As a microarray manufacturer Illumina fully appreciates the importance of MIAME as a standard necessary for meaningful interpretation and verification of microarray experiments. Our goal is to become a MIAME compliant platform when we release our gene expression related products. Therefore, I created this document to introduce the MIAME workgroup to some concepts which we think can extend the current version of the standard and enable compliance by our customers. I include a description of the Illumina array, and I would like to seek from your workgroup comments of whether they consider such description to be MIAME compliant.

Terms Needed in the Description of the Illumina Microarray Platform

- Random Array – this concept implies that features assemble at random locations on each physical array manufactured. In terms of changes to the Array Description part of MIAME, it would require an additional array property – 'Array Geometry'. Describing array geometry as Random will allow us to introduce a Virtual Array Geometry placing all features present in the array onto a fixed virtual rectangular grid which makes it analogous to an ordered array. This will spare our customers the need to submit array description data with every physical array they use (it is impossible to reproduce feature locations so it does not affect verification of experiments ability). Also, virtual arrays can be helpful in describing flow cytometry gene expression data.
- Universal Array – this concept is not unique to Illumina and implies that probes on the array are used indirectly as address readout reporters rather than direct hybridization target quantifiers. This notion can help to create a placeholder for additional sequence information required to interpret microarray data. In case of Illumina we refer to address reporters as IllumiCode capture sequences.
- Generalized Reporter – this concept allows characterization of various parts of the sequence generated on the array. In particular, we propose to split reporter sequence into hybridizing part and auxiliary part. In figure 1a below, the reporter's hybridizing sequence coincides with the gene specific portion while the auxiliary sequence functions as an IllumiCode capture sequence. Another possible function for the auxiliary sequence is to provide padding between the substrate and the gene specific sequence.

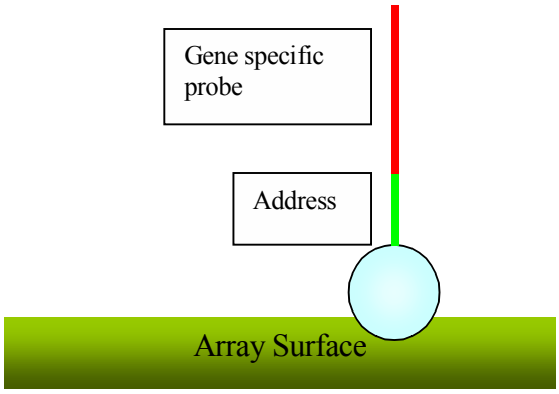

Figure 1.a

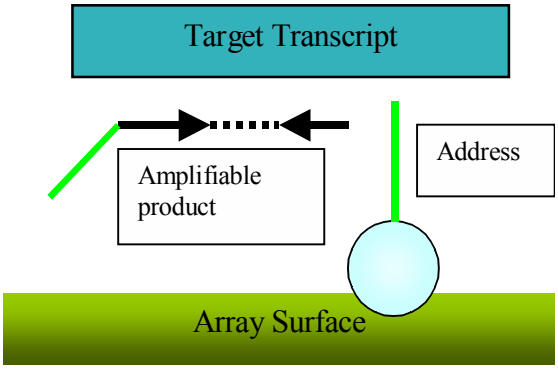

Figure 1.b

Figure 1b represents a typical universal array. In this case the sequence hybridizing to the gene specific part is not attached to the array. The reporter sequence does not have an auxiliary part, however, it serves as IllumiCode capture sequences. This could be described through an additional reporter sequence property - 'hybridization type'. It would be included in the experiment description part of MIAME. Illumina plans to release probe sequence information for gene specific portions, however, for intellectual property considerations we cannot release sequences for our IllumiCode capture sequences.

- Probe orientation – As a minor addition, we think MIAME should add additional reporter sequence property describing orientation (3' outward or inward) of probes with respect to the array surface.

To summarize these concepts here is the proposed content of Illumina's array description: Blue font represents additions either new to the current MIAME content or having new interpretation.

1. **Array related information**
   Array Design Name= IlluminaTest
   Platform Type= DNA Oligonucleotide
   Surface and Coating Specification= silica beads, diameter=3µm.
   Physical Dimension= hexagon, diameter=1.5mm
   Array Geometry= Random (means that locations described below are virtual)
   Number of Features on the Array= Number of distinct feature types (number of replicates per distinct feature is random)
   Availability= provided by Illumina Inc.
2. **(a) For each reporter type**
   Reporter Type = synthetic DNA oligonucleotide
   Number of strands = 1
   (b) For each reporter type
   Identifier = illumina's unique identifier
   Hybridization type = Universal or Specific
   Auxiliary sequence function = (None, padding, address, etc.)
   Auxiliary sequence position = proximal to surface (if exists)

Auxiliary sequence length= sequence length if exists, 0 otherwise
Specific sequence=AGCT… (gene specific portion of the sequence if exists)
Sequence orientation= 5' proximal to surface

3. **(a) For each feature type**
   Dimension = diameter 3µm
   (b) For each feature
   Reporter identifier
   Location on the virtual array.
4. **For each composite sequence**
   List of reporters
   Reference sequence
   Gene name
5. **Control Elements**
   Position of the feature on virtual array
   Control type
   Control qualifier

In the Experiment Description section we propose to introduce the following descriptor:

For each universal reporter:
Specific sequence
Element of the array expected to hybridize to it

In terms of raw data, Illumina will provide customers with gene expression data tables which would list for all features described in array descriptor:

- Feature Identifier
- Average Intensity
- Number of Beads
- Bead to Bead Standard Deviation
- Detection p-value.

In addition, customers will have raw images, so they will be able to examine images for scratches, intensity gradients, etc. We will make our extraction algorithms and normalization routines public.