



Manufacturing Quality Control and Illumina® BeadArray™ Technology

Introduction

This communication presents an overview of Illumina's BeadArray technology and how it is used to deliver industry-leading content and quality in Illumina's Gene Expression Solutions. The aim is to communicate the fundamental differences from other array technologies, both in design and in manufacture, and how these differences contribute to a superior product. We explain how we can guarantee 100% Quality Control (QC) of every array, every time, and why no other manufacturer can make this statement.

Array Design: Built-In Quality

Central to Illumina's product portfolio is a different way of building arrays: the self-assembly of beads into ordered microwells. The array elements are 3- μm silica beads to which 50mer, transcript-specific oligonucleotide probes have been immobilized. To make a Gene Expression BeadChip, we first synthesize tens of thousands of unique oligonucleotides, each complementary to a different target sequence. These oligos are then immobilized in separate reactions to the beads, generating a corresponding number of unique bead types. The beads are pooled together, and then loaded by a self-assembly process into microwells that have been etched into the surface of a BeadChip. A key feature of this process that leads to quality data is feature redundancy. Each feature (i.e., each bead type) is represented on the array surface an average of more than thirty times. Other commercial microarray platforms represent their features only once per

array. This redundancy results in higher data quality, because it means that each hybridization intensity value in your microarray experiment is based on the average of many independent readings. This also enables the removal of outlier data points without compromising overall results. The power of outlier removal is bolstered by the fact that replicate beads for any given feature are distributed throughout the array surface, essentially eliminating the possibility of dust or other local defects affecting the outcome of a reading.

Decoding: Hybridization-Based QC of Every Feature on Every Array

After array assembly, Illumina's manufacturing process uses a hybridization-based procedure, termed decoding, to map the contents of the array. This mapping is made possible by the two-part design of the oligonucleotides immobilized to the beads. Transcript-specific 50mers are concatenated during synthesis to a 29-base address sequence. During manufacturing, pools of labeled, complementary sequences are hybridized to the address on each bead in a series of steps designed to identify and error-check the identity of each feature. Beyond simply being necessary for Illumina's process, decoding also provides a key advantage over all other array technologies: 100% hybridization-based QC of every feature on every array. At the end of the decoding process, each feature has not only had its identity validated, but has also proven itself to be capable of robust hybridization. This is what we mean when we say, "Every gene, every array,

every time.” When we design a probe for a given target, we know that it is present on the array and that it is functional before it ever leaves the production line.

Incoming Materials Testing

A number of components come together to make Illumina’s unique products. Incoming materials are tested and passed or failed by the Quality Assurance department prior to incorporation into the product. In addition, QC measurements are taken throughout the process of building every component.

The oligonucleotide probes attached to Illumina’s beads are central to the performance of the platform. Oligo synthesis is performed using proprietary, high-throughput robotics to ensure a completely consistent and controlled environment. This is in sharp contrast to industry-standard, non-automated oligo production lines requiring multiple technicians and manual interventions. Illumina synthesizers are operated by a team of highly-skilled and trained manufacturing laboratory personnel in a Laboratory Information Management System (LIMS) controlled process. During synthesis, a comprehensive QC process tied to the LIMS ensures that each liquid dispensing step is successful and that the resulting oligo is full length. Once synthesis is complete, Capillary Electrophoresis (CE) is used to estimate percent full-length product. Optical Density (OD₂₆₀) measurements are carried out to estimate product yield. In addition, all QC metrics are pushed to the LIMS and are used to generate control charts and feedback mechanisms, including pass and fail calls for every oligo produced. This proprietary oligo manufacturing process allows us to synthesize millions of bases

daily under the most stringent quality control.

Probe Selection from Public Sources: Public Scrutiny, Field Consensus and Perpetual Availability

Illumina bases its probe design on 100% publicly-available content. This philosophy ensures that the content will reflect the latest revisions of the public sequencing efforts and the consensus annotation of splice variants, also permitting us to make all sources and probe sequences freely available. Commercial microarrays containing proprietary content do not offer this convenience. An example of array content design is described below for the Sentrix® Human-6 and HumanRef-8 Expression BeadChips.

The HumanRef-8 Expression BeadChip contains > 24,000 unique probe sequences based on the National Center of Biotechnology Information (NCBI) Reference Sequence (RefSeq) database, while the Human-6 Expression BeadChip contains both this RefSeq core content plus additional content from other public sources. Because no proprietary data sources were used, access to the source sequences used in Illumina’s content is freely available. The RefSeq database, from which core content was developed for both products has many benefits over other sources, including high-quality sequence information, curation by field experts and tracking of sequence and annotation changes. To ensure that quality sequences were selected for the supplementary content on the Human-6 BeadChip, the following criteria were applied:

- Match to human genome sequence
- Multiple hits to expressed sequence tag (EST) databases
- Hits to cDNA library databases

- Open reading frame (ORF) prediction to ensure correct strand choice
- Cross-referencing among multiple data sources
- Avoidance of duplicating records
- Avoidance of pseudogenes

After preliminary selection of source sequence records, the 50mer probe sequences are carefully selected bioinformatically using a multi-step algorithm that scores with the following parameters:

- Similarity to other genes
- Absence of highly repeated sequence in the genome
- Sequence complexity
- EST coverage (Genome Annotation-RefSeq genes)
- Self-complementarity for hairpin structure prediction
- Melting temperature for hybridization uniformity
- Distance from 3' end of the transcript

Illumina's use of the latest public data sources combined with the extensive quality checks used in probe design minimizes the errors in probe sequence or strand association that have been

reported in the literature with other array platforms.

Summary

Illumina's Gene Expression Solutions provide industry-leading data quality, current content and the industry's only 100% hybridization-based array QC. These factors combined deliver the quality you need for the confidence you deserve in your gene expression microarray research.

References

Array Decoding:

Gunderson, K., Kruglyak, S., Graige M.S., Garcia, F., Kermani, B., Zhao, C., Che, D., Dickinson, T., Wickham, E., Bierle, J., Doucet, D., Milewski, M., Yang, R., Siegmund, C., Haas, J., Zhou, L., Oliphant, A., Fan, J.B., Barnard, S., Chee, M.S. Decoding randomly ordered DNA arrays. *Genome Res.* (2004) 14: 870-877.

Gene Expression Data Quality:

Kuhn, K, Baker, SC, Chudin, E, Lieu, M-H, Oeser, S., Bennett, H., Rigault, P, McDaniel, TK, Chee, MS. A Novel, High-Performance Random Array Platform for Quantitative Gene Expression Profiling. *Genome Res.* (2004) 14: 2347-2356.