provide a clear label. In those circumstances, GMO labeling is analogous to labeling for religious reasons (as in kosher or halal labeling), where personal belief systems are given priority over the reasonable-risk standard.

Conversely, as yet unknown GMO-containing products might pose a dire risk of future immune dysfunction to newborns or present an unsuspected allergenic risk to persons with a unique genetic makeup. In this instance, a label would serve a crucial public-health purpose, allowing epidemiologists to pursue a trail of GMO ingredients in the wake of a suspected outbreak. Either way, labeling is a declaration of the public's right to know.

The virtue of labeling is that it satisfies the ethical requirement of meeting transparency rules, and that it honors all concerns about GMOs along a spectrum of need-to-know, from the health or allergen perspective to the moral and religious viewpoint. Avoiding GMOs on principle, out of disapproval of this technology in modern agriculture, is just as appropriate a reason to encourage labeling as is the need to know from a general consumer perspective. To declare, as Reiss does, that GMO labeling should be discretionary for the manufacturer or seller assumes that the producer has the same right of choice as do consumers. Freedom of choice does not extend to placing consumers at risk, nor to obfuscating product ingredients. The notion of caveat emptor long ago ceased to be a legitimate part of the free enterprise system, even in the United States.

Marc Lappé
Director, Center for Ethics & Toxics
Gualala, CA 95445
(mlappe@mcn.org)

Michael Reiss responds:
I am grateful to Marc Lappé for his letter. There are clearly some points on which he and I agree, including the fact that those who make and those who sell foods have certain duties to consumers. However, we differ on several points. I am not convinced by his argument that labels should be required so as to permit "epidemiologists to pursue a trail of GMO ingredients in the wake of a suspected outbreak." First, I doubt that such adverse health consequences in well regulated countries are any more likely to result from GM than from non-GM foods; secondly, food tracing in a health emergency is generally much more fine-grained—relying on batch details and so forth rather than broad labeling.

More fundamentally, while I accept his point that "labeling is a declaration of the public's right to know," few, if any, rights are absolute. A person who incites others to

violence forfeits the right to freedom of speech. Mandatory labeling, as I argued in my commentary, has counterbalancing disadvantages: in particular, it costs money and reduces the freedom of retailers.

Finally, there is the rather practical problem of how much we should require to be on a label. Contrary to Marc Lappé, I am not in most cases in favor of labeling being required for religious reasons. Those who wish to restrict their food purchase and consumption for religious reasons do not, in my view, have a general right to insist on such labeling (and I write as a priest in the Church of England). Indeed, if they did, we might soon run out of all space on many food products for legible labeling as various minority groups—religious and lifestyle—claimed the statutory right to their particular labeling requirements.

An open-source oligomicroarray standard for human and mouse

To the editor:
DNA microarrays have transformed biology, allowing the expression levels of thousands of genes to be measured simultaneously[1,2]. However, the variety of chip designs and the lack of information about the probes used severely limits the use of the data. As a way of addressing these problems, we have designed a probe standard, a set of 121,000 70-base oligonucleotides for the mouse and human transcriptomes, together with a probe selection algorithm. We make the probe sequences and the code for the algorithm freely available on our web site, http://arep.med.harvard.edu/ probes.htm.

Open-source standards of this sort have many advantages. With sequence and design information available, probe quality can be closely monitored and the community can suggest improvements. Probe sequences can be directly linked to microarray measurements, allowing detailed troubleshooting and analysis of biologically meaningful fine structure[3,4]. By annotating the probe sequences, the probes and design methods can be optimized by iteration of experiment and probe design.

Our algorithm terminates once it has found probes that satisfy a set of conditions for sensitivity, specificity, and uniformity, rather than searching for optimal probes as, for 70-mers, many equally specific and sensitive probes may be chosen for the same target[5]. Using this strategy, our algorithm can select probes for human and mouse where algorithms that demand optimality are unsuccessful[6].

The probes are 70-mers, selected from the representative transcripts of the UniGene database (http://www.ncbi.nlm.nih/UniGene; August 3, 2001, build #138). We identify them by the GenBank identifier and the UniGene cluster of their target so that they can be tracked through future UniGene updates. Our criteria are a BLAST[7] threshold for specificity[5,8], secondary structure prediction[9], and sequence complexity calculations for sensitivity (J. Deris, personal communication), and a melting temperature window (see http://www.operon.com/oligos/webfaq.php#calculations) for uniformity. We also select probes from within 1,000 bases of the 3′ end of the transcript to reflect the 3′ bias generated by poly(A) priming during

Table 1. Summary statistics for human and mouse probe sets and comparison with human probe set from Operon Technologies*

Probe set	Human (*n* = 65,0652)[a]				Mouse (*n* = 56,037)			
Property	Min	Max	Average	Stdev	Min	Max	Average	Stdev
T_m[b]	73	83	76.2	2.9	70	78	74	2.7
	71	83	77.1	2.4	–	–	–	–
LZ[c]	21	39	29.3	1.2	26	31	29.2	0.9
	10	41	29.2	1.2	–	–	–	–
RNA[d]	0.1	33.8	10.2	4.6	0.1	32.6	7.8	3.7
	0.1	36.5	13.8	4.7	–	–	–	–
Distance[e]	1	1,000	334.7	242.3	1	1,000	219	197.7
	1	4,297	244.8	179.1	–	–	–	–
BLAST[f]	0	0	0	0	0	0	0	0
	0	70	14.5	25.9	–	–	–	–

*Shaded rows, our oligos; unshaded rows, Operon Technology oligos. Min, minimum score for property; Max, maximum score for property; Average, average score for property; Stdev, standard deviation. [a]Number of transcripts from which probes were selected. [b]Melting temperature calculated by the formula $T_m = 81.5 + (16.6\log[Na^+]) + 41[(G+C)/length] - (500/length)$ with $[Na^+]$ (sodium ion concentration) taken to be 0.1 (see http://www.operon.com/oligos/webfaq.php#calculations). Other methods of T_m calculation exist (notably that of SantaLucia *et al.*[10]), but for 70-mers, the difference between the T_m values calculated using these methods is negligible. [c]Sequence complexity score calculated by comparing the length of the string (in characters) before and after compression with the Lempel-Ziv algorithm gzip. [d]RNAfold score ($-\Delta G$) calculated by the Vienna group's RNAfold algorithm. [e]Distance of 3′-most base of the probe from the 3′ end of the transcript. [f]Maximum number of identities in a BLAST alignment (with bit score over 50) with a transcript in a different cluster of UniGene.

the cDNA synthesis step of many microarray protocols. Although there are alternative ways of predicting specificity[6,8] and melting temperature[10], none of these alternatives has been shown to be significantly more valuable for 70-mer selection. More details of our methods are available at http://arep.med.harvard.edu/probes.htm.

Table 1 summarizes data for our probe sets and compares human probes selected using our design criteria with those selected by Operon Technologies (Alameda, CA; http://www.operon.com/arrays/arraysets.php). The comparison was carried out between probes designed for the same UniGene cluster by calculating the score of each probe for each of the criteria used in our algorithm. Our human probe set is much larger than Operon's (65,062 versus 13,975) and contains no significant BLAST hits (bit score > 50), whereas Operon's contains 1,496.

Using our initial criteria, we were able to choose probes from 56,037 mouse and 65,062 human UniGene clusters[11]. Our algorithm did not find probes that met these criteria from among the remaining one-third of the UniGene clusters. We are currently working to choose probes from these remaining clusters by making the criteria less stringent, specifically the 3′ distance restriction. It should be noted that some UniGene clusters may contain multiple splice variants. In these cases, the probes target a particular exon or splice junction in the representative transcript and are linked to this transcript by their position in the sequence and GenBank identifier.

In the future, as knowledge of the human and mouse transcriptomes increases, the list of probes will include each exon and splice junction of each transcript from which subsets could be chosen for particular applications. The list presented here is currently the largest freely available list of probes for which the selection criteria are rigorous and publicly disclosed. Synthesis and testing of the probes is currently underway. They are intended as a community resource, and input about the probes or the design methods is welcome. This is a starting point from which to develop an optimized and comprehensive transcriptional oligonucleotide standard. Indeed, our designs are currently being tested by the Programs in Genomics Applications initiative of the National Heart, Lung, and Blood Institute, a large-scale collaborative functional genomics effort that involves groups from across the United States. In choosing our criteria and developing our algorithm, we are grateful to Joe DeRisi and to Operon Technologies, as well as to Affymetrix (Santa Clara, CA), which has released sequence information for its *Escherichia coli* chip.

Matthew A. Wright and George M. Church
Department of Genetics
Harvard Medical School
Boston, MA 02115
(church@arep.med.harvard.edu)

1. Schena, M. *et al. Science* 270, 467–470 (1995).
2. Gress, T.M. *et al. Mamm. Genome* 3, 609–619 (1992).
3. Badarinarayana, V. *et al. Nat. Biotechnol.* 19, 1060–1065 (2001).
4. Selinger, D.W. *et al. Nat. Biotechnol.* 18, 1262–1268 (2000).
5. Hughes, T.R. *et al. Nat. Biotechnol.* 19, 342–347 (2001).
6. Li, F. & Stormo, G.D. *Bioinformatics* 17, 1067–1076 (2001).
7. Altschul, S.F. *et al. J. Mol. Biol.* 215, 403–410 (1990).
8. Kane, M.D. *et al. Nucleic Acids Res.* 28, 4552–4557 (2000).
9. Wuchty, S. *et al. Biopolymers.* 49, 145–165 (1999).
10. SantaLucia, J., Jr., Allawi, H.T. & Seneviratne, P.A. *Biochemistry* 35, 3555–3562 (1996).
11. Boguski, M.S. & Schuler, G.D. *Nat. Genet.* 10, 369–371 (1995).