48% | 3.3e-03 * | 2.3e-01 | | E |
| LINC01618 | 6 | 1 | 7 | 1228.48% | 3.3e-03 * | 2.3e-01 | | E |
| NEMP2 | 7 | 1 | 7 | 1228.48% | 3.3e-03 * | 2.3e-01 | | E |
| NFE2 | 8 | 1 | 7 | 1228.48% | 3.3e-03 * | 2.3e-01 | | E |
| NME6 | 7 | 1 | 7 | 1228.48% | 3.3e-03 * | 2.3e-01 | | E |
| SNAP91 | 7 | 1 | 7 | 1228.48% | 3.3e-03 * | 2.3e-01 | | E |
| TRAM2 | 8 | 1 | 7 | 1228.48% | 3.3e-03 * | 2.3e-01 | | E |
| DHX15 | 63 | 52 | 50 | 82.48% | 3.3e-03 * | 2.3e-01 | | E |
| ROCK1 | 99 | 134 | 104 | 47.29% | 3.3e-03 * | 2.3e-01 | | EL |
| FLJ42627 | 22 | 8 | 15 | 255.84% | 3.4e-03 * | 2.3e-01 | | E |
| NTAQ1 | 14 | 5 | 12 | 355.48% | 3.4e-03 * | 2.3e-01 | | E |
| DAZAP1 | 104 | 164 | 123 | 42.34% | 3.5e-03 * | 2.3e-01 | | EL |
| MYO10 | 21 | 10 | 17 | 222.63% | 3.6e-03 * | 2.4e-01 | | EL |
| TMPRSS11E | 22 | 10 | 17 | 222.63% | 3.6e-03 * | 2.4e-01 | | E |

Genes associated with the Depleted category have lower integration frequencies at later time points compared to earlier time points which suggests that integration near these genes may be detrimental to cell survival.

| | |
|---------------|--|
| gene | Gene symbol. |
| subjects | Total number of subjects with an integration near gene. |
| earlyCount | Number of integration sites recovered from earlier time points (≤ 0 days). |
| lateCount | Number of integration sites recovered from later time points (> 0 days). |
| percentChange | Percent increase in integration frequency compared to earlier time period (> 0 days). |
| pVal | p-value from Fisher's Exact test. |
| pVal.adj | BH corrected p-value from Fisher's Exact test. |
| oncoGene | Gene is found in a broad lists of oncogenes. |
| categories | DEAL categories associated with gene. |

Table 4. Top 100 depleted genes. p-Values are marked with an * if ≤ 0.05 .

| gene | subjects | earlyCount | lateCount | percentChange | pVal | pVal.adj | oncoGene | categories |
|-----------------|----------|------------|-----------|---------------|-----------|-----------|----------|------------|
| RNPS1 | 135 | 590 | 146 | -53.04% | 2.5e-18 * | 4.1e-14 * | | D |
| CD4 | 85 | 167 | 24 | -72.73% | 7.6e-12 * | 6.2e-08 * | | D |
| PACS1 | 145 | 1881 | 751 | -24.23% | 6.1e-11 * | 2.5e-07 * | | DA |
| KDM2A | 143 | 1378 | 531 | -26.87% | 4.0e-10 * | 1.1e-06 * | | D |
| EXOC2 | 111 | 314 | 81 | -51.04% | 1.2e-09 * | 2.5e-06 * | | D |
| IP6K1 | 139 | 609 | 199 | -37.99% | 1.4e-09 * | 2.5e-06 * | | D |
| SF3B2 | 80 | 142 | 25 | -66.59% | 1.9e-08 * | 2.8e-05 * | | D |
| TANK | 144 | 641 | 221 | -34.57% | 2.5e-08 * | 3.4e-05 * | | D |
| MROH1 | 140 | 1305 | 518 | -24.67% | 2.7e-08 * | 3.4e-05 * | | D |
| RUVBL2 | 61 | 96 | 12 | -76.28% | 3.9e-08 * | 4.5e-05 * | yes | D |
| LONP1 | 71 | 98 | 13 | -74.82% | 6.2e-08 * | 6.8e-05 * | | D |
| QRICH1 | 134 | 461 | 149 | -38.66% | 8.9e-08 * | 8.6e-05 * | | D |
| LINC02569 | 107 | 313 | 91 | -44.82% | 1.9e-07 * | 1.7e-04 * | | D |
| TBC1D10C | 56 | 87 | 12 | -73.82% | 6.6e-07 * | 5.4e-04 * | | D |
| ASH1L | 142 | 789 | 300 | -27.84% | 1.0e-06 * | 7.6e-04 * | | D |
| ARHGDIA | 120 | 281 | 82 | -44.62% | 1.0e-06 * | 7.6e-04 * | | D |
| UBE2J2 | 112 | 283 | 84 | -43.67% | 1.6e-06 * | 1.1e-03 * | | D |
| ANK3 | 94 | 186 | 47 | -52.04% | 2.3e-06 * | 1.6e-03 * | | D |
| NPLOC4 | 148 | 2175 | 956 | -16.58% | 2.4e-06 * | 1.6e-03 * | | D |
| WNK1 | 124 | 395 | 132 | -36.58% | 3.4e-06 * | 2.1e-03 * | | D |
| EIF2B3 | 124 | 344 | 111 | -38.76% | 4.1e-06 * | 2.5e-03 * | | D |
| KPNB1 | 83 | 129 | 28 | -58.81% | 4.7e-06 * | 2.7e-03 * | yes | D |
| NOSIP | 133 | 655 | 247 | -28.43% | 5.0e-06 * | 2.7e-03 * | | DA |
| TNFSF12-TNFSF13 | 46 | 57 | 6 | -80.02% | 7.3e-06 * | 3.7e-03 * | | D |
| PRR12 | 125 | 447 | 157 | -33.34% | 8.0e-06 * | 4.0e-03 * | | D |
| TNFSF12 | 45 | 56 | 6 | -79.67% | 1.1e-05 * | 5.3e-03 * | | D |
| SDF4 | 75 | 122 | 27 | -58.00% | 1.3e-05 * | 6.1e-03 * | | D |
| STAG2 | 61 | 82 | 14 | -67.60% | 2.1e-05 * | 9.0e-03 * | yes | D |
| VMP1 | 123 | 443 | 159 | -31.88% | 2.1e-05 * | 9.0e-03 * | yes | DA |
| RBM6 | 134 | 511 | 189 | -29.81% | 2.4e-05 * | 9.7e-03 * | | D |
| TONSL | 103 | 214 | 63 | -44.13% | 2.7e-05 * | 1.1e-02 * | | D |
| LSM2 | 99 | 238 | 73 | -41.79% | 2.8e-05 * | 1.1e-02 * | | D |
| CSGALNACT1 | 69 | 112 | 25 | -57.64% | 3.1e-05 * | 1.1e-02 * | | D |
| UBTF | 53 | 74 | 12 | -69.22% | 3.2e-05 * | 1.1e-02 * | | D |
| IL6R | 50 | 69 | 11 | -69.74% | 4.7e-05 * | 1.6e-02 * | | D |
| RETREG3 | 82 | 135 | 34 | -52.20% | 4.7e-05 * | 1.6e-02 * | | D |
| CCDC57 | 139 | 651 | 255 | -25.66% | 4.9e-05 * | 1.6e-02 * | | D |
| TBX21 | 35 | 42 | 4 | -81.93% | 7.4e-05 * | 2.4e-02 * | yes | D |
| PSMB9 | 104 | 294 | 99 | -36.09% | 8.2e-05 * | 2.6e-02 * | | D |
| WIPF1 | 63 | 81 | 16 | -62.51% | 1.0e-04 * | 3.1e-02 * | | D |
| LOC105369632 | 73 | 100 | 23 | -56.35% | 1.3e-04 * | 3.8e-02 * | | D |

Table 4. Top 100 depleted genes. p-Values are marked with an * if ≤ 0.05 . (continued)

| gene | subjects | earlyCount | lateCount | percentChange | pVal | pVal.adj | oncoGene | categories |
|-----------|----------|------------|-----------|---------------|-----------|-----------|----------|------------|
| CISH | 59 | 80 | 16 | -62.04% | 1.4e-04 * | 4.0e-02 * | | D |
| CD27-AS1 | 102 | 241 | 79 | -37.79% | 1.6e-04 * | 4.3e-02 * | | D |
| HCG20 | 118 | 494 | 190 | -27.01% | 1.8e-04 * | 4.5e-02 * | | D |
| HLA-DMB | 32 | 40 | 4 | -81.02% | 1.8e-04 * | 4.5e-02 * | | D |
| LINC02332 | 30 | 32 | 2 | -88.14% | 1.8e-04 * | 4.5e-02 * | | D |
| LPCAT3 | 100 | 252 | 84 | -36.74% | 1.9e-04 * | 4.5e-02 * | | D |
| SH3GL1 | 69 | 105 | 25 | -54.81% | 2.0e-04 * | 4.7e-02 * | yes | D |
| LINC01970 | 35 | 43 | 5 | -77.93% | 2.0e-04 * | 4.8e-02 * | | D |
| GATAD2B | 117 | 312 | 110 | -33.09% | 2.2e-04 * | 5.1e-02 | | D |
| EMBP1 | 52 | 57 | 9 | -70.03% | 2.2e-04 * | 5.1e-02 | | D |
| HLA-DPA1 | 41 | 53 | 8 | -71.35% | 2.3e-04 * | 5.1e-02 | | D |
| NCAPH2 | 93 | 148 | 42 | -46.14% | 2.3e-04 * | 5.2e-02 | | D |
| STAT3 | 113 | 276 | 95 | -34.68% | 2.5e-04 * | 5.3e-02 | yes | D |
| ERN1 | 73 | 127 | 34 | -49.19% | 2.5e-04 * | 5.3e-02 | | D |
| ENTHD1 | 80 | 133 | 36 | -48.63% | 2.5e-04 * | 5.3e-02 | | D |
| ANTXR2 | 102 | 175 | 53 | -42.52% | 2.7e-04 * | 5.6e-02 | | D |
| VRK3 | 106 | 229 | 75 | -37.84% | 2.8e-04 * | 5.6e-02 | | D |
| GSDMD | 21 | 25 | 1 | -92.41% | 3.1e-04 * | 6.0e-02 | | D |
| EIF4A3 | 47 | 59 | 10 | -67.83% | 3.2e-04 * | 6.0e-02 | | D |
| MAP3K14 | 90 | 137 | 38 | -47.36% | 3.2e-04 * | 6.0e-02 | | D |
| RELA | 105 | 191 | 60 | -40.38% | 3.2e-04 * | 6.0e-02 | yes | D |
| EIF3B | 48 | 55 | 9 | -68.94% | 3.2e-04 * | 6.0e-02 | | D |
| ZNF764 | 28 | 34 | 3 | -83.25% | 3.7e-04 * | 6.7e-02 | | D |
| KCTD13 | 73 | 104 | 26 | -52.55% | 4.1e-04 * | 7.1e-02 | | D |
| PTPN11 | 48 | 61 | 11 | -65.78% | 4.3e-04 * | 7.4e-02 | yes | D |
| FNBP1 | 130 | 516 | 204 | -24.97% | 4.7e-04 * | 7.8e-02 | yes | DA |
| MDS2 | 33 | 40 | 5 | -76.28% | 4.8e-04 * | 7.8e-02 | yes | D |
| EIF3L | 80 | 150 | 44 | -44.33% | 4.8e-04 * | 7.8e-02 | | D |
| HORMAD2 | 125 | 563 | 226 | -23.82% | 4.9e-04 * | 7.8e-02 | | D |
| AKAP8 | 41 | 50 | 8 | -69.63% | 4.9e-04 * | 7.8e-02 | | D |
| HSPA1B | 35 | 51 | 8 | -70.23% | 5.0e-04 * | 7.8e-02 | | D |
| PFKL | 24 | 24 | 1 | -92.09% | 5.1e-04 * | 7.9e-02 | | D |
| ADCK5 | 102 | 212 | 70 | -37.34% | 5.5e-04 * | 8.4e-02 | | D |
| ABCF1 | 87 | 180 | 57 | -39.90% | 6.1e-04 * | 9.1e-02 | | D |
| LINC00824 | 45 | 59 | 11 | -64.62% | 6.1e-04 * | 9.1e-02 | | D |
| NCR3 | 73 | 115 | 31 | -48.84% | 6.3e-04 * | 9.3e-02 | | D |
| PTK2 | 64 | 85 | 20 | -55.35% | 6.4e-04 * | 9.3e-02 | yes | D |
| MYH9 | 97 | 175 | 55 | -40.35% | 6.4e-04 * | 9.3e-02 | yes | D |
| SNRNP200 | 48 | 53 | 9 | -67.77% | 6.9e-04 * | 9.8e-02 | | D |
| LINC02570 | 43 | 49 | 8 | -69.01% | 7.2e-04 * | 9.9e-02 | | D |
| HGS | 84 | 117 | 32 | -48.09% | 7.2e-04 * | 9.9e-02 | | D |
| TYK2 | 70 | 111 | 30 | -48.71% | 7.2e-04 * | 9.9e-02 | | D |
| RAB40C | 106 | 229 | 78 | -35.36% | 7.4e-04 * | 1.0e-01 | | D |
| UNK | 112 | 238 | 82 | -34.61% | 7.7e-04 * | 1.0e-01 | | D |
| PPP6R2 | 137 | 720 | 302 | -20.40% | 7.8e-04 * | 1.0e-01 | | D |
| FLT3LG | 103 | 224 | 76 | -35.61% | 8.0e-04 * | 1.0e-01 | | D |
| DKC1 | 22 | 23 | 1 | -91.75% | 8.4e-04 * | 1.1e-01 | yes | D |
| H2BC5 | 22 | 23 | 1 | -91.75% | 8.4e-04 * | 1.1e-01 | | D |
| GPD2 | 70 | 84 | 20 | -54.81% | 8.7e-04 * | 1.1e-01 | | D |
| PRKCH | 90 | 143 | 43 | -42.93% | 8.8e-04 * | 1.1e-01 | | D |
| OBSCN | 60 | 89 | 22 | -53.09% | 9.1e-04 * | 1.1e-01 | yes | D |
| ALG9 | 29 | 31 | 3 | -81.63% | 9.1e-04 * | 1.1e-01 | | D |
| IFT140 | 126 | 401 | 155 | -26.64% | 9.4e-04 * | 1.1e-01 | | DA |
| PYM1 | 80 | 122 | 35 | -45.55% | 1.0e-03 * | 1.2e-01 | | D |
| ABCA7 | 48 | 63 | 13 | -60.84% | 1.0e-03 * | 1.2e-01 | | D |
| MECP2 | 131 | 703 | 296 | -20.09% | 1.1e-03 * | 1.2e-01 | | D |
| SPHK2 | 66 | 88 | 22 | -52.55% | 1.2e-03 * | 1.3e-01 | | D |
| FKBP5 | 128 | 601 | 249 | -21.37% | 1.3e-03 * | 1.4e-01 | yes | D |
| TAP2 | 88 | 165 | 53 | -39.04% | 1.3e-03 * | 1.4e-01 | yes | D |

Genes associated with the Abundant category reached high levels of clonal abundance measured by the sonic abundance method at later time points which suggests that integration near these genes may bolster cell division. For this analysis, the threshold for inclusion is an estimated abundances ≥ 24 cells.

| | |
|-------------|--|
| gene | Gene symbol. |
| subjects | Total number of subjects with an integration near gene. |
| lateCount | Number of integration sites recovered from earlier time points (≤ 0 days). |
| maxAbund | Maximum estimated clonal abundance observed > 0 days. |
| maxRelAbund | Maximum relative sample clonal abundance observed > 0 days. |
| oncoGene | Gene is found in a broad lists of oncogenes. |
| categories | DEAL categories associated with gene. |

Table 5. Top 100 abundant genes.

| gene | subjects | lateCount | maxAbund | maxRelAbund | oncoGene | categories |
|--------------|----------|-----------|----------|-------------|----------|------------|
| TET2 | 73 | 60 | 858 | 99.00% | yes | AL |
| TET2-AS1 | 71 | 48 | 858 | 99.00% | | AL |
| PATL1 | 49 | 31 | 629 | 27.09% | | A |
| PIKFYVE | 53 | 33 | 460 | 28.64% | | A |
| NELL2 | 103 | 137 | 441 | 14.85% | | EAL |
| GLCCI1 | 88 | 61 | 425 | 14.31% | | A |
| SRCAP | 105 | 126 | 402 | 38.00% | | AL |
| MTMR3 | 106 | 95 | 280 | 8.16% | yes | AL |
| C1ORF159 | 119 | 125 | 176 | 11.21% | | AL |
| IFNGR2 | 9 | 9 | 173 | 45.58% | yes | EA |
| RC3H1 | 82 | 75 | 164 | 4.42% | | A |
| PCNX1 | 111 | 112 | 156 | 1.06% | | A |
| PPP6R3 | 124 | 298 | 154 | 5.43% | | EAL |
| UHRF1 | 81 | 58 | 150 | 9.55% | | AL |
| SSH2 | 122 | 170 | 146 | 6.29% | | AL |
| RSRC1 | 75 | 55 | 115 | 1.47% | | A |
| WDR7 | 111 | 96 | 113 | 44.84% | | AL |
| MAPK14 | 96 | 93 | 98 | 1.28% | yes | AL |
| SNHG12 | 5 | 4 | 98 | 5.74% | | A |
| ZZEF1 | 127 | 247 | 92 | 50.00% | | AL |
| MGA | 105 | 128 | 91 | 5.05% | | AL |
| RPA3 | 54 | 31 | 90 | 4.68% | | AL |
| UMAD1 | 79 | 53 | 90 | 4.68% | | AL |
| AQR | 59 | 29 | 88 | 7.66% | | A |
| LEF1 | 98 | 91 | 86 | 3.70% | yes | A |
| MAN1B1 | 106 | 114 | 85 | 18.93% | | A |
| ZNF573 | 50 | 20 | 85 | 63.43% | | AL |
| LINC01473 | 17 | 8 | 84 | 8.18% | | A |
| CARD8 | 132 | 244 | 82 | 8.17% | yes | DA |
| BCAS3 | 96 | 69 | 81 | 32.14% | yes | AL |
| IQCB1 | 46 | 19 | 80 | 2.86% | | A |
| KANSL1 | 110 | 134 | 79 | 3.30% | | AL |
| WWOX | 68 | 51 | 78 | 69.03% | yes | AL |
| LOC100294362 | 90 | 69 | 76 | 2.47% | | AL |
| RNF213 | 126 | 241 | 76 | 2.86% | yes | EAL |
| DNAJC13 | 77 | 60 | 73 | 0.56% | | A |
| EXOSC10 | 60 | 31 | 69 | 0.77% | | A |
| ATP2A2 | 77 | 48 | 67 | 2.89% | | AL |
| TRIO | 35 | 28 | 66 | 2.57% | yes | EAL |
| SEC31A | 64 | 40 | 65 | 3.23% | | A |
| SMAP2 | 70 | 48 | 65 | 0.46% | | A |
| GPN1 | 44 | 15 | 63 | 1.82% | | A |
| EARS2 | 16 | 7 | 60 | 28.17% | | A |

Table 5. Top 100 abundant genes. *(continued)*

| gene | subjects | lateCount | maxAbund | maxRelAbund | oncoGene | categories |
|--------------|----------|-----------|----------|-------------|----------|------------|
| LINC01322 | 3 | 1 | 57 | 3.63% | | A |
| JMJD6 | 32 | 22 | 54 | 1.56% | | A |
| CLK4 | 102 | 147 | 53 | 6.08% | | EAL |
| DERL2 | 64 | 38 | 52 | 14.29% | | A |
| MEMO1 | 60 | 24 | 52 | 1.22% | | A |
| PTBP1 | 102 | 116 | 52 | 4.48% | yes | EA |
| FOXP1 | 120 | 180 | 49 | 36.57% | yes | EAL |
| PPP3CC | 120 | 117 | 49 | 52.13% | | DA |
| DNM2 | 111 | 118 | 48 | 4.40% | yes | AL |
| UBR1 | 94 | 82 | 48 | 42.11% | | AL |
| EIF2AK4 | 50 | 28 | 47 | 0.85% | | A |
| RASEF | 6 | 5 | 47 | 0.60% | | EA |
| DYNC1H1 | 77 | 59 | 46 | 0.38% | | A |
| UXT-AS1 | 11 | 4 | 45 | 6.69% | | A |
| GRB2 | 131 | 218 | 44 | 4.41% | yes | AL |
| NGDN | 36 | 19 | 44 | 0.57% | | A |
| TAC3 | 34 | 10 | 44 | 1.86% | | A |
| ZNF92 | 27 | 20 | 44 | 3.12% | | EA |
| ADD1 | 121 | 168 | 43 | 16.28% | | EAL |
| CPSF1 | 104 | 73 | 43 | 9.89% | | DA |
| OPA1 | 49 | 28 | 43 | 1.40% | | A |
| ZNF251 | 120 | 265 | 43 | 19.00% | | AL |
| PHF12 | 52 | 18 | 42 | 2.21% | | DA |
| ACTL6A | 17 | 7 | 41 | 0.27% | | A |
| POLG2 | 24 | 17 | 41 | 0.54% | | EA |
| DIDO1 | 101 | 104 | 40 | 18.78% | | AL |
| MICAL2 | 41 | 16 | 40 | 2.86% | | A |
| LUC7L | 135 | 357 | 39 | 22.41% | | AL |
| PHF20 | 93 | 60 | 39 | 1.96% | yes | DA |
| RNF157 | 134 | 485 | 39 | 7.58% | yes | AL |
| DNAJC1 | 64 | 36 | 38 | 1.03% | | A |
| ELMO1 | 88 | 45 | 38 | 0.98% | yes | A |
| MED13L | 112 | 150 | 38 | 2.38% | | EAL |
| PA2G4 | 61 | 44 | 38 | 1.25% | yes | A |
| HERC4 | 77 | 42 | 37 | 2.14% | | A |
| MAD1L1 | 83 | 62 | 37 | 1.82% | yes | AL |
| VAV1 | 133 | 416 | 37 | 15.69% | yes | EAL |
| HRH1 | 10 | 6 | 36 | 16.90% | | AL |
| KMT2B | 97 | 77 | 36 | 28.07% | | A |
| PDE3B | 91 | 93 | 36 | 2.13% | | AL |
| STAG3 | 87 | 70 | 36 | 0.53% | | A |
| SUZ12 | 77 | 66 | 36 | 1.21% | yes | A |
| CRTAP | 6 | 5 | 35 | 0.35% | | A |
| HSF5 | 81 | 46 | 35 | 1.92% | | AL |
| KHDC4 | 95 | 83 | 35 | 2.23% | | AL |
| LOC101927151 | 22 | 8 | 35 | 0.53% | | A |
| MYH11 | 11 | 7 | 34 | 1.75% | yes | A |
| NDE1 | 42 | 18 | 34 | 1.75% | | A |
| PAM | 61 | 25 | 34 | 3.77% | | A |
| RHOD | 7 | 3 | 34 | 23.40% | | A |
| ZNF34 | 130 | 182 | 34 | 28.16% | | AL |
| ZNF568 | 34 | 15 | 34 | 41.98% | | A |
| C5 | 16 | 5 | 33 | 71.74% | | A |
| DCUN1D4 | 62 | 31 | 33 | 4.88% | | A |
| DNAJB5 | 41 | 23 | 33 | 21.71% | | A |
| KMT2D | 74 | 110 | 33 | 0.82% | yes | EA |
| PAFAH1B1 | 121 | 234 | 33 | 5.17% | | EAL |

Genes are categorized as Longitudinal if ≥ 3 different integrations across ≥ 3 different subjects are observed ≥ 90 days post-transduction.

| | |
|------------------------|--|
| gene | Gene symbol. |
| totalSites | Total number of unique integrations associated with gene. |
| longitudinalSites | Total number of unique integrations associated with gene recovered ≥ 90 days. |
| longitudinalSubjects | Number of subjects associated with integrations associated with gene recovered ≥ 90 days. |
| longitudinalTimePoints | Number of time points sampled ≥ 90 days. |
| latestTimePointDays | Longest time point sampled. |
| oncoGene | Gene is found in a broad lists of oncogenes. |
| categories | DEAL categories associated with gene. |

Table 6. Top 100 longitudinally persistent genes.

| gene | totalSites | longitudinalSites | longitudinalSubjects | longitudinalTimePoints | latestTimePointDays | oncoGene |
|--------------|------------|-------------------|----------------------|------------------------|---------------------|----------|
| ZNF251 | 719 | 17 | 10 | 9 | 4015 | |
| RAB11FIP3 | 589 | 11 | 8 | 6 | 4015 | |
| KANSL1 | 419 | 10 | 5 | 6 | 4015 | |
| NELL2 | 316 | 6 | 5 | 3 | 4015 | |
| RFX2 | 217 | 6 | 6 | 7 | 4015 | yes |
| CD96 | 163 | 5 | 4 | 6 | 4015 | |
| AP2A2 | 229 | 4 | 3 | 2 | 4015 | |
| EHBP1 | 112 | 4 | 4 | 4 | 4015 | |
| HIVEP3 | 149 | 4 | 3 | 4 | 4015 | yes |
| FOXP1 | 431 | 5 | 4 | 12 | 3285 | yes |
| CENPP | 100 | 4 | 4 | 4 | 3285 | |
| DYM | 205 | 4 | 3 | 3 | 3285 | yes |
| GAK | 176 | 4 | 3 | 3 | 3285 | yes |
| SHANK2 | 22 | 4 | 4 | 4 | 3285 | |
| LOC730100 | 16 | 3 | 3 | 3 | 3285 | |
| MGAT4C | 8 | 3 | 3 | 3 | 3285 | |
| MYO3B | 16 | 3 | 3 | 3 | 3285 | |
| SLC9A7 | 97 | 3 | 3 | 4 | 3285 | |
| UHRF1 | 174 | 7 | 4 | 4 | 3042 | |
| USP25 | 342 | 7 | 6 | 6 | 3042 | |
| ACACA | 262 | 6 | 5 | 4 | 3042 | |
| TRAT1 | 100 | 6 | 3 | 6 | 3042 | |
| CLEC16A | 308 | 5 | 5 | 4 | 3042 | |
| GXYLT1 | 49 | 5 | 3 | 4 | 3042 | |
| MACROD2 | 73 | 5 | 3 | 4 | 3042 | |
| NTM | 16 | 5 | 4 | 5 | 3042 | |
| ABCD2 | 208 | 4 | 4 | 4 | 3042 | |
| HRAS | 19 | 4 | 3 | 4 | 3042 | yes |
| KIFC1 | 159 | 4 | 4 | 4 | 3042 | |
| KLHL6 | 35 | 4 | 3 | 3 | 3042 | |
| SDHC | 100 | 4 | 3 | 4 | 3042 | yes |
| SHISA6 | 10 | 4 | 4 | 3 | 3042 | |
| WDR45B | 170 | 4 | 3 | 4 | 3042 | |
| ZNF254 | 55 | 4 | 4 | 5 | 3042 | |
| HROB | 45 | 3 | 3 | 2 | 3042 | |
| LCMT1 | 21 | 3 | 3 | 3 | 3042 | |
| LOC100289333 | 78 | 3 | 3 | 5 | 3042 | |
| MKKS | 35 | 3 | 3 | 3 | 3042 | |
| PTPRM | 64 | 3 | 3 | 3 | 3042 | |
| SDCCAG8 | 73 | 3 | 3 | 2 | 3042 | yes |
| UBE2H | 135 | 3 | 3 | 3 | 3042 | |
| RNF157 | 1361 | 15 | 7 | 7 | 2920 | yes |
| VAPB | 39 | 4 | 3 | 2 | 2920 | |
| FANCA | 1757 | 19 | 12 | 5 | 2555 | yes |

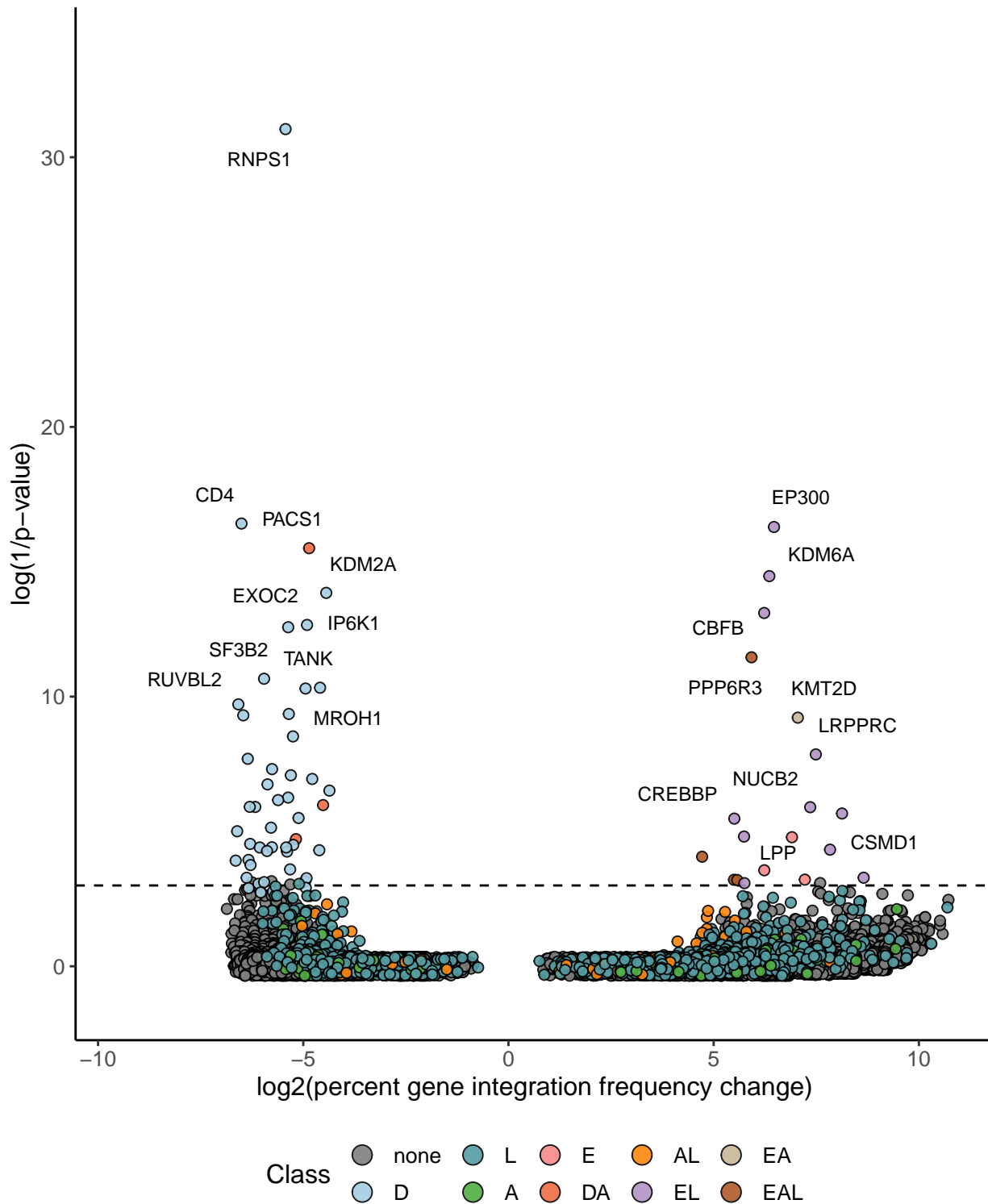
Table 6. Top 100 longitudinally persistent genes. *(continued)*

| gene | totalSites | longitudinalSites | longitudinalSubjects | longitudinalTimePoints | latestTimePointDays | oncoGene |
|--------------|------------|-------------------|----------------------|------------------------|---------------------|----------|
| FCHSD2 | 704 | 17 | 9 | 7 | 2555 | |
| ANKRD11 | 807 | 14 | 10 | 9 | 2555 | |
| SMARCC1 | 736 | 14 | 9 | 8 | 2555 | |
| PPP6R3 | 642 | 13 | 9 | 4 | 2555 | |
| SMG1P5 | 500 | 10 | 7 | 6 | 2555 | |
| VAV1 | 981 | 9 | 5 | 8 | 2555 | yes |
| AP3B1 | 177 | 8 | 7 | 3 | 2555 | |
| MTOR | 290 | 8 | 6 | 7 | 2555 | yes |
| EYS | 39 | 7 | 4 | 5 | 2555 | |
| MPP7 | 148 | 7 | 6 | 3 | 2555 | |
| CAMKMT | 86 | 6 | 3 | 5 | 2555 | |
| HHAT | 20 | 6 | 6 | 4 | 2555 | |
| KMT2C | 271 | 6 | 6 | 5 | 2555 | yes |
| RAP1GDS1 | 129 | 6 | 4 | 4 | 2555 | yes |
| ZBTB8OS | 49 | 6 | 3 | 5 | 2555 | |
| ZZEF1 | 706 | 6 | 5 | 19 | 2555 | |
| BAZ1A | 182 | 5 | 4 | 3 | 2555 | yes |
| ACOX1 | 332 | 4 | 4 | 3 | 2555 | |
| ATAD2B | 128 | 4 | 3 | 3 | 2555 | |
| DNAAF4-CCPG1 | 25 | 4 | 4 | 4 | 2555 | |
| ECPAS | 83 | 4 | 3 | 3 | 2555 | |
| MED13L | 358 | 4 | 4 | 4 | 2555 | |
| NUP160 | 84 | 4 | 4 | 4 | 2555 | |
| PHF21A | 247 | 4 | 4 | 3 | 2555 | |
| PPHLN1 | 84 | 4 | 3 | 4 | 2555 | |
| TENM2 | 19 | 4 | 4 | 3 | 2555 | |
| UNKL | 199 | 4 | 3 | 4 | 2555 | |
| ARSK | 27 | 3 | 3 | 3 | 2555 | |
| ATAD3B | 196 | 3 | 3 | 2 | 2555 | |
| COL21A1 | 10 | 3 | 3 | 3 | 2555 | |
| CUX1 | 109 | 3 | 3 | 3 | 2555 | yes |
| FAM114A2 | 29 | 3 | 3 | 2 | 2555 | |
| GRK6 | 69 | 3 | 3 | 3 | 2555 | |
| IQCJ-SCHIP1 | 20 | 3 | 3 | 3 | 2555 | |
| KCNB1 | 7 | 3 | 3 | 3 | 2555 | |
| LRPPRC | 121 | 3 | 3 | 3 | 2555 | |
| MACO1 | 207 | 3 | 3 | 3 | 2555 | |
| MALRD1 | 21 | 3 | 3 | 3 | 2555 | |
| MAP3K20 | 33 | 3 | 3 | 3 | 2555 | |
| POLR3K | 154 | 3 | 3 | 3 | 2555 | |
| SH3BP4 | 7 | 3 | 3 | 3 | 2555 | |
| STX2 | 33 | 3 | 3 | 3 | 2555 | |
| TIAM1 | 117 | 3 | 3 | 3 | 2555 | yes |
| UBR3 | 151 | 3 | 3 | 2 | 2555 | |
| SMG1P2 | 418 | 10 | 5 | 6 | 2372 | |
| RBFOX1 | 50 | 7 | 5 | 6 | 2372 | |
| ARHGAP15 | 533 | 5 | 4 | 3 | 2372 | |
| IQGAP1 | 385 | 5 | 5 | 4 | 2372 | |
| POLR2A | 729 | 5 | 3 | 4 | 2372 | |
| RIGI | 91 | 5 | 4 | 3 | 2372 | |
| SARNP | 468 | 4 | 4 | 3 | 2372 | yes |
| A1BG | 16 | 3 | 3 | 2 | 2372 | |
| AKAP10 | 124 | 3 | 3 | 2 | 2372 | |
| HYCC1 | 105 | 3 | 3 | 3 | 2372 | |
| TCERG1 | 111 | 3 | 3 | 3 | 2372 | |
| EHMT1 | 858 | 14 | 9 | 6 | 2190 | yes |

Volcano plot analysis

Changes in gene integration frequencies between pre- and post- transfusion samples and their significance is shown below in Figure 2. Changes in integration frequency is defined as $(f - fo) / fo$ where f is the integration frequency post-transfusion and fo is the integration of preinfused transduction material. The significance of these changes is assessed using Fisher's Exact tests corrected for multiple comparisons.

Figure 2.



References

1. Denlinger N, Bond D, Jaglowski S. CAR T-cell therapy for B-cell lymphoma. Curr Probl Cancer. 2022 Feb;46(1):100826. doi: 10.1016/j.crrproblcancer.2021.100826. Epub 2021 Dec 25
2. Berry CC, Gillet NA, Melamed A, Gormley N, Bangham CR, Bushman FD. Estimating abundances of retroviral insertion sites from DNA fragment length data. Bioinformatics. 2012 Mar 15;28(6):755-62. doi: 10.1093/bioinformatics/bts004. Epub 2012 Jan 11.

Software parameters

| parameter | value |
|--------------------------------|--|
| earlyVsLateCutoffDays | 0 |
| inputDataPath | data/intSiteData.tsv.gz |
| minSampleAbund | 25 |
| minGeneSubjects | 2 |
| maxDistNearestGene | 50000 |
| longitudinal__minNumSubjects | 3 |
| longitudinal__minNumSites | 3 |
| longitudinal__minNumTimepoints | 3 |
| longitudinal__minTimeDays | 90 |
| volcanoPlot__numTopGeneLabels | 10 |
| COSMIC__oncogene_table | data/COSMIC__oncogenes.txt |
| COSMIC__tsg_table | data/COSMIC__oncogenes_tumor_suppressors.txt |
| allOnco__oncogene_table | data/allOnco.txt |