TECHNICAL BULLETIN

RNA Profiling with the DASL® Assay

1. INTRODUCTION

Microarray analysis of gene expression has proven to be a remarkable tool for biologists, and has the potential to be widely adopted by clinical researchers. One challenge to clinical applications has been the need for high quality, cryopreserved RNA for standard microarray sample labeling. Such samples are of limited availability and must be collected over the course of a disease. In contrast, formalin-fixed, paraffin-embedded (FFPE) tissue samples are readily available and have the added advantage of known patient outcomes. However, the poor integrity of RNA derived from FFPE samples has largely precluded their use in microarray analysis. Illumina has developed a novel gene expression assay for microarrays that is capable of utilizing partially degraded RNA.

The cDNA-mediated Annealing, Selection, extension and Ligation (DASL) Assay can monitor RNA expression of up to 1536 sequence targets derived from RNA in FFPE samples stored up to 12 years. Together with the high-throughput capability of Illumina's microarray platforms, the DASL Assay offers researchers the opportunity to analyze hundreds to thousands of RNA transcripts derived from previously collected, preserved samples. This technical bulletin explains the features of the DASL Assay, describes its performance for measuring RNA derived from both fresh-frozen and FFPE samples and demonstrates its utility in analyzing clinical samples.

2. BEADARRAY™ TECHNOLOGY

At the heart of Illumina's expression products lies a fundamentally different way of building arrays: the random self-assembly of microspheres into ordered microwell substrates. Illumina has used technological advances in both the fiber-optics and MEMS industries to build substrates that contain thousands of wells in an ordered, addressable pattern. Quantitatively pooled libraries of beads are then introduced to the etched microwell substrates. The beads automatically assemble into the wells, resulting in the highest-density array platform currently available. To meet the broadest possible range of researchers' needs, Illumina has developed two different Sentrix® products in Array of Arrays™ formats: the Array Matrix and the BeadChip.

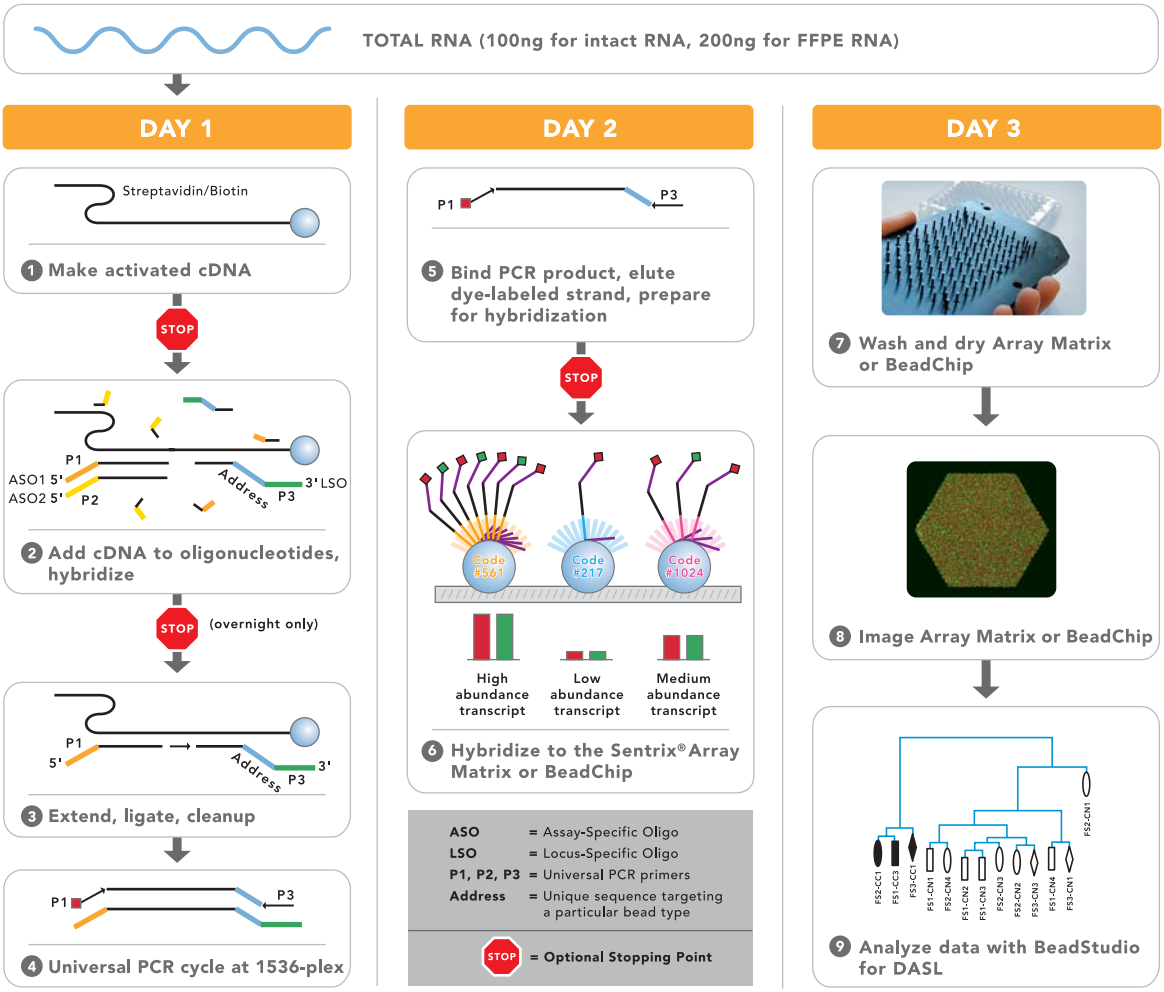
FIGURE 1: SENTRIX® ARRAY MATRIX AND BEADCHIPS

Sentrix Universal-96 Array Matrix (top left) and Sentrix Universal-16 BeadChip formats for 96 or 16 samples, respectively, for use with Illumina's DASL Assay.

The Sentrix Array Matrix (SAM) uses fiber-optic bundles comprising nearly 50,000 individual light-conducting fiber strands that are chemically etched to create a 3 µm well at the end of each strand. Array bundles are grouped together into a 96-array configuration that matches the well spacing of standard 96-well plates. This unique format allows users to conduct experiments simply and quickly on 96 arrays at a time (*Figure 1*). Moreover, the platform can be readily incorporated into automation routines with standard robotic equipment, leading to reduced error, labor and resource requirements.

For users with moderate sample throughput demands, Illumina has introduced the BeadChip format (*Figure 1*). This platform, approximately the size of a glass slide, allows processing of up to 16 samples at a time, and can be scanned in the same BeadArray Reader as the SAM. Independent of format, multiple copies of oligonucleotide probes are covalently attached to each bead in Illumina arrays. Up to 1536 unique probe sequences are represented in each array, with an average 30-fold redundancy of each bead type. After bead assembly, a hybridization-based procedure is used to map the array, which identifies the bead in each well. This final process validates the performance of each bead type and provides a level of quality control unmatched in the microarray industry.

FIGURE 2: DASL WORKFLOW



3. THE DASL ASSAY

The DASL Assay monitors gene expression by targeting sequences in cDNAs with probe groups. Other than the initial preparation of samples, the core biochemical approach is the same as that used in the Illumina GoldenGate® genotyping assay, offering researchers compatibility with existing reagents and arrays. The probe groups are composed of multiple parts. In addition to gene-specific sequences, the probe groups contain primer landing sites for PCR amplification (P1, P2 and P3) and an address sequence for hybridization to the universal array. To allow for future applications using the same biochemical approach (such as allele-specific expression analysis), the DASL Assay is configured to use both Cy3 and Cy5 fluorescence in targeting non-polymorphic sequences.

In the DASL Assay, probes are designed to target exonic sequences. This design requirement means genomic DNA may be used as a positive control sample. Because genomic DNA consists of an essentially equimolar representation of all sequence targets, the performance of any individual assay can be assessed. This capability offers a unique method for verification of probe performance as a means to decrease false negatives.

In addition to a requirement for exonic target sequences, probe groups are selected for uniqueness and designed to provide optimal performance in the DASL Assay. Custom oligo probes are designed by Illumina's bioinformatics engines and are accompanied by a "score" that reflects each target site for specificity in the

transcriptome (RefSeq), the genome and predicted behavior in assay biochemistry.

To begin the assay, total RNA is converted to cDNA using both biotinylated random nonamers (biotin-d$(N)_9$) and biotinylated oligo d$(T)_{18}$. The probe groups are annealed to the biotinylated cDNA, followed by selection of the duplexes on streptavidin beads to remove unhybridized oligos. Only those oligos that are correctly annealed are extended and ligated to generate amplifiable products. These templates are labeled during PCR amplification by including fluorescent primers in the reaction. Finally, the labeled strand is recovered and hybridized to a universal BeadArray which is scanned to generate fluorescence intensity data. Because randomers are used in the cDNA synthesis, and because the probe groups target cDNA sequences spanning only about 50 bases, partially degraded RNA can be used in the assay.

In its design, the DASL Assay most resembles RT-PCR with highly multiplexed templates but with only three PCR primers. Because the oligos all share the same primers, and the amplicons are of a uniform size, the amplification step maintains an unbiased representation of transcript abundance[1].

An important feature of the DASL Assay approach is that all the ligated oligos compete for the shared primers, and therefore share the available fluorescent label. This biochemical approach has several consequences that are important when interpreting array data. First, the labeled products only report the transcripts targeted with probe groups, so any genes not targeted will not be monitored. This feature can be advantageous when the oligo pool is designed to detect small differences among samples, or when monitoring gene expression in a subset of cells in a heterogeneous mixture. Second, because the assay uses a fixed cycle number to amplify ligated oligos, it is relatively insensitive to RNA input levels above a minimum. If one sample has double the mass of another but has no change in expression, then the higher input sample will exhaust the PCR sooner than the lower input sample, but there will be no difference in relative expression. Finally, the labeling of one product is not independent of the labeling of another. Therefore, the intensity for any given transcript may be influenced by a change in expression of another transcript. One of the key benefits of this arrangement is that the DASL Assay is particularly sensitive to differences between samples.

4. DASL ASSAY WORKFLOW

An overview of the DASL Assay workflow is shown in *Figure 2*. After cDNA synthesis, the assay shares its core processes and reagents with those used in the Illumina GoldenGate genotyping assay. The assay can be stopped at several points, as indicated, and requires a single overnight hybridization to the array.

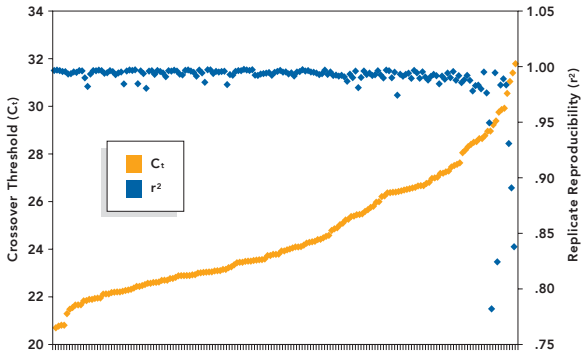
The DASL Assay starts with reverse transcription of total RNA. Intact RNA may be purified using any standard method, so long as the RNA is free from contaminants that would inhibit reverse transcription, such as alcohols or guanidine salts. RNA derived from FFPE tissues performed best in the DASL Assay when the High Pure RNA Paraffin Kit from Roche® Molecular Biochemicals (catalog #3270289) was used for purification. Important parameters for RNA purification from FFPE samples include using tissue sections not exceeding 5 µm in thickness and using freshly prepared Proteinase K for tissue homogenization.

The amount of RNA required for the DASL Assay can vary depending on the RNA integrity. For intact RNA, Illumina recommends 100 ng of total RNA per assay. For partially degraded RNA, a minimum of 200 ng of total RNA is recommended to compensate for decreased amounts of intact sequence targets. Lower levels of RNA input can result in increased variability.

5. EVALUATING RNA SAMPLES

RNA derived from FFPE tissues can vary considerably in integrity and purity. Illumina has developed a method based on real-time RT-PCR for assessing the expected performance of RNA in the DASL Assay. Primers directed against the highly expressed RPL13a gene are used to amplify a 90 base pair fragment. The crossover

FIGURE3: CORRELATION BETWEEN SAMPLE QC AND DATA QUALITY



threshold (C_t) for detection is a measurement of the level of intact RNA for this transcript in the sample. *Figure 3* shows a comparison of the crossover threshold to the reproducibility of duplicate DASL Assays for 143 independent FFPE-derived RNA transcripts. These data indicate that RNA with a C_t of 29 cycles or less can be used successfully in the DASL Assay. For comparison, the crossover threshold for the same amount of intact reference RNA under the same conditions is 17 cycles. The data suggest the DASL Assay can tolerate a wide range of RNA degradation, up to a 12 cycle difference between intact and degraded samples.

6. THE DASL CANCER PANEL

The Illumina DASL Cancer Panel is a pool of selected probe groups that targets 502 genes collected from ten publicly available cancer gene lists (*Table 1*). Genes were chosen based on the frequency of appearance on these lists and the frequency of literature citations of these genes in association with cancer. Each gene in the Cancer Panel is targeted in 3 locations with 3 probe groups per gene.

Assay performance was tested using the DASL Cancer Panel with commercially available RNA and FFPE tissues (*Table 2*). Excellent sample replicate reproducibility was observed for intact RNA, both at the individual probe level and at the gene level. For RNA derived from FFPE samples, excellent reproducibility was demonstrated at the probe level, and the reproducibility at the gene level was equivalent to that for intact RNA. Highest r^2 coefficients were obtained for samples hybridized to the same SAM.

The false positive rate for the DASL Assay was determined by testing sample replicates for detection of differentially expressed genes. For both intact and

TABLE 1: DASL PANEL CANCER GENES

ABCB1	BRCA2	CHEK1	E2F2	FGF2	HIF1A	KIT	MMP1	NUMA1	RAD23A	SPP1	TNFRSF6
ABCC2	BTK	COL18A1	E2F3	FGF3	HLF	KRAS2	MMP10	OGG1	RAD50	SRC	TNFSF10
ABCG2	CASP10	COL1A1	E2F5	FGF5	HMMR	L1CAM	MMP14	OSM	RAD51	STAT1	TNFSF6
ABL1	CASP2	COL4A3	EGF	FGF6	HOXA9	LAF4	MMP2	PBX1	RAD52	STAT3	TNFSF8
ADPRT	CASP3	COMT	EGFR	FGF7	HRAS	LAMB1	MMP3	PCNA	RAD54B	STAT5B	TOP1
AHR	CASP8	COPEB	EGR1	FGF8	ICAM1	LCK	MMP7	PCTK1	RAD54L	STK11	TOP2A
AIM2	CAV1	CREBBP	ELK1	FGF9	IFNG	LCN2	MMP9	PDGFA	RAF1	SYK	TP53
AKT1	CBFA2T1	CRK	ELK3	FGFR1	IFNGR1	LIF	MOS	PDGFB	RALB	TAL1	TP73
AKT2	CBL	CRKL	ELL	FGFR2	IFNGR2	LIG1	MPL	PDGFRA	RAN	TCF7L2	TPR
ALK	CBLB	CSF1R	EMS1	FGFR3	IGF1	LIG3	MRE11A	PDGFRB	RAP1A	TDGF1	TRAF3
ALOX12	CCNA2	CSF2	ENC1	FGFR4	IGF1R	LIG4	MSF	PGF	RAP1GDS1	TEK	TRAF4
APAF1	CCNC	CSF3	EPHA1	FGR	IGF2	LMO1	MSH2	PGR	RAP2A	TERT	TSC1
APC	CCND1	CSF3R	EPHB4	FHIT	IGF2R	LMO2	MSH3	PIK3CA	RARA	TFAP2C	TSC2
AR	CCND2	CSK	EPO	FLI1	IGFBP1	LTA	MSH6	PIM1	RARB	TFDP1	TSG101
ARAF1	CCND3	CSPG2	EPS15	FLT1	IGFBP2	LYN	MST1R	PLA2G2A	RASA1	TFE3	TYMS
AREG	CCNE1	CTGF	EPS8	FLT3	IGFBP3	MAD	MTA1	PLAG1	RB1	TFF1	TYRO3
ARHA	CCNH	CTNNA1	ERBB2	FLT4	IGFBP5	MADH2	MTHFR	PLAT	RBBP1	TFG	VAV1
ARHGDIB	CD34	CTNNB1	ERBB3	FOLR1	IGFBP6	MADH4	MUC1	PLAUR	RBBP2	TFRC	VAV2
ARHH	CD44	CTSD	ERBB4	FOS	IL11	MAF	MXI1	PLG	RBBP5	TGFA	VBP1
ARHI	CD59	CTSL	ERCC1	FOSB	IL12A	MALT1	MYB	PML	RBBP6	TGFB1	VEGF
ARNT	CD9	CUL2	ERCC2	FOSL2	IL12B	MAP3K8	MYBL2	PMS1	RBL2	TGFB2	VEGFB
ATF1	CDC2	CXCL9	ERCC3	FRAP1	IL13	MAPK10	MYC	PNUTL1	RECQL	TGFB3	VHL
ATM	CDC25A	CYP1A1	ERCC4	FRZB	IL1A	MAPK14	MYCL1	PPARD	REL	TGFBI	VIL2
AXL	CDC25B	CYP1B1	ERCC5	FVT1	IL1B	MAS1	MYCL2	PPARG	RELA	TGFBR1	WEE1
BAD	CDC25C	DAB2	ERCC6	FYN	IL1RN	MATK	MYCN	PPP2R1B	RET	TGFBR2	WNT1
BAG1	CDH1	DAP3	ERG	FZD7	IL2	MBD2	NAT2	PRCC	RIPK1	TGFBR3	WNT10B
BAK1	CDH11	DAPK1	ESR1	G22P1	IL3	MCAM	NBS1	PRKAR1A	RLF	THBS2	WNT2
BARD1	CDH13	DCC	ETS1	GADD45A	IL4	MCC	NEO1	PRKR	ROS1	THPO	WNT2B
BCL2	CDK10	DCN	ETS2	GAS1	IL6	MCF2	NF1	PTCH	RRAS	TIAM1	WNT5A
BCL2A1	CDK2	DDB2	ETV1	GAS7	IL8	MCL1	NFKB1	PTCH2	S100A4	TIMP1	WNT8B
BCL2L1	CDK4	DDIT3	ETV6	GFI1	ILK	MDM4	NFKB2	PTEN	SEMA3F	TIMP2	WRN
BCL3	CDK6	DDX6	EVI1	GLI	ING1	MDS1	NFKBIA	PTGS1	SERPINE1	TIMP3	WT1
BCL6	CDK7	DEK	EVI2A	GLI2	INHA	MEL	NGFR	PTGS2	SH3BP2	TK1	XPA
BCR	CDK9	DKC1	EXT1	GLI3	IRF1	MEN1	NOS3	PTHLH	SHH	TNF	XPC
BIRC2	CDKN1A	DLC1	EXT2	GML	ITGB1	MET	NOTCH1	PTK2	SIAH1	TNFAIP1	XRCC1
BIRC3	CDKN1B	DLEU1	FANCA	GRB2	ITGB4	MLF1	NOTCH2	PTK7	SKI	TNFRSF10	XRCC2
BIRC5	CDKN2A	DLG3	FANCG	GRB7	JAK2	MLF2	NOTCH4	PTPRF	SKIL	A	XRCC4
BLM	CDKN2B	DMBT1	FAT	GRPR	JUN	MLH1	NQO1	PTPRG	SMARCA4	TNFRSF10	XRCC5
BMI1	CDKN2C	DSP	FER	GSTP1	JUNB	MLL	NRAS	PTPRH	SMARCB1	B	YES1
BMP4	CDKN2D	DTR	FES	HCK	JUND	MLLT3	NTRK1	PURA	SOD1	TNFRSF1A	YY1
BRAF	CEACAM1	DVL3	FGF1	HDAC1	KAI1	MLLT4	NTRK2	PXN	SPARC	TNFRSF1B	ZNF146
BRCA1	CEBPA	E2F1	FGF12	HDGF	KDR	MLLT6	NTRK3	QARS	SPI1	TNFRSF5	ZNFN1A1

TABLE 2: DASL ASSAY METRICS

INTACT RNA

	Reproducibility Among Replicates		False Positive Rate
	Probe Level	**Gene Level**	**Gene Level**
Within SAM	0.990	0.994	Less than 0.22%
Across SAMs	0.980	0.986	Less than 0.66%
Within BeadChip	0.990	0.995	
Across BeadChips	0.985	0.992	
SAM vs BeadChip	0.972	0.983	

FFPE-DERIVED RNA

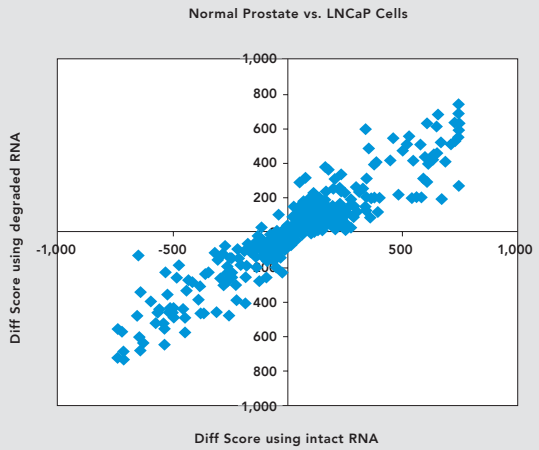
	Reproducibility Among Replicates		False Positive Rate
	Probe Level	**Gene Level**	**Gene Level**
Within SAM	0.975	0.991	Less than 0.20%

FFPE-derived RNA, the false positive rate within a SAM was ~0.22% (1 gene out of the 502 monitored). When comparing across SAMs, the false positive rate was ~0.66% (3-4 of the 502 genes).

In other tests, the DASL Assay has exhibited a dynamic range of 2.5 to 3 logs, a limit of detection of about $1x10^4$ molecules and measurement precision of 1.3 fold, using synthetic RNA targets[1]. Due to the competition among ligated templates for labeling in the PCR step, the sensitivity of the assay depends on the abundance of all the targeted transcripts, including those present in the background RNA in which the synthetic RNA was spiked. Therefore, measurements of dynamic range, limit of detection and measurement precision in the DASL Assay are sensitive to the oligo pool (i.e., probe groups) used and the proportion of synthetic to background RNA targeted. Given these considerations, the dynamic range, sensitivity and measurement precision of the DASL Assay using defined conditions are comparable to those determined using traditional microarray expression analyses.

When analyzing differential expression profiles of partially degraded RNA, it is important that the data faithfully reflect differential expression measured using intact RNA. To model such a comparison, RNA from both normal prostate tissue and the LNCaP prostate cancer cell line were intentionally partially degraded using heat to simulate FFPE-derived RNA. Both intact and partially degraded RNA were analyzed using the Illumina DASL Cancer Panel, and scores for differential expression were determined using the Illumina BeadStudio analysis software. Briefly, differential expression scores were derived from a p-value of the difference between samples for a given gene. *Figure 4* illustrates a comparison of the use of intact RNA vs. degraded RNA for the determination of differential expression scores between normal prostate RNA and LNCaP cell RNA. The comparison indicates that genes showing expression differences in partially degraded RNA are also genes that show expression differences

FIGURE 4: DIFFERENTIAL EXPRESSION OF CANCER GENES IN INTACT VS. DEGRADED RNA

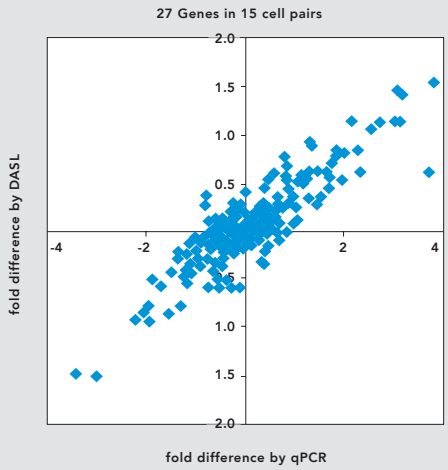


in intact RNA. This observation suggests expression analysis using degraded RNA will properly reflect biological differences measured using intact RNA.

7. COMPARISON TO qPCR

Quantitative PCR is commonly used for validation of microarray expression results. Among the 502 genes monitored using the DASL Cancer Panel in six cell types, 27 were chosen for analysis using qPCR. These genes were selected to cover a range of expression patterns including low, medium or high expression levels, and large, moderate or small expression differences. Fold differences were calculated pairwise among all six RNAs for both C_t in qPCR and for array intensity in the DASL Assay (*Figure* 5). The comparison of fold-difference detection between these platforms gives a correlation coefficient (r^2) of 0.87, a value comparable to many standard microarrays. Given the competition among amplifiable templates in the DASL Assay method, this result suggests an excellent correlation between expression differences as determined using the DASL Assay and those determined using a reference method.

FIGURE 5: COMPARISON BETWEEN DASL AND qPCR



8. DISTINGUISHING AMONG CLOSELY RELATED SAMPLES.

The sensitivity of the DASL Assay to changes in expression suggests this approach will be especially well suited to sample classification. To test this capability, RNA from normal prostate tissue was mixed with RNA from HL-60 cells at different ratios, then assayed using the DASL Cancer Panel. The correlation coefficient (r^2) for the RNA mixture is 0.981, consistent with very little difference in expression for these mixtures.

The data for all replicates of each RNA mixture were

FIGURE 6: CANCER GENE EXPRESSION IN FFPE SAMPLES

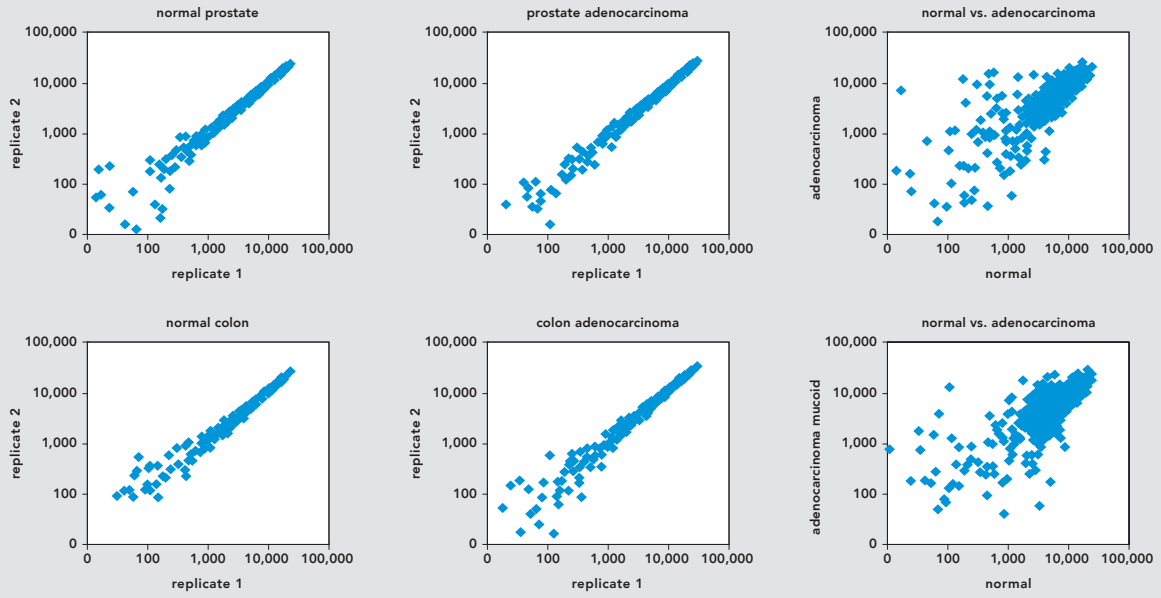
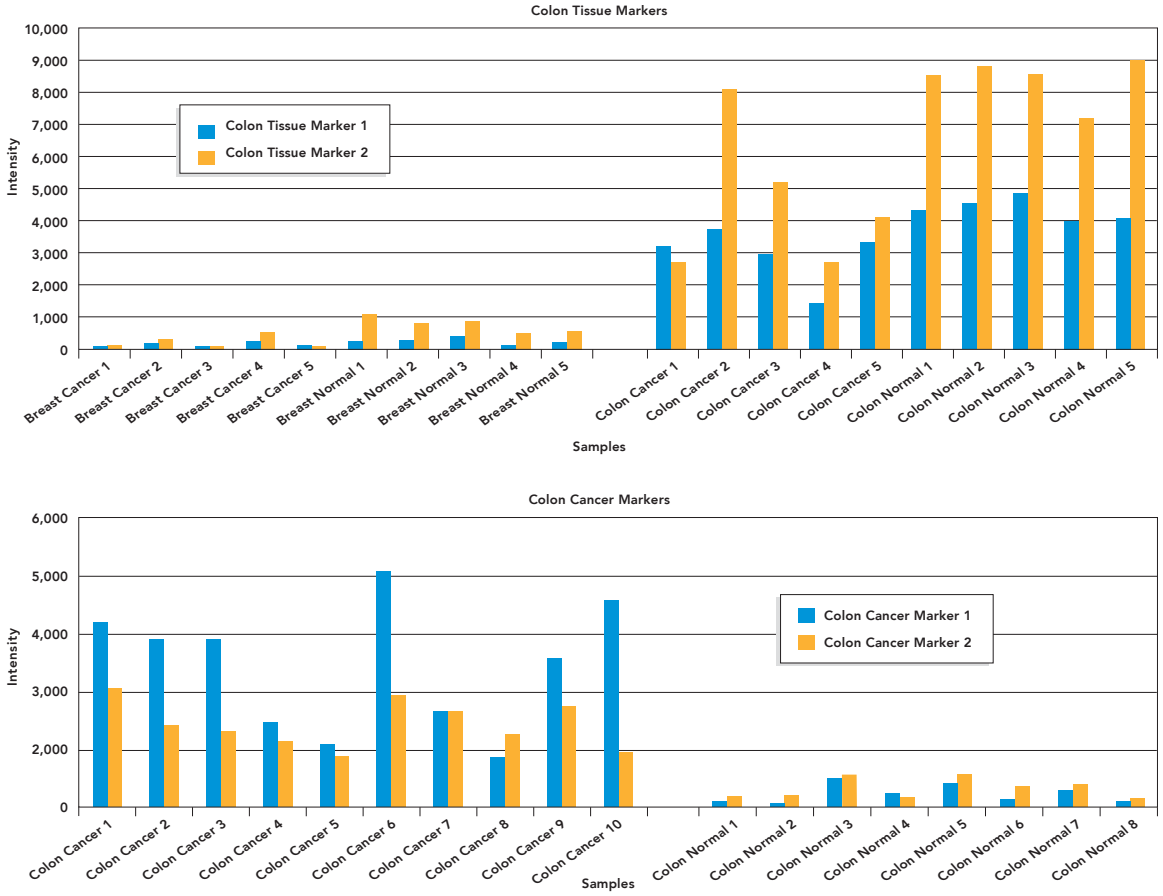


FIGURE 7: VALIDATION OF TISSUE-SPECIFIC AND CANCER-SPECIFIC MARKERS IN FFPE SAMPLES



analyzed by clustering the samples. Despite the close similarity of the 67% and 75% mixtures, all replicates of all mixtures were cleanly distinguished in a dendrogram (based on Ward clustering, generated using the Illumina BeadStudio software). This result suggests the DASL Assay can be used to classify samples with transcripts differing by less than 1.1 fold (75%/67% = 1.1).

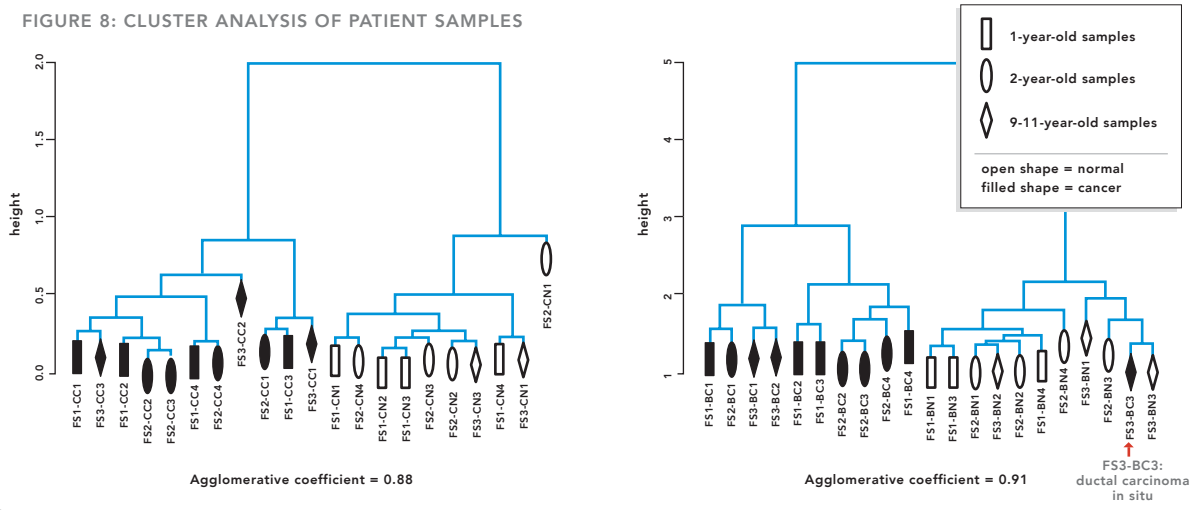
9. APPLICATION TO CLINICAL SAMPLES

To determine proper classification of clinical samples by the DASL Assay, RNA was first derived from commercially available FFPE tissue sections (BioChain Institute). Analysis of RNA from normal prostate, cancerous prostate and colon using the DASL Cancer Panel is shown in *Figure* 6. These data show excellent reproducibility among sample replicates, and dramatic differences in expression when comparing normal to diseased samples[2].

In other experiments using FFPE clinical samples, differences in expression were analyzed for known tissue and cancer markers[3] (*Figure 7*). The expression of two colon tissue markers was significantly higher in colon-derived samples than in breast-derived samples, regardless of whether the samples were derived from normal or cancerous tissues (top). In addition, the expression of two cancer markers was much higher in colon cancer samples than in normal colon samples (bottom). These data suggest the DASL Assay correctly reports biologically relevant results.

To address the applicability of the DASL Assay to clinical samples of varied quality, RNA was analyzed from FFPE tissues stored for 1 to 11 years. *Figure 8* shows

FIGURE 8: CLUSTER ANALYSIS OF PATIENT SAMPLES



patient samples were properly segregated into normal and cancerous groups for both colon and breast tissues with one exception. The one breast cancer sample not distinguishable from normal was derived in situ from an early stage ductal carcinoma that had been stored for 11 years[3]. These data illustrate the capability of the DASL Assay to classify existing FFPE clinical samples that have been stored for at least a decade.

10. SUMMARY

The DASL Assay is a novel approach to gene expression analysis that offers researchers the ability to interrogate partially degraded RNA, such as those derived from formalin-fixed, paraffin-embedded tissues. The unique features of the DASL Assay are particularly well suited for the discernment of very closely related samples, while reporting biologically relevant expression data. The assay takes advantage of Illumina's universal arrays, offering the additional flexibility to customize expression analysis without requiring the manufacture of a custom array.

DASL Gene Expression provides a high-throughput expression profiling solution that includes proprietary assay technology, standard or custom assay panels and user-friendly analysis tools to analyze 16 or 96 samples simultaneously using one of Illumina's multi-sample array formats. This solution allows researchers an unparalleled capacity to take advantage of existing tissue banks and known patient outcomes for clinical discovery.

REFERENCES

(1) Fan, J., Yeakley, J.M., Bibikova, M.,Chudin, E.W., Chen, J., Doucet, D., Rigault, P., Zhang, B., Shen, R., McBride, C., Li, H., Fu, X., Oliphant, A., Barker, D.L., Chee, M.S. (2004). A Versatile Assay for High-Throughput Gene Expression Profiling on Universal Array Martices. *Genome Res* **14**, 878-885.

(2) Bibikova, M., Yeakley, J.M., Chudin, E., Chen, J., Wickham, E., Wang-Rodriguez, J., Fan, J.B. (2004). Gene expression profiles in formalin-fixed, paraffin-embedded tissues obtained with a novel assay for microarray analysis. *Clin Chem* **50**, 2384-2386.

(3) Bibikova, M., Talantov, D., Chudin, E., Yeakley, J.M., Chen, J., Doucet, D., Wickham, E., Atkins, D., Barker, D, Chee, M.S., Wang, Y., Fan, J.B. (2004). Quantitative gene expres sion profiling in formalin-fixed, paraffin-embedded tissues using Universal Bead Arrays. *Am J of Pathol* **165**, 1799-1807.

ADDITIONAL INFORMATION

To learn more about DASL Gene Expression or about other Illumina products and services, contact us.

Illumina, Inc.
Customer Solutions
9885 Towne Centre Drive
San Diego, CA 92121-1975
1.800.809.4566 (toll free)
1.858.202.4566 (outside the U.S.)
techsupport@illumina.com
www.illumina.com

Pub. No. 470-2005-003 16Nov05

illumina®
making sense out of life