Integrated Transcriptomics and Metabolomics Study of Retinoblastoma Using Agilent Microarrays and LC/MS/GC/MS Platforms

Application Note

Authors

Nilanjan Guha, Deepak S.A.,
Syed Lateef, Seetaraman Gundimeda,
Arunkumar Padmanaban, and
Upendra Simha
Agilent Technologies, Inc.

Ashwin Mallipatna,
Vishnu Suresh Babu, and
Arkasubhra Ghosh
Narayana Nethralaya

Abstract

This Application Note illustrates a multi-omics approach combining transcriptomics and metabolomics to study molecular events in the progression of retinoblastoma (Rb). On a set of tissue samples affected with Rb, we performed mRNA and miRNA gene expression using microarrays followed by pathway analysis to identify gene enrichment that would enable functional characterization of Rb. Transcriptomics data were collected using Agilent SurePrint G3 Human microarrays and a SureScan microarray scanner. Metabolomics data were obtained from aqueous humor, vitreous humor, and tear samples of Rb using Agilent 7200 GC/Q-TOF and Agilent 6550 iFunnel Q-TOF LC/MS systems. After feature extraction and processing using Agilent MassHunter Software, differential and multi-omics analyses were performed using Agilent GeneSpring software suite. This study demonstrates how a comprehensive understanding of a disease mechanism can be obtained using Agilent mass spectrometry (GC/MS and LC/MS) and microarray technologies along with Agilent Mass Profiler Professional and GeneSpring GX software.

Agilent Technologies

Introduction

Retinoblastoma occurs when both copies of the RB1 gene in a child's retina are inactivated. While much literature exists regarding pathogenesis and genetic changes in retinoblastoma, there is still a lack of understanding of disease mechanism. In this study, we acquired multi-omics data using a combination of an Agilent portfolio of instruments, software tools, and sample preparation kits to acquire, integrate, and analyze multi-omics data.

The study samples were classified with a risk number based on the stage of disease progression. Total RNA extracted from nine fresh frozen Rb and two control tissue samples were analyzed using Agilent SurePrint G3 Gene Expression microarrays to define a gene expression pattern. The Rb sample expressions were compared to two control samples. Metabolite extraction was carried out from aqueous humor, vitreous humor, and tears of Rb and control samples. The extracted metabolites were subjected to analysis using LC/MS and GC/MS. Figure 1 shows the various Agilent products used in this multi-omic study.

Data generated from transcriptomics (mRNA and miRNA) and metabolomics analyses were analyzed using Agilent GeneSpring 13.1 software.

Experimental

Gene expression and miRNA microarray analysis

Total RNA was extracted from Rb and control tissues using the Agilent Absolutely RNA miRNA Kit (Cat# 400814). The quality of isolated RNA was determined on an Agilent 2200 TapeStation system (G2964AA) using an Agilent RNA ScreenTape assay (5067-5576). mRNA labeling and microarray processing was performed as detailed in the "One-Color Microarray-Based Gene Expression Analysis" (version 6.9, p/n G4140-90040). miRNA labeling was done using an Agilent miRNA Complete Labeling and Hyb Kit (Cat# 5190-0456).

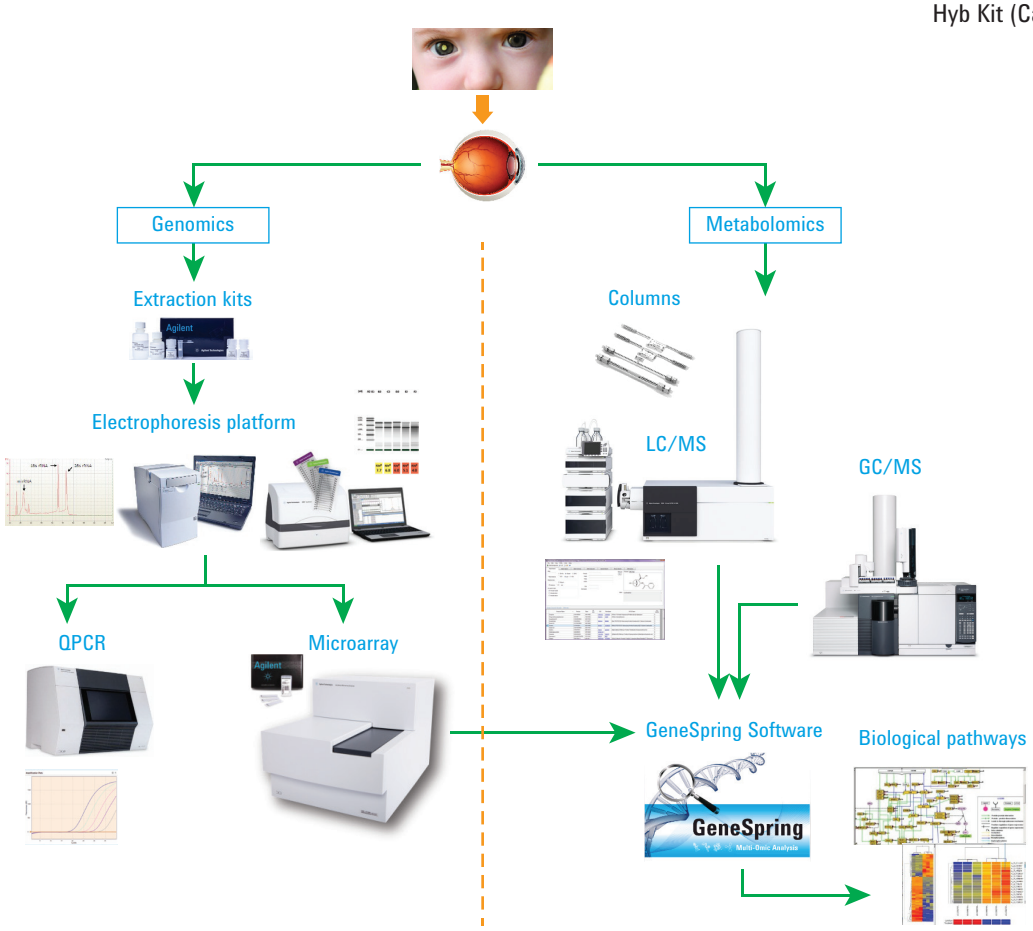

Figure1. Agilent products used in the study.

The gene expression and miRNA data were extracted using Agilent Feature Extraction Software (11.5.1.1) and analyzed using Agilent GeneSpring GX 13.1. In both mRNA and miRNA analyses, transcripts exhibiting $P \leq 0.05$ and fold changes greater than or equal to two were considered to be differentially expressed.

Figure 2 outlines the genomics workflow.

Differential and pathway analysis of gene expression and miRNA data

Gene Expression microarray data were analyzed using the mRNA and miRNA workflows in GeneSpring GX 13.1. Signal intensities for each probe were normalized to 75th percentile without baseline transformation for gene expression analysis. No normalization and baseline transformation was performed for miRNA analysis. Differential gene and miRNA expression analysis was performed using the Filter on Volcano Plot option in GeneSpring. The analysis was carried out using a Moderated T-test unpaired statistical method with the Benjamini Hochberg FDR method; p-values were computed asymptotically. The differential miRNA list was used to identify the gene targets by using TargetScan, a miRNA target identification database incorporated within GeneSpring. The validated targets of the differential miRNAs were identified using miRWalk[1].

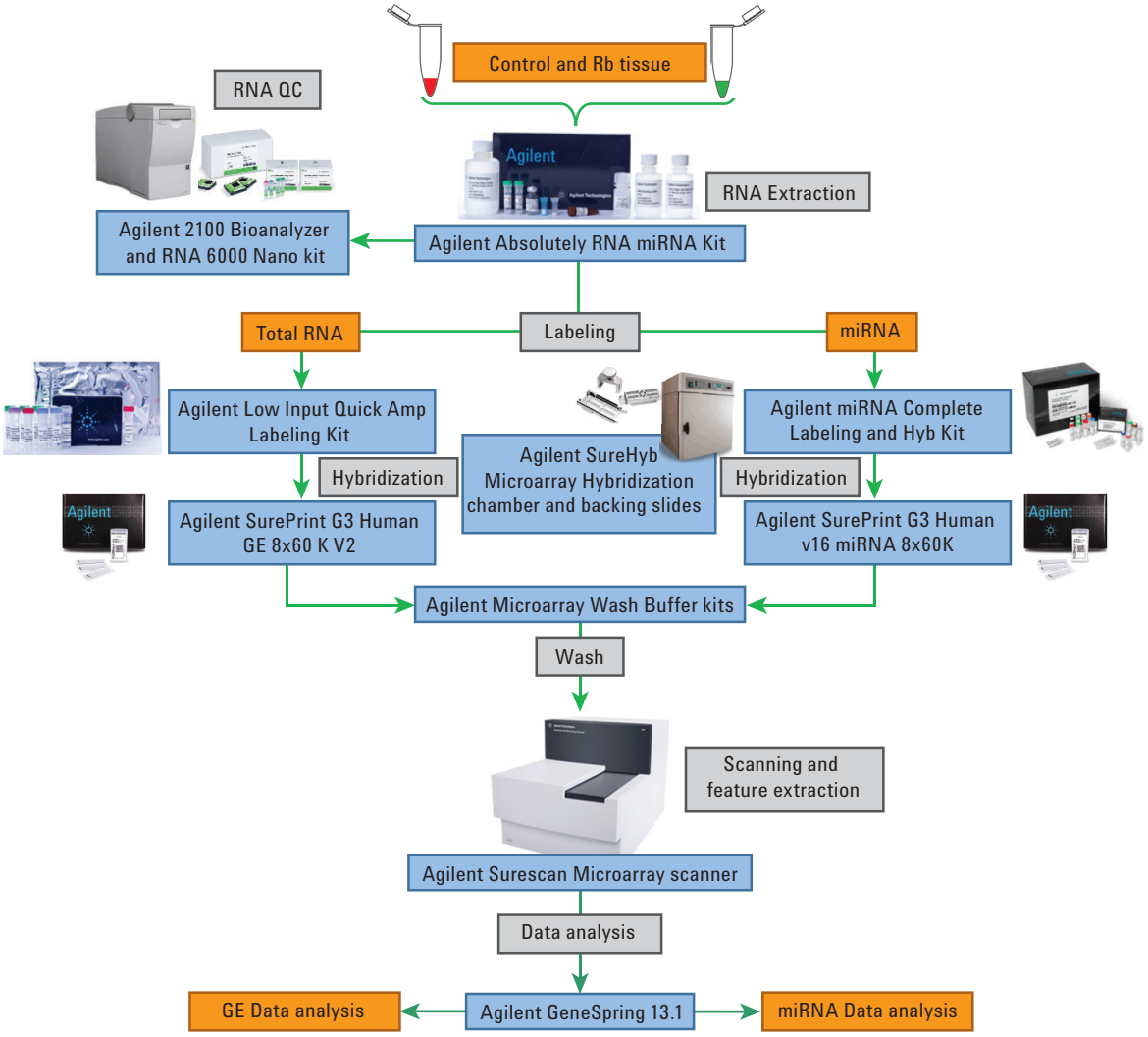

Figure 2. Schematic showing gene expression and miRNA workflow. Agilent products used in the workflow are highlighted in blue boxes.

Pathway analysis was carried out using the Pathway Architect Module in GeneSpring 13.1. The differentially expressed gene entity list ($p \leq 0.05$ and fold change ≥ 2.0) was selected for pathway analysis. The list of validated targets was used for miRNA pathway analysis. Curated pathways from KEGG were used for pathway analysis.

Metabolomics analysis using LC/MS and GC/MS

The metabolites from each of the samples were extracted using a mixture of methanol and ethanol (1:1 v/v). A 500 µL aliquot of this solvent mixture was added to 25 µL of each aqueous humor/vitreous humor/tear sample with appropriate amounts of internal standards, glucose (1-^{13}C), tyrosine (1-^{13}C), cholesterol (3,4-$^{13}C_2$), and palmitic acid (1-^{13}C). The extract containing internal standards was vortexed for 20 seconds followed by vigorous shaking for 15 minutes at 4 °C. The samples were then centrifuged at 4 °C for 5 minutes at 10,000 rpm. Two 180 µL aliquots of this supernatant were placed into separate glass vials and dried. One vial was used for GC/MS analysis, and the other vial was used for LC/MS analysis. Figure 3 shows a summary of the workflow for monophasic extraction followed by analysis using LC/MS and GC/MS platforms.

LC/MS sample preparation and data analysis

To the dried sample, 60 µL of 0.5 ppm epicatechin in methanol was added, and vortexed for 10 seconds followed by sonication for 30 seconds. Another 60 µL of 0.2 % acetic acid in water was added to the sample, vortexed for 10 seconds, and sonicated for 30 seconds. The reconstituted samples were subjected to MS and data-dependent MS/MS acquisition using an Agilent 1290 Infinity LC System coupled to an Agilent 6550 Accurate mass QTOF LC-MS system with a dual Agilent Jet Stream source. The Q-TOF was tuned for low mass and fragile molecule tune.

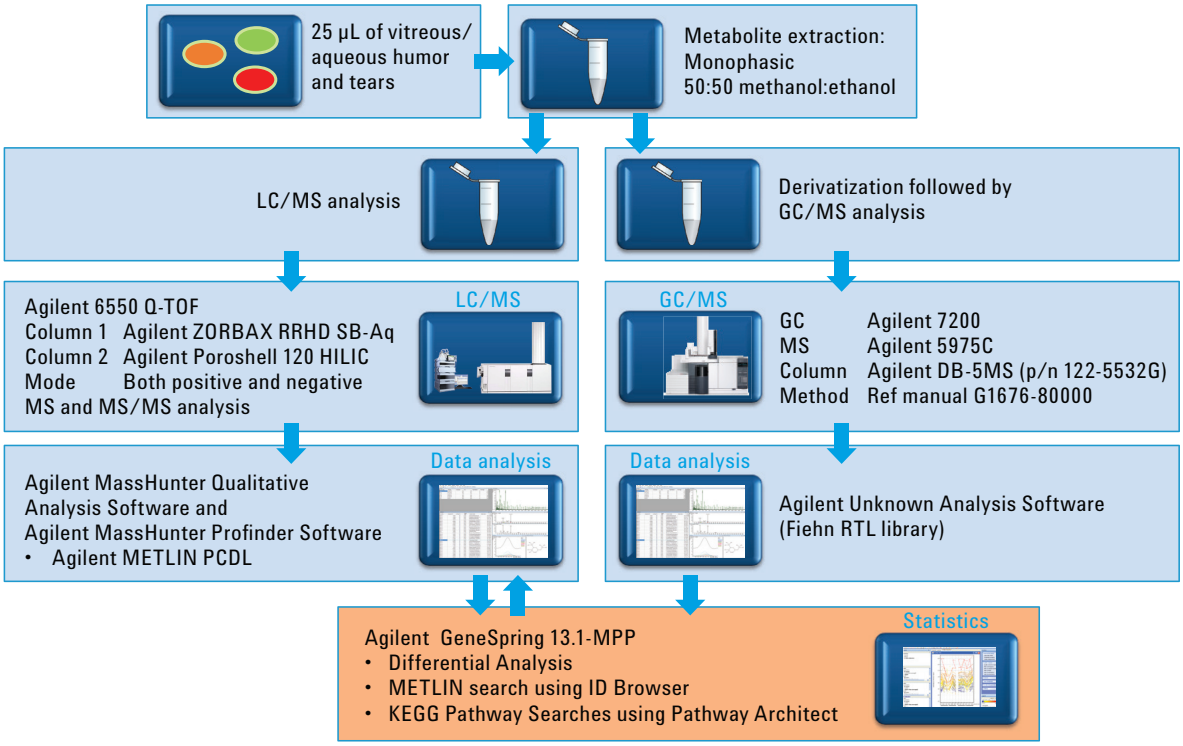

Figure 3. Summary of the workflow for monophasic extraction followed by LC/MS and GC/MS.

The reference solution was prepared in ethanol:acetonitrile:water (750:200:50(v/v)) containing 0.1 % acetic acid by adding an appropriate amount of Agilent HP921 reference solution. The reference solution was added continuously during the run using an external pump at the flow rate of 0.1 mL/min. Ions at *m/z* 64.0158 and *m/z* 922.0098 were used as reference ions for calibration in positive mode. Ions at *m/z* 68.9957, *m/z* 119.0363, and *m/z* 980.0164 were used as reference ions in negative mode. Data were acquired using electrospray ionization (ESI) in positive and negative ion modes using modified polar reverse phase C18 and HILIC columns. Tables 1A and 1B show the details of chromatographic and mass spectrometric parameters. Features were found using Agilent MassHunter Profinder, and were searched against the Agilent METLIN database and confirmed by matching them against the Agilent METLIN MS/MS library.

Agilent MassHunter Profinder (v. B.06.00) software was used for processing LC/MS data. The accurate mass MS data were processed using the recursive MFE tool in the Profinder software. The aligned compounds were exported to an Agilent Mass Profiler Professional module of the Agilent GeneSpring Multi-Omics Analysis software. Agilent Mass Profiler Professional software (MPP) was used for statistical comparison of the LC/MS data from Rb and control samples. The differential list was exported as a .cef file. The .cef file, which contains the mass and retention time of the differential compounds data, was used as a database to process the data-dependent MS/MS files. The MS/MS data were processed using the Find by Formula algorithm of the Agilent MassHunter Qualitative Analysis Software (v B.07.00), pulling out those features found only in the differential list. The spectral pattern generated was searched against the Agilent METLIN MS/MS library.

Table 1A. The chromatographic parameters used in the LC/MS analysis.

LC parameters				
Column	Agilent ZORBAX RRHD SB-Aq 3.0 × 50 mm, 1.8 µm column (p/n 857700-314)		Agilent Poroshell 120 HILIC Plus, 3.0 × 50 mm, 2.7 µm column (p/n 699975-301)	
Ionization mode	Positive and negative MS and MS/MS		Positive and negative MS and MS/MS	
Mobile phase	A) Water with 0.2 % acetic acid B) Methanol with 0.2 % acetic acid		A) (9:1) Acetonitrile:water with 50 mM ammonium acetate B) (5:4:1) Acetonitrile:water:water with 50 mM ammonium acetate	
LC Gradient	Time (min)	% mobile phase B	Time (min)	% mobile phase B
	1.0	5.0	3.0	0.0
	12.0	35.0	12.0	100.0
	13.0	95.0	15.0	100.0
	15.0	95.0	15.1	0.0
	15.1	5.0	20.0	0.0
	20.0	5.0		

Table 1B. The MS source parameters used in MS and MS/MS data acquisition.

Agilent 6550 MS parameters	
Injection volume	5 µL
Flow rate	0.3 mL/min
Thermostated column temperature	40 °C
Gas temperature	175 °C
Drying gas flow	15 L/min
Nebulilzer	45 psig
Sheath gas temperature	200 °C
Sheath gas flow	12 L/min
VCap	3,500 V
Nozzle voltage	1,000 V
Fragmentor	175
Skimmer	65
Octopole RF peak	750
Min range	50 *m/z*
Max range	1,200 *m/z*
MS Scan rate in MS mode	1 spectra/sec
MS Scan rate in MS/MS mode	6 spectra/sec
MS/MS scan rate	3 spectra/sec
Isolation width (MS/MS)	Medium (~4 amu)
Collision energy	10, 20, and 40 V

GC/MS sample preparation and data analysis

The dried samples were subjected to derivatization as described elsewhere[2]. An Agilent Fiehn GC/MS Metabolomics Standards Kit (p/n 400505) was used to perform derivatization. An Agilent 7890C GC coupled to an Agilent 7200 GC/Q-TOF system was used for data acquisition. D27 myristic acid was used for retention time locking. Data were acquired using an EI source and an Agilent DB-5ms column (p/n 122-5532G). Table 2 shows the GC/MS chromatographic and acquisitions parameters.

The 7200 GC/Q-TOF data were processed using MassHunter Unknowns Analysis Software (version B.07.00). This software uses mass spectral deconvolution, which automatically finds peaks and deconvolutes spectra from coeluting compounds using model ion traces. The spectral information was matched with the retention time locked Agilent-Fiehn library and retention time index with respect to FAME mix (Agilent Fiehn GC-MS Metabolomics Standards Kit, p/n 400505). The matched compound list was exported to MPP for further analysis.

Protein estimation

Three microliters of each ocular fluid sample was added to 7 µL of water containing 0.9% sodium chloride to estimate the protein concentration using the Bio-Rad DC Protein Assay following the manufacturer guidelines. A standard curve was generated using BSA, and the total protein of the extract was determined using UV measurement at 750 nm. This protein concentration was used to normalize the LC/MS metabolomics data during MPP analysis.

Multi-omics data analysis

The metabolomics and gene microarray results were combined and analyzed using a pathway-centric approach. Genomics and metabolomics data were covisualized in the pathway context using the Multi-Omics Analysis tool of GeneSpring 13.1. This enabled simultaneous viewing of the differential entities from both gene expression and metabolomics. The data integration of different omics data in GeneSpring helps to gain a better understanding of the interrelationship between changes in expression of individual biochemical entities.

Reagents and materials

LC/MS grade isopropanol, methanol, and acetonitrile were purchased from Fluka (Germany). Milli-Q water (Millipore Elix 10 model, USA) was used for mobile phase preparation. The additives, ammonium fluoride, acetic acid, ammonium formate, formic acid, and ammonium acetate, were procured from Fluka (Germany).

Table 2. The Agilent 7200 GC/Q-TOF chromatography and MS experimental parameters.

GC parameters	
Column	Agilent DB-5MS, 30 mm × 0.2 5mm, 0.25 µm, Guard column 10 m (p/n 122-5532G)
2nd transfer column	Deactivated fused silica, 0.7 m × 150 µm, 0 µm (p/n 122-5532) at constant pressure of 3 psi
Injection volume	1 µL
Inlet	Multimode operated in split less mode
Inlet temperature	250 °C
Carried gas and flow	Helium at 1.8636 mL/min, contact flow
Oven temperature program	60 °C for 1 minute 10 °C/min to 325 °C 10 minute hold
Transfer line temperature	290 °C
Agilent 7200 Q-TOF parameters	
Ionization mode	EI
Source temperature	230 °C
Quadrupole temperature	150 °C
m/z scan	50 to 600 *m/z*
Spectral acquisition rate	5 spectra/sec, 2,679 transients/spectrum

Results and Discussion

mRNA

Approximately 1,600 genes were differentially expressed ($p \leq 0.05$ and $FC \geq 10$) between control and Rb samples. Pathway analysis revealed many key pathways to be differentially altered in Rb affected samples. Figure 4 displays many key entities of the photo transduction pathway to be differentially expressed. The pathway in GeneSpring is illustrated in two ways: first by coloring the rectangles surrounding the differentially expressed genes, and second by displaying histograms near entities to indicate relative levels of genes between Rb and controls. For example, Figure 4 shows down-regulation of transducine (Gt), with red and blue bars representing the expression levels in control and Rb samples, respectively.

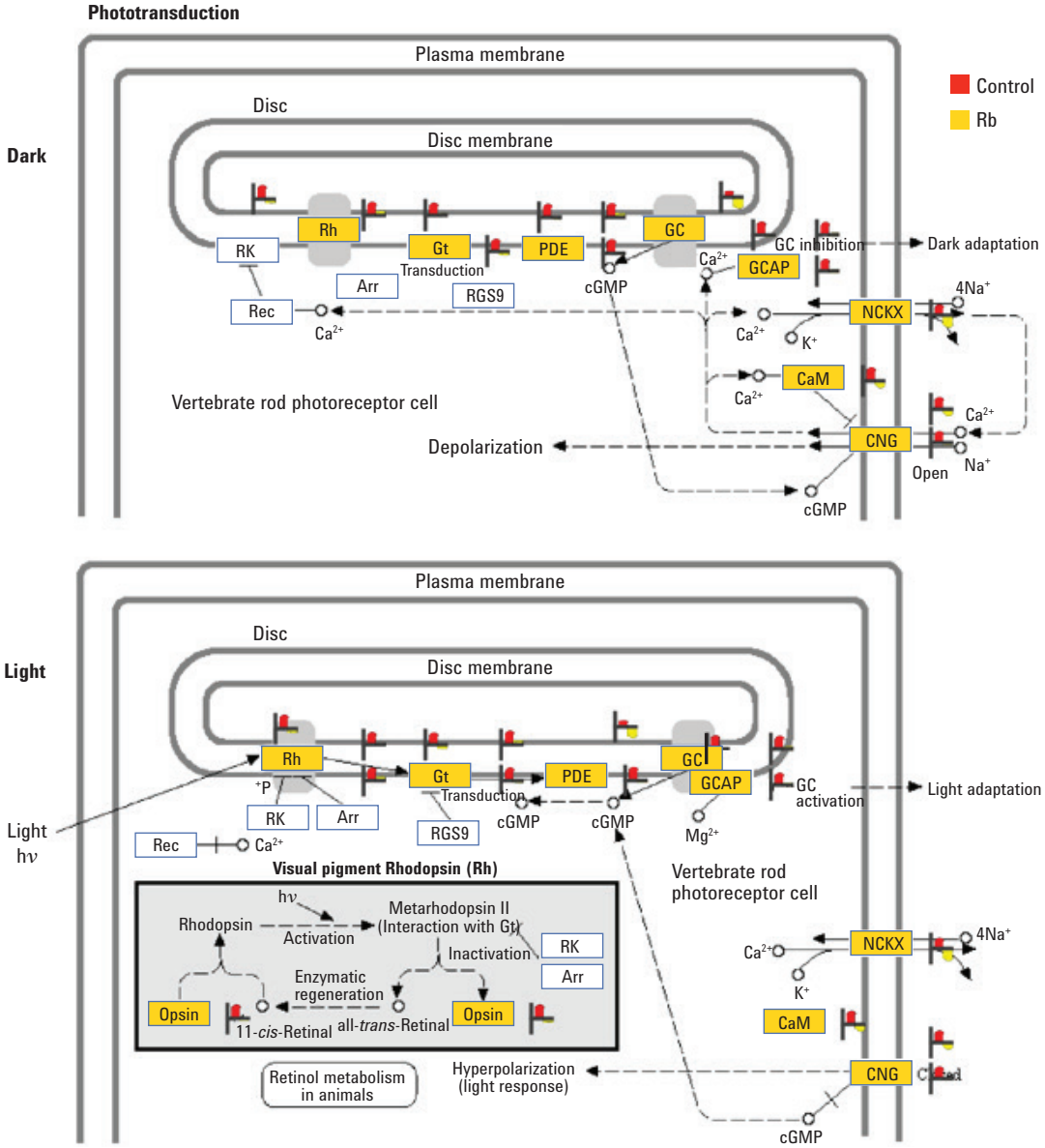

Figure 4. Gene expression pathway analysis showing the photo transduction pathway to be significantly affected.

miRNA

We could identify 18 miRNAs to be significantly regulated ($p \leq 0.05$ and $FC \geq 10$) in Rb samples relative to the controls. Figure 5 displays a volcano plot showing differentially expressed miRNAs.

Metabolomics analysis

Metabolomics analysis was conducted using both LC/MS (positive and negative, HILIC and C18 columns) and GC/MS methods. LC/MS and GC/MS produce nonoverlapping results, allowing different classes of compounds to be identified with confidence. For LC/MS, the choice of two different columns along with both ionization modes gave broader coverage of metabolites. HILIC columns are an ideal choice for the analysis of polar compounds, and C18 columns for nonpolar compounds. The metabolomics experiments were conducted on aqueous humor, vitreous humor, and tear samples from the same set of samples used for gene expression and miRNA.

Of the 102, 117, and 75 total entities identified from aqueous, vitreous, and tear samples, respectively by GC/MS, 32, 32, and 23 differential metabolites were identified by spectral search against the Agilent Fiehn GC/MS library. The identified metabolites included mostly carbohydrates and amino acids. GC/MS and gene expression multi-omics analysis was performed using the differential gene entity list from Rb and control tissues samples, and identified differential metabolite list from aqueous humor, vitreous humor, and tear samples independently. Table 3 shows the predominate pathways revealed by combined GC/MS and gene expression multi-omics in tears.

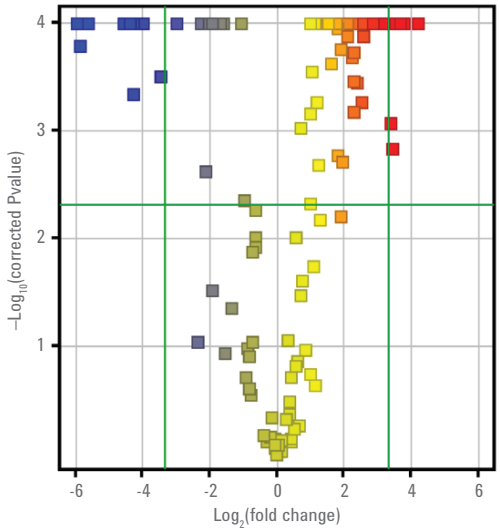

Figure 5. Volcano plot showing 18 differentially regulated miRNA between Rb and controls. Of the 1,000 validated and reported targets of these 18 miRNAs, 12 were found to be differentially expressed in the mRNA data from this study.

Table 3. Predominant pathways revealed by combined GC-MS and gene expression multi-omic analysis in tears

Pathway	p-value (Gene expression)	Matched entities (Gene expression)	Pathway entities of experiment type (Gene expression)	Matched entities (Tears_GC_ Normalization)	Pathway entities of experiment type (Tears_ GC_Normalization)
Fructose and mannose metabolism	0.0882	3	32	3	53
Starch and sucrose metabolism	0.0651	4	56	3	51
ABC transporters	0.1777	3	44	3	122
Aminoacyl-tRNA biosynthesis	0.7726	1	66	2	52
Protein digestion and absorption	0.3250	4	89	2	47
Carbohydrate digestion and absorption	0.0031	6	45	2	27

Pathway analysis and visualization can be extended to multi-omics analysis as well. Figure 6 shows the combined analysis of transcriptomics data from tissues and metabolomics using GC/MS from aqueous humor of the same samples visualized on the valine biosynthesis pathway. The histograms next to the yellow bar and the blue bar show the relative levels of genes and metabolites, respectively. GC/MS studies showed high levels of valine to be present in aqueous humor Rb samples. The valine synthesis pathway also shows significantly low levels of the branched chain amino-acid transaminase 1 enzyme (BCAT1)(EC 2.6.1.42), in Rb samples. The deficiency of the enzyme is responsible for the up-regulation of valine, which is also evident from the pathway figure.

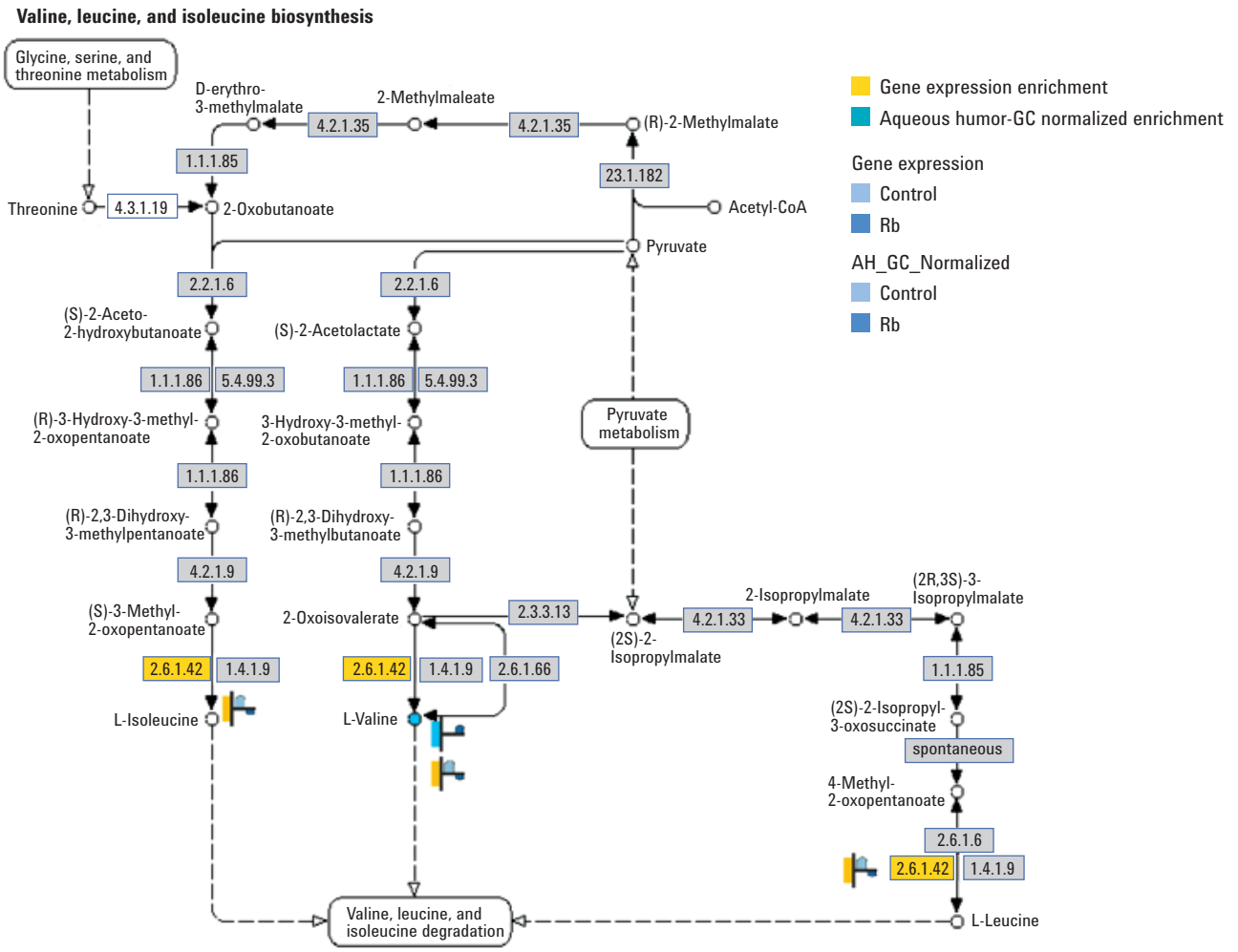

Figure 6. Multi-omics analysis of aqueous humor GC/MS and tissue gene expression reveals the valine, leucine and isoleucine biosynthesis pathway to be significantly down-regulated.

The differentially expressed entities between Rb and control samples across aqueous humor, vitreous humor, and tear samples were matched by searching against the METLIN LC/MS accurate mass database, resulting in approximately 1,300 annotated metabolites. Figure 7 shows the combined multi-omics analysis of LC/MS and gene expression data visualized on the purine metabolism pathway. LC/MS studies of the tear samples showed significantly lower levels of inosine and uric acid in the tears of Rb versus the control group.

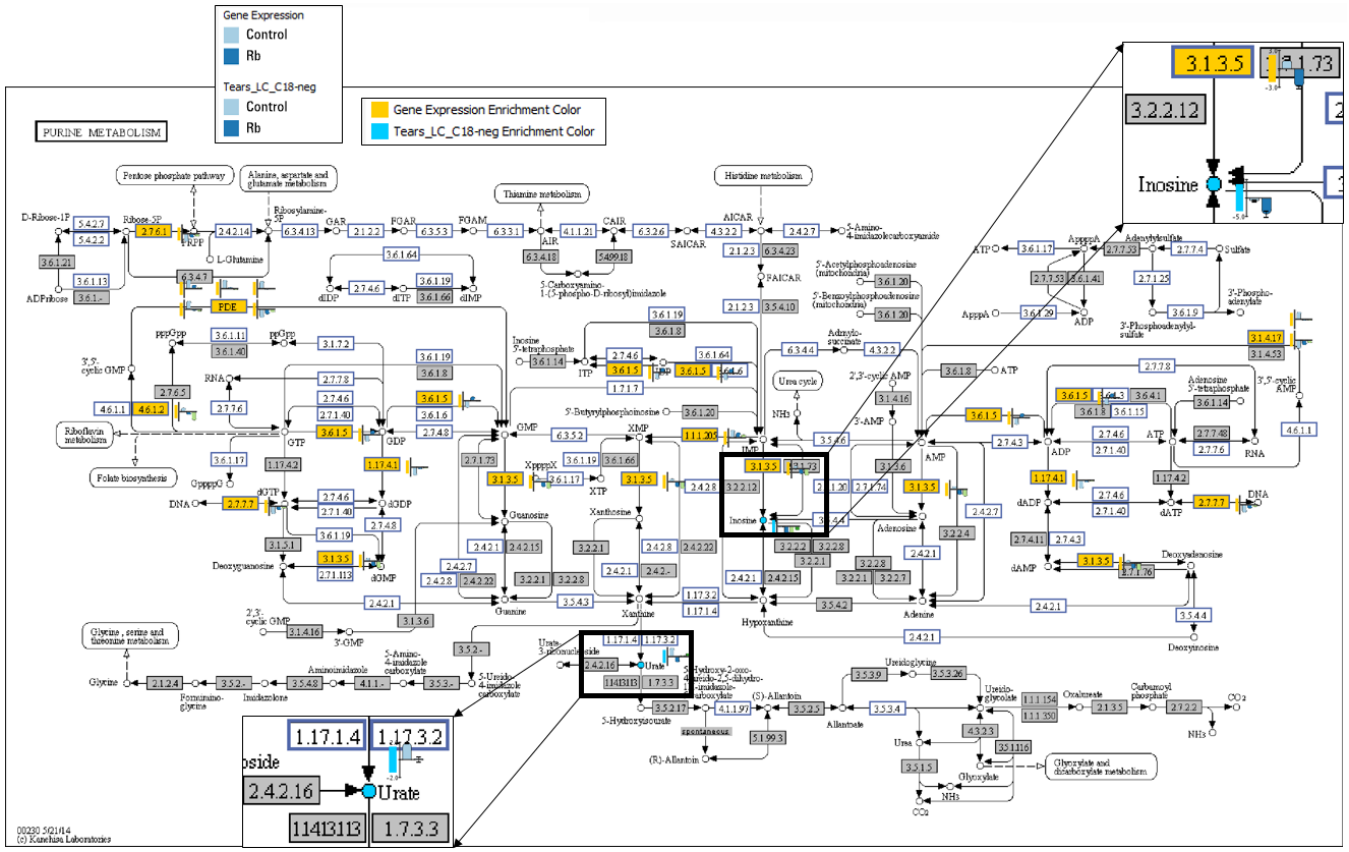

Figure 7. LC/MS and gene expression multi-omics analysis shows purine metabolism to be significantly affected in tumor samples.

Figure 8 shows the MS/MS spectra of inosine as identified and confirmed using the Agilent METLIN MS/MS library. Several associated genes in the pathway were also found to be differentially regulated.

Conclusions

This study demonstrates the value of using a multi-omics approach to understanding the biological pathways in disease progression. The data were acquired using an Agilent multi-omics portfolio including reagents, instrumentation, and software tools designed for transcriptomics and metabolomics. Agilent GeneSpring, Agilent Mass Profiler Professional, and Agilent Pathway Architect software offer new biological insights from complex biological datasets. The software enables researchers to perform discovery experiments by co-analyzing data from different omics such as transcriptomics and metabolomics.

Combining data from transcriptomics and metabolomics revealed previously unknown pathways related to retinoblastoma. For example, the valine, leucine, and isoleucine biosynthesis pathway has never been implicated in retinoblastoma. In addition to the high levels of valine detected by our GC/MS experiments, gene expression results also showed the enzyme branched chain amino-acid transaminase 1 (BCAT1), an enzyme required for valine metabolism, to be significantly lower in Rb samples. Our study opens a possibility for follow-up studies, and demonstrates a unified approach to discovering new biological insights – an approach made possible by integrating data from different omics in the same sample.

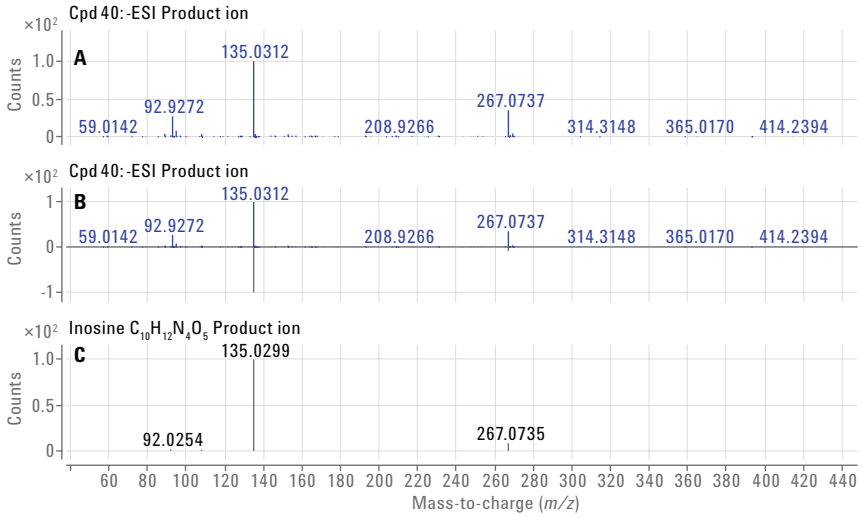

Figure 8. MS/MS Library spectral match of inosine A) spectra taken from the sample, B) mirror plot combining the MS/MS from the library and from the unknown sample C) spectra from Agilent METLIN MS/MS library.

References

1. Dweep, H., *et al.* miRWalk - database: prediction of possible miRNA binding sites by "walking" the genes of 3 genomes, *J. of Biomedical Informatics* **2011**, *44*, pp 839-7.

2. Palazoglu, M., and Fiehn, O. Metabolite Identification in Blood Plasma Using GC/MS and the Agilent Fiehn GC/MS Metabolomics RTL Library. *Agilent Technologies Application Note*, publication number 5990-3638EN, **2009**.

www.agilent.com/chem

Published in the USA, October 1, 2015
5991-6215EN

Agilent Technologies