Supplementary Material

In the supplementary material, we first present the Supplementary Figures, then Supplementary Tables. These are followed by four reports on genes of interest called as enriched or depleted by multiple criteria in different comparisons. We conclude with a short Supplementary Analysis.

Supplementary Figures

Figure S1. Favored integration in active transcription units. (Top) Favoring of integration in CR/PRtd in active transcription units. RNA-seq data was used to annotate gene activity (as reads per kilobase per million; RPKM), then genes were distributed into 10 bins with equal numbers of genes, ranked from lowest to highest RPKM (x-axis). Integration sites (fragments called by sonicAbundance) were then distributed into those bins and scored (y-axis). (Bottom) As in (Top), but using RNA-seq data and integration site data for PR/NR. RNA-seq data used in this study was from Fraietta et al., Nat. Med 24, 563-571, 2018.

Figure S2. Assessment of genes shared by expanded clones from infusion products (TDN), day 28 samples from CR/PRtd, and day 28 samples from PR/NR patients. The top 50 clones were selected from each population (data from Figure 3 A-C). Genes associated with expanded clones only with TDN products: NEMF, ZNF34, RMDN1, TYROBP, OTUD5, SURF4, AQR, ZSWIM1, PPP6C, SFMBT2, TSR1, ARAF, KIF2C,

WDR82, DNM2, LARP4B, TMEM135, THOC6, ABCA7, MYO9B, NFKBIL1, VPS51, PMEL, MUM1, LPIN2, GAB3, IRAK1BP1, SPDYE3, HAGH, NVL, EIF2B3, STARD7, GARS, FCHSD2, DLEU2, LIPC, ZNF682, STAU1, SKI, LZTR1, GMDS, EHD1, NLRC3, SMCHD1, KDM4A, MSH4, ETF1, FAM193B, MED13L, SPTSSA, SMG6, CD3EAP, MAP4K1, PPP6R2, SEC14L1, ETV3, POLH, FOXO3, ZNF213, CBFA2T3, SLC6A16, STAT5B. Genes associated with expanded clones only with day 28 products from CR/PRtd patients: KCTD3, PATL1, SRCAP, SSH2, SNHG12, ATP2A2, CARD8, PIKFYVE, ZZEF1, VAV1, PDCD11, RABEP1, PPP1CB, GFOD2, AP4M1, OGFOD1, SPATA5, CDK11B, FOPNL, CPEB2, YLPM1, MVB12B, MARK4, JAK3, KANTR, ATAT1, PPP6R3, SLC38A10, COL4A3BP, ZNF251, XPO5, ATP6V0C, USP14, ZNF563, ZNF565, PGD, NUP214, CCP110, CBFB, TRUB1, TMTC2, MED13, MCM3AP-AS1, SRSF11, MRPS10, SUPT3H, LOC102724064, CTCF. Genes associated with expanded clones only with day 28 products from PR/NR patients: IFNGR2, LTA4H, TRIM37, GUF1, STK4, LOC100128568, DPYD, TULP2, POLR2A, EIF3E, BUB1B, SLC3A2, MAN1B1, LRRK2, CD96, ATF7IP, AKT2, MAN2A1, MCF2L2, CD69, SBNO2, AKT3, OXR1, NCOR1, CASK, ZNF609, TNIP3, TPR, LINC02015, SMARCA2, EDRF1, NPRL3, IL21R, CTC1, CPNE1, CHIC1, TRAPPC12, SLC8A1-AS1, DAG1, ANP32B, CWF19L2, NAA30, SF3A2, FAM217B, GGA3, KANSL3, CDC73, NSD1, SYNE1, UBAP2, GNB1, SF1, CCDC137, TENM1, KDM1A, TMEM57, MACF1, DHX9, C1orf27, FAM168B, SLC16A14, LOC101929411, GNL1, PCM1, CHORDC1, NCAPD3, NECAP1, RPAP3, SCYL2, TTC8, LOC100128108, NFATC2IP, GINS2, PIGN, RP2. Genes associated in expanded clones in both TDN products and day 28 products from CR/PRtd patients: LEF1, NPLOC4, SAFB, NAA38, PRKD2. Genes

associated with expanded clones in both TDN products and day 28 products from PR/NR patients: RBX1. Genes associated with expanded clones in both CR/PRtd and PR/NR patients on day 28 samples: CLEC16A.

Figure S3. Integration sites in cells undergoing clonal expansion are found recurrently in specific host genes across multiple patients. We sought to test the possibility that the same clones expanded in abundance in multiple CART19 patients. To assess this, we first selected the top 5% most abundant clones from each patient (based on SonicAbundance). We then plotted how often the genes marked by integration were found in the top 5% most abundant clones in other patients. In the figure, the number of patients sharing an expanded clone is shown on the x-axis, and the y-axis shows the number of clones shared overall for the given number of patients sharing clones. For example, a total of 18 genes (y-axis) were shared across 7 patients (x-axis). To analyze the background of sharing expected by chance (taking account of differential integration frequency in different transcription units), we sampled a random 5% of genes from each patient, and repeated the overlap calculation. This was repeated 1000 times, and the resulting permutation background plotted (red dots). Error bars indicate Standard Deviation. Thus the observed sharing is far greater than that expected by chance as reported by the permutation background. Specific genes detected are as follows: Observed across 7 patients: CBFB, CBLB, CREBBP, EPB41, GNB1, HN1L, LPP, MECP2, NFATC3, PPP6R2, PPP6R3, RPTOR, SAFB, SAFB2, SMG1, TNRC6B, VAV1, XPO5. Observed across 8 patients: CLEC16A, EP300, HNRNPUL2- BSCL2, MIR5096,

NSD1, UTRN. Observed across 9 patients: DNMT1, KDM2A, RNF157. Clones within NPLOC4 were observed in 11 patients and clones within FANCA were observed in 12 patients.

Figure S4. Gene ontology analysis of genes called as influencing T cell proliferation. Heatmaps indicating the proportion of each gene ontology term (**A**) or KEGG pathway (**B**). The four criteria for calling genes were each analyzed separately and are shown in the columns. Asterisks indicate significant enrichment for the term over random distributions (Fisher's exact test based on noncentral hypergeometric distributions). The rows show GO terms (**A**) or KEGG pathways (**B**) called as enriched by at least one criteria.

Figure S5. Comparing clonal diversity between TCR-beta sequencing results from CART19 sorted T-cells and vector integration sites from peripheral blood (measured by Shannon Index). P-values were determined from the Spearman correlation. Blue points indicate data from CR/PRtd patients, red points indicate data from PR/NR patients. Circles represent patients with CLL, triangles represent patients with ALL (upward pointing, Adult ALL; downward pointing, Pediatric ALL). All TCR-beta sequencing data was collected from CAR+ sorted T-cells. T cell subsets analyzed included CD3+, CART19+, and in some cases CD4+-sorted or CD8+-sorted. An comparison of TCR diversity before transplantation and after transplantation yielded a non-significant Spearman correlation.

Figure S6. Model optimization based on vector integration sites from infusion products or day 28 samples. A total of 91 features (Table S6) spanning population metrics, genomic features, and epigenetic features from 29 patients were used in least absolute shrinkage and selection operator (LASSO) logistic regression to build a classification model. Results from leave-one-out cross-validation of models based on transduction/pre-infusion products (A) and day 28 peripheral blood samples (B). The top horizontal numbers indicate the number of principal components used in the classification model, while the x-axis indicates the lambda constant used for tuning the model, and the y-axis shows misclassification error. Error bars indicated standard error. The minimum value of misclassification is indicated to the right of the plots. A similar comparison of data for ALL subjects versus CLL subjects is in Figure S8.

Figure S7. Central Memory T-cell proportions within Infusion Products. Proportion of central memory T-cells are shown for infusion products of CLL and ALL patients. Central Memory T-cells frequencies were identified at apheresis by the following markers: CD8+, CCR7+, and CD45RO+. Mean and standard error of the associated data are shown. No significant difference was identified between response groups within ALL or CLL as determined by a two-tailed T-test.

Figure S8. Disease specific analysis of integration site profiles. A LASSO regression was used to identify variables within integration site profiles that would separate CLL and ALL patients. Analysis was performed on infusion products, and the percent contribution to the overall variance between the groups is displayed for values greater than 1%.

Supplementary Tables

Table S1. Patient metadata and results of integration site sequencing.

Table S2. Summary of patient ages in the CLL and ALL cohorts

Table S3. Patient DNA samples analyzed by integration site sequencing.

Table S4. Gene sets used for analysis, including cancer associated genes, human lymphoid cancer genes, and genes associated with clonal hematopoiesis, and overlaps between lists. Origins of gene lists are as follows: Bushman Lab allOnco List: (Sadelain et al., 2011, http://www.bushmanlab.org/links/genelists); Lymphoid cancer list: (personal communication, Marina Cavazzana, http://www.bushmanlab.org/links/genelists); COSMIC: (Sondka, Z. *et al.* 2018 Nature Cancer Reviews), TCGA_Xie and Clonal_Hema_Xie (Xie, M. *et al.* 2014 Nature Medicine).

Table S5. Statistical analysis of overlaps among genes called at integration sites and cancer associated gene lists in Table S4. The references for each list are included in the table.

Table S6. 91 model variables used in the LASSO regression analysis.

Table S7. Sources of antibodies used.

Specificity	Clone	Conjugate	Supplier	Catalog Number
CD3	OKT3	Brilliant Violet (BV) 605	BioLegend	317322
CAR19	136.20.1	Alexa Fluor (AF) 647	Kind gift of B. Jena and L. Cooper (MD Anderson Cancer Center)	--
CD8	RPA-T8	BV650	BioLegend	301042
CD4	OKT4	BV785	BioLegend	317442
CD45	2D1	Phycoerythrin BE	BioLegend	368510

Supplementary Reports.

Supplementary report 1: Genes of interest marked by vector integration from all patients.

Supplementary report 2: Genes of interest marked by vector integration from CLL patients.

Supplementary report 3: Genes of interest marked by vector integration from ALL patients.

Supplementary report 4: Genes of interest marked by vector integration from CR/PRtd patients.

Supplementary Analysis

Supplementary analysis: effects of age. We compared data on age and the proportion of central memory cells and found that younger patients have lower proportions of

central memory cells (Figure S7). However, within ALL or CLL there was no association of age versus response and nonresponse. Regarding integration site distributions, a significant difference could be detected between the ALL and CLL subjects using a LASSO regression model (Figure S8).

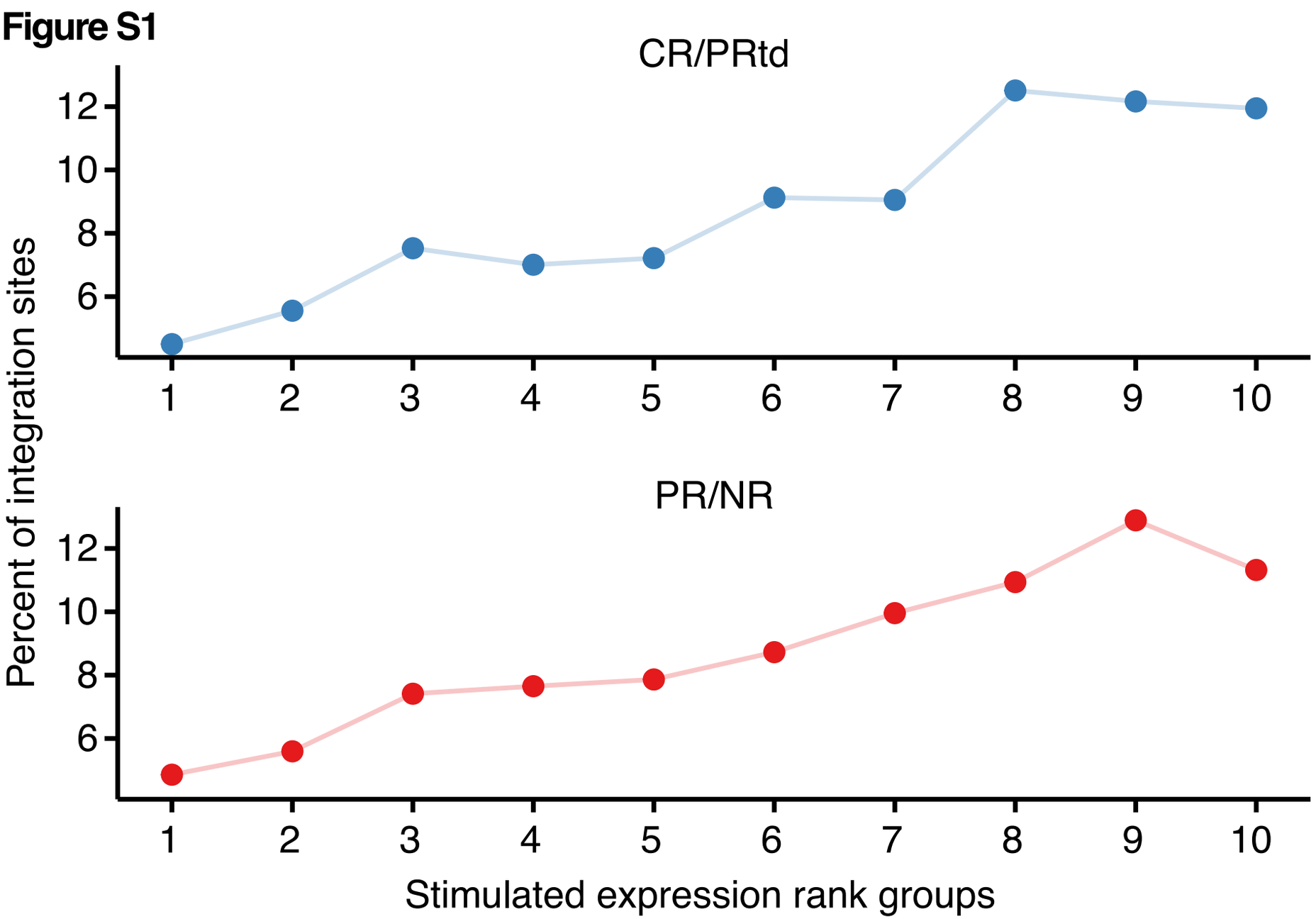

Figure S1. Favored integration in active transcription units. Figure S1. Favored integration in active transcription units. (Top) Favoring of integration in CR/PRtd in active transcription units. RNA-seq data was used to annotate gene activity (as reads per kilobase per million; RPKM), then genes were distributed into 10 bins with equal numbers of genes, ranked from lowest to highest RPKM (x-axis). Integration sites (fragments called by sonicAbundance) were then distributed into those bins and scored (y-axis). (Bottom) As in (Top), but using RNA-seq data and integration site data for PR/NR. RNA-seq data used in this study was from Fraietta et al., Nat. Med 24, 563-571, 2018.

Figure S2

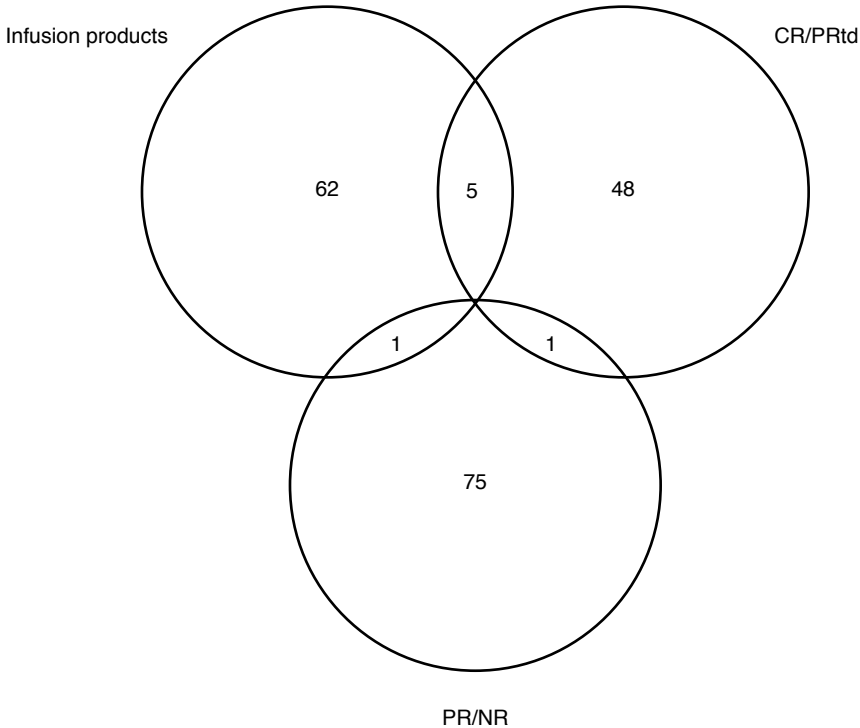

Figure S2. Assessment of genes shared by expanded clones from infusion products (TDN), day 28 samples from CR/PRtd, and day 28 samples from PR/NR patients. The top 50 clones were selected from each population (data from Figure 3 A-C). Genes associated with expanded clones only with TDN products: NEMF, ZNF34, RMDN1, TYROBP, OTUD5, SURF4, AQR, ZSWIM1, PPP6C, SFMBT2, TSR1, ARAF, KIF2C, WDR82, DNM2, LARP4B, TMEM135, THOC6, ABCA7, MYO9B, NFKBIL1, VPS51, PMEL, MUM1, LPIN2, GAB3, IRAK1BP1, SPDYE3, HAGH, NVL, EIF2B3, STARD7, GARS, FCHSD2, DLEU2, LIPC, ZNF682, STAU1, SKI, LZTR1, GMDS, EHD1, NLRC3, SMCHD1, KDM4A, MSH4, ETF1, FAM193B, MED13L, SPTSSA, SMG6, CD3EAP, MAP4K1, PPP6R2, SEC14L1, ETV3, POLH, FOXO3, ZNF213, CBFA2T3, SLC6A16, STAT5B. Genes associated with expanded clones only with day 28 products from CR/PRtd patients: KCTD3, PATL1, SRCAP, SSH2, SNHG12, ATP2A2, CARD8, PIKFYVE, ZZEF1, VAV1, PDCD11, RABEP1, PPP1CB, GFOD2, AP4M1, OGFOD1, SPATA5, CDK11B, FOPNL, CPEB2, YLPM1, MVB12B, MARK4, JAK3, KANTR, ATAT1, PPP6R3, SLC38A10, COL4A3BP, ZNF251, XPO5, ATP6V0C, USP14, ZNF563, ZNF565, PGD, NUP214, CCP110, CBFB, TRUB1, TMTC2, MED13, MCM3AP-AS1, SRSF11, MRPS10, SUPT3H, LOC102724064, CTCF. Genes associated with expanded clones only with day 28 products from PR/NR patients: IFNGR2, LTA4H, TRIM37, GUF1, STK4, LOC100128568, DPYD, TULP2, POLR2A, EIF3E, BUB1B, SLC3A2, MAN1B1, LRRK2, CD96, ATF7IP, AKT2, MAN2A1, MCF2L2, CD69, SBNO2, AKT3, OXR1, NCOR1, CASK, ZNF609, TNIP3, TPR, LINC02015, SMARCA2, EDRF1, NPRL3, IL21R, CTC1, CPNE1, CHIC1, TRAPPC12, SLC8A1-AS1, DAG1, ANP32B, CWF19L2, NAA30, SF3A2, FAM217B, GGA3, KANSL3, CDC73, NSD1, SYNE1, UBAP2, GNB1, SF1, CCDC137, TENM1, KDM1A, TMEM57, MACF1, DHX9, C1orf27, FAM168B, SLC16A14, LOC101929411, GNL1, PCM1, CHORDC1, NCAPD3, NECAP1, RPAP3, SCYL2, TTC8, LOC100128108, NFATC2IP, GINS2, PIGN, RP2. Genes associated in expanded clones in both TDN products and day 28 products from CR/PRtd patients: LEF1, NPLOC4, SAFB, NAA38, PRKD2. Genes associated with expanded clones in both TDN products and day 28 products from PR/NR patients: RBX1. Genes associated with expanded clones in both CR/PRtd and PR/NR patients on day 28 samples: CLEC16A.

Figure S3

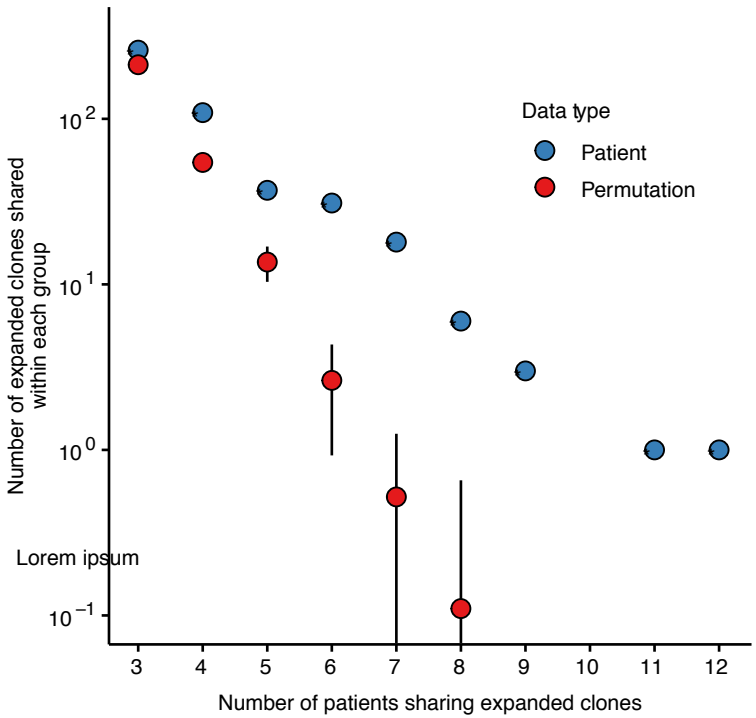

Figure S3. Integration sites in cells undergoing clonal expansion are found recurrently in specific host genes across multiple patients. We sought to test the possibility that the same clones expanded in abundance in multiple CART19 patients. To assess this, we first selected the top 5% most abundant clones from each patient (based on SonicAbundance). We then plotted how often the genes marked by integration were found in the top 5% most abundant clones in other patients. In the figure, the number of patients sharing an expanded clone is shown on the x-axis, and the y-axis shows the number of clones shared overall for the given number of patients sharing clones. For example, a total of 18 genes (y-axis) were shared across 7 patients (x-axis). To analyze the background of sharing expected by chance (taking account of differential integration frequency in different transcription units), we sampled a random 5% of genes from each patient, and repeated the overlap calculation. This was repeated 1000 times, and the resulting permutation background plotted (red dots). Error bars indicate Standard Deviation. Thus the observed sharing is far greater than that expected by chance as reported by the permutation background. Specific genes detected are as follows: Observed across 7 patients: CBFB, CBLB, CREBBP, EPB41, GNB1, HN1L, LPP, MECP2, NFATC3, PPP6R2, PPP6R3, RPTOR, SAFB, SAFB2, SMG1, TNRC6B, VAV1, XPO5. Observed across 8 patients: CLEC16A, EP300, HNRNPUL2- BSCL2, MIR5096, NSD1, UTRN. Observed across 9 patients: DNMT1, KDM2A, RNF157. Clones within NPLOC4 were observed in 11 patients and clones within FANCA were observed in 12 patients.

Figure S4

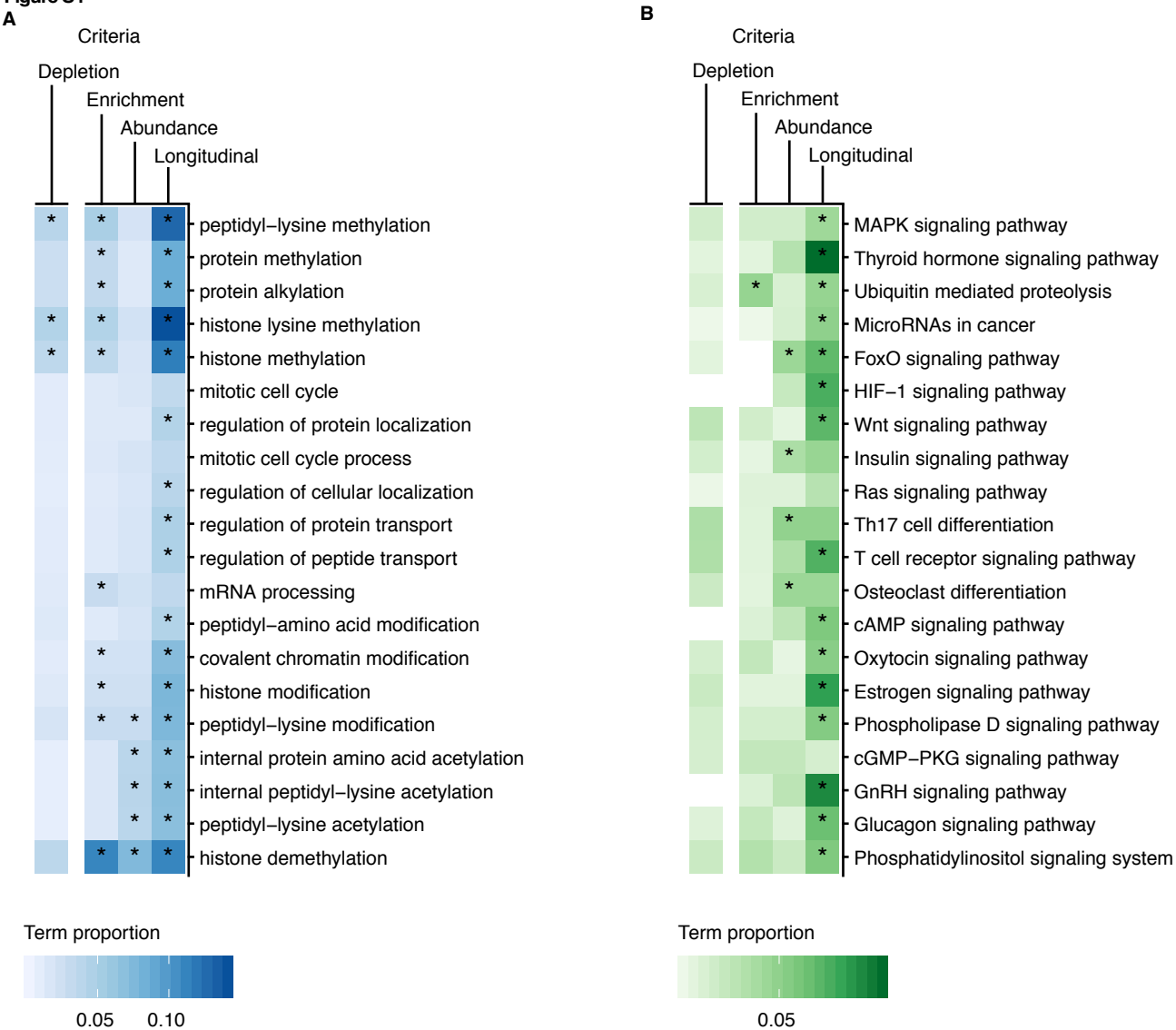

Figure S4. Gene ontology analysis of genes called as influencing T cell proliferation. Heatmaps indicating the proportion of each gene ontology term (A) or KEGG pathway (B). The four criteria for calling genes were each analyzed separately and are shown in the columns. Asterisks indicate significant enrichment for the term over random distributions (Fisher's exact test based on noncentral hypergeometric distributions). The rows show GO terms (A) or KEGG pathways (B) called as enriched by at least one criteria.

Figure S5

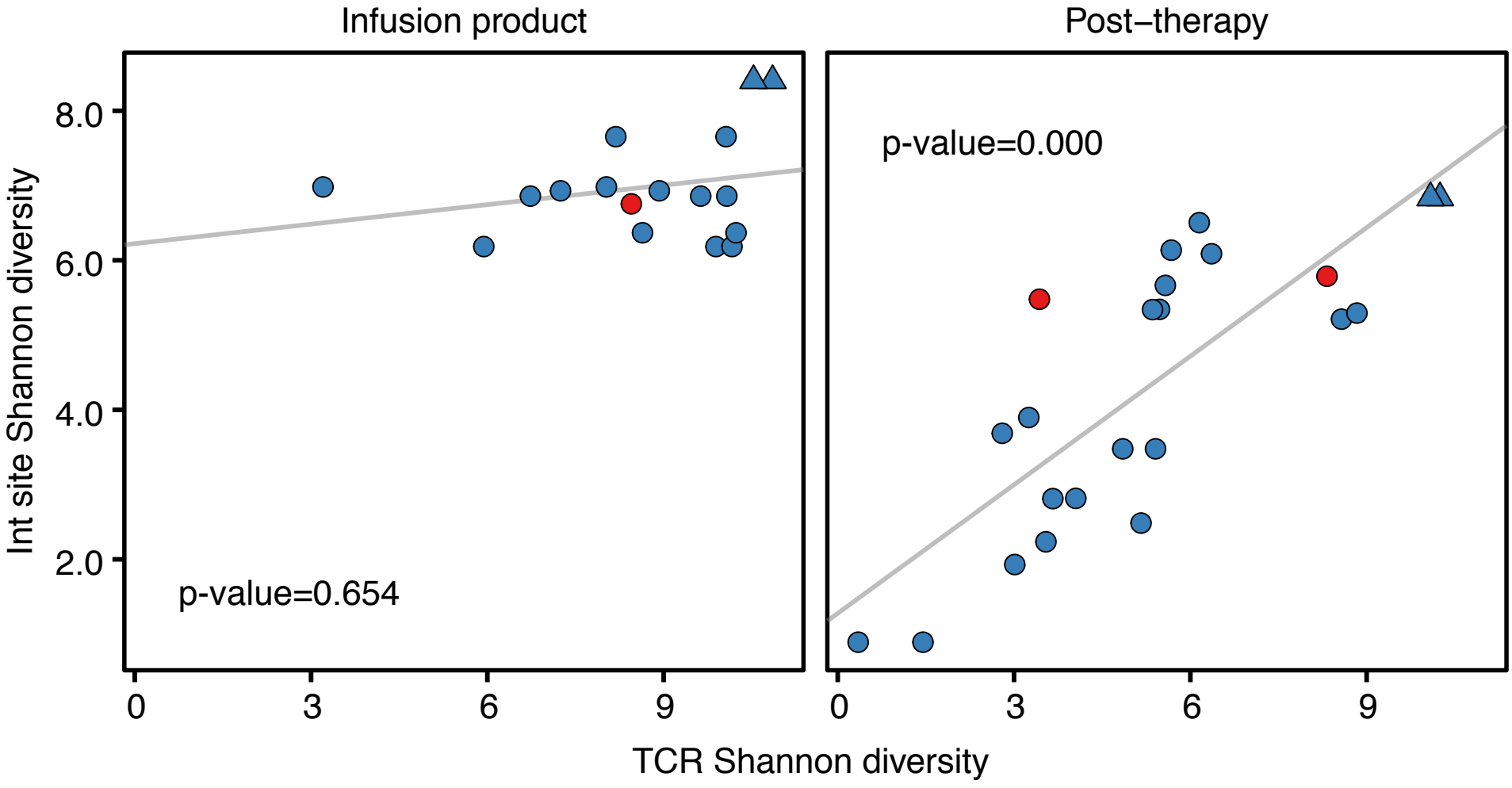

Figure S5. Comparing clonal diversity between TCR-beta sequencing results from CART19 sorted T-cells and vector integration sites from peripheral blood (measured by Shannon Index). P-values were determined from the Spearman correlation. Blue points indicate data from CR/PRtd patients, red points indicate data from PR/NR patients. Circles represent patients with CLL, triangles represent patients with ALL (upward pointing, Adult ALL; downward pointing, Pediatric ALL). All TCR-beta sequencing data was collected from CAR+ sorted T-cells. T cell subsets analyzed included CD3+, CART19+, and in some cases CD4+-sorted or CD8+-sorted. An comparison of TCR diversity before transplantation and after transplantation yielded a non-significant Spearman correlation.

Figure S6

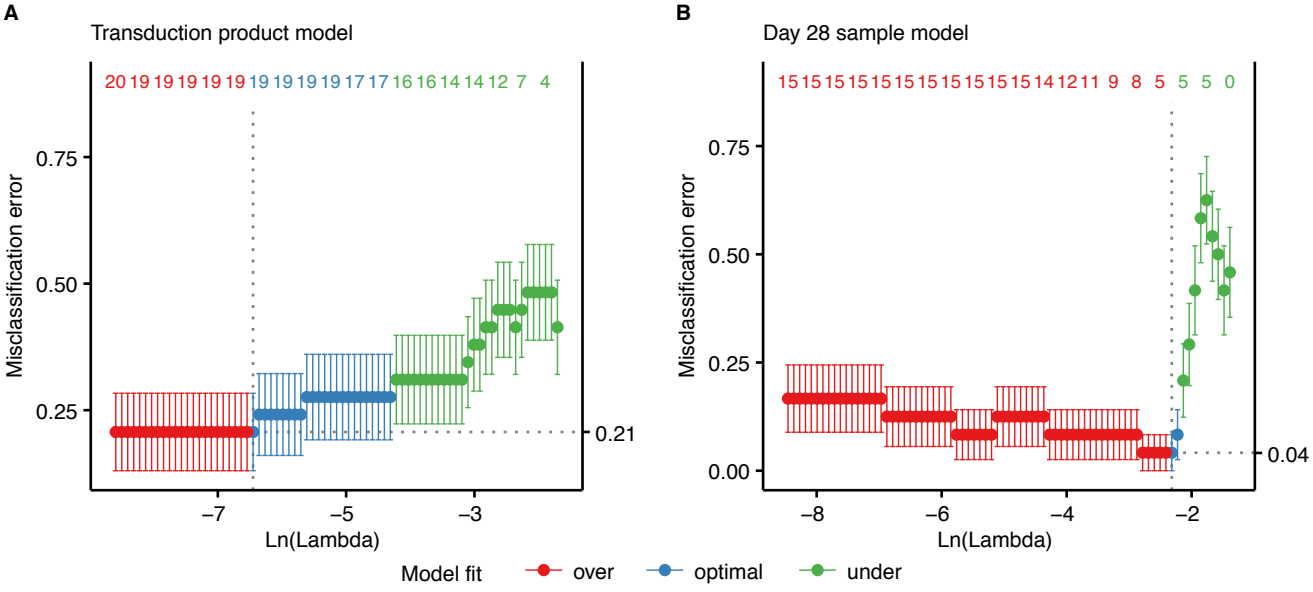

Figure S6. Model optimization based on vector integration sites from infusion products or day 28 samples. A total of 91 features (Table S6) spanning population metrics, genomic features, and epigenetic features from 29 patients were used in least absolute shrinkage and selection operator (LASSO) logistic regression to build a classification model. Results from leave-one-out cross-validation of models based on transduction/pre-infusion products (A) and day 28 peripheral blood samples (B). The top horizontal numbers indicate the number of principal components used in the classification model, while the x-axis indicates the lambda constant used for tuning the model, and the y-axis shows misclassification error. Error bars indicated standard error. The minimum value of misclassification is indicated to the right of the plots.

Figure S7

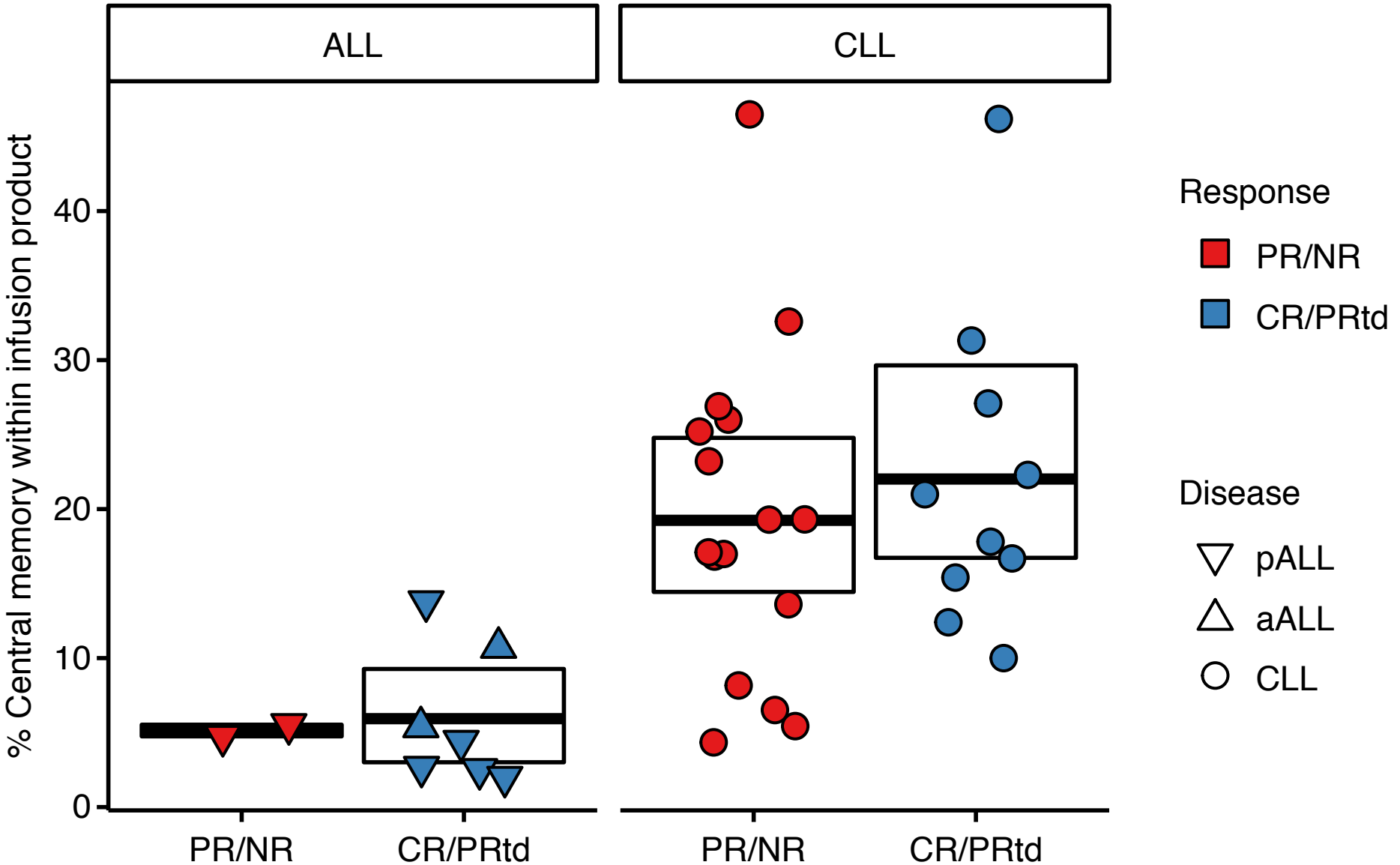

Figure S7. Central Memory T-cell proportions within Infusion Products. Proportion of central memory T-cells are shown for infusion products of CLL and ALL patients. Central Memory T-cells frequencies were identified at apheresis by the following markers: CD8+, CCR7+, and CD45RO+. Mean and standard error of the associated data are shown. No significant difference was identified between response groups within ALL or CLL as determined by a two-tailed T-test.

Figure S8

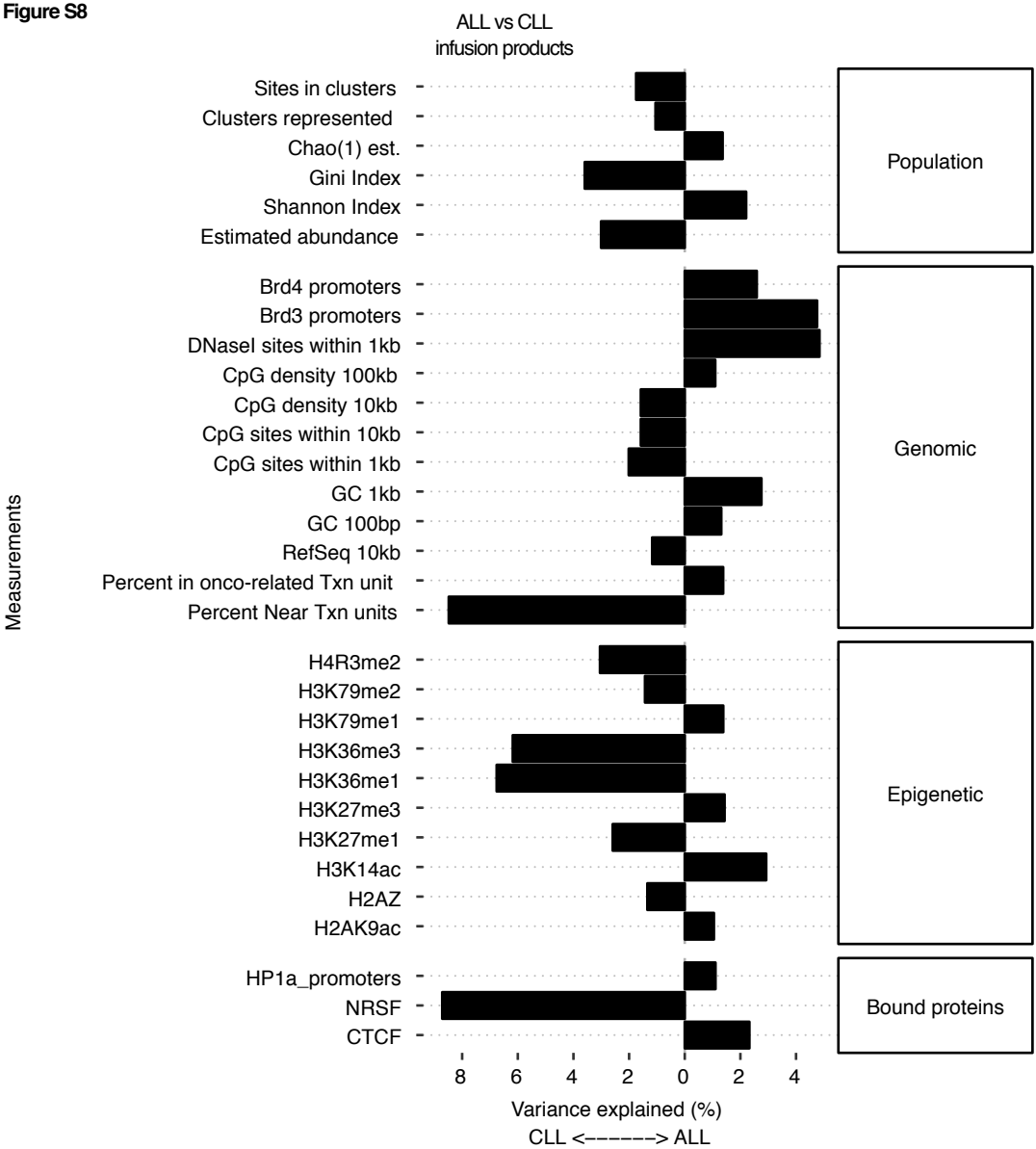

Figure S8. Disease specific analysis of integration site profiles. A LASSO regression was used to identify variables within integration site profiles that would separate CLL and ALL patients. Analysis was performed on infusion products, and the percent contribution to the overall variance between the groups is displayed for values greater than 1%.

Supplemental Tables

Supplemental Table S1. Patient Information.

Table S1. Patient metadata and results of integration site sequencing.

Patient ID	Disease	Clin. Trial	Response	Resp. Class	Reads	Cells Sampled	Unique Int. Sites
p959-100	pALL	CHP959	Complete	CR/PRtd	1929626	6832	6532
p959-101	pALL	CHP959	Complete w/ relapse	CR/PRtd	2161985	9088	7386
p959-103	pALL	CHP959	None	PR/NR	289620	1423	1355
p959-107	pALL	CHP959	Complete w/ relapse	CR/PRtd	763129	11076	10325
p959-112	pALL	CHP959	Complete	CR/PRtd	1311546	9749	9476
p959-117	pALL	CHP959	None	PR/NR	3420512	2578	2074
p959-133	pALL	CHP959	Complete w/ relapse	CR/PRtd	1118577	12533	12308
p03712-03	CLL	UPCC03712	Complete	CR/PRtd	5722000	28632	12140
p03712-04	CLL	UPCC03712	Complete	CR/PRtd	1887292	3698	2474
p03712-06	CLL	UPCC03712	Complete	CR/PRtd	8791604	7358	5890
p03712-08	CLL	UPCC03712	None	PR/NR	200363	497	331
p03712-11	CLL	UPCC03712	None	PR/NR	1436827	2510	1895
p03712-12	CLL	UPCC03712	None	PR/NR	632238	1969	1373
p03712-14	CLL	UPCC03712	None	PR/NR	85904	1034	931
p03712-16	CLL	UPCC03712	Partial	PR/NR	2511087	2951	2616
p03712-17	CLL	UPCC03712	None	PR/NR	97734	690	568
p03712-18	CLL	UPCC03712	Partial w/ TnDz	CR/PRtd	2391322	7596	5340
p03712-20	CLL	UPCC03712	None	PR/NR	507722	3332	2479
p03712-22	CLL	UPCC03712	Partial w/ TnDz	CR/PRtd	1803690	2592	2066
p03712-25	CLL	UPCC03712	None	PR/NR	820143	3409	2524
p03712-29	CLL	UPCC03712	Partial	PR/NR	2202312	4703	4095
p03712-45	CLL	UPCC03712	Complete	CR/PRtd	4296437	7404	4419
p04409-01	CLL	UPCC04409	Complete	CR/PRtd	4656521	11470	7104
p04409-02	CLL	UPCC04409	Complete	CR/PRtd	3889474	7979	2200
p04409-03	CLL	UPCC04409	Partial	PR/NR	1506913	3552	2973
p04409-05	CLL	UPCC04409	Partial	PR/NR	3349759	1566	1154
p04409-06	CLL	UPCC04409	None	PR/NR	508559	4596	4272
p04409-07	CLL	UPCC04409	None	PR/NR	532722	2489	2345
p04409-09	CLL	UPCC04409	Complete	CR/PRtd	5423944	12090	8427
p04409-10	CLL	UPCC04409	Complete	CR/PRtd	4519487	4109	1887
p04409-12	CLL	UPCC04409	Partial	PR/NR	2267562	3216	2765
p04409-14	CLL	UPCC04409	None	PR/NR	554038	763	608
p04409-17	CLL	UPCC04409	None	PR/NR	927522	2578	2110
p04409-18	CLL	UPCC04409	None	PR/NR	148258	393	275
p04409-22	CLL	UPCC04409	Partial w/ TnDz	CR/PRtd	2413960	5755	4833
p04409-23	aALL	UPCC04409	Complete	CR/PRtd	203455	1587	1296
p04409-25	CLL	UPCC04409	None	PR/NR	800419	871	830

(continued)

Patient ID	Disease	Clin. Trial	Response	Resp. Class	Reads	Cells Sampled	Unique Int. Sites
p04409-26	aALL	UPCC04409	Complete w/ relapse	CR/PRtd	119563	982	969
p04409-27	aALL	UPCC04409	Complete	CR/PRtd	122586	1128	1110
p04409-29	aALL	UPCC04409	Complete w/ relapse	CR/PRtd	172742	1483	1459
p18415-01	CLL	UPCC18415	Complete	CR/PRtd	944888	1925	1812
p18415-02	CLL	UPCC18415	Complete	CR/PRtd	871295	1438	1351
p18415-03	CLL	UPCC18415	Complete	CR/PRtd	944334	2914	2815
p18415-04	CLL	UPCC18415	Complete	CR/PRtd	640111	3934	3833
p18415-05	CLL	UPCC18415	Complete	CR/PRtd	771094	1539	1447
p18415-06	CLL	UPCC18415	Complete	CR/PRtd	850718	2917	2803
p18415-07	CLL	UPCC18415	Complete	CR/PRtd	591275	2510	2436
p18415-08	CLL	UPCC18415	Complete	CR/PRtd	874944	3036	2890
p18415-09	CLL	UPCC18415	Complete w/ relapse	CR/PRtd	918815	2335	2212
p18415-10	CLL	UPCC18415	Complete	CR/PRtd	948901	2311	2216
p18415-11	CLL	UPCC18415	Complete	CR/PRtd	1054162	3885	3721
p18415-12	CLL	UPCC18415	Complete	CR/PRtd	771060	2686	2543
p18415-13	CLL	UPCC18415	Complete	CR/PRtd	626232	3901	3809
p18415-15	CLL	UPCC18415	Complete	CR/PRtd	1558046	3285	2984
p18415-16	CLL	UPCC18415	Complete	CR/PRtd	875474	2000	1900
p18415-18	CLL	UPCC18415	None	PR/NR	481737	3682	3593
p18415-19	CLL	UPCC18415	Complete	CR/PRtd	544293	4236	4160
p18415-20	CLL	UPCC18415	Complete	CR/PRtd	686532	4382	4232

Supplemental Table S2. Summary of patient ages in the CLL and ALL cohorts.

Table S2. Summary of patient ages in the CLL and ALL cohorts. We did not observe a significant correlation between age and clinical reponse within disease groups.

Disease	Mean	Median	Std. Dev.	Std. Err.	Minimum	Maximum	Count
ALL	24.3	16	19.9	6.0	6	60	11
CLL	63.9	62	6.9	1.3	50	77	29

Supplemental Table S3. Patient DNA samples analyzed by integration site sequencing.

Table S3. Table of specimens analyzed by vector integration site analysis.

GTSP ID	Patient ID	Clin. Trial	Cell Type	Time Point	Reads	Cells Sampled	Unique Int. Sites
GTSP0560	p04409-10	UPCC04409	Tcells:CAR+CD8+	d121	685759	920	105
GTSP0561	p04409-01	UPCC04409	Tcells:CAR+CD8-	d14	732156	450	338
GTSP0562	p04409-01	UPCC04409	Tcells:CAR+CD8+	d14	609593	3552	1585
GTSP0563	p04409-02	UPCC04409	Tcells:CAR+CD8-	d28	498286	1044	321
GTSP0564	p04409-02	UPCC04409	Tcells:CAR+CD8+	d28	416315	2222	237
GTSP0565	p04409-09	UPCC04409	Tcells:CAR+CD8-	d14	269567	258	221
GTSP0566	p04409-09	UPCC04409	Tcells:CAR+CD8+	d14	704330	4109	3029
GTSP0567	p03712-03	UPCC03712	Tcells:CAR+CD8-	d14	341591	7878	3621
GTSP0568	p03712-03	UPCC03712	Tcells:CAR+CD8+	d14	1891206	15778	6990
GTSP0569	p03712-06	UPCC03712	Tcells:CAR+CD8-	d14	802600	444	276
GTSP0570	p03712-06	UPCC03712	Tcells:CAR+CD8+	d14	834877	2438	2095
GTSP0571	p03712-45	UPCC03712	Tcells:CAR+CD8-	d14	315375	34	26
GTSP0572	p03712-45	UPCC03712	Tcells:CAR+CD8+	d14	724342	3208	1630
GTSP0573	p03712-03	UPCC03712	Tcells	d0	261340	748	559
GTSP0574	p03712-03	UPCC03712	PBL	d28	143181	108	72
GTSP0575	p03712-04	UPCC03712	Tcells	d0	412136	2052	1547
GTSP0576	p03712-04	UPCC03712	PBL	d28	60461	174	107
GTSP0577	p03712-06	UPCC03712	Tcells	d0	75761	1101	1084
GTSP0578	p03712-06	UPCC03712	PBL	d28	91347	166	157
GTSP0579	p03712-08	UPCC03712	Tcells	d0	200362	496	330
GTSP0580	p03712-08	UPCC03712	PBL	d28	1	1	1
GTSP0581	p03712-11	UPCC03712	Tcells	d0	316702	2011	1621
GTSP0582	p03712-11	UPCC03712	PBL	d21	158508	18	5
GTSP0583	p03712-11	UPCC03712	PBL	d23	398869	223	177
GTSP0584	p03712-11	UPCC03712	PBL	d25	502693	170	56
GTSP0585	p03712-11	UPCC03712	PBL	d28	40148	83	45
GTSP0586	p03712-11	UPCC03712	PBL	d36	19907	5	5
GTSP0587	p03712-12	UPCC03712	Tcells	d0	631320	1950	1358
GTSP0588	p03712-12	UPCC03712	PBL	d28	918	19	15
GTSP0589	p03712-14	UPCC03712	Tcells	d0	82386	986	886
GTSP0590	p03712-14	UPCC03712	PBL	d28	3518	48	45
GTSP0591	p03712-16	UPCC03712	Tcells	d0	83471	879	865
GTSP0592	p03712-16	UPCC03712	PBL	d28	197785	230	202
GTSP0593	p03712-17	UPCC03712	Tcells	d0	97323	661	540
GTSP0594	p03712-17	UPCC03712	PBL	d28	411	29	28
GTSP0595	p03712-18	UPCC03712	Tcells	d0	120382	602	588
GTSP0596	p03712-18	UPCC03712	PBL	d28	105352	362	292

(continued)

GTSP ID	Patient ID	Clin. Trial	Cell Type	Time Point	Reads	Cells Sampled	Unique Int. Sites
GTSP0597	p03712-20	UPCC03712	Tcells	d0	507625	3305	2452
GTSP0598	p03712-20	UPCC03712	PBL	d28	97	27	27
GTSP0599	p03712-22	UPCC03712	Tcells	d0	280944	2013	1580
GTSP0600	p03712-22	UPCC03712	PBL	d28	5080	31	29
GTSP0601	p03712-25	UPCC03712	Tcells	d0	819562	3387	2510
GTSP0602	p03712-25	UPCC03712	PBL	d28	581	22	14
GTSP0603	p04409-09	UPCC04409	Tcells	d0	1304491	4140	2777
GTSP0604	p04409-09	UPCC04409	PBL	d28	714619	1289	857
GTSP0605	p04409-10	UPCC04409	Tcells	d0	328324	1381	1076
GTSP0606	p04409-10	UPCC04409	PBL	d10	1371	18	15
GTSP0607	p04409-10	UPCC04409	PBL	d28	285	7	6
GTSP0608	p04409-10	UPCC04409	PBL	d120	52261	113	17
GTSP0609	p04409-10	UPCC04409	PBL	d204	23391	54	10
GTSP0610	p04409-10	UPCC04409	PBL	d442	1892	14	9
GTSP0611	p04409-10	UPCC04409	PBL	d801	21	8	8
GTSP0613	p04409-12	UPCC04409	PBL	d28	102039	390	334
GTSP0614	p04409-17	UPCC04409	Tcells	d0	309822	2089	1715
GTSP0615	p04409-17	UPCC04409	PBL	d28	847	25	25
GTSP0616	p04409-18	UPCC04409	Tcells	d0	146580	364	257
GTSP0617	p04409-18	UPCC04409	PBL	d28	1676	27	16
GTSP0619	p04409-22	UPCC04409	PBL	d28	107564	197	187
GTSP0620	p04409-23	UPCC04409	Tcells	d0	201134	1553	1270
GTSP0621	p04409-23	UPCC04409	PBL	d28	2321	34	26
GTSP0624	p04409-26	UPCC04409	Tcells	d0	100927	975	963
GTSP0625	p04409-26	UPCC04409	PBL	d28	18636	7	6
GTSP0626	p04409-27	UPCC04409	Tcells	d0	75831	1122	1106
GTSP0627	p04409-27	UPCC04409	PBL	d28	46755	6	4
GTSP0628	p04409-29	UPCC04409	Tcells	d0	88354	203	194
GTSP0629	p04409-29	UPCC04409	PBL	d28	84388	1280	1265
GTSP0630	p959-100	CHP959	Tcells	d0	97109	187	167
GTSP0631	p959-100	CHP959	PBL	d13	824668	1072	971
GTSP0632	p959-100	CHP959	PBL	d15	829124	887	784
GTSP0633	p959-100	CHP959	PBL	d28	2374	121	121
GTSP0634	p959-103	CHP959	Tcells	d0	186658	1356	1308
GTSP0635	p959-103	CHP959	PBL	d14	102426	66	46
GTSP0636	p959-103	CHP959	PBL	d28	536	1	1

(continued)

GTSP ID	Patient ID	Clin. Trial	Cell Type	Time Point	Reads	Cells Sampled	Unique Int. Sites
GTSP0638	p959-107	CHP959	PBL	d14	64690	47	40
GTSP0639	p959-107	CHP959	PBL	d28	40270	25	12
GTSP0640	p959-117	CHP959	Tcells	d0	870169	1390	1163
GTSP0641	p959-117	CHP959	PBL	d14	764624	693	572
GTSP0642	p959-117	CHP959	PBL	d25	847373	184	156
GTSP0643	p959-100	CHP959	Tcells:CAR+	d0	176351	4565	4492
GTSP0644	p959-101	CHP959	Tcells:CAR+	d0	180832	3299	3213
GTSP0645	p959-101	CHP959	Tcells	d0	1015501	5476	3954
GTSP0646	p959-101	CHP959	PBL	m2	899	11	9
GTSP0647	p959-107	CHP959	Tcells:CAR+	d0	416274	7613	7095
GTSP0648	p959-107	CHP959	Tcells	d0	239525	3252	3051
GTSP0649	p959-107	CHP959	BM:CAR+	m9	170	34	34
GTSP0650	p959-107	CHP959	BM	m9	176	43	41
GTSP0651	p959-107	CHP959	PBL	m20	2024	62	59
GTSP0652	p959-112	CHP959	Tcells:CAR+	d0	232831	5341	5238
GTSP0653	p959-112	CHP959	Tcells	d0	206973	4248	4137
GTSP0654	p959-133	CHP959	Tcells:CAR+	d0	156806	6156	6065
GTSP0655	p959-133	CHP959	Tcells	d0	147520	6212	6150
GTSP0734	p04409-01	UPCC04409	PBL	d13	751709	2770	1354
GTSP0735	p04409-02	UPCC04409	PBL	d28	643993	954	213
GTSP0736	p04409-03	UPCC04409	PBL	d11	62603	31	26
GTSP0737	p04409-05	UPCC04409	Tcells	d0	893878	945	845
GTSP0738	p04409-05	UPCC04409	PBL	d17	752836	277	136
GTSP0739	p04409-06	UPCC04409	Tcells	d0	508559	4596	4272
GTSP0741	p04409-07	UPCC04409	Tcells	d0	438594	2468	2333
GTSP0742	p04409-07	UPCC04409	PBL	d1	76316	13	8
GTSP0744	p04409-09	UPCC04409	PBL	d10	437746	1396	1125
GTSP0746	p04409-10	UPCC04409	PBL	d121	138379	362	4
GTSP0747	p04409-12	UPCC04409	Tcells	d0	68770	1180	1171
GTSP0748	p04409-12	UPCC04409	PBL	d14	804839	1347	1120
GTSP0749	p04409-14	UPCC04409	Tcells	d0	502045	759	605
GTSP0750	p04409-14	UPCC04409	PBL	d14	51993	4	3
GTSP0752	p04409-17	UPCC04409	PBL	d10	616853	464	370
GTSP0754	p04409-18	UPCC04409	PBL	d1	2	2	2
GTSP0755	p04409-22	UPCC04409	Tcells	d0	83535	1043	1028
GTSP0757	p04409-25	UPCC04409	Tcells	d0	80342	589	581

(continued)

GTSP ID	Patient ID	Clin. Trial	Cell Type	Time Point	Reads	Cells Sampled	Unique Int. Sites
GTSP0758	p04409-25	UPCC04409	PBL	d14	720077	282	250
GTSP1166	p04409-02	UPCC04409	PBL	m6	539335	158	71
GTSP1167	p04409-02	UPCC04409	PBL	m12	321638	54	17
GTSP1173	p04409-02	UPCC04409	Tcells	d0	245654	1242	1216
GTSP1175	p04409-05	UPCC04409	PBL	d21	694693	232	157
GTSP1177	p04409-09	UPCC04409	PBL	m9	26664	22	20
GTSP1178	p04409-09	UPCC04409	PBL	m12	676120	139	35
GTSP1180	p04409-12	UPCC04409	PBL	d9	510064	130	102
GTSP1183	p04409-12	UPCC04409	PBL	m2	781850	169	133
GTSP1187	p04409-22	UPCC04409	PBL	d10	272730	212	164
GTSP1188	p04409-22	UPCC04409	PBL	d14	631464	3762	3112
GTSP1190	p04409-22	UPCC04409	PBL	m2	529828	276	223
GTSP1191	p04409-22	UPCC04409	PBL	m3	650179	165	104
GTSP1192	p04409-22	UPCC04409	PBL	m4	137375	78	63
GTSP1193	p04409-22	UPCC04409	PBL	m5	1285	22	22
GTSP1196	p04409-01	UPCC04409	Tcells	d0	579187	3539	3378
GTSP1197	p04409-01	UPCC04409	PBL	m9	445150	457	371
GTSP1198	p04409-01	UPCC04409	PBL	m12	712135	91	55
GTSP1203	p03712-03	UPCC03712	PBL	d14	826260	2262	1576
GTSP1204	p03712-03	UPCC03712	PBL	d21	967241	1164	690
GTSP1206	p03712-04	UPCC03712	PBL	d14	803350	376	286
GTSP1207	p03712-04	UPCC03712	PBL	d21	611345	1096	595
GTSP1209	p03712-22	UPCC03712	PBL	d10	843793	295	257
GTSP1210	p03712-22	UPCC03712	PBL	d14	673873	253	203
GTSP1211	p03712-45	UPCC03712	Tcells	d0	860010	2334	2171
GTSP1212	p03712-45	UPCC03712	PBL	d14	758985	956	536
GTSP1213	p03712-45	UPCC03712	PBL	d21	766176	439	223
GTSP1215	p03712-16	UPCC03712	PBL	d10	757798	308	267
GTSP1216	p03712-16	UPCC03712	PBL	d14	708497	1146	1036
GTSP1219	p03712-18	UPCC03712	PBL	d14	635449	2823	2544
GTSP1220	p03712-18	UPCC03712	PBL	d21	623823	3760	2254
GTSP1222	p03712-29	UPCC03712	Tcells	d0	772885	3597	3313
GTSP1223	p03712-29	UPCC03712	PBL	d14	711525	331	233
GTSP1224	p03712-29	UPCC03712	PBL	d28	717902	775	577
GTSP1225	p04409-01	UPCC04409	PBL	m2	230	12	12
GTSP1226	p04409-01	UPCC04409	PBL	m3	813608	563	517

(continued)

GTSP ID	Patient ID	Clin. Trial	Cell Type	Time Point	Reads	Cells Sampled	Unique Int. Sites
GTSP1228	p04409-02	UPCC04409	PBL	d50	402159	2183	397
GTSP1229	p04409-02	UPCC04409	PBL	m5	822094	122	33
GTSP1232	p04409-09	UPCC04409	PBL	m2	572716	435	337
GTSP1233	p04409-09	UPCC04409	PBL	m5	599535	280	240
GTSP1234	p04409-09	UPCC04409	PBL	m6	118156	22	18
GTSP1235	p03712-03	UPCC03712	PBL	m2	717287	413	273
GTSP1236	p03712-03	UPCC03712	PBL	m4	573894	281	215
GTSP1238	p03712-06	UPCC03712	PBL	d14	879621	1178	938
GTSP1240	p03712-16	UPCC03712	PBL	m2	763536	388	287
GTSP1241	p03712-18	UPCC03712	PBL	m2	906316	49	32
GTSP1405	p03712-45	UPCC03712	PBL	d28	871549	433	227
GTSP1406	p04409-01	UPCC04409	PBL	d28	12753	36	35
GTSP1407	p04409-03	UPCC04409	Tcells	d0	1115510	2733	2186
GTSP1408	p04409-03	UPCC04409	PBL	d28	328800	788	761
GTSP1409	p04409-05	UPCC04409	PBL	d28	1008352	112	53
GTSP1411	p04409-07	UPCC04409	PBL	d28	17812	8	5
GTSP1413	p959-101	CHP959	PBL	d28	964753	302	213
GTSP1414	p959-112	CHP959	PBL	d28	871742	160	104
GTSP1415	p959-117	CHP959	PBL	d25	938346	311	193
GTSP1416	p959-133	CHP959	PBL	d28	814251	165	106
GTSP1603	p04409-10	UPCC04409	PBL	d28	465128	455	396
GTSP1604	p04409-10	UPCC04409	PBL	d63	865690	143	88
GTSP1605	p04409-10	UPCC04409	PBL	d92	33434	17	14
GTSP1606	p04409-10	UPCC04409	PBL	d147	735867	397	27
GTSP1607	p04409-10	UPCC04409	PBL	d442	306698	80	42
GTSP1608	p04409-10	UPCC04409	PBL	d801	646210	92	54
GTSP1609	p04409-10	UPCC04409	PBL	d1584	234777	48	28
GTSP2275	p03712-06	UPCC03712	BM	d28	585462	299	244
GTSP2276	p03712-06	UPCC03712	PBL	m6	521819	322	276
GTSP2277	p03712-06	UPCC03712	PBL	m9	632212	139	101
GTSP2278	p03712-06	UPCC03712	BM	m9	643889	121	76
GTSP2279	p03712-06	UPCC03712	PBL	m12	485129	200	159
GTSP2280	p03712-06	UPCC03712	BM	m12	340995	232	200
GTSP2281	p03712-06	UPCC03712	PBL	y1.5	584295	347	220
GTSP2282	p03712-06	UPCC03712	PBL	y1.5	697221	69	34
GTSP2283	p03712-06	UPCC03712	PBL	y2.5	497894	114	52

(continued)

GTSP ID	Patient ID	Clin. Trial	Cell Type	Time Point	Reads	Cells Sampled	Unique Int. Sites
GTSP2284	p03712-06	UPCC03712	PBL	y3	537768	149	93
GTSP2285	p03712-06	UPCC03712	PBL	y5	580714	39	20
GTSP2648	p18415-01	UPCC18415	Tcells	d0	944888	1925	1812
GTSP2649	p18415-02	UPCC18415	Tcells	d0	871295	1438	1351
GTSP2650	p18415-03	UPCC18415	Tcells	d0	944334	2914	2815
GTSP2651	p18415-04	UPCC18415	Tcells	d0	640111	3934	3833
GTSP2652	p18415-05	UPCC18415	Tcells	d0	771094	1539	1447
GTSP2653	p18415-06	UPCC18415	Tcells	d0	850718	2917	2803
GTSP2654	p18415-07	UPCC18415	Tcells	d0	591275	2510	2436
GTSP2655	p18415-08	UPCC18415	Tcells	d0	874944	3036	2890
GTSP2656	p18415-09	UPCC18415	Tcells	d0	918815	2335	2212
GTSP2657	p18415-10	UPCC18415	Tcells	d0	948901	2311	2216
GTSP2658	p18415-11	UPCC18415	Tcells	d0	1054162	3885	3721
GTSP2659	p18415-12	UPCC18415	Tcells	d0	771060	2686	2543
GTSP2660	p18415-13	UPCC18415	Tcells	d0	626232	3901	3809
GTSP2661	p18415-15	UPCC18415	Tcells	d0	1558046	3285	2984
GTSP2662	p18415-16	UPCC18415	Tcells	d0	875474	2000	1900
GTSP2664	p18415-18	UPCC18415	Tcells	d0	481737	3682	3593
GTSP2665	p18415-19	UPCC18415	Tcells	d0	544293	4236	4160
GTSP2666	p18415-20	UPCC18415	Tcells	d0	686532	4382	4232

Supplemental Table S4. Cancer-associated gene lists.

Table S4. Gene sets used for analysis, including cancer associated genes, human lymphoid cancer genes, and genes associated with clonal hematopoiesis, and overlaps between lists. Origins of gene lists are as follows: Bushman Lab allOnco List: (Sadelain et al., 2011, http://www.bushmanlab.org/links/genelists); Lymphoid cancer list: (personal communication, Marina Cavazzana, http://www.bushmanlab.org/links/genelists); COSMIC: (Sondka, Z. et al. 2018 Nature Cancer Reviews), TCGA_Xie and Clonal_Hema_Xie (Xie, M. et al. 2014 Nature Medicine).

List	Size	allOnco	TSGene	humanLymph	COSMIC-CGC	TCGA List	PMID
allOnco (v3, 2017)	2027						22129804
TSGene (v2, 2016)	1217	408					26590405
humanLymph	38	38	8				personal corr.
COSMIC-CGC (v89)	723	560	179	34			30293088
TCGA List	556	311	146	12	289		25326804
Clonal Hema.	30	19	12	1	18	30	25326804

Supplmental Table S5. Statistical analysis of overlaps among genes called at integration sites and cancer associated gene lists.

Table S5. Statistical analysis of overlaps among genes called at integration sites and cancer associated gene lists in Table S4.

Criteria	Size	allOnco (v3, 2017)	TSGene (v2, 2016)	humanLymph	COSMIC-CGC (v89)	TCGA List	Clonal Hema.
Enrichment	102	0.001 (20)*	0.001 (14)*	1.000 (0)	0.012 (9)*	0.001 (10)*	0.191 (1)
Depletion	98	0.003 (19)*	0.030 (10)*	1.000 (0)	0.180 (6)	0.014 (8)*	0.184 (1)
Abundance	127	0.000 (27)*	0.444 (7)	1.000 (0)	0.013 (11)*	0.245 (6)	0.292 (1)
Longitudinal	230	0.000 (57)*	0.000 (27)*	0.306 (1)	0.000 (27)*	0.000 (26)*	0.008 (3)*
Composite	399	0.000 (84)*	0.000 (38)*	0.471 (1)	0.000 (41)*	0.000 (36)*	0.035 (3)*

Supplemental Table S6. Table of model variables

Table S6. 91 model variables used in the LASSO regression analysis.

Group	Abbv.	Description	Size (Kb)	Source
Population	numUniqSites	Unique Sites	NA	Calculated
Population	estAbund	Mean Estimated Abundance	NA	Calculated
Population	relAbund	Mean Relative Abundance	NA	Calculated
Population	ShannonIndex	Shannon Index	NA	Calculated
Population	GiniIndex	Gini Index	NA	Calculated
Population	Chao1	Chao 1 Estimate	NA	Calculated
Population	UC50	UC50 - Minimal number of unique clones making up 50% of the overall abundance	NA	Calculated
Population	clustersRepresented	Integration sites clusters represented within the specimens	NA	Calculated
Population	numSitesInClusters	Number of sites within clusters	NA	Calculated
Population	abundInClusters	Abundance of clones found within clusters	NA	Calculated
Genomic	pctTxnUnit	Percent of clones within Transcription Units	NA	RefSeq-NCBI
Genomic	pctSameOrt	Percent of clones in same orientation as gene transcription	NA	RefSeq-NCBI
Genomic	pctNearTxnUn	Percent of clones near transcription units	NA	RefSeq-NCBI
Genomic	pctInOnco	Percent clones within transcription units of onco-related genes	NA	Bushman Lab
Genomic	refSeq_counts.10k	RefSeq Genes within X window	10.0	RefSeq-NCBI
Genomic	refSeq_counts.100k	RefSeq Genes within X window	100.0	RefSeq-NCBI
Genomic	refSeq_counts.1M	RefSeq Genes within X window	1000.0	RefSeq-NCBI
Genomic	GC.100	GC Percent within X window	0.1	Calculated
Genomic	GC.1k	GC Percent within X window	1.0	Calculated
Genomic	GC.10k	GC Percent within X window	10.0	Calculated
Genomic	GC.100k	GC Percent within X window	100.0	Calculated
Genomic	GC.1M	GC Percent within X window	1000.0	Calculated
Genomic	CpG_counts.1k	CpG counts within X window	1.0	Calculated
Genomic	CpG_counts.10k	CpG counts within X window	10.0	Calculated
Genomic	CpG_density.10k	CpG Density within X window	10.0	Calculated
Genomic	CpG_density.100k	CpG Density within X window	100.0	Calculated
Genomic	CpG_density.1M	CpG Density within X window	1000.0	Calculated
Genomic	DNaseI_count.1k	DNaseI counts within X window	1.0	UCSC
Genomic	DNaseI_count.10k	DNaseI counts within X window	10.0	UCSC
Genomic	DNaseI_count.100k	DNaseI counts within X window	100.0	UCSC
Genomic	DNaseI_count.1M	DNaseI counts within X window	1000.0	UCSC
Epigenetic	H2AK5ac.10k	H2AK5ac counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H2AK9ac.10k	H2AK9ac counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H2AZ.10k	H2AZ counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H2BK120ac.10k	H2BK120ac counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H2BK12ac.10k	H2BK12ac counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H2BK20ac.10k	H2BK20ac counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H2BK5ac.10k	H2BK5ac counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H2BK5me1.10k	H2BK5me1 counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3K14ac.10k	H3K14ac counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3K18ac.10k	H3K18ac counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3K23ac.10k	H3K23ac counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3K27ac.10k	H3K27ac counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3K27me1.10k	H3K27me1 counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3K27me2.10k	H3K27me2 counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3K27me3.10k	H3K27me3 counts within X window	10.0	UCSC - hg18 LiftOver

(continued)

Group	Abbv.	Description	Size (Kb)	Source
Epigenetic	H3K36ac.10k	H3K36ac counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3K36me1.10k	H3K36me1 counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3K36me3.10k	H3K36me3 counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3K4ac.10k	H3K4ac counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3K4me1.10k	H3K4me1 counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3K4me2.10k	H3K4me2 counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3K4me3.10k	H3K4me3 counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3K79me1.10k	H3K79me1 counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3K79me2.10k	H3K79me2 counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3K79me3.10k	H3K79me3 counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3K9ac.10k	H3K9ac counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3K9me1.10k	H3K9me1 counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3K9me2.10k	H3K9me2 counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3K9me3.10k	H3K9me3 counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3R2me1.10k	H3R2me1 counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H3R2me2.10k	H3R2me2 counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H4K12ac.10k	H4K12ac counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H4K16ac.10k	H4K16ac counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H4K20me1.10k	H4K20me1 counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H4K20me3.10k	H4K20me3 counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H4K5ac.10k	H4K5ac counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H4K8ac.10k	H4K8ac counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H4K91ac.10k	H4K91ac counts within X window	10.0	UCSC - hg18 LiftOver
Epigenetic	H4R3me2.10k	H4R3me2 counts within X window	10.0	UCSC - hg18 LiftOver
Bound Proteins	CTCF.10k	CTCF sites within X window	10.0	UCSC - hg18 LiftOver
Bound Proteins	NRSF.10k	NRSF or REST sites within X window	10.0	UCSC - hg18 LiftOver
Bound Proteins	PolII.10k	PolII binding sites within X window	10.0	UCSC - hg18 LiftOver
Bound Proteins	Brd2_promoters.10k	Brd2 promotors within X window	10.0	UCSC - hg18 LiftOver
Bound Proteins	Brd3_promoters.10k	Brd3 promotors within X window	10.0	UCSC - hg18 LiftOver
Bound Proteins	Brd4_promoters.10k	Brd4 promotors within X window	10.0	UCSC - hg18 LiftOver
Bound Proteins	HP1a_promoters.10k	HP1aphla promotors within X window	10.0	UCSC - hg18 LiftOver
Bound Proteins	HP1b_promoters.10k	HP1beta promotors within X window	10.0	UCSC - hg18 LiftOver
Bound Proteins	Act_CD4_HDAC6.10k	HDAC6 binding in Activated CD4 Tcells within X window	10.0	UCSC - hg18 LiftOver
Bound Proteins	Act_CD4_Tip60.10k	Tip60 bindiing in Activated CD4 Tcells within X window	10.0	UCSC - hg18 LiftOver
Bound Proteins	ActivatedNucleosomes.10k	Activated Nucleosomes	10.0	UCSC - hg18 LiftOver
Bound Proteins	RestingNucleosomes.10k	Resting Nucleosomes	10.0	UCSC - hg18 LiftOver
Bound Proteins	Rest_CD4_CBP.10k	CBP binding in Resting CD4 Tcells within X window	10.0	UCSC - hg18 LiftOver
Bound Proteins	Rest_CD4_HDAC1.10k	HDAC1 binding in Resting CD4 Tcells within X window	10.0	UCSC - hg18 LiftOver
Bound Proteins	Rest_CD4_HDAC2.10k	HDAC2 binding in Resting CD4 Tcells within X window	10.0	UCSC - hg18 LiftOver
Bound Proteins	Rest_CD4_HDAC3.10k	HDAC3 binding in Resting CD4 Tcells within X window	10.0	UCSC - hg18 LiftOver
Bound Proteins	Rest_CD4_HDAC6.10k	HDAC6 binding in Resting CD4 Tcells within X window	10.0	UCSC - hg18 LiftOver
Bound Proteins	Rest_CD4_MOF.10k	MOF binding in Resting CD4 Tcells within X window	10.0	UCSC - hg18 LiftOver
Bound Proteins	Rest_CD4_p300.10k	p300 binding in Resting CD4 Tcells within X window	10.0	UCSC - hg18 LiftOver
Bound Proteins	Rest_CD4_PCAF.10k	PCAF binding in Resting CD4 Tcells within X window	10.0	UCSC - hg18 LiftOver
Bound Proteins	Rest_CD4_Tip60.10k	Tip60 binding in Resting CD4 Tcells within X window	10.0	UCSC - hg18 LiftOver

Supplementary Report 1:
Genes of interest marked by vector integration
All Patients (CLL & ALL) and
Response Groups (CR/PRtd & PR/NR)

Contents

Summary

Lentiviral vectors integrate into genomes of targeted host cells (Tcells). 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 mined the data collected from integration site sequencing for 40 CART treated subjects. We constructed 4 gene lists based on: 1 & 2) increased / decreased integration site occurrence in patient samples relative to the initial transduction product, 3) peak clonal abundance, and 4) longitudinal clonal persistence. More about each of these criteria is below:

- **Integration Frequency** is the rate at which integration sites are observed within a gene. This is compared between patient samples and the initial transduction product to score enrichment or depletion during growth in patients. The top of genes with higher patient sample integration frequency over transduction samples were chosen for study (p-value $<= 0.05$ after exclusion of genes with clones from less than 2 patients and less than 10 observed clones).
- **Clonal Abundance** can be determined during analysis by quantifying the number of sites of linker ligation associated with each unique integration site. This method is further described in **Berry *et al.* 2012**. This allows clonal expansion to be quantified. The top 1% of the genes were selected for study based on their maximal peak clonal abundance.
- **Longitudinal Observation** of clones is the quantification of observed timespans and last observed timepoints. The maximum value for clones within a gene were considered for characterization of the gene in this analysis. Genes were only considered if there were 10 or more integration sites isolated from at least two different patient samples. Genes were also not considered if they only consited of clones which were observed once or the last observed timepoint was less than 90 days from initial infusion.

A point to keep in mind through all this analysis is that integration sites are sampled from a larger population. It would be rare for all integration sites in a sample to be represented in the sequence data.

Table 1: Summary of each filtering criteria.

	Gene	Onco	Tumor	Lymphoma	COSMIC	TCGA	Clonal Hema.
Criteria	Count	Related (%)	Suppressors (%)	Related (%)	Related (%)	Related (%)	Related (%)
Enrichment	102	*/* 19.6	*/* 13.73	/ 0.000	*/* 8.82	*/* 9.80	/ 0.980
Depletion	93	*/* 20.4	*/* 10.75	/ 0.000	/ 6.45	*/* 8.60	/ 1.075
Abundance	132	*/* 20.5	/ 5.30	/ 0.000	*/* 8.33	/ 4.55	/ 0.758
Longitudinal	226	*/* 25.2	*/* 11.95	/ 0.442	*/* 11.95	*/* 11.50	*/* 1.327
Composite	404	*/* 20.8	*/* 9.41	/ 0.248	*/* 10.15	*/* 8.91	*/ 0.743

Table 1 summarizes the size and contents of each criteria gene list identified by the various methods. Significance of overlap between lists are displayed by asterisks before the percent of genes identifed 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 Fishers Exact test and multiple comparison corrections were made using a Benjamini-Hochberg (FDR) method for each criteria based list.

Percent of all analyzed transcription units associated with each list as as follows:

- Onco Related: 9.1%
- Tumor Suppressors: 4.83%
- Lymphoma Related: 0.16%
- COSMIC Related: 3.55%
- TCGA Related: 2.73%
- Clonal Hematopoiesis Related: 0.17%

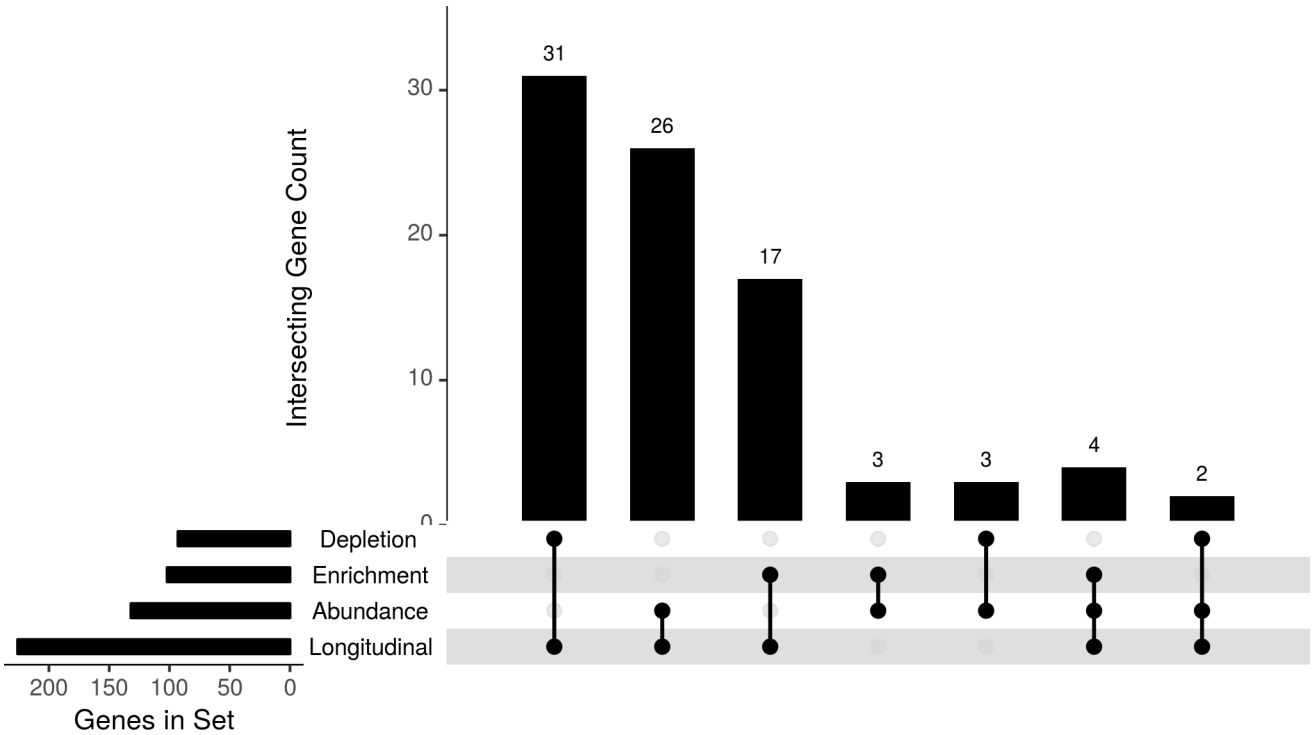

Figure 1: Intersecting gene lists identified through the various selection criteria.

Table 2: The most consistently observed genes from filtering by various criteria. The 'Criteria.' column is a count of how many times the gene was identified by these methods, while the 'Patients' column notes how many specimens collected from patients have had integration sites within the noted gene.

Gene	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
AKAP13	12	75.1	27	360	EAL
UBAP2L	12	65.9	30	180	EAL
PTBP1	8	106.0	47	360	EAL
TET2	6	196.7	814	1584	EAL

Table 3: GO Biological Process. Top 6 per group. Total genes considered: 400

Group	GO ID	GO Term	Term Size	Gene Count	Adjusted P-value
1	GO:0016569	covalent chromatin modification	465	44	0.0000000
	GO:0016570	histone modification	377	39	0.0000000
	GO:0018205	peptidyl-lysine modification	317	35	0.0000000
	GO:0043414	macromolecule methylation	232	18	0.0053317
	GO:0006479	protein methylation	152	16	0.0003283
	GO:0033044	regulation of chromosome organization	256	16	0.0448756
2	GO:1903827	regulation of cellular protein localization	462	30	0.0022168
	GO:0080135	regulation of cellular response to stress	489	29	0.0105064
	GO:0032386	regulation of intracellular transport	388	27	0.0015907
	GO:0006281	DNA repair	454	27	0.0142276
	GO:0006913	nucleocytoplasmic transport	369	22	0.0236988
	GO:0051169	nuclear transport	376	22	0.0291111
3	GO:0010256	endomembrane system organization	496	39	0.0000019
	GO:0030155	regulation of cell adhesion	473	29	0.0067485
	GO:0032101	regulation of response to external stimulus	472	27	0.0225913
	GO:0030036	actin cytoskeleton organization	451	25	0.0324282
	GO:0051493	regulation of cytoskeleton organization	352	23	0.0100700
	GO:0097435	supramolecular fiber organization	423	23	0.0448756
4	GO:0051640	organelle localization	415	26	0.0090197
	GO:0051656	establishment of organelle localization	352	25	0.0019782
	GO:0048285	organelle fission	384	23	0.0225913
	GO:0000280	nuclear division	351	22	0.0195459
	GO:0016050	vesicle organization	310	21	0.0102878
	GO:0048193	Golgi vesicle transport	294	18	0.0379356
5	GO:0046649	lymphocyte activation	472	25	0.0452453
	GO:0071396	cellular response to lipid	369	23	0.0170187
	GO:0048729	tissue morphogenesis	396	23	0.0266749
	GO:0071407	cellular response to organic cyclic compound	414	23	0.0417615
	GO:0009792	embryo development ending in birth or egg hatching	362	22	0.0225913
	GO:0006914	autophagy	402	22	0.0475427
6	GO:0006397	mRNA processing	409	30	0.0002568
	GO:0008380	RNA splicing	360	22	0.0225913
	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	277	17	0.0435616
	GO:0000398	mRNA splicing, via spliceosome	277	17	0.0435616
	GO:0000375	RNA splicing, via transesterification reactions	279	17	0.0448756
	GO:0050684	regulation of mRNA processing	97	8	0.0458860
7	GO:0006511	ubiquitin-dependent protein catabolic process	497	30	0.0069438
	GO:0010498	proteasomal protein catabolic process	371	22	0.0251384
	GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic process	342	20	0.0405641
	GO:0061136	regulation of proteasomal protein catabolic process	134	12	0.0106185
	GO:0009895	negative regulation of catabolic process	174	12	0.0448756
	GO:0032434	regulation of proteasomal ubiquitin-dependent protein catabolic process	101	11	0.0034931

Table 4: KEGG Pathway analysis. Top 10 per group. Total genes considered: 190

Group	KEGG ID	Description	Term Size	Gene Count	Adjusted P-value
1	path:hsa04024	cAMP signaling pathway	141	12	0.0230257
	path:hsa04120	Ubiquitin mediated proteolysis	127	11	0.0230257
	path:hsa04720	Long-term potentiation	51	9	0.0004645
	path:hsa04114	Oocyte meiosis	99	8	0.0713955
	path:hsa04921	Oxytocin signaling pathway	114	8	0.0936522
	path:hsa04916	Melanogenesis	61	7	0.0230257
	path:hsa04310	Wnt signaling pathway	103	7	0.1074085
	path:hsa05152	Tuberculosis	118	7	0.1360859
	path:hsa04360	Axon guidance	142	7	0.2296122
	path:hsa04020	Calcium signaling pathway	136	6	0.3099265
2	path:hsa05200	Pathways in cancer	358	17	0.1360859
	path:hsa04144	Endocytosis	211	13	0.0771521
	path:hsa04919	Thyroid hormone signaling pathway	92	11	0.0026018
	path:hsa05165	Human papillomavirus infection	235	11	0.2021271
	path:hsa05203	Viral carcinogenesis	162	10	0.1065141
	path:hsa05206	MicroRNAs in cancer	189	10	0.1412888
	path:hsa05166	Human T-cell leukemia virus 1 infection	173	9	0.1568472
	path:hsa04068	FoxO signaling pathway	90	8	0.0463230
	path:hsa05202	Transcriptional misregulation in cancer	112	8	0.0872195
	path:hsa05016	Huntington disease	147	7	0.2507228
3	path:hsa04070	Phosphatidylinositol signaling system	89	6	0.1195906
	path:hsa00562	Inositol phosphate metabolism	67	5	0.1160011
4	path:hsa03018	RNA degradation	70	5	0.1195906
5	path:hsa05205	Proteoglycans in cancer	133	12	0.0187863
	path:hsa04140	Autophagy - animal	114	10	0.0285422
	path:hsa00310	Lysine degradation	51	9	0.0004645
	path:hsa05225	Hepatocellular carcinoma	107	9	0.0463230
	path:hsa04714	Thermogenesis	167	9	0.1471983
	path:hsa04915	Estrogen signaling pathway	87	7	0.0771521
	path:hsa04910	Insulin signaling pathway	108	7	0.1185385
	path:hsa04072	Phospholipase D signaling pathway	110	7	0.1195906
	path:hsa04014	Ras signaling pathway	158	7	0.2964885
	path:hsa05168	Herpes simplex virus 1 infection	382	7	0.9259280
6	path:hsa04010	MAPK signaling pathway	208	13	0.0771521
	path:hsa04141	Protein processing in endoplasmic reticulum	139	11	0.0402779
	path:hsa05163	Human cytomegalovirus infection	166	10	0.1074085
	path:hsa04810	Regulation of actin cytoskeleton	156	9	0.1263724
	path:hsa04218	Cellular senescence	116	8	0.0960844
	path:hsa04022	cGMP-PKG signaling pathway	115	7	0.1346511
	path:hsa05167	Kaposi sarcoma-associated herpesvirus infection	131	7	0.1791738
	path:hsa04660	T cell receptor signaling pathway	84	6	0.1103395
	path:hsa04380	Osteoclast differentiation	90	6	0.1195906
	path:hsa04650	Natural killer cell mediated cytotoxicity	90	6	0.1195906

Integration Frequency (Enrichment)

Table 5: Table of top 50 genes with the most frequent clonal enrichment.

Gene	Num. Patients	TDN Sites	Patient Sites	Onco-Related	Frequency Increase (%)
HERC2	6	3	10	FALSE	518.1
PIP5K1A	9	4	13	FALSE	502.7
RAB11FIP2	9	4	12	FALSE	456.3
NUP107	9	4	12	FALSE	456.3
HSF2	8	5	14	FALSE	419.2
PDCD10	9	4	11	FALSE	410.0
LRPPRC	10	6	15	FALSE	363.6
RAD23B	7	4	10	TRUE	363.6
RBM27	6	5	12	FALSE	345.1
PIKFYVE	10	5	12	FALSE	345.1
ABLIM1	6	5	11	FALSE	308.0
CAMKMT	6	5	11	FALSE	308.0
TMTC3	6	6	13	FALSE	301.8
ARHGAP12	6	5	10	FALSE	270.9
ATG5	9	10	19	FALSE	252.3
GNA12	6	6	11	TRUE	240.0
ATE1	7	6	11	FALSE	240.0
PPP4R2	7	6	11	FALSE	240.0
BCKDHB	7	6	11	FALSE	240.0
FRG1BP	9	8	14	FALSE	224.5
MACROD2	7	8	14	FALSE	224.5
FUS	5	7	12	TRUE	217.9
UCHL3	7	7	12	FALSE	217.9
CPEB2	11	10	17	FALSE	215.2
HELLS	10	10	17	FALSE	215.2
KDM4A	9	10	17	FALSE	215.2
SNRPA	5	6	10	FALSE	209.1
LUC7L2	8	6	10	FALSE	209.1
USP9Y	7	6	10	FALSE	209.1
CDK8	7	6	10	FALSE	209.1
BZW2	6	6	10	FALSE	209.1
RBPJ	9	6	10	FALSE	209.1
IKZF2	8	8	13	TRUE	201.3
FUNDC2	6	8	13	FALSE	201.3
PDE12	7	8	13	FALSE	201.3
URI1	9	13	21	FALSE	199.6
TET2	6	10	16	TRUE	196.7
FANCL	5	7	11	FALSE	191.4
PRKN	7	7	11	FALSE	191.4
LOC101929095	12	13	20	FALSE	185.3
ASCC3	8	22	33	FALSE	178.2
WWP1	9	15	22	TRUE	172.0
GMDS	12	13	19	FALSE	171.0
BRWD3	6	9	13	TRUE	167.9
ECD	6	9	13	FALSE	167.9
KIF20B	9	9	13	FALSE	167.9
PHF3	10	14	20	FALSE	164.9
RBM39	11	14	20	TRUE	164.9
NDFIP2	6	12	17	FALSE	162.7
MTREX	9	12	17	FALSE	162.7

Integration Frequency (Depletion)

Table 6: Table of top 50 genes with the most frequent clonal depletion.

Gene	Num. Patients	TDN Sites	Patient Sites	Onco-Related	Frequency Increase (%)
RNPS1	6	146	17	FALSE	-78.4
EXOC2	5	66	10	FALSE	-71.9
LSM2	7	73	13	FALSE	-67.0
NOSIP	10	158	31	FALSE	-63.6
SFI1	6	56	11	FALSE	-63.6
ZNF598	5	56	11	FALSE	-63.6
UBE2J2	7	69	14	FALSE	-62.4
ZBTB4	6	49	10	FALSE	-62.2
WDR90	8	53	11	FALSE	-61.5
IP6K1	11	139	29	FALSE	-61.3
EIF2B3	10	81	17	FALSE	-61.1
STK11	7	47	10	TRUE	-60.5
WNK1	6	70	15	FALSE	-60.3
PLEC	10	105	23	FALSE	-59.4
NARFL	9	54	12	FALSE	-58.8
PRKAR2A	6	58	13	FALSE	-58.4
HAGH	5	53	12	FALSE	-58.0
GRAP2	6	44	10	FALSE	-57.9
IFT140	8	83	19	FALSE	-57.5
CCND3	10	82	19	TRUE	-57.0
CNOT6	5	47	11	FALSE	-56.6
DIDO1	6	42	10	FALSE	-55.8
QRICH1	8	92	22	FALSE	-55.7
PSMB9	9	54	13	FALSE	-55.4
PCBP3	4	41	10	FALSE	-54.8
PCED1B	6	45	11	TRUE	-54.7
TAP2	8	49	12	TRUE	-54.6
HORMAD2	8	93	23	FALSE	-54.1
TSC2	9	88	22	TRUE	-53.6
RPRD2	6	44	11	FALSE	-53.6
ZGPAT	11	98	25	FALSE	-52.7
HCG20	9	90	23	FALSE	-52.6
RAB40C	7	43	11	FALSE	-52.6
FAM222B	4	39	10	FALSE	-52.5
FKBP5	13	126	33	FALSE	-51.4
RBM14-RBM4	7	42	11	FALSE	-51.4
ABHD16A	9	38	10	FALSE	-51.2
MCM3AP	6	38	10	FALSE	-51.2
MIR5096	6	49	13	FALSE	-50.8
MTMR3	6	45	12	TRUE	-50.5
MROH1	13	291	78	FALSE	-50.3
SEPT2	7	41	11	TRUE	-50.2
TC2N	5	37	10	FALSE	-49.9
TRAF2	10	103	28	FALSE	-49.6
CEACAM21	9	73	20	FALSE	-49.2
RBM4	9	47	13	FALSE	-48.7
ADCK5	6	47	13	FALSE	-48.7
RNF216	9	54	15	TRUE	-48.5
PRRC2A	7	61	17	FALSE	-48.3
ASCC1	7	43	12	FALSE	-48.2

Genes with the Most Abundant Clones

Table 7: Table of top 50 Genes containing the highest abundant clones.

Gene	Num. Patients	Peak Abundance	Peak Rel. Abund.	Clonal Gini Index	Onco-Related
TET2	8	814	0.989	0.923	TRUE
KCTD3	4	589	0.265	0.745	FALSE
PATL1	4	578	0.260	0.793	FALSE
PIKFYVE	10	410	0.273	0.890	FALSE
SRCAP	11	373	0.357	0.896	FALSE
MTMR3	6	261	0.041	0.876	TRUE
PCNX1	11	153	0.010	0.827	FALSE
PPP6R3	15	149	0.040	0.717	FALSE
SSH2	10	137	0.062	0.792	FALSE
RSRC1	9	109	0.014	0.812	FALSE
SNHG12	2	96	0.057	0.646	FALSE
MAPK14	9	91	0.018	0.774	TRUE
RPA3	5	87	0.020	0.783	FALSE
ZNF573	3	86	0.610	0.677	FALSE
MGA	13	85	0.013	0.746	FALSE
AQR	5	84	0.022	0.798	FALSE
LEF1	9	84	0.038	0.765	TRUE
LINC01473	3	82	0.075	0.643	FALSE
CARD8	14	79	0.056	0.681	TRUE
IQCB1	5	79	0.028	0.752	FALSE
DNAJC13	9	71	0.004	0.764	FALSE
EXOSC10	4	70	0.008	0.776	FALSE
ATP2A2	8	67	0.030	0.749	FALSE
SEC31A	6	66	0.004	0.752	FALSE
GPN1	2	62	0.017	0.711	FALSE
SMAP2	6	61	0.004	0.768	FALSE
TRIO	6	61	0.025	0.769	TRUE
ZZEF1	13	56	0.333	0.614	FALSE
CLK4	8	53	0.036	0.653	FALSE
IFNGR2	2	53	0.722	0.635	TRUE
JMJD6	2	53	0.015	0.755	FALSE
KDM5D	8	51	0.017	0.745	FALSE
UBR1	10	48	0.421	0.686	FALSE
MEMO1	6	47	0.006	0.741	FALSE
PTBP1	8	47	0.043	0.660	TRUE
DYNC1H1	8	44	0.003	0.709	FALSE
NGDN	3	44	0.005	0.623	FALSE
EIF2AK4	3	43	0.003	0.659	FALSE
MSH5-SAPCD1	4	43	0.039	0.708	FALSE
POLG2	2	43	0.003	0.708	FALSE
RASEF	2	43	0.005	0.622	FALSE
UXT-AS1	2	43	0.039	0.477	FALSE
ADD1	10	42	0.011	0.594	FALSE
GRB2	12	42	0.017	0.554	TRUE
KIFC1	7	42	0.003	0.694	FALSE
TAC3	2	42	0.018	0.477	FALSE
ZNF92	3	42	0.031	0.707	FALSE
ACTL6A	1	40	0.003	0.000	FALSE
ATP6V1G2-DDX39B	11	40	0.005	0.621	FALSE
PHF12	2	40	0.014	0.670	FALSE

Longitudinal Observation

Table 8: Table of top 50 genes identified by longitudinal observations.

Gene	Time Span	Longest Time	Obs. Count	Num. Patients	Patient Sites	Peak Abund.	Onco-Related
FKBP5	1555.0	1825.0	4	13	33	15	FALSE
PTPRA	1555.0	1825.0	3	9	36	4	FALSE
TET2	1464.0	1584.0	7	6	16	814	TRUE
UBR1	1277.5	1825.0	4	10	18	48	FALSE
COX6B1	825.0	1095.0	3	10	18	5	FALSE
CCDC57	642.5	912.5	2	15	40	6	FALSE
KMT5B	642.5	912.5	2	13	32	7	FALSE
MACF1	519.5	547.5	3	14	34	11	TRUE
DNMT1	365.0	912.5	2	14	74	13	TRUE
STXBP5	350.0	360.0	4	11	21	8	FALSE
CASK	346.0	547.5	2	8	17	5	FALSE
RPTOR	346.0	360.0	2	19	96	11	FALSE
DIP2A	346.0	360.0	2	14	41	25	FALSE
PTBP1	346.0	360.0	2	8	20	47	TRUE
MIR4745	346.0	360.0	2	7	13	47	FALSE
ZZEF1	332.0	360.0	5	13	51	56	FALSE
SRCAP	332.0	360.0	5	11	23	373	FALSE
SNORA30	332.0	360.0	5	7	10	373	FALSE
OGDH	332.0	360.0	4	5	12	17	FALSE
WDR82	277.5	547.5	3	10	25	7	TRUE
PIP5K1A	277.5	547.5	2	9	13	3	FALSE
EP400P1	260.0	270.0	3	6	18	5	FALSE
HSF1	256.0	270.0	3	13	48	10	FALSE
BOP1	256.0	270.0	3	12	35	10	TRUE
FNBP1	256.0	270.0	2	10	35	5	TRUE
ACOX1	256.0	270.0	2	8	19	2	FALSE
PDS5B	256.0	270.0	2	8	16	10	TRUE
PIK3C3	180.0	360.0	3	12	33	5	FALSE
IQGAP1	166.0	180.0	3	11	22	5	FALSE
SNAPC4	166.0	180.0	2	10	21	7	FALSE
UBE2J2	166.0	180.0	2	7	14	2	FALSE
SSH2	152.0	1095.0	4	10	30	137	FALSE
CARD8	152.0	270.0	4	14	44	79	TRUE
MED13	152.0	270.0	4	13	27	21	FALSE
LEF1	152.0	180.0	4	9	18	84	TRUE
VAV1	152.0	180.0	3	14	80	37	TRUE
STAG1	136.0	912.5	2	9	14	6	TRUE
PPP6R2	136.0	180.0	2	14	45	15	FALSE
RTTN	136.0	150.0	2	9	16	6	FALSE
MAPK8IP3	130.0	270.0	2	12	34	5	FALSE
SMG1	122.0	150.0	3	14	41	7	FALSE
INPP4B	122.0	150.0	3	13	32	3	FALSE
PIAS1	122.0	150.0	3	11	24	6	FALSE
DDX60	122.0	150.0	3	10	16	22	FALSE
ZNRD1ASP	122.0	150.0	3	6	14	4	FALSE
DPYD	122.0	150.0	2	14	36	22	FALSE
RUNX1	110.0	360.0	2	9	19	1	TRUE
ASH1L	106.0	1825.0	2	14	51	5	FALSE
WWOX	106.0	1095.0	3	6	12	5	TRUE
RFX2	106.0	360.0	2	7	14	9	TRUE

Reference Data

The NCBI RefGenes data set was used to identify gene regions (hg38) while genes identified as onco-related were from the Bushman Lab curated list of **onco-related genes**.

Gene Ontologies were extracted from the `GO.db` R-package (v3.4.1). KEGG pathways were acquired via interfacing with the KEGG web-server API through the `KEGGREST` R-package (v1.16.1). Gene lists, including RefSeq genes used for annotation of integration sites, were standardized to HGNC gene symbols (date: 2018-02-07). Groups identified in GO and KEGG analyses were determined from Jaccard distances between identified terms, followed by modularity-optimizing clustering from a weighted-undirected graph using a Louvain algorithm (**Blondel *et al.* 2008**). Terms within groups of GO or KEGG terms have greater overlap of gene lists between themselves that between terms found in other groups. This method was implemented to help reduce the functional redundancy commonly observed in GO and overlapping pathways observed with KEGG.

Comprehensive Genes of Interest Table

Table 9: Table of all genes identified within analysis.

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
AKAP13	chr15	85,375,615	85,754,358	12	75.1	27	7.0	EAL
UBAP2L	chr1	154,215,171	154,276,510	12	65.9	30	7.0	EAL
PTBP1	chr19	792,391	817,327	8	106.0	47	346.0	EAL
TET2	chr4	105,140,874	105,284,803	6	196.7	814	1464.0	EAL
LUC7L	chr16	183,968	234,482	17	-20.9	30	7.0	AL
PPP3CA	chr4	101,018,429	101,352,471	17	118.2	6	46.0	EL
JPT2	chr16	1,673,276	1,707,072	16	2.7	23	46.0	AL
ANKRD11	chr16	89,262,620	89,495,561	15	10.4	23	7.0	AL
PPP6R3	chr11	68,455,717	68,620,333	15	56.3	149	14.0	EA
RNF157	chr17	76,137,452	76,245,311	15	-21.7	28	14.0	AL
SEC16A	chr9	136,435,095	136,488,759	15	-8.9	29	46.0	AL
CARD8	chr19	48,203,085	48,260,946	14	-30.3	79	152.0	DAL
CRAMP1	chr16	1,609,639	1,682,908	14	13.8	30	18.0	AL
DIP2A	chr21	46,453,948	46,575,013	14	-20.8	25	346.0	AL
PAFAH1B1	chr17	2,588,628	2,690,615	14	-4.7	30	7.0	AL
SMG1P1	chr16	22,432,007	22,497,220	14	135.4	3	100.0	EL
VAV1	chr19	6,767,667	6,862,366	14	26.8	37	152.0	AL
KDM6A	chrX	44,868,174	45,117,612	13	100.9	9	106.0	EL
ZZEF1	chr17	3,999,444	4,147,959	13	5.1	56	332.0	AL
GMDS	chr6	1,618,799	2,250,634	12	171.0	8	14.0	EL
GRB2	chr17	75,313,075	75,410,709	12	-14.8	42	14.0	AL
JMJD1C	chr10	63,162,220	63,527,075	12	107.7	5	7.0	EL
PIK3C3	chr18	41,950,197	42,086,482	12	144.8	5	180.0	EL
PRKACB	chr1	84,072,974	84,243,498	12	73.1	4	7.0	EL
XPO5	chr6	43,517,329	43,581,075	12	0.2	26	22.0	AL
EYA3	chr1	27,965,343	28,093,637	11	117.4	7	32.0	EL
PIAS1	chr15	68,049,178	68,196,466	11	93.5	6	122.0	EL
SRCAP	chr16	30,694,140	30,745,129	11	4.0	373	332.0	AL
ST13	chr22	40,819,534	40,862,008	11	53.5	29	7.0	AL
USP25	chr21	15,725,024	15,885,071	11	98.2	10	46.0	EL
HELLS	chr10	94,540,766	94,607,099	10	215.2	15	106.0	EL
PIKFYVE	chr2	208,261,266	208,363,751	10	345.1	410	14.0	EA
SSH2	chr17	29,620,938	29,935,228	10	-20.5	137	152.0	AL
UBR1	chr15	42,937,899	43,111,088	10	-22.4	48	1277.5	AL
FAM13A	chr4	88,720,953	89,062,195	9	136.0	1	7.0	EL
LEF1	chr4	108,042,544	108,173,956	9	-1.8	84	152.0	AL
MAPK14	chr6	36,022,676	36,116,236	9	21.5	91	7.0	AL
MCPH1	chr8	6,401,591	6,653,505	9	159.6	9	7.0	EL
PIP5K1A	chr1	151,193,543	151,254,531	9	502.7	3	277.5	EL
RSRC1	chr3	158,105,051	158,549,835	9	23.6	109	106.0	AL
SMURF2	chr17	64,539,616	64,667,268	9	-36.8	27	7.0	AL

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
CLK4	chr5	178,597,663	178,632,053	8	52.3	53	106.0	AL
HERC4	chr10	67,916,898	68,080,346	8	33.9	35	106.0	AL
PDCD4	chr10	110,866,794	110,905,006	8	36.6	26	14.0	AL
MIR4745	chr19	799,939	810,001	7	141.1	47	346.0	EL
BRWD3	chrX	80,664,487	80,814,734	6	167.9	6	7.0	EL
ECD	chr10	73,129,523	73,173,095	6	167.9	24	1.0	EA
MAD1L1	chr7	1,810,791	2,237,948	6	-2.9	35	46.0	AL
MTMR3	chr22	29,878,168	30,035,868	6	-50.5	261	106.0	DAL
MTOR	chr1	11,101,530	11,267,551	6	-7.3	29	90.0	AL
NDFIP2	chr13	79,476,123	79,561,077	6	162.7	8	46.0	EL
PA2G4	chr12	56,099,318	56,118,910	6	48.4	38	14.0	AL
FANCA	chr16	89,732,550	89,821,657	20	16.0	21	15.0	L
NPLOC4	chr17	81,551,884	81,642,153	19	-33.7	16	46.0	DL
RPTOR	chr17	80,539,824	80,971,373	19	0.0	11	346.0	L
KDM2A	chr11	67,114,268	67,263,079	18	-35.4	7	50.0	DL
CBFB	chr16	67,024,146	67,106,055	16	66.9	14	22.0	E
EP300	chr22	41,087,609	41,185,077	16	13.1	6	46.0	L
PACS1	chr11	66,065,352	66,249,747	16	-41.1	5	46.0	DL
TRAPPC10	chr21	44,007,324	44,111,551	16	-17.4	7	14.0	L
UTRN	chr6	144,286,736	144,858,034	16	13.8	6	22.0	L
CCDC57	chr17	82,096,469	82,217,829	15	-46.2	6	642.5	DL
CREBBP	chr16	3,720,054	3,885,120	15	33.3	5	46.0	L
EHMT1	chr9	137,613,991	137,841,126	15	-5.3	3	50.0	L
NSD1	chr5	177,128,078	177,305,213	15	-32.0	6	62.0	DL
ASH1L	chr1	155,330,260	155,567,533	14	-34.8	5	106.0	DL
ATF7IP	chr12	14,360,631	14,507,935	14	18.4	8	7.0	L
DNMT1	chr19	10,128,343	10,200,135	14	-24.6	13	365.0	DL
DPYD	chr1	97,072,743	97,926,059	14	-7.3	22	122.0	L
EPB41	chr1	28,882,090	29,125,046	14	52.7	7	14.0	E
FCHSD2	chr11	72,831,744	73,147,098	14	-14.7	5	22.0	L
MACF1	chr1	39,079,166	39,492,138	14	-35.7	11	519.5	DL
PPP6R2	chr22	50,338,316	50,450,089	14	-32.7	15	136.0	DL
RABEP1	chr17	5,277,262	5,391,339	14	-4.1	26	14.0	A
SETD2	chr3	47,011,407	47,168,977	14	2.8	17	22.0	L
SMARCC1	chr3	47,580,887	47,786,915	14	-4.4	3	7.0	L
SMG1	chr16	18,799,852	18,931,404	14	-22.4	7	122.0	L
SUPT3H	chr6	44,821,729	45,383,051	14	45.5	8	14.0	L
VPS13D	chr1	12,225,038	12,517,046	14	54.5	3	0.0	E
VPS8	chr3	184,807,142	185,057,614	14	79.0	10	46.0	E
CYTH1	chr17	78,669,046	78,787,342	13	-35.4	4	7.0	DL
ELP4	chr11	31,504,728	31,789,525	13	113.3	4	0.0	E

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
FKBP5	chr6	35,568,584	35,733,583	13	-51.4	15	1555.0	DL
HSF1	chr8	144,286,568	144,319,726	13	-39.0	10	256.0	DL
INPP4B	chr4	142,018,159	142,851,535	13	-11.4	3	122.0	L
KMT5B	chr11	68,149,862	68,218,772	13	7.9	7	642.5	L
LRBA	chr4	150,259,658	151,020,497	13	42.4	17	106.0	L
MECP2	chrX	154,016,812	154,102,731	13	-31.0	12	99.0	DL
MED13	chr17	61,937,604	62,070,282	13	-1.8	21	152.0	L
MROH1	chr8	144,143,015	144,266,940	13	-50.3	5	46.0	DL
NF1	chr17	31,089,926	31,382,677	13	17.3	14	14.0	L
PBRM1	chr3	52,540,351	52,690,850	13	-24.7	14	7.0	L
PELP1	chr17	4,666,383	4,709,337	13	73.5	3	14.0	E
RAB11FIP3	chr16	420,667	527,481	13	-40.3	7	46.0	DL
SAFB2	chr19	5,581,998	5,627,927	13	-3.7	13	22.0	L
SF1	chr11	64,759,603	64,783,844	13	27.2	16	14.0	L
ARHGAP15	chr2	143,124,329	143,773,352	12	-1.7	7	5.0	L
BOP1	chr8	144,257,045	144,296,438	12	-32.4	10	256.0	L
CAPN1	chr11	65,176,214	65,217,006	12	-26.4	2	46.0	L
CHD4	chr12	6,565,081	6,612,433	12	26.3	19	7.0	L
CUX1	chr7	101,810,903	102,288,958	12	151.7	3	0.0	E
GBE1	chr3	81,484,698	81,766,799	12	155.0	9	46.0	E
LOC101929095	chr4	14,999,941	15,432,914	12	185.3	14	22.0	E
MAPK8IP3	chr16	1,701,182	1,775,317	12	-41.1	5	130.0	DL
MGA	chr15	41,655,411	41,774,943	12	11.3	85	46.0	A
MIR5096	chr17	4,136,088	4,245,637	12	-29.4	9	14.0	DL
PARP8	chr5	50,660,898	50,851,522	12	0.8	4	14.0	L
RABGAP1L	chr1	174,154,413	175,000,308	12	51.1	3	46.0	L
SMG6	chr17	2,054,838	2,308,775	12	-30.9	16	7.0	DL
SRRM2	chr16	2,747,328	2,776,412	12	45.3	32	1.0	A
USP15	chr12	62,255,339	62,414,721	12	49.4	14	15.0	L
ZNF34	chr8	144,767,223	144,792,345	12	-25.8	23	1.0	A
AP3B1	chr5	77,997,325	78,299,755	11	111.9	5	0.0	E
CDKAL1	chr6	20,529,456	21,237,403	11	6.0	5	106.0	L
CLEC16A	chr16	10,939,487	11,187,189	11	70.8	9	0.0	E
CPEB2	chr4	14,997,673	15,075,153	11	215.2	14	22.0	E
CSNK1D	chr17	82,237,660	82,278,742	11	-25.8	8	7.0	L
DDX42	chr17	63,769,188	63,824,317	11	-7.3	6	14.0	L
DIP2B	chr12	50,499,984	50,753,667	11	-34.0	4	7.0	L
DLG1	chr3	197,037,559	197,304,272	11	-2.4	8	7.0	L
GLCCI1	chr7	7,963,742	8,094,079	11	37.4	4	84.0	L
HNRNPUL2	chr11	62,707,624	62,732,385	11	-20.1	9	76.0	L
IQGAP1	chr15	90,383,240	90,507,243	11	-9.3	5	166.0	L

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
KMT2C	chr7	152,129,924	152,441,005	11	-1.5	5	7.0	L
MED13L	chr12	115,953,575	116,282,186	11	15.4	38	7.0	A
NCOA1	chr2	24,579,476	24,775,701	11	-0.5	3	22.0	L
PCNX1	chr14	70,902,404	71,120,382	11	-9.4	153	14.0	A
POT1	chr7	124,817,385	124,934,983	11	46.8	26	0.0	A
RBM39	chr20	35,698,608	35,747,336	11	164.9	2	14.0	E
SMCHD1	chr18	2,650,886	2,810,017	11	-17.6	4	7.0	L
STK4	chr20	44,961,473	45,084,977	11	-7.3	23	11.0	A
STXBP5	chr6	147,199,357	147,395,476	11	49.8	8	350.0	L
UBR4	chr1	19,069,505	19,215,252	11	11.3	8	7.0	L
ZGPAT	chr20	63,702,441	63,741,142	11	-52.7	5	7.0	DL
ZNF251	chr8	144,715,908	144,760,585	11	-22.9	21	22.0	L
ZNF407	chr18	74,625,962	75,070,672	11	109.6	3	1.0	E
ADD1	chr4	2,838,856	2,935,075	10	-5.6	42	7.0	A
ARIH1	chr15	72,469,325	72,591,555	10	85.4	5	7.0	L
ATF7	chr12	53,502,855	53,631,415	10	-9.8	32	1.0	A
CAMK2D	chr4	113,446,031	113,766,927	10	17.1	9	106.0	L
CLTC	chr17	59,614,688	59,701,956	10	107.3	3	1.0	E
COX6B1	chr19	35,643,222	35,663,784	10	-4.6	5	825.0	L
DDX60	chr4	168,211,290	168,323,807	10	111.9	22	122.0	L
DENND1B	chr1	197,499,748	197,780,493	10	-15.2	3	7.0	L
FAM117B	chr2	202,630,177	202,774,757	10	68.6	6	14.0	E
FNBP1	chr9	129,882,186	130,048,194	10	-27.1	5	256.0	L
FOXJ3	chr1	42,171,538	42,340,877	10	29.2	12	7.0	L
FRYL	chr4	48,492,362	48,785,299	10	27.2	24	14.0	A
IL4I1	chr19	49,884,655	49,934,539	10	-36.7	5	46.0	DL
KIAA1468	chr18	62,182,290	62,312,122	10	34.3	17	46.0	L
LOC101926943	chr7	74,683,936	74,733,918	10	106.0	7	0.0	E
LRPPRC	chr2	43,881,223	44,001,005	10	363.6	7	0.0	E
MOB3A	chr19	2,066,035	2,101,270	10	-12.2	2	22.0	L
NBEAL1	chr2	203,009,878	203,222,994	10	72.2	11	14.0	L
NELL2	chr12	44,503,274	44,918,928	10	39.1	2	4.0	L
NOSIP	chr19	49,550,467	49,585,572	10	-63.6	29	0.0	DA
PHF3	chr6	63,630,801	63,720,522	10	164.9	7	4.0	E
PLEC	chr8	143,910,146	143,981,745	10	-59.4	11	22.0	DL
PLEKHA5	chr12	19,124,691	19,381,399	10	41.8	2	7.0	L
COP1	chr1	175,939,825	176,212,244	10	125.2	3	14.0	E
RUNX2	chr6	45,323,316	45,556,082	10	94.3	8	14.0	E
SLC6A16	chr19	49,284,634	49,330,217	10	-8.7	2	14.0	L
SNAPC4	chr9	136,370,568	136,403,437	10	29.8	7	166.0	L
SNX13	chr7	17,785,760	17,945,508	10	85.4	5	14.0	E

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
TANC2	chr17	63,004,536	63,432,706	10	105.0	11	7.0	E
TCF20	chr22	42,155,012	42,288,927	10	-7.3	10	7.0	L
TRAPPC8	chr18	31,824,172	31,948,128	10	49.8	4	7.0	L
UBAC2	chr13	99,195,424	99,391,499	10	36.0	4	14.0	L
VMP1	chr17	59,702,464	59,847,255	10	-36.2	28	46.0	DA
VPS52	chr6	33,245,271	33,276,965	10	-9.2	10	106.0	L
WDR82	chr3	52,249,421	52,283,643	10	3.0	7	277.5	L
ZFC3H1	chr12	71,604,600	71,668,969	10	136.0	4	0.0	E
ADK	chr10	74,146,184	74,714,303	9	-16.1	3	7.0	L
AP2B1	chr17	35,582,262	35,731,417	9	39.1	13	90.0	L
ASXL2	chr2	25,728,752	25,883,516	9	-41.0	3	7.0	L
ATG5	chr6	106,179,476	106,330,820	9	252.3	2	0.0	E
BCAS3	chr17	60,672,774	61,397,838	9	-34.8	2	7.0	L
C6orf106	chr6	34,582,279	34,701,850	9	-11.9	3	46.0	L
CAMK4	chr5	111,218,652	111,499,884	9	4.3	9	7.0	L
DAP3	chr1	155,684,090	155,744,009	9	131.8	2	0.0	E
DAZAP1	chr19	1,402,568	1,440,687	9	-32.6	2	4.0	L
DNAJC13	chr3	132,412,659	132,544,032	9	98.7	71	7.0	A
ERC1	chr12	986,207	1,500,933	9	3.5	4	7.0	L
FOCAD	chr9	20,653,308	21,000,955	9	140.0	17	7.0	E
FRG1BP	chr20	30,372,163	30,424,842	9	224.5	4	7.0	E
GAK	chr4	844,274	937,390	9	156.8	4	14.0	E
GANAB	chr11	62,619,825	62,651,726	9	-19.6	6	7.0	L
GPBP1L1	chr1	45,622,303	45,691,630	9	-38.2	2	46.0	L
GPHN	chr14	66,502,406	67,186,808	9	131.8	2	0.0	E
HNRNPUL1	chr19	41,257,475	41,312,783	9	68.6	4	14.0	L
HTT	chr4	3,069,680	3,248,960	9	6.0	25	1.0	A
KDM4A	chr1	43,645,125	43,710,518	9	215.2	7	0.0	E
KIF20B	chr10	89,696,589	89,779,943	9	167.9	13	0.0	E
LCOR	chr10	96,827,259	96,991,212	9	113.3	4	0.0	E
MUM1	chr19	1,349,976	1,383,431	9	-43.7	4	90.0	DL
NAA38	chr17	7,851,680	7,890,388	9	-36.3	15	14.0	DL
NDUFV2	chr18	9,097,629	9,139,345	9	119.2	15	22.0	L
NEAT1	chr11	65,417,797	65,450,538	9	2.0	4	22.0	L
NEMP1	chr12	57,050,642	57,083,791	9	-16.9	10	46.0	L
NUP107	chr12	68,681,950	68,750,814	9	456.3	4	0.0	E
NUP214	chr9	131,120,560	131,239,670	9	-2.1	8	106.0	L
PDCD10	chr3	167,678,905	167,739,863	9	410.0	4	0.0	E
PHF20L1	chr8	132,770,357	132,853,807	9	54.5	4	22.0	L
POGZ	chr1	151,397,723	151,464,465	9	23.6	26	0.0	A
POLA2	chr11	65,256,851	65,303,685	9	-20.5	9	7.0	L

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
PTPRA	chr20	2,859,194	3,043,669	9	-13.3	4	1555.0	L
RAB11FIP2	chr10	117,999,915	118,051,884	9	456.3	19	1.0	E
RBPJ	chr4	26,314,709	26,440,130	9	209.1	2	0.0	E
RNF216	chr7	5,615,040	5,786,730	9	-48.5	15	106.0	DL
ROCK1	chr18	20,944,741	21,116,851	9	93.9	13	14.0	E
RTTN	chr18	69,998,805	70,210,726	9	2.3	6	136.0	L
RUNX1	chr21	34,782,800	35,054,298	9	6.8	1	110.0	L
SEPT7	chr7	35,795,985	35,912,105	9	107.3	10	7.0	E
MTREX	chr5	55,302,747	55,430,581	9	162.7	6	0.0	E
SNTB1	chr8	120,530,744	120,817,069	9	147.3	3	4.0	E
SPEN	chr1	15,842,863	15,945,455	9	2.3	2	7.0	L
STAG1	chr3	136,332,156	136,757,403	9	29.8	6	136.0	L
TARSL2	chr15	101,648,751	101,729,442	9	71.2	22	7.0	L
THEMIS	chr6	127,703,193	127,923,631	9	-16.0	5	106.0	L
TONSL	chr8	144,423,779	144,449,429	9	-29.9	7	7.0	L
TTC21B	chr2	165,868,361	165,958,838	9	15.9	5	46.0	L
TUT1	chr11	62,570,044	62,596,637	9	-31.3	5	7.0	L
URI1	chr19	29,918,643	30,021,612	9	199.6	6	1.0	E
WWP1	chr8	86,337,764	86,472,949	9	172.0	4	7.0	E
ZC3H13	chr13	45,949,464	46,057,778	9	152.9	6	46.0	E
ABCD2	chr12	39,546,219	39,625,041	8	131.8	8	14.0	E
ACOX1	chr17	75,936,510	75,984,434	8	-21.7	2	256.0	L
ASCC3	chr6	100,503,194	100,886,372	8	178.2	4	7.0	E
ATP2A2	chr12	110,276,226	110,356,092	8	12.9	67	22.0	A
ATP8A1	chr4	42,403,374	42,662,105	8	33.5	4	14.0	L
AUH	chr9	91,208,814	91,366,969	8	29.0	23	7.0	A
CASK	chrX	41,509,935	41,928,034	8	-4.5	5	346.0	L
DOT1L	chr19	2,159,148	2,237,578	8	-18.1	6	14.0	L
DYNC1H1	chr14	101,959,527	102,055,798	8	71.2	44	7.0	A
EED	chr11	86,239,383	86,283,810	8	136.0	6	14.0	E
HSF2	chr6	122,394,550	122,438,119	8	419.2	15	14.0	E
IKZF2	chr2	212,994,685	213,156,609	8	201.3	1	0.0	E
KLRG1	chr12	8,945,043	9,015,744	8	-11.0	3	7.0	L
LUC7L2	chr7	139,335,358	139,428,457	8	209.1	4	14.0	E
MAP4K3	chr2	39,244,265	39,442,312	8	29.8	15	46.0	L
MMP23A	chr1	1,627,779	1,706,808	8	5.1	17	14.0	L
NBAS	chr2	15,161,907	15,566,348	8	6.0	4	106.0	L
NCOA3	chr20	47,496,856	47,661,877	8	-35.1	6	7.0	L
NUP62	chr19	49,901,825	49,934,731	8	-40.1	5	46.0	DL
PDS5B	chr13	32,581,426	32,783,020	8	85.4	10	256.0	L
PPP1R16A	chr8	144,472,981	144,507,121	8	-36.6	4	7.0	L

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
RAD51B	chr14	67,814,778	68,688,106	8	-21.5	2	22.0	L
RBL2	chr16	53,429,419	53,496,648	8	102.3	8	14.0	L
SLX4IP	chr20	10,430,302	10,633,034	8	102.3	3	7.0	L
SPG7	chr16	89,503,387	89,562,768	8	-30.5	10	7.0	L
SYNE2	chr14	63,847,964	64,231,451	8	65.9	8	22.0	L
UBE2L3	chr22	21,544,446	21,629,034	8	52.7	3	7.0	L
UBR5	chr8	102,247,273	102,417,689	8	21.2	19	7.0	L
VPS28	chr8	144,418,600	144,433,563	8	9.1	7	7.0	L
VRK3	chr19	49,971,466	50,030,548	8	-34.6	2	7.0	L
YTHDF3	chr8	63,163,552	63,217,788	8	70.0	13	7.0	L
ANXA1	chr9	73,146,730	73,175,394	7	155.0	5	14.0	L
ATE1	chr10	121,735,420	121,933,801	7	240.0	14	46.0	E
BCKDHB	chr6	80,101,609	80,351,270	7	240.0	3	0.0	E
CCDC47	chr17	63,740,249	63,778,728	7	-42.1	6	22.0	L
CDC73	chr1	193,116,957	193,259,812	7	150.9	19	7.0	E
CDK8	chr13	26,249,103	26,410,236	7	209.1	1	0.0	E
CHMP2B	chr3	87,222,262	87,260,548	7	85.4	24	7.0	A
CLASP2	chr3	33,491,245	33,723,213	7	36.6	5	7.0	L
DERL2	chr17	5,466,250	5,491,230	7	48.4	37	0.0	A
DNAJC1	chr10	21,751,547	22,008,721	7	4.3	38	7.0	A
GATAD2B	chr1	153,799,906	153,927,975	7	-43.3	2	7.0	DL
GTDC1	chr2	143,941,013	144,337,534	7	138.4	9	7.0	E
INO80	chr15	40,973,880	41,121,246	7	-25.8	3	14.0	L
KMT2D	chr12	49,013,974	49,060,324	7	12.6	33	0.0	A
LSM2	chr6	31,792,391	31,811,984	7	-67.0	2	14.0	DL
MACROD2	chr20	13,990,499	16,058,196	7	224.5	1	0.0	E
MATR3	chr5	139,268,751	139,336,677	7	125.2	6	1.0	E
MIR5096	chr1	15,866,148	15,910,467	7	11.3	2	7.0	DL
NFKBIL1	chr6	31,541,850	31,563,829	7	-47.0	4	7.0	DL
OPRM1	chr6	154,005,495	154,251,867	7	20.0	3	14.0	L
PAG1	chr8	80,962,810	81,117,068	7	-13.9	2	32.0	L
PCNT	chr21	46,319,121	46,450,769	7	6.0	5	4.0	L
PDE12	chr3	57,551,246	57,661,480	7	201.3	3	0.0	E
PDE7A	chr8	65,709,333	65,846,734	7	-10.5	4	7.0	L
PHACTR4	chr1	28,364,581	28,505,369	7	-38.2	3	22.0	L
PPP4R2	chr3	72,991,742	73,074,201	7	240.0	2	7.0	E
PRKCA	chr17	66,297,807	66,815,744	7	62.3	7	7.0	L
PRKN	chr6	161,342,557	162,732,802	7	191.4	3	0.0	E
RAD23B	chr9	107,278,235	107,337,194	7	363.6	2	0.0	E
RASA1	chr5	87,263,252	87,396,926	7	140.0	3	0.0	E
RFX2	chr19	5,988,163	6,115,653	7	-33.4	9	106.0	L

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
RIPOR2	chr6	24,799,280	25,047,288	7	-16.3	10	7.0	L
SNORA30	chr16	30,705,536	30,715,665	7	54.5	373	332.0	L
SPPL3	chr12	120,757,509	120,909,352	7	-35.5	9	14.0	L
SYNE1	chr6	152,116,683	152,642,399	7	36.6	22	46.0	L
TCF25	chr16	89,868,585	89,916,384	7	-35.1	5	106.0	L
UBE2J2	chr1	1,248,911	1,278,854	7	-62.4	2	166.0	DL
UCHL3	chr13	75,544,479	75,611,020	7	217.9	6	7.0	E
UNKL	chr16	1,358,204	1,419,720	7	15.9	7	15.0	L
USP9Y	chrY	12,696,230	12,865,843	7	209.1	4	18.0	E
ZNF473	chr19	50,020,892	50,053,774	7	-7.3	3	22.0	L
ABLIM1	chr10	114,426,109	114,773,225	6	308.0	6	14.0	E
ARHGAP12	chr10	31,800,397	31,933,876	6	270.9	3	14.0	E
ATP9B	chr18	79,064,274	79,383,282	6	13.3	2	7.0	L
BAG6	chr6	31,634,027	31,657,700	6	-32.2	36	46.0	A
BZW2	chr7	16,641,133	16,711,523	6	209.1	6	0.0	E
CAMKMT	chr2	44,356,903	44,777,592	6	308.0	9	0.0	E
CEP85L	chr6	118,455,771	118,715,075	6	-15.7	3	46.0	L
EP400P1	chr12	132,079,282	132,131,340	6	66.9	5	260.0	L
EVL	chr14	99,966,474	100,149,236	6	-23.3	4	7.0	L
FUNDC2	chrX	155,021,788	155,061,916	6	201.3	9	46.0	E
GNA12	chr7	2,723,105	2,849,325	6	240.0	2	0.0	E
HERC2	chr15	28,106,036	28,327,152	6	518.1	4	1.0	E
HSF5	chr17	58,415,166	58,493,401	6	-7.3	34	1.0	A
MARF1	chr16	15,589,368	15,648,166	6	75.7	17	106.0	L
KIFC1	chr6	33,386,535	33,414,922	6	-2.9	42	7.0	A
MBD3	chr19	1,571,670	1,597,761	6	58.9	2	7.0	L
MIR5096	chr22	37,663,025	38,029,093	6	-50.8	6	7.0	DL
N4BP1	chr16	48,533,725	48,615,209	6	85.4	32	106.0	A
OXCT1	chr5	41,725,064	41,875,689	6	54.5	3	7.0	L
PAPOLA	chr14	96,497,375	96,572,116	6	128.2	3	0.0	E
PHF20	chr20	35,767,000	35,955,366	6	-24.7	38	7.0	A
PPFIA1	chr11	70,265,699	70,389,501	6	23.6	20	106.0	L
PPP1CB	chr2	28,746,747	28,807,940	6	-25.8	23	22.0	A
PTGES3	chr12	56,658,340	56,693,408	6	-47.0	23	1.0	A
RAB18	chr10	27,499,173	27,547,237	6	29.8	24	7.0	A
RBM27	chr5	146,198,599	146,294,221	6	345.1	3	0.0	E
RPRD2	chr1	150,359,110	150,481,565	6	-53.6	6	14.0	DL
RSBN1L	chr7	77,691,425	77,784,803	6	48.4	31	1.0	A
SEC23A	chr14	39,026,918	39,108,528	6	33.9	7	7.0	L
SEC31A	chr4	82,813,508	82,905,571	6	29.8	66	7.0	A
SFI1	chr22	31,491,138	31,623,551	6	-63.6	24	7.0	DA

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
SMAP2	chr1	40,368,705	40,428,326	6	-7.3	61	7.0	A
TMTC3	chr12	88,137,295	88,204,887	6	301.8	2	0.0	E
TNKS	chr8	9,550,934	9,787,346	6	162.7	8	7.0	E
TRIM33	chr1	114,387,776	114,516,160	6	104.0	17	106.0	L
UBE2F-SCLY	chr2	237,961,944	238,104,413	6	-7.3	6	46.0	L
WWOX	chr16	78,094,412	79,217,667	6	39.1	5	106.0	L
ZNRD1ASP	chr6	29,996,010	30,066,189	6	-7.3	4	122.0	L
ANKRD46	chr8	100,504,751	100,564,786	5	-14.4	24	7.0	A
AQR	chr15	34,851,350	34,974,794	5	23.6	84	14.0	A
CNOT6	chr5	180,489,398	180,583,405	5	-56.6	2	7.0	DL
CSNK1G1	chr15	64,160,516	64,361,259	5	-19.4	8	7.0	L
ELMO1	chr7	36,847,905	37,454,326	5	-38.2	34	46.0	A
FANCL	chr2	58,154,242	58,246,380	5	191.4	6	0.0	E
FUS	chr16	31,175,109	31,199,871	5	217.9	12	0.0	E
IQCB1	chr3	121,764,760	121,840,079	5	85.4	79	15.0	A
KDM5D	chrY	19,700,414	19,749,939	5	178.2	2	14.0	A
MAP2K2	chr19	4,085,321	4,129,129	5	-32.0	2	106.0	L
MEMO1	chr2	31,862,809	32,016,052	5	-51.2	47	7.0	A
MIR5096	chr1	235,507,822	235,723,113	5	116.3	1	0.0	DL
NAP1L1	chr12	76,039,744	76,090,033	5	196.7	23	7.0	A
NHLRC2	chr10	113,849,631	113,917,506	5	44.2	23	7.0	A
OGDH	chr7	44,601,521	44,714,070	5	-44.4	17	332.0	L
POM121	chr7	72,874,334	72,956,440	5	224.5	29	1.0	A
QKI	chr6	163,409,642	163,583,596	5	45.7	3	7.0	L
RMND5A	chr2	86,715,290	86,783,041	5	15.9	26	0.0	A
RPA3	chr7	7,631,562	7,723,607	5	85.4	87	106.0	A
SEPT9	chr17	77,276,409	77,505,596	5	-57.6	27	7.0	A
SNRPA	chr19	40,745,853	40,770,392	5	209.1	19	7.0	E
STAG3	chr7	100,172,723	100,219,387	5	-11.3	35	0.0	A
STX8	chr17	9,245,470	9,580,958	5	-2.4	4	7.0	L
XPO1	chr2	61,472,933	61,543,283	5	85.4	25	106.0	A
AKAP9	chr7	91,935,874	92,115,673	4	-1.8	25	0.0	A
CAAP1	chr9	26,835,684	26,897,828	4	-7.3	27	46.0	A
EHMT1	chr9	137,758,021	137,769,772	4	131.8	1	0.0	L
EXOSC10	chr1	11,061,612	11,104,910	4	-0.1	70	7.0	A
GOLPH3L	chr1	150,641,224	150,702,196	4	-53.6	24	7.0	A
ITM2B	chr13	48,228,137	48,267,096	4	-7.3	23	7.0	A
MSH5-SAPCD1	chr6	31,734,947	31,769,847	4	-62.9	43	7.0	A
PATL1	chr11	59,631,715	59,674,038	4	-7.3	578	332.0	A
PDCD11	chr10	103,391,654	103,451,262	4	6.0	27	152.0	A
PDE3B	chr11	14,638,722	14,877,058	4	2.3	35	7.0	A

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
RABGAP1	chr9	122,936,008	123,109,868	4	6.0	29	7.0	A
TRIO	chr5	14,138,701	14,515,204	4	332.7	61	106.0	A
CHD1L	chr1	147,168,193	147,300,766	3	641.8	25	1.0	A
DCUN1D4	chr4	51,837,999	51,921,837	3	-22.7	32	106.0	A
EIF2AK4	chr15	39,929,123	40,040,596	3	54.5	43	7.0	A
GPN1	chr2	27,623,647	27,655,846	3	11.3	62	1.0	A
KCTD3	chr1	215,562,378	215,626,821	3	39.1	1	0.0	A
LOC101927151	chr19	27,788,466	27,811,780	3	456.3	31	46.0	A
NGDN	chr14	23,464,688	23,483,193	3	39.1	44	0.0	A
POLG2	chr17	64,472,784	64,502,066	3	116.3	43	7.0	A
PRKD2	chr19	46,669,315	46,722,127	3	-45.5	24	22.0	A
SNAP29	chr22	20,854,003	20,896,213	3	147.3	27	7.0	A
ZNF573	chr19	37,733,301	37,784,590	3	-25.8	86	735.0	A
C20orf196	chr20	5,745,386	5,869,407	2	-47.0	1	0.0	A
CRTAP	chr3	33,108,957	33,152,773	2	270.9	35	7.0	A
GRSF1	chr4	70,810,781	70,844,910	2	Inf	23	0.0	A
JMJD6	chr17	76,707,831	76,731,799	2	164.9	53	1.0	A
LINC01473	chr2	186,028,533	186,091,317	2	Inf	82	7.0	A
PHF12	chr17	28,900,252	28,956,490	2	-42.9	40	106.0	A
RASEF	chr9	82,974,584	83,068,128	2	Inf	43	46.0	A
SNHG12	chr1	28,573,537	28,586,854	2	456.3	96	332.0	A
TAC3	chr12	57,004,996	57,021,560	2	-71.5	42	7.0	A
TGFBR2	chr3	30,601,501	30,699,141	2	23.6	31	7.0	A
ACTL6A	chr3	179,557,879	179,593,405	1	-69.1	40	0.0	A
C19orf48	chr19	50,792,692	50,809,853	1	-53.6	28	46.0	A
CD109	chr6	73,691,084	73,833,317	1	-85.7	32	0.0	A
IFNGR2	chr21	33,397,895	33,442,521	1	Inf	53	15.0	A
KARS	chr16	75,622,723	75,652,687	1	-7.3	24	0.0	A
LOC101927501	chrX	43,171,993	43,231,598	1	Inf	23	1.0	A
MICAL2	chr11	12,105,575	12,268,790	1	-73.5	39	15.0	A
RBAK-RBAKDN	chr7	5,040,820	5,078,223	1	-53.6	28	7.0	A
RTCA-AS1	chr1	100,259,741	100,271,174	1	Inf	27	0.0	A
UXT-AS1	chrX	47,653,832	47,665,111	1	85.4	43	7.0	A
ZNF92	chr7	65,368,798	65,406,135	1	Inf	42	7.0	A
TNRC6B	chr22	40,039,816	40,340,808	17	-29.3	10	46.0	D
RBM6	chr3	49,935,043	50,082,252	15	-31.8	7	1.0	D
MIR1268A	chr9	128,347,046	128,667,136	14	-29.8	4	1.0	D
NFATC3	chr16	68,080,365	68,234,259	14	-30.6	4	4.0	D
CCNL2	chr1	1,380,710	1,404,338	13	-32.6	8	22.0	D
NUP188	chr9	128,942,692	129,012,096	13	-38.9	12	1.0	D
IKZF3	chr17	39,752,714	39,869,188	12	-38.2	4	46.0	D

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
UBE2G1	chr17	4,264,216	4,371,674	12	-33.8	3	14.0	D
FOXK2	chr17	82,514,717	82,609,607	11	-40.2	19	0.0	D
IP6K1	chr3	49,719,294	49,791,540	11	-61.3	2	1.0	D
RABL6	chr9	136,802,921	136,846,187	11	-36.7	11	0.0	D
CCND3	chr6	41,929,932	42,053,894	10	-57.0	3	0.0	D
EIF2B3	chr1	44,845,521	44,991,722	10	-61.1	6	1.0	D
R3HDM2	chr12	57,248,763	57,436,005	10	-41.0	2	0.0	D
RERE	chr1	8,347,403	8,822,640	10	-46.7	5	0.0	D
SP1	chr12	53,375,194	53,421,442	10	-47.9	10	0.0	D
STAT5B	chr17	42,194,176	42,281,406	10	-42.2	2	0.0	D
TRAF2	chr9	136,881,512	136,931,615	10	-49.6	4	46.0	D
ABHD16A	chr6	31,681,948	31,708,360	9	-51.2	6	0.0	D
CEACAM21	chr19	41,544,517	41,591,844	9	-49.2	2	0.0	D
HCG20	chr6	30,761,824	30,797,250	9	-52.6	4	22.0	D
ITGAL	chr16	30,467,661	30,528,185	9	-45.5	14	14.0	D
NARFL	chr16	724,754	746,038	9	-58.8	3	22.0	D
PSMB9	chr6	32,849,160	32,864,851	9	-55.4	2	0.0	D
RBM4	chr11	66,633,616	66,673,386	9	-48.7	6	11.0	D
TSC2	chr16	2,042,894	2,093,720	9	-53.6	1	22.0	D
HORMAD2	chr22	30,075,068	30,182,075	8	-54.1	4	7.0	D
IFT140	chr16	1,505,426	1,617,108	8	-57.5	3	7.0	D
PPP3CC	chr8	22,435,969	22,546,144	8	-48.1	3	0.0	D
QRICH1	chr3	49,024,706	49,099,373	8	-55.7	7	1.0	D
TAP2	chr6	32,816,832	32,843,823	8	-54.6	4	7.0	D
VARS	chr6	31,772,519	31,800,935	8	-47.3	5	46.0	D
WDR90	chr16	644,362	672,829	8	-61.5	3	0.0	D
ASCC1	chr10	72,091,031	72,222,134	7	-48.2	2	0.0	D
PRRC2A	chr6	31,615,672	31,642,777	7	-48.3	13	14.0	D
RAB40C	chr16	584,356	634,273	7	-52.6	3	0.0	D
RBM14-RBM4	chr11	66,611,581	66,651,473	7	-51.4	6	11.0	D
SEPT2	chr2	241,310,186	241,359,026	7	-50.2	8	0.0	D
STK11	chr19	1,200,798	1,233,435	7	-60.5	6	0.0	D
ADCK5	chr8	144,369,014	144,398,238	6	-48.7	2	0.0	D
BLM	chr15	90,712,326	90,820,462	6	-41.3	11	14.0	D
CPSF1	chr8	144,388,230	144,414,349	6	-42.2	4	0.0	D
DIDO1	chr20	62,872,737	62,942,952	6	-55.8	2	0.0	D
GRAP2	chr22	39,896,081	39,978,342	6	-57.9	4	4.0	D
MCM3AP	chr21	46,230,124	46,290,394	6	-51.2	8	22.0	D
PCED1B	chr12	47,074,602	47,241,663	6	-54.7	1	0.0	D
PRKAR2A	chr3	48,741,578	48,852,850	6	-58.4	2	7.0	D
RNPS1	chr16	2,248,115	2,273,412	6	-78.4	2	0.0	D

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
WASF2	chr1	27,399,225	27,495,187	6	-48.2	3	0.0	D
WNK1	chr12	747,922	916,452	6	-60.3	4	1.0	D
ZBTB4	chr17	7,454,365	7,489,249	6	-62.2	3	0.0	D
EXOC2	chr6	480,137	698,141	5	-71.9	4	0.0	D
HAGH	chr16	1,804,102	1,832,194	5	-58.0	6	14.0	D
MIR1268A	chr19	2,997,812	3,069,714	5	8.2	3	0.0	D
TC2N	chr14	91,774,751	91,872,536	5	-49.9	4	0.0	D
ZNF598	chr16	1,992,651	2,014,821	5	-63.6	3	0.0	D
FAM222B	chr17	28,750,977	28,847,839	4	-52.5	5	0.0	D
PCBP3	chr21	45,638,724	45,947,454	4	-54.8	2	0.0	D
MIR1268A	chr15	28,320,482	28,505,841	2	Inf	1	0.0	D

Supplementary Report 2:
Genes of interest marked by vector integration
CLL Patients only (CR/PRtd & PR/NR Response Groups)

Contents

Summary

Lentiviral vectors integrate into genomes of targeted host cells (Tcells). 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 mined the data collected from integration site sequencing for 29 CART treated subjects. We constructed 4 gene lists based on: 1 & 2) increased / decreased integration site occurrence in patient samples relative to the initial transduction product, 3) peak clonal abundance, and 4) longitudinal clonal persistence. More about each of these criteria is below:

- **Integration Frequency** is the rate at which integration sites are observed within a gene. This is compared between patient samples and the initial transduction product to score enrichment or depletion during growth in patients. The top of genes with higher patient sample integration frequency over transduction samples were chosen for study (p-value <= 0.05 after exclusion of genes with clones from less than 2 patients and less than 10 observed clones).
- **Clonal Abundance** can be determined during analysis by quantifying the number of sites of linker ligation associated with each unique integration site. This method is further described in **Berry *et al.* 2012**. This allows clonal expansion to be quantified. The top 1% of the genes were selected for study based on their maximal peak clonal abundance.
- **Longitudinal Observation** of clones is the quantification of observed timespans and last observed timepoints. The maximum value for clones within a gene were considered for characterization of the gene in this analysis. Genes were only considered if there were 10 or more integration sites isolated from at least two different patient samples. Genes were also not considered if they only consited of clones which were observed once or the last observed timepoint was less than 90 days from initial infusion.

A point to keep in mind through all this analysis is that integration sites are sampled from a larger population. It would be rare for all integration sites in a sample to be represented in the sequence data.

Table 1: Summary of each filtering criteria.

	Gene	Onco	Tumor	Lymphoma	COSMIC	TCGA	Clonal Hema.
Criteria	Count	Related (%)	Suppressors (%)	Related (%)	Related (%)	Related (%)	Related (%)
Enrichment	64	*/* 25.00	*/* 12.50	/ 0.000	/ 7.81	*/* 9.38	/ 1.562
Depletion	38	/ 7.89	/ 10.53	/ 0.000	/ 2.63	/ 7.89	/ 0.000
Abundance	120	*/* 20.83	/ 5.83	/ 0.000	*/* 9.17	/ 5.00	/ 0.833
Longitudinal	208	*/* 25.48	*/* 12.02	/ 0.481	*/* 12.50	*/* 12.02	*/* 1.442
Composite	343	*/* 22.16	*/* 8.75	/ 0.292	*/* 10.50	*/* 9.33	*/ 0.875

Table 1 summarizes the size and contents of each criteria gene list identified by the various methods. Significance of overlap between lists are displayed by asterisks before the percent of genes identifed 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 Fishers Exact test and multiple comparison corrections were made using a Benjamini-Hochberg (FDR) method for each criteria based list.

Percent of all analyzed transcription units associated with each list as as follows:

- Onco Related: 9.41%
- Tumor Suppressors: 4.91%
- Lymphoma Related: 0.16%
- COSMIC Related: 3.78%
- TCGA Related: 2.88%
- Clonal Hematopoiesis Related: 0.18%

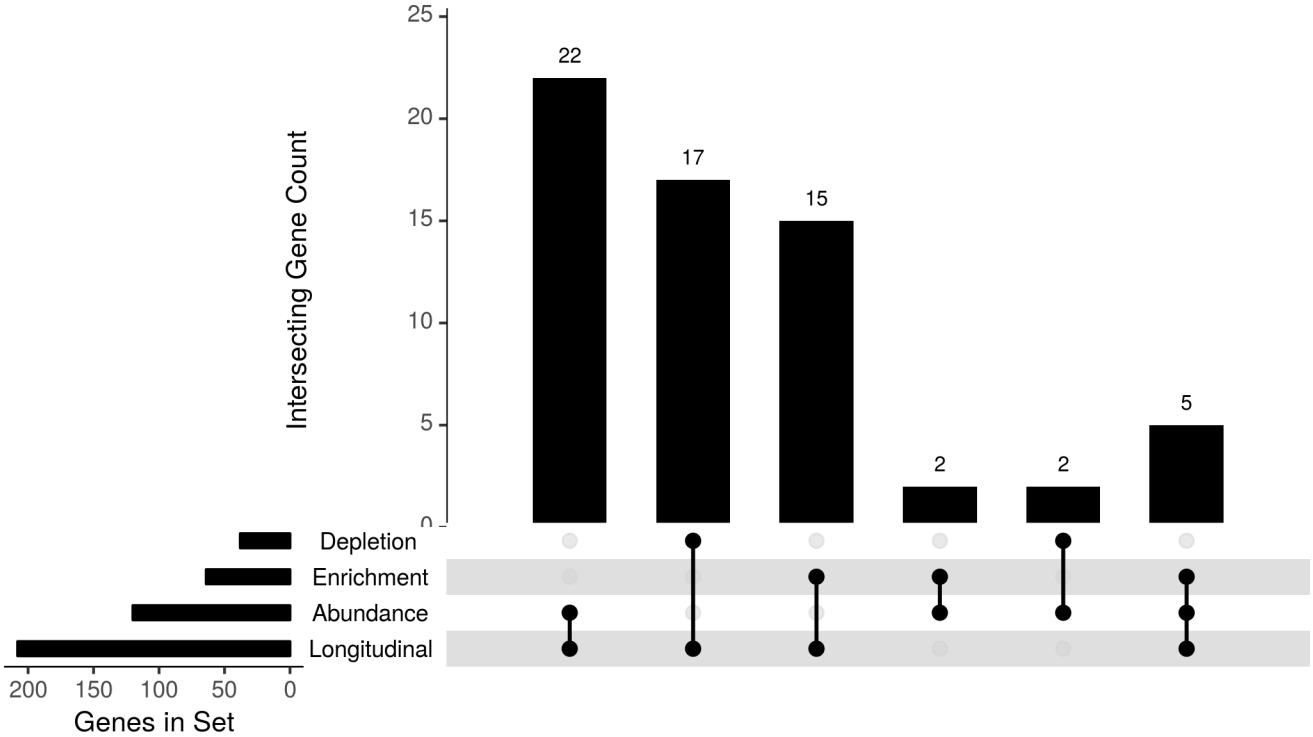

Figure 1: Intersecting gene lists identified through the various selection criteria.

Table 2: The most consistently observed genes from filtering by various criteria. The 'Criteria.' column is a count of how many times the gene was identified by these methods, while the 'Patients' column notes how many specimens collected from patients have had integration sites within the noted gene.

Gene	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
AKAP13	11	191.5	27	360	EAL
UBAP2L	10	122.1	30	180	EAL
CLK4	7	137.5	53	120	EAL
PTBP1	7	175.3	47	360	EAL
TET2	5	264.4	814	1584	EAL

Table 3: GO Biological Process. Top 11 per group. Total genes considered: 314

Group	GO ID	GO Term	Term Size	Gene Count	Adjusted P-value
	GO:0016569	covalent chromatin modification	436	41	0.0000000
	GO:0016570	histone modification	353	36	0.0000000
	GO:0018205	peptidyl-lysine modification	295	30	0.0000000
	GO:0071396	cellular response to lipid	338	22	0.0049811
	GO:0006397	mRNA processing	394	21	0.0406881
1	GO:0048545	response to steroid hormone	248	16	0.0206796
	GO:0043414	macromolecule methylation	219	15	0.0196128
	GO:0061458	reproductive system development	236	15	0.0272872
	GO:0033044	regulation of chromosome organization	243	15	0.0345353
	GO:0030522	intracellular receptor signaling pathway	220	14	0.0342112
	GO:0048608	reproductive structure development	234	14	0.0460150
	GO:0010256	endomembrane system organization	460	33	0.0000274
	GO:0051640	organelle localization	378	23	0.0092226
	GO:0030031	cell projection assembly	442	23	0.0379157
	GO:0044770	cell cycle phase transition	449	23	0.0407726
	GO:0051656	establishment of organelle localization	322	22	0.0028015
2	GO:0044772	mitotic cell cycle phase transition	425	22	0.0407726
	GO:0000226	microtubule cytoskeleton organization	377	20	0.0427348
	GO:0051493	regulation of cytoskeleton organization	323	19	0.0224037
	GO:0000280	nuclear division	336	19	0.0330693
	GO:0048193	Golgi vesicle transport	273	18	0.0128083
	GO:0044782	cilium organization	307	18	0.0285928
	GO:0051223	regulation of protein transport	495	33	0.0001378
	GO:1903827	regulation of cellular protein localization	427	26	0.0046629
	GO:0032386	regulation of intracellular transport	351	25	0.0006210
	GO:0080135	regulation of cellular response to stress	458	25	0.0206796
	GO:1902582	single-organism intracellular transport	484	24	0.0461773
3	GO:0006913	nucleocytoplasmic transport	342	23	0.0026024
	GO:0051169	nuclear transport	347	23	0.0031293
	GO:0006281	DNA repair	432	23	0.0304191
	GO:0033157	regulation of intracellular protein transport	259	20	0.0011311
	GO:0034504	protein localization to nucleus	257	17	0.0160576
	GO:0040029	regulation of gene expression, epigenetic	218	16	0.0077303
	GO:0032101	regulation of response to external stimulus	426	28	0.0008850
	GO:0030155	regulation of cell adhesion	424	27	0.0019013
	GO:0001816	cytokine production	419	24	0.0156509
	GO:0031347	regulation of defense response	453	24	0.0272872
	GO:0046649	lymphocyte activation	432	22	0.0463553
4	GO:0042110	T cell activation	314	19	0.0206796
	GO:0044089	positive regulation of cellular component biogenesis	349	19	0.0407726
	GO:0061061	muscle structure development	348	19	0.0407726
	GO:0002764	immune response-regulating signaling pathway	354	19	0.0445747
	GO:0030099	myeloid cell differentiation	233	18	0.0023140
	GO:0009314	response to radiation	321	18	0.0407726

Table 4: KEGG Pathway analysis. Top 10 per group. Total genes considered: 152

Group	KEGG ID	Description	Term Size	Gene Count	Adjusted P-value
1	path:hsa04144	Endocytosis	200	10	0.1487712
	path:hsa04070	Phosphatidylinositol signaling system	85	7	0.0530465
	path:hsa04810	Regulation of actin cytoskeleton	148	7	0.2174690
	path:hsa00562	Inositol phosphate metabolism	63	6	0.0505688
2	path:hsa05200	Pathways in cancer	338	15	0.1677885
	path:hsa04919	Thyroid hormone signaling pathway	89	11	0.0011710
	path:hsa04140	Autophagy - animal	109	10	0.0229892
	path:hsa05225	Hepatocellular carcinoma	100	9	0.0384941
	path:hsa04714	Thermogenesis	153	9	0.1048879
	path:hsa05206	MicroRNAs in cancer	168	9	0.1317062
	path:hsa05165	Human papillomavirus infection	223	9	0.2976213
	path:hsa04915	Estrogen signaling pathway	78	7	0.0483593
	path:hsa04072	Phospholipase D signaling pathway	107	7	0.1048879
	path:hsa04151	PI3K-Akt signaling pathway	225	6	0.6006387
3	path:hsa04010	MAPK signaling pathway	196	12	0.0689070
	path:hsa05163	Human cytomegalovirus infection	154	9	0.1048879
	path:hsa00310	Lysine degradation	46	8	0.0011710
	path:hsa04141	Protein processing in endoplasmic reticulum	130	8	0.1048879
	path:hsa05167	Kaposi sarcoma-associated herpesvirus infection	118	7	0.1280169
	path:hsa04659	Th17 cell differentiation	78	6	0.0841115
	path:hsa04660	T cell receptor signaling pathway	80	6	0.0878337
	path:hsa04380	Osteoclast differentiation	85	6	0.1030673
	path:hsa04650	Natural killer cell mediated cytotoxicity	85	6	0.1030673
	path:hsa05202	Transcriptional misregulation in cancer	104	6	0.1486516
4	path:hsa05205	Proteoglycans in cancer	128	10	0.0483593
	path:hsa05161	Hepatitis B	118	7	0.1280169
	path:hsa04912	GnRH signaling pathway	63	6	0.0505688
	path:hsa04926	Relaxin signaling pathway	89	6	0.1048879
	path:hsa04722	Neurotrophin signaling pathway	87	6	0.1048879
	path:hsa04068	FoxO signaling pathway	90	6	0.1088547
	path:hsa04015	Rap1 signaling pathway	137	6	0.2860101
	path:hsa04014	Ras signaling pathway	149	6	0.3294764
	path:hsa04664	Fc epsilon RI signaling pathway	48	5	0.0505688
	path:hsa04540	Gap junction	62	5	0.0902081
5	path:hsa04024	cAMP signaling pathway	130	10	0.0483593
	path:hsa05203	Viral carcinogenesis	147	9	0.1030673
	path:hsa05166	Human T-cell leukemia virus 1 infection	165	9	0.1280169
	path:hsa04720	Long-term potentiation	49	8	0.0011710
	path:hsa04114	Oocyte meiosis	95	8	0.0483593
	path:hsa04066	HIF-1 signaling pathway	77	7	0.0483593
	path:hsa04310	Wnt signaling pathway	96	7	0.0841115
	path:hsa05164	Influenza A	110	7	0.1048879
	path:hsa05152	Tuberculosis	109	7	0.1048879
	path:hsa04120	Ubiquitin mediated proteolysis	122	7	0.1317062

Integration Frequency (Enrichment)

Table 5: Table of top 50 genes with the most frequent clonal enrichment.

Gene	Num. Patients	TDN Sites	Patient Sites	Onco-Related	Frequency Increase (%)
RBM27	6	0	12	FALSE	Inf
AGL	4	1	10	FALSE	871.8
MCPH1	8	2	13	TRUE	531.7
BACH2	7	2	12	TRUE	483.1
NDFIP2	6	3	17	FALSE	450.7
GAK	8	3	17	TRUE	450.7
GNA12	6	2	11	TRUE	434.5
FANCL	5	2	11	FALSE	434.5
IKZF2	6	2	11	TRUE	434.5
PIKFYVE	9	2	11	FALSE	434.5
PAPOLA	6	3	16	FALSE	418.3
FAM117B	6	5	25	FALSE	385.9
CD55	9	3	14	TRUE	353.5
EP400P1	6	4	18	FALSE	337.3
FUNDC2	6	3	13	FALSE	321.1
NDUFV2	9	3	13	FALSE	321.1
GPHN	8	3	13	TRUE	321.1
RAB11FIP2	9	3	12	FALSE	288.7
PDS5B	8	4	16	TRUE	288.7
TAF2	7	4	16	FALSE	288.7
UHRF1BP1	8	3	12	FALSE	288.7
RBM39	10	5	19	TRUE	269.3
TET2	5	4	15	TRUE	264.4
LRPPRC	10	4	15	FALSE	264.4
HSF2	8	4	14	FALSE	240.1
MACROD2	7	4	14	FALSE	240.1
SLK	9	4	14	FALSE	240.1
PELP1	11	8	26	FALSE	215.8
ZFAND3	9	4	13	FALSE	215.8
KDM4A	8	5	16	FALSE	211.0
HELLS	9	5	16	FALSE	211.0
RNF10	8	5	16	FALSE	211.0
RASA1	7	7	22	TRUE	205.4
TANC2	8	6	18	FALSE	191.5
YWHAE	8	7	21	TRUE	191.5
ATG5	8	6	18	FALSE	191.5
AKAP13	11	11	33	TRUE	191.5
TNKS	5	5	15	FALSE	191.5
SACM1L	8	6	18	FALSE	191.5
ASCC3	7	11	32	FALSE	182.7
SEPT7	7	6	17	FALSE	175.3
PTBP1	7	6	17	TRUE	175.3
PUM1	9	7	19	FALSE	163.8
CLEC16A	10	12	32	FALSE	159.1
FOCAD	8	8	21	FALSE	155.1
PIAS1	8	8	21	FALSE	155.1
CHD1	6	7	18	FALSE	149.9
JMJD1C	12	11	28	FALSE	147.4
GBE1	10	8	20	FALSE	142.9
ABCD2	8	10	25	FALSE	142.9

Integration Frequency (Depletion)

Table 6: Table of top 38 genes with the most frequent clonal depletion.

Gene	Num. Patients	TDN Sites	Patient Sites	Onco-Related	Frequency Increase (%)
RNPS1	5	69	16	FALSE	-77.5
EXOC2	5	34	10	FALSE	-71.4
EIF2B3	8	42	14	FALSE	-67.6
NARFL	7	25	10	FALSE	-61.1
USP34	5	25	10	FALSE	-61.1
TRAF2	8	53	22	FALSE	-59.7
QRICH1	6	45	19	FALSE	-59.0
UBE2J2	7	33	14	FALSE	-58.8
HAGH	5	28	12	FALSE	-58.4
IP6K1	9	60	26	FALSE	-57.9
BCAS3	6	23	10	TRUE	-57.7
NOSIP	9	64	28	FALSE	-57.5
SFI1	5	22	10	FALSE	-55.8
ZNF598	4	22	10	FALSE	-55.8
NPEPPS	9	26	12	FALSE	-55.1
WDR90	8	23	11	FALSE	-53.5
PPP3CC	7	27	13	FALSE	-53.2
LSM2	7	27	13	FALSE	-53.2
WNK1	5	28	14	FALSE	-51.4
IFT140	6	32	16	FALSE	-51.4
TSC2	8	42	21	TRUE	-51.4
MROH1	8	142	71	FALSE	-51.4
PLEC	8	37	19	FALSE	-50.1
HCG20	7	38	20	FALSE	-48.9
MIR5096	8	35	19	FALSE	-47.2
NFKBIL1	6	31	17	FALSE	-46.7
ANKFY1	8	36	20	FALSE	-46.0
RAB11FIP3	10	54	30	FALSE	-46.0
ZGPAT	10	41	23	FALSE	-45.5
FKBP5	11	53	30	FALSE	-45.0
SIRT3	10	37	21	FALSE	-44.8
HSF1	12	76	45	FALSE	-42.5
MAPK8IP3	10	52	31	FALSE	-42.1
CCDC57	11	53	33	FALSE	-39.5
ASH1L	12	66	46	FALSE	-32.3
PACS1	12	150	105	FALSE	-32.0
DNMT1	11	91	65	TRUE	-30.6
NPLOC4	14	217	155	FALSE	-30.6

Genes with the Most Abundant Clones

Table 7: Table of top 50 Genes containing the highest abundant clones.

Gene	Num. Patients	Peak Abundance	Peak Rel. Abund.	Clonal Gini Index	Onco-Related
TET2	6	814	0.989	0.918	TRUE
KCTD3	3	589	0.265	0.663	FALSE
PATL1	4	578	0.260	0.793	FALSE
PIKFYVE	9	410	0.273	0.885	FALSE
SRCAP	10	373	0.357	0.896	FALSE
MTMR3	4	261	0.041	0.859	TRUE
PCNX1	10	153	0.010	0.825	FALSE
PPP6R3	12	149	0.040	0.733	FALSE
SSH2	8	137	0.062	0.805	FALSE
RSRC1	8	109	0.014	0.812	FALSE
SNHG12	2	96	0.057	0.646	FALSE
MAPK14	7	91	0.006	0.784	TRUE
RPA3	4	87	0.020	0.752	FALSE
ZNF573	3	86	0.610	0.677	FALSE
MGA	11	85	0.013	0.754	FALSE
AQR	4	84	0.022	0.790	FALSE
LEF1	7	84	0.038	0.771	TRUE
LINC01473	3	82	0.075	0.643	FALSE
CARD8	11	79	0.056	0.702	TRUE
IQCB1	5	79	0.028	0.752	FALSE
DNAJC13	8	71	0.004	0.765	FALSE
EXOSC10	4	70	0.008	0.776	FALSE
ATP2A2	8	67	0.030	0.749	FALSE
SEC31A	5	66	0.004	0.736	FALSE
GPN1	2	62	0.017	0.711	FALSE
SMAP2	5	61	0.004	0.758	FALSE
TRIO	6	61	0.025	0.769	TRUE
ZZEF1	11	56	0.333	0.626	FALSE
CLK4	7	53	0.036	0.659	FALSE
IFNGR2	2	53	0.722	0.635	TRUE
JMJD6	2	53	0.015	0.755	FALSE
KDM5D	7	51	0.017	0.747	FALSE
UBR1	7	48	0.421	0.722	FALSE
MEMO1	5	47	0.006	0.733	FALSE
PTBP1	7	47	0.043	0.672	TRUE
DYNC1H1	8	44	0.003	0.709	FALSE
NGDN	3	44	0.005	0.623	FALSE
EIF2AK4	3	43	0.003	0.659	FALSE
MSH5-SAPCD1	4	43	0.039	0.708	FALSE
POLG2	1	43	0.003	0.609	FALSE
RASEF	2	43	0.005	0.622	FALSE
UXT-AS1	2	43	0.039	0.477	FALSE
ADD1	8	42	0.011	0.608	FALSE
GRB2	9	42	0.017	0.592	TRUE
KIFC1	6	42	0.003	0.697	FALSE
TAC3	2	42	0.018	0.477	FALSE
ZNF92	3	42	0.031	0.707	FALSE
ACTL6A	1	40	0.003	0.000	FALSE
ATP6V1G2-DDX39B	11	40	0.005	0.621	FALSE
PHF12	2	40	0.014	0.670	FALSE

Longitudinal Observation

Table 8: Table of top 50 genes identified by longitudinal observations.

Gene	Time Span	Longest Time	Obs. Count	Num. Patients	Patient Sites	Peak Abund.	Onco-Related
FKBP5	1555.0	1825.0	4	11	30	15	FALSE
PTPRA	1555.0	1825.0	3	8	33	4	FALSE
TET2	1464.0	1584.0	7	5	15	814	TRUE
UBR1	1277.5	1825.0	4	7	13	48	FALSE
COX6B1	825.0	1095.0	3	9	17	5	FALSE
CCDC57	642.5	912.5	2	11	33	6	FALSE
KMT5B	642.5	912.5	2	10	26	7	FALSE
MACF1	519.5	547.5	3	12	32	11	TRUE
DNMT1	365.0	912.5	2	11	65	13	TRUE
STXBP5	350.0	360.0	4	8	15	8	FALSE
CASK	346.0	547.5	2	7	16	5	FALSE
RPTOR	346.0	360.0	2	15	85	11	FALSE
DIP2A	346.0	360.0	2	10	35	25	FALSE
PTBP1	346.0	360.0	2	7	17	47	TRUE
MIR4745	346.0	360.0	2	6	10	47	FALSE
ZZEF1	332.0	360.0	5	11	47	56	FALSE
SRCAP	332.0	360.0	5	10	21	373	FALSE
OGDH	332.0	360.0	4	4	10	17	FALSE
WDR82	277.5	547.5	3	7	20	7	TRUE
PIP5K1A	277.5	547.5	2	7	11	3	FALSE
EP400P1	260.0	270.0	3	6	18	5	FALSE
HSF1	256.0	270.0	3	12	45	10	FALSE
BOP1	256.0	270.0	3	11	34	10	TRUE
FNBP1	256.0	270.0	2	9	32	5	TRUE
PDS5B	256.0	270.0	2	8	16	10	TRUE
ACOX1	256.0	270.0	2	7	15	2	FALSE
PIK3C3	180.0	360.0	3	10	30	5	FALSE
IQGAP1	166.0	180.0	3	9	16	5	FALSE
SNAPC4	166.0	180.0	2	7	18	7	FALSE
UBE2J2	166.0	180.0	2	7	14	2	FALSE
SSH2	152.0	1095.0	4	8	26	137	FALSE
MED13	152.0	270.0	4	12	26	21	FALSE
CARD8	152.0	180.0	4	11	38	79	TRUE
LEF1	152.0	180.0	4	7	15	84	TRUE
VAV1	152.0	180.0	3	13	76	37	TRUE
STAG1	136.0	912.5	2	8	13	6	TRUE
PPP6R2	136.0	180.0	2	11	40	15	FALSE
RTTN	136.0	150.0	2	8	15	6	FALSE
MAPK8IP3	130.0	270.0	2	10	31	5	FALSE
SMG1	122.0	150.0	3	12	39	7	FALSE
INPP4B	122.0	150.0	3	11	28	3	FALSE
PIAS1	122.0	150.0	3	8	21	6	FALSE
DDX60	122.0	150.0	3	9	15	22	FALSE
ZNRD1ASP	122.0	150.0	3	6	14	4	FALSE
DPYD	122.0	150.0	2	12	33	22	FALSE
RUNX1	110.0	360.0	2	9	19	1	TRUE
ASH1L	106.0	1825.0	2	12	46	5	FALSE
WWOX	106.0	1095.0	3	5	11	5	TRUE
RFX2	106.0	360.0	2	7	14	9	TRUE
PPFIA1	106.0	270.0	4	5	11	20	FALSE

Reference Data

The NCBI RefGenes data set was used to identify gene regions (hg38) while genes identified as onco-related were from the Bushman Lab curated list of **onco-related genes**.

Gene Ontologies were extracted from the `GO.db` R-package (v3.4.1). KEGG pathways were acquired via interfacing with the KEGG web-server API through the `KEGGREST` R-package (v1.16.1). Gene lists, including RefSeq genes used for annotation of integration sites, were standardized to HGNC gene symbols (date: 2018-02-07). Groups identified in GO and KEGG analyses were determined from Jaccard distances between identified terms, followed by modularity-optimizing clustering from a weighted-undirected graph using a Louvain algorithm (**Blondel *et al.* 2008**). Terms within groups of GO or KEGG terms have greater overlap of gene lists between themselves that between terms found in other groups. This method was implemented to help reduce the functional redundancy commonly observed in GO and overlapping pathways observed with KEGG.

Comprehensive Genes of Interest Table

Table 9: Table of all genes identified within analysis.

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
AKAP13	chr15	85,375,615	85,754,358	11	191.5	27	7.0	EAL
UBAP2L	chr1	154,215,171	154,276,510	10	122.1	30	7.0	EAL
CLK4	chr5	178,597,663	178,632,053	7	137.5	53	106.0	EAL
PTBP1	chr19	792,391	817,327	7	175.3	47	346.0	EAL
TET2	chr4	105,140,874	105,284,803	5	264.4	814	1464.0	EAL
PPP3CA	chr4	101,018,429	101,352,471	14	84.1	5	46.0	EL
LUC7L	chr16	183,968	234,482	13	-10.1	30	7.0	AL
VAV1	chr19	6,767,667	6,862,366	13	39.4	37	152.0	AL
JMJD1C	chr10	63,162,220	63,527,075	12	147.4	5	7.0	EL
PPP6R3	chr11	68,455,717	68,620,333	12	87.2	149	14.0	EA
RNF157	chr17	76,137,452	76,245,311	12	17.6	28	14.0	AL
SMG1P1	chr16	22,432,007	22,497,220	12	115.2	3	100.0	EL
CARD8	chr19	48,203,085	48,260,946	11	15.4	79	152.0	AL
SUPT3H	chr6	44,821,729	45,383,051	11	90.1	8	14.0	EL
XPO5	chr6	43,517,329	43,581,075	11	31.9	26	22.0	AL
ZZEF1	chr17	3,999,444	4,147,959	11	8.7	56	332.0	AL
CRAMP1	chr16	1,609,639	1,682,908	10	1.3	30	18.0	AL
DIP2A	chr21	46,453,948	46,575,013	10	-15.0	25	346.0	AL
EYA3	chr1	27,965,343	28,093,637	10	122.1	7	32.0	EL
NF1	chr17	31,089,926	31,382,677	10	94.4	14	14.0	EL
PAFAH1B1	chr17	2,588,628	2,690,615	10	39.4	30	7.0	AL
PIK3C3	chr18	41,950,197	42,086,482	10	142.9	5	180.0	EL
SRCAP	chr16	30,694,140	30,745,129	10	57.0	373	332.0	AL
GRB2	chr17	75,313,075	75,410,709	9	-22.3	42	14.0	AL
HELLS	chr10	94,540,766	94,607,099	9	211.0	15	106.0	EL
NDUFV2	chr18	9,097,629	9,139,345	9	321.1	15	22.0	EL
PIKFYVE	chr2	208,261,266	208,363,751	9	434.5	410	14.0	EA
SMURF2	chr17	64,539,616	64,667,268	9	32.5	27	7.0	AL
USP25	chr21	15,725,024	15,885,071	9	110.6	10	46.0	EL
MCPH1	chr8	6,401,591	6,653,505	8	531.7	9	7.0	EL
PDCD4	chr10	110,866,794	110,905,006	8	36.0	26	14.0	AL
PDS5B	chr13	32,581,426	32,783,020	8	288.7	10	256.0	EL
PIAS1	chr15	68,049,178	68,196,466	8	155.1	6	122.0	EL
RSRC1	chr3	158,105,051	158,549,835	8	-2.8	109	106.0	AL
SSH2	chr17	29,620,938	29,935,228	8	-15.8	137	152.0	AL
ST13	chr22	40,819,534	40,862,008	8	67.9	29	7.0	AL
LEF1	chr4	108,042,544	108,173,956	7	-2.8	84	152.0	AL
MAPK14	chr6	36,022,676	36,116,236	7	37.7	91	7.0	AL
UBR1	chr15	42,937,899	43,111,088	7	-45.1	48	1277.5	AL
EP400P1	chr12	132,079,282	132,131,340	6	337.3	5	260.0	EL
HERC4	chr10	67,916,898	68,080,346	6	-10.9	35	106.0	AL

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
MTOR	chr1	11,101,530	11,267,551	6	2.9	29	90.0	AL
NDFIP2	chr13	79,476,123	79,561,077	6	450.7	8	46.0	EL
PA2G4	chr12	56,099,318	56,118,910	5	52.7	38	14.0	AL
FANCA	chr16	89,732,550	89,821,657	15	28.6	21	15.0	L
JPT2	chr16	1,673,276	1,707,072	15	36.8	23	46.0	L
RPTOR	chr17	80,539,824	80,971,373	15	16.3	11	346.0	L
KDM2A	chr11	67,114,268	67,263,079	14	-24.2	7	50.0	L
NPLOC4	chr17	81,551,884	81,642,153	14	-30.6	16	46.0	DL
ANKRD11	chr16	89,262,620	89,495,561	13	28.5	23	7.0	L
EHMT1	chr9	137,613,991	137,841,126	13	39.7	3	50.0	L
EP300	chr22	41,087,609	41,185,077	13	71.5	6	46.0	L
NSD1	chr5	177,128,078	177,305,213	13	-25.8	6	62.0	L
RABEP1	chr17	5,277,262	5,391,339	13	15.6	26	14.0	A
UTRN	chr6	144,286,736	144,858,034	13	-7.0	6	22.0	L
VPS8	chr3	184,807,142	185,057,614	13	118.7	10	46.0	E
ASH1L	chr1	155,330,260	155,567,533	12	-32.3	5	106.0	DL
DPYD	chr1	97,072,743	97,926,059	12	-25.4	22	122.0	L
HSF1	chr8	144,286,568	144,319,726	12	-42.5	10	256.0	DL
MACF1	chr1	39,079,166	39,492,138	12	-29.3	11	519.5	L
MED13	chr17	61,937,604	62,070,282	12	68.4	21	152.0	L
PACS1	chr11	66,065,352	66,249,747	12	-32.0	5	46.0	DL
SMG1	chr16	18,799,852	18,931,404	12	-29.8	7	122.0	L
BOP1	chr8	144,257,045	144,296,438	11	-28.2	10	256.0	L
CCDC57	chr17	82,096,469	82,217,829	11	-39.5	6	642.5	DL
CREBBP	chr16	3,720,054	3,885,120	11	21.5	5	46.0	L
DNMT1	chr19	10,128,343	10,200,135	11	-30.6	13	365.0	DL
FKBP5	chr6	35,568,584	35,733,583	11	-45.0	15	1555.0	DL
GMDS	chr6	1,618,799	2,250,634	11	118.7	8	14.0	L
INPP4B	chr4	142,018,159	142,851,535	11	-17.5	3	122.0	L
PELP1	chr17	4,666,383	4,709,337	11	215.8	3	14.0	E
PPP6R2	chr22	50,338,316	50,450,089	11	-28.0	15	136.0	L
SEC16A	chr9	136,435,095	136,488,759	11	-12.9	29	46.0	A
SF1	chr11	64,759,603	64,783,844	11	22.2	16	14.0	L
TRAPPC10	chr21	44,007,324	44,111,551	11	-0.3	7	14.0	L
USP15	chr12	62,255,339	62,414,721	11	45.8	14	15.0	L
VPS13D	chr1	12,225,038	12,517,046	11	74.9	3	0.0	E
ZNF407	chr18	74,625,962	75,070,672	11	110.6	3	1.0	E
ARHGAP15	chr2	143,124,329	143,773,352	10	45.8	7	5.0	L
ATF7IP	chr12	14,360,631	14,507,935	10	9.9	8	7.0	L
CHD4	chr12	6,565,081	6,612,433	10	53.4	19	7.0	L
CLEC16A	chr16	10,939,487	11,187,189	10	159.1	9	0.0	E

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
CYTH1	chr17	78,669,046	78,787,342	10	-23.6	4	7.0	L
DDX42	chr17	63,769,188	63,824,317	10	42.0	6	14.0	L
GBE1	chr3	81,484,698	81,766,799	10	142.9	9	46.0	E
GLCCI1	chr7	7,963,742	8,094,079	10	8.6	4	84.0	L
KDM6A	chrX	44,868,174	45,117,612	10	86.3	9	106.0	L
KMT5B	chr11	68,149,862	68,218,772	10	-6.4	7	642.5	L
LRBA	chr4	150,259,658	151,020,497	10	32.2	17	106.0	L
LRPPRC	chr2	43,881,223	44,001,005	10	264.4	7	0.0	E
MAPK8IP3	chr16	1,701,182	1,775,317	10	-42.1	5	130.0	DL
MECP2	chrX	154,016,812	154,102,731	10	-9.6	12	99.0	L
MGA	chr15	41,655,411	41,774,943	10	64.5	85	46.0	A
NCOA1	chr2	24,579,476	24,775,701	10	-7.2	3	22.0	L
PBRM1	chr3	52,540,351	52,690,850	10	-16.2	14	7.0	L
PCNX1	chr14	70,902,404	71,120,382	10	-23.1	153	14.0	A
PRKACB	chr1	84,072,974	84,243,498	10	39.7	4	7.0	L
RAB11FIP3	chr16	420,667	527,481	10	-46.0	7	46.0	DL
RBM39	chr20	35,698,608	35,747,336	10	269.3	2	14.0	E
SAFB2	chr19	5,581,998	5,627,927	10	-6.9	13	22.0	L
SETD2	chr3	47,011,407	47,168,977	10	32.8	17	22.0	L
SMARCC1	chr3	47,580,887	47,786,915	10	-11.1	3	7.0	L
SMG6	chr17	2,054,838	2,308,775	10	-16.7	16	7.0	L
SRRM2	chr16	2,747,328	2,776,412	10	87.4	32	1.0	A
ZGPAT	chr20	63,702,441	63,741,142	10	-45.5	5	7.0	DL
ZNF251	chr8	144,715,908	144,760,585	10	-22.6	21	22.0	L
CD55	chr1	207,316,471	207,365,966	9	353.5	4	0.0	E
COX6B1	chr19	35,643,222	35,663,784	9	3.3	5	825.0	L
CSNK1D	chr17	82,237,660	82,278,742	9	-19.0	8	7.0	L
DDX60	chr4	168,211,290	168,323,807	9	62.0	22	122.0	L
DENND1B	chr1	197,499,748	197,780,493	9	-27.1	3	7.0	L
DIP2B	chr12	50,499,984	50,753,667	9	2.6	4	7.0	L
DLG1	chr3	197,037,559	197,304,272	9	5.0	8	7.0	L
FNBP1	chr9	129,882,186	130,048,194	9	-18.2	5	256.0	L
FOXJ3	chr1	42,171,538	42,340,877	9	27.5	12	7.0	L
FRYL	chr4	48,492,362	48,785,299	9	17.6	24	14.0	A
GANAB	chr11	62,619,825	62,651,726	9	9.9	6	7.0	L
HNRNPUL2	chr11	62,707,624	62,732,385	9	-10.0	9	76.0	L
IL4I1	chr19	49,884,655	49,934,539	9	-36.8	5	46.0	L
IQGAP1	chr15	90,383,240	90,507,243	9	-29.3	5	166.0	L
KIAA1468	chr18	62,182,290	62,312,122	9	23.1	17	46.0	L
MED13L	chr12	115,953,575	116,282,186	9	35.0	38	7.0	A
NBEAL1	chr2	203,009,878	203,222,994	9	66.6	11	14.0	L

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
NOSIP	chr19	49,550,467	49,585,572	9	-57.5	29	0.0	DA
NUP214	chr9	131,120,560	131,239,670	9	8.6	8	106.0	L
PARP8	chr5	50,660,898	50,851,522	9	-21.5	4	14.0	L
PLEKHA5	chr12	19,124,691	19,381,399	9	16.6	2	7.0	L
POLA2	chr11	65,256,851	65,303,685	9	-16.7	9	7.0	L
POT1	chr7	124,817,385	124,934,983	9	10.1	26	0.0	A
PUM1	chr1	30,926,505	31,070,717	9	163.8	5	0.0	E
RAB11FIP2	chr10	117,999,915	118,051,884	9	288.7	19	1.0	E
RNF216	chr7	5,615,040	5,786,730	9	-41.7	15	106.0	L
RUNX1	chr21	34,782,800	35,054,298	9	130.8	1	110.0	L
SLC6A16	chr19	49,284,634	49,330,217	9	20.5	2	14.0	L
SLK	chr10	103,962,184	104,034,233	9	240.1	13	1.0	E
SMCHD1	chr18	2,650,886	2,810,017	9	-33.5	4	7.0	L
TARSL2	chr15	101,648,751	101,729,442	9	191.5	22	7.0	L
TCF20	chr22	42,155,012	42,288,927	9	-2.8	10	7.0	L
TRAPPC8	chr18	31,824,172	31,948,128	9	21.5	4	7.0	L
UBAC2	chr13	99,195,424	99,391,499	9	45.8	4	14.0	L
UBR4	chr1	19,069,505	19,215,252	9	136.0	8	7.0	L
VPS52	chr6	33,245,271	33,276,965	9	1.6	10	106.0	L
ZFAND3	chr6	37,814,530	38,159,623	9	215.8	5	46.0	E
ABCD2	chr12	39,546,219	39,625,041	8	142.9	8	14.0	E
ADD1	chr4	2,838,856	2,935,075	8	-6.6	42	7.0	A
AP2B1	chr17	35,582,262	35,731,417	8	62.0	13	90.0	L
ARIH1	chr15	72,469,325	72,591,555	8	51.2	5	7.0	L
ASXL2	chr2	25,728,752	25,883,516	8	-39.8	3	7.0	L
ATG5	chr6	106,179,476	106,330,820	8	191.5	2	0.0	E
ATP2A2	chr12	110,276,226	110,356,092	8	70.1	67	22.0	A
CAMK2D	chr4	113,446,031	113,766,927	8	2.0	9	106.0	L
CAMK4	chr5	111,218,652	111,499,884	8	3.3	9	7.0	L
CDKAL1	chr6	20,529,456	21,237,403	8	-13.1	5	106.0	L
DAZAP1	chr19	1,402,568	1,440,687	8	-10.9	2	4.0	L
DNAJC13	chr3	132,412,659	132,544,032	8	94.4	71	7.0	A
DYNC1H1	chr14	101,959,527	102,055,798	8	94.4	44	7.0	A
ERC1	chr12	986,207	1,500,933	8	8.6	4	7.0	L
FAM13A	chr4	88,720,953	89,062,195	8	80.5	1	7.0	L
FOCAD	chr9	20,653,308	21,000,955	8	155.1	17	7.0	E
GAK	chr4	844,274	937,390	8	450.7	4	14.0	E
GPHN	chr14	66,502,406	67,186,808	8	321.1	2	0.0	E
HNRNPUL1	chr19	41,257,475	41,312,783	8	84.6	4	14.0	L
HSF2	chr6	122,394,550	122,438,119	8	240.1	15	14.0	E
KDM4A	chr1	43,645,125	43,710,518	8	211.0	7	0.0	E

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
KMT2C	chr7	152,129,924	152,441,005	8	-23.6	4	7.0	L
MIR5096	chr17	4,136,088	4,245,637	8	-47.2	9	14.0	DL
MOB3A	chr19	2,066,035	2,101,270	8	55.5	2	22.0	L
MROH1	chr8	144,143,015	144,266,940	8	-51.4	5	46.0	DL
MUM1	chr19	1,349,976	1,383,431	8	-26.0	4	90.0	L
NCOA3	chr20	47,496,856	47,661,877	8	2.0	6	7.0	L
NEAT1	chr11	65,417,797	65,450,538	8	-2.8	4	22.0	L
NELL2	chr12	44,503,274	44,918,928	8	20.0	2	4.0	L
PHF20L1	chr8	132,770,357	132,853,807	8	36.0	4	22.0	L
PLEC	chr8	143,910,146	143,981,745	8	-50.1	11	22.0	DL
POGZ	chr1	151,397,723	151,464,465	8	52.7	26	0.0	A
PTPRA	chr20	2,859,194	3,043,669	8	10.6	4	1555.0	L
RNF10	chr12	120,529,328	120,582,594	8	211.0	7	1.0	E
RTTN	chr18	69,998,805	70,210,726	8	-8.9	6	136.0	L
SACM1L	chr3	45,684,240	45,750,425	8	191.5	11	0.0	E
SLX4IP	chr20	10,430,302	10,633,034	8	66.6	3	7.0	L
STAG1	chr3	136,332,156	136,757,403	8	152.7	6	136.0	L
STXBP5	chr6	147,199,357	147,395,476	8	-8.9	8	350.0	L
TANC2	chr17	63,004,536	63,432,706	8	191.5	11	7.0	E
TONSL	chr8	144,423,779	144,449,429	8	-32.4	7	7.0	L
UHRF1BP1	chr6	34,787,016	34,882,514	8	288.7	4	0.0	E
VMP1	chr17	59,702,464	59,847,255	8	-2.8	28	46.0	A
YWHAE	chr17	1,339,539	1,405,262	8	191.5	2	0.0	E
ACOX1	chr17	75,936,510	75,984,434	7	-30.6	2	256.0	L
ASCC3	chr6	100,503,194	100,886,372	7	182.7	4	7.0	E
ATF7	chr12	53,502,855	53,631,415	7	12.1	32	1.0	A
ATP8A1	chr4	42,403,374	42,662,105	7	19.6	4	14.0	L
BACH2	chr6	89,921,527	90,301,908	7	483.1	2	0.0	E
C6orf106	chr6	34,582,279	34,701,850	7	-2.8	3	46.0	L
CASK	chrX	41,509,935	41,928,034	7	29.6	5	346.0	L
CCDC47	chr17	63,740,249	63,778,728	7	-42.8	6	22.0	L
CDC73	chr1	193,116,957	193,259,812	7	123.5	19	7.0	E
GATAD2B	chr1	153,799,906	153,927,975	7	6.9	2	7.0	L
HTT	chr4	3,069,680	3,248,960	7	-2.8	25	1.0	A
KLRG1	chr12	8,945,043	9,015,744	7	-40.6	3	7.0	L
LSM2	chr6	31,792,391	31,811,984	7	-53.2	2	14.0	DL
MACROD2	chr20	13,990,499	16,058,196	7	240.1	1	0.0	E
MAP4K3	chr2	39,244,265	39,442,312	7	-2.8	15	46.0	L
NAA38	chr17	7,851,680	7,890,388	7	-34.5	15	14.0	L
NBAS	chr2	15,161,907	15,566,348	7	-19.0	4	106.0	L
NEMP1	chr12	57,050,642	57,083,791	7	113.8	10	46.0	L

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
NUP62	chr19	49,901,825	49,934,731	7	-42.3	5	46.0	L
PIP5K1A	chr1	151,193,543	151,254,531	7	256.3	3	277.5	L
PPP1R16A	chr8	144,472,981	144,507,121	7	-10.3	4	7.0	L
PRKCA	chr17	66,297,807	66,815,744	7	51.2	7	7.0	L
RAD51B	chr14	67,814,778	68,688,106	7	-2.8	2	22.0	L
RASA1	chr5	87,263,252	87,396,926	7	205.4	3	0.0	E
RFX2	chr19	5,988,163	6,115,653	7	13.4	9	106.0	L
SEPT7	chr7	35,795,985	35,912,105	7	175.3	10	7.0	E
SNAPC4	chr9	136,370,568	136,403,437	7	34.6	7	166.0	L
SPEN	chr1	15,842,863	15,945,455	7	4.7	2	7.0	L
SPG7	chr16	89,503,387	89,562,768	7	-28.7	10	7.0	L
SYNE1	chr6	152,116,683	152,642,399	7	-2.8	22	46.0	L
SYNE2	chr14	63,847,964	64,231,451	7	12.1	8	22.0	L
TAF2	chr8	119,725,773	119,837,834	7	288.7	8	1.0	E
THEMIS	chr6	127,703,193	127,923,631	7	-35.2	5	106.0	L
TTC21B	chr2	165,868,361	165,958,838	7	-9.8	5	46.0	L
UBE2J2	chr1	1,248,911	1,278,854	7	-58.8	2	166.0	DL
UBR5	chr8	102,247,273	102,417,689	7	29.6	19	7.0	L
VRK3	chr19	49,971,466	50,030,548	7	-43.0	2	7.0	L
WDR82	chr3	52,249,421	52,283,643	7	-11.7	7	277.5	L
YTHDF3	chr8	63,163,552	63,217,788	7	8.0	13	7.0	L
ANXA1	chr9	73,146,730	73,175,394	6	62.0	5	14.0	L
ATP9B	chr18	79,064,274	79,383,282	6	18.8	2	7.0	L
BAG6	chr6	31,634,027	31,657,700	6	-19.0	36	46.0	A
BCAS3	chr17	60,672,774	61,397,838	6	-57.7	2	7.0	DL
BRWD3	chrX	80,664,487	80,814,734	6	80.5	6	7.0	L
CHD1	chr5	98,850,203	98,931,534	6	149.9	8	1.0	E
CHMP2B	chr3	87,222,262	87,260,548	6	24.9	24	7.0	A
CLASP2	chr3	33,491,245	33,723,213	6	57.9	5	7.0	L
DERL2	chr17	5,466,250	5,491,230	6	55.5	37	0.0	A
DNAJC1	chr10	21,751,547	22,008,721	6	-22.3	38	7.0	A
DOT1L	chr19	2,159,148	2,237,578	6	-18.2	6	14.0	L
FAM117B	chr2	202,630,177	202,774,757	6	385.9	6	14.0	E
FUNDC2	chrX	155,021,788	155,061,916	6	321.1	9	46.0	E
GNA12	chr7	2,723,105	2,849,325	6	434.5	2	0.0	E
IKZF2	chr2	212,994,685	213,156,609	6	434.5	1	0.0	E
MIR4745	chr19	799,939	810,001	6	142.9	47	346.0	L
MIR5096	chr1	15,866,148	15,910,467	6	18.8	2	7.0	DL
MIR5096	chr22	37,663,025	38,029,093	6	-25.7	6	7.0	DL
MMP23A	chr1	1,627,779	1,706,808	6	4.7	17	14.0	L
NFKBIL1	chr6	31,541,850	31,563,829	6	-46.7	4	7.0	DL

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
OPRM1	chr6	154,005,495	154,251,867	6	-2.8	3	14.0	L
PAPOLA	chr14	96,497,375	96,572,116	6	418.3	3	0.0	E
PCNT	chr21	46,319,121	46,450,769	6	113.8	5	4.0	L
PHACTR4	chr1	28,364,581	28,505,369	6	-33.2	3	22.0	L
RBM27	chr5	146,198,599	146,294,221	6	Inf	3	0.0	E
RIPOR2	chr6	24,799,280	25,047,288	6	-2.8	10	7.0	L
RSBN1L	chr7	77,691,425	77,784,803	6	94.4	31	1.0	A
SEC23A	chr14	39,026,918	39,108,528	6	14.8	7	7.0	L
SPPL3	chr12	120,757,509	120,909,352	6	4.1	9	14.0	L
TCF25	chr16	89,868,585	89,916,384	6	-45.1	5	106.0	L
TRIM33	chr1	114,387,776	114,516,160	6	167.2	17	106.0	L
UBE2F-SCLY	chr2	237,961,944	238,104,413	6	-10.9	6	46.0	L
UBE2L3	chr22	21,544,446	21,629,034	6	94.4	3	7.0	L
UNKL	chr16	1,358,204	1,419,720	6	18.8	7	15.0	L
ZNF473	chr19	50,020,892	50,053,774	6	-30.6	3	22.0	L
ZNRD1ASP	chr6	29,996,010	30,066,189	6	51.2	4	122.0	L
CSNK1G1	chr15	64,160,516	64,361,259	5	-11.7	8	7.0	L
ECD	chr10	73,129,523	73,173,095	5	191.5	24	1.0	A
ELMO1	chr7	36,847,905	37,454,326	5	-29.3	34	46.0	A
EVL	chr14	99,966,474	100,149,236	5	-25.2	4	7.0	L
FANCL	chr2	58,154,242	58,246,380	5	434.5	6	0.0	E
HSF5	chr17	58,415,166	58,493,401	5	-2.8	34	1.0	A
IQCB1	chr3	121,764,760	121,840,079	5	62.0	79	15.0	A
MARF1	chr16	15,589,368	15,648,166	5	106.5	17	106.0	L
KIFC1	chr6	33,386,535	33,414,922	5	-2.8	42	7.0	A
KMT2D	chr12	49,013,974	49,060,324	5	-28.7	33	0.0	A
MIR5096	chr1	235,507,822	235,723,113	5	70.1	1	0.0	DL
N4BP1	chr16	48,533,725	48,615,209	5	159.1	32	106.0	A
PAG1	chr8	80,962,810	81,117,068	5	-33.2	2	32.0	L
PDE7A	chr8	65,709,333	65,846,734	5	18.8	4	7.0	L
PHF20	chr20	35,767,000	35,955,366	5	-28.7	38	7.0	A
POM121	chr7	72,874,334	72,956,440	5	240.1	29	1.0	A
PPFIA1	chr11	70,265,699	70,389,501	5	78.2	20	106.0	L
SEC31A	chr4	82,813,508	82,905,571	5	-2.8	66	7.0	A
SEPT9	chr17	77,276,409	77,505,596	5	-44.5	27	7.0	A
SFI1	chr22	31,491,138	31,623,551	5	-55.8	24	7.0	DA
SMAP2	chr1	40,368,705	40,428,326	5	45.8	61	7.0	A
STAG3	chr7	100,172,723	100,219,387	5	78.2	35	0.0	A
STX8	chr17	9,245,470	9,580,958	5	38.8	4	7.0	L
TNKS	chr8	9,550,934	9,787,346	5	191.5	8	7.0	E
WWOX	chr16	78,094,412	79,217,667	5	-17.8	5	106.0	L

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
AGL	chr1	99,845,083	99,929,023	4	871.8	4	0.0	E
ANKRD46	chr8	100,504,751	100,564,786	4	29.6	24	7.0	A
AQR	chr15	34,851,350	34,974,794	4	-15.0	84	14.0	A
CAAP1	chr9	26,835,684	26,897,828	4	-39.3	27	46.0	A
CNOT6	chr5	180,489,398	180,583,405	4	-19.0	2	7.0	L
EHMT1	chr9	137,758,021	137,769,772	4	62.0	1	0.0	L
EXOSC10	chr1	11,061,612	11,104,910	4	13.4	70	7.0	A
GOLPH3L	chr1	150,641,224	150,702,196	4	-51.4	24	7.0	A
KDM5D	chrY	19,700,414	19,749,939	4	62.0	2	14.0	A
MAD1L1	chr7	1,810,791	2,237,948	4	-20.5	35	46.0	A
MAP2K2	chr19	4,085,321	4,129,129	4	-2.8	2	106.0	L
MEMO1	chr2	31,862,809	32,016,052	4	-64.7	47	7.0	A
MSH5-SAPCD1	chr6	31,734,947	31,769,847	4	-67.6	43	7.0	A
MTMR3	chr22	29,878,168	30,035,868	4	-45.3	261	106.0	A
OGDH	chr7	44,601,521	44,714,070	4	-48.9	17	332.0	L
PATL1	chr11	59,631,715	59,674,038	4	29.6	578	332.0	A
PDCD11	chr10	103,391,654	103,451,262	4	94.4	27	152.0	A
PDE3B	chr11	14,638,722	14,877,058	4	29.6	35	7.0	A
RAB18	chr10	27,499,173	27,547,237	4	-22.3	24	7.0	A
RMND5A	chr2	86,715,290	86,783,041	4	Inf	26	0.0	A
RPA3	chr7	7,631,562	7,723,607	4	21.5	87	106.0	A
TRIO	chr5	14,138,701	14,515,204	4	580.2	61	106.0	A
XPO1	chr2	61,472,933	61,543,283	4	21.5	25	106.0	A
AKAP9	chr7	91,935,874	92,115,673	3	-22.3	25	0.0	A
CHD1L	chr1	147,168,193	147,300,766	3	Inf	25	1.0	A
DCUN1D4	chr4	51,837,999	51,921,837	3	-19.0	32	106.0	A
EIF2AK4	chr15	39,929,123	40,040,596	3	385.9	43	7.0	A
NGDN	chr14	23,464,688	23,483,193	3	45.8	44	0.0	A
PRKD2	chr19	46,669,315	46,722,127	3	62.0	24	22.0	A
RABGAP1	chr9	122,936,008	123,109,868	3	-32.0	29	7.0	A
SNAP29	chr22	20,854,003	20,896,213	3	288.7	27	7.0	A
ZNF573	chr19	37,733,301	37,784,590	3	-22.3	86	735.0	A
C20orf196	chr20	5,745,386	5,869,407	2	-35.2	1	0.0	A
CRTAP	chr3	33,108,957	33,152,773	2	94.4	35	7.0	A
GPN1	chr2	27,623,647	27,655,846	2	-2.8	62	1.0	A
JMJD6	chr17	76,707,831	76,731,799	2	94.4	53	1.0	A
KCTD3	chr1	215,562,378	215,626,821	2	-51.4	1	0.0	A
LINC01473	chr2	186,028,533	186,091,317	2	Inf	82	7.0	A
LOC101927151	chr19	27,788,466	27,811,780	2	94.4	31	46.0	A
PHF12	chr17	28,900,252	28,956,490	2	-22.3	40	106.0	A
POLG2	chr17	64,472,784	64,502,066	2	21.5	43	7.0	A

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
RASEF	chr9	82,974,584	83,068,128	2	Inf	43	46.0	A
SNHG12	chr1	28,573,537	28,586,854	2	191.5	96	332.0	A
TAC3	chr12	57,004,996	57,021,560	2	-67.6	42	7.0	A
TGFBR2	chr3	30,601,501	30,699,141	2	94.4	31	7.0	A
ACTL6A	chr3	179,557,879	179,593,405	1	-80.6	40	0.0	A
C19orf48	chr19	50,792,692	50,809,853	1	-2.8	28	46.0	A
CD109	chr6	73,691,084	73,833,317	1	-87.9	32	0.0	A
IFNGR2	chr21	33,397,895	33,442,521	1	Inf	53	15.0	A
KARS	chr16	75,622,723	75,652,687	1	-2.8	24	0.0	A
MICAL2	chr11	12,105,575	12,268,790	1	-51.4	39	15.0	A
RBAK-RBAKDN	chr7	5,040,820	5,078,223	1	-2.8	28	7.0	A
RTCA-AS1	chr1	100,259,741	100,271,174	1	Inf	27	0.0	A
UXT-AS1	chrX	47,653,832	47,665,111	1	Inf	43	7.0	A
ZNF92	chr7	65,368,798	65,406,135	1	Inf	42	7.0	A
SIRT3	chr11	210,029	241,362	10	-44.8	4	1.0	D
IP6K1	chr3	49,719,294	49,791,540	9	-57.9	2	1.0	D
NPEPPS	chr17	47,517,956	47,628,276	9	-55.1	3	14.0	D
ANKFY1	chr17	4,158,817	4,268,995	8	-46.0	9	14.0	D
EIF2B3	chr1	44,845,521	44,991,722	8	-67.6	6	1.0	D
TRAF2	chr9	136,881,512	136,931,615	8	-59.7	4	46.0	D
TSC2	chr16	2,042,894	2,093,720	8	-51.4	1	22.0	D
WDR90	chr16	644,362	672,829	8	-53.5	3	0.0	D
HCG20	chr6	30,761,824	30,797,250	7	-48.9	4	22.0	D
NARFL	chr16	724,754	746,038	7	-61.1	3	22.0	D
PPP3CC	chr8	22,435,969	22,546,144	7	-53.2	3	0.0	D
IFT140	chr16	1,505,426	1,617,108	6	-51.4	3	7.0	D
QRICH1	chr3	49,024,706	49,099,373	6	-59.0	7	1.0	D
EXOC2	chr6	480,137	698,141	5	-71.4	4	0.0	D
HAGH	chr16	1,804,102	1,832,194	5	-58.4	6	14.0	D
RNPS1	chr16	2,248,115	2,273,412	5	-77.5	2	0.0	D
USP34	chr2	61,182,454	61,475,714	5	-61.1	2	0.0	D
WNK1	chr12	747,922	916,452	5	-51.4	4	1.0	D
ZNF598	chr16	1,992,651	2,014,821	4	-55.8	3	0.0	D

Supplementary Report 3:
Genes of interest marked by vector integration
ALL Patients only (CR/PRtd & PR/NR Response Groups)

Contents

Summary

Lentiviral vectors integrate into genomes of targeted host cells (Tcells). 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 mined the data collected from integration site sequencing for 11 CART treated subjects. We constructed 4 gene lists based on: 1 & 2) increased / decreased integration site occurrence in patient samples relative to the initial transduction product, 3) peak clonal abundance, and 4) longitudinal clonal persistence. More about each of these criteria is below:

- **Integration Frequency** is the rate at which integration sites are observed within a gene. This is compared between patient samples and the initial transduction product to score enrichment or depletion during growth in patients. The top of genes with higher patient sample integration frequency over transduction samples were chosen for study (p-value $<= 0.05$ after exclusion of genes with clones from less than 2 patients and less than 10 observed clones).
- **Clonal Abundance** can be determined during analysis by quantifying the number of sites of linker ligation associated with each unique integration site. This method is further described in **Berry *et al.* 2012**. This allows clonal expansion to be quantified. The top 1% of the genes were selected for study based on their maximal peak clonal abundance.
- **Longitudinal Observation** of clones is the quantification of observed timespans and last observed timepoints. The maximum value for clones within a gene were considered for characterization of the gene in this analysis. Genes were only considered if there were 10 or more integration sites isolated from at least two different patient samples. Genes were also not considered if they only consited of clones which were observed once or the last observed timepoint was less than 90 days from initial infusion.

A point to keep in mind through all this analysis is that integration sites are sampled from a larger population. It would be rare for all integration sites in a sample to be represented in the sequence data.

Table 1: Summary of each filtering criteria.

Criteria	Gene Count	Onco Related (%)	Tumor Suppressors (%)	Lymphoma Related (%)	COSMIC Related (%)	TCGA Related (%)	Clonal Hema. Related (%)
Enrichment	2	*/* 100.0	/ 50.0	/ 0	*/* 100.0	/ 50.0	*/* 50.00
Depletion	0	NA/NA	NA/NA	NA/NA	NA/NA	NA/NA	NA/NA
Abundance	40	*/* 22.5	/ 12.5	/ 0	*/* 15.0	*/* 12.5	/ 0.00
Longitudinal	0	NA/NA	NA/NA	NA/NA	NA/NA	NA/NA	NA/NA
Composite	41	*/* 24.4	*/ 14.6	/ 0	*/* 17.1	*/* 14.6	/ 2.44

Table 1 summarizes the size and contents of each criteria gene list identified by the various methods. Significance of overlap between lists are displayed by asterisks before the percent of genes identifed 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 Fishers Exact test and multiple comparison corrections were made using a Benjamini-Hochberg (FDR) method for each criteria based list.

Percent of all analyzed transcription units associated with each list as as follows:

- Onco Related: 10.52%
- Tumor Suppressors: 5.13%
- Lymphoma Related: 0.21%
- COSMIC Related: 4.32%
- TCGA Related: 3.19%
- Clonal Hematopoiesis Related: 0.22%

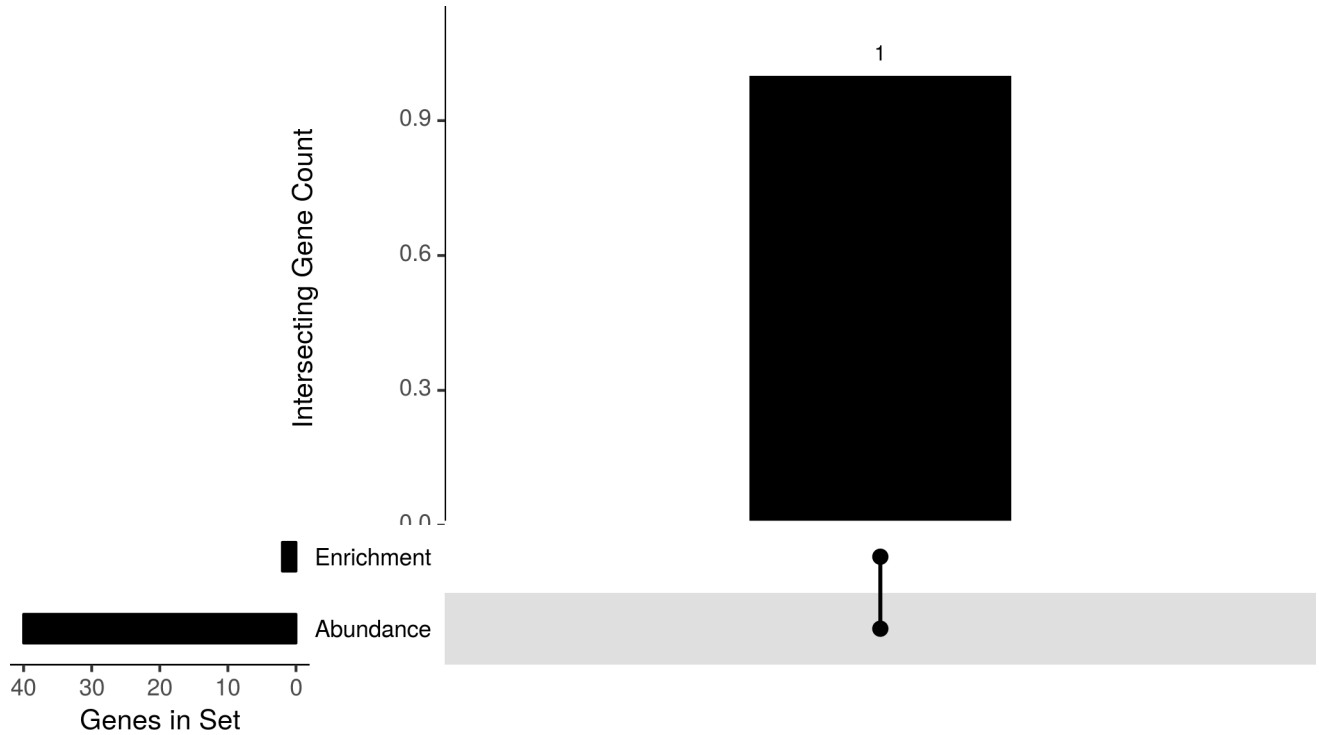

Figure 1: Intersecting gene lists identified through the various selection criteria.

Table 2: The most consistently observed genes from filtering by various criteria. The 'Criteria.' column is a count of how many times the gene was identified by these methods, while the 'Patients' column notes how many specimens collected from patients have had integration sites within the noted gene.

Gene	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
RNF213	3	136.9	5	28	EA

Table 3: GO Biological Process. Top 1 per group. Total genes considered: 35

Group	GO ID	GO Term	Term Size	Gene Count	Adjusted P-value
1	GO:0006325	chromatin organization	478	8	0.0169989
2	GO:0048534	hematopoietic or lymphoid organ development	471	7	0.0169989

Integration Frequency (Enrichment)

Table 4: Table of top 2 genes with the most frequent clonal enrichment.

Gene	Num. Patients	TDN Sites	Patient Sites	Onco-Related	Frequency Increase (%)
CREBBP	4	36	11	TRUE	218.6
RNF213	3	44	10	TRUE	136.9

Genes with the Most Abundant Clones

Table 5: Table of top 40 Genes containing the highest abundant clones.

Gene	Num. Patients	Peak Abundance	Peak Rel. Abund.	Clonal Gini Index	Onco-Related
ATAT1	1	13	0.520	0.000	FALSE
KIF1B	3	12	0.038	0.550	FALSE
PIAS2	1	10	0.054	0.000	FALSE
DSE	2	9	0.029	0.400	FALSE
PGD	1	9	0.030	0.400	FALSE
DUSP16	1	8	0.121	0.000	FALSE
AHRR	2	7	0.042	0.444	FALSE
BAZ2A	1	7	0.023	0.000	FALSE
HNRNPUL2-BSCL2	2	7	0.023	0.400	FALSE
PPP6R3	3	7	0.022	0.408	FALSE
ERP44	1	6	0.019	0.000	FALSE
FANCA	5	6	0.009	0.301	TRUE
HOPX	1	6	0.036	0.000	FALSE
LYPLAL1	1	6	0.037	0.000	FALSE
MRE11	1	6	0.037	0.000	TRUE
PPP3CA	3	6	0.013	0.400	FALSE
UBR2	4	6	0.019	0.357	FALSE
ZNF354B	1	6	0.019	0.000	FALSE
ABHD17A	1	5	0.031	0.000	FALSE
AP3B1	2	5	0.006	0.381	FALSE
APOF	1	5	0.031	0.000	FALSE
ARNT	1	5	0.005	0.333	TRUE
C3orf58	2	5	0.016	0.333	FALSE
DNMT1	3	5	0.007	0.311	TRUE
EML5	1	5	0.030	0.000	FALSE
IPO7	2	5	0.139	0.333	FALSE
KMT2C	3	5	0.016	0.333	TRUE
MAN1A2	2	5	0.027	0.333	FALSE
NLRC3	1	5	0.030	0.000	FALSE
PHF20	1	5	0.006	0.333	TRUE
PIP4K2A	1	5	0.030	0.000	FALSE
POT1	2	5	0.016	0.333	FALSE
PSMD13	2	5	0.005	0.375	FALSE
RCAN3	2	5	0.016	0.333	FALSE
RNF157	3	5	0.076	0.257	TRUE
RNF213	3	5	0.016	0.257	TRUE
SET	1	5	0.016	0.000	TRUE
SLC44A2	2	5	0.017	0.346	FALSE
SNX13	3	5	0.007	0.333	FALSE
UTY	2	5	0.016	0.361	FALSE

Reference Data

The NCBI RefGenes data set was used to identify gene regions (hg38) while genes identified as onco-related were from the Bushman Lab curated list of **onco-related genes**.

Gene Ontologies were extracted from the `GO.db` R-package (v3.4.1). KEGG pathways were acquired via interfacing with the KEGG web-server API through the `KEGGREST` R-package (v1.16.1). Gene lists, including RefSeq genes used for annotation of integration sites, were standardized to HGNC gene symbols (date: 2018-02-07). Groups identified in GO and KEGG analyses were determined from Jaccard distances between identified terms, followed by modularity-optimizing clustering from a weighted-undirected graph using a Louvain algorithm (**Blondel *et al.* 2008**). Terms within groups of GO or KEGG terms have greater overlap of gene lists between themselves that between terms found in other groups. This method was implemented to help reduce the functional redundancy commonly observed in GO and overlapping pathways observed with KEGG.

Comprehensive Genes of Interest Table

Table 6: Table of all genes identified within analysis.

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
RNF213	chr17	80,255,860	80,403,781	3	136.9	5	0	EA
FANCA	chr16	89,732,550	89,821,657	5	18.5	6	0	A
CREBBP	chr16	3,720,054	3,885,120	4	218.6	2	0	E
UBR2	chr6	42,559,021	42,698,505	4	32.7	6	0	A
DNMT1	chr19	10,128,343	10,200,135	3	0.9	5	0	A
KIF1B	chr1	10,205,705	10,386,603	3	189.6	12	0	A
KMT2C	chr7	152,129,924	152,441,005	3	247.5	5	0	A
PPP3CA	chr4	101,018,429	101,352,471	3	178.0	6	0	A
PPP6R3	chr11	68,455,717	68,620,333	3	69.7	7	0	A
RNF157	chr17	76,137,452	76,245,311	3	-20.4	5	0	A
SNX13	chr7	17,785,760	17,945,508	3	468.7	5	0	A
AHRR	chr5	299,175	443,290	2	-13.1	7	0	A
AP3B1	chr5	77,997,325	78,299,755	2	247.5	5	0	A
ATAT1	chr6	30,621,841	30,651,823	2	197.9	13	0	A
DSE	chr6	116,249,151	116,443,291	2	1985.1	9	0	A
IPO7	chr11	9,379,621	9,453,127	2	131.7	5	0	A
MAN1A2	chr1	117,362,462	117,530,698	2	247.5	5	0	A
POT1	chr7	124,817,385	124,934,983	2	131.7	5	0	A
PSMD13	chr11	231,807	257,984	2	-19.8	5	0	A
RCAN3	chr1	24,497,350	24,542,020	2	4.3	5	0	A
SLC44A2	chr19	10,597,444	10,649,559	2	78.7	5	0	A
UTY	chrY	13,243,378	13,485,670	2	681.9	2	0	A
ARNT	chr1	150,804,704	150,881,768	1	-9.3	5	0	A
BAZ2A	chr12	56,590,595	56,641,379	1	-25.5	7	0	A
DUSP16	chr12	12,468,281	12,567,514	1	-19.8	8	0	A
EML5	chr14	88,609,829	88,797,752	1	421.3	5	0	A
ERP44	chr9	99,974,180	100,104,052	1	108.5	6	0	A
HOPX	chr4	56,642,987	56,686,706	1	Inf	6	0	A
LYPLAL1	chr1	219,168,830	219,217,865	1	73.8	6	0	A
MRE11	chr11	94,412,300	94,498,908	1	73.8	6	0	A
NLRC3	chr16	3,534,035	3,582,404	1	48.9	5	0	A
PGD	chr1	10,393,991	10,425,511	1	Inf	9	0	A
PHF20	chr20	35,767,000	35,955,366	1	22.7	5	0	A
PIAS2	chr18	46,798,224	46,925,167	1	48.9	10	0	A
PIP4K2A	chr10	22,529,836	22,719,574	1	108.5	5	0	A
SET	chr9	128,678,654	128,701,396	1	15.8	5	11	A
ZNF354B	chr5	178,854,952	178,889,423	1	247.5	6	0	A

Supplementary Report 4:
Genes of interest marked by vector integration
Response Group CR/PRtd only (CLL & ALL Patients)

Contents

Summary

Lentiviral vectors integrate into genomes of targeted host cells (Tcells). 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 mined the data collected from integration site sequencing for 20 CART treated subjects. We constructed 4 gene lists based on: 1 & 2) increased / decreased integration site occurrence in patient samples relative to the initial transduction product, 3) peak clonal abundance, and 4) longitudinal clonal persistence. More about each of these criteria is below:

- **Integration Frequency** is the rate at which integration sites are observed within a gene. This is compared between patient samples and the initial transduction product to score enrichment or depletion during growth in patients. The top of genes with higher patient sample integration frequency over transduction samples were chosen for study (p-value <= 0.05 after exclusion of genes with clones from less than 2 patients and less than 10 observed clones).
- **Clonal Abundance** can be determined during analysis by quantifying the number of sites of linker ligation associated with each unique integration site. This method is further described in **Berry *et al.* 2012**. This allows clonal expansion to be quantified. The top 1% of the genes were selected for study based on their maximal peak clonal abundance.
- **Longitudinal Observation** of clones is the quantification of observed timespans and last observed timepoints. The maximum value for clones within a gene were considered for characterization of the gene in this analysis. Genes were only considered if there were 10 or more integration sites isolated from at least two different patient samples. Genes were also not considered if they only consited of clones which were observed once or the last observed timepoint was less than 90 days from initial infusion.

A point to keep in mind through all this analysis is that integration sites are sampled from a larger population. It would be rare for all integration sites in a sample to be represented in the sequence data.

Table 1: Summary of each filtering criteria.

Criteria	Gene Count	Onco Related (%)	Tumor Suppressors (%)	Lymphoma Related (%)	COSMIC Related (%)	TCGA Related (%)	Clonal Hema. Related (%)
Enrichment	278	*/* 21.2	*/* 11.15	/ 0.360	*/* 12.23	*/* 9.35	*/* 1.079
Depletion	4	/ 0.0	/ 0.00	/ 0.000	/ 0.00	/ 0.00	/ 0.000
Abundance	119	*/* 20.2	/ 5.88	/ 0.000	*/* 9.24	/ 5.04	/ 0.840
Longitudinal	208	*/* 25.5	*/* 12.02	/ 0.481	*/* 12.50	*/* 12.02	*/* 1.442
Composite	501	*/* 20.2	*/* 9.38	/ 0.200	*/* 10.78	*/* 8.18	/ 0.599

Table 1 summarizes the size and contents of each criteria gene list identified by the various methods. Significance of overlap between lists are displayed by asterisks before the percent of genes identifed 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 Fishers Exact test and multiple comparison corrections were made using a Benjamini-Hochberg (FDR) method for each criteria based list.

Percent of all analyzed transcription units associated with each list as as follows:

- Onco Related: 9.41%
- Tumor Suppressors: 4.91%
- Lymphoma Related: 0.16%
- COSMIC Related: 3.78%
- TCGA Related: 2.88%
- Clonal Hematopoiesis Related: 0.18%

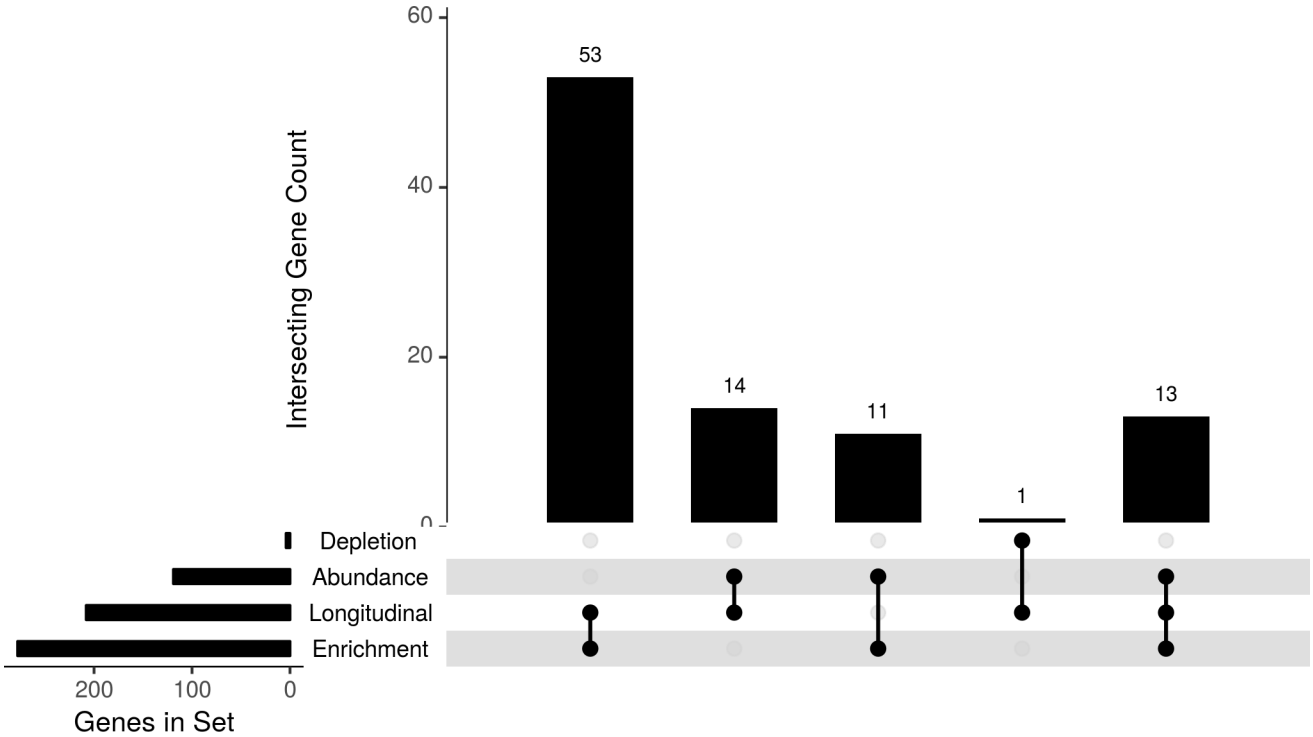

Figure 1: Intersecting gene lists identified through the various selection criteria.

Table 2: The most consistently observed genes from filtering by various criteria. The 'Criteria.' column is a count of how many times the gene was identified by these methods, while the 'Patients' column notes how many specimens collected from patients have had integration sites within the noted gene.

Gene	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
VAV1	13	110.6	37	180	EAL
RNF157	12	77.8	28	180	EAL
AKAP13	11	340.6	27	360	EAL
CARD8	11	74.4	79	180	EAL
ZZEF1	11	64.4	56	360	EAL
CRAMP1	10	53.1	30	548	EAL
PAFAH1B1	10	110.7	30	360	EAL
SRCAP	10	137.3	373	360	EAL
UBAP2L	10	235.7	30	180	EAL
ST13	8	153.7	29	180	EAL
CLK4	7	259.0	53	120	EAL
PTBP1	7	316.1	47	360	EAL
TET2	5	450.8	814	1584	EAL

Table 3: GO Biological Process. Top 7 per group. Total genes considered: 446

Group	GO ID	GO Term	Term Size	Gene Count	Adjusted P-value
1	GO:0016569	covalent chromatin modification	436	52	0.0000000
	GO:0016570	histone modification	353	45	0.0000000
	GO:0018205	peptidyl-lysine modification	295	41	0.0000000
	GO:0033044	regulation of chromosome organization	243	22	0.0095290
	GO:0043414	macromolecule methylation	219	21	0.0063044
	GO:0006479	protein methylation	142	19	0.0001415
	GO:0008213	protein alkylation	142	19	0.0001415
2	GO:0071407	cellular response to organic cyclic compound	390	32	0.0049931
	GO:0032101	regulation of response to external stimulus	426	30	0.0393697
	GO:0060322	head development	430	30	0.0432584
	GO:0071396	cellular response to lipid	338	29	0.0045652
	GO:0007420	brain development	407	29	0.0384799
	GO:0043009	chordate embryonic development	323	25	0.0326732
	GO:0009792	embryo development ending in birth or egg hatching	326	25	0.0328006
3	GO:0043632	modification-dependent macromolecule catabolic process	477	39	0.0014114
	GO:0006511	ubiquitin-dependent protein catabolic process	464	38	0.0016770
	GO:0019941	modification-dependent protein catabolic process	469	38	0.0020971
	GO:0010498	proteasomal protein catabolic process	350	28	0.0136100
	GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic process	325	26	0.0188227
	GO:0034976	response to endoplasmic reticulum stress	203	17	0.0407180
	GO:0032434	regulation of proteasomal ubiquitin-dependent protein catabolic process	97	10	0.0431798
4	GO:0051223	regulation of protein transport	495	41	0.0007062
	GO:0006397	mRNA processing	394	37	0.0001124
	GO:1903827	regulation of cellular protein localization	427	37	0.0006504
	GO:0010608	posttranscriptional regulation of gene expression	412	34	0.0032284
	GO:1902582	single-organism intracellular transport	484	33	0.0419393
	GO:0051169	nuclear transport	347	32	0.0006481
	GO:0032386	regulation of intracellular transport	351	32	0.0007989
5	GO:0010256	endomembrane system organization	460	41	0.0001305
	GO:0051640	organelle localization	378	32	0.0031081
	GO:0044770	cell cycle phase transition	449	31	0.0442715
	GO:0051656	establishment of organelle localization	322	30	0.0009015
	GO:0000226	microtubule cytoskeleton organization	377	28	0.0328006
	GO:0097435	supramolecular fiber organization	388	28	0.0373534
	GO:0051493	regulation of cytoskeleton organization	323	26	0.0174580
6	GO:0030155	regulation of cell adhesion	424	34	0.0050392
	GO:0043547	positive regulation of GTPase activity	490	34	0.0328006
	GO:0007265	Ras protein signal transduction	249	20	0.0366259
	GO:0001667	ameboidal-type cell migration	199	19	0.0102806
	GO:0018105	peptidyl-serine phosphorylation	197	18	0.0202218
	GO:0018209	peptidyl-serine modification	207	18	0.0326384
	GO:0034329	cell junction assembly	139	13	0.0393697

Table 4: KEGG Pathway analysis. Top 10 per group. Total genes considered: 204

Group	KEGG ID	Description	Term Size	Gene Count	Adjusted P-value
1	path:hsa04070	Phosphatidylinositol signaling system	85	9	0.0926317
	path:hsa00562	Inositol phosphate metabolism	63	7	0.0964835
2	path:hsa03018	RNA degradation	69	7	0.1097398
3	path:hsa05200	Pathways in cancer	338	16	0.4929677
	path:hsa04144	Endocytosis	200	14	0.1831382
	path:hsa04919	Thyroid hormone signaling pathway	89	13	0.0016727
	path:hsa05166	Human T-cell leukemia virus 1 infection	165	12	0.1831382
	path:hsa05206	MicroRNAs in cancer	168	12	0.1886212
	path:hsa05203	Viral carcinogenesis	147	11	0.1831382
	path:hsa05202	Transcriptional misregulation in cancer	104	10	0.0926317
	path:hsa05225	Hepatocellular carcinoma	100	10	0.0926317
	path:hsa05165	Human papillomavirus infection	223	10	0.5742798
	path:hsa05016	Huntington disease	134	8	0.3527942
4	path:hsa04024	cAMP signaling pathway	130	12	0.0926317
	path:hsa00310	Lysine degradation	46	9	0.0016727
	path:hsa04114	Oocyte meiosis	95	9	0.1006508
	path:hsa04310	Wnt signaling pathway	96	9	0.1010288
	path:hsa04720	Long-term potentiation	49	8	0.0130078
	path:hsa04921	Oxytocin signaling pathway	109	8	0.2405338
	path:hsa04916	Melanogenesis	56	7	0.0926317
	path:hsa04020	Calcium signaling pathway	124	7	0.4040354
	path:hsa05152	Tuberculosis	109	6	0.4262190
	path:hsa04360	Axon guidance	127	6	0.5393572
5	path:hsa04120	Ubiquitin mediated proteolysis	122	16	0.0016727
	path:hsa04010	MAPK signaling pathway	196	14	0.1831382
	path:hsa04141	Protein processing in endoplasmic reticulum	130	13	0.0525680
	path:hsa05205	Proteoglycans in cancer	128	13	0.0525680
	path:hsa04810	Regulation of actin cytoskeleton	148	11	0.1831382
	path:hsa05163	Human cytomegalovirus infection	154	10	0.2720810
	path:hsa03013	RNA transport	119	9	0.2034353
	path:hsa04660	T cell receptor signaling pathway	80	8	0.1006508
	path:hsa04015	Rap1 signaling pathway	137	8	0.3606783
	path:hsa04022	cGMP-PKG signaling pathway	107	7	0.2960050
6	path:hsa04140	Autophagy - animal	109	10	0.1006508
	path:hsa04714	Thermogenesis	153	9	0.3526443
	path:hsa04915	Estrogen signaling pathway	78	7	0.1831382
	path:hsa04371	Apelin signaling pathway	95	7	0.2405338
	path:hsa04910	Insulin signaling pathway	101	7	0.2720810
	path:hsa04072	Phospholipase D signaling pathway	107	7	0.2960050
	path:hsa04014	Ras signaling pathway	149	7	0.5409949
	path:hsa04722	Neurotrophin signaling pathway	87	6	0.2929396
	path:hsa04150	mTOR signaling pathway	104	6	0.4017740
	path:hsa04151	PI3K-Akt signaling pathway	225	6	0.8841293

Integration Frequency (Enrichment)

Table 5: Table of top 50 genes with the most frequent clonal enrichment.

Gene	Num. Patients	TDN Sites	Patient Sites	Onco-Related	Frequency Increase (%)
RBM27	6	0	12	FALSE	Inf
AGL	4	1	10	FALSE	1368.7
MCPH1	8	2	13	TRUE	854.7
BACH2	7	2	12	TRUE	781.2
NDFIP2	6	3	17	FALSE	732.3
GAK	8	3	17	TRUE	732.3
IKZF2	6	2	11	TRUE	707.8
GNA12	6	2	11	TRUE	707.8
FANCL	5	2	11	FALSE	707.8
PIKFYVE	9	2	11	FALSE	707.8
PAPOLA	6	3	16	FALSE	683.3
FAM117B	6	5	25	FALSE	634.4
CD55	9	3	14	TRUE	585.4
EP400P1	6	4	18	FALSE	560.9
FUNDC2	6	3	13	FALSE	536.4
GPHN	8	3	13	TRUE	536.4
NDUFV2	9	3	13	FALSE	536.4
UHRF1BP1	8	3	12	FALSE	487.5
RAB11FIP2	9	3	12	FALSE	487.5
TAF2	7	4	16	FALSE	487.5
PDS5B	8	4	16	TRUE	487.5
RBM39	10	5	19	TRUE	458.1
LRPPRC	10	4	15	FALSE	450.8
TET2	5	4	15	TRUE	450.8
FUS	4	3	11	TRUE	438.5
PIP5K1A	7	3	11	FALSE	438.5
LPXN	5	3	11	FALSE	438.5
RBMS1	7	3	11	FALSE	438.5
EEF2	8	3	11	FALSE	438.5
SLK	9	4	14	FALSE	414.0
MACROD2	7	4	14	FALSE	414.0
HSF2	8	4	14	FALSE	414.0
SIPA1L1	6	3	10	FALSE	389.6
VPS9D1	7	3	10	FALSE	389.6
GARS	4	3	10	FALSE	389.6
NSRP1	6	3	10	FALSE	389.6
SLBP	7	3	10	FALSE	389.6
PDCD10	8	3	10	FALSE	389.6
BZW2	6	3	10	FALSE	389.6
NAA16	6	3	10	FALSE	389.6
UBR3	5	3	10	FALSE	389.6
PELP1	11	8	26	FALSE	377.3
ZFAND3	9	4	13	FALSE	377.3
RNF10	8	5	16	FALSE	370.0
KDM4A	8	5	16	FALSE	370.0
HELLS	9	5	16	FALSE	370.0
RASA1	7	7	22	TRUE	361.6
PCMTD2	6	4	12	FALSE	340.6
ZNF512	9	4	12	FALSE	340.6
SYNJ1	5	4	12	FALSE	340.6

Integration Frequency (Depletion)

Table 6: Table of top 4 genes with the most frequent clonal depletion.

Gene	Num. Patients	TDN Sites	Patient Sites	Onco-Related	Frequency Increase (%)
RNPS1	5	69	16	FALSE	-65.9
EXOC2	5	34	10	FALSE	-56.8
EIF2B3	8	42	14	FALSE	-51.0
MROH1	8	142	71	FALSE	-26.6

Genes with the Most Abundant Clones

Table 7: Table of top 50 Genes containing the highest abundant clones.

Gene	Num. Patients	Peak Abundance	Peak Rel. Abund.	Clonal Gini Index	Onco-Related
TET2	7	814	0.989	0.920	TRUE
KCTD3	3	589	0.265	0.663	FALSE
PATL1	3	578	0.260	0.745	FALSE
PIKFYVE	8	410	0.273	0.878	FALSE
SRCAP	10	373	0.357	0.896	FALSE
MTMR3	5	261	0.041	0.872	TRUE
PCNX1	10	153	0.010	0.828	FALSE
PPP6R3	11	149	0.040	0.745	FALSE
SSH2	8	137	0.062	0.805	FALSE
RSRC1	7	109	0.014	0.811	FALSE
SNHG12	2	96	0.057	0.646	FALSE
MAPK14	8	91	0.018	0.784	TRUE
RPA3	4	87	0.011	0.767	FALSE
ZNF573	3	86	0.610	0.677	FALSE
MGA	10	85	0.013	0.762	FALSE
AQR	4	84	0.022	0.790	FALSE
LEF1	8	84	0.038	0.770	TRUE
LINC01473	2	82	0.075	0.488	FALSE
CARD8	11	79	0.056	0.700	TRUE
IQCB1	4	79	0.028	0.713	FALSE
DNAJC13	6	71	0.004	0.764	FALSE
EXOSC10	4	70	0.008	0.776	FALSE
ATP2A2	8	67	0.030	0.749	FALSE
SEC31A	6	66	0.004	0.752	FALSE
GPN1	2	62	0.017	0.711	FALSE
SMAP2	6	61	0.004	0.768	FALSE
TRIO	5	61	0.025	0.770	TRUE
ZZEF1	10	56	0.333	0.632	FALSE
CLK4	8	53	0.036	0.653	FALSE
JMJD6	2	53	0.015	0.755	FALSE
KDM5D	5	51	0.017	0.741	FALSE
UBR1	8	48	0.421	0.715	FALSE
MEMO1	4	47	0.006	0.722	FALSE
PTBP1	6	47	0.043	0.680	TRUE
DYNC1H1	6	44	0.003	0.721	FALSE
NGDN	3	44	0.005	0.623	FALSE
EIF2AK4	3	43	0.003	0.659	FALSE
MSH5-SAPCD1	4	43	0.039	0.708	FALSE
POLG2	2	43	0.003	0.708	FALSE
RASEF	1	43	0.005	0.477	FALSE
UXT-AS1	2	43	0.039	0.477	FALSE
ADD1	8	42	0.006	0.605	FALSE
GRB2	10	42	0.017	0.566	TRUE
KIFC1	7	42	0.003	0.694	FALSE
TAC3	2	42	0.018	0.477	FALSE
ZNF92	2	42	0.003	0.690	FALSE
ACTL6A	1	40	0.003	0.000	FALSE
ATP6V1G2-DDX39B	8	40	0.005	0.656	FALSE
PHF12	2	40	0.014	0.670	FALSE
MICAL2	1	39	0.028	0.000	FALSE

Longitudinal Observation

Table 8: Table of top 50 genes identified by longitudinal observations.

Gene	Time Span	Longest Time	Obs. Count	Num. Patients	Patient Sites	Peak Abund.	Onco-Related
FKBP5	1555.0	1825.0	4	11	30	15	FALSE
PTPRA	1555.0	1825.0	3	8	33	4	FALSE
TET2	1464.0	1584.0	7	5	15	814	TRUE
UBR1	1277.5	1825.0	4	7	13	48	FALSE
COX6B1	825.0	1095.0	3	9	17	5	FALSE
CCDC57	642.5	912.5	2	11	33	6	FALSE
KMT5B	642.5	912.5	2	10	26	7	FALSE
MACF1	519.5	547.5	3	12	32	11	TRUE
DNMT1	365.0	912.5	2	11	65	13	TRUE
STXBP5	350.0	360.0	4	8	15	8	FALSE
CASK	346.0	547.5	2	7	16	5	FALSE
RPTOR	346.0	360.0	2	15	85	11	FALSE
DIP2A	346.0	360.0	2	10	35	25	FALSE
PTBP1	346.0	360.0	2	7	17	47	TRUE
MIR4745	346.0	360.0	2	6	10	47	FALSE
ZZEF1	332.0	360.0	5	11	47	56	FALSE
SRCAP	332.0	360.0	5	10	21	373	FALSE
OGDH	332.0	360.0	4	4	10	17	FALSE
WDR82	277.5	547.5	3	7	20	7	TRUE
PIP5K1A	277.5	547.5	2	7	11	3	FALSE
EP400P1	260.0	270.0	3	6	18	5	FALSE
HSF1	256.0	270.0	3	12	45	10	FALSE
BOP1	256.0	270.0	3	11	34	10	TRUE
FNBP1	256.0	270.0	2	9	32	5	TRUE
PDS5B	256.0	270.0	2	8	16	10	TRUE
ACOX1	256.0	270.0	2	7	15	2	FALSE
PIK3C3	180.0	360.0	3	10	30	5	FALSE
IQGAP1	166.0	180.0	3	9	16	5	FALSE
SNAPC4	166.0	180.0	2	7	18	7	FALSE
UBE2J2	166.0	180.0	2	7	14	2	FALSE
SSH2	152.0	1095.0	4	8	26	137	FALSE
MED13	152.0	270.0	4	12	26	21	FALSE
CARD8	152.0	180.0	4	11	38	79	TRUE
LEF1	152.0	180.0	4	7	15	84	TRUE
VAV1	152.0	180.0	3	13	76	37	TRUE
STAG1	136.0	912.5	2	8	13	6	TRUE
PPP6R2	136.0	180.0	2	11	40	15	FALSE
RTTN	136.0	150.0	2	8	15	6	FALSE
MAPK8IP3	130.0	270.0	2	10	31	5	FALSE
SMG1	122.0	150.0	3	12	39	7	FALSE
INPP4B	122.0	150.0	3	11	28	3	FALSE
PIAS1	122.0	150.0	3	8	21	6	FALSE
DDX60	122.0	150.0	3	9	15	22	FALSE
ZNRD1ASP	122.0	150.0	3	6	14	4	FALSE
DPYD	122.0	150.0	2	12	33	22	FALSE
RUNX1	110.0	360.0	2	9	19	1	TRUE
ASH1L	106.0	1825.0	2	12	46	5	FALSE
WWOX	106.0	1095.0	3	5	11	5	TRUE
RFX2	106.0	360.0	2	7	14	9	TRUE
PPFIA1	106.0	270.0	4	5	11	20	FALSE

Reference Data

The NCBI RefGenes data set was used to identify gene regions (hg38) while genes identified as onco-related were from the Bushman Lab curated list of **onco-related genes**.

Gene Ontologies were extracted from the `GO.db` R-package (v3.4.1). KEGG pathways were acquired via interfacing with the KEGG web-server API through the `KEGGREST` R-package (v1.16.1). Gene lists, including RefSeq genes used for annotation of integration sites, were standardized to HGNC gene symbols (date: 2018-02-07). Groups identified in GO and KEGG analyses were determined from Jaccard distances between identified terms, followed by modularity-optimizing clustering from a weighted-undirected graph using a Louvain algorithm (**Blondel *et al.* 2008**). Terms within groups of GO or KEGG terms have greater overlap of gene lists between themselves that between terms found in other groups. This method was implemented to help reduce the functional redundancy commonly observed in GO and overlapping pathways observed with KEGG.

Comprehensive Genes of Interest Table

Table 9: Table of all genes identified within analysis.

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
VAV1	chr19	6,767,667	6,862,366	13	110.6	37	152.0	EAL
RNF157	chr17	76,137,452	76,245,311	12	77.8	28	14.0	EAL
AKAP13	chr15	85,375,615	85,754,358	11	340.6	27	7.0	EAL
CARD8	chr19	48,203,085	48,260,946	11	74.4	79	152.0	EAL
ZZEF1	chr17	3,999,444	4,147,959	11	64.4	56	332.0	EAL
CRAMP1	chr16	1,609,639	1,682,908	10	53.1	30	18.0	EAL
PAFAH1B1	chr17	2,588,628	2,690,615	10	110.7	30	7.0	EAL
SRCAP	chr16	30,694,140	30,745,129	10	137.3	373	332.0	EAL
UBAP2L	chr1	154,215,171	154,276,510	10	235.7	30	7.0	EAL
ST13	chr22	40,819,534	40,862,008	8	153.7	29	7.0	EAL
CLK4	chr5	178,597,663	178,632,053	7	259.0	53	106.0	EAL
PTBP1	chr19	792,391	817,327	7	316.1	47	346.0	EAL
TET2	chr4	105,140,874	105,284,803	5	450.8	814	1464.0	EAL
FANCA	chr16	89,732,550	89,821,657	15	94.3	21	15.0	EL
JPT2	chr16	1,673,276	1,707,072	15	106.7	23	46.0	EL
RPTOR	chr17	80,539,824	80,971,373	15	75.8	11	346.0	EL
PPP3CA	chr4	101,018,429	101,352,471	14	178.3	5	46.0	EL
ANKRD11	chr16	89,262,620	89,495,561	13	94.2	23	7.0	EL
EHMT1	chr9	137,613,991	137,841,126	13	111.1	3	50.0	EL
EP300	chr22	41,087,609	41,185,077	13	159.2	6	46.0	EL
LUC7L	chr16	183,968	234,482	13	35.9	30	7.0	AL
RABEP1	chr17	5,277,262	5,391,339	13	74.7	26	14.0	EA
JMJD1C	chr10	63,162,220	63,527,075	12	273.9	5	7.0	EL
MED13	chr17	61,937,604	62,070,282	12	154.6	21	152.0	EL
PPP6R3	chr11	68,455,717	68,620,333	12	182.9	149	14.0	EA
SMG1P1	chr16	22,432,007	22,497,220	12	225.2	3	100.0	EL
CREBBP	chr16	3,720,054	3,885,120	11	83.6	5	46.0	EL
GMDS	chr6	1,618,799	2,250,634	11	230.5	8	14.0	EL
SF1	chr11	64,759,603	64,783,844	11	84.6	16	14.0	EL
SUPT3H	chr6	44,821,729	45,383,051	11	187.4	8	14.0	EL
USP15	chr12	62,255,339	62,414,721	11	120.3	14	15.0	EL
XPO5	chr6	43,517,329	43,581,075	11	99.3	26	22.0	AL
ARHGAP15	chr2	143,124,329	143,773,352	10	120.3	7	5.0	EL
CHD4	chr12	6,565,081	6,612,433	10	131.9	19	7.0	EL
DDX42	chr17	63,769,188	63,824,317	10	114.7	6	14.0	EL
DIP2A	chr21	46,453,948	46,575,013	10	28.5	25	346.0	AL
EYA3	chr1	27,965,343	28,093,637	10	235.7	7	32.0	EL
KDM6A	chrX	44,868,174	45,117,612	10	181.5	9	106.0	EL
LRBA	chr4	150,259,658	151,020,497	10	99.7	17	106.0	EL
MGA	chr15	41,655,411	41,774,943	10	148.6	85	46.0	EA
NF1	chr17	31,089,926	31,382,677	10	193.7	14	14.0	EL

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
PIK3C3	chr18	41,950,197	42,086,482	10	267.2	5	180.0	EL
PRKACB	chr1	84,072,974	84,243,498	10	111.1	4	7.0	EL
SETD2	chr3	47,011,407	47,168,977	10	100.7	17	22.0	EL
SRRM2	chr16	2,747,328	2,776,412	10	183.3	32	1.0	EA
DDX60	chr4	168,211,290	168,323,807	9	144.8	22	122.0	EL
GRB2	chr17	75,313,075	75,410,709	9	17.5	42	14.0	AL
HELLS	chr10	94,540,766	94,607,099	9	370.0	15	106.0	EL
MED13L	chr12	115,953,575	116,282,186	9	104.0	38	7.0	EA
NDUFV2	chr18	9,097,629	9,139,345	9	536.4	15	22.0	EL
PIKFYVE	chr2	208,261,266	208,363,751	9	707.8	410	14.0	EA
RUNX1	chr21	34,782,800	35,054,298	9	248.8	1	110.0	EL
SLC6A16	chr19	49,284,634	49,330,217	9	82.1	2	14.0	EL
SMURF2	chr17	64,539,616	64,667,268	9	100.3	27	7.0	AL
TARSL2	chr15	101,648,751	101,729,442	9	340.6	22	7.0	EL
UBAC2	chr13	99,195,424	99,391,499	9	120.3	4	14.0	EL
UBR4	chr1	19,069,505	19,215,252	9	256.7	8	7.0	EL
USP25	chr21	15,725,024	15,885,071	9	218.2	10	46.0	EL
AP2B1	chr17	35,582,262	35,731,417	8	144.8	13	90.0	EL
ATP2A2	chr12	110,276,226	110,356,092	8	157.0	67	22.0	EA
DNAJC13	chr3	132,412,659	132,544,032	8	193.7	71	7.0	EA
DYNC1H1	chr14	101,959,527	102,055,798	8	193.7	44	7.0	EA
FAM13A	chr4	88,720,953	89,062,195	8	172.8	1	7.0	EL
HNRNPUL1	chr19	41,257,475	41,312,783	8	179.1	4	14.0	EL
MCPH1	chr8	6,401,591	6,653,505	8	854.7	9	7.0	EL
MOB3A	chr19	2,066,035	2,101,270	8	135.0	2	22.0	EL
PDCD4	chr10	110,866,794	110,905,006	8	105.6	26	14.0	AL
PDS5B	chr13	32,581,426	32,783,020	8	487.5	10	256.0	EL
PIAS1	chr15	68,049,178	68,196,466	8	285.5	6	122.0	EL
RSRC1	chr3	158,105,051	158,549,835	8	46.9	109	106.0	AL
SSH2	chr17	29,620,938	29,935,228	8	27.3	137	152.0	AL
STAG1	chr3	136,332,156	136,757,403	8	281.9	6	136.0	EL
LEF1	chr4	108,042,544	108,173,956	7	46.9	84	152.0	AL
MAPK14	chr6	36,022,676	36,116,236	7	108.1	91	7.0	AL
NEMP1	chr12	57,050,642	57,083,791	7	223.1	10	46.0	EL
PIP5K1A	chr1	151,193,543	151,254,531	7	438.5	3	277.5	EL
UBR1	chr15	42,937,899	43,111,088	7	-17.0	48	1277.5	AL
BRWD3	chrX	80,664,487	80,814,734	6	172.8	6	7.0	EL
EP400P1	chr12	132,079,282	132,131,340	6	560.9	5	260.0	EL
HERC4	chr10	67,916,898	68,080,346	6	34.6	35	106.0	AL
MIR4745	chr19	799,939	810,001	6	267.2	47	346.0	EL
MTOR	chr1	11,101,530	11,267,551	6	55.5	29	90.0	AL

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
NDFIP2	chr13	79,476,123	79,561,077	6	732.3	8	46.0	EL
PCNT	chr21	46,319,121	46,450,769	6	223.1	5	4.0	EL
TRIM33	chr1	114,387,776	114,516,160	6	303.9	17	106.0	EL
UBE2L3	chr22	21,544,446	21,629,034	6	193.7	3	7.0	EL
ECD	chr10	73,129,523	73,173,095	5	340.6	24	1.0	EA
MARF1	chr16	15,589,368	15,648,166	5	212.1	17	106.0	EL
PA2G4	chr12	56,099,318	56,118,910	5	130.8	38	14.0	AL
PPFIA1	chr11	70,265,699	70,389,501	5	169.3	20	106.0	EL
STAG3	chr7	100,172,723	100,219,387	5	169.3	35	0.0	EA
EHMT1	chr9	137,758,021	137,769,772	4	144.8	1	0.0	EL
KDM2A	chr11	67,114,268	67,263,079	14	14.5	7	50.0	L
NPLOC4	chr17	81,551,884	81,642,153	14	4.9	16	46.0	L
SARNP	chr12	55,747,462	55,822,756	14	136.6	5	22.0	E
SMG1P5	chr16	30,280,017	30,340,374	14	88.2	5	14.0	E
TNRC6C	chr17	77,999,236	78,113,835	14	75.2	6	22.0	E
CBFB	chr16	67,024,146	67,106,055	13	98.3	14	22.0	E
NSD1	chr5	177,128,078	177,305,213	13	12.2	6	62.0	L
UTRN	chr6	144,286,736	144,858,034	13	40.5	6	22.0	L
VPS8	chr3	184,807,142	185,057,614	13	230.5	10	46.0	E
ASH1L	chr1	155,330,260	155,567,533	12	2.4	5	106.0	L
DPYD	chr1	97,072,743	97,926,059	12	12.7	22	122.0	L
HSF1	chr8	144,286,568	144,319,726	12	-13.0	10	256.0	L
MACF1	chr1	39,079,166	39,492,138	12	6.8	11	519.5	L
PACS1	chr11	66,065,352	66,249,747	12	2.8	5	46.0	L
PSMD13	chr11	231,807	257,984	12	77.8	18	1.0	E
SAFB	chr19	5,618,034	5,673,478	12	58.2	17	0.0	E
SMG1	chr16	18,799,852	18,931,404	12	6.1	7	122.0	L
ZNF276	chr16	89,715,367	89,745,924	12	77.5	3	1.0	E
BOP1	chr8	144,257,045	144,296,438	11	8.6	10	256.0	L
CCDC57	chr17	82,096,469	82,217,829	11	-8.6	6	642.5	L
DIAPH2	chrX	96,679,662	97,605,598	11	95.8	20	46.0	E
DNMT1	chr19	10,128,343	10,200,135	11	4.9	13	365.0	L
FKBP5	chr6	35,568,584	35,733,583	11	-16.9	15	1555.0	L
FOXP1	chr3	70,949,713	71,588,989	11	88.0	4	14.0	E
FXR2	chr17	7,586,229	7,619,897	11	86.4	12	14.0	E
INPP4B	chr4	142,018,159	142,851,535	11	24.6	3	122.0	L
PELP1	chr17	4,666,383	4,709,337	11	377.3	3	14.0	E
PPP6R2	chr22	50,338,316	50,450,089	11	8.8	15	136.0	L
SEC16A	chr9	136,435,095	136,488,759	11	31.7	29	46.0	A
SNORD117	chr6	31,531,373	31,541,449	11	98.7	40	14.0	E
TRAPPC10	chr21	44,007,324	44,111,551	11	50.6	7	14.0	L

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
VPS13D	chr1	12,225,038	12,517,046	11	164.4	3	0.0	E
ZNF407	chr18	74,625,962	75,070,672	11	218.2	3	1.0	E
ATF7IP	chr12	14,360,631	14,507,935	10	66.0	8	7.0	L
CBLB	chr3	105,650,460	105,874,422	10	100.3	4	7.0	E
CLEC16A	chr16	10,939,487	11,187,189	10	291.7	9	0.0	E
CUX1	chr7	101,810,903	102,288,958	10	212.1	3	0.0	E
CYTH1	chr17	78,669,046	78,787,342	10	15.4	4	7.0	L
DDX10	chr11	108,660,024	108,945,930	10	193.7	3	14.0	E
EPB41	chr1	28,882,090	29,125,046	10	140.3	7	14.0	E
FCHSD2	chr11	72,831,744	73,147,098	10	72.0	5	22.0	E
GBE1	chr3	81,484,698	81,766,799	10	267.2	9	46.0	E
GLCCI1	chr7	7,963,742	8,094,079	10	64.2	4	84.0	L
KLF12	chr13	73,681,011	74,138,929	10	133.7	12	14.0	E
KMT5B	chr11	68,149,862	68,218,772	10	41.4	7	642.5	L
LRPPRC	chr2	43,881,223	44,001,005	10	450.8	7	0.0	E
MAPK8IP3	chr16	1,701,182	1,775,317	10	-12.4	5	130.0	L
MECP2	chrX	154,016,812	154,102,731	10	36.6	12	99.0	L
MGEA5	chr10	101,779,442	101,823,465	10	218.2	4	0.0	E
MIR6767	chr16	2,440,391	2,450,457	10	140.3	3	1.0	E
NCOA1	chr2	24,579,476	24,775,701	10	40.2	3	22.0	L
PBRM1	chr3	52,540,351	52,690,850	10	26.6	14	7.0	L
PCNX1	chr14	70,902,404	71,120,382	10	16.3	153	14.0	A
PHF14	chr7	10,968,871	11,174,623	10	125.2	3	7.0	E
PRPF6	chr20	63,976,077	64,038,100	10	120.3	9	0.0	E
RAB11FIP3	chr16	420,667	527,481	10	-18.4	7	46.0	L
RANBP9	chr6	13,616,497	13,716,564	10	167.0	6	7.0	E
RBM39	chr20	35,698,608	35,747,336	10	458.1	2	14.0	E
RBM5	chr3	50,083,907	50,123,964	10	135.0	10	1.0	E
RNF213	chr17	80,255,860	80,403,781	10	66.5	6	15.0	E
SAFB2	chr19	5,581,998	5,627,927	10	40.8	13	22.0	L
SIN3A	chr15	75,364,378	75,460,783	10	124.6	6	0.0	E
SMARCC1	chr3	47,580,887	47,786,915	10	34.4	3	7.0	L
SMG6	chr17	2,054,838	2,308,775	10	25.9	16	7.0	L
SUPT5H	chr19	39,440,545	39,481,668	10	157.0	3	0.0	E
VPS13B	chr8	99,008,265	99,882,586	10	67.9	5	0.0	E
XPO6	chr16	28,092,975	28,216,918	10	146.7	2	7.0	E
ZGPAT	chr20	63,702,441	63,741,142	10	-17.6	5	7.0	L
ZNF251	chr8	144,715,908	144,760,585	10	17.0	21	22.0	L
AKAP8L	chr19	15,375,047	15,424,121	9	157.0	3	0.0	E
AP3B1	chr5	77,997,325	78,299,755	9	157.0	3	0.0	E
CD55	chr1	207,316,471	207,365,966	9	585.4	4	0.0	E

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
CHD2	chr15	92,895,320	93,033,007	9	109.8	7	0.0	E
COX6B1	chr19	35,643,222	35,663,784	9	56.1	5	825.0	L
CPEB2	chr4	14,997,673	15,075,153	9	242.7	14	22.0	E
CSNK1D	chr17	82,237,660	82,278,742	9	22.4	8	7.0	L
DENND1B	chr1	197,499,748	197,780,493	9	10.2	3	7.0	L
DIP2B	chr12	50,499,984	50,753,667	9	55.0	4	7.0	L
DLG1	chr3	197,037,559	197,304,272	9	58.6	8	7.0	L
EIF4G3	chr1	20,801,291	21,181,888	9	102.8	3	0.0	E
ELP4	chr11	31,504,728	31,789,525	9	140.3	4	0.0	E
FNBP1	chr9	129,882,186	130,048,194	9	23.7	5	256.0	L
FOXJ3	chr1	42,171,538	42,340,877	9	92.8	12	7.0	L
FRYL	chr4	48,492,362	48,785,299	9	77.8	24	14.0	A
GANAB	chr11	62,619,825	62,651,726	9	66.0	6	7.0	L
GTF2I	chr7	74,652,664	74,765,692	9	193.7	7	0.0	E
HERC1	chr15	63,603,617	63,838,948	9	132.5	4	1.0	E
HNRNPUL2	chr11	62,707,624	62,732,385	9	36.0	9	76.0	L
IL4I1	chr19	49,884,655	49,934,539	9	-4.5	5	46.0	L
IQGAP1	chr15	90,383,240	90,507,243	9	6.8	5	166.0	L
KIAA1468	chr18	62,182,290	62,312,122	9	86.0	17	46.0	L
LOC101926943	chr7	74,683,936	74,733,918	9	193.7	7	0.0	E
LOC101929095	chr4	14,999,941	15,432,914	9	235.7	14	22.0	E
MARK3	chr14	103,380,363	103,508,829	9	80.3	15	14.0	E
NBEAL1	chr2	203,009,878	203,222,994	9	151.8	11	14.0	L
NOSIP	chr19	49,550,467	49,585,572	9	-35.7	29	0.0	A
NUP214	chr9	131,120,560	131,239,670	9	64.2	8	106.0	L
PAN3	chr13	28,133,505	28,300,338	9	127.0	8	14.0	E
PARP8	chr5	50,660,898	50,851,522	9	18.6	4	14.0	L
PLEKHA5	chr12	19,124,691	19,381,399	9	76.2	2	7.0	L
POLA2	chr11	65,256,851	65,303,685	9	25.9	9	7.0	L
POT1	chr7	124,817,385	124,934,983	9	66.5	26	0.0	A
PUM1	chr1	30,926,505	31,070,717	9	298.7	5	0.0	E
RAB11FIP2	chr10	117,999,915	118,051,884	9	487.5	19	1.0	E
COP1	chr1	175,939,825	176,212,244	9	172.8	3	14.0	E
RNF216	chr7	5,615,040	5,786,730	9	-11.9	15	106.0	L
SENP6	chr6	75,596,508	75,723,285	9	101.9	3	0.0	E
MTREX	chr5	55,302,747	55,430,581	9	212.1	6	0.0	E
SLK	chr10	103,962,184	104,034,233	9	414.0	13	1.0	E
SMCHD1	chr18	2,650,886	2,810,017	9	0.5	4	7.0	L
SYNRG	chr17	37,509,796	37,614,438	9	120.3	10	46.0	E
TCF20	chr22	42,155,012	42,288,927	9	46.9	10	7.0	L
TRAPPC8	chr18	31,824,172	31,948,128	9	83.6	4	7.0	L

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
TRPC4AP	chr20	34,997,403	35,097,815	9	212.1	3	0.0	E
UBE2I	chr16	1,304,152	1,332,018	9	120.3	4	50.0	E
VPS52	chr6	33,245,271	33,276,965	9	53.5	10	106.0	L
VPS53	chr17	503,667	719,856	9	226.4	6	0.0	E
ZCCHC7	chr9	37,115,471	37,363,148	9	148.6	7	46.0	E
ZFAND3	chr6	37,814,530	38,159,623	9	377.3	5	46.0	E
ZMYM4	chr1	35,263,966	35,426,944	9	157.0	3	0.0	E
ZNF512	chr2	27,577,968	27,628,215	9	340.6	2	0.0	E
ZNF609	chr15	64,494,419	64,691,067	9	193.7	6	0.0	E
ABCD2	chr12	39,546,219	39,625,041	8	267.2	8	14.0	E
ACSF3	chr16	89,088,808	89,160,846	8	115.4	4	0.0	E
ADD1	chr4	2,838,856	2,935,075	8	41.2	42	7.0	A
ANKHD1	chr5	140,396,813	140,544,856	8	125.2	6	7.0	E
ARID4B	chr1	235,161,894	235,333,219	8	167.0	6	0.0	E
ARIH1	chr15	72,469,325	72,591,555	8	128.5	5	7.0	L
ASXL2	chr2	25,728,752	25,883,516	8	-9.1	3	7.0	L
ATG5	chr6	106,179,476	106,330,820	8	340.6	2	0.0	E
BRWD1	chr21	39,180,477	39,318,786	8	153.7	4	14.0	E
CAMK2D	chr4	113,446,031	113,766,927	8	54.2	9	106.0	L
CAMK4	chr5	111,218,652	111,499,884	8	56.1	9	7.0	L
CDKAL1	chr6	20,529,456	21,237,403	8	31.4	5	106.0	L
CHAF1A	chr19	4,397,662	4,448,397	8	89.7	7	7.0	E
CHD3	chr17	7,879,804	7,917,757	8	179.1	3	0.0	E
CLTC	chr17	59,614,688	59,701,956	8	161.1	3	1.0	E
DAP3	chr1	155,684,090	155,744,009	8	252.5	2	0.0	E
DAZAP1	chr19	1,402,568	1,440,687	8	34.6	2	4.0	L
EED	chr11	86,239,383	86,283,810	8	311.2	6	14.0	E
EEF2	chr19	3,971,055	3,990,463	8	438.5	3	0.0	E
ERC1	chr12	986,207	1,500,933	8	64.2	4	7.0	L
FOCAD	chr9	20,653,308	21,000,955	8	285.5	17	7.0	E
GAK	chr4	844,274	937,390	8	732.3	4	14.0	E
GCN1	chr12	120,122,209	120,199,710	8	144.8	3	1.0	E
GIGYF2	chr2	232,692,304	232,865,577	8	214.7	2	0.0	E
GPHN	chr14	66,502,406	67,186,808	8	536.4	2	0.0	E
HNRNPR	chr1	23,299,689	23,349,364	8	235.7	7	46.0	E
HSF2	chr6	122,394,550	122,438,119	8	414.0	15	14.0	E
KDM4A	chr1	43,645,125	43,710,518	8	370.0	7	0.0	E
KIAA1109	chr4	122,165,602	122,367,759	8	120.3	19	1.0	E
KMT2C	chr7	152,129,924	152,441,005	8	15.4	4	7.0	L
KMT2E	chr7	105,009,189	105,119,085	8	140.3	3	0.0	E
LCOR	chr10	96,827,259	96,991,212	8	115.4	4	0.0	E

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
MIR5096	chr17	4,136,088	4,245,637	8	-20.3	9	14.0	L
MROH1	chr8	144,143,015	144,266,940	8	-26.6	5	46.0	DL
MUM1	chr19	1,349,976	1,383,431	8	11.9	4	90.0	L
NAA35	chr9	85,936,141	86,030,462	8	267.2	8	1.0	E
NCOA3	chr20	47,496,856	47,661,877	8	54.2	6	7.0	L
NEAT1	chr11	65,417,797	65,450,538	8	46.9	4	22.0	L
NELL2	chr12	44,503,274	44,918,928	8	81.4	2	4.0	L
PCM1	chr8	17,917,856	18,034,948	8	107.3	11	14.0	E
PDCD10	chr3	167,678,905	167,739,863	8	389.6	4	0.0	E
PDLIM5	chr4	94,446,856	94,673,227	8	252.5	3	4.0	E
PHF20L1	chr8	132,770,357	132,853,807	8	105.6	4	22.0	L
PLEC	chr8	143,910,146	143,981,745	8	-24.6	11	22.0	L
POGZ	chr1	151,397,723	151,464,465	8	130.8	26	0.0	A
PPIP5K2	chr5	103,115,247	103,209,911	8	172.8	3	0.0	E
PTPRA	chr20	2,859,194	3,043,669	8	67.1	4	1555.0	L
RASA2	chr3	141,482,046	141,620,363	8	115.4	7	14.0	E
RNF10	chr12	120,529,328	120,582,594	8	370.0	7	1.0	E
RRN3P2	chr16	29,069,841	29,121,717	8	144.8	2	4.0	E
RTTN	chr18	69,998,805	70,210,726	8	37.7	6	136.0	L
SACM1L	chr3	45,684,240	45,750,425	8	340.6	11	0.0	E
SEC24A	chr5	134,643,784	134,732,911	8	193.7	1	0.0	E
SLX4IP	chr20	10,430,302	10,633,034	8	151.8	3	7.0	L
SRP68	chr17	76,033,774	76,077,526	8	115.4	12	0.0	E
STIM2	chr4	26,855,690	27,030,381	8	132.5	2	4.0	E
STRN3	chr14	30,888,798	31,031,401	8	169.3	3	7.0	E
STXBP5	chr6	147,199,357	147,395,476	8	37.7	8	350.0	L
SYMPK	chr19	45,810,441	45,868,290	8	193.7	3	1.0	E
TANC2	chr17	63,004,536	63,432,706	8	340.6	11	7.0	E
TONSL	chr8	144,423,779	144,449,429	8	2.2	7	7.0	L
TOP1	chr20	41,023,817	41,129,486	8	242.7	2	0.0	E
UHRF1BP1	chr6	34,787,016	34,882,514	8	487.5	4	0.0	E
URI1	chr19	29,918,643	30,021,612	8	193.7	6	1.0	E
VAV3	chr1	107,566,159	107,969,923	8	135.0	1	0.0	E
VMP1	chr17	59,702,464	59,847,255	8	46.9	28	46.0	A
YLPM1	chr14	74,758,365	74,842,310	8	214.7	19	14.0	E
YWHAE	chr17	1,339,539	1,405,262	8	340.6	2	0.0	E
ZC3H13	chr13	45,949,464	46,057,778	8	311.2	6	46.0	E
ZFC3H1	chr12	71,604,600	71,668,969	8	340.6	4	0.0	E
ZNF148	chr3	125,220,668	125,380,354	8	153.7	6	7.0	E
ACOX1	chr17	75,936,510	75,984,434	7	4.9	2	256.0	L
ASCC3	chr6	100,503,194	100,886,372	7	327.3	4	7.0	E

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
ATF7	chr12	53,502,855	53,631,415	7	69.5	32	1.0	A
ATP8A1	chr4	42,403,374	42,662,105	7	80.8	4	14.0	L
BACH2	chr6	89,921,527	90,301,908	7	781.2	2	0.0	E
C6orf106	chr6	34,582,279	34,701,850	7	46.9	3	46.0	L
CASK	chrX	41,509,935	41,928,034	7	95.8	5	346.0	L
CCDC47	chr17	63,740,249	63,778,728	7	-13.6	6	22.0	L
CDC73	chr1	193,116,957	193,259,812	7	237.8	19	7.0	E
CHD8	chr14	21,380,193	21,442,298	7	223.1	16	14.0	E
CLEC2D	chr12	9,664,707	9,704,555	7	214.7	6	1.0	E
COG5	chr7	107,196,743	107,569,514	7	256.7	7	46.0	E
CSNK2A1	chr20	477,693	548,838	7	135.0	7	0.0	E
DEPDC5	chr22	31,748,950	31,912,034	7	223.1	4	0.0	E
DNAJC5	chr20	63,890,101	63,941,031	7	120.3	2	7.0	E
FRG1BP	chr20	30,372,163	30,424,842	7	340.6	4	7.0	E
GATAD2B	chr1	153,799,906	153,927,975	7	61.6	2	7.0	L
GMCL1	chr2	69,824,605	69,886,395	7	311.2	4	0.0	E
HTT	chr4	3,069,680	3,248,960	7	46.9	25	1.0	A
KDM7A	chr7	140,079,745	140,181,941	7	242.7	5	7.0	E
KLRG1	chr12	8,945,043	9,015,744	7	-10.2	3	7.0	L
LSM2	chr6	31,792,391	31,811,984	7	-29.3	2	14.0	L
MACROD2	chr20	13,990,499	16,058,196	7	414.0	1	0.0	E
MAN1B1	chr9	137,081,926	137,114,187	7	83.6	20	5.0	E
MAP4K3	chr2	39,244,265	39,442,312	7	46.9	15	46.0	L
MATR3	chr5	139,268,751	139,336,677	7	212.1	6	1.0	E
MIA2	chr14	39,228,909	39,356,193	7	161.1	10	7.0	E
NAA15	chr4	139,296,466	139,395,781	7	212.1	11	0.0	E
NAA38	chr17	7,851,680	7,890,388	7	-0.9	15	14.0	L
NBAS	chr2	15,161,907	15,566,348	7	22.4	4	106.0	L
NUMA1	chr11	71,997,863	72,085,693	7	135.0	3	0.0	E
NUP62	chr19	49,901,825	49,934,731	7	-12.8	5	46.0	L
NUP88	chr17	5,379,832	5,424,739	7	129.5	4	0.0	E
PAPD4	chr5	79,607,419	79,691,648	7	281.9	2	0.0	E
PHF3	chr6	63,630,801	63,720,522	7	144.8	7	4.0	E
PMS2P1	chr7	100,315,639	100,341,307	7	193.7	10	0.0	E
PPP1R16A	chr8	144,472,981	144,507,121	7	35.6	4	7.0	L
PPP4R2	chr3	72,991,742	73,074,201	7	303.9	2	7.0	E
PRKCA	chr17	66,297,807	66,815,744	7	128.5	7	7.0	L
PRKN	chr6	161,342,557	162,732,802	7	303.9	3	0.0	E
PRPSAP1	chr17	76,305,735	76,359,149	7	149.7	2	46.0	E
RAD51B	chr14	67,814,778	68,688,106	7	46.9	2	22.0	L
RASA1	chr5	87,263,252	87,396,926	7	361.6	3	0.0	E

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
RBMS1	chr2	160,267,150	160,498,807	7	438.5	5	1.0	E
RFX2	chr19	5,988,163	6,115,653	7	71.3	9	106.0	L
ROCK1	chr18	20,944,741	21,116,851	7	105.6	13	14.0	E
RUNX2	chr6	45,323,316	45,556,082	7	248.8	8	14.0	E
SENP3-EIF4A1	chr17	7,556,991	7,584,006	7	157.0	5	46.0	E
SEPT7	chr7	35,795,985	35,912,105	7	316.1	10	7.0	E
SLBP	chr4	1,687,730	1,717,741	7	389.6	2	0.0	E
SMAD2	chr18	47,828,094	47,936,146	7	172.8	7	0.0	E
SNAPC4	chr9	136,370,568	136,403,437	7	103.4	7	166.0	L
SNTB1	chr8	120,530,744	120,817,069	7	172.8	3	4.0	E
SPEN	chr1	15,842,863	15,945,455	7	58.2	2	7.0	L
SPG7	chr16	89,503,387	89,562,768	7	7.7	10	7.0	L
SYNE1	chr6	152,116,683	152,642,399	7	46.9	22	46.0	L
SYNE2	chr14	63,847,964	64,231,451	7	69.5	8	22.0	L
TAF2	chr8	119,725,773	119,837,834	7	487.5	8	1.0	E
THEMIS	chr6	127,703,193	127,923,631	7	-2.1	5	106.0	L
TTC21B	chr2	165,868,361	165,958,838	7	36.4	5	46.0	L
UBE2J2	chr1	1,248,911	1,278,854	7	-37.7	2	166.0	L
UBE3A	chr15	25,332,248	25,444,028	7	149.7	10	0.0	E
UBR5	chr8	102,247,273	102,417,689	7	95.8	19	7.0	L
VPS9D1	chr16	89,702,132	89,725,986	7	389.6	5	0.0	E
VRK3	chr19	49,971,466	50,030,548	7	-13.9	2	7.0	L
WDR82	chr3	52,249,421	52,283,643	7	33.5	7	277.5	L
YTHDF3	chr8	63,163,552	63,217,788	7	63.2	13	7.0	L
ZNF81	chrX	47,831,901	47,927,256	7	175.4	7	7.0	E
ZNRF2	chr7	30,279,306	30,372,692	7	144.8	3	14.0	E
ANXA1	chr9	73,146,730	73,175,394	6	144.8	5	14.0	L
ARHGEF6	chrX	136,660,550	136,786,344	6	193.7	5	7.0	E
ATE1	chr10	121,735,420	121,933,801	6	267.2	14	46.0	E
ATP9B	chr18	79,064,274	79,383,282	6	79.5	2	7.0	L
BAG6	chr6	31,634,027	31,657,700	6	22.4	36	46.0	A
BCAS3	chr17	60,672,774	61,397,838	6	-36.1	2	7.0	L
BCKDHB	chr6	80,101,609	80,351,270	6	267.2	3	0.0	E
BIRC6	chr2	32,352,027	32,623,898	6	311.2	3	0.0	E
BUB1B	chr15	40,156,008	40,226,136	6	267.2	7	14.0	E
BZW2	chr7	16,641,133	16,711,523	6	389.6	6	0.0	E
CD96	chr3	111,537,078	111,670,991	6	252.5	6	0.0	E
CHD1	chr5	98,850,203	98,931,534	6	277.7	8	1.0	E
CHMP2B	chr3	87,222,262	87,260,548	6	88.8	24	7.0	A
CLASP2	chr3	33,491,245	33,723,213	6	138.7	5	7.0	L
CUL3	chr2	224,465,149	224,590,397	6	252.5	8	22.0	E

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
DERL2	chr17	5,466,250	5,491,230	6	135.0	37	0.0	A
DNAJC1	chr10	21,751,547	22,008,721	6	17.5	38	7.0	A
DOT1L	chr19	2,159,148	2,237,578	6	23.7	6	14.0	L
ERGIC2	chr12	29,335,645	29,386,210	6	169.3	3	0.0	E
FAM117B	chr2	202,630,177	202,774,757	6	634.4	6	14.0	E
FUNDC2	chrX	155,021,788	155,061,916	6	536.4	9	46.0	E
GNA12	chr7	2,723,105	2,849,325	6	707.8	2	0.0	E
GTDC1	chr2	143,941,013	144,337,534	6	256.7	9	7.0	E
IKZF2	chr2	212,994,685	213,156,609	6	707.8	1	0.0	E
MED12L	chr3	151,081,797	151,441,677	6	169.3	7	7.0	E
MIR5096	chr1	15,866,148	15,910,467	6	79.5	2	7.0	L
MIR5096	chr22	37,663,025	38,029,093	6	12.3	6	7.0	L
MMP23A	chr1	1,627,779	1,706,808	6	58.2	17	14.0	L
NAA16	chr13	41,306,204	41,382,030	6	389.6	2	1.0	E
NFKBIL1	chr6	31,541,850	31,563,829	6	-19.5	4	7.0	L
NSRP1	chr17	30,111,806	30,191,475	6	389.6	2	0.0	E
NUCB1	chr19	48,895,049	48,928,283	6	303.9	2	0.0	E
NUCB1-AS1	chr19	48,905,929	48,923,891	6	303.9	2	0.0	E
OPRM1	chr6	154,005,495	154,251,867	6	46.9	3	14.0	L
PAPOLA	chr14	96,497,375	96,572,116	6	683.3	3	0.0	E
PCMTD2	chr20	64,250,694	64,281,226	6	340.6	17	1.0	E
PDE12	chr3	57,551,246	57,661,480	6	303.9	3	0.0	E
PHACTR4	chr1	28,364,581	28,505,369	6	1.0	3	22.0	L
RAB28	chr4	13,362,722	13,489,365	6	172.8	12	0.0	E
RBM27	chr5	146,198,599	146,294,221	6	Inf	3	0.0	E
RIPOR2	chr6	24,799,280	25,047,288	6	46.9	10	7.0	L
RSBN1L	chr7	77,691,425	77,784,803	6	193.7	31	1.0	A
SEC23A	chr14	39,026,918	39,108,528	6	73.6	7	7.0	L
SHPRH	chr6	145,879,808	145,969,097	6	267.2	4	0.0	E
SIPA1L1	chr14	71,524,311	71,746,229	6	389.6	22	7.0	E
SPPL3	chr12	120,757,509	120,909,352	6	57.4	9	14.0	L
TCF25	chr16	89,868,585	89,916,384	6	-17.0	5	106.0	L
UBE2F-SCLY	chr2	237,961,944	238,104,413	6	34.6	6	46.0	L
UNKL	chr16	1,358,204	1,419,720	6	79.5	7	15.0	L
WWP1	chr8	86,337,764	86,472,949	6	177.4	3	7.0	E
ZNF473	chr19	50,020,892	50,053,774	6	4.9	3	22.0	L
ZNRD1ASP	chr6	29,996,010	30,066,189	6	128.5	4	122.0	L
BMP2K	chr4	78,771,377	78,917,187	5	303.9	3	0.0	E
CLEC2B	chr12	9,847,368	9,874,859	5	252.5	7	0.0	E
CSNK1G1	chr15	64,160,516	64,361,259	5	33.5	8	7.0	L
CYLD	chr16	50,737,049	50,806,935	5	172.8	3	7.0	E

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
ELMO1	chr7	36,847,905	37,454,326	5	6.8	34	46.0	A
EVL	chr14	99,966,474	100,149,236	5	13.0	4	7.0	L
FANCL	chr2	58,154,242	58,246,380	5	707.8	6	0.0	E
HIVEP1	chr6	12,007,490	12,169,999	5	169.3	1	0.0	E
HS2ST1	chr1	86,909,651	87,114,998	5	303.9	12	7.0	E
HSF5	chr17	58,415,166	58,493,401	5	46.9	34	1.0	A
IQCB1	chr3	121,764,760	121,840,079	5	144.8	79	15.0	A
KIFC1	chr6	33,386,535	33,414,922	5	46.9	42	7.0	A
KMT2D	chr12	49,013,974	49,060,324	5	7.7	33	0.0	A
LPXN	chr11	58,521,870	58,583,239	5	438.5	9	46.0	E
MBD3	chr19	1,571,670	1,597,761	5	267.2	2	7.0	E
MIR5096	chr1	235,507,822	235,723,113	5	157.0	1	0.0	L
N4BP1	chr16	48,533,725	48,615,209	5	291.7	32	106.0	A
PAG1	chr8	80,962,810	81,117,068	5	1.0	2	32.0	L
PDE7A	chr8	65,709,333	65,846,734	5	79.5	4	7.0	L
PHF20	chr20	35,767,000	35,955,366	5	7.7	38	7.0	A
PLPPR3	chr19	807,487	826,952	5	169.3	6	1.0	E
POM121	chr7	72,874,334	72,956,440	5	414.0	29	1.0	A
RELB	chr19	44,996,448	45,043,198	5	223.1	2	0.0	E
SEC31A	chr4	82,813,508	82,905,571	5	46.9	66	7.0	A
SEPT9	chr17	77,276,409	77,505,596	5	-16.1	27	7.0	A
SFI1	chr22	31,491,138	31,623,551	5	-33.2	24	7.0	A
SMAP2	chr1	40,368,705	40,428,326	5	120.3	61	7.0	A
STX8	chr17	9,245,470	9,580,958	5	109.8	4	7.0	L
SYNJ1	chr21	32,623,758	32,733,040	5	340.6	4	0.0	E
TNKS	chr8	9,550,934	9,787,346	5	340.6	8	7.0	E
TTC3	chr21	37,068,183	37,208,118	5	223.1	2	0.0	E
UBR3	chr2	169,822,507	170,089,129	5	389.6	11	46.0	E
WWOX	chr16	78,094,412	79,217,667	5	24.3	5	106.0	L
AGL	chr1	99,845,083	99,929,023	4	1368.7	4	0.0	E
ANKRD46	chr8	100,504,751	100,564,786	4	95.8	24	7.0	A
AQR	chr15	34,851,350	34,974,794	4	28.5	84	14.0	A
CAAP1	chr9	26,835,684	26,897,828	4	-8.2	27	46.0	A
CNOT6	chr5	180,489,398	180,583,405	4	22.4	2	7.0	L
DENND6A	chr3	57,620,453	57,698,089	4	267.2	3	0.0	E
EXOSC10	chr1	11,061,612	11,104,910	4	71.3	70	7.0	A
FAM13B	chr5	137,932,915	138,038,113	4	223.1	3	7.0	E
FUS	chr16	31,175,109	31,199,871	4	438.5	12	0.0	E
GARS	chr7	30,589,734	30,639,032	4	389.6	3	0.0	E
GOLPH3L	chr1	150,641,224	150,702,196	4	-26.6	24	7.0	A
KDM5D	chrY	19,700,414	19,749,939	4	144.8	2	14.0	A

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
MAD1L1	chr7	1,810,791	2,237,948	4	20.2	35	46.0	A
MAP2K2	chr19	4,085,321	4,129,129	4	46.9	2	106.0	L
MEMO1	chr2	31,862,809	32,016,052	4	-46.6	47	7.0	A
MSH5-SAPCD1	chr6	31,734,947	31,769,847	4	-51.0	43	7.0	A
MTMR3	chr22	29,878,168	30,035,868	4	-17.4	261	106.0	A
OGDH	chr7	44,601,521	44,714,070	4	-22.7	17	332.0	L
PATL1	chr11	59,631,715	59,674,038	4	95.8	578	332.0	A
PDCD11	chr10	103,391,654	103,451,262	4	193.7	27	152.0	A
PDE3B	chr11	14,638,722	14,877,058	4	95.8	35	7.0	A
PEX5	chr12	7,184,162	7,223,573	4	340.6	2	0.0	E
RAB18	chr10	27,499,173	27,547,237	4	17.5	24	7.0	A
RMND5A	chr2	86,715,290	86,783,041	4	Inf	26	0.0	A
RPA3	chr7	7,631,562	7,723,607	4	83.6	87	106.0	A
TMTC3	chr12	88,137,295	88,204,887	4	223.1	2	0.0	E
TRIO	chr5	14,138,701	14,515,204	4	928.1	61	106.0	A
XPO1	chr2	61,472,933	61,543,283	4	83.6	25	106.0	A
ZC3H7A	chr16	11,745,585	11,802,258	4	311.2	11	1.0	E
AKAP9	chr7	91,935,874	92,115,673	3	17.5	25	0.0	A
CHD1L	chr1	147,168,193	147,300,766	3	Inf	25	1.0	A
DCUN1D4	chr4	51,837,999	51,921,837	3	22.4	32	106.0	A
EIF2AK4	chr15	39,929,123	40,040,596	3	634.4	43	7.0	A
NGDN	chr14	23,464,688	23,483,193	3	120.3	44	0.0	A
PRKD2	chr19	46,669,315	46,722,127	3	144.8	24	22.0	A
RABGAP1	chr9	122,936,008	123,109,868	3	2.8	29	7.0	A
SNAP29	chr22	20,854,003	20,896,213	3	487.5	27	7.0	A
STT3B	chr3	31,527,500	31,642,622	3	252.5	6	0.0	E
ZNF573	chr19	37,733,301	37,784,590	3	17.5	86	735.0	A
C20orf196	chr20	5,745,386	5,869,407	2	-2.1	1	0.0	A
CRTAP	chr3	33,108,957	33,152,773	2	193.7	35	7.0	A
GPN1	chr2	27,623,647	27,655,846	2	46.9	62	1.0	A
JMJD6	chr17	76,707,831	76,731,799	2	193.7	53	1.0	A
KCTD3	chr1	215,562,378	215,626,821	2	-26.6	1	0.0	A
LINC01473	chr2	186,028,533	186,091,317	2	Inf	82	7.0	A
LOC101927151	chr19	27,788,466	27,811,780	2	193.7	31	46.0	A
PHF12	chr17	28,900,252	28,956,490	2	17.5	40	106.0	A
POLG2	chr17	64,472,784	64,502,066	2	83.6	43	7.0	A
RASEF	chr9	82,974,584	83,068,128	2	Inf	43	46.0	A
SNHG12	chr1	28,573,537	28,586,854	2	340.6	96	332.0	A
TAC3	chr12	57,004,996	57,021,560	2	-51.0	42	7.0	A
TGFBR2	chr3	30,601,501	30,699,141	2	193.7	31	7.0	A
ACTL6A	chr3	179,557,879	179,593,405	1	-70.6	40	0.0	A

Table 9: Table of all genes identified within analysis. *(continued)*

Gene	Chromosome	Start Pos.	End Pos.	Patients	Freq. Change (%)	Peak Abund.	Long. Obs.	Criteria
C19orf48	chr19	50,792,692	50,809,853	1	46.9	28	46.0	A
CD109	chr6	73,691,084	73,833,317	1	-81.6	32	0.0	A
KARS	chr16	75,622,723	75,652,687	1	46.9	24	0.0	A
MICAL2	chr11	12,105,575	12,268,790	1	-26.6	39	15.0	A
RBAK-RBAKDN	chr7	5,040,820	5,078,223	1	46.9	28	7.0	A
RTCA-AS1	chr1	100,259,741	100,271,174	1	Inf	27	0.0	A
UXT-AS1	chrX	47,653,832	47,665,111	1	Inf	43	7.0	A
ZNF92	chr7	65,368,798	65,406,135	1	Inf	42	7.0	A
EIF2B3	chr1	44,845,521	44,991,722	8	-51.0	6	1.0	D
EXOC2	chr6	480,137	698,141	5	-56.8	4	0.0	D
RNPS1	chr16	2,248,115	2,273,412	5	-65.9	2	0.0	D