DATA SHEET

Oligator® "MEEBO" Mouse Genome Set

Developed in collaboration with leading researchers, the MEEBO (Mouse Exonic Evidence Based Oligonucleotide) Set contains a collection of oligo probes, largely derived from constitutively expressed exons, allowing interrogation of almost 25,000 mouse genes. The set was designed to enable study of mouse transcription patterns and, as broadly as possible, alternative splicing.

An exon-centric design was selected to allow the differentiation of constitutively expressed versus alternatively expressed exons. This design also supports comparative genome hybridization (CGH) analysis. In addition to the exon-centric probes, the set contains an extensive assortment of controls that facilitate accurate evaluation of expression results.

PRODUCT INFORMATION AND SPECIFICATIONS

The MEEBO Set contains 38,467 70mer oligonucleotide probes with an amino modification on the 5' end. Probes are provided at a final yield of 200pmol each, and delivered in 101 384-well microarray print plates (Genetix model X7020). Probes are Tm normalized, and are designed to minimize cross-hybridization and secondary structure.

Mouse Probes

The set contains a total of 35,302 probes targeting mouse genes. The probes fall into the following categories:

- **Constitutive Exonic Probes (30,125 probes):** A probe that will recognize all known transcripts of a gene.
- **Alternatively Spliced / Skipped Exonic Probes (4,201 probes):** Probes that will recognize exons that are present in some, but not all transcripts of a gene.
- **Non Coding RNA Probes (196 probes):** Probes recognizing non-protein coding transcripts (ribosomal RNAs, miRNAs).
- **BCR / TCR Genic / Regional Probes (372 probes):** Probes recognizing transcripts from genes that undergo somatic rearrangement.
- **Mitochondrial Probes (13 probes):** Probes recognizing mouse mitochondrion derived DNA sequences.
- **Transgenic / Cassette Probes (37 probes):** Probes recognizing elements commonly used for transgenic constructs (e.g. GFP, YFP).
- **Murine Viral Probes (358 probes):** Probes recognizing mouse viral pathogen sequences.

Controls

The set contains a total of 3,482 controls. The controls fall into the following categories:

- **Negative Controls (317 empty wells and 97 probes):** Empty wells and 97 random sequences are positioned throughout the set to assist in determining background.
- **Positive Controls (1,152 probes):** Probes recognizing a small subset of mouse transcripts.
- **Doped Controls (1,916 probes):** Probes recognizing non-mouse sequences that can be spiked into RNA samples.

DESIGN APPROACH

Mouse Sequence Selection and Probe Design

A systematic methodology was applied to identify the exons, generate all possible 70mer candidate probes, and select the optimal probe from the candidates. The pipeline for sequence selection and probe design included three steps:

1. Collect and curate exon sequences, supplement as needed with transcript sequences.
2. Design the candidate 70mer probes for exon or transcript sequences.
   - ArrayOligoSelector, an open source tool for selecting 70mer oligo probes from a defined set of sequence data, was used to generate a list of candidate probes for each exon or transcript sequence.
   - Multiple filters including uniqueness, self-binding, complexity, GC, content, and user defined parameters were used to narrow and rank the list of candidate probes.
3. Pick the best probe from the list of candidate probes. Several criteria were used to identify the optimal probe from the list of candidate probes:
   - **Uniqueness:** Probes that had binding energies of > -35 kcal / mol for other sequences were preferred.
   - **3' Proximity:** Probes that were less than 1,000 bases from the 3' end of the transcript were preferred.
   - **Constitutive:** Probe should be present in all transcripts.

If the above three criteria could not be met, more than one probe would be selected for the exon or transcript sequence.

Annotation Details

Probe annotation information is provided in the OLIGATOR_MEEBO_MOUSE_SET.XLS file. The file contains the following data elements:

- **Plate_Name:** Contains the name that will appear on the plate label. MCC plates contain mouse probes; non-MCC plates contain controls.
- **Plate_Num:** Contains the plate number used to identify which packing box a plate is located.
- **Row:** Oligo plate row position.
- **Column:** Oligo plate column position.
- **Probe_Name:** Contains the oligo name. The following codes can assist with interpreting the oligo name:
  - Rockefeller MouSDB3 constitutive exons / islands (oligo names start with 'scl' followed by a number >0)
  - LocusLink constitutive exons / islands (oligo names start with 'scl0' followed by a number >0)
  - mRNA derived 70mers which may span intron/exon boundaries (oligo names start with 'scl00' followed by a number >0)
  - A collection of alternative spliced / skipped exons generated through extensive curation of 5 published datasets by Max Diehn, Ash Alizadeh, Jean Yang, and Catherine Foo (oligo names start with 'scl000' followed by a number >0)
  - Syntenic orthologs of human loci exhibiting cis-antisense transcription based on Yelin et al *Nature Biotech* 2003. (oligo names start with 'scl0000' followed by a number >0)
- **LocusLink_ID:** LocusLink ID number.
- **Gid:** GI number.
- **Accession:** Accession number.
- **Symbol:** Gene symbol.
- **Probe Sequence:** 70mer oligo probe sequence.
- **T_m:** Oligo melting temperature.
- **GC:** Percentage of GC content within the probe sequence.
- **Product:** Probe target definition.
- **Description:** The specific annotation for each 70mer is detailed and is carat (^) delimited as follows for MCC plates:
  - MCC Details: [LLID (060504) ^ Design type ^ cluster ^ Long Oligoname ^ outputset ^ worstxhyb ^ oligo_3_marg]
  - Where LLID refers to LocusLink ID, which can be linked with gene annotation data through a variety of tools including BatchSource.

OLIGONUCLEOTIDE SYNTHESIS AND QUALITY CONTROL

Oligos were synthesized on Illumina's proprietary Oligator synthesis platform.

Illumina utilizes multiple quality control methods to assess oligo quality including:

- **Real Time Digital Trityl Monitoring:** Illumina monitors coupling success of each base addition for every oligo in every plate synthesized. Monitoring is performed real-time during the synthesis process using its proprietary digital trityl monitoring system.
- **Capillary Electrophoresis (CE):** Illumina's uses CE to achieve single base resolution of synthesis success for long oligos; precise abundance measurements and stepwise coupling efficiency values are derived from CE.
- **OD260 Analysis:** Illumina uses OD260 analysis (absorbance measurement) to quantify oligo yield.

CREDITS

Illumina thanks the following individuals for their effort in defining and designing the contents of this set:

University of California, San Francisco: Ash Alizadeh, Mike Hagen, Catherine Foo, Jess Leber, David Erle, Jean Yang, Andrea Barczak, Joseph DeRisi, Jing Zhu

Stowers Institute: Chris Seidel

Stanford University: Pat Brown, Max Diehn, Kate Rubins, Stephen Popper, Joseph Marquis, Mike Fero, John Coller,, Nicki Chin, Elena Seraia, Peng Zhang, Jon Pollack, Young Kim

Rockefeller University: Terry Gaasterland, Alexey Novoradovsky, Ben Snyder

University of Basel, Switzerland: Mihaela Zavolan

ADDITIONAL RESOURCES

A genome browser file and other resources for this set are available at the following link: http://arrays.ucsf.edu/meebo.html

CONTACTING ILLUMINA

For questions or support, please call 800.809.4566
(1.858.202.4566 from outside the United States)
or e-mail us at techsupport@illumina.com.

Illumina, Inc.
9885 Towne Centre Drive
San Diego, CA 92121-1975
1.800.809.4566 (toll free)
1.858.202.4566 (outside the U.S.)
techsupport@illumina.com
www.illumina.com

FOR RESEARCH USE ONLY

Pub. No. 670-2004-008

illumina®
making sense out of life