

Simplified global gene expression profiling

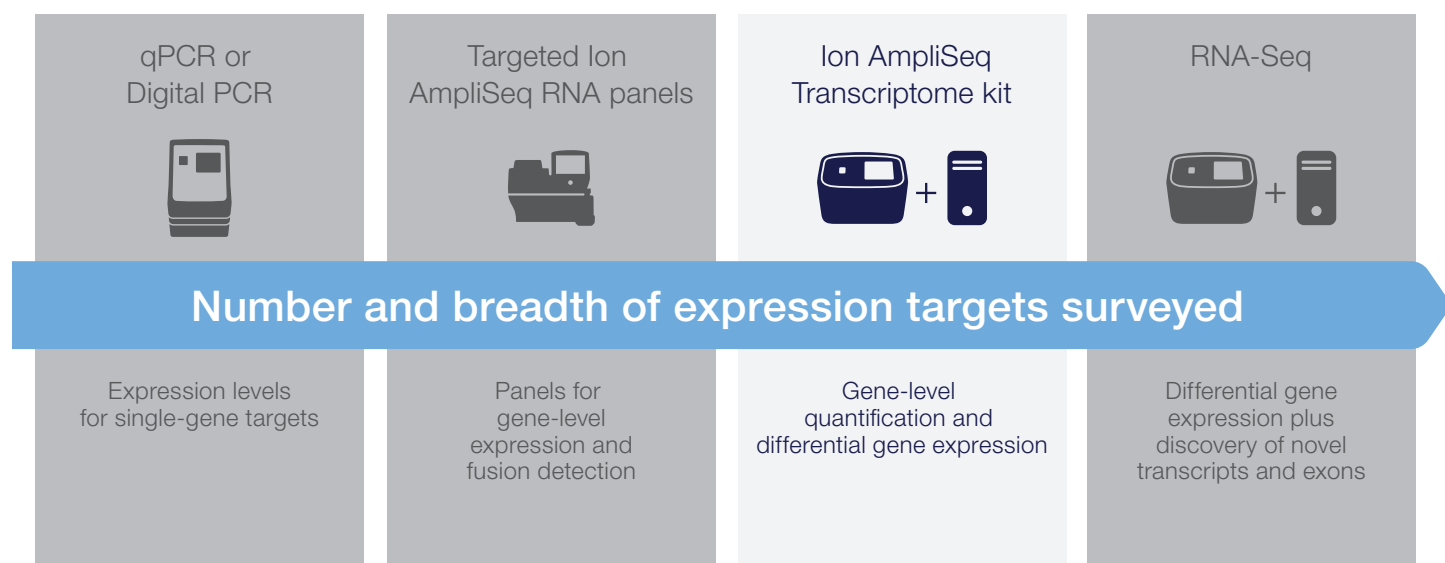


Figure 1. Transcript profiling options. RNA transcript analysis solutions range from multiplex PCR and RNA sequencing of individual targets to unbiased, global transcript profiling by means of whole-transcriptome analysis.

Key findings

- **Highly correlated results**—differentially expressed genes (DEGs) identified using the Ion AmpliSeq™ Transcriptome Human Gene Expression Kit were highly correlated with MicroArray Quality Control (MAQC) data, RNA-Seq, and qPCR data with Pearson correlation coefficients (R) of 0.88, 0.95, and 0.96, respectively, when using the Ion S5™ XL and Ion Chef™ Systems.
- **Simple workflow and analysis**—the simplicity and speed of PCR enables a rapid method for gene expression profiling with a ≤2-day workflow from sample to analysis. Intuitive bioinformatics through point-and-click run setup and automated secondary data analysis in Torrent Suite™ Software with an associated plug-in, the AmpliSeqRNA, enables unambiguous gene expression analysis and visualization from sequence-read data (Figures 7–9, Supplementary information).
- **Robust performance with limited or degraded samples**—the power of Ion AmpliSeq™ technology enables reproducible detection of high- and low-abundance transcripts from difficult samples such as formalin-fixed, paraffin-embedded (FFPE) tissue sections or from samples with limited input—as little as 10 ng total RNA.
- **Cost effective**—reduced cost and complexity of gene expression analysis through the targeted sequencing of 20,802 genes and up to 8 multiplexed samples per Ion 540™ Chip.

Introduction

The power of massively multiplexed amplicon sequencing

Scalable and simple end-to-end workflows enabled by Ion AmpliSeq technology and Ion Torrent™ instruments have established ultrahigh-multiplex PCR and RNA sequencing as a simple and powerful method for discovering variants of interest, measuring gene expression and copy number changes, or detecting gene fusion events. Now Ion AmpliSeq technology has been expanded to profile the gene expression of nearly 21,000 distinct human RNA targets simultaneously in a single primer pool. The Ion AmpliSeq Transcriptome Human Gene Expression Kit completes a continuum of RNA sequencing solutions between targeted Ion AmpliSeq™ RNA panels and unbiased, global transcript profiling by means of whole-transcriptome analysis (Figure 1). Ready-to-use research panels for gene expression analysis include the Ion AmpliSeq™ RNA Apoptosis Panel for gene expression profiling of 267 genes involved in the cellular apoptosis pathway, including genes associated with death receptor-mediated apoptosis, *c-Myc*, and *TP53*-mediated apoptosis.

Custom Ion AmpliSeq RNA panels, targeting up to 1,200 genes, can be created using the free assay design tool, the Ion AmpliSeq™ Designer, for the creation of primer pools targeting any customer-defined region of interest within human and nonhuman reference genomes.

Ion AmpliSeq technology has the additional advantages of accommodating FFPE and low-input RNA samples, down to 10 ng of total RNA input. With the Ion Chef System for automated library and template preparation, only 45 minutes of hands-on time is needed—from cDNA all the way to final gene expression data (Figure 2).

Next-generation sequencing solution for gene expression studies

The Ion AmpliSeq Transcriptome Human Gene Expression Kit covers >95% of human RefSeq genes (UCSC *refGene*) with a single amplicon designed per gene target. Coupled with the simplicity of a PCR-based workflow and an automated analysis solution, the Ion AmpliSeq Transcriptome Human Gene Expression Kit provides an attractive next-generation sequencing (NGS) solution for researchers who have traditionally used expression microarrays to obtain gene expression profiles from both FFPE-derived and high-quality (unfixed) RNA samples. Moreover, compared to microarrays, sequencing-based gene expression profiling provides a digital readout with greater quantitative linear dynamic range and improved detection at the extremes of the quantitation spectrum. In addition, sequencing coverage depth can be adjusted to accommodate gene expression studies with sample size variability.

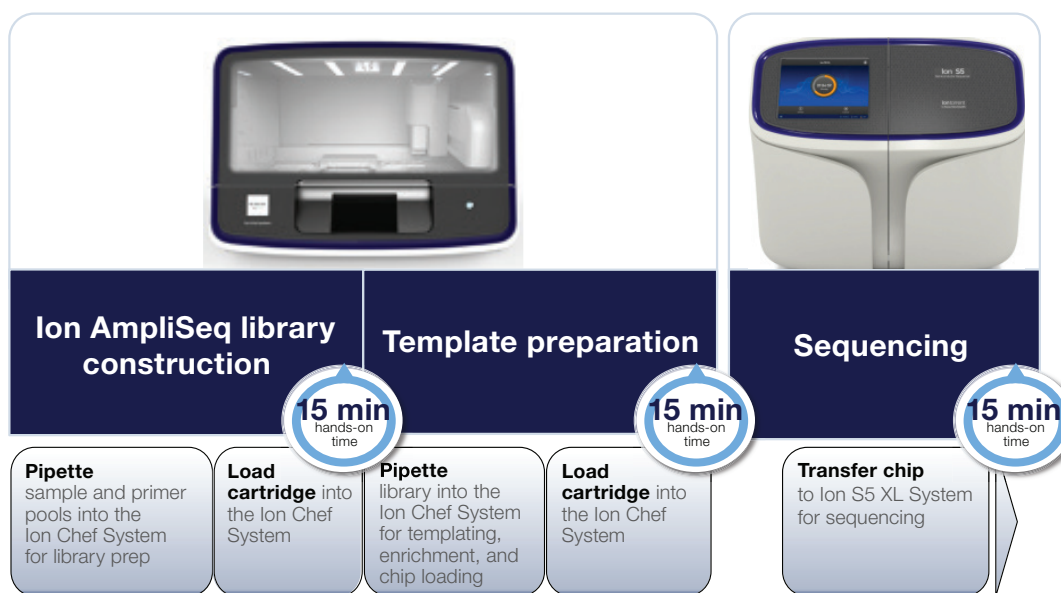


Figure 2. Streamlined workflow for gene expression analysis. Fully automated workflow from cDNA to gene expression data requires only 45 minutes of hands-on time using the Ion Chef System and Ion S5 XL System.

Convenient evaluation of gene expression analysis metrics in Torrent Suite Software

A summary table of mapping statistics per barcode of total mapped reads, percentage on target, and percentage of panel genes detected ("Targets Detected") is viewable in Torrent Suite Software to quickly evaluate run and library performance. Also available are links to the full plug-in output and to tables with absolute and normalized counts per gene, as well as rich visualizations including, gene level expression correlation between samples, clustering of expression profiles, and categorical distributions. Unlike other RNA-Seq analyses that require normalization approaches such as RPKM (reads per kilobase of transcript per million reads mapped), it is unnecessary to normalize data obtained with the Ion AmpliSeq Transcriptome Human Gene Expression Kit to account for transcript length bias. Ion AmpliSeq RNA normalization for a given sample is automatically calculated by the plug-in as the number of reads mapped per gene per million reads mapped or RPM (Figures 7–9 and Table 1, Supplementary information). In addition, the plug-in produces pairwise scatter plots of RPM-normalized gene count correlations for all barcoded samples. These correlations can be used to measure the reproducibility between replicate libraries. Additional plug-in analyses include output for each barcode of the number of genes (amplicons) with at least 1, 10, 100, 1,000, and 10,000 counts to enable determination of the dynamic range and sensitivity per sample.

Methods

RNA extraction

Applied Biosystems™ FirstChoice™ Human Brain Reference RNA (HBRR) and Stratagene™ Universal Human Reference RNA (UHRR) were used to assess the performance of the Ion S5™ XL System and the Ion AmpliSeq Transcriptome Human Gene Expression Kit. The HBRR and UHRR samples were used in the MAQC study, thus allowing comparison of gene expression profiling data from the Ion AmpliSeq Transcriptome Human Gene Expression Kit to the microarray and qPCR data included in the MAQC publication [1]. To assess the sensitivity and reproducibility of results from degraded RNA, total RNA was isolated using a 1 x 10 µm section from 2 different lung FFPE samples using the Applied Biosystems™ MagMAX™ FFPE DNA/RNA Ultra Kit (Cat. No. A31881) with the RNA-only protocol. The matching fresh-frozen (FF) samples were processed with the Invitrogen™ *mirVana*™ miRNA Isolation Kit (Cat. No. AM1561).

Library preparation

10 ng of total RNA was reverse-transcribed using the Invitrogen™ SuperScript™ VILO™ cDNA Synthesis Kit (Cat. No. 11754050). Transcriptome libraries were generated using automated library preparation with the Ion AmpliSeq™ Transcriptome Human Gene Expression Panel, Chef-Ready Kit (Cat. No. A31446) for UHRR and HBRR samples. The kit enables users to generate 8 equalized Ion AmpliSeq™ libraries utilizing the robust liquid handling capabilities of the Ion Chef System. The Ion S5 XL System combined with automated library and template preparation on the Ion Chef System enables a simple solution for gene expression analysis requiring only 45 minutes of hands-on time from 10 ng of FFPE RNA (Figure 2). Libraries from FFPE and fresh-frozen samples were generated using the Ion AmpliSeq Transcriptome Human Gene Expression Kit (Cat. No. A26325). Transcriptome libraries were barcoded, templated, and sequenced on the Ion S5 XL System. Multiplexed libraries were templated using the automated Ion Chef System and sequenced using the Ion 540™ Kit-Chef (Cat. No. A27759) and Ion 540 Chip Kit (Cat. No. A27766) as one 8-plex library pool.

Data analysis

Raw data were processed on the Ion S5 XL System and transferred to the Torrent Server for primary data analysis performed using the AmpliSeqRNA plug-in with gene-level transcript quantification from sequence-read data. This plug-in is included in Torrent Suite Software, which comes with each Ion Torrent™ sequencer. The AmpliSeqRNA plug-in can be automatically executed upon completion of a sequencing run and its output provides the user with simple QC metrics, visualization plots, and normalized counts per gene that corresponds to gene expression information that includes a link to a downloadable file detailing the read counts per gene in a tab-delimited text file. When setting up a sequencing run, the mapping reference and the AmpliSeqRNA plug-in may be configured in the run plan such that all data are automatically mapped and analyzed once the raw sequencing data have been collected. The number of reads aligning to a given gene target represents an expression value referred to as "counts". Seamlessly, and without any extra steps, a researcher can initiate a sequencing run on the Ion S5 XL System, and immediately after completion, the normalized, gene-level count data from that run are available for download or further off-line analysis.

Results

High reproducibility, read alignment, and RefSeq target detection

Multiplexing 8 samples on a single Ion 540 Chip, the Ion AmpliSeq Transcriptome Human Gene Expression Kit demonstrated high technical reproducibility, read alignment, and detection of target transcripts. A total of 85,034,348 and 86,309,879 reads were generated, with a mean read count of 10.1 and 10.4 million per barcode for sequencing runs of libraries from high-quality (UHRR, HBRR) and FFPE-derived (human lung) RNA, respectively. In addition, mean mapping rate for high-quality RNA and FFPE samples was 99.8% and 98.7%, respectively. Output from the AmpliSeqRNA analysis plug-in for each of 4 UHRR and HBRR replicates revealed a very high mapping rate of >98% with an average of 92.2% reads on target and an average of 70% (14,563) targets detected (Table 1A, Supplementary information). Likewise, using compromised RNA samples, 2 replicates each for 2 distinct FFPE samples yielded a high mapping rate of >98% with an average of 76.3% reads on target, and an average of 68% (14,061) targets were detected (Table 1B, Supplementary information). In addition, the FFPE samples have only slightly fewer genes detected, indicating robust gene detection sensitivity from degraded FFPE samples. For evaluating technical reproducibility, pairwise Pearson

correlations of normalized read counts were calculated across high-quality RNA (Figure 3) and FFPE sample replicates. Although there is slightly more variability for less-abundant genes in libraries generated from degraded FFPE samples, the Pearson correlations were >0.91, whereas HBRR and UHRR replicates were highly correlated at >0.99. In addition, correlations of normalized counts between matched FFPE and fresh-frozen lung tissue RNA were >0.9.

Inter-platform concordance for gene-level transcript detection and differentially expressed genes

Inter-platform concordance was used to illustrate the sensitivity of gene expression profiling data obtained using the Ion S5 XL System and Ion AmpliSeq Transcriptome Human Gene Expression Kit. Data from 8 samples multiplexed on a single Ion 540 Chip were compared to microarray data. The Ion AmpliSeq Transcriptome Human Gene Expression Kit enabled the detection of a significantly higher number of genes, >15,900 targets for UHRR and HBRR samples, compared to >9,700 detected by microarray analysis (Table 2, Supplementary information). Between the two gene lists, 9,393 were in common between the platforms, with 6,564 detected by the Ion AmpliSeq Transcriptome Human Gene Expression Kit only. In contrast, only 322 genes were uniquely detected by microarray (Figure 4).

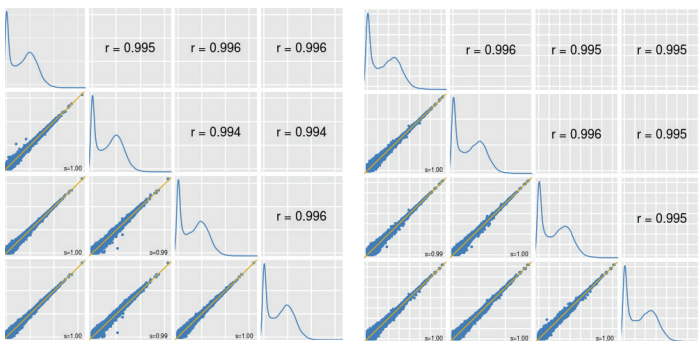


Figure 3. Correlation of technical replicates from Ion AmpliSeq Transcriptome Human Gene Expression Kit-mapped reads using the Ion S5 XL System. Scatter plot comparisons of \log_2 RPM gene counts from the RefSeq set are shown illustrating a wide dynamic range spanning 5 \log_{10} units (2^{17}), sufficient to measure the full range of expression in cells. All pairwise comparisons between technical replicates for high-quality RNA libraries from UHRR (left) and HBRR (right) on the Ion 540 Chip demonstrated a Pearson correlation coefficient (R) of ≥ 0.994 .

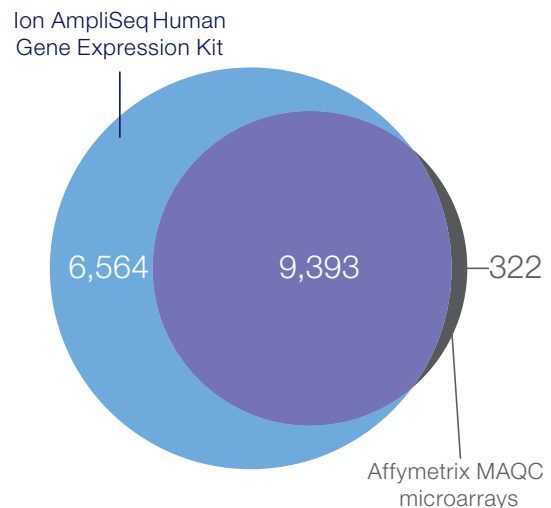


Figure 4. Improved gene-level transcript detection using the Ion AmpliSeq Transcriptome Human Gene Expression Kit and the Ion S5 XL System. A Venn diagram is shown demonstrating the concordance of genes identified by MAQC microarray data and the Ion AmpliSeq Transcriptome Human Gene Expression Kit.

While gene-level detection describes the sensitivity for a given platform, DEGs between HBRR and UHRR samples were determined to gain insight into quantitative accuracy of relative expression analysis. With reference to DEGs, the relative gene expression by Ion AmpliSeq technology-based detection was compared to data from Illumina™ whole-transcriptome RNA-Seq from the SEQC study [2], microarrays, and qPCR. A total of 9,393 RefSeq genes were differentially expressed and shared between the MAQC microarray and Ion AmpliSeq Transcriptome Human Gene Expression Kit datasets; of these, 8,719 genes were identified by the Ion AmpliSeq Transcriptome Human Gene Expression Kit to be significantly differentially expressed vs. 5,428 in the MAQC microarray data. For a gene to be considered significantly differentially expressed, the fold change between UHRR and HBRR samples had to be ≥ 2 with a P value ≤ 0.05 . A scatter plot comparing the fold change demonstrates that the datasets were correlated,

with a Pearson correlation (R) of 0.88 between the Ion AmpliSeq Transcriptome Human Gene Expression Kit data and microarrays (Figure 5A).

A fold-change comparison of Ion AmpliSeq Transcriptome Human Gene Expression Kit data and qPCR data from the MAQC study (a set of 754 Applied Biosystems™ TaqMan® Gene Expression Assays) also showed high correlation, with a Pearson correlation coefficient (R) of 0.96 (Figure 5B). A total of 14,643 differentially expressed RefSeq genes were shared between Illumina™ RNA-Seq and Ion AmpliSeq Transcriptome Human Gene Expression Kit datasets. A scatter plot comparing the fold change demonstrates that the datasets were strongly correlated, with a Pearson correlation (R) of 0.95 between the Ion AmpliSeq Transcriptome Human Gene Expression Kit count data and Illumina RNA-Seq data (Figure 5C).

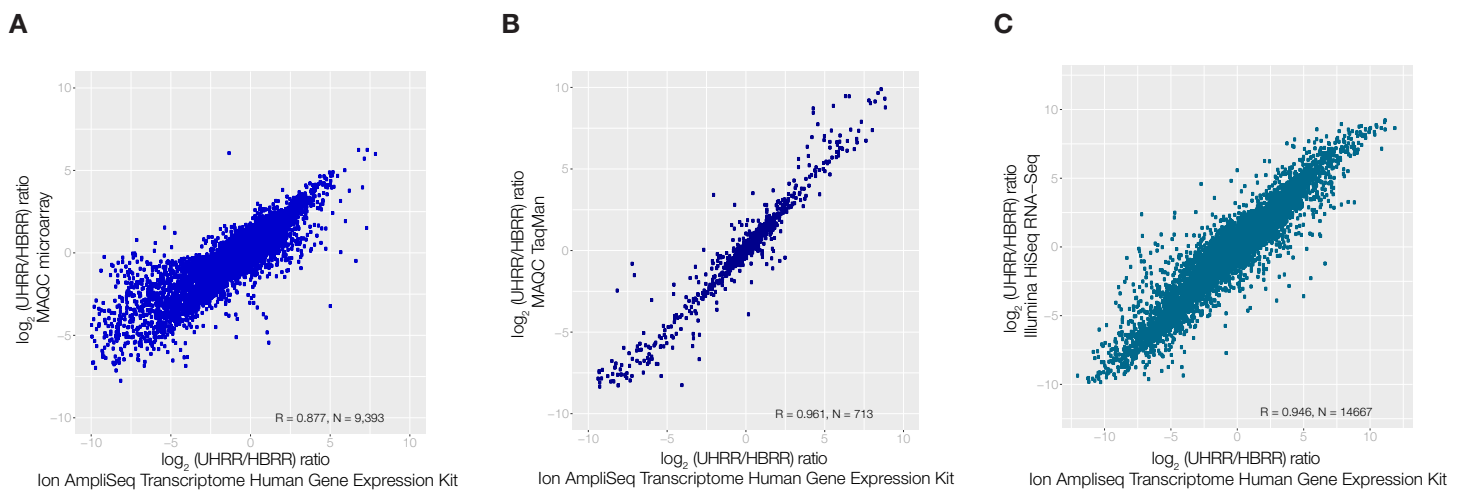


Figure 5. Representation of strong correlation for DEGs among Ion AmpliSeq Transcriptome Human Gene Expression Kit data with whole-transcriptome data, relative to MAQC array and qPCR data. Data from 8 UHRR and HBRR samples multiplexed on a single Ion 540 Chip were used to calculate differential gene expression for Ion AmpliSeq Transcriptome Human Gene Expression Kit data. DEGs between HBRR and UHRR samples for both Illumina™ HiSeq™ whole-transcriptome RNA-Seq and Ion AmpliSeq Transcriptome Human Gene Expression Kit datasets were determined with the DESeq package [3]. Affymetrix™ microarray data were normalized and expression values were calculated using PLIER. All microarray calculations reflect the same methods used in the MAQC publication [1]. **(A)** Scatter plot comparison of \log_2 (UHRR/HBRR) ratios from the Ion AmpliSeq Transcriptome Human Gene Expression Kit and MAQC microarray expression data. A total of 9,393 RefSeq genes had detectable differential expression on both platforms; the datasets were highly correlated, with Pearson correlation coefficients (R) of 0.88 and 0.91 for Ion AmpliSeq gene expression kit and MAQC data, respectively. **(B)** Scatter plot comparison of \log_2 (UHRR/HBRR) ratios from the Ion AmpliSeq Transcriptome Human Gene Expression Kit and MAQC qPCR data. Differential expression for 713 genes was compared for the two platforms with a Pearson correlation coefficient (R) of 0.96, demonstrating that the qPCR and Ion AmpliSeq Transcriptome Human Gene Expression Kit datasets are highly correlated. **(C)** Scatter plot comparison of \log_2 (UHRR/HBRR) ratios from Illumina HiSeq whole-transcriptome RNA-Seq and Ion AmpliSeq Transcriptome Human Gene Expression Kit data. For whole-transcriptome data, the transcript detection threshold was ≥ 10 counts. Differential expression for the two methods demonstrated a Pearson correlation coefficient (R) of 0.95, demonstrating that whole-transcriptome RNA-Seq and Ion AmpliSeq Transcriptome Human Gene Expression Kit datasets are highly correlated.

The Venn diagram in Figure 6 illustrates that, out of 9,770 significant DEGs detected by either method, 45% (4,387) of the Ion AmpliSeq Transcriptome Human Gene Expression Kit DEGs were concordant with microarray DEGs, while 44% (4,332) were detected only by Ion AmpliSeq Transcriptome Human Gene Expression Kit sequencing. In contrast, only 10% (1,051) of DEGs detected by microarray were not detected by Ion AmpliSeq Transcriptome Human Gene Expression Kit sequencing.

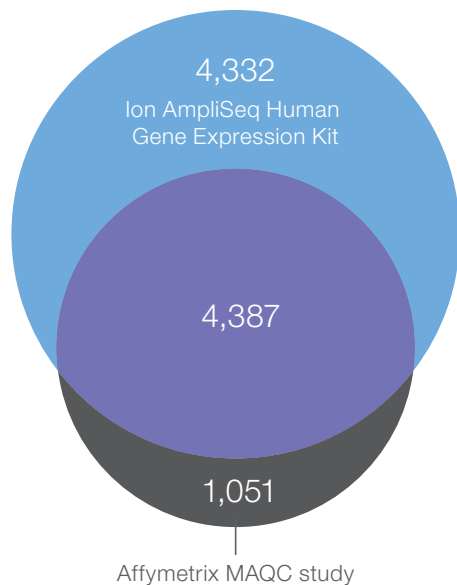


Figure 6. Detection of significantly differentially expressed genes (DEGs) by using the Ion AmpliSeq Transcriptome Human Gene Expression Kit. The Venn diagram shows the concordance of significant DEGs between MAQC data and Ion AmpliSeq Transcriptome Human Gene Expression Kit data.

Conclusion

The Ion AmpliSeq Transcriptome Human Gene Expression Kit, in combination with the Ion Chef and Ion S5 XL Systems, enables a simple workflow solution to simultaneously measure the expression of human genes in a massively multiplex panel from a single primer pool. This automated solution enables researchers to move from cDNA to gene expression data in just 45 minutes of hands-on time and with as little as 10 ng RNA input.

Multiplexing 8 samples (8–12 million reads/library) per Ion 540 or Ion PI™ Chip minimizes the cost of differential gene expression. The dynamic range exceeds microarray sensitivity, enabling the detection of gene expression levels spanning more than 5 log₁₀ units, sufficient to measure the range of expression in most cells. The Ion AmpliSeq Transcriptome Human Gene Expression Kit exceeded microarray sensitivity for gene-level transcript detection and, using MAQC samples for comparison, the detected expression levels of differentially expressed genes were highly correlated with results from microarrays and qPCR. In addition, the Ion AmpliSeq Transcriptome Human Gene Expression Kit enabled the reproducible detection of high- and low-abundance transcripts from difficult samples such as FFPE tissue sections. With the simplicity and speed of PCR incorporated into the sequencing workflow and accessible bioinformatics in Torrent Suite Software, researchers can confidently go from RNA sample to differential gene expression analysis in less than 2 days.

References

1. Shi L, Reid LH, Jones WD et al. (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 24:1151–1161.
2. SEQC/MAQC-III Consortium (2014) A comprehensive assessment of RNA-Seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat Biotechnol* 32:903–14.
3. Gentleman RC, Carey VJ, Bates DM et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5:R80.

Supplementary information*

Table 1. Ion AmpliSeq Transcriptome Human Gene Expression Kit sequencing results from (A) UHRR and HBRR samples and (B) FFPE samples, using the Ion S5 XL System and the Ion 540 Chip. Sequence reads were aligned to the Ion AmpliSeq Transcriptome reference file (hg19_AmpliSeq_Transcriptome_ERCC_v1.fasta) available in Torrent Suite Software using the Ion Torrent Mapping Alignment Program (TMAP) (<https://github.com/iontorrent/TMAP>). The reference file contains the entire set of RefSeq transcripts from which all 20,802 Ion AmpliSeq Transcriptome panel primer pairs were designed. Following alignment, the AmpliSeqRNA plug-in examines the number of reads mapping to the expected amplicon ranges and assigns counts per gene for reads that align to these regions as defined in the BED file also available in Torrent Suite Software (hg19_AmpliSeq_Transcriptome_21K_v1.bed). The Ion AmpliSeq Transcriptome BED file is a formatted file containing the nucleotide positions of each amplicon per transcript in the mapping reference. Reads aligning to the expected amplicon locations and meeting filtering criteria such as minimum alignment length are reported as percent “valid” reads in the these summary tables generated by the AmpliSeqRNA plug-in. “Targets Detected” is defined as the number of genes detected (≥ 10 read counts) as a percentage of the total number of gene targets in the Ion AmpliSeq Transcriptome panel (20,802 genes). As shown in (B), FFPE and FF samples correspond to FFPE lung and fresh-frozen lung samples, respectively.

A. UHRR and HBRR samples

Barcode Name	Sample	Mapped Reads	Valid Reads	Targets Detected
IonCode_0125	HBRR rep 1	11,103,007	92.75%	70.74%
IonCode_0126	HBRR rep 2	9,607,324	92.33%	70.13%
IonCode_0127	HBRR rep 3	9,883,575	93.16%	69.58%
IonCode_0128	HBRR rep 4	10,412,331	93.09%	70.06%
IonCode_0129	UHRR rep 1	9,239,764	91.66%	67.40%
IonCode_0130	UHRR rep 2	10,653,629	91.57%	68.40%
IonCode_0131	UHRR rep 3	10,576,488	91.54%	68.13%
IonCode_0132	UHRR rep 4	9,116,596	91.97%	67.03%

B. FFPE and FF samples

Barcode Name	Sample	Mapped Reads	Valid Reads	Targets Detected
IonXpress_009	FFPE A rep 1	11,187,122	84.66%	67.18%
IonXpress_010	FFPE A rep 2	12,276,062	79.42%	69.49%
IonXpress_011	FFPE B rep 1	9,961,060	70.93%	67.59%
IonXpress_012	FFPE B rep 2	7,928,665	70.04%	65.99%
IonXpress_013	FF A rep 1	10,637,546	94.68%	65.32%
IonXpress_014	FF A rep 2	10,596,244	94.58%	65.45%
IonXpress_015	FF B rep 1	9,797,036	90.42%	73.75%
IonXpress_016	FF B rep 2	11,040,279	90.17%	74.72%

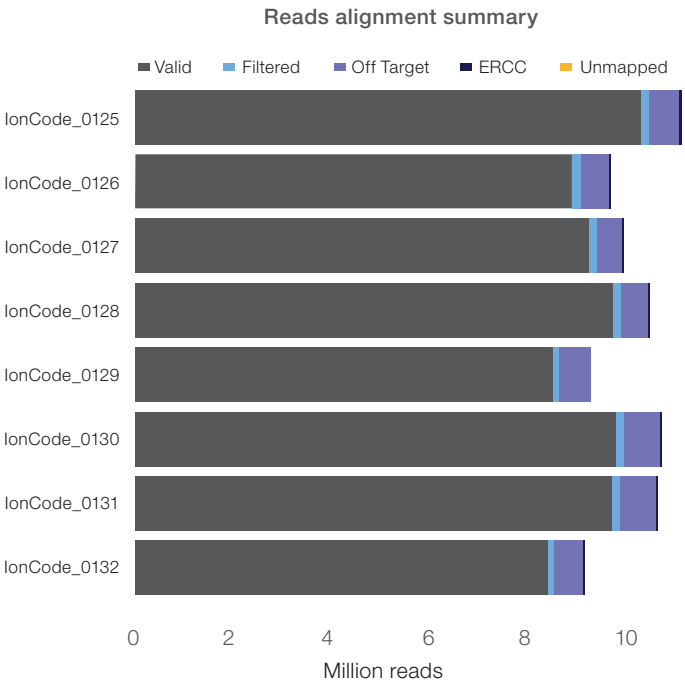


Figure 7. A graphical summary of the number of all read assignment categories across barcodes. Data represented here are from UHRR and HBRR reference standard materials.

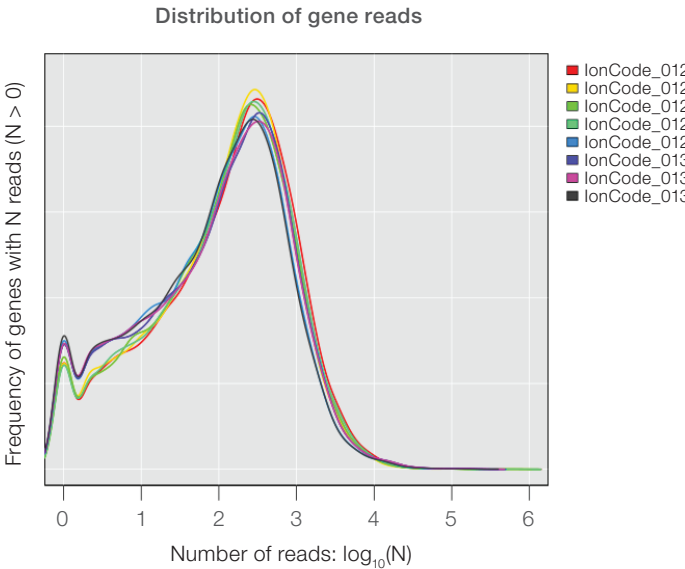


Figure 8. Density-based frequency distributions of genes across barcodes showing the frequency of number of genes having similar \log_{10} read counts. Data shown here are from UHRR and HBRR reference standard materials. Barcodes with more genes represented will have proportionally the larger areas under these curves.

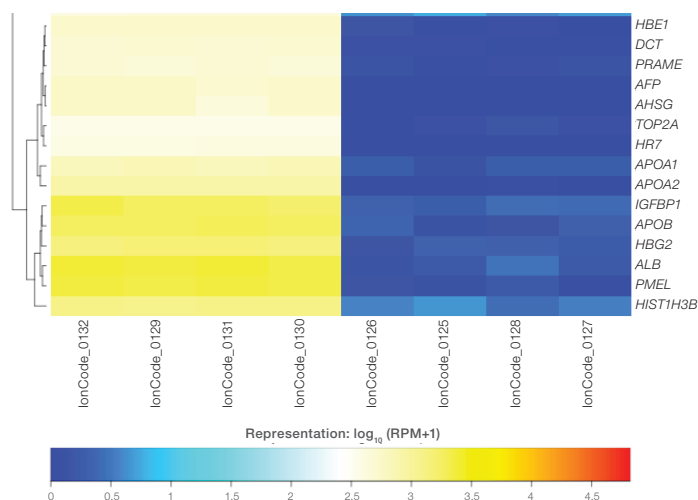


Figure 9. Gene representation heatmap. The most variation in gene representation is shown across barcodes as measured by the coefficient of variation (CV) of normalized read counts for genes that have at least one barcode with at least 100 RPKM reads, plotted using \log_{10} of those counts.

* All figures in the 'Supplementary information' section were generated by the AmpliseqRNA plug-in.

Ordering information

Product	Cat. No.
Ion AmpliSeq Transcriptome Human Gene Expression Panel, Chef-Ready Kit	A31446
Ion AmpliSeq Transcriptome Human Gene Expression Kit	A26325
SuperScript VILO cDNA Synthesis Kit	11754050
MagMAX FFPE DNA/RNA Ultra Kit	A31881
Ion Chef System	4484177
Ion S5 XL System	A27214
Ion 540 Chip Kit	A27766
Ion 540 Kit-Chef	A27759

Find out more at
thermofisher.com/ampliseqtranscriptome

ThermoFisher
 S C I E N T I F I C

The content provided herein may relate to products that have not been officially released and is subject to change without notice.

For Research Use Only. Not for use in diagnostic procedures. © 2016 Thermo Fisher Scientific Inc. All rights reserved.

All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified. Stratagene is a trademark of Agilent Technologies, Inc. TaqMan is a registered trademark of Roche Molecular Systems, Inc., used under permission and license. **COL21530 1016**