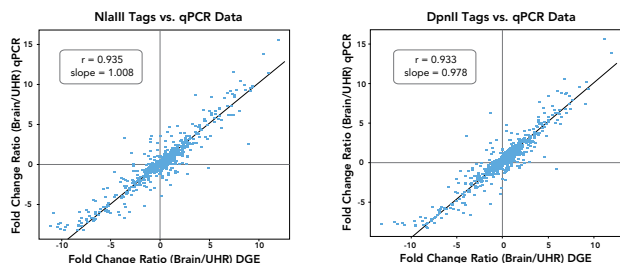


Figure 8: Tag Profiling is Highly Concordant with qPCR



Experiments using MAQC samples show that the data correlation between qPCR and the Tag Profiling protocol is greater than 0.93. After being assayed by qPCR, 629 and 625 RefSeq genes were quantified using the NlaIII and DpnII protocols, respectively. Unlike microarray data, there is no observed “ratio compression” in the data as evidenced by the slopes being approximately 1.

content selected from widely used databases such as NCBI RefSeq, and more specialized sources. Illumina Whole-Genome Expression BeadChips are ideal for applications such as differential expression analysis, disease classification, pathway analysis, and eQTL studies when discovery of novel transcripts is not essential and samples are from human, mouse, or rat.

Expression BeadChips are part of a complete gene expression solution that includes instrumentation, software, and reagent kits. Data analysis is straightforward, since known biologically relevant transcripts are annotated from heavily curated databases and probes are designed and validated by Illumina scientists. With a streamlined workflow and multi-sample BeadChip format, researchers can profile up to twelve samples in parallel on a single BeadChip, dramatically increasing throughput while decreasing experimental variability. The 100% hybridization-based QC of every probe ensures that BeadChips deliver outstanding performance and reproducibility. In addition, this unrivaled data quality comes at a lower cost per sample than other microarrays, allowing researchers to expand the scope of their science.

Comprehensive Array Content

Each address and probe sequence combination on Illumina Expression BeadChips was carefully selected bioinformatically. Gene-specific probes were designed using a multi-step algorithm to optimize several parameters:

- Lack of similarity to other genes
- Absence of highly repeated sequence in the genome
- Sequence complexity
- Self-complementarity for hairpin structure prediction
- Melting temperature for hybridization uniformity
- Distance from 3' end of the transcript

Table 1: Expression BeadChip Content

PROBES	DESCRIPTION	HUMAN WG-6	HUMAN REF-8	HUMAN HT-12†	MOUSE WG-6	MOUSE REF-8	RAT REF-12
RefSeq Content*							
NM	Coding transcript, well-established annotation	27,455	23,811	27,455	26,766	24,854	6,277
XM	Coding transcript, provisional annotation	7,870	426	7,870	6,856	796	15,983
NR	Non-coding transcript, well-established annotation	446	263	446	56	47	1
XR	Non-coding transcript, provisional annotation	196	26	196			12
Supplementary Content							
UniGene (Build 199)	Experimentally confirmed mRNA sequences that align to EST clusters	12,837		12,837			250
RIKEN FANTOM2	Exemplar protein-coding sequences from the RIKEN FANTOM2 database				5,659		
RefSeq Release 5	Transcripts with NM and XM annotation in RefSeq Release 5 (Build 33.1)				3,573		
MEEBO	Probes to transcripts that do not align with 100% accuracy to RefSeq, but are confirmed as valid mRNA mapping to clusters in Expressed Sequence Tag databases ⁶				2,371		
Total		48,804	24,526	48,804	45,281	25,697	22,523

*Human RefSeq Build 36.2 Rel 22, mouse RefSeq Build 36 Rel 22, rat RefSeq Rel 16

† > 99.99% of the bead types are present on any HumanHT-12 array

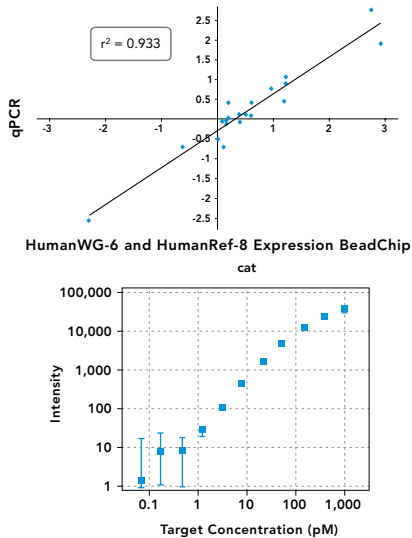
Table 2: Whole-Genome Expression BeadChips Product Specifications

Parameter	Specification
Probe Length	50-mer gene-specific probe plus 29-mer address sequence
Sensitivity	≤ 1:250,000
Dynamic Range	≥ 3 logs
Precision	≤ 1.35 fold
Input RNA Required	50–100 ng

All Illumina Expression BeadChips feature up-to-date content largely derived from a recent release of the RefSeq database. Regularly curated and updated by field experts and annotated according to strict guidelines, this widely used database serves as the scientific community's most comprehensive and stable reference for genomic DNA, transcript, and protein products. The HumanRef-8, MouseRef-8, and RatRef-12 BeadChips are developed exclusively from the RefSeq database content and target 18,631, 18,122, and 21,910 unique genes, respectively (Table 1).

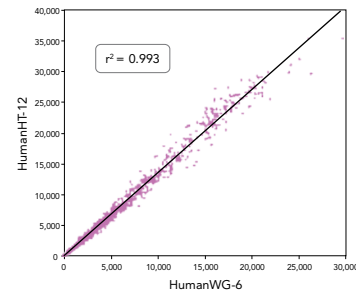
The HumanWG-6 and MouseWG-6 BeadChips contain the full set of HumanRef-8 and MouseRef-8 BeadChip probes plus supplemental content derived from additional databases. The HumanWG-6 BeadChip includes 12,837 probes targeting EST clusters from the UniGene database (Build 199). As a result, the HumanWG-6 BeadChip targets a total of 25,440 annotated genes with more than 48,000 probes (Table 1). The MouseWG-6 BeadChip includes 11,603 ad-

Figure 9: High-Quality Data Generated with Illumina Gene Expression BeadChips



Illumina Gene Expression BeadChips show high concordance with qPCR assay results (left), and have a wide dynamic range (right).

Figure 10: High Concordance Between HumanHT-12 and HumanWG-6 Data



Data generated from the higher throughput HumanHT-12 BeadChip are highly concordant with the HumanWG-6 BeadChip.

ditional probes derived from RIKEN FANTOM2, RefSeq rel 5, and MEEBO databases.

The HumanHT-12 contains the same comprehensive panel of probes as the HumanWG-6 BeadChip, but provides higher throughput processing of 12 samples per BeadChip. With this BeadChip, expression information can easily be incorporated in genome-wide association studies (GWAS), and large gene expression studies can be completed more quickly and economically. Illumina guarantees that more than 99.99% of the bead types will be present on any given HumanHT-12 array. This means up to five HumanWG-6 probes may be represented with only 0, 1, or 2 copies on each HumanHT-12 array.

In sum, the high-value content on human, mouse, and rat Expression BeadChips provides genome-wide transcriptional coverage of well-characterized genes, gene candidates, and splice variants, targeting well-established sequences supported by peer-reviewed literature.

High-Quality Data

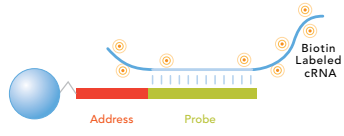
Illumina has compiled performance data for all Expression BeadChips2. Reproducibility has been demonstrated by high concordance between hybridization replicates. Industry-leading performance specifications for sensitivity, dynamic range, and fold-change detection precision dramatically minimize false discovery rates for differential expression analysis (Table 2, Figure 9).

As expected, HumanHT-12 assay performance is equally high, and shows very high concordance with HumanWG-6 BeadChip data (Figure 10).

Streamlined Assay Workflow

Illumina Expression BeadChips are designed using BeadArray technology. BeadChips consist of bead-linked oligonucleotides held in microwells on the surface of a slide-sized substrate. During the manufacturing process, beads self-assemble into the microwells of the BeadChips. Each bead type contains hundreds of thousands of copies of a covalently attached, full-length oligonucleotide probe. Data quality and reproducibility are supported in part by the high level of bead type redundancy (up to an average of 30 beads per probe) on every array. After random bead assembly, 29-mer address sequences present on each bead are used for a hybridization-based procedure to map the array, identifying the location of each bead. This

Figure 11: Direct HYB Gene Expression Profiling Bead Design



Gene-specific 50-mer probes are attached to beads assembled on BeadArray substrates.

final process also validates the hybridization performance of every bead on every BeadChip, ensuring 100% array QC.

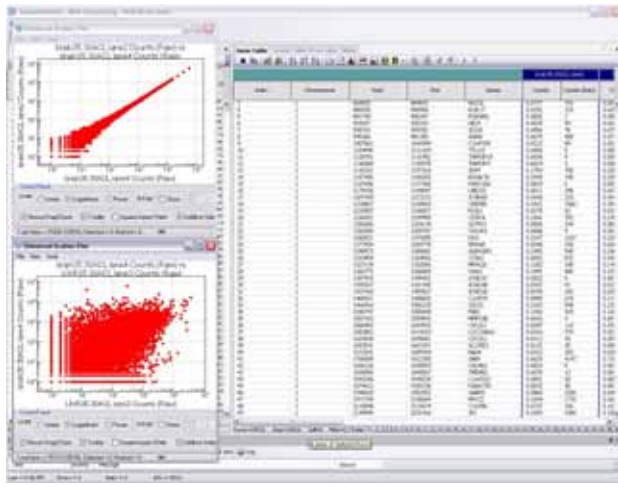
Expression BeadChip arrays are arranged in a multi-sample format for higher throughput and virtual elimination of array-to-array variability. All steps after hybridization are performed in parallel on each BeadChip, significantly reducing experimental variation and handling requirements. Labeled sample cRNA are detected by hybridization to 50-mer probes on the BeadChip. After washing and staining steps, BeadChips are scanned on the Illumina iScan™, HiScanSQ, or BeadArray Reader.

For the highest sample throughput (up to 216 arrays at a time), researchers can automate BeadChip loading and scanning with the Illumina AutoLoader.

DASL Assay

The DASL Assay (cDNA-mediated annealing, selection, extension, and ligation) is highly robust for gene expression profiling of traditionally difficult to assay samples, such as those having low RNA abundance or with RNA degradation due to FFPE (formalin-fixed paraffin-embedded) processing.

Figure 12: Genomestudio RNA-Sequencing Module



The Genomestudio RNA-Sequencing Module includes tools for data visualization, including graphical plotting and table views for analyzing genes, exons, exon junctions, and alleles (for SNP detection).

The Whole-Genome DASL Assay⁵ covers more than 24,000 transcripts, using same content as HumanRef-8 v3.0 BeadChip.

For focused panels of up to 1,536 genes, the DASL Assay is deployed on multi-sample Universal Arrays³. Very high throughput of low- to mid-multiplex assays are possible when the DASL Assay is used in combination with VeraCode technology⁴.

Systems and Software Sequencing Data Analysis

mRNA-Seq and Tag Profiling generate data using open architecture software, allowing researchers to tailor Illumina sequencing system data analysis to address their specific needs. The Analysis Pipeline is responsible for performing primary data acquisition, determining base calls, and calculating confidence scores from the fluorescence signals on Illumina's sequencing systems. Higher level analyses, such as aligning reads to a reference, determining expression levels of genes and exons, and identifying variants, are performed using the tools in the Analysis Pipeline and integrated algorithms. The results from these analyses are parsed to the RNA-Sequencing Module in Illumina's GenomeStudio™ analysis software for display (Figure 12). Results are easily visualized on the integrated graphical genome viewer that includes annotation information (Figure 3). Users can also zoom down to single-base sequence resolution to identify cSNPs and exon-exon junctions (Figure 5). Exon and gene table views allow for the examination of expression levels in unprecedented detail (Figure 12).

For the Tag Profiling application, image analysis, base calling, and standard filtering by the Analysis Pipeline generate a list of sequence tags and counts. This list of tags can be annotated with genomic information and used to analyze differential gene expression. The software provides canonical sequences for human and mouse genomes. Expression profiles for other species can be compared easily against public databases like the NCBI RefSeq database and the University of California Santa Cruz (UCSC) genome browser.

Expression Array Data Analysis

Illumina's GenomeStudio Gene Expression Module (Figure 13) enables simplified data management for hierarchical organization of samples, groups, groupsets, and all associated project analysis. It offers gene-level statistical analysis tools for differential analysis, heat map visualization, and clustering. With Gene Expression Modules 3.6.2 or higher, researchers can combine Expression BeadChip data generated from different major product versions (e.g., HumanRef-8 v2.0 and HumanRef-8 v3.0).

Researchers can easily combine Expression BeadChip data with either methylation or miRNA profiling data in a single GenomeStudio gene expression project. This enables powerful integrated approaches to studying epigenetic impacts on gene expression. Important for eQTL studies, the flexible data management architecture of the GenomeStudio software supports integrating gene expression probe annotation information with SNP location coordinates. Using APIs, researchers can export genotyping and expression data from GenomeStudio software to third-party applications to perform eQTL-like integrated analysis.

Ordering Information

Product	Quantity	Catalog No	
mRNA-Seq			
mRNA-Seq 8-Sample Prep Kit	8 Samples	RS-100-0801	
mRNA-Seq Cluster Generation Kit (GALL)	1 Flow Cell	RS-110-0101	
	10 Flow Cells	RS-110-1001	
mRNA-Seq 36-Cycle Sequencing Kit	1 Flow Cell	RS-120-3601	
Tag Profiling		Enzyme	
Tag Profiling Sample Prep Kit	Nlalll	8 Samples/1 Flow Cell	FC-102-1005
		40 Samples/5 Flow Cells	FC-102-1006
	Dpnll	8 Samples/1 Flow Cell	FC-102-1007
		40 Samples/5 Flow Cells	FC-102-1008
Tag Profiling Cluster Generation Kit	Nlalll	Up to 8 Samples	FC-103-1004
		Up to 80 Samples	FC-103-1005
	Dpnll	Up to 8 Samples	FC-103-1006
		Up to 80 Samples	FC-103-1007
18-cycle Illumina Sequencing Kit		1 Flow Cell	FC-104-1001
Sequencing Instruments			
Illumina Cluster Station			SY-301-2001
• Includes computer, software, installation, training, and 1-year warranty			
Illumina Genome Analyzer			SY-301-1001
• Includes computer, software, installation, training, and 1-year warranty			
Illumina HiScanSQ			SY-301-1001-PRE
• Includes computer, software, installation, training, and 1-year warranty			

AAAGAATGATAACAGTAAACACACTTCTGTTAAACCTTAAGATTACTTGATCCACTGATTC AACGTACCCTAAACAGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTGCAACGACGAAAAGAAATGATAACAGTAAACACACTTCTGTTAAAC
AATCAACGTACCCTAACGAAACGTATCAATTAAAGATTACTTGATCCACTGATTC AACGTACCCTAAACAGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTGCAACGACGAAAAGAAATGATAACAGTAAACACACTTCTGTTAAAC
AACGACGAAAAGAATGATAACAGTAAACACACTTCTGTTAAACCTTAAGATTACTTGATCCACTGATTC AACGTACCCTAAACAGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTT
TTAAGGTACCATTAAAGAGCTACCGTGCAACAGTAAACACACTTCTGTTAAACCTTAAGATTACTTGATCCACTGATTC AACGTACCCTAAACAGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTGCAACGACGAAAAGAAATGAT
AAAGAAATGATAACAGTAAACACACTTCTGTTAAACCTTAAGATTACTTGATCCACTGATTC AACGTACCCTAAACAGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTGCAACGACGAAAAGAAATGATAAC
AAGATTACTTGATCCACTGATTC AACGTACCCTTAAGATTACTTGATCCACTGATTC AACGTACCCTAAACAGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTGCAACGACGAAAAGAAATGATAAC
AACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGTCTGTTAAACCTTAAGATTACTTGATCCACTGATTC AACGTACCCTAAACAGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTGCAACGACGAAAAGAAATGATAAC

