Design of a bovine metabolism oligonucleotide gene array

B.E. Etchebarne[1], W. Nobis, M.S. Allen and M.J. Van de Haar

Department of Animal Science, Michigan State University
East Lansing, Michigan 48824,USA

ABSTRACT

We have designed a bovine metabolism-related microarray chip containing known genes with well-studied functional properties. Using publicly available genomic internet database resources provided by NCBI, TIGR, KEGG and BioCarta, a set of bovine gene sequences was compiled to focus on research in metabolism. Multiple oligonucleotide probes for each gene were designed for spotting on glass slides at sufficient replication to facilitate accurate detection of changes in gene expression. By looking at known genes rather than expressed sequence tags, we hope to gain a better understanding of metabolism and its regulation in cattle.

KEY WORDS: metabolism, gene expression, microarray, statistics

INTRODUCTION

First generation microarrays employing extensive cDNA libraries have allowed high numbers of both known and unidentified genes to be surveyed. Many of these arrays have only one spot per gene, giving no information on within-plate variance. The Human Genome Project has provided extensively annotated databases, such as LocusLink, the Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), The Institute for Genomics Research (TIGR), and BioCarta. These publicly available resources, paired with recent price reductions in oligonucleotide synthesis, allow researchers to feasibly design and produce microarrays with gene sets tailored to specific research areas. Using these databases, we identified approximately 2000 bovine genes representing enzymes of metabolic pathways, metabolic regulators and receptors, transport and binding proteins, intracellular signaling cascades, and cell cycle and apoptotic pathways. Three individual 70mer oligonucleotide probes per gene were

---

[1] Corresponding author: e-mail: madcow@msu.edu

designed for triplicate spotting onto glass slides, giving nine spots per gene. Each oligonucleotide was designed within specific parameters to standardize hybridization behaviour. Use of multiple oligonucleotides per gene improves representation of the expressed fraction of each gene, including splice variants. Spot replication improves within-array quality control and increases the statistical power of accurately detecting small changes in expression at a lower cost than slide replication. Reduction in technical error to increase statistical power is especially important for metabolic research, in which changes in gene expression are often subtle. In addition, our focus on only those genes that are relevant to metabolism improves downstream bioinformatics and data analysis for integration of metabolic gene networks. Because all genes included in this design are annotated with corresponding human homologs, the design can be applied to other species to promote our understanding of comparative metabolism. In conclusion, our design of a focused oligonucleotide microarray with multiple spots per gene will facilitate research in the metabolic genomics of cattle and can be easily applied to other species and disciplines.

MATERIAL AND METHODS

We started with several functional categories: metabolic enzymes, mitogen and mitogen binding proteins, growth factors and cytokines, intracellular signaling proteins, transcriptional control regulatory proteins, apoptotic regulators, and cell-cycle regulatory proteins. A gene list of human DNA sequences pertaining to metabolic genes and genes specific to pathways of interest were extracted from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/ kegg/pathway.html), and BioCarta (http://cgap. nci.nih.gov/ Pathways/BioCarta) sites. Resulting sequences and their annotations to NCBI LocusLink (http: //www.ncbi.nlm.nih.gov/LocusLink/), Swiss-Prot (http://us.expasy.org/), Gene Ontology (www.geneontology.org/) and Online Mendelian Inheritance in Man (OMIM) (www.ncbi.nlm.nih.gov /entrez/query.fcgi? db=OMIM) were gathered and BLASTed against bovine sequences in the GenBank database. Those genes with an NCBI Reference Sequence (RefSeq), complete sequence, or 3'-end sequence were saved for oligonucleotide design. Next, we used keywords corresponding to remaining categories of interest to search the Gene Ontology, UniGene and Swiss-Prot databases to gather sequences of known genes and collect their bovine sequences in the same manner. This process generated a large list of genes corresponding to a UniGene cluster unique ID number with a specific accession number as an identifier. A cluster is a compilation of sequences of overlapping sequences that represent one gene (http://www.ncbi.nlm.nih.gov /entrez/query.fcgi?db=unigene).We decided to use only those bovine genes with a RefSeq link, complete sequence, 3'-end sequencing, or strong TIGR

(http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=cattle) cluster match to facilitate oligonucleotide probe design. A verified gene sequence can be used to design oligonucleotides to be used as gene probes on a chip.

RESULTS

Out of the 2180 target human sequences for which we searched, our final dataset for the BMET array contained 1972 genes. This dataset included 426 bovine RefSeq genes, 109 completely sequenced genes, and 773 3'-end United States Department of Agriculture sequenced genes and 664 genes with a TIGR cluster match. All matches had an expectation value of better than e^{-35}. Of the 1207 KEGG metabolism genes extracted, greater than 1100 are represented in this dataset.

DISCUSSION

Production of an oligonucleotide gene array begins with the designing of oligonucleotide probes corresponding to each gene of interest. As a preliminary screen, we used the OligoPicker software (pga.mgh.harvard.edu/oligopicker) to design up to three 70 mer oligonucleotides per gene for slide production. Specific parameters were set to ensure similar probe binding properties for each oligonucleotide. Three oligonucleotides per gene were successfully designed for all 1972 genes in the dataset. The use of oligonucleotide microarrays involves ordering of custom 70 mer oligonucleotides from a commercial company to be printed on Poly-L-lysine coated glass slides. Triplicate spotting of each of the three oligonucleotides designed (9 spots per gene) will ensure adequate spot replication for downstream data analysis. This custom set of oligonucleotides will be ordered from a commercial company. The BMET array will contain approximately 19.000 spots when replicate spots (n=10) of both positive and negative control genes (including GAPDH and Beta-actin) are added. After a series of postprocessing fixation steps, the printed array is ready for hybridization of Cy3- and Cy5-labeled cDNA target samples from two biological samples being compared for differential gene expression. Following hybridization the array is analysed using a fluorescence scanner that is able to quantify the relative amounts of mRNA present in each of the two samples based on the abundance of the two flourophores bound to a homologous array spot.
As part of our effort to build a strong bovine metabolic genome resource, we are currently developing a web-based analytical tool to help with downstream analysis of the BMET array. Known human metabolic and cell regulatory pathways represented within KEGG and BioCarta will be adapted to this bovine-specific gene set using the GenMAPP tool (http://www.gen mapp. org/) and a GeneLink resource. This will be housed at our website www.meta bolism.msu.edu.

CONCLUSIONS

Array technology has drawn criticism for inadequate reproducibility, which arises due to technical issues in target preparation, and array fabrication and design. Typical arrays are spotted with between one and three spots per gene. This poses a problem for detection of statistical differences between samples. Our use of multiple oligonucleotides per gene is intended to improve representation of the expressed fraction of each gene, including splice variants. Additionally, spot replication within slides will improve within-array quality control and increase the statistical power of detecting small changes in expression at a lower cost than slide replication. Thus, biological variation should be the major determinant of the number of arrays needed in a study. This maximizing of statistical power is especially important for metabolic research, in which changes in gene expression are often subtle.

By our feed-forward approach to designing a gene array specific to bovine metabolism and its related cell regulatory pathways using only well-defined sequences and gene signalling pathways, we will increase the quality and utility of gene expression data that can be acquired from microarray technology.

WEBSITE REFERENCES

http://www.genome.jp/kegg/pathway.html; Kyoto Encyclopedia of Genes and Genomes pathways website

http://cgap.nci.nih.gov/Pathways/BioCarta; Pathways obtained from Biocarta via the Cancer Genome Anatomy Project

http://www.ncbi.nlm.nih.gov/LocusLink/; NCBI LocusLink

http://us.expasy.org/; The ExPASy (Expert Protein Analysis System) proteomics server of the Swiss Institute of Bioinformatics

http://www.geneontology.org/; The Gene Ontology consortium website

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM; The Online Mendelian Inheritance in Man database. A catalog of human genes and genetic disorders

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene; NCBI's UniGene link. This system automatically partitions GenBank sequences into a non-redundant set of gene-oriented clusters

http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=cattle; The Institute for Genome Research's bovine gene index

http://pga.mgh.harvard.edu/oligopicker; A publicly-available oligonucleotide design program

http://www.genmapp.org; GenMAPP is a free computer application designed to visualize gene expression data on maps representing biological pathways and groupings of genes

http://www.metabolism.msu.edu/; Home of the BMET microarray database