Global changes in processing of mRNA 3' untranslated regions characterize clinically distinct cancer subtypes: Supplemental Materials

Priyam Singh*, Travis L Alley, Sarah M Wright, Sonya Kamdar, William Schott, Robert Y Wilpan, Kevin D Mills, Joel H Graber
The Jackson Laboratory, 600 Main Street, Bar Harbor, Maine USA
*Bioinformatics Program, Boston University, 24 Cummington Street, Boston MA USA

SUPPLEMENTAL METHODS

Mice: Mice containing germline, gene targeted knockout alleles of *Lig4*, *Dclre1c/Art*, or *Trp53* were rederived by *in vitro* fertilization (IVF) in a C57BL/6J strain background, and then backcrossed with C57BL/6J mice for at least 4 generations to generate heterozygous animal with a minimum of 95% C57BL/6J genetic content. *Trp53* heterozygotes were intercrossed with either *Lig4* or *Dclre1c/Art* heterozygotes, and subsequently incrossed to generate double mutant (*Lig4 Trp53* or *Art Trp53*). All animal work was carried out in accordance with IACUC approved protocols.

Quantitative Reverse Transcriptase PCR: Briefly, tumor tissues were stored in RNAlater® (Ambion Austin, TX) per manufacturer's instructions and homogenized in Trizol® Reagent (Invitrogen, Carlsbad, CA). Total RNA was isolated by standard Trizol methods, and quality was assessed using a 2100 Bioanalyzer instrument. The RNA was then treated with RNase-free DNase I (Qiagen, Valencia, CA) according to manufacturer's protocol. 2µg of total RNA was reverse transcribed employing standard Oligo dT priming methods and Superscript III enzyme (Invitrogen, Carlsbad, CA) according to the manufacturer's protocols. Diluted reaction products were then used in a subsequent PCR reaction containing Power SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA) and gene-specific primers. Real-Time PCR reactions were performed using the ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems) with recommended thermal cycling protocols and 40 cycles of amplification. Threshold cycle (Ct) values were determined using the supplied Sequence Detection System (SDS v2.2) software package.

Two sets of primers were designed for each gene (Table S3), probing both the isoforms (called common probe) and the extended 3'-UTR (called extended probe). Ratio difference was calculated relative to normal pro-B-cells using the following formula:

$$\Delta\Delta Ct = \Delta Ct_{\text{extended}} - \Delta Ct_{\text{common}} = (Ct_{\text{ext,tumor}} - Ct_{\text{ext,proB}}) - (Ct_{\text{common,tumor}} - Ct_{\text{common,proB}})$$

Sample isolation, preparation, and histopathology: Single-cell suspensions for flow cytometric analysis of pro-B-cells, B-cells, and tumor samples were prepared by dispersion of bone marrow, spleen, or lymph nodes (respectively) using fine, sterile mesh. Cell suspensions were prepared in RPMI 1640 medium containing with 10%fetal calf serum. Cell preparations were then stained with an antibody cocktail specific for B-lymphocytes, containing antibodies directed against the pan-B cell maker B220, and the developmental stage markers CD19, CD43, and IgM. Flow cytometry was carried out using a Becton Dickinson FACSCalibur cytometer with CellQuest Pro acquisition software, and analyzed using FlowJo 2.0 software. Cell sorting to obtain purified populations of progenitor or mature B cells was carried out on a

FACSvantage SE/DiVa cell sorter after staining of whole bone marrow or splenocyte preparations (respectively) with the same B-cell cocktail used for flow cytometry.

For histopathology, lymph node-derived tumor samples were treated with Bouin's fixative and embedded in paraffin. After sectioning, slides were stained with hematoxylin & eosin (H&E) prior to imaging. At necropsy, lymphoma tissue samples were obtained from cervical lymphnode or thymus lymphoma foci, submerged in RNALater, and stored until processing for total RNA recovery. Normal pro-B-cells were sorted by fluorescence activated cell sorting (FACS) to obtain B220+ CD43+ IgM- cells from bone marrow of normal C57BL/6J mice and B-cells were the B220+ fraction from spleen of normal C57BL/6J mice, sorted by FACS.

Western Blot Analysis: Tumors and cultured cells from APN, APC and LPC mice, Normal pro-B-cells (B220+ CD43+ IgM- cells from bone marrow of normal C57BL/6J mice) and mature B-cells (B220+ fraction from spleen of normal C57BL/6J mice, sorted by FACS) were lysed on ice in RIPA buffer (25 mM Tris pH 7.5, 150 mM NaCl, 1% Triton X-100, 0.1% SDS, 1.0% Na-Deoxycholate) with "Complete Mini" protease inhibitor cocktail from Roche Diagnostics, 11.836.153.001. The protein content of the soluble fractions was quantitated using the Bradford Reagent from Sigma, B6919. Ten micrograms of protein were loaded per lane of a NuPage Bis-Tris SDS-PAGE gel, 4-12%, from Invitrogen, NP0329, and were transferred to Invitrolon Polyvinylidene fluoride (PVDF) membrane from invitrogen, LC2005, and blocked with 10% milk/Tris-buffered saline, 1% tween-20 (TBST) for an hour at room temperature. Blots were incubated with primary antibodies in 5% milk/TBST overnight at 4^0C, HRP conjugated secondary antibodies for an hour at room temperature, and developed with the SuperSignal West Dura Extended Duration Substrate from Pierce, prod. #34076. The UBE2A antibody was from Abcam, ab31917 and CSTF3 antibody was from Novus Biologicals, NB100-60463. Beta-Actin antibody, AB8227-50 from Abcam was used as the loading control. Beta-Actin results for each lane were quantitated using the software "Quantiy One" from BioRad and used to normalize the UBE2A and CSTF3 results.

Selection of gene-segmentation pairs for heatmap and crossvalidation: Segmentation points within a probeset are defined as points where the flanking probes have a difference in the change in expression based on a comparison of the three probes immediately up and downstream. Differential segmentations were determined in the following manner. All segmentation point scores were assessed individually for each array of each sample, as compared to the average intensity of the control samples. Gene-segmentation pairs included in the heatmap and crossvalidation were selected with a modified *t*-test comparison that compared all ratio-difference (rdiff) scores for that particular segmentation point for all arrays within each sample. Thresholds were chosen empirically to optimize the crossvalidation or hierarchical clustering quality.

Determining over/underrepresentation of statistically differentially processed genes: The overlap of genes with alternative processing events between different tumor classes was determined by collecting the processing pattern across all four comparisons (restricting to genes that displayed alternative processing in

at least one comparison). The overlap in alternatively processed genes between the lymphoma subtypes and mature B-cells was determined using A Venn diagram (created using the online tool Venny http://bioinfogp.cnb.csic.es/tools/venny/index.html). Over and under representation of the overlap categories was determined by a permutation analysis that used the same total number of genes across all tests, and independently assigned membership to each sample class according to the empirically observed probability that any given gene was alternatively processed in that sample. Significance was determined as a value that was not observed in 1000 random fills of the Venn diagram.

Comparison of 3'-UTR characteristics in genes with no alternative processing: Genes that are differentially expressed without evidence of differential processing were investigated to identify systematic characteristics of 3'-UTR sequences. At least 5 lymphoma samples were compared with two pro-B-cell samples. A *t*-test was performed on the means of probe intensities of all genes between lymphoma samples and progenitor samples. The method of Benjamini and Hochberg (1) was used to correct for multiple testing. Genes without evidence of alternative processing were ranked order of their relative expression compared to progenitor B-cells, and the top and bottom 5% of this sorted list was extracted resulting in ~700 genes in each data set. The 3′-UTR lengths of these genes were calculated by using the 3′-most probe for that gene in the Affymetrix gene chip. The location of the 3′ most probe was used to interrogate into PACdb (2) to extract the polyadenylation site that is immediately downstream of the chosen probe. In case of genes that had evidence of multiple polyadenylation sites downstream of the 3′ most probes, it is computationally not possible to determine the correct polyadenylation. To be stringent in our analysis, the polyadenylation site that was immediately downstream of the 3′ most probes was selected to calculate the 3′-UTR length**.**

Gene ontology analysis: The set of genes that are differentially processed in lymphoma subtypes and mature B-cells were used to identify over/under-representation in biological processes using GOstat (3), which uses a Fisher's exact test with multiple test corrections to assess over and underrepresented GO categories in query list of genes relative to a reference set. Several reference sets were used, including all genes, and only those annotated as present on the microarray of interest. Empirical investigation showed that the returned categories were largely constant, but p-values fluctuated. Ontology reference sets used in this analysis were drawn from the September 4th, 2009 MGI (4) mouse release and the September 1st, 2009 EBI GOA human release (5).

Determining the contribution of miRNA targeting towards destabilizing alternative isoforms in APA genes: Contribution of putative miRNAs towards changes in isoform abundance in the three subtypes of lymphomas was investigated. Only genes with evidence of differential processing that occurred in the 3′-UTR was used for this analysis. The portion of the 3′-UTR that is differentially included or excluded (based on processing changes) extends from the first base of the probe immediately downstream of the segmentation point to the most distal polyadenylation site. This sequence was extracted for all differentially processed genes (method flowchart is illustrated in Figure S3).

Statistical analysis (differential word analysis, DWA) was performed on the differentially excluded sequences (DES) and the differentially included sequence (DIS) in each lymphoma subtype to determine hexamers that are comparatively overrepresented in the first set (and correspondingly underrepresented in the second set) and vice-versa. When counting the number of sequences that have a specific hexamer, multiple occurrences of the hexamer in a single sequence was counted as a single occurrence. A *p*-value was determined through a permutation analysis that randomized the assignment to sequence set, while preserving the sizes of the starting sets. The permutation was repeated 1000 times, and the resulting *p*-values represent the fraction of permutations with an equal or greater difference.

Hexamers that were present with significant *p*-values in the DWA analysis were used to query into a database of computationally predicted miRNA target sequences, miRBase (6). Previous microarray studies have identified a number of miRNAs that are differentially expressed in B-cell malignancies in mouse and human data (7). MiRNAs that are significantly differentially expressed in B-cell malignancies and B-cell development when compared to normal B-cells was compiled using 2 different studies (7); (8). Hexamers with statistically significant differences were matched with the seed regions of the miRNAs compiled to determine putative miRNAs that contribute to changes in isoform abundance in the three subtypes of lymphomas.

Polyadenylation *trans*-factor mRNA expression analysis: From a total of 29 polyA factors (9), all with probesets present in our custom CDF were used. Probe-level data (background corrected and normalized) were used to calculate mean probe intensities in the 3 subtypes of lymphoma samples, mature B-cells and normal pro-B-cell samples. Mean probe intensities from the three lymphoma samples and mature B-cells were divided by the mean probe intensities in the pro-B-cell samples to compare mRNA abundance of *trans*-factors. Processing change *t*-values and ratio differences were calculated as described in the main text. Results are displayed in Figures S8 and S9.

Estimation of FDR: False Discovery Rates (FDR values were estimated as the ratio of above-threshold segmentations in a null model to above-threshold segmentations in the true distribution. The null model was generated using comparisons between replicate arrays rather than between the samples, without randomization of the probeset order. The estimated FDR for the *GV-MII* was 0.05 based on the observed number of events in each set.

SUPPLEMENTAL RESULTS

Specific differentially processed sets of genes are prognostic in lymphoma subtypes. In light of the finding that posttranscriptional processing could vary in specific genes (e.g., *Pik3ap1*) between subtypes of lymphomas, we compared the patterns for all genes, limiting the analysis to genes with evidence of a change in at least one of the samples, and selected the gene/segmentation combinations that were most differential. Mature B-cells were included in this analysis, as alternative polyadenylation is known to play an important role during B-cell development (10). A heatmap representation of all genes with a single segmentation point (Main Text Figure 5) demonstrates the common patterns within and differences between subtypes of lymphoma. The sets of genes that indicate changes in mRNA processing can be used for classification of lymphoma subtype, as shown below. An alternative analysis with a 4-way Venn diagram was performed to assess the overlap in sets of genes with statistically significant alternatively processing within each tumor or mature B-cells. Comparison against a null model that assumes independence of each gene in each sample revealed that all tumor subtypes and mature B-cells had significantly more genes unique to the sample type than expected (Figure S4).

Expression of additional 3'-processing trans-factors are de-regulated in pro-B-cell lymphoma. Given the findings of changes in CSTF3 protein expression, we investigated the microarray-measured

changes in the transcripts of other known RNA 3'-processing trans-factors (9). Analysis of the microarray data revealed uniform up- or downregulation of a significant number of polyA factors (Figure S8). Detailed analysis of the probe-level data revealed putative evidence for alternative mRNA processing in several genes, though not all passing our threshold of |t-value| > 6.0 and |ratio difference| > 1.5 (Figure S9). In several of the polyadenylation factors, the changes in expression are distinct from the changes observed in the natural progression to mature B-cells, suggesting tumor-specific polyadenylation activity.

In the absence of alternative processing, genes with long 3'-UTRs are preferentially downregulated in lymphoma samples. Genes that are not alternatively processed can be subject to changes in expression, possibly mediated by the same mechanisms that change isoform distribution. A search for systematic differences in these genes revealed that the 3′-UTR length distribution of genes with increased regulation (compared to pro-B-cells) is significantly shifted to shorter lengths when compared with genes with decreased regulation for all tumor types (Figure S10). In contrast, there was no significant difference in the 3′-UTR length distribution for genes with increased or decreased expression in mature B-cells (Figure S10). Differences in sequence content, specifically focusing on hexamers miRNA seed regions, were measured in the same manner as described above, and were dominated by the reduced length of the genes with increased expression (data not shown). The finding of a reduced number of putative miRNA target sites in genes with increased expression (but no alternative processing) is consistent with the findings of another recent study that compared proliferating and non-proliferating cells (11).

SUPPLEMENTAL DISCUSSION

Differences in methodology with Sandberg et al (11). A recent study, utilizing similar microarray analysis, reported that proliferating cells express mRNAs with shorter 3′-UTRs in a variety of cell types (11). Aside from small differences in array normalization and filtering, the mathematical approaches are essentially equivalent, using a modified *t*-test (J. Salisbury, *unpublished*) of logarithmic expression ratios on either side of a putative segmentation point. Nonetheless, several differences in methodology are relevant. The current study compares malignant proliferating cells with normal proliferating cells rather than proliferating and senescent cells, as in the prior study. In addition, our analysis searches de novo for segmentation points in the microarray probesets, whereas the prior study focused on sites with existing EST or mRNA evidence of a polyadenylation site. The limitation to known polyadenylation sites limits the analysis of both rare polyadenylation sites, as well as transcription initiation within the 3'-UTR (12). A final critical difference is that Sandberg *et al.* (2008) used the Affymetrix mouse exon ST 1.0 microarray, which specifically uses cDNA generated from random primers. Our analysis, In contrast, is based on the Affymetrix mouse 430v2 microarray, which uses cDNA generated from oligo-dT primers. MicroRNA-mediated cleavage of some genes results in stabilization of the 5′-end of the cleaved mRNA without a polyA tail (13). Random primed microarray data can capture these miRNA-mediated cleavages and subsequently indicate elevated truncation of processed transcripts (J. Salisbury, *unpublished*). Oligo-dT primed

microarrays cannot detect these miRNA-truncated isoforms and are more applicable in assessing processing changes in isoforms ending with a polyA tail.

Benefits of comparison with mature B-cells. The common comparison of both mature B-cells and pro-B-cell derived lymphomas with pro-B-cells revealed overlap in the genes with changes in mRNA processing, providing a possibility for distinguishing the changes that are specific to the tumorigenesis process, rather than those arising from the intended progression of the initiating cell type. Previous studies of B-cell maturation have demonstrated a role for programmed changes in polyadenylation for a small number of genes (10). Our analysis extends these studies, revealing system-wide mRNA processing changes during B-cell maturation and also in lymphoma tumorigenesis, albeit on a distinct set of target genes, potentially confirming changes in the specific molecular mechanisms involved.

The microarray analysis of *CstF3* revealed a common notable shift to the truncated, non-functional transcript, yet the Western blot analysis revealed a discrepancy between the mature B-cells and the lymphoma subtypes (Figure 4). Nearly all lymphoma samples showed equivalent or increased CSTF3 expression, while mature B-cells showed a significant reduction in CSTF3 expression.

AUTHOR CONTRIBUTIONS

The study was conceived and designed by JHG, PS, and KDM. The manuscript was written by PS, JHG and KDM. Computational analysis and associated programming were performed my PS and JHG. Western blots were performed by RYW. Flow Cytometry was performed by WS. RNA extraction and microarrays were prepared by SK. Mouse work and tumor collection were performed by TLA and SMW.

SUPPLEMENTAL FIGURES

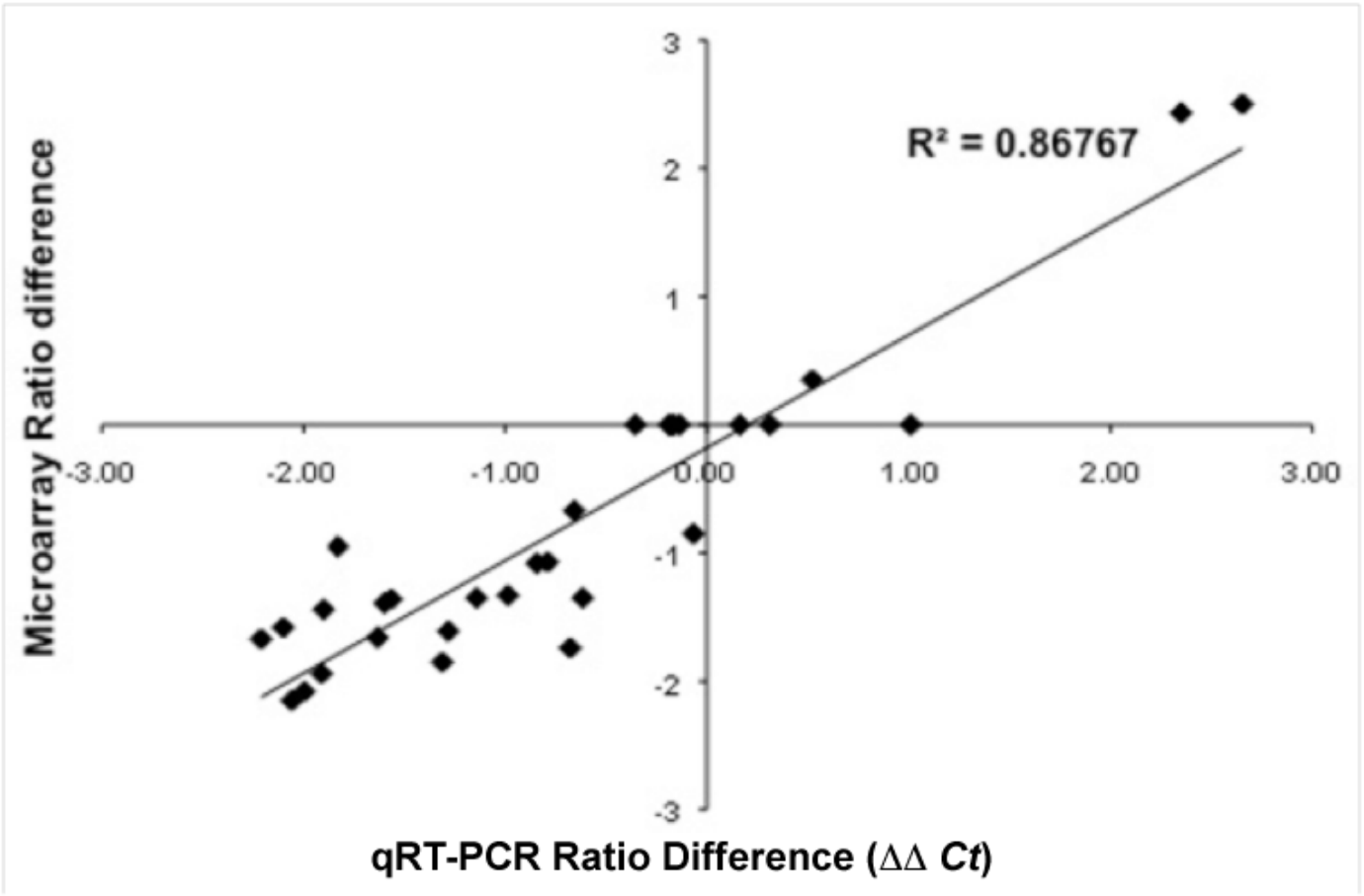

Figure S1. Quantitative RT-PCR validates the microarray-based changes in processing. The microarray "Ratio difference" is the difference between the average probe-level change on either side of the segmentation point (shown graphically in Figure 2). An equivalent measure was determined from pairs of qPCR primers generating products that flank the putative segmentation point.

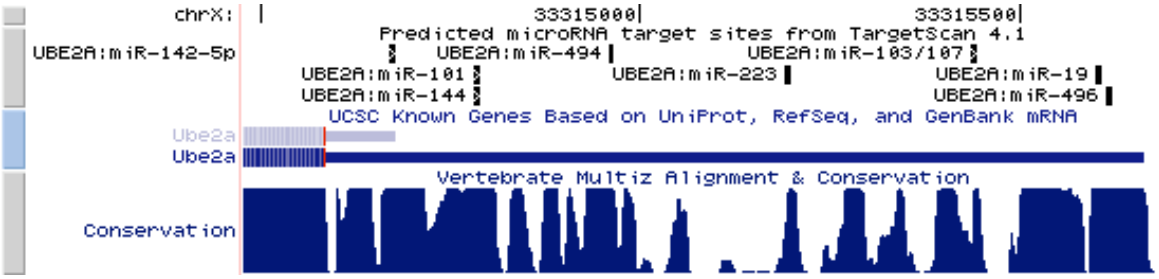

Figure S2. The extended 3′-UTR of Ube2a contains several putative miRNA binding sites. The seed regions of miRNA-101, miRNA-19 and miRNA-103/107 are conserved in human, mouse, rat and dog. These miRNA-binding sites could potentially be involved in the translational repression of *Ube2a* in normal pro-B-cells. The truncation of *Ube2a* isoform putatively de-represses the *Ube2a* transcript resulting in elevated UBE2A protein levels in the lymphomas.

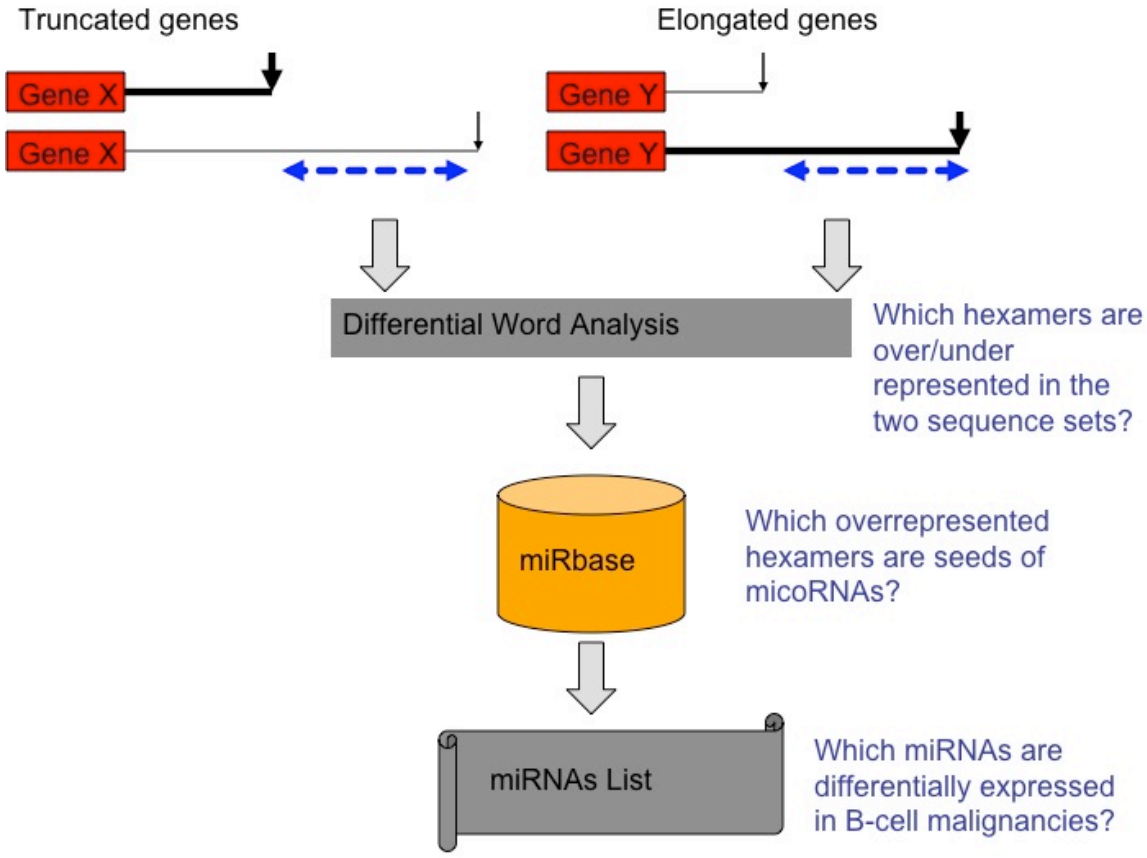

Figure S3. Flowchart of miRNA search methodology. The differential portion of sequences that are preferentially elongated was compared with the differential portion of sequences that are preferentially truncated to search for patterns present in one but absent in the other. The presence of such sites would support a model where differential expression of isoforms was driven by trans-factors acting on differentially included sequence rather than direct changes in the formation of the transcript (such as alternative polyadenylation). Hexamers found to be differential at a statistically significantly level were intersected with seed sequences of known miRNAs to identify miRNAs involved in tumorigenesis.

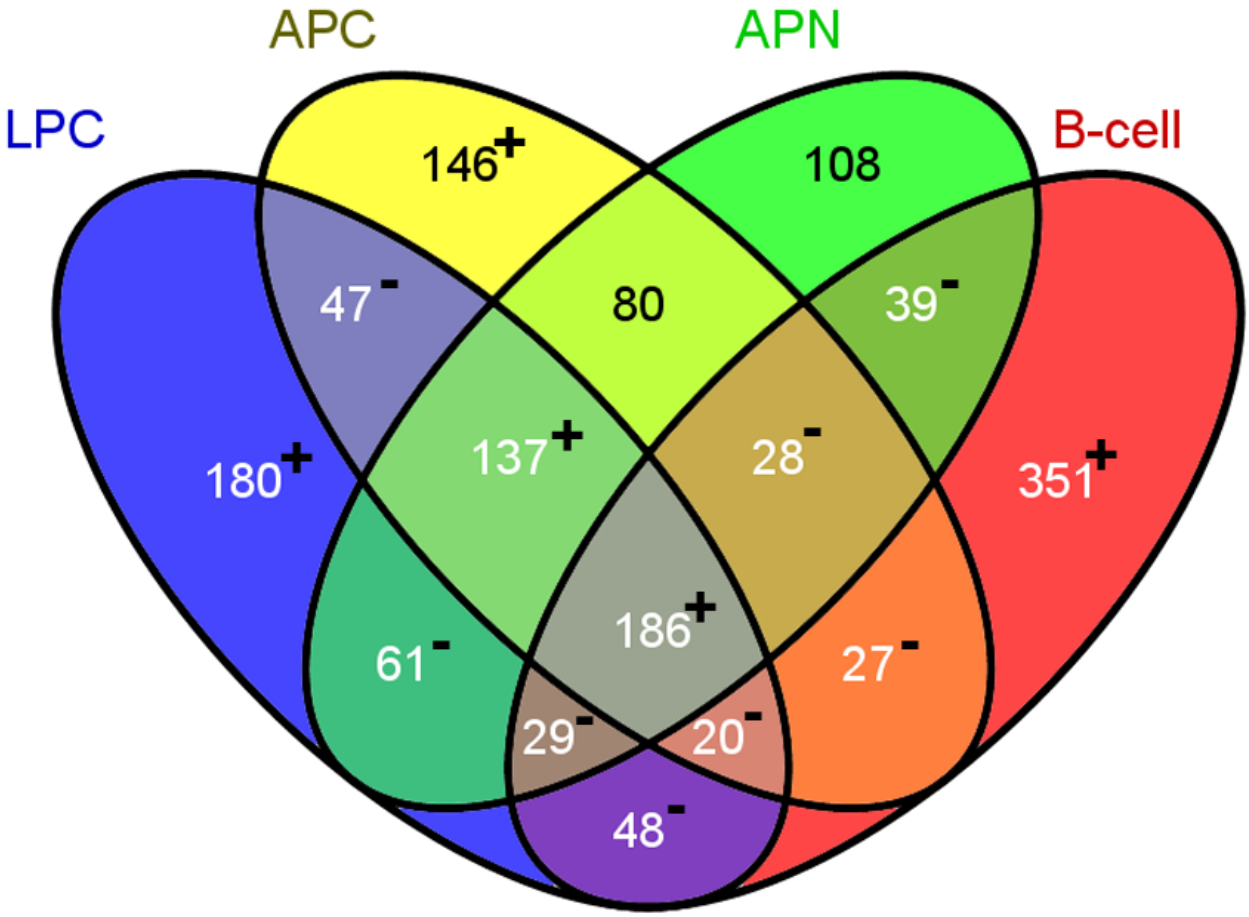

Figure S4. Lymphoma-subtype-specific alternative processing events occur more frequently than expected and provide diagnostic signatures. A 4-way Venn diagram reveals the number of genes with evidence of alternative processing (when compared to the pro-B-cell reference) shared between all examined cell types. Fields with counts marked '**+**' and '**-**' indicate counts that were significantly larger, or smaller than expected, respectively, based on a simulation that assigns alternative processing in each sample via independent draws of a pseudorandom variables according to the observed frequency of events in each sample.

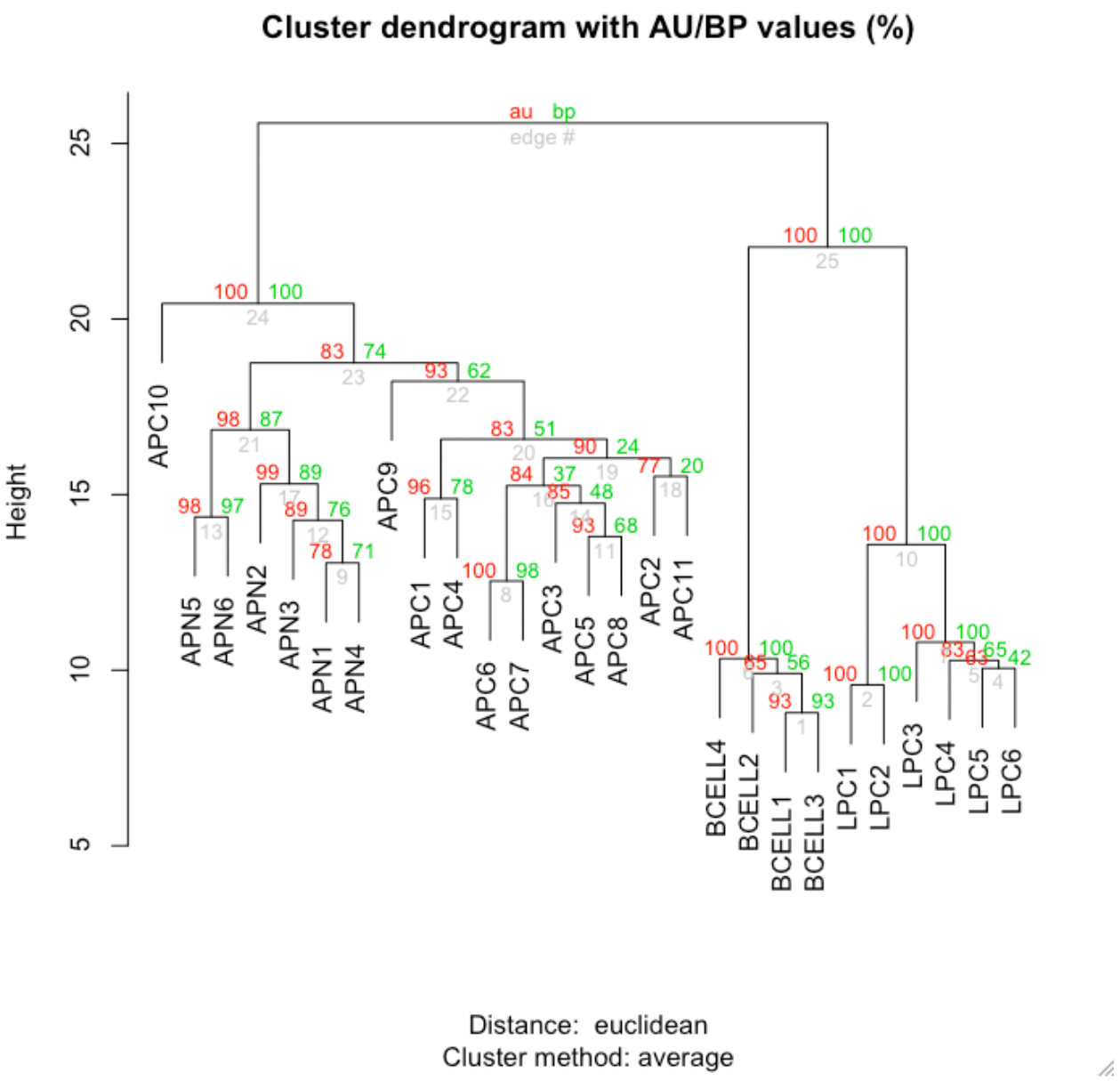

Figure S5. Bootstrap tree for the mouse B-cell lymphomas, generated with the R package, pvclust (14). This tree was generated based on the gene/probe combinations indicating segmentations that are significantly different among the tumor types. Green numbers represent bootstrapping probabilities for the inclusion of the indicated cluster in the true tree, red numbers represent "approximately unbiased" **(14)** probabilities of inclusion of the indicated cluster.

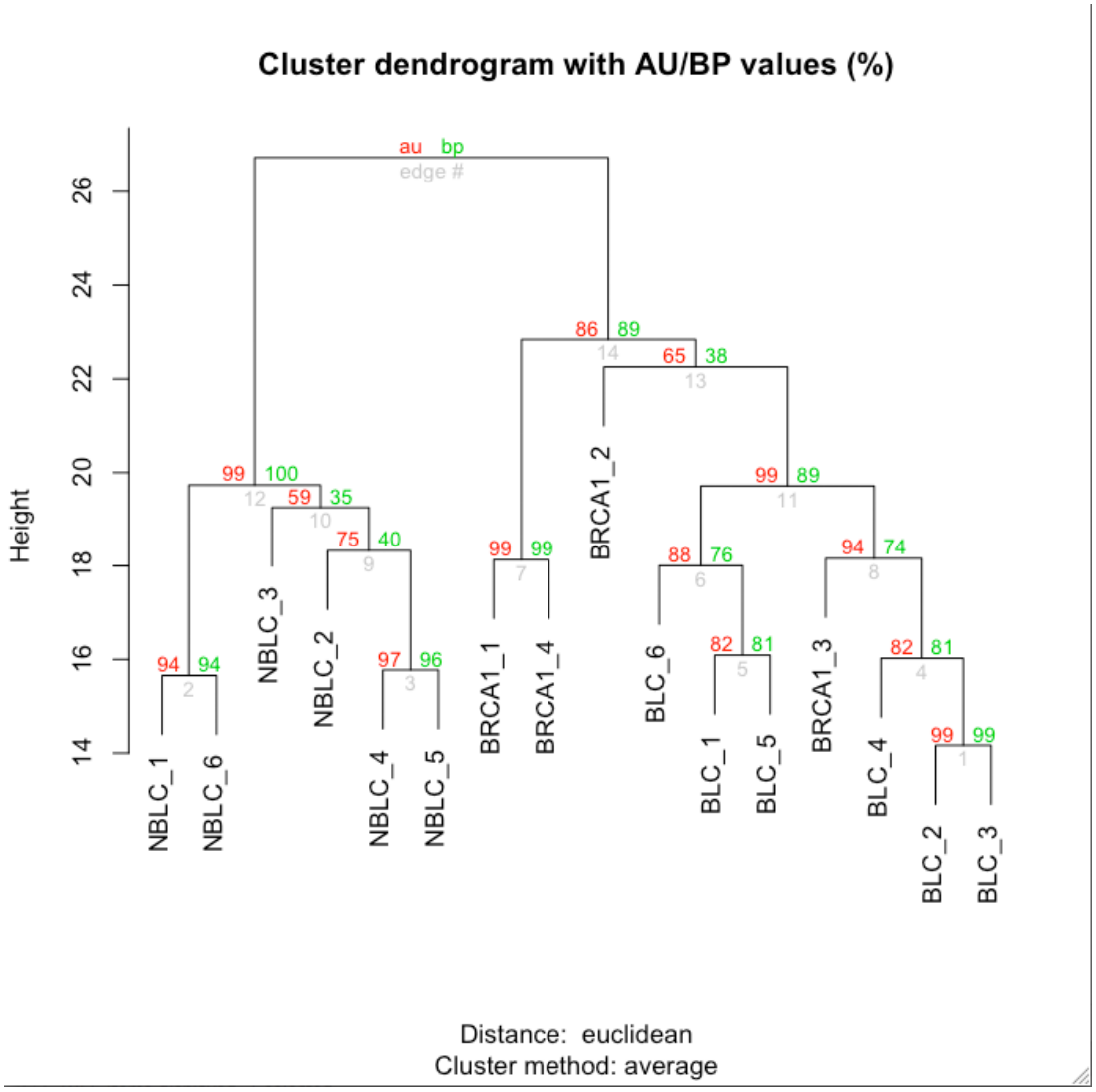

Figure S6. Bootstrap tree for the human breast cancer samples, generated with the R package, pvclust (14). This tree was generated based on the gene/probe combinations indicating segmentations that are significantly different among the tumor types. Green numbers represent bootstrapping probabilities for the inclusion of the indicated cluster in the true tree, red numbers represent “approximately unbiased” **(14)** probabilities of inclusion of the indicated cluster.

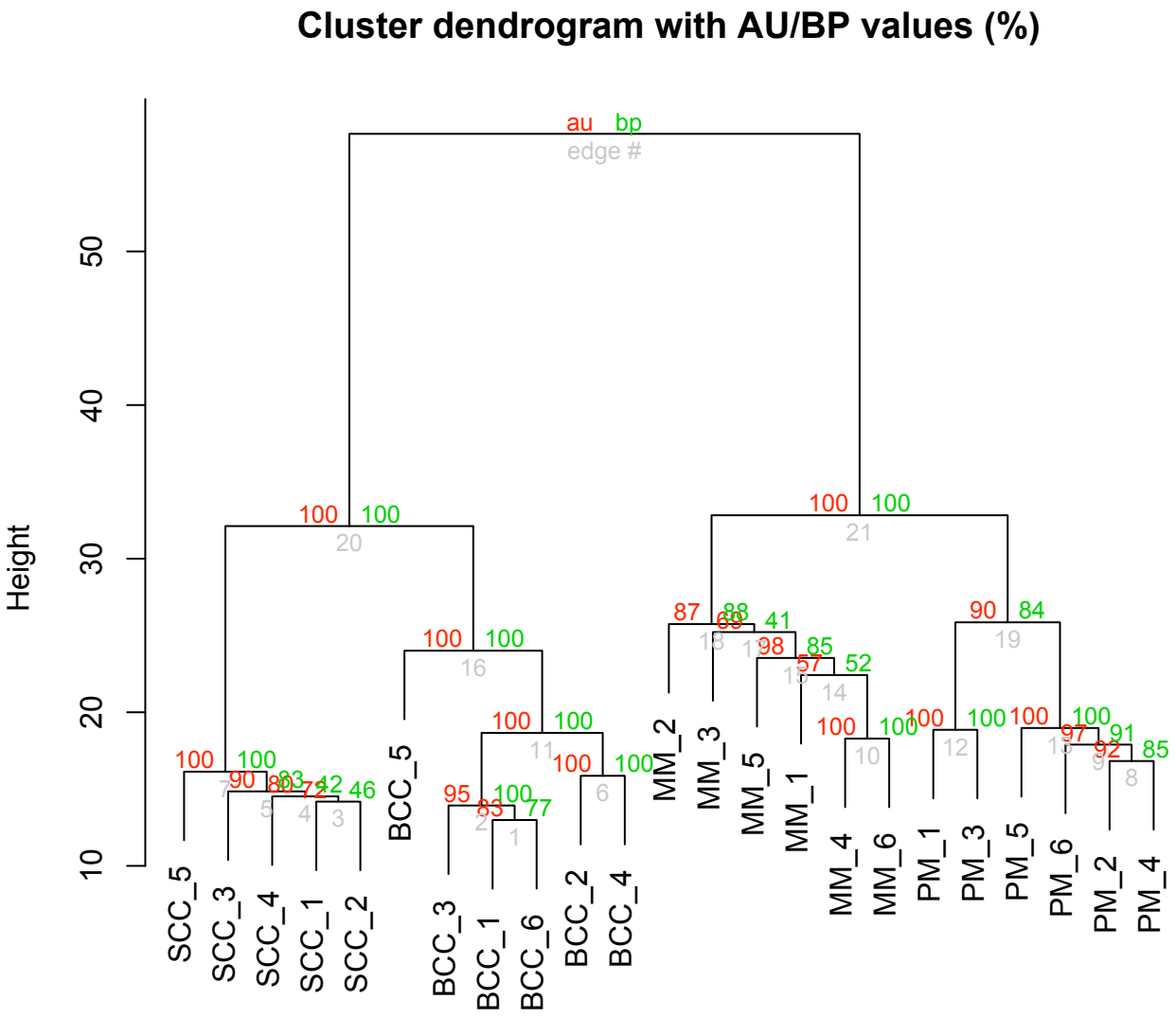

Figure S7. Bootstrap tree for the human skin cancer samples, generated with the R package, pvclust (14). This tree was generated based on the gene/probe combinations indicating segmentations that are significantly different among the tumor types. Green numbers represent bootstrapping probabilities for the inclusion of the indicated cluster in the true tree, red numbers represent "approximately unbiased" **(14)** probabilities of inclusion of the indicated cluster.

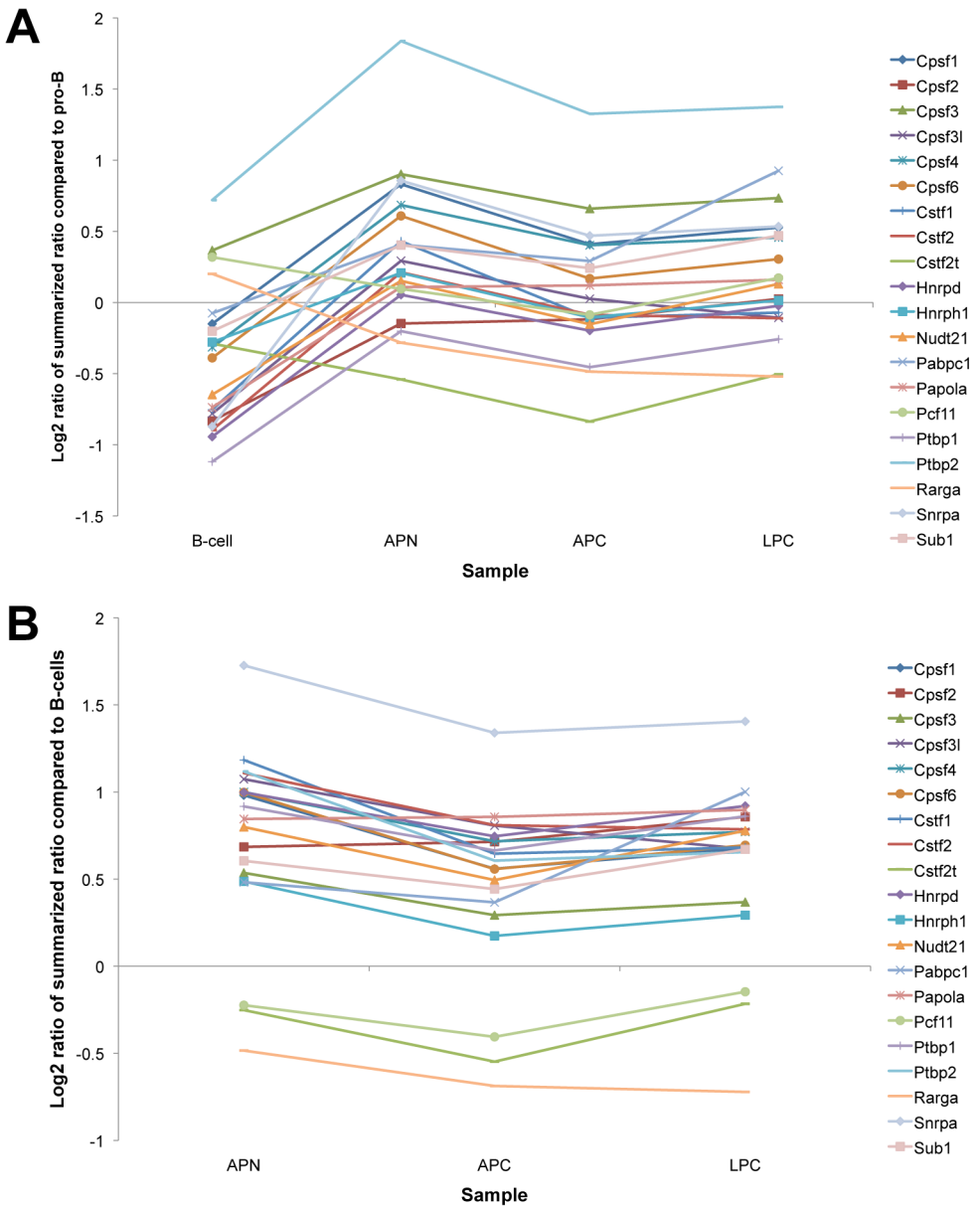

Figure S8. Analysis of microarray data shows that polyA trans-factors without definitive evidence of alternative processing changes are expressed at different levels in lymphoma samples compared to either pro-B-cells or mature B-cells. Summarized expression values area shown for each factor in each sample, given as a log2-ratio compared to either (A) pro-B-cells, or (B) mature B-cells. Changes in these factors support a model of alternative polyadenylation site selection and processing as a result of lymphoma specific molecular processes.

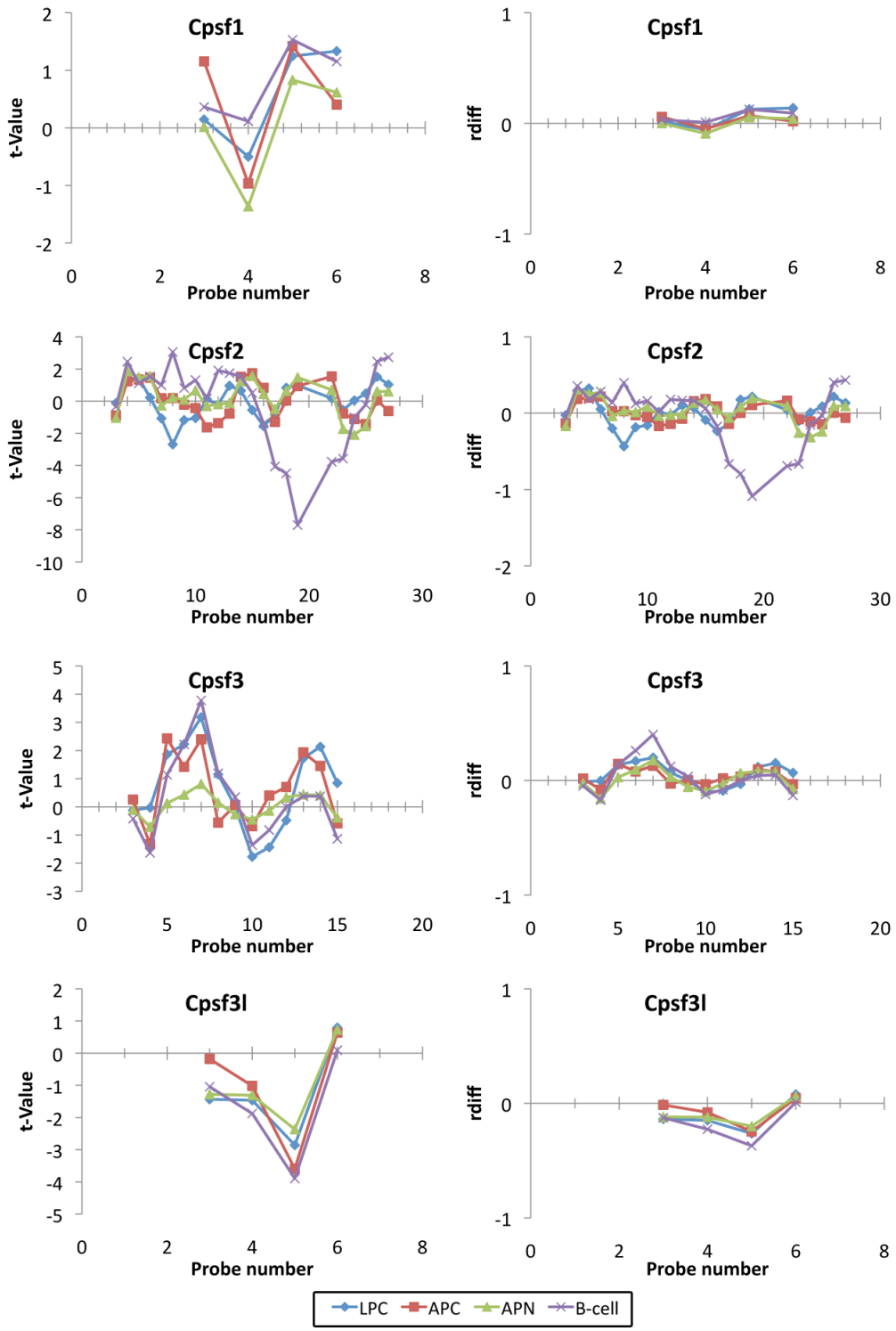

Figure S9. Segmentation analysis of polyA trans-factors shows some evidence of alternative processing changes. Each pair of plots represents the *t*-value and *rdiff* (ratio difference) for each possible segmentation in the probeset.. Putative segmentations are represented by the upstream probe number (x-axis). The requirement of at least three probes on each side of a segmentation leads to each plot beginning with probe number 3. Given the threshold of $|t| > 6$, $|rdiff| > 1.5$, a common elongation event is shared between all samples for *Cpsf6*, while a truncation in *Cpsf2* normally observed in mature B-cells is lost in all three tumor types. Other putative events that don't reach statistical significance at our thresholds can be seen for several other transcripts.

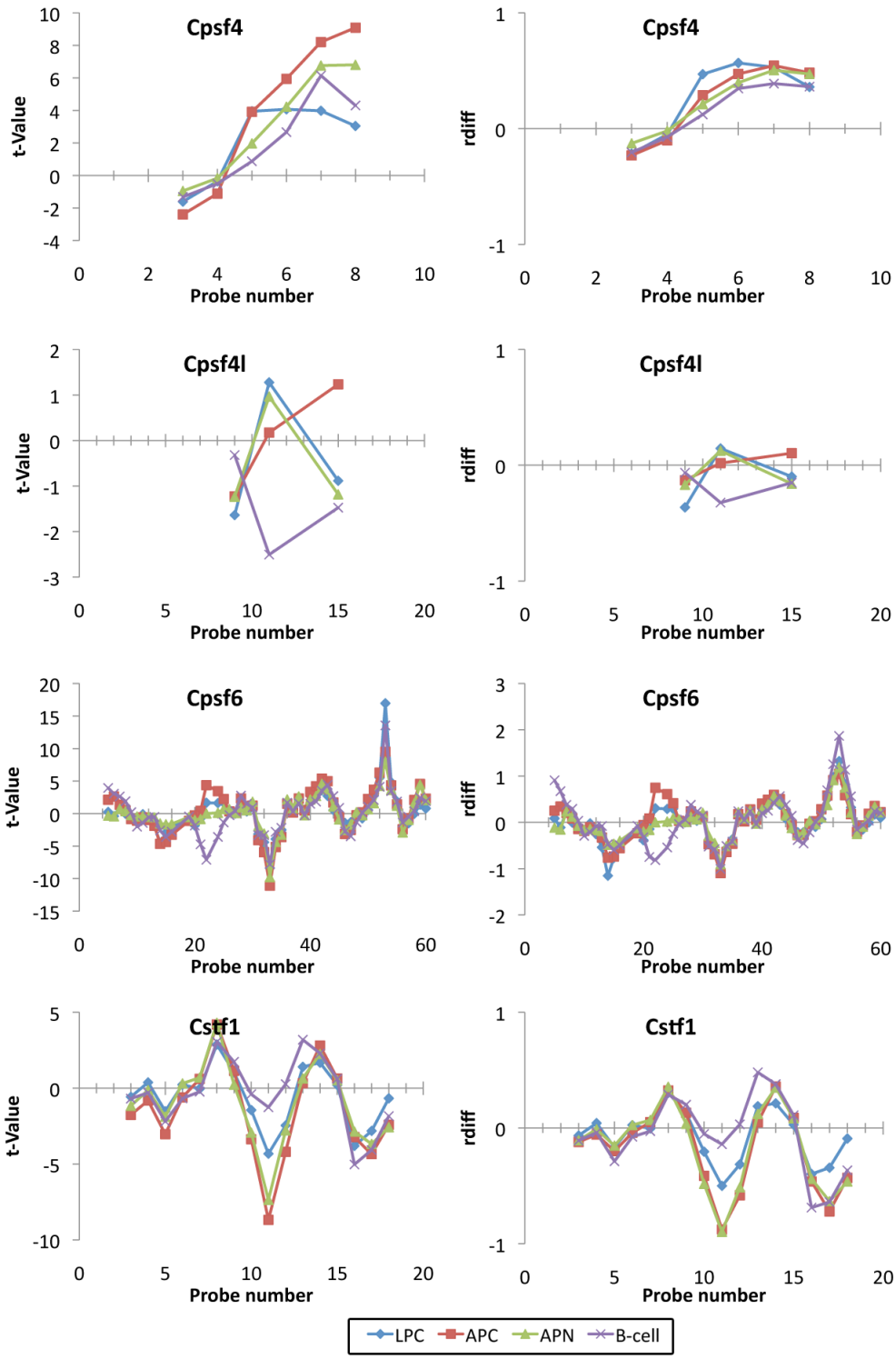

Figure S9 (continued).

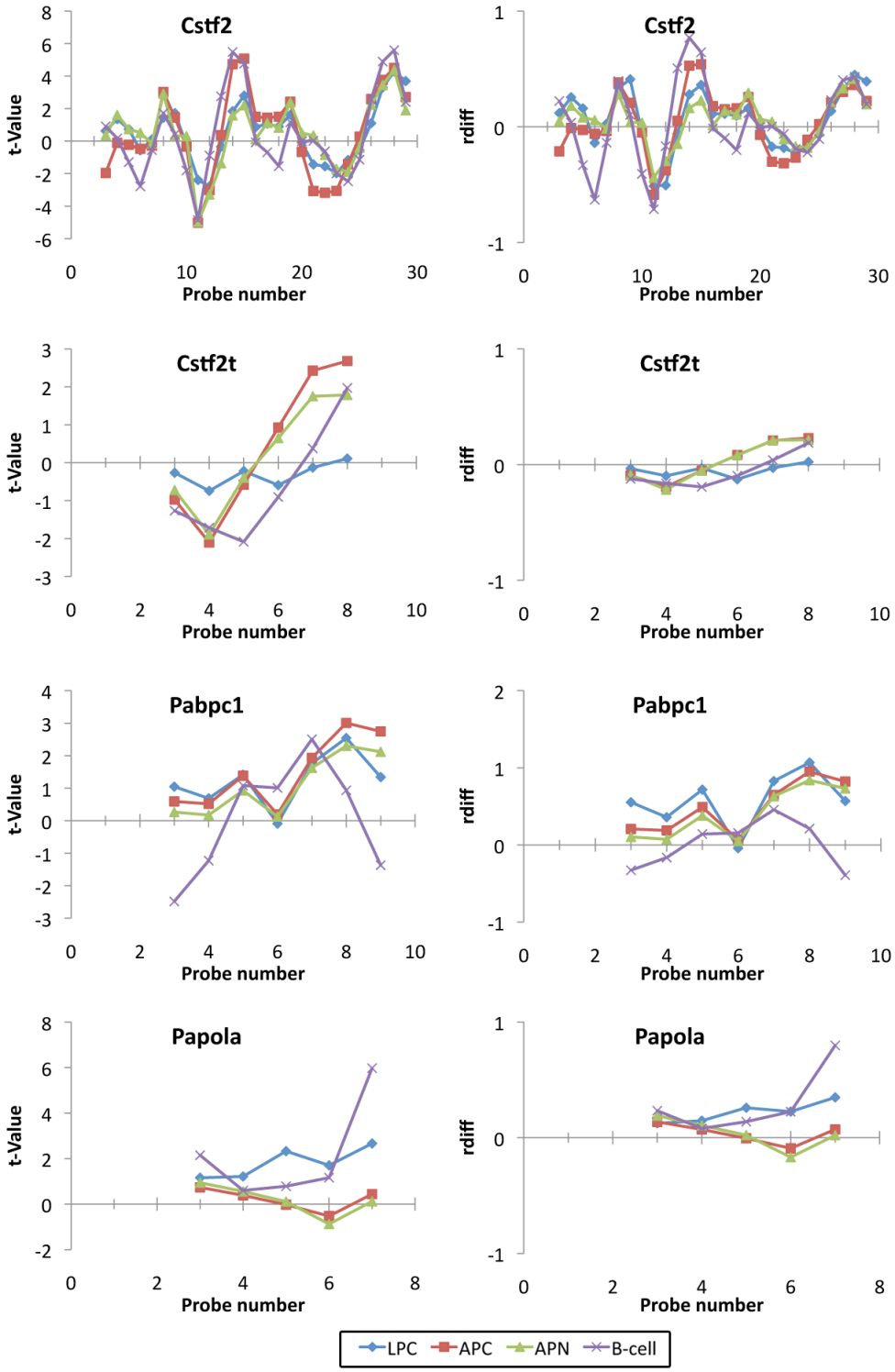

Figure S9 (continued).

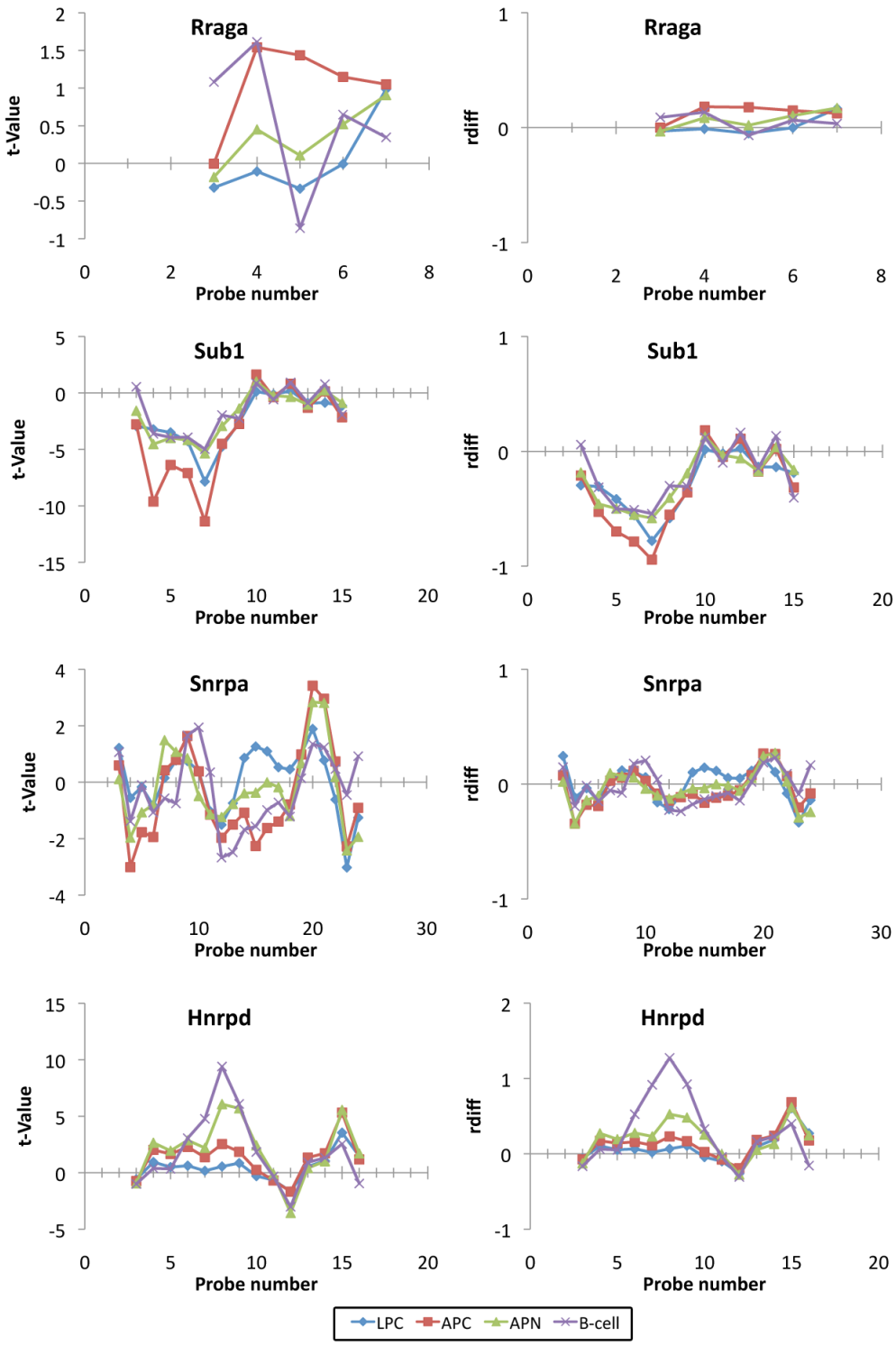

Figure S9 (continued).

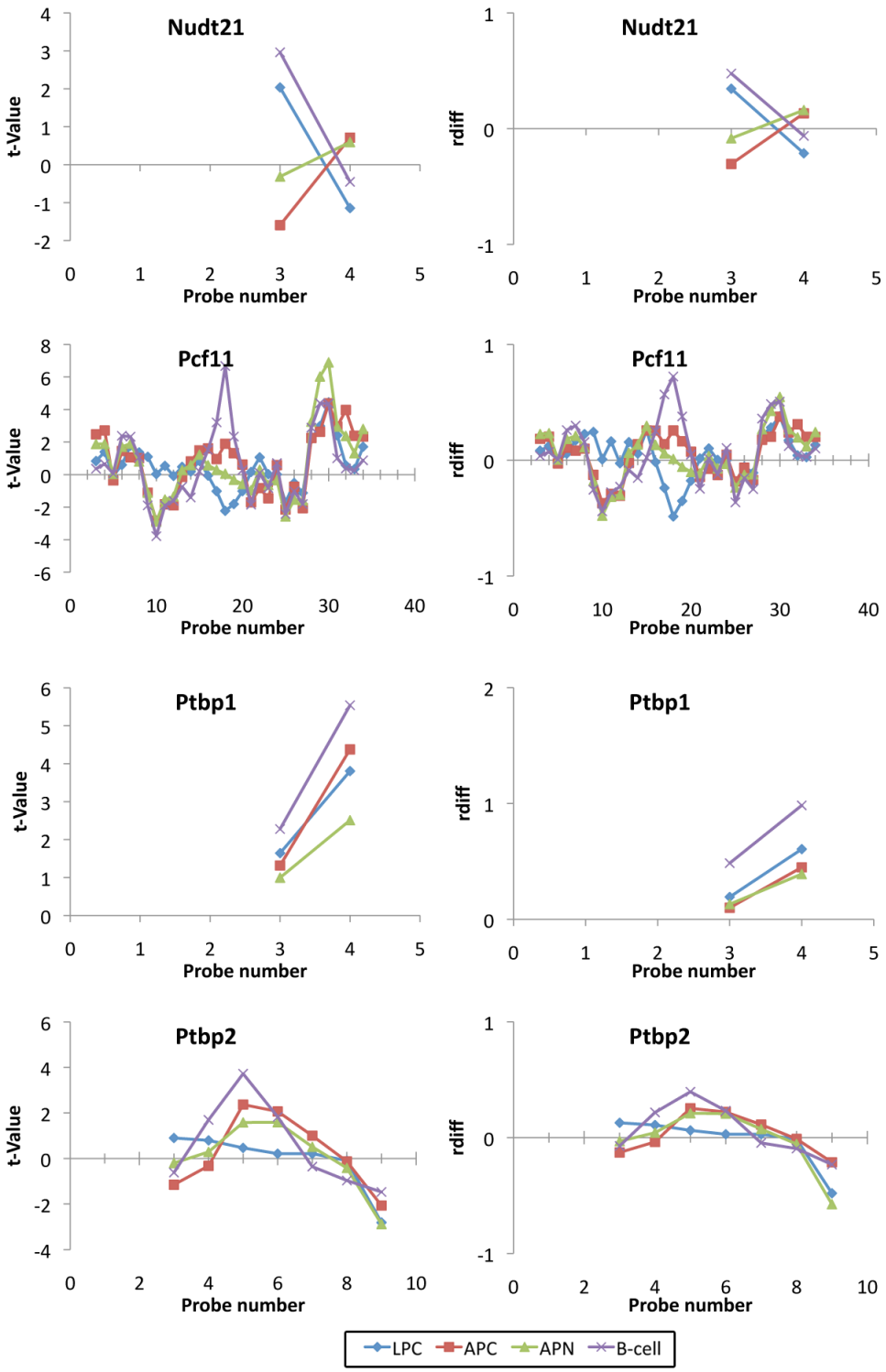

Figure S9 (continued).

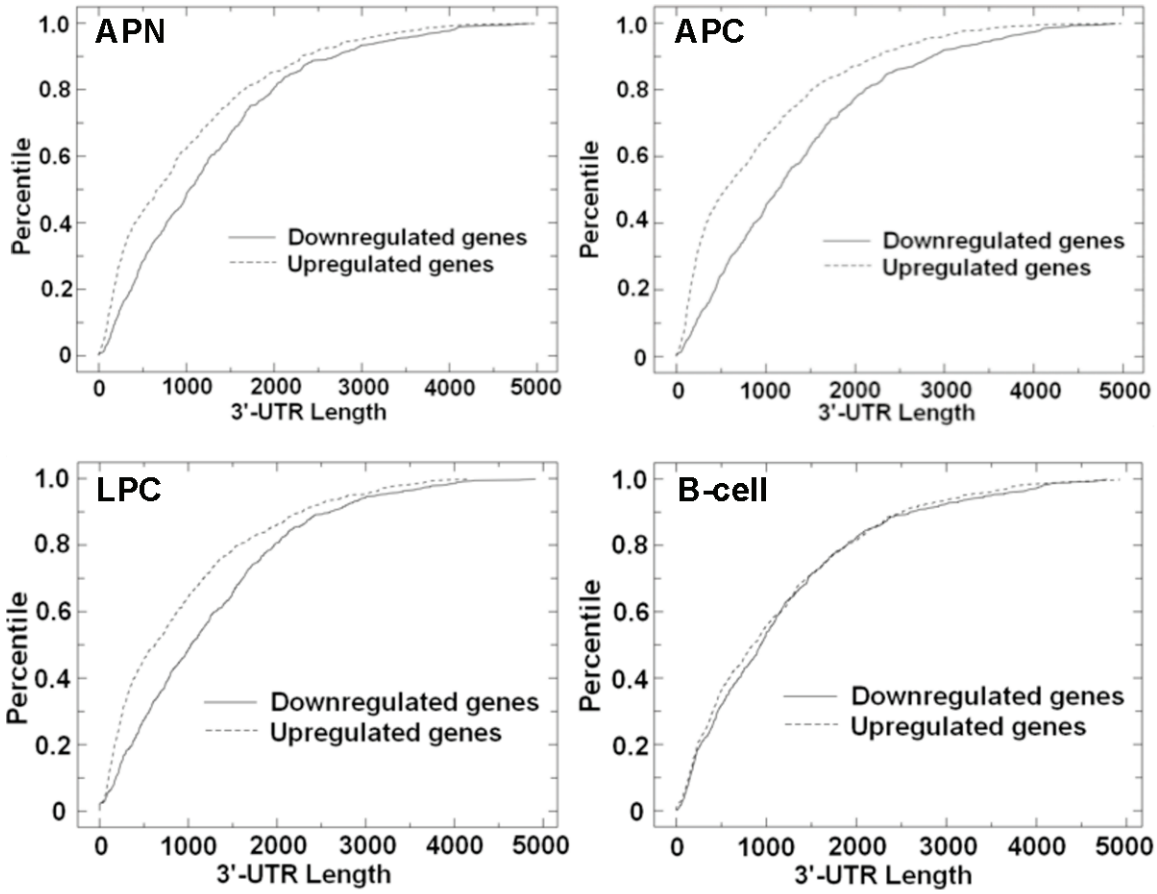

Figure S10. Genes with no evidence of alternative processing display a 3'-UTR bias in lymphoma samples only. The cumulative length distributions of the most significantly up- and downregulated genes (5% at each end of the distribution) are plotted for APN, APC, LPC, and mature B-cells. Putative 3'-UTR lengths were obtained from PACDB (2).

SUPPLEMENTAL TABLES

Table S1: A majority of probesets that have any processing events have only one. Events were limited with the conservative |*t*| ≥ 6.0 and |ratio difference| ≥ 1.5 criteria.

Count	LPC	APC	APN	B-cell
1	708	671	668	728
2	111	128	117	99
3	16	25	18	14
4	4	15	8	1
5	2	1	0	0

Table S2: Counts of probesets identified as differentially expressed (FDR < 0.05) in comparisons between APC, APN, and LPC tumors.

		Sample 2	
		LPC	APC
Sample 1	APC	56	
	APN	596	35

Table S3. Primers used in quantitative PCR validation of probeset validation.

Ube2a	Common_F	CGGGAATATGAAAAGCGTGT
	Common_R	TTATGGCCAGCCTCTTCTTC
	Extended_F	AATGCTGCATGCTTTCAGTG
	Extended_R	TTACAAGCCTTGCCCCATAC
Cstf3	Common_F	CATAACATCCTGTATGCACACTCAAG
	Common_R	ATCATAATTGCCACACCATGAAAT
	Extended_F	AGCCTGTGTGGATTTTGAATTTG
	Extended_R	GGAACAAGATGTGGCGTTGTC
Serbp1	Common_F	GATTCAGTTATGGACCATCATTTCC
	Common_R	AGTCTCCAAAATTGATCTCCAGTTG
	Extended_F	AGACTTGAGCTGGTGGCTTACAT
	Extended_R	CTAATCCTTGCCTCTGAAAAGCA
Sfrs7	Common_F	TCGTTCCCCATCAGGAAGTC
	Common_R	TCCCTAACGGGTGAACTTGAGA
	Extended_F	CTGTTGCACTTGGTTAGCTTGGT
	Extended_R	TAACACTTTGCTTGACATCTCTTTATACC
Pik3ap1	Common_F	GTGTAGCCCTGCTGGTTTGG
	Common_R	TCGGCAACGTTGAACAAAGA
	Extended_F	TGTCCCTCAACCCTCTTGTCA
	Extended_R	ATGGTCAAGAACAACCACACTGA
Sf3b1	Common_F	TCGGAAAGTCAGAGACGTATATTGG
	Common_R	TAGATTCGTGGGTAATGTGCTATGA
	Extended_F	CACAGGAGACTGCTTACCCCTTA
	Extended_R	GATTGCAGCCTTCAGCATTTT

Table S4. Comparison of the ratio difference score as measured by microarray data or quantitative RT-PCR. This data is plotted graphically in the Figure S1.

Gene	Sample	Microarray Ratio Difference (Mean)	qRT-PCR Ratio Difference (Mean)
Cstf3	APN3	-1.67	-2.21
	APC1	-0.95	-1.83
	APC4	-1.58	-2.1
	LPC1	-1.94	-1.91
	LPC4	-2.08	-2.0
Ube2a	APN3	-0.67	-2.21
	APC1	-1.66	-1.63
	APC4	-1.36	-1.56
	LPC1	-1.33	-1.0
	LPC4	-1.44	-1.9
Pik3ap1	APN3	2.5	2.65
	APC1	2.43	2.35
	APC4	0.35	0.52
	LPC1	0	0.16
	LPC4	0	-0.13
Serbp1	APN3	0	-0.35
	APC1	-1.07	-0.8
	APC4	-1.61	-1.29
	LPC1	-1.08	-0.84
	LPC4	-1.39	-1.6
Sf3b1	APN3	0	0.31
	APC1	0	-0.19
	APC4	0	-0.16
	LPC1	-0.85	-0.07
	LPC4	-1.74	-0.7
Sfrs7	APN3	0	1.0
	APC1	-1.85	-1.31
	APC4	-1.35	-1.14
	LPC1	-1.35	-0.62
	LPC4	-2.15	-2.06

Table S5. Biological processes over-represented in differentially processed genes that show truncation compared to pro-B-cells. The set of differentially processed genes in each of the four comparisons respectively were analyzed using GOstat (3) to determine over-representation of gene ontology categories. Complete GOstat output for each run is available at http://harlequin.jax.org/cancerAPA/.

Sample	GO term	Description	count	all	Pvalue
LPC	GO:0065007	biological regulation	150	3853	4.48E-09
	GO:0050789	regulation of biological process	139	3497	4.48E-09
	GO:0050794	regulation of cellular process	128	3140	4.48E-09
	GO:0043412	biopolymer modification	74	1490	4.63E-09
	GO:0043283	biopolymer metabolic process	153	4055	1.45E-08
	GO:0006464	protein modification process	70	1435	3.68E-08
	GO:0043170	macromolecule metabolic process	184	5327	2.32E-07
	GO:0007242	intracellular signaling cascade	51	985	9.43E-07
	GO:0006512	ubiquitin cycle	29	437	1.13E-06
	GO:0043687	post-translational protein modification	60	1246	1.13E-06
APC	GO:0065007	biological regulation	133	3853	3.65E-08
	GO:0050790	regulation of catalytic activity	21	267	6.53E-08
	GO:0065009	regulation of a molecular function	22	310	6.19E-07
	GO:0032502	developmental process	100	2804	9.24E-07
	GO:0050789	regulation of biological process	118	3497	9.24E-07
	GO:0050794	regulation of cellular process	108	3140	1.41E-06
	GO:0007242	intracellular signaling cascade	46	985	1.65E-06
	GO:0048869	cellular developmental process	66	1677	5.28E-06
	GO:0030154	cell differentiation	66	1677	5.28E-06
	GO:0043283	biopolymer metabolic process	125	4055	6.82E-05
APN	GO:0065007	biological regulation	155	3853	3.42E-14
	GO:0050789	regulation of biological process	143	3497	9.97E-14
	GO:0050794	regulation of cellular process	132	3140	1.04E-13
	GO:0007242	intracellular signaling cascade	57	985	5.01E-12
	GO:0050790	regulation of catalytic activity	23	267	1.14E-09
	GO:0065009	regulation of a molecular function	25	310	1.73E-09
	GO:0043283	biopolymer metabolic process	146	4055	4.50E-09
	GO:0008283	cell proliferation	30	462	1.31E-07
	GO:0009966	regulation of signal transduction	24	331	1.84E-07
	GO:0032502	developmental process	106	2804	2.50E-07
B-cell	GO:0043170	macromolecule metabolic process	193	5327	6.69E-07
	GO:0043412	biopolymer modification	73	1490	6.69E-07
	GO:0043283	biopolymer metabolic process	155	4055	6.69E-07
	GO:0044238	primary metabolic process	214	6147	7.41E-07
	GO:0065007	biological regulation	148	3853	8.03E-07
	GO:0050789	regulation of biological process	137	3497	8.15E-07
	GO:0050794	regulation of cellular process	123	3140	5.79E-06
	GO:0006464	protein modification process	67	1435	1.05E-05
	GO:0044237	cellular metabolic process	208	6170	1.40E-05
	GO:0019538	protein metabolic process	106	2714	6.68E-05

Table S6. Biological processes over-represented in differentially processed genes that show elongation compared to pro-B-cells. The set of differentially processed genes in each of the four comparisons respectively were analyzed using GOstat (3) to determine over-representation of gene ontology categories. Complete GOstat output for each run is available at http://harlequin.jax.org/cancerAPA/.

Sample	GO term	Description	count	all	P-value
LPC	GO:0022610	biological adhesion	19	588	5.31E-06
	GO:0007155	cell adhesion	19	588	5.31E-06
	GO:0009653	anatomical structure morphogenesis	26	1109	2.66E-04
	GO:0048856	anatomical structure development	36	1904	1.51E-04
	GO:0048514	blood vessel morphogenesis	9	173	5.51E-03
	GO:0007167	enzyme linked receptor protein signaling pathway	11	266	5.51E-03
	GO:0009887	organ morphogenesis	15	498	7.76E-03
	GO:0048731	system development	30	1633	1.10E-02
	GO:0001568	blood vessel development	9	203	1.10E-02
	GO:0001944	vasculature development	9	206	1.11E-02
APC	GO:0048856	anatomical structure development	49	1904	1.72E-08
	GO:0048731	system development	43	1633	7.28E-08
	GO:0009653	anatomical structure morphogenesis	33	1109	9.58E-08
	GO:0007275	multicellular organismal development	49	2131	1.12E-06
	GO:0022610	biological adhesion	20	588	8.02E-06
	GO:0007155	cell adhesion	20	588	8.02E-06
	GO:0032502	developmental process	56	2804	2.80E-05
	GO:0009887	organ morphogenesis	17	498	6.95E-05
	GO:0048513	organ development	32	1323	1.05E-04
	GO:0007167	enzyme linked receptor protein signaling pathway	13	266	4.91E-04
APN	GO:0002376	immune system process	17	718	2.52E-02
B-cell	GO:0007049	cell cycle	20	697	2.57E-05
	GO:0043283	biopolymer metabolic process	61	4055	8.32E-04
	GO:0043170	macromolecule metabolic process	72	5327	4.37E-04
	GO:0022402	cell cycle process	16	552	6.41E-03
	GO:0044238	primary metabolic process	78	6147	1.24E-02
	GO:0044237	cellular metabolic process	78	6170	1.24E-02
	GO:0043412	biopolymer modification	27	1490	1.24E-02
	GO:0022403	cell cycle phase	10	270	1.27E-02
	GO:0050789	regulation of biological process	50	3497	1.27E-02
	GO:0000278	mitotic cell cycle	9	224	1.32E-02

Table S7: Overrepresented GO terms in genes with truncated transcripts in at least two types of B-cell lymphoma, but not in mature B-cells. Complete GOstat output is available at http://harlequin.jax.org/cancerAPA/.

GO term	Description	count	all	P-value
GO:0050794	regulation of cellular process	56	3140	2.27E-05
GO:0050789	regulation of biological process	60	3497	2.27E-05
GO:0043283	biopolymer metabolic process	65	4055	8.19E-05
GO:0010468	regulation of gene expression	39	2001	9.41E-05
GO:0065007	biological regulation	62	3853	9.41E-05
GO:0044237	cellular metabolic process	87	6170	1.08E-04
GO:0044238	primary metabolic process	85	6147	3.93E-04
GO:0043170	macromolecule metabolic process	76	5327	4.86E-04
GO:0031323	regulation of cellular metabolic process	38	2083	5.95E-04
GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	48	2928	9.44E-04
GO:0019222	regulation of metabolic process	38	2160	1.47E-03
GO:0010467	gene expression	44	2653	1.53E-03
GO:0007242	intracellular signaling cascade	21	985	4.89E-03
GO:0045449	regulation of transcription	33	1874	4.93E-03
GO:0032502	developmental process	44	2804	6.54E-03
GO:0019219	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	33	1911	6.94E-03
GO:0007399	nervous system development	15	618	6.94E-03
GO:0006350	transcription	33	1936	8.63E-03
GO:0001947	heart looping	3	14	0.011
GO:0007275	multicellular organismal development	35	2131	0.012
GO:0006793	phosphorus metabolic process	17	782	0.012
GO:0042541	hemoglobin biosynthetic process	2	4	0.018
GO:0006355	regulation of transcription, DNA-dependent	30	1797	0.022
GO:0016070	RNA metabolic process	35	2217	0.024
GO:0043412	biopolymer modification	26	1490	0.024
GO:0020027	hemoglobin metabolic process	2	5	0.025
GO:0006351	transcription, DNA-dependent	30	1819	0.025
GO:0032774	RNA biosynthetic process	30	1823	0.025
GO:0006464	protein modification process	25	1435	0.027
GO:0048869	cellular developmental process	28	1677	0.027
GO:0030154	cell differentiation	28	1677	0.027
GO:0042035	regulation of cytokine biosynthetic process	4	50	0.028
GO:0008354	germ cell migration	2	6	0.029
GO:0060017	parathyroid gland development	2	6	0.029
GO:0006417	regulation of translation	6	131	0.030
GO:0048522	positive regulation of cellular process	15	719	0.034
GO:0031326	regulation of cellular biosynthetic process	6	139	0.037
GO:0006468	protein amino acid phosphorylation	13	592	0.039
GO:0016310	phosphorylation	14	661	0.039
GO:0042089	cytokine biosynthetic process	4	60	0.042
GO:0045086	positive regulation of interleukin-2 biosynthetic process	2	8	0.042
GO:0042981	regulation of apoptosis	10	373	0.042
GO:0042107	cytokine metabolic process	4	61	0.042
GO:0051251	positive regulation of lymphocyte activation	4	61	0.042
GO:0043066	negative regulation of apoptosis	6	149	0.043
GO:0043067	regulation of programmed cell death	10	377	0.043
GO:0043069	negative regulation of programmed cell death	6	151	0.045

Table S8: Overrepresented GO terms in genes with elongated transcripts in at least two types of B-cell lymphoma, but not in mature B-cells. Complete GOstat output is available at http://harlequin.jax.org/cancerAPA/.

GO term	Description	count	all	P-value
GO:0048856	anatomical structure development	27	1904	4.32E-04
GO:0022610	biological adhesion	15	588	4.32E-04
GO:0007155	cell adhesion	15	588	4.32E-04
GO:0007167	enzyme linked receptor protein signaling pathway	10	266	6.78E-04
GO:0009653	anatomical structure morphogenesis	18	1109	1.33E-03
GO:0048731	system development	23	1633	1.59E-03
GO:0032502	developmental process	32	2804	5.99E-03
GO:0048513	organ development	19	1323	5.99E-03
GO:0002376	immune system process	14	718	6.38E-03
GO:0007275	multicellular organismal development	26	2131	7.19E-03
GO:0002449	lymphocyte mediated immunity	5	92	0.013
GO:0002443	leukocyte mediated immunity	5	99	0.017
GO:0051093	negative regulation of developmental process	5	104	0.020
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	6	163	0.020
GO:0048534	hemopoietic or lymphoid organ development	7	232	0.021
GO:0006955	immune response	10	477	0.022
GO:0048514	blood vessel morphogenesis	6	173	0.023
GO:0001570	vasculogenesis	3	29	0.024
GO:0006817	phosphate transport	4	67	0.024
GO:0002520	immune system development	7	249	0.024
GO:0009887	organ morphogenesis	10	498	0.024
GO:0048646	anatomical structure formation	6	186	0.026
GO:0002252	immune effector process	5	135	0.034
GO:0001525	angiogenesis	5	135	0.034
GO:0001568	blood vessel development	6	203	0.034
GO:0048869	cellular developmental process	20	1677	0.034
GO:0030154	cell differentiation	20	1677	0.034
GO:0001944	vasculature development	6	206	0.034
GO:0045596	negative regulation of cell differentiation	4	85	0.038
GO:0042269	regulation of natural killer cell mediated cytotoxicity	2	12	0.049

Table S9. A selection of hexamers that match known microRNAs occur at statistically significantly different rates in the comparison of Differentially Included Sequence (DIS) and Differentially Excluded Sequence (DES). DIS Fraction and DES Fraction show the fractions of sequences with at least one match to the hexamer in each set. P-Values were evaluated with a permutation analysis that randomizes the assignment of sequence to sequence set. Only selected hits are shown. The total results are dominated by the shorter sequences in the set of DIS compared to the DES sequences. Complete results are available at http://harlequin.jax.org/cancerAPA/.

Sample	Hexamer	DIS Fraction	DES fraction	P-val	miRNA	Known expression of miRNA
LPC	CUGUUG	59/240=0.25	174/427=0.41	<0.001	miR-421	Upregulated DLBCL (GCB-type cell line)
LPC	ACUGUU	65/240=0.27	171/427=0.40	<0.001	miR-132	Upregulated in PPBL, CLL, also in B-cell activation
APN	CAGUGU	77/249=0.31	181/448=0.40	0.05	miR-141	upregulated in CLL cells
APN	ACUUGA	50/249=0.20	131/448=0.29	0.007	miR-26a	Downregulated by *Myc*
APC	UGUGGU	75/326=0.23	166/476= 0.35	0.02	miR-220	downregulated in CLL cells
APC	CUGUUG	90/326=0.28	175/476= 0.37	0.005	miR-196b	
APC	CAGUGU	99/326=0.11	203/476= 0.43	<0.001	miR-141	upregulated in CLL cells

*Multiple occurrences of a hexamer was calculated as a single occurrence

Table S10. Diagnosis of incorrectly predicted subtypes of lymphomas. LPC samples were predicted correctly without fail, APN predictions were the weakest, and APC predictions were intermediate. A detailed breakdown on a sample-by-sample basis is available below as Table S7. Sn: Sensitivity; Sp: Specificity; TP: True Positive; FP: False Positive; TN: True Negative; FN: False Negative.

		True Subtype		
		LPC	APC	APN
Predicted Subtype	LPC	100%	0%	0%
	APC	0%	92.2%	25.8%
	APN	0%	7.8%	74.2%
Sn (TP/(TP +FN))		1.0	0.92	0.76
Sp (TP/(TP + FP))		1.0	0.79	0.91

Table S11. Detailed breakdown of the lymphoma cross-validation. The missed predictions in the cross-validation were not uniformly distributed among samples. Each tumor sample was tested against every possible model built from all but one of each type of tumor. Given 6 LPC, 11 APC, and 6 APN samples, LPC and APN samples were each tested 66 times, and APC samples were tested 36 times.

Tumor		Cross-validation Prediction Percentage		
Type	Number	LPC	APC	APN
LPC	1	100	0	0
	2	100	0	0
	3	100	0	0
	4	100	0	0
	5	100	0	0
	6	100	0	0
APC	1	0	83.3	16.7
	2	0	38.9	61.1
	3	0	100	0
	4	0	100	0
	5	0	100	0
	6	0	100	0
	7	0	100	0
	8	0	100	0
	9	0	94.4	5.6
	10	0	100	0
	11	0	100	0
APN	1	0	0	100
	2	0	100	0
	3	0	0	100
	4	0	3	97
	5	0	40.9	59.1
	6	0	0	100

Table S12. GO terms in the biological process ontology that are over-represented in differentially processed genes among the mouse B-cell lymphoma samples. GO analysis done with GOstat. Complete output is available at http://harlequin.jax.org/cancerAPA/.

GO term	Description	count	all	P-value
GO:0007049	cell cycle	32	697	3.92E-04
GO:0022402	cell cycle process	26	552	1.25E-03
GO:0046777	protein amino acid autophosphorylation	6	33	4.87E-03
GO:0016485	protein processing	8	66	4.87E-03
GO:0035295	tube development	14	201	4.87E-03
GO:0016540	protein autoprocessing	6	35	5.28E-03
GO:0043170	macromolecule metabolic process	137	5327	0.020
GO:0006281	DNA repair	12	187	0.025
GO:0032502	developmental process	80	2804	0.025
GO:0016043	cellular component organization and biogenesis	68	2313	0.029
GO:0043009	chordate embryonic development	13	221	0.029
GO:0045786	negative regulation of progression through cell cycle	9	116	0.029
GO:0009792	embryonic development ending in birth or egg hatching	13	224	0.029
GO:0065007	biological regulation	103	3853	0.029
GO:0019538	protein metabolic process	77	2714	0.030
GO:0006469	negative regulation of protein kinase activity	5	36	0.030
GO:0033673	negative regulation of kinase activity	5	36	0.030
GO:0006464	protein modification process	46	1435	0.030
GO:0006974	response to DNA damage stimulus	13	231	0.030
GO:0035239	tube morphogenesis	10	148	0.030
GO:0048513	organ development	43	1323	0.030
GO:0051348	negative regulation of transferase activity	5	38	0.033
GO:0001838	embryonic epithelial tube formation	6	57	0.033
GO:0006950	response to stress	27	737	0.044
GO:0009719	response to endogenous stimulus	13	250	0.049
GO:0044238	primary metabolic process	150	6147	0.049

Table S13. GO terms in the biological process ontology that are over-represented in differentially processed genes among the human melanoma samples. GO analysis done with GOstat. Complete output is available at http://harlequin.jax.org/cancerAPA/.

GO term	Description	count	all	P-value
GO:0007275	multicellular organismal development	80	2299	2.48E-28
GO:0032502	developmental process	99	3347	1.61E-26
GO:0048856	anatomical structure development	71	2005	5.01E-26
GO:0048731	system development	60	1605	6.09E-24
GO:0032501	multicellular organismal process	103	3822	4.50E-23
GO:0048513	organ development	43	1141	6.36E-17
GO:0009653	anatomical structure morphogenesis	40	1047	4.50E-16
GO:0007399	nervous system development	31	716	3.74E-15
GO:0022610	biological adhesion	32	960	9.94E-10
GO:0007155	cell adhesion	32	960	9.94E-10
GO:0065007	biological regulation	120	6731	1.35E-08
GO:0048869	cellular developmental process	46	1810	4.42E-08
GO:0030154	cell differentiation	46	1810	4.42E-08
GO:0000902	cell morphogenesis	19	478	1.13E-07
GO:0032989	cellular structure morphogenesis	19	478	1.13E-07
GO:0050789	regulation of biological process	105	6140	3.62E-06
GO:0009887	organ morphogenesis	17	362	2.84E-05
GO:0031325	positive regulation of cellular metabolic process	17	365	3.01E-05
GO:0048732	gland development	7	46	3.36E-05
GO:0045935	positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	15	289	3.36E-05
GO:0006366	transcription from RNA polymerase II promoter	20	640	3.90E-05
GO:0048518	positive regulation of biological process	28	1062	3.90E-05
GO:0045893	positive regulation of transcription, DNA-dependent	13	223	4.59E-05
GO:0009893	positive regulation of metabolic process	17	391	5.78E-05
GO:0045740	positive regulation of DNA replication	3	3	5.78E-05

Table S14. GO terms in the biological process ontology that are over-represented in differentially processed genes among the human breast cancer samples. GO analysis done with GOstat. Complete output is available at http://harlequin.jax.org/cancerAPA/.

GO term	Description	count	all	P-value
GO:0007275	multicellular organismal development	64	2299	5.41E-23
GO:0032501	multicellular organismal process	81	3822	1.82E-17
GO:0032502	developmental process	72	3347	7.59E-16
GO:0048856	anatomical structure development	50	2005	2.32E-14
GO:0048731	system development	43	1605	4.72E-14
GO:0000165	MAPKKK cascade	14	150	5.79E-09
GO:0048513	organ development	29	1141	4.21E-08
GO:0007243	protein kinase cascade	19	376	6.20E-08
GO:0009653	anatomical structure morphogenesis	27	1047	8.68E-08
GO:0065007	biological regulation	95	6731	8.22E-07
GO:0048869	cellular developmental process	37	1810	9.83E-07
GO:0030154	cell differentiation	37	1810	9.83E-07
GO:0007399	nervous system development	20	716	1.35E-06
GO:0006366	transcription from RNA polymerase II promoter	18	640	6.21E-06
GO:0045893	positive regulation of transcription, DNA-dependent	12	223	2.75E-05
GO:0048518	positive regulation of biological process	24	1062	2.91E-05
GO:0050789	regulation of biological process	84	6140	2.91E-05
GO:0043405	regulation of MAP kinase activity	8	93	6.11E-05
GO:0031098	stress-activated protein kinase signaling pathway	7	68	7.98E-05
GO:0050794	regulation of cellular process	78	5704	7.98E-05
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	9	139	1.19E-04
GO:0031325	positive regulation of cellular metabolic process	14	365	1.19E-04
GO:0045860	positive regulation of protein kinase activity	9	141	1.26E-04
GO:0033674	positive regulation of kinase activity	9	144	1.43E-04
GO:0051347	positive regulation of transferase activity	9	147	1.63E-04
GO:0045941	positive regulation of transcription	12	279	1.65E-04
GO:0007154	cell communication	75	5560	1.86E-04
GO:0045859	regulation of protein kinase activity	11	237	1.86E-04
GO:0048522	positive regulation of cellular process	21	954	1.88E-04
GO:0009893	positive regulation of metabolic process	14	391	1.92E-04
GO:0045935	positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	12	289	1.97E-04
GO:0043549	regulation of kinase activity	11	242	1.99E-04

Table S15. GO terms identified as overrepresented in all differential processing gene sets.

GO Term	Description
GO:0001944	vasculature development
GO:0006468	protein amino acid phosphorylation
GO:0007275	multicellular organismal development
GO:0007507	heart development
GO:0007517	muscle development
GO:0009790	embryonic development
GO:0009792	embryonic development ending in birth or egg hatching
GO:0016043	cellular component organization and biogenesis
GO:0030154	cell differentiation
GO:0032502	developmental process
GO:0043009	chordate embryonic development
GO:0043283	biopolymer metabolic process
GO:0043549	regulation of kinase activity
GO:0045859	regulation of protein kinase activity
GO:0048513	organ development
GO:0048522	positive regulation of cellular process
GO:0048731	system development
GO:0048856	anatomical structure development
GO:0048869	cellular developmental process
GO:0050789	regulation of biological process
GO:0051338	regulation of transferase activity
GO:0065007	biological regulation

Table S16. Genes identified as alternatively processed in the comparison of cell lines derived from human ovarian cancer that differ in resistance to the chemotherapeutic cisplatin. Truncation and elongation refer, respectively, to loss or gain of hybridization signal in in the 3'-end of the gene compared to the 5'-end in resistant cells compared to sensitive cells. To focus on likely changes in 3'-processing, this list is limited to genes whose transcripts have only one statistically significant processing event.

Elongated			**Truncated**		
ABCB9	FOXN3	NOL4	ALS2CR4	FUNDC2	PTCD2
AF070541	GALNS	NRG1	ANKRD22	GGA1	PTCH1
ANKRD15	GALNT2	OPA3	ANTXR1	GLUL	QSOX1
ANKRD46	GINS2	PDGFRA	ANXA1	GRK5	RECK
ASXL3	GPT2	PHF21B	APOC1	HOXA9	RPA1
BPTF	GYPA	POLG	BCL11A	ICAM2	SLC10A3
C10orf110	HAPLN1	POSTN	BMP5	KCNE4	TAF5
C9orf3	HAR1A	PPA2	C11orf30	LGALS8	TAF8
CCT8	HNRNPA2B1	RAD54B	C1orf151	LOXL2	TBCD
CD36	HSPA14	RCAN1	C1orf43	LSS	TCL1B
CDKN2C	HTR1F	RTN1	C7orf44	MFAP3L	TDRKH
CHRM3	IDI1	SFRS4	CAMK2D	MGC16121	TH1L
CNN1	IGFBP5	SPARCL1	CAPN3	MIB2	TM4SF18
COL16A1	IQCE	SRGN	CCDC132	MRPS18C	TMED4
COQ2	KIF26A	SYDE1	CCDC80	MYBL1	TMTC2
CTSC	KISS1R	SYNJ1	CDH2	MYL9	TPI1
CXorf38	KLHL14	TJP2	CLDN11	NDST1	TUBA4A
CXorf56	LOC283663	TLE4	CSF1R	NDUFB2	TUBB6
DMRT3	MAB21L1	TNFRSF11B	CSTF3	NEFL	TXNDC17
DNAJB5	MCAM	TPCN2	DCLK1	NOVA1	UBE2U
E2F7	MCM10	TPD52	DEF8	NPHP1	UCHL5IP
EDG2	MDFIC	TPM2	DNMT3A	PDGFC	WBSCR22
EPC1	MLLT3	TRPC7	EIF4EBP3	PDLIM7	ZFYVE20
EXOSC4	MYL4	UCHL5	EPHA7	PHF19	
FAM126B	NDUFS7	ZFPL1	FANCM	PIR	
FANCD2	NF2	ZNF423	FBN2	POLR2L	
FASN	NLGN2	ZNF503	FLJ13236	PPP1R9A	

Table S17. Significantly enriched GO terms from the "biological process" ontology from the set of genes that show any processing event in the comparison of cisplatin-sensitive and resistance cell lines (Table S8).

GO term	Description	count	all	P-value
GO:0048856	anatomical structure development	32	2005	3.10E-10
GO:0048731	system development	25	1605	2.34E-07
GO:0007275	multicellular organismal development	31	2299	3.63E-07
GO:0032502	developmental process	39	3347	8.07E-07
GO:0048519	negative regulation of biological process	18	1182	6.91E-05
GO:0032501	multicellular organismal process	39	3822	8.32E-05
GO:0009892	negative regulation of metabolic process	11	436	9.95E-04
GO:0007399	nervous system development	14	716	9.95E-04
GO:0048523	negative regulation of cellular process	16	1137	1.07E-03
GO:0031324	negative regulation of cellular metabolic process	10	381	1.30E-03
GO:0065007	biological regulation	53	6731	5.69E-03
GO:0009653	anatomical structure morphogenesis	14	1047	7.37E-03
GO:0044237	cellular metabolic process	85	12668	8.25E-03
GO:0045934	negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	8	319	8.32E-03
GO:0008283	cell proliferation	12	745	0.012
GO:0016043	cellular component organization and biogenesis	30	3277	0.012
GO:0040008	regulation of growth	6	201	0.018
GO:0048513	organ development	14	1141	0.020
GO:0016481	negative regulation of transcription	7	291	0.020
GO:0022610	biological adhesion	13	960	0.027
GO:0007155	cell adhesion	13	960	0.027
GO:0006937	regulation of muscle contraction	3	39	0.028
GO:0001558	regulation of cell growth	5	162	0.036
GO:0044238	primary metabolic process	82	12764	0.038
GO:0043283	biopolymer metabolic process	56	7940	0.040
GO:0016337	cell-cell adhesion	7	347	0.040
GO:0043170	macromolecule metabolic process	73	11144	0.043
GO:0007219	Notch signaling pathway	3	50	0.043
GO:0006996	organelle organization and biogenesis	16	1526	0.048

Table S18. Significantly enriched GO terms from the "biological process" ontology from the set of genes that show relative truncation in cisplatin-sensitive cell-lines compared to cisplatin-resistance cell line.

GO term	Description	count	all	P-value
GO:0007275	multicellular organismal development	12	2299	0.088
GO:0048856	anatomical structure development	11	2005	0.088
GO:0032502	developmental process	15	3347	0.088
GO:0032501	multicellular organismal process	16	3822	0.116
GO:0016337	cell-cell adhesion	4	347	0.116
GO:0007018	microtubule-based movement	3	180	0.116
GO:0048731	system development	9	1605	0.116
GO:0043687	post-translational protein modification	11	2235	0.119
GO:0030705	cytoskeleton-dependent intracellular transport	3	196	0.119
GO:0051641	cellular localization	7	1126	0.119
GO:0019538	protein metabolic process	21	5858	0.119
GO:0006464	protein modification process	12	2704	0.119
GO:0046907	intracellular transport	6	910	0.119
GO:0043170	macromolecule metabolic process	34	11144	0.119
GO:0044267	cellular protein metabolic process	20	5584	0.119
GO:0044260	cellular macromolecule metabolic process	20	5656	0.119
GO:0043412	biopolymer modification	12	2815	0.119
GO:0016043	cellular component organization and biogenesis	13	3277	0.143
GO:0051649	establishment of cellular localization	6	1098	0.143
GO:0007017	microtubule-based process	3	315	0.143
GO:0065007	biological regulation	22	6731	0.152
GO:0043283	biopolymer metabolic process	25	7940	0.152
GO:0019226	transmission of nerve impulse	3	330	0.152
GO:0044237	cellular metabolic process	36	12668	0.152
GO:0048519	negative regulation of biological process	6	1182	0.163
GO:0044238	primary metabolic process	36	12764	0.163
GO:0016192	vesicle-mediated transport	4	606	0.163
GO:0006281	DNA repair	3	392	0.183
GO:0022610	biological adhesion	5	960	0.183
GO:0007155	cell adhesion	5	960	0.183
GO:0007010	cytoskeleton organization and biogenesis	4	686	0.192
GO:0051246	regulation of protein metabolic process	3	422	0.196

Table S19. Significantly enriched GO terms from the "biological process" ontology from the set of genes that show relative elongation in cisplatin-sensitive cell-lines compared to cisplatin-resistance cell line.

GO term	Description	count	all	P-value
GO:0032502	developmental process	21	3347	4.59E-05
GO:0048856	anatomical structure development	18	2005	4.59E-05
GO:0032501	multicellular organismal process	21	3822	7.58E-04
GO:0007275	multicellular organismal development	17	2299	7.58E-04
GO:0048731	system development	14	1605	7.58E-04
GO:0031324	negative regulation of cellular metabolic process	7	381	1.64E-03
GO:0009892	negative regulation of metabolic process	7	436	3.31E-03
GO:0045934	negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	6	319	4.43E-03
GO:0007399	nervous system development	8	716	9.04E-03
GO:0048519	negative regulation of biological process	10	1182	0.011
GO:0016481	negative regulation of transcription	5	291	0.017
GO:0048523	negative regulation of cellular process	9	1137	0.026
GO:0009653	anatomical structure morphogenesis	8	1047	0.027
GO:0048518	positive regulation of biological process	8	1062	0.027
GO:0016043	cellular component organization and biogenesis	15	3277	0.035
GO:0045941	positive regulation of transcription	4	279	0.035
GO:0045935	positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	4	289	0.038
GO:0022414	reproductive process	4	305	0.041
GO:0048522	positive regulation of cellular process	7	954	0.041
GO:0022610	biological adhesion	7	960	0.041
GO:0007155	cell adhesion	7	960	0.041
GO:0050789	regulation of biological process	23	6140	0.041
GO:0001558	regulation of cell growth	3	162	0.045
GO:0006259	DNA metabolic process	8	1249	0.045
GO:0006996	organelle organization and biogenesis	9	1526	0.045
GO:0008610	lipid biosynthetic process	4	333	0.048
GO:0044237	cellular metabolic process	39	12668	0.048
GO:0000074	regulation of progression through cell cycle	4	353	0.049
GO:0006260	DNA replication	4	355	0.049
GO:0051726	regulation of cell cycle	4	359	0.049
GO:0065007	biological regulation	24	6731	0.049
GO:0031325	positive regulation of cellular metabolic process	4	365	0.049
GO:0006631	fatty acid metabolic process	3	191	0.049

Supplemental References

1. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society 1995; Series B/ *85*: 289-300.
2. Brockman JM, Singh P, Liu D, Quinlan S, Salisbury J, Graber JH. PACdb: PolyA Cleavage Site and 3'-UTR Database. Bioinformatics 2005; 21: 3691-3.
3. Beissbarth T, Speed TP. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. Bioinformatics 2004; 20: 1464-5.
4. Blake JA, Bult CJ, Eppig JT, Kadin JA, Richardson JE. The Mouse Genome Database genotypes::phenotypes. Nucleic Acids Res 2009; 37: D712-9.
5. Barrell D, Dimmer E, Huntley RP, Binns D, O'Donovan C, Apweiler R. The GOA database in 2009--an integrated Gene Ontology Annotation resource. Nucleic Acids Res 2009; 37: D396-403.
6. Griffiths-Jones S. miRBase: the microRNA sequence database. Methods Mol Biol 2006; 342: 129-38.
7. Mi S, Lu J, Sun M, et al. MicroRNA expression signatures accurately discriminate acute lymphoblastic leukemia from acute myeloid leukemia. Proc Natl Acad Sci U S A 2007; 104: 19971-6.
8. Chang TC, Yu D, Lee YS, et al. Widespread microRNA repression by Myc contributes to tumorigenesis. Nat Genet 2008; 40: 43-50.
9. Zhang H, Lee JY, Tian B. Biased alternative polyadenylation in human tissues. Genome Biol 2005; 6: R100.
10. Edwalds-Gilbert G, Milcarek C. Regulation of poly(A) site use during mouse B-cell development involves a change in the binding of a general polyadenylation factor in a B-cell stage-specific manner. Mol Cell Biol 1995; 15: 6420-9.
11. Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. Science 2008; 320: 1643-7.
12. Carninci P, Sandelin A, Lenhard B, et al. Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet 2006; 38: 626-35.
13. Song MG, Kiledjian M. 3' Terminal oligo U-tract-mediated stimulation of decapping. RNA 2007; 13: 2356-65.
14. Suzuki R, Shimodaira H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 2006; 22: 1540-2.