Notes on Illumina Technology for Whole-Genome Expression

Kevin R. Coombes and Jiexin Zhang

2 July 2007

1 Illumina Products

Illumina provides products for the following kinds of assays:

1. DNA analysis
   (a) SNP genotyping
   (b) Copy number/LOH
   (c) Methylation
2. RNA analysis
   (a) Genome-wide expression
   (b) Focused arrays for expression: custom focused arrays
   (c) DASL (cDNA-mediated annealing, selection, extension, and ligation assay) gene expression: reproducible profiles.

2 Basic Illumina Technology

Illumina's BeadArray Technology is based on 3-micron silica beads that self assemble in microwells on a substrate. When randomly assembled on the substrate, the beads have a uniform spacing of about 5.7 microns. Each bead is covered with hundreds of thousands (200K to 400K) of copies of a specific oligonucleotide that act as the capture sequence in one of Illumina's assays. Because the micro-wells outnumber the probe sequences, multiple copies (20-30) of each bead type are present in the array. This build-in redundancy improves robustness and measurement precision. For example, researchers can profile six samples per Mouse-6 BeadChip with about 47,000 transcripts per sample.

2.1 Controls

The controls used include the following:

1. Controls for biological specimen: 7 housekeeping genes.

2. Controls for sample labeling: 4 RNA spike (lysA, pheA, thrB, thpF)

3. Controls for hybridization: 6 probes for Cy3-labeled hyb; 8 probes for low stringency hyb; 1 probe for high stringency hyb.

4. Controls for signal generation: 2 probes with complementary biotin-tagged oligo.

5. Negative controls: hundreds of probes of random sequence selected to have no corresponding targets in the genomes.

3 Generating Gene Profile Data Using Bead Studio

In order to analyze Illumina data in R, it must first be exported from the Illumina Bead Studio program. For completeness, we include instructions on data export.

1. File → New Project → Gene Expression → Direct Hyb → Specify the name and location for your project → Specify the directory where your arrays are located → Select arrays for your project and move them into "Project Data" window → Group samples into groups of replicates (Figure 1).

2. Choose analysis type. ("Gene expression" in our case. the other option is "differential expression".) → Give content descriptor (e.g., MOUSE-6_V1.XML) → Finish.

3. Export "Sample Gene Profile" to GeneSpring GX format. (Other files that can be generated are "group gene profile", "group probe profile", and "sample probe profile"). Figure 2 shows part of the Sample_Gene_Profile.

4. In order to obtain a mapping from the "TargetID" values that are present in SGP to gene annotation, you need to do the following:

   - Select the "Sample Gene Profile" tab (one of four display options above the table);
   - Click on "column chooser" and make sure "TargetID" and one of all of "ACCESSION"/"SYMBOL"/"GID" are in the "Displayed Columns" (in my case, I choose "ACCESSION");
   - Click the export icon (folder with the out arrow) on the tool bar, rather than the export to GeneSpring function. Using the GeneSpring function strips the accession numbers, whereas the straight export tool leaves all of the viewable columns, including accession, definition, etc;
   - give name and save the file (in my case I called it "targetID_accession.txt").

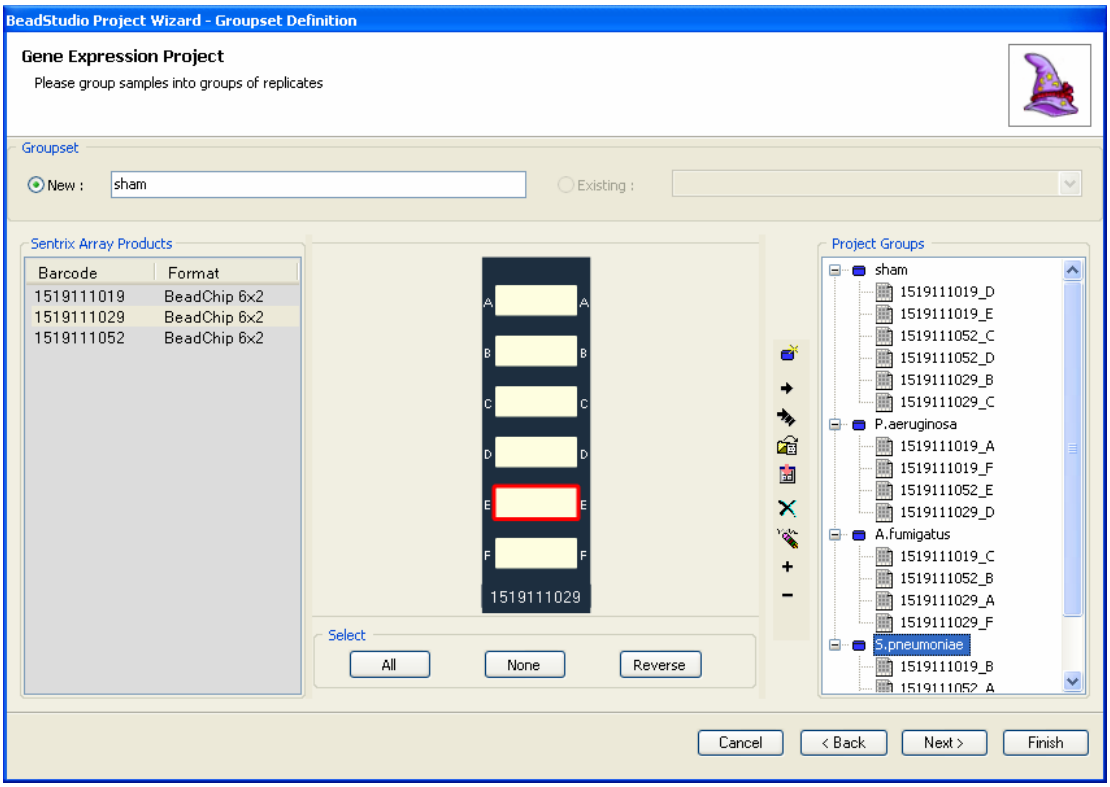

Figure 1: Bead Studio Snapshot

TargetID	MIN_Signal-1519111019_D	AVG_Signal-1519111019_D	MAX_Signal-1519111019_D	NARRAYS-1519111019_D	ARRAY_STDEV-1519111019_D	BEAD_STDEV-1519111019_D	Avg 151
10181072_239_rc-S	273.4	273.4	273.4	1	NaN	91.266	
10181072_290_rc-S	303.4	303.4	303.4	1	NaN	105.965	
10181072_290-S	289.6	289.6	289.6	1	NaN	87.265	
10181072_311_rc-S	275.5	275.5	275.5	1	NaN	81.663	
10181072_311-S	293	293	293	1	NaN	92.046	
10181072_418_rc-S	305.8	305.8	305.8	1	NaN	74.414	
10181072_418-S	215.2	215.2	215.2	1	NaN	81.93	
10181072_486_rc-S	260.8	260.8	260.8	1	NaN	75.407	
10181072_486-S	283.1	283.1	283.1	1	NaN	88.995	
17974913_3385_rc-S	265.2	265.2	265.2	1	NaN	62.203	
17974913_3385-S	284.1	284.1	284.1	1	NaN	83.912	
17974913_3999_rc-S	254.6	254.6	254.6	1	NaN	76.387	
17974913_4071_rc-S	254.1	254.1	254.1	1	NaN	89.89	
17974913_4071-S	242.5	242.5	242.5	1	NaN	80.838	
17974913_4426_rc-S	307.9	307.9	307.9	1	NaN	69.707	
17974913_4426-S	277.8	277.8	277.8	1	NaN	70	
17974913_4962_rc-S	248.2	248.2	248.2	1	NaN	70.738	
17974913_4962-S	230	230	230	1	NaN	54.9	
18S_rRNA_X00686_301-S	35158.9	35158.9	35158.9	1	NaN	7792.206	
18S_rRNA_X00686_523-S	13188.8	13188.8	13188.8	1	NaN	5482.807	
18S_rRNA_X00686_849-S	8878.7	8878.7	8878.7	1	NaN	3679.028	
20198505_2519-S	287.3	287.3	287.3	1	NaN	78.862	
20198505_4692_rc-S	259.4	259.4	259.4	1	NaN	96.548	
20198505_4826_rc-S	284	284	284	1	NaN	66.851	
20198505_4826-S	244.6	244.6	244.6	1	NaN	55.642	
20198505_5027_rc-S	281.7	281.7	281.7	1	NaN	99.569	
20198505_5027-S	274.7	274.7	274.7	1	NaN	52.716	
20198505_5605-S	276.4	276.4	276.4	1	NaN	83.178	
21716071_3233_rc-S	269.6	269.6	269.6	1	NaN	77.893	
21716071_3233-S	314.4	314.4	314.4	1	NaN	83.719	

Figure 2: Example of Sample Gene Profile file.