Supporting Text

Microarray Hybridization and Analysis. RNA from cells cultured in the presence or absence of Sonic hedgehog (Shh)-N for 6 h was converted to double-stranded cDNA by using the Superscript Choice cDNA synthesis kit (Invitrogen) and a T7-dT(24) primer (Genset/Proligo, Boulder, CO). cDNA was transcribed to generate cRNA by using a T7-transcription/labeling kit from Enzo Life Sciences. Fifteen-microgram cRNA samples were hybridized to Affymetrix (Santa Clara, CA) Mu11K GeneChips, chips were scanned, and hybridization data were acquired by using Affymetrix MICROARRAY SUITE 5.0 software. Data were then subjected to probe-level analysis with the R package (1), which carries out normalization, estimates gene expression levels, and determines statistical significance. Briefly, to account for local background differences within each microarray, raw hybridization intensities were corrected by using mismatch probe hybridization data and then normalized by using a variance-stabilizing algorithm (VSN) (2). The level of each transcript was determined by modeling the data from all the corresponding array oligonucleotides by using the Li–Wong model (3). The difference in transcript level between untreated and Shh-treated samples was scale-normalized and averaged across five replicate experiments to obtain an estimate of the difference in expression for each gene (average expression difference, mbar). To determine the significance of differences in gene expression from the five replicate experiments, empirical Bayesian analysis (4) was conducted by using a hyper-parameter of $P = 0.015$. For each gene, the statistical confidence [logarithm of odds ratio (lod)] was calculated as follows:

$$\text{lod} = \log_{10} \frac{\text{Probability (the gene is differentially expressed} \mid \text{observed data)}}{\text{Probability (the gene is not differentially expressed} \mid \text{observed data)}}$$

Logarithm of odds scores were used to rank the statistical confidence of differentially expressed genes, with the highest lod scores considered the most significant. We identified 134 genes with lod > 0 (Fig. 6). By randomly permuting gene expression data with 500 iterations, we estimated the false-discovery rate among these genes to be <1% (5). The identities of the differentially expressed genes were verified and updated on October 16, 2002, by integrating data from the Affymetrix and Unigene databases. Information regarding gene function was obtained by using

Gene Ontology data from the mouse and human Unigene databases, supplemented with literature searches. Numerical analyses were run on UNIX-based IBM p690 server and Microsoft Windows NT-based Intel Pentium III platforms.

1. Irizarry, R. A., Gautier, L. & Cope, L. (2003) in *The Analysis of Gene Expression Data: Methods and Software*, eds. Parmigiani, G., Garrett, E. S., Irizarry, R. A. & Zeger, S. L. (Springer, New York).
2. Huber, W., Von Heydebreck, A., Sultmann, H., Poustka, A. & Vingron, M. (2002) *Bioinformatics* **18,** Suppl. 1, S96–S104.
3. Li, C. & Wong, W. H. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 31–36.
4. Lonnstedt, I. & Speed, T. (2002) *Statistica Sinica* **12,** 31–46.
5. Tusher, V. G., Tibshirani, R. & Chu, G. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 5116–5121.