

# myGOI report

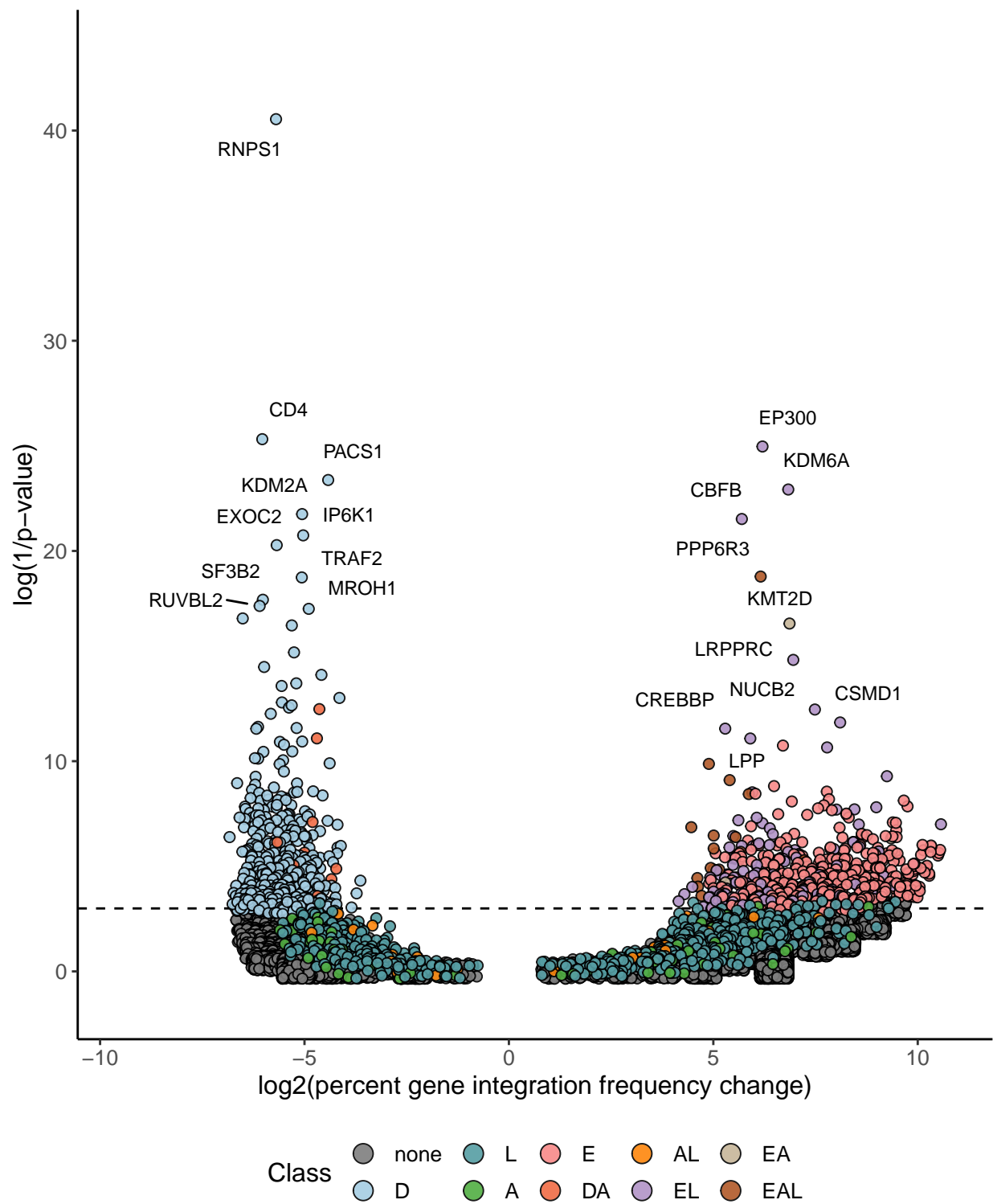
John K. Everett, Ph.D.

March 22, 2024

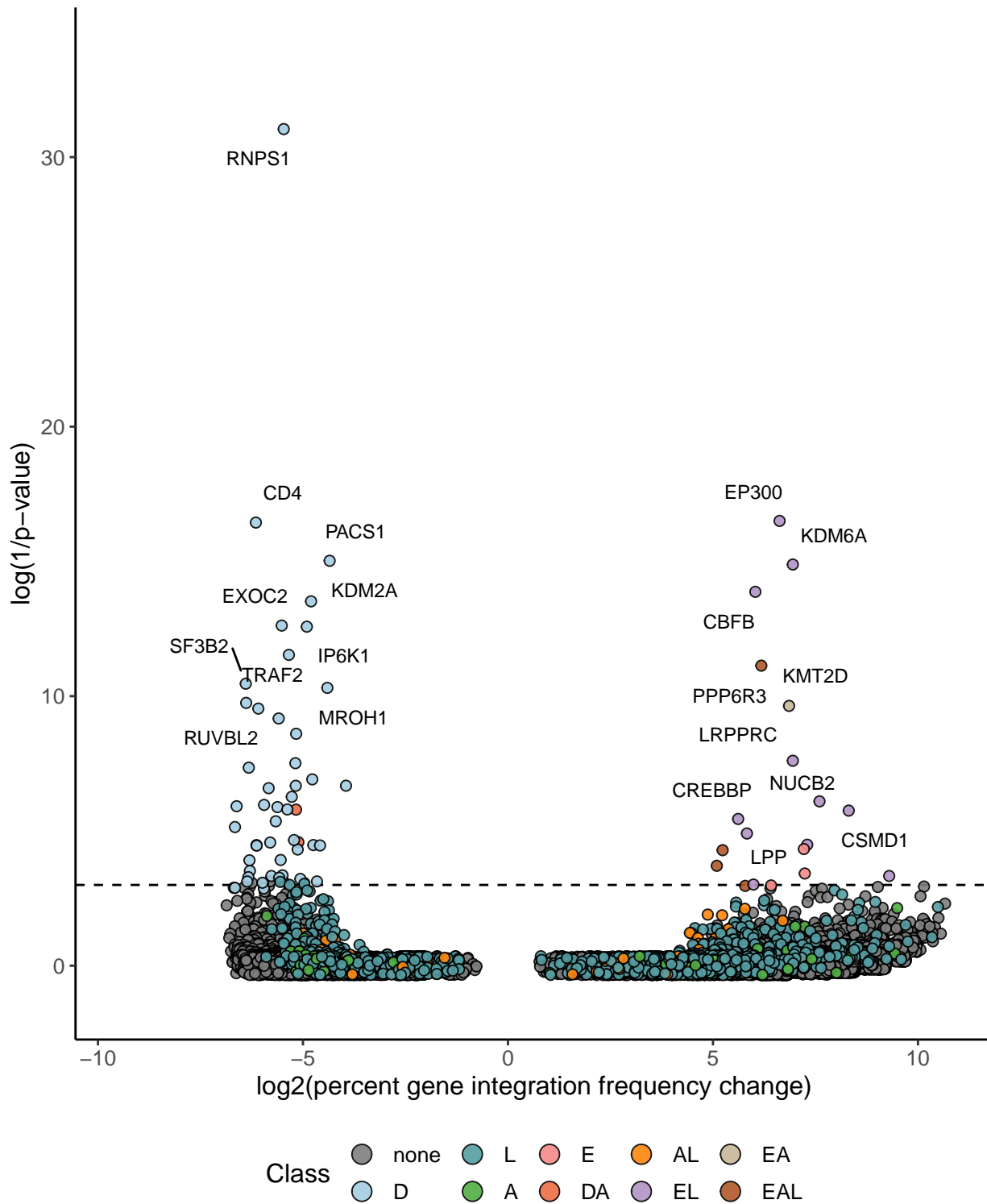
<https://github.com/helixscript/myGOI>

parameter	value
earlyVsLateCutoffDays	0
inputDataPath	expandedIntSiteData.tsv.gz
minSampleAbund	25
minGeneSubjects	2
maxDistNearestGene	50000
longitudinal__minNumSubjects	3
longitudinal__minTimeDays	90
volcanoPlot__numTopGeneLabels	10

The volcano plot below depicts changes in integration frequency of genes between early and later time points. Integration frequency is defined as the number of unique integration sites near a gene divided by the total number of integration sites recovered within the early or later time periods. The change in integration frequency is defined as  $(f - f_0) / f_0$  where  $f$  is the frequency during later time points and  $f_0$  is the frequency during earlier time points. The significance of enrichment or depletion of integration events for each gene is assessed using Fisher's Exact tests. This plot does not correct for multiple comparisons.



Alternatively, the volcano plot can be drawn where the p-values from the multiple Fisher's Exact tests are corrected for multiple comparisons using the Benjamini & Hochberg method.



Genes of interest can be associated with four possible categories:

Depleted (D)	Depleted genes show a significant decrease ( $p < 0.05$ ) in integration frequency at later time points.
Enriched (E)	Enriched genes show a significant increase ( $p < 0.05$ ) 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 at least 3 unique integrations in at least 3 subjects recovered after 90 days.

The assignment of genes to the Enriched and Depleted categories is dependent on gene specific Fisher’s Exact tests (uncorrected p-values  $\leq 0.05$ ). Genes associated with the Enriched, Abundant, and Longitudinal categories are of particular interest.

gene	Gene symbol.
subjects	Total number of subjects with an integration near gene.
percentChange	Percent increase in integration frequency compared to earlier time period.
maxAbund	Max. abundance estimate observed $> 0$ days.
longitudinalSites	Number of integration sites observed $\geq 90$ days.
categories	DEAL categories associated with gene.

Table 1. Genes annotated as EAL.

gene	subjects	totalSites	percentChange	maxAbund	longitudinalSites	categories
TRIO	35	52	121.4%	66	4	EAL
PPP6R3	124	642	58.4%	154	13	EAL
CLK4	102	324	52.4%	53	6	EAL
SRRM2	101	232	46.2%	33	6	EAL
NELL2	103	316	43.6%	441	6	EAL
SUPT3H	119	589	38.6%	31	4	EAL
MED13L	112	358	36.8%	38	4	EAL
FOXP1	120	431	35.5%	49	5	EAL
AKAP13	109	344	33.7%	27	4	EAL
VAV1	133	981	30.5%	37	9	EAL
ANKRD11	136	807	26.8%	25	14	EAL
ADD1	121	416	25.5%	43	5	EAL
RNF213	126	603	24.6%	76	7	EAL
PAFAH1B1	121	579	24.4%	33	11	EAL
CRAMP1	127	634	20.3%	31	7	EAL

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 ( $\leq 0$ 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' Exact test.
categories	DEAL categories associated with gene.

Table 2. Top 100 enriched genes.

gene	subjects	earlyCount	lateCount	percentChange	pVal	categories
EP300	115	249	243	85.18%	1.54e-11	EL
KDM6A	107	178	185	97.21%	1.44e-10	EL
CBFB	130	389	328	59.99%	5.73e-10	EL
PPP6R3	124	357	298	58.39%	7.39e-09	EAL
KMT2D	74	99	110	110.83%	8.14e-08	EA
LRPPRC	68	52	69	151.78%	4.45e-07	EL
NUCB2	47	28	44	198.18%	4.90e-06	EL
CSMD1	45	12	28	342.75%	6.08e-06	EL
CREBBP	123	300	233	47.37%	1.15e-05	EL
LPP	116	184	155	59.84%	2.18e-05	EL
PPP4R2	47	27	40	181.11%	2.82e-05	EL
SMG1P7	68	56	64	116.86%	2.91e-05	E
VAV1	133	605	416	30.47%	3.73e-05	EAL
SUPT3H	119	345	252	38.60%	1.03e-04	EAL
MICAL3	20	5	16	507.20%	1.17e-04	EL
ATG5	77	78	77	87.32%	1.27e-04	E
HELQ	40	22	33	184.63%	1.61e-04	E
SMG1P1	110	127	110	64.35%	1.62e-04	EL
CLK4	102	183	147	52.42%	1.62e-04	EAL
SLC1A1	10	1	10	1797.50%	1.81e-04	E
LCOR	81	102	92	71.15%	2.66e-04	E
FANCL	46	27	36	153.00%	2.78e-04	E
CSDE1	37	18	28	195.17%	2.86e-04	E
A4GALT	12	2	11	943.63%	3.02e-04	E
USP16	39	19	29	189.62%	3.49e-04	E
DLG2	27	13	23	235.71%	3.53e-04	E
MIR4435-2HG	19	6	16	406.00%	3.80e-04	EL
CALN1	25	10	20	279.50%	3.81e-04	EL
PTPRT	10	3	12	659.00%	4.13e-04	E
NRXN3	24	9	19	300.58%	4.42e-04	EL
TLL1	8	1	9	1607.75%	4.79e-04	E
RHEB	25	11	21	262.25%	5.29e-04	E
AGAP1	21	8	18	326.94%	5.68e-04	EL
FRG1CP	57	48	51	101.61%	6.34e-04	E
TCF12	84	110	94	62.15%	6.75e-04	EL
FDX1	21	7	16	333.71%	7.06e-04	E
LOC101927550	10	2	10	848.75%	7.46e-04	E
RLIM	19	6	15	374.38%	7.57e-04	E
EIF4H	29	16	25	196.48%	7.61e-04	E
TAF2	62	54	55	93.26%	7.95e-04	EL
PHF3	97	121	101	58.39%	8.61e-04	EL
TIAM1	76	59	58	86.53%	8.76e-04	EL
UTY	72	95	83	65.78%	8.88e-04	E
GOLT1B	26	11	20	245.00%	9.21e-04	E
FAM81A	13	3	11	595.75%	9.67e-04	E
GFRA1	12	3	11	595.75%	9.67e-04	E
ANKRD11	136	488	326	26.76%	1.02e-03	EAL

Table 2. Top 100 enriched genes. *(continued)*

gene	subjects	earlyCount	lateCount	percentChange	pVal	categories
DNAAF4-CCPG1	25	8	17	303.22%	1.02e-03	EL
FMR1	44	25	32	142.88%	1.13e-03	E
WAC	72	67	63	78.42%	1.16e-03	EL
ESRRG	8	1	8	1418.00%	1.26e-03	EL
PPP3CA	108	191	144	43.06%	1.27e-03	EL
CHD8	65	75	68	72.04%	1.46e-03	E
CNTN5	16	6	14	342.75%	1.51e-03	E
CRAT37	17	6	14	342.75%	1.51e-03	E
HHAT	16	6	14	342.75%	1.51e-03	EL
TTC7B	19	6	14	342.75%	1.51e-03	E
NELL2	103	181	137	43.62%	1.72e-03	EAL
KIAA1217	14	5	13	393.35%	1.73e-03	E
KIF13A	15	5	13	393.35%	1.73e-03	EL
PDCD6	14	5	13	393.35%	1.73e-03	E
IFNGR2	9	2	9	753.88%	1.82e-03	EA
PTPRN2	9	2	9	753.88%	1.82e-03	E
ORC4	48	38	41	104.73%	1.88e-03	E
FOXP1	120	252	180	35.54%	2.00e-03	EAL
ATXN1	84	91	78	62.64%	2.01e-03	EL
PPWD1	25	10	18	241.55%	2.11e-03	E
CDK17	82	103	86	58.43%	2.12e-03	E
ZNF148	78	103	86	58.43%	2.12e-03	E
USP11	54	43	44	94.16%	2.19e-03	E
LINC01687	12	3	10	532.50%	2.23e-03	E
LTBP1	13	3	10	532.50%	2.23e-03	E
NECAB1	13	3	10	532.50%	2.23e-03	E
VAMP7	8	3	10	532.50%	2.23e-03	E
RBPJ	50	33	37	112.75%	2.27e-03	E
RASA2	114	185	137	40.52%	2.76e-03	E
FRG1DP	35	23	28	131.00%	3.00e-03	EL
ATP9A	16	6	13	311.13%	3.00e-03	E
LINGO2	18	6	13	311.13%	3.00e-03	E
TENM2	18	6	13	311.13%	3.00e-03	EL
TPST1	15	6	13	311.13%	3.00e-03	E
CDK8	37	26	31	126.24%	3.03e-03	E
CEP128	71	74	65	66.67%	3.11e-03	EL
PLEKHA1	65	65	59	72.23%	3.23e-03	EL
ADCYAP1	7	1	7	1228.25%	3.26e-03	E
DOK5	7	1	7	1228.25%	3.26e-03	E
FAM83A	8	1	7	1228.25%	3.26e-03	E
LINC01618	6	1	7	1228.25%	3.26e-03	E
NEMP2	7	1	7	1228.25%	3.26e-03	E
NFE2	8	1	7	1228.25%	3.26e-03	E
NME6	7	1	7	1228.25%	3.26e-03	E
SNAP91	7	1	7	1228.25%	3.26e-03	E
TRAM2	8	1	7	1228.25%	3.26e-03	E
DHX15	63	52	50	82.45%	3.31e-03	E
ROCK1	99	134	104	47.27%	3.33e-03	EL
FLJ42627	22	8	15	255.78%	3.37e-03	E
WDYHV1	14	5	12	355.40%	3.38e-03	E
DAZAP1	104	164	123	42.31%	3.46e-03	EL
MYO10	21	10	17	222.58%	3.59e-03	EL
TMPRSS11E	22	10	17	222.58%	3.59e-03	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 ( $\leq 0$ 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' Exact test.
categories	DEAL categories associated with gene.

Table 3. Top 100 depleted genes.

gene	subjects	earlyCount	lateCount	percentChange	pVal	categories
RNPS1	135	590	146	-53.04%	2.54e-18	D
CD4	85	167	24	-72.73%	7.55e-12	D
PACS1	145	1880	751	-24.20%	6.08e-11	D
KDM2A	143	1378	531	-26.88%	4.00e-10	D
EXOC2	111	314	81	-51.05%	1.21e-09	D
IP6K1	139	609	199	-38.00%	1.39e-09	D
TRAF2	140	575	190	-37.30%	8.20e-09	D
SF3B2	80	142	25	-66.59%	1.89e-08	D
MROH1	140	1305	518	-24.68%	2.74e-08	D
RUVBL2	61	96	12	-76.28%	3.87e-08	D
LONP1	71	98	13	-74.83%	6.21e-08	D
QRICH1	134	461	149	-38.67%	8.94e-08	D
LINC02569	107	313	91	-44.83%	1.89e-07	D
TBC1D10C	56	87	12	-73.83%	6.62e-07	D
ASH1L	142	789	300	-27.85%	1.01e-06	D
ARHGDIA	120	281	82	-44.63%	1.02e-06	D
UBE2J2	112	283	84	-43.68%	1.57e-06	D
ANK3	94	186	47	-52.05%	2.29e-06	D
NPLOC4	148	2174	956	-16.56%	2.43e-06	D
WNK1	124	395	132	-36.59%	3.40e-06	D
EIF2B3	124	344	111	-38.77%	4.11e-06	D
KPNB1	83	129	28	-58.81%	4.69e-06	D
NOSIP	133	655	247	-28.45%	5.02e-06	DA
TNFSF12-TNFSF13	46	57	6	-80.03%	7.27e-06	D
PRR12	125	447	157	-33.35%	8.01e-06	D
TNFSF12	45	56	6	-79.67%	1.13e-05	D
SDF4	75	122	27	-58.01%	1.34e-05	D
STAG2	61	82	14	-67.60%	2.06e-05	D
VMP1	123	443	159	-31.90%	2.10e-05	DA
RBM6	134	511	189	-29.82%	2.39e-05	D
TONSL	103	214	63	-44.14%	2.74e-05	D
LSM2	99	238	73	-41.80%	2.77e-05	D
CSGALNACT1	69	112	25	-57.65%	3.11e-05	D
UBTF	53	74	12	-69.23%	3.22e-05	D
IL6R	50	69	11	-69.75%	4.70e-05	D
RETREG3	82	135	34	-52.21%	4.72e-05	D
CCDC57	139	651	255	-25.67%	4.86e-05	D
TBX21	35	42	4	-81.93%	7.40e-05	D
PSMB9	104	294	99	-36.10%	8.17e-05	D
WIPF1	63	81	16	-62.52%	1.03e-04	D
LOC105369632	73	100	23	-56.36%	1.34e-04	D
CISH	59	80	16	-62.05%	1.44e-04	D
CD27-AS1	102	241	79	-37.80%	1.62e-04	D
HCG20	118	494	190	-27.02%	1.79e-04	D
HLA-DMB	32	40	4	-81.02%	1.80e-04	D
LINC02332	30	32	2	-88.14%	1.82e-04	D

Table 3. Top 100 depleted genes. (*continued*)

gene	subjects	earlyCount	lateCount	percentChange	pVal	categories
LPCAT3	100	252	84	-36.75%	1.85e-04	D
SH3GL1	69	105	25	-54.82%	1.96e-04	D
LINC01970	35	43	5	-77.94%	2.01e-04	D
GATAD2B	117	312	110	-33.10%	2.19e-04	D
EMBP1	52	57	9	-70.04%	2.25e-04	D
HLA-DPA1	41	53	8	-71.36%	2.25e-04	D
NCAPH2	93	148	42	-46.15%	2.34e-04	D
STAT3	113	276	95	-34.69%	2.45e-04	D
ERN1	73	127	34	-49.20%	2.46e-04	D
ENTHD1	80	133	36	-48.64%	2.47e-04	D
ANTXR2	102	175	53	-42.53%	2.75e-04	D
VRK3	106	229	75	-37.85%	2.80e-04	D
GSDMD	21	25	1	-92.41%	3.13e-04	D
EIF4A3	47	59	10	-67.84%	3.16e-04	D
MAP3K14	90	137	38	-47.37%	3.17e-04	D
RELA	105	191	60	-40.39%	3.21e-04	D
EIF3B	48	55	9	-68.95%	3.24e-04	D
ZNF764	28	34	3	-83.26%	3.72e-04	D
KCTD13	73	104	26	-52.56%	4.10e-04	D
PTPN11	48	61	11	-65.78%	4.29e-04	D
FNBP1	130	516	204	-24.98%	4.73e-04	D
MDS2	33	40	5	-76.28%	4.75e-04	D
EIF3L	80	150	44	-44.34%	4.78e-04	D
HORMAD2	125	563	226	-23.83%	4.86e-04	D
AKAP8	41	50	8	-69.64%	4.94e-04	D
HSPA1B	35	51	8	-70.24%	4.95e-04	D
PFKL	24	24	1	-92.09%	5.07e-04	D
ADCK5	102	212	70	-37.35%	5.47e-04	D
ABCF1	87	180	57	-39.91%	6.06e-04	D
LINC00824	45	59	11	-64.62%	6.06e-04	D
NCR3	73	115	31	-48.85%	6.33e-04	D
PTK2	64	85	20	-55.35%	6.42e-04	D
MYH9	97	175	55	-40.36%	6.44e-04	D
SNRNP200	48	53	9	-67.78%	6.95e-04	D
LINC02570	43	49	8	-69.02%	7.15e-04	D
HGS	84	117	32	-48.10%	7.19e-04	D
TYK2	70	111	30	-48.72%	7.22e-04	D
RAB40C	106	229	78	-35.37%	7.45e-04	D
UNK	112	238	82	-34.62%	7.74e-04	D
PPP6R2	137	720	302	-20.41%	7.78e-04	D
FLT3LG	103	224	76	-35.62%	8.05e-04	D
DKC1	22	23	1	-91.75%	8.39e-04	D
HIST1H2BD	22	23	1	-91.75%	8.39e-04	D
GPD2	70	84	20	-54.82%	8.73e-04	D
PRKCH	90	143	43	-42.94%	8.76e-04	D
OBSCN	60	89	22	-53.10%	9.07e-04	D
ALG9	29	31	3	-81.64%	9.07e-04	D
IFT140	126	401	155	-26.66%	9.43e-04	DA
PYM1	80	122	35	-45.56%	1.02e-03	D
ABCA7	48	63	13	-60.85%	1.02e-03	D
MECP2	131	703	296	-20.11%	1.09e-03	D
SPHK2	66	88	22	-52.56%	1.22e-03	D
FKBP5	128	601	249	-21.38%	1.30e-03	D
TAP2	88	165	53	-39.05%	1.31e-03	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.

gene	Gene symbol.
subjects	Total number of subjects with an integration near gene.
lateCount	Number of integration sites recovered after 0 days.
maxAbund	Maximum estimated clonal abundance observed after 0 days.
maxRelAbund	Maximum relative sample clonal abundance observed after 0 days.
categories	DEAL categories associated with gene.

Table 4. Top 100 abundant genes.

gene	subjects	lateCount	maxAbund	maxRelAbund	categories
TET2	73	60	858	99.00%	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%	AL
C1orf159	119	125	176	11.21%	AL
IFNGR2	9	9	173	45.58%	EA
RC3H1	82	75	164	4.42%	A
PCNX1	111	113	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%	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	92	86	3.70%	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%	DA
BCAS3	96	69	81	32.14%	AL
IQCB1	46	19	80	2.86%	A
KANSL1	110	134	79	3.30%	AL
WWOX	68	51	78	69.03%	AL
LOC100294362	90	69	76	2.47%	AL
RNF213	126	241	76	2.86%	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%	EAL
SEC31A	64	40	65	3.23%	A
SMAP2	70	48	65	0.46%	A
GPN1	44	16	63	1.82%	A
EARS2	16	7	60	28.17%	A
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

Table 4. Top 100 abundant genes. (*continued*)

gene	subjects	lateCount	maxAbund	maxRelAbund	categories
MEMO1	60	24	52	1.22%	A
PTBP1	102	116	52	4.48%	EA
FOXP1	120	180	49	36.57%	EAL
PPP3CC	120	117	49	52.13%	DA
DNM2	111	118	48	4.40%	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%	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	36	14	40	2.86%	A
LUC7L	135	357	39	22.41%	AL
PHF20	93	60	39	1.96%	DA
RNF157	134	485	39	7.58%	AL
DNAJC1	64	36	38	1.03%	A
ELMO1	88	45	38	0.98%	A
MED13L	112	150	38	2.38%	EAL
PA2G4	61	45	38	1.25%	A
CXorf67	25	16	37	2.17%	A
HERC4	77	42	37	2.14%	A
MAD1L1	83	62	37	1.82%	AL
VAV1	133	416	37	15.69%	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%	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%	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%	EA

Genes are categorized as Longitudinal if  $\geq 3$  different integrations across at least 3 different subjects are observed  $\geq 90$  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 $\geq 90$ days.
longitudinalSubjects	number of subjects associated with integrations associated with gene recovered $\geq 90$ days.
longitudinalTimePoints	number of time points sampled $\geq 90$ days.
latestTimePointDays	last time point sampled containing integrations in gene.

Table 5. Top 100 longitudinally persistent genes.

gene	totalSites	longitudinalSites	longitudinalSubjects	longitudinalTimePoints	latestTimePointDays
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
CD96	163	5	4	6	4015
AP2A2	229	4	3	2	4015
EHBP1	112	4	4	4	4015
HIVEP3	149	4	3	4	4015
FOXP1	431	5	4	12	3285
CENPP	100	4	4	4	3285
DYM	205	4	3	3	3285
GAK	176	4	3	3	3285
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
KIFC1	159	4	4	4	3042
KLHL6	35	4	3	3	3042
SDHC	100	4	3	4	3042
SHISA6	10	4	4	3	3042
WDR45B	170	4	3	4	3042
ZNF254	55	4	4	5	3042
C17orf53	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
UBE2H	135	3	3	3	3042
RNF157	1361	15	7	7	2920
VAPB	39	4	3	2	2920
FANCA	1757	19	12	5	2555
FCHSD2	704	17	9	7	2555
ANKRD11	807	14	10	9	2555

Table 5. Top 100 longitudinally persistent genes. (*continued*)

gene	totalSites	longitudinalSites	longitudinalSubjects	longitudinalTimePoints	latestTimePointDays
SMARCC1	736	14	9	8	2555
PPP6R3	642	13	9	4	2555
SMG1P5	500	10	7	6	2555
VAV1	981	9	5	8	2555
AP3B1	177	8	7	3	2555
MTOR	290	8	6	7	2555
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
RAP1GDS1	129	6	4	4	2555
ZBTB8OS	49	6	3	5	2555
ZZEF1	706	6	5	19	2555
BAZ1A	182	5	4	3	2555
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
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
UBR3	151	3	3	2	2555
SMG1P2	418	10	5	6	2372
RBFOX1	50	7	5	6	2372
ARHGAP15	533	5	4	3	2372
DDX58	91	5	4	3	2372
IQGAP1	385	5	5	4	2372
POLR2A	729	5	3	4	2372
SARNP	469	4	4	3	2372
A1BG	16	3	3	2	2372
AKAP10	124	3	3	2	2372
FAM126A	105	3	3	3	2372
TCERG1	111	3	3	3	2372
EHMT1	858	14	9	6	2190