

Table B: Oligosets for Printing Microarrays

| | <u>Compugen</u> | <u>Qiagen Operon</u> | <u>MWG Biotech</u> | <u>Clontech</u> |
|--------------|---|--|--|---|
| Human | <p>Human Oligo Library Number of oligos: 18,861. LEADS human clusters represented: 18,664 (plus 197 controls). UniGene human clusters represented: 15,223. Public domain human genes not found as well-characterized genes in the UniGene database: 3,441. Number of oligos per LEADS cluster: 1. Oligo length: 60 bases (5'-C6-amino modification). Distance from the 3' end: 95% of oligos are within 61–1309 bases from 3' end. The average distance from the 3' end is 397 bases. Cross-homology: 95% of oligos contain a predicted cross-homology of <66.2%. The average cross-homology is 33.8%.</p> <p>Human UniGene Build #138 was used for mapping oligos to UniGene clusters.</p> | <p>Human Genome Oligo Set Total number of oligos: 21,353. Number of human genes represented: 21,329 (plus 24 controls). Oligo length: 70 bases (5'-C6-amino modification). Distance from the 3' end: oligos are within 1000 bases from the 3' end.</p> <p>All oligonucleotides in the Human Genome Oligo Set were designed from the representative sequences in the UniGene database, Hs build # 147 (Feb. 2002), and the Human Reference Sequence (RefSeq) Database.</p> | <p>Pan Human 10K Oligo Set Number of oligos: 9,984. UniGene clusters represented: 9,850. Number of oligos per UniGene cluster: 1. Oligo length: 50 bases (5'-C6-amino modification).</p> <p>Pan Human Cancer Oligo Set Number of oligos: 1,920. UniGene clusters represented: 1,900. Number of oligos per UniGene cluster: 1. Oligo length: 50 bases (5'-C6-amino modification).</p> | <p>Atlas Human Ready-To-Print Long Oligos Number of oligos: 12,800 LocusLink clusters represented: 12,546 Unique genes not represented in LocusLink: 105 Total number of unique genes: 12,651 Oligo length: 68 bases + 12 base tag Distance from 3' end: ~95% of the Oligos are within 1,500 bases of the 3' end. Of these, ~80% of the sequences are within 500 bases from the 3' end. Cross-homology: More than 93% of oligos have unique sequences (i.e. no sequence match longer than 24 bases (35%) to other LocusLink entries) The human oligo set was designed using curated NCBI Refseq (88%) and GenBank (12%) sequences. RefSeq/LocusLink version 5/22/2002 was used for validation of oligos (LocusLink representation, cross-homology).</p> |

| | | | | |
|---------------------|---|---|---|--|
| <p>Mouse</p> | <p>Mouse OligoLibrary Number of oligos: 7,524. LEADS mouse clusters represented: 7,445 (plus 79 controls). UniGene mouse clusters represented: 6,544. Public domain mouse genes not found as well-characterized genes in the UniGene database: 901. Number of oligos per LEADS cluster: 1. Oligo length: 65 bases (5'-C6-amino modification). Distance from the 3' end: 95% of oligos are within 65–1133 bases from 3' end. The average distance from the 3' end is 396 bases. Cross-homology: 95% of oligos contain a predicted cross-homology of <66.5%. Average cross-homology is 30.5%.</p> <p>Mouse UniGene Build #93 was used for mapping oligos to UniGene clusters.</p> <p>Mouse Extension Oligo Library Number of oligos: 14,532. LEADS mouse clusters represented: 14,380.</p> <p>The Mouse OligoLibrary Extension is designed to reflect the rapid accumulation of data in the public domain since the release of the Mouse OligoLibrary. Oligos in the Extension were selected to represent all of the mouse genes associated with public mRNA sequences found in GenBank 126 and not represented in the Mouse OligoLibrary.</p> | <p>Mouse Genome Oligo Set Number of oligos: 16,487. Number of mouse genes represented: 16,463 (plus 24 controls). Oligo length: 70 bases (5'-C6-amino modification). Distance from the 3' end: oligos are within 1000 bases from the 3' end.</p> <p>The oligonucleotides in the Mouse Genome Oligo Set were designed from the representative sequences in the Mouse UniGene database, Mm build #102, and the Mouse Reference Sequence (RefSeq) Database.</p> | <p>Pan Mouse 10K Oligo Set Number of oligos: 9,984. UniGene clusters represented: 9,850. Number of oligos per UniGene cluster: 1. Oligo length: 50 bases (5'-C6-amino modification).</p> | <p>Atlas Mouse Ready-To-Print Long Oligos Number of oligos: 5,082 LocusLink clusters represented: 5002 Total number of unique genes: 5002 Oligo length: 68 bases + 12 base tag Distance from 3' end: As for human set Cross-homology: More than 96% of oligos have unique sequences (i.e. no sequence match longer than 24 bases (35%) to other LocusLink entries) The mouse oligo set was designed using curated NCBI Refseq (78%) and GenBank (22%) sequences. RefSeq/LocusLink version 5/22/2002 was used for validation of oligos (LocusLink representation, cross-homology).</p> |
|---------------------|---|---|---|--|

| | | | | |
|--------------------------|--|---|---|---|
| <p>Rat</p> | <p>Rat OligoLibrary Number of oligos: 4,854. LEADS rat clusters represented: 4,803 (plus 51 controls). UniGene rat clusters represented: 3,821. Public domain rat genes not found as well-characterized genes in the UniGene database: 982. Number of oligos per LEADS cluster: 1. Oligo length: 65 bases (5'-C6-amino modification). Distance from the 3' end: 95% of oligos are within 65–928 bases from the 3' end. The average distance from the 3' end is 302 bases. Cross-homology: 95% of oligos contain a predicted cross-homology of <66.2%. Average cross-homology is 29.2%. Rat UniGene Build #90 was used for mapping oligos to UniGene clusters.</p> | <p>Rat Genome Oligo Set Total number of oligos: 4283. Number of rat genes represented: 4273 (plus 10 controls). Oligo length: 70 bases (5'-C6-amino modification). Distance from the 3'end: oligos are within 750 bases from the 3' end.</p> <p>The oligonucleotides in the Rat Genome Oligo Set were designed from the representative sequences in the Rat UniGene database, Rn build #90.</p> | <p>Pan Rat Liver Oligo Set Number of oligos: 1,408. UniGene clusters represented: 1,353. Number of oligos per UniGene cluster: 1. Oligo length: 50 bases (5'-C6-amino modification).</p> <p>Pan Rat 5K Oligo Set Number of oligos: 5,760. UniGene clusters represented: 5,535. Number of oligos per UniGene cluster: 1. Oligo length 50 bases (5'-C6-amino modification).</p> | <p>Atlas Rat Ready-To-Print Long Oligos Number of oligos: 3,959 LocusLink clusters represented: 3442 Unique genes not represented in LocusLink: 447 Total number of unique genes: 3889 Oligo length: 68 bases + 12 base tag Distance from 3' end: As for human set Cross-homology: More than 98% of oligos have unique sequences (i.e. no sequence match longer than 24 bases (35%) to other LocusLink entries) The rat oligo set was designed using representative sequences from the Rat Unigene database. RefSeq/LocusLink version 5/22/2002 was used for validation of oligos (LocusLink representation, cross-homology).</p> |
| <p>Drosophila</p> | | <p>Drosophila Genome Oligo Set Total number of oligos: 14,617. Number of <i>Drosophila</i> genes represented: 14,593 (plus 24 controls) Oligo length: 70 bases (5'-C6-amino modification). Distance from the 3' end: oligos are within 1000 bases from the 3' end. 14,335 probes are designed from the database na_gadfly.dros.RELEASE2, (March 2002), and developed by the Berkeley <i>Drosophila</i> Genome Project (BDGP), FlyBase, and Celera. 240 probes are designed from the database na_cDNA.dros of full-length cDNA sequences from the BDGP <i>Drosophila</i> Gene Collection 1 and 2 (March 2002). 18 probes are designed from cDNA reference sequences from the <i>Drosophila</i> UniGene database, Dm build #1 (April 2002).</p> | | |

| | | | | |
|--------------------|--|---|--|--|
| <i>Arabidopsis</i> | | <p><i>Arabidopsis</i> Genome Oligo Set Total number of oligos: 26,114. Number of <i>Arabidopsis</i> genes represented: 26,090 (plus 24 controls). Oligo length: 70 bases (5'-C6-amino modification). Distance from the 3' end: oligos are within 1000 bases from the 3' end.</p> <p>The oligonucleotides in the <i>Arabidopsis</i> Genome Oligo Set were designed from the representative sequences in the <i>Arabidopsis</i> UniGene database, At build #4.</p> | | |
| <i>C. elegans</i> | | <p><i>C. elegans</i> Genome Oligo Set Total number of oligos: 7,137. Number of <i>C. elegans</i> ORFs represented: 7,125 (plus 12 controls). Distance from the 3' end: oligos are within 1000 bases from the 3' end.</p> <p>The oligonucleotides in the <i>C. elegans</i> Genome Oligo Set were designed from the WormPep DNA Release 75 (March 2002) database, maintained and developed by the Sanger Institute.</p> | | |
| Yeast | | <p><i>S. cerevisiae</i> Genome Oligo Set Total number of oligos: 6,317. Number of <i>S. cerevisiae</i> ORFs represented: 6,307 (plus 10 controls). Oligo length: 70 bases (5'-C6-amino modification). Distance from the 3' end: oligos are within 750 bases from the 3' end.</p> <p>The oligo set represents the open reading frames (ORFs) of <i>Saccharomyces cerevisiae</i> from the <i>Saccharomyces</i> Genome Database (SGD) at Stanford University.</p> | <p>Pan Yeast Oligo Set Number of oligos: 6,368. UniGene clusters represented: 6,250. Number of oligos per UniGene cluster: 1. Oligo length 40 bases (5'-C6-amino modification).</p> | |

| | | | | |
|--------------------|--|---|--|--|
| <i>C. albicans</i> | | <p><i>C. albicans</i> Genome Oligo Set Total number of oligos: 6276. Number of <i>C. albicans</i> ORFs represented: 6266 (plus 10 controls). Oligo length: 70 bases (5'-C6-amino modification). Distance from the 3' end: oligos are within 1000 bases from the 3' end.</p> <p>The oligonucleotides in <i>the C. albicans</i> Genome Oligo Set were mainly designed from the <i>C. albicans</i>-predicted Open Reading Frame Set Assembly 6 (5921 oligonucleotides). 345 oligonucleotides were designed from cloned gene sequences obtained from GenBank.</p> | | |
| Malaria | | <p><i>Plasmodium falciparum</i> Genome Oligo Set Total number of oligos: 6,369. Number of <i>P. falciparum</i> genes represented: 6,231 (plus 138 controls). Oligo length: 70 bases. Distance from the 3' end: oligos are within 750 bases from the 3' end.</p> <p>The oligo set represents the open reading frames from <i>P. falciparum</i>. The genomic sequence for <i>P. falciparum</i> was obtained from the Sanger Center, The Institute of Genome Research, and Stanford University. From the genomic sequence, open reading frames were predicted using the gene prediction software GLIMMR for malaria, developed at TIGR.</p> | | |

| | | | | |
|---------------------|--|--|--|--|
| <i>H. pylori</i> | | | <p>Pan <i>H. pylori</i> Oligo Set Number of oligos: 1,920. UniGene clusters represented: 1,877. Number of oligos per UniGene cluster: 1. Oligo length: 50 bases (5'-C6-amino modification).</p> | |
| <i>E. coli</i> | | | <p>Pan <i>E. coli</i> K12 Oligo Set Number of oligos: 4,416. UniGene clusters represented: 4,289. Number of oligos per UniGene cluster: 1. Oligo length: 40 bases (5'-C6-amino modification).</p> | |
| <i>B. Subtilis</i> | <p><i>B. Subtilis</i> Oligo Library Number of oligos: 4,128. LEADS clusters represented: 4,106 from <i>B. subtilis</i> genome data release R16.1 (Apr 26, 2001). Oligo length: 65 bases (5' -C6-amino modification).</p> | | | |
| Tuberculosis | | <p>Tuberculosis Genome Oligo Set Total number of oligos: 4,321. Number of <i>M. tuberculosis</i> genes represented: 4,295 (plus 26 controls). Oligo length: 70 bases. Distance from the 3' end: oligos are within 750 bases from the 3' end.</p> <p>A major portion of the oligos were designed from the Sanger Center's set of 3,924 predicted protein-coding gene sequences for the <i>Mycobacterium tuberculosis</i> genome. This set of predicted gene sequences is from the strain H37RV (lab strain), which was entirely sequenced at the Sanger Center. A smaller portion of the oligo set was designed from the strain CDC-1551, sequenced and predicted by The Institute for</p> | | |

| | | | | |
|---------------------|---|---|---|--|
| | | Genomic Research. Note that 371 predicted genes from this strain were found to have less than 97% homology to H37RV predicted genes, and these are included in the oligo set. | | |
| Oligo Design | <p>The production of the oligo libraries is based on clustering and assembly of GenBank EST and mRNA sequences (GenBank version #121 was used for the design). The clustering algorithms take into account known repetitive elements, low complexity regions, chimeric sequences, vector contamination, and correct sequence orientation. The oligo sequences are optimized to minimize the distance from the 3' end, cross-hybridization, and secondary structures, and to normalize the homogeneity of GC content and melting temperature across the collection.</p> <p>Assembly algorithms correct for sequencing errors, predict antisense genes, and use expressed and genomic data to model alternative splicing and predict transcripts for each gene. Using these criteria, each oligo was selected from a sequence segment common to a maximal number of transcripts predicted for a gene.</p> | <p>A proprietary program developed by QIAGEN Operon is used to design oligonucleotides 70 bases in length with normalized melting temperature; that is, having a Tm range of ± 5 °C. The formula used to calculate the Tm is: $T_m = 81.5 + 16.6 \times \log[Na^+] + 41 \times (\#G + \#C)/\text{length} - 500/\text{length}$ where $[Na^+] = 0.1$ M</p> <p>The oligos are designed to hybridize within 1000 bases of the 3' end of the gene (ORF) to accommodate incomplete reverse-transcription reactions using oligo-dT priming. The sequences are optimized to have minimal homology (thus, minimal cross-hybridization) with other genes, by using BLAST against a large data set, including the genome database UniGene. The design allows for the differentiation of overlapping or alternatively spliced genes.</p> | <p>Oligos are selected on the basis of 3' bias and on preferential coding sequence. Redundant sequence entries for the same gene are assembled and the most complete coding sequence is entered into the MWG CodeSeq database. CodeSeq includes alternative splice forms of the same gene. Each species-specific gene index can be searched to find the gene of interest. The same search tells which family group a given gene belongs to and which MWG oligos are needed to detect it, its alternative splice form, or family members. Oligos are checked for cross-hybridization in extensive alignments based upon BLAST and global Smith-Waterman searches. Secondary structure is minimized by testing for reverse-complementarity. The sequences are optimized using BLAST against EST and genomic databases (the UniGene build used is the current release). The MWG CodeSeq database is used for all checks.</p> <p>For <i>in silico</i> design, oligos are selected based on physical parameters including GC content, melting temperature, secondary structure, primer dimer formation, unique binding sites on sequence, etc.</p> | <p>The criteria for designing 68-mer oligonucleotides include, most importantly, gene specificity (lack of crosshybridization potential) and location close to the 3' end of the gene. Extended BLAST searches against GenBank and repeat databases (LINE, SINE, low complexity) are used to define unique regions (if available). Within such regions, oligos are picked to minimize distance to the 3' end, maximize coverage of possible alternative splice forms, and optimize oligo sequence properties. In addition, all oligonucleotides are functionally tested. Hybridization experiments with enzymatic and synthetic antisense oligos are performed, and oligos that fail to produce a strong and specific signal are resynthesized or redesigned. Most failures appear to be due to synthesis quality issues</p> |

| | | | | |
|--|---|--|--|---|
| <p>Oligo Synthesis</p> | <p>Oligo libraries are manufactured by Sigma-Genosys using the proprietary Abacus oligo synthesis platform that provides a coupling efficiency of 98.5%. The minimum amount of full-length sequence represented is 41% for human and 38% for mouse and rat.</p> | <p>All Genome Sets except for the Malaria and the Tuberculosis Genome Oligo Sets carry the 5'-C6-amino modification. The coupling efficiency of synthesis is up to 99.7%, providing an expected yield of about 80% of full-length product. The capping efficiency during synthesis is observed to be up to 100%, resulting in almost no internal deletion products. (The mutation rate is reported to be 1 out of 1000.)</p> | <p>All oligos are synthesized with a proprietary HPSF technology, ensuring a high percent of full-length oligos and no contaminants (salt, metal, etc.). All oligos are 5'-C6-amino-modified 50-mers (some non-mammalian arrays use 40mers where appropriate).</p> | <p>Desalted oligonucleotides were ordered from commercial oligo suppliers (Sigma-Genosys, Operon, others) and are not 5'C6-amino modified. All oligos are made in pairs of sense/anti-sense. The sequence of the oligos is checked by printing sense oligos on an array and hybridising with 1) 'minus'-chain DNA made by using the sense oligo as a DNA polymerase template and 2) synthetic anti-sense oligo mixes.</p> |
| <p>Quality Control Procedures</p> | <p>Quality control includes OD measurements on every oligo. MALDI-TOF mass spectrometry is performed on every batch of synthesized oligos to confirm molecular weight and to ensure correct well location and oligo sequence.</p> | <p>Monitoring is performed for trityl peaks, final trityl peaks and iodine peaks to assess whether the synthesis process is running correctly. Iodine and trityl peaks are measured for all oligos together, not for a single oligo. The final trityl peaks are measured for every oligo to provide information about the last coupling step. The oligos are routinely assessed by UV spectroscopy. (mass spectrometry is not possible for 70-mers). A statistical sample of the oligonucleotides are run on polyacrylamide gel electrophoresis to verify that the individual oligonucleotides have the expected length. Capillary electrophoresis can be performed with a portion of oligos to assess the yield of full-length oligos. UV spectroscopy is used a direct assessment to check the coupling efficiency, because the yield will drop drastically if the coupling efficiency is decreased by 1% or more.</p> | <p>Trityl monitoring and MALDI-TOF mass spectrometry analysis are performed on all oligo sets to ensure 90% purity. Bar coding and tracking ensures 100% identity confirmation of the plated product.</p> | <p>Oligos are functionally tested by printing the entire set on an array and performing hybridizations with separately synthesized antisense oligos. Typically, 15-30% of oligos are reordered due to synthesis failures, e.g. non-homogeneous randomised sequences or sequences with deletions/insertions. Cross-hybridization is tested by hybridizing printed (sense) oligos with subsets of antisense oligos: i.e. a mix of antisense oligos for one array quadrant should produce signals in that quadrant only. Since the oligos are not amino modified, they are chemically stable and can be stored indefinitely in dry form and for at least 1 year in solution at -20C.</p> |

| | | | | |
|-----------------------------|---|---|--|--|
| Supply Format | Oligo library sets are provided as 96-well plates or custom 384-well plates. Oligo library subsets are provided as individual 384-well plates or sets of four 96-well plates. Oligos are arranged on the plates according to standard Gene Ontology (GO) assignments for easy selection of oligos derived from genes with similar function. Concentration: 50 μ M (in water) or 1 nmole. Recommended printing concentration: 20 μ M. Recommended printing surface: all major surfaces. | The genome oligo sets are provided in 384-well plates at 600 pmol per well. Oligonucleotides are shipped as lyophilised samples (dried from water). Recommended printing concentration: 40 μ M. | The oligo sets are provided in 96-well plates or 384-well plates either at 1400 pmoles of oligo per well frozen in water or at the concentration and in the buffer of the customer's choice. The company supplies its own optimized buffers for pin & ring and split-pin arrayers, or customers can supply another buffer the company to use. Recommended printing surface: epoxy and all major surfaces. | The oligo sets are provided in 96-well plates, 400pmol or 1000pmol per well. Hybridization solutions, glass/plastic slides, and hybridisation chambers are provided. Recommended surface for printing: DNA ready glass slides Type II and plastic slides. |
| Information Provided | Interactive Web interface with search capabilities provides gene lists, plate and well locations, accession numbers, sequence information, oligo properties, gene annotation and GO assignments, links to NCBI Entrez and UniGene, proprietary LEADS data on mRNA splice variants, homology reports, SAGE expression reports and more. Gene lists and more information on the technology is available at http://www.labonweb.com/chips . | The genome oligo sets are accompanied by a user manual and disc containing a gene list with accession numbers and well positions, a T _m distribution graph, a readme file with product information, and a resuspension and printing protocol. More information on the Array-Ready Oligo Sets is available via the Operon Microarray database (OMAD) on the Web. This database provides links to the corresponding genome databases, the gene sequence, chromosomal location, and the BLAST search results for each corresponding oligo. | Gene lists and more information on the technology is available at http://www.mwg-biotech.com . | Protocols for printing. Gene lists on CD disk and on the web linked to NCBI web sites LocusLink, GenBank and to Swiss-Prot. Cross-hybridization information for each oligo. Oligo sequence location within the target sequence. Antisense oligos. Hands-on training in Clontech lab. Bioinformatics Internet tools: search engine, interactive oligo browser, lots more at http://bioinfo.clontech.com/atlasinfo/ |

Modified with permission from Bowtell, D.D. & Sambrook, J.F. Eds. *DNA Microarrays: A Molecular Cloning Manual*, (© Cold Spring Harbor Laboratory Press, New York, 2002).