



## Noncoding Sequences Conserved in a Limited Number of Mammals in the *SIM2* Interval are Frequently Functional

Kelly A. Frazer, Heng Tao, Kazutoyo Osoegawa, et al.

*Genome Res.* 2004 14: 367-372

Access the most recent version at doi:[10.1101/gr.1961204](https://doi.org/10.1101/gr.1961204)

---

**References** This article cites 21 articles, 11 of which can be accessed free at:  
<http://genome.cshlp.org/content/14/3/367.full.html#ref-list-1>

### License

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Cold Spring Harbor Laboratory Press

# Noncoding Sequences Conserved in a Limited Number of Mammals in the *SIM2* Interval are Frequently Functional

Kelly A. Frazer,<sup>1,3</sup> Heng Tao,<sup>1</sup> Kazutoyo Osoegawa,<sup>2</sup> Pieter J. de Jong,<sup>2</sup> Xiyin Chen,<sup>1</sup> Mark F. Doherty,<sup>1</sup> and David R. Cox<sup>1</sup>

<sup>1</sup>Perlegen Sciences, Mountain View, California 95051, USA; <sup>2</sup>Children's Hospital and Research Center at Oakland, Oakland, California 94609, USA

Cross-species DNA sequence comparison is a fundamental method for identifying biologically important elements, because functional sequences are evolutionarily conserved, whereas nonfunctional sequences drift. A recent genome-wide comparison of human and mouse DNA discovered over 200,000 conserved noncoding sequences with unknown function. Multispecies DNA comparison has been proposed as a method to prioritize these conserved noncoding sequences for functional analysis based on the hypothesis that elements present in many species are more likely to be functional than elements present in limited numbers of species. Here, we perform a comparative analysis of the single-minded 2 (*SIM2*) gene interval on human chromosome 21 with horse, cow, pig, dog, cat, and mouse DNA. We classify conserved sequences based on the number of mammals in which they are present, and experimentally test sequences in each class for function. As hypothesized, conserved sequences present in many mammals are frequently functional. Additionally, we demonstrate that sequences conserved in a limited number of mammals are also frequently functional. Examination of genomic deletions in chimpanzee and rhesus macaque DNA showed that several putatively functional conserved noncoding human sequences were absent in these primates. These findings suggest that functional conserved noncoding human sequences can be missing in other mammals, even closely related primate species.

[Supplemental material is available online at [www.genome.org](http://www.genome.org).]

Previous comparative DNA analyses have shown that the human chromosome 21 (21q) interval surrounding *SIM2*, a transcription factor that is a key regulator of central nervous system development (Nambu et al. 1991; Ema et al. 1999), contains a large number of noncoding sequences that are conserved between humans and mice (Frazer et al. 2001; Dermitzakis et al. 2002; Mural et al. 2002). It is generally well proven that noncoding conserved regions can represent functional regulatory elements (Loots et al. 2000; Schwartz et al. 2003; Thomas et al. 2003). Thus, these conserved noncoding elements are likely involved in the transcriptional regulation of *SIM2*, but experimental functional examination of them all is difficult. Here, we analyze the 365-kb *SIM2* interval to test the utility of using multispecies sequence conservation for prioritization of conserved noncoding sequences for functional characterization.

## RESULTS AND DISCUSSION

### Isolation of Syntenic BAC Sequences

When performing cross-species DNA comparisons to identify functional elements on the basis of evolutionary conservation, it is important to use orthologous sequences (DNA sequences in different species that are derived from the same genomic interval in the last common ancestral species). We isolated horse, cow, pig, dog, cat, and mouse sequences orthologous to the human 21q *SIM2* interval in parallel using 14 universal oligonucleotide

hybridization probes (overgos) designed on the basis of sequences conserved between humans and mice (Thomas et al. 2002) to screen arrayed BAC libraries (<http://bacpac.chori.org/>). We selected a minimally overlapping set of BAC clones that span the *SIM2* interval in the horse, cow, pig, dog, cat, and mouse genomes by STS-content mapping and restriction enzyme digest-based DNA fingerprinting (Marra et al. 1997; Supplemental Fig. 1 available online at [www.genome.org](http://www.genome.org)).

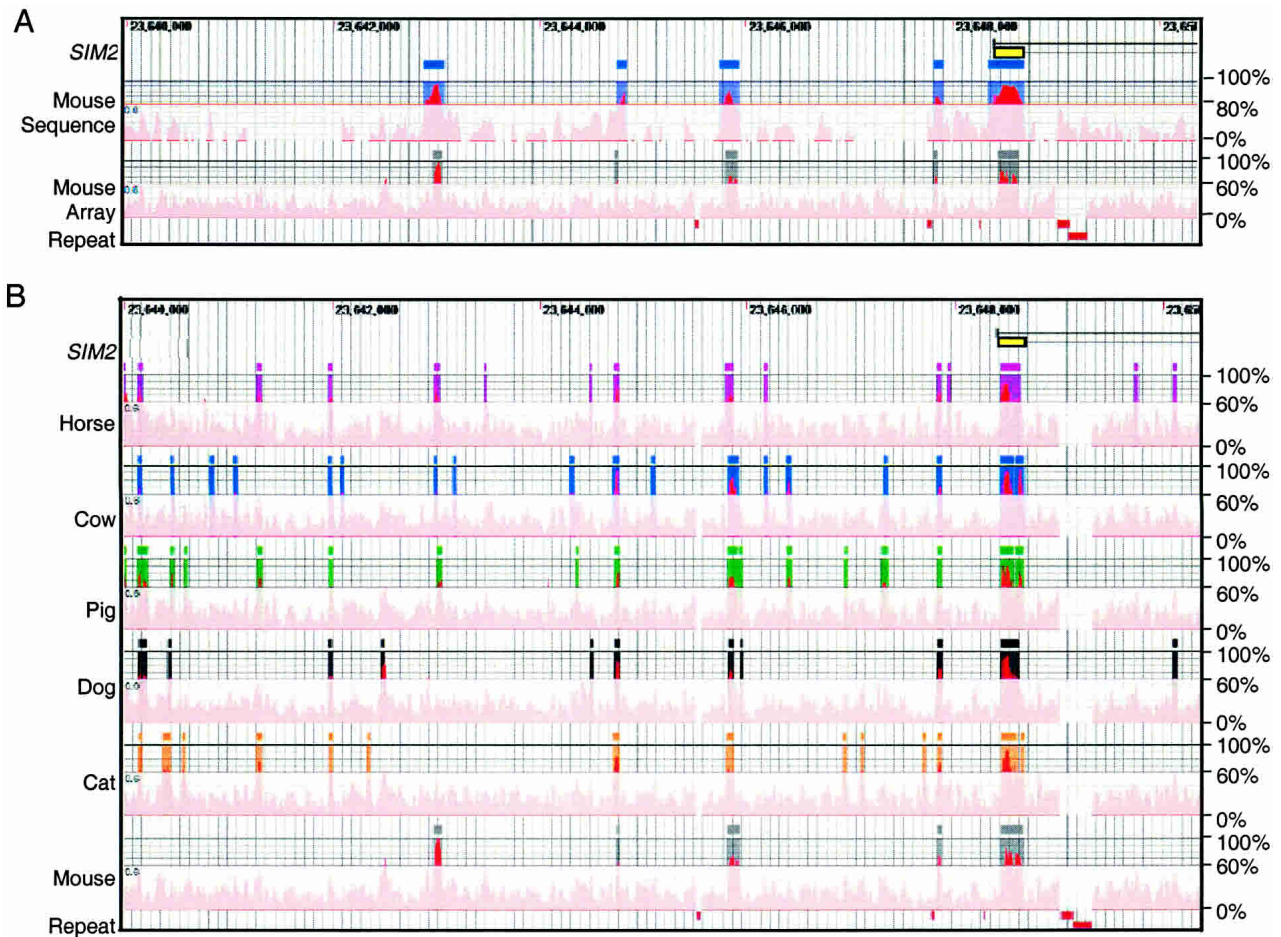
### Identification and Analysis of Evolutionarily Conserved Human Sequences

To identify sequences that are evolutionarily conserved between humans and mice, DNA isolated from the set of minimally overlapping mouse BAC clones was pooled, fluorescently labeled, and hybridized to high-density oligonucleotide arrays containing probes for the unique sequences in the *SIM2* interval on human 21q (Frazer et al. 2001). Sequences in the 365-kb 21q *SIM2* interval were classified as evolutionarily conserved on the basis of the analysis of the comparative array data using an algorithm described previously (Frazer et al. 2001) that we modified for increased sensitivity (see Supplemental method). To estimate the percent of conserved elements that we failed to detect by array hybridization, we compared the array data with conserved elements identified by analyzing orthologous human chromosome 21 and mouse chromosome 16 sequences aligned using the BLASTZ algorithm (Fig. 1A; Schwartz et al. 2003). We identified 72 conserved human–mouse elements ( $\geq 100$  nt length and  $\geq 80\%$  identity) by sequence alignments in the 365-kb *SIM2* interval, of which 86% overlap conserved human–mouse elements

### <sup>3</sup>Corresponding author.

E-MAIL [kelly\\_frazer@perlegen.com](mailto:kelly_frazer@perlegen.com); FAX (650) 625-4510.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.1961204>. Article published online before print in February 2004.



**Figure 1** Conserved human sequences in the 10-kb interval surrounding the first exon of *SIM2* (yellow rectangle). The conserved elements are shown relative to their position in the human reference sequence (horizontal axis), and their percent conformances (vertical axis). (Position numbers) The position on human 21q (NCBI contig NT\_002836). (A) Rectangles rising above the veil indicate conserved human–mouse elements identified by two methods; (blue) sequence alignment ( $\geq 100$  bp and 80% identity); (gray) 21q array data ( $\geq 30$  bp and 60% conformance). Red rectangles at the bottom indicate the positions of interspersed repeats, which were not tiled on the arrays, and therefore, conformance information is absent. (B) Rectangles indicate conserved human elements ( $\geq 30$ -nt length and  $\geq 60\%$  conformance) identified by comparisons of horse, cow, pig, dog, cat, and mouse DNA by hybridization to human 21q arrays.

( $\geq 30$  nt length and  $\geq 60\%$  conformance) in the comparative human–mouse array data.

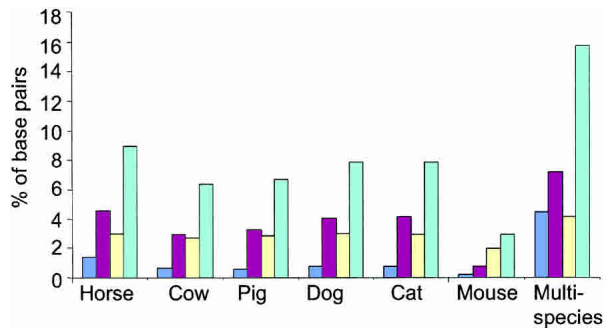
Similarly, to identify conserved elements ( $\geq 30$  nt length and  $\geq 60\%$  conformance) between humans and horses, cows, pigs, dogs, and cats in the *SIM2* interval, we hybridized the BAC clones selected for each of these species to the human 21q high-density arrays. For all five mammalian species (horse/cow/pig/dog/cat), the comparative array data detected  $>93\%$  of the 72 conserved human elements identified by human–mouse sequence alignments. These results reflect the greater similarity at the nucleotide level between humans and these five species than between humans and mice. Thus, analysis of the array data for these other mammals identifies a greater number of conserved elements (Supplemental Tables 2–7 and Supplemental Fig. 2), and the false-negative rate is expected to be lower.

We determined the relative levels of sequence conservation between humans and each of the species examined in this study by analyzing the six DNA comparisons (human–horse, human–cow, human–pig, human–dog, human–cat, and human–mouse) individually. The number of base pairs within conserved elements ( $\geq 30$  nt length and  $\geq 60\%$  conformance) ranged from

$\sim 3.0\%$  in the human–mouse DNA comparison to  $\sim 9.0\%$  in the human–horse DNA comparison (Fig. 2).

### Classification of Evolutionarily Conserved Sequences on the Basis of the Number of Mammals in Which They Are Present

We next classified the conserved human sequences on the basis of the frequency of conservation in the six mammals examined; unique (present in humans and only one of six mammals), limited (present in humans and between two to five of the mammals), or common (found in humans and all six of the mammals analyzed; Figs. 1B, 2). Interestingly, the majority of the conserved base pairs identified in the human–mouse comparison are in the common class. This contrasts with the horse, cow, pig, dog, and cat DNA comparisons for which the limited class contains the largest percentage of conserved base pairs (Supplemental Table 8). For example, in the 10-kb interval shown in Figure 1B, four human–mouse conserved elements are common and one is in the limited class, whereas four human–dog conserved elements are common, six are limited, and one is in the unique class. These results suggest that a considerably larger fraction of the human



**Figure 2** Evolutionarily conserved base pairs in a 365-kb region on human chromosome 21 containing the *SIM2* locus. Six species are compared with human genomic DNA for evolutionary conservation. The multispecies data set represents the composite of conserved sequences identified in all six species. The bars represent the percent of the 365,000 bp that are in elements conserved ( $\geq 30$ -nt length and  $\geq 60\%$  conformance) between humans and different numbers of the six species. Base pairs conserved only between the indicated species and humans (unique, dark blue); base pairs conserved between humans and two to five species (limited, red); base pairs conserved in all mammals analyzed (common, yellow). Pale blue bars represent the total percent of base pairs conserved. For details see Supplemental information.

genome is under evolutionary constraint than that detected by human–mouse DNA comparisons (Waterston et al. 2002), and thus comparing human sequences with the DNA of multiple species will be important for generating a comprehensive list of evolutionarily conserved sequences in the human genome.

To determine how much of the total DNA sequence in the human 365-kb *SIM2* region is conserved, we collapsed the six comparative analyses into a single composite data set (human–multispecies). Approximately 57.5 kb of the base pairs are conserved between humans and at least one of the six mammalian species analyzed (Fig. 2). Of these conserved human sequences, ~46% is in the limited class, ~28% is in the unique class, and ~26% is in the common class. Further analysis of the human–multispecies data set revealed that ~17% of the conserved human base pairs overlap either *SIM2* or holocarboxylase synthetase (*HLCs*) exons, whereas ~83% correspond to sequences not in known exons (data not shown). Thus, the majority of the conserved sequences in the human 21q *SIM2* interval examined are noncoding and present in a limited number of mammals.

### Functional Analysis of Conserved Noncoding Sequences in the Limited and Common Classes

To address the question of whether conserved noncoding sequences present in all mammals are more likely to be functional than those present in a limited number of mammals, we used two *in vitro* assays. Ten evolutionarily conserved sequences (three common and seven limited) and six nonconserved sequences were chosen for functional characterization (Fig. 3A). We first tested the 16 noncoding sequences for their ability to interact with the *SIM2* promoter (Yamaki et al. 2001) to drive the expression of a luciferase reporter gene in transient infections of a human glioblastoma cell line (Fig. 4A). The 10 reporter constructs containing conserved noncoding sequences have, on average, a significantly greater level of luciferase expression than the six reporter constructs containing nonconserved noncoding sequences ( $P \sim 0.0047$ ). Two of the conserved noncoding regions (c4 and c7) tested were composed of multiple conserved noncoding elements. Both of these reporter constructs had higher luciferase expression levels than when the individual conserved sequences were examined separately, suggesting that the conserved elements are acting cooperatively and their effect is additive.

However, there were no differences in the levels of expression between the luciferase reporter constructs containing common conserved sequences and those containing limited conserved sequences, suggesting that elements in both classes are equally likely to functionally interact with the *SIM2* promoter.

Noncoding transcriptional regulatory elements typically bind proteins (Gill 2001). We examined eight of the 10 conserved elements and the six nonconserved elements for their ability to bind proteins when incubated with glioblastoma nuclear extract by electrophoretic gel-shift analysis. Of the eight conserved sequences tested, all six of the limited class elements and one of the two common-class elements bound protein(s). None of the six nonconserved sequences bound protein(s). Interestingly, several of the conserved elements in close proximity to one another (those in interval 1 as well as those in interval 3, see Fig. 3) have highly similar mobility-shift patterns and may or may not be binding the same protein(s) (Fig. 4B). These data indicate that elements conserved in a limited number of mammals are just as likely to bind proteins as elements conserved in many mammals.

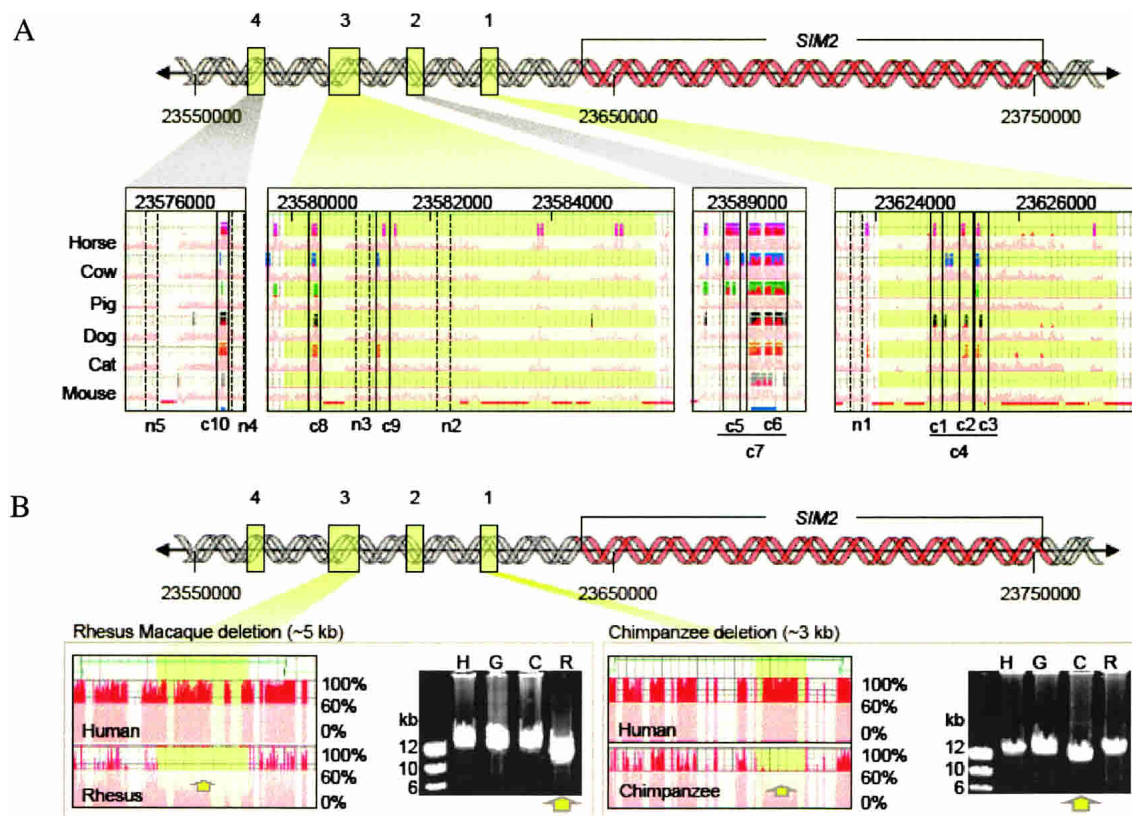
### Deletion of Putatively Functional Human Sequences in Other Primates

To determine whether any of the conserved elements identified by sequence comparisons between humans and mammals (horse, cow, pig, dog, cat, and mouse) are absent in some primate species, we examined two genomic DNA deletions (one in the chimpanzee genome and one in the rhesus macaque genome) that we observed previously (Frazer et al. 2003) in DNA sequences upstream of the *SIM2* gene (Fig. 3B). The chimpanzee DNA deletion (~3 kb in length) results in the loss of three limited conserved elements in interval 1, and the rhesus macaque deletion (~5 kb in length) results in the loss of two limited conserved elements in interval 2. All five of these limited conserved elements demonstrated functionality in at least one of the two *in vitro* assays (luciferase expression and protein binding). Our results indicate that conserved noncoding human sequences that are putatively functional can be absent even in evolutionarily closely related primate species.

### Conclusions

Previous multispecies comparative DNA analyses have shown that mammals share different evolutionarily conserved noncoding sequences in common with each other (Frazer et al. 2001; Chureau et al. 2002; Boffelli et al. 2003; Hare and Palumbi 2003; Thomas et al. 2003). These studies have largely proposed that sequences conserved in most species are more likely to be functional than sequences conserved in only some species. However, we show that elements conserved in a limited number of mammals are just as likely to be functional as elements conserved in many mammals. Thus, our data suggests that one should not use the criteria of sequence conservation in most mammals for prioritization of noncoding sequences for functional characterization.

The limited class of conserved human elements is composed of rapidly evolving functional sequences on the basis of the observation that conserved sequences in the limited class tend to be shorter in length than conserved sequences in the common class (see Supplemental Table 8), and the fact that they are only found in a subset of the mammalian species. We propose that the rapidly evolving limited class of noncoding functional sequences may be responsible for gene expression differences between species. For example, the *SIM2* gene is likely to be regulated by numerous conserved sequences having redundant and/or additive effects on the expression pattern of the gene, with the evolution-



**Figure 3** Noncoding sequences upstream of *SIM2* chosen for functional characterization. (A) Conserved sequences (c1–c10) and nonconserved sequences (n1–n5) are located within four different intervals (coordinates based on NCBI contig NT\_002836) (n6 is located ~94 kb upstream). Visualization plots (see Fig. 1) show the species in which elements c1–c10 are conserved. (B) Conserved sequences deleted in the chimpanzee and rhesus macaque genomes. Comparison of syntenic human (H), gorilla (G), chimpanzee (C), and rhesus macaque (R) long-range PCR (LR-PCR) products by gel electrophoresis shows the deletion of interval 1 conserved sequences in chimpanzee genomic DNA and the deletion of interval 3 conserved sequences in rhesus macaque genomic DNA (yellow arrows). The visualization plots, generated by hybridization of the primate LR-PCR products to the human 21q arrays, indicate the positions of the human sequences (highlighted in yellow in A and B) deleted in the chimpanzee and rhesus macaque genomes. The deletions correspond to the drop in percent conformance, plotted on the vertical scale relative to the position in the human reference sequence. Green horizontal lines at the top of the plots indicate the positions of the LR-PCR products.

ary loss or gain of some regulatory elements resulting in expression differences ranging from small to large.

## METHODS

### BAC Library Screening and Contig Assembly

Sequences conserved between humans and mice (ranging from 50–200 bp in length and 90%–99% identity) in the 280-kb region surrounding *SIM2* were used to design overgo hybridization probes. Paired oligonucleotides (24 nt) containing eight base overlapping regions at the 3'-end (Supplemental Table 1) were annealed and labeled with  $^{32}\text{P}$  at A or C position with Klenow. High-density replica filters were prepared from horse (CHORI-241), cow (RPCI-42), pig (RPCI-44; Fahrenkrug et al. 2001), cat (RPCI-86), and mouse (RPCI-23; Osoegawa et al. 2000) BAC libraries. The overgo probes were hybridized to the filters as described previously (Osoegawa et al. 2000). All positive BACs were analyzed by standard fingerprinting methods (Marra et al. 1997). Overlapping contig maps were built using FPC software from each library.

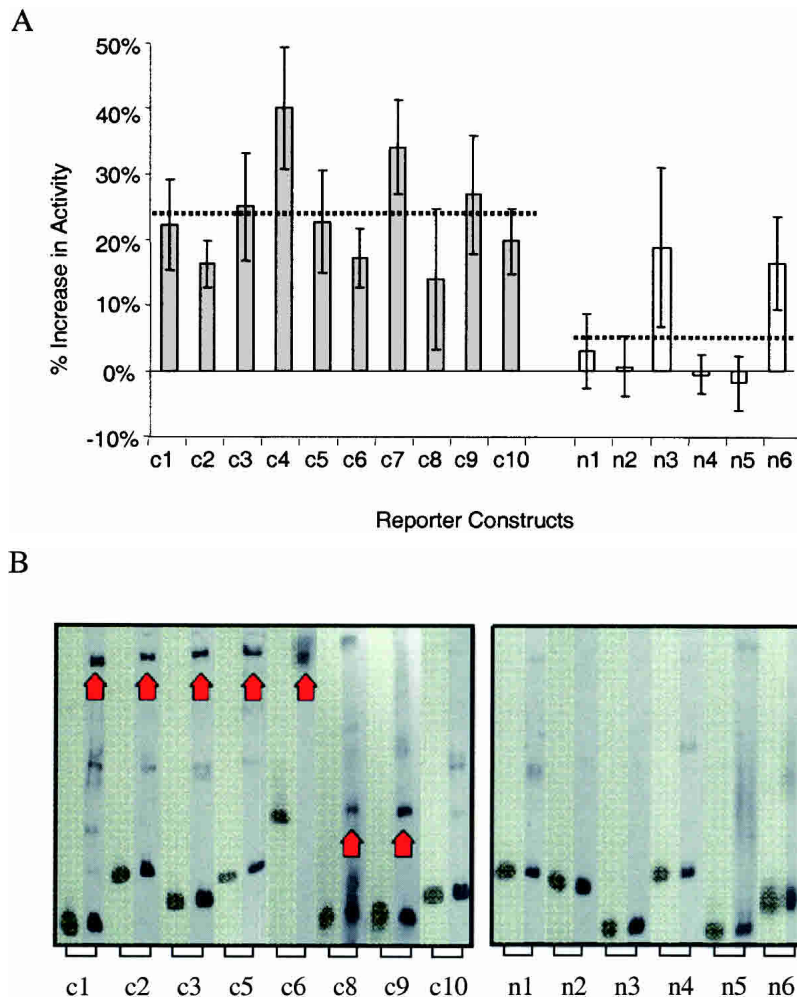
### Identification of Conserved Elements Using Human 21q High-Density Arrays

Human 21q high-density arrays were designed such that each unique base of the 365-kb region surrounding *SIM2* was interro-

gated by eight unique oligonucleotides (25 mers) as described previously (Frazer et al. 2001). For each mammal, DNA isolated from the set of minimally overlapping BACs was pooled, labeled with biotin-N6-ddATP, and hybridized to human 21q high-density arrays as described previously (Frazer et al. 2001). The average conformance of individual nucleotides in 30-nt length windows was calculated. Conforming nucleotides are those for which the probe complementary to the human reference sequence has greater fluorescent intensity than the corresponding noncomplementary probe. Sequences were classified as conserved on the basis of high conformance using an algorithm described previously (Frazer et al. 2001), modified here for increased sensitivity as described (see Supplemental method for description).

### Specificity of Detecting Conserved Elements Using High-Density Arrays

To estimate the false-positive rate of the modified algorithm, we analyzed a data set consisting of ~600 kb of human chromosome 21 sequence hybridized with random mouse BAC DNA, and identified 14 conserved elements, covering 501 nt. Hybridization of the same ~600 kb of chromosome 21 with orthologous mouse BAC DNA resulted in identification of 377 conserved elements covering a total of 23,375 nt. Thus, we estimate that ~3.7% of the elements and ~2.1% of the nucleotides identified as conserved by the modified algorithm are false positives.



**Figure 4** Functional characterization of noncoding sequences. (A) Expression values of 16 independent luciferase reporter constructs as assayed by transfection analysis; (filled bars) conserved noncoding sequences; (open bars) nonconserved noncoding sequences inserted in front of the *SIM2* promoter (see Fig. 3A). Luciferase expression values, normalized against  $\beta$ -galactosidase expression values, are averages of 12 individual experiments (each construct was analyzed in triplicate on four different days) and are expressed as the percent increase in activity over the control (the luciferase reporter construct containing only the *SIM2* promoter). The means across conserved and nonconserved sequences (dotted lines) were significantly different ( $P \sim 0.0047$ , two sample *t*-test). Error bars, one SD. (B) Electrophoretic mobility-shift patterns of noncoding DNA fragments. (Lane 1, highlighted) DNA fragment alone; (lane 2) DNA fragment incubated with nuclear extract. (Red arrows) Band-shift indicating DNA-protein binding. Conserved elements c4 and c7 were too long for gel-shift analysis.

### Analysis of Human-Mouse Sequence Alignments

Local alignments of human chromosome 21 sequences with orthologous mouse chromosome 16 sequences generated by the BLASTZ program (Schwartz et al. 2003) were downloaded from the Pennsylvania State University "Whole Genome Human/Mouse Homology" Web site at <http://bio.cse.psu.edu/genome/hummus/>; data set number 3, corresponding to the December 2001 Golden Path human assembly from the University of California at Santa Cruz aligned with the February 2002 Arachne mouse assembly from the Whitehead Institute. The December 2001 Golden Path human assembly was also aligned to the source contigs from which Perlegen high-density arrays were designed (Hattori et al. 2000) using the BLAT program (default parameters; Kent 2002). This resulted in 95 (120,267 bp) of the existing 108 (121,424bp) human-mouse BLASTZ alignments (in the range 34520115–34885293 on Chromosome 21, December 2001 Golden Path assembly) being mapped onto the source contig (NT\_002836; 23520000–23885000). Percent identities were

calculated for 100-base windows within each alignment, sliding 25 bases at a time. Windows with >80% identity were identified, and overlapping windows meeting this criterion were merged into single elements.

### Generation of Luciferase Reporter Constructs

The *SIM2* promoter region (Yamaki et al. 2001), 10 conserved and six nonconserved noncoding regions, were PCR amplified from human genomic DNA using PCR primers shown in Supplemental Table 9. The *SIM2* promoter luciferase reporter construct was generated by ligation of the *SIM2* promoter region into the BglIII and HindIII sites of the pGL3-basic vector (Promega). The 16 noncoding DNA fragments were ligated into KpnI- and XhoI-digested pGL3-*SIM2* promoter vector to produce the luciferase reporter constructs.

### Transfection Analysis of Reporter Constructs

The T98G human glioblastoma cells (ATCC) were transiently transfected with 3.2  $\mu$ g of reporter constructs and 0.8 mg of pSV- $\beta$ -galactosidase (Promega) control constructs using lipofectamine. Luciferase and  $\beta$ -galactosidase expression were assayed with the Bright-Glo luciferase assay system (Promega) and the Galactosidase enzyme assay system (Promega), respectively, after 48-h incubation at 37°C. Relative luciferase activity was obtained by normalizing the raw luminescence units by  $\beta$ -galactosidase activity.

### Electrophoretic Mobility Shift Assay

Each of the 16 noncoding DNA fragments used to generate the luciferase reporter constructs was end labeled with biotin-ddUTP, and incubated for 20 min at room temperature in the following reaction: 2  $\mu$ L glioblastoma cell (T98G) nuclear protein extract, 20–40 fm-labeled DNA fragments, 1 $\times$  binding buffer, and 1L poly dI-dC polymers. The DNA was detected using the Light-shift Biotin detection kit (Pierce Biotechnology) after transfer from 4% acrylamide gel to positively charged nylon membrane.

### ACKNOWLEDGMENTS

We thank E.J. Beilharz for discussions and assistance with the manuscript; I.M. Jen for data visualization; K. Pant for assistance with human/mouse sequence alignment; D.A. Hinds for assistance with statistical analysis; C.R. Kautzer for artwork; G.M. Vessere and T. Ren for technical assistance, and B. Zhu and Y. Wang for creating the horse and cat BAC libraries. This work was supported in part by grants from NHGRI and NIGMS to K.A.F. and from NHGRI and USDA-CSREES to P.J.deJ. and K.O.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

### REFERENCES

- Boffelli, D., McAuliffe, J., Ovcharenko, D., Lewis, K.D., Ovcharenko, L., Pachter, L., and Rubin, E.M. 2003. Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science* **299**: 1391–1394.

- Chureau, C., Prissette, M., Bourdet, A., Barbe, V., Cattolico, L., Jones, L., Eggen, A., Avner, P., and Duret, L. 2002. Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. *Genome Res.* **12**: 894–908.
- Dermitzakis, E.T., Reymond, A., Lyle, R., Scamuffa, N., Ucla, C., Deutsch, S., Stevenson, B.J., Flegel, V., Bucher, P., Jongeneel, C.V., et al. 2002. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. *Nature* **420**: 578–582.
- Emma, M., Ikegami, S., Hosoya, T., Mimura, J., Ohtani, H., Nakao, K., Inokuchi, K., Katsuki, M., and Fujii-Kuriyama, Y. 1999. Mild impairment of learning and memory in mice overexpressing the mSim2 gene located on chromosome 16: An animal model of Down's syndrome. *Hum. Mol. Genet.* **8**: 1409–1415.
- Fahrenkrug, S.C., Rohrer, G.A., Freking, B.A., Smith, T.P., Osoegawa, K., Shu, C.L., Catanese, J.J., and de Jong, P.J. 2001. A porcine BAC library with tenfold genome coverage: A resource for physical and genetic map integration. *Mamm. Genome* **12**: 472–474.
- Frazer, K.A., Sheehan, J.B., Stokowski, R.P., Chen, X., Hosseini, R., Cheng, J.F., Fodor, S.P., Cox, D.R., and Patil, N. 2001. Evolutionarily conserved sequences on human chromosome 21. *Genome Res.* **11**: 1651–1659.
- Frazer, K.A., Chen, X., Hinds, D.A., Pant, P.V., Patil, N., and Cox, D.R. 2003. Genomic DNA insertions and deletions occur frequently between humans and nonhuman primates. *Genome Res.* **13**: 341–346.
- Gill, G. 2001. Regulation of the initiation of eukaryotic transcription. *Essays Biochem.* **37**: 33–43.
- Hare, M.P. and Palumbi, S.R. 2003. High intron sequence conservation across three mammalian orders suggests functional constraints. *Mol. Biol. Evol.* **20**: 969–978.
- Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T., Park, H.S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D.K., et al. 2000. The DNA sequence of human chromosome 21. *Nature* **405**: 311–319.
- Kent, W.J. 2002. BLAT—The BLAST-like alignment tool. *Genome Res.* **12**: 656–664.
- Loots, G.G., Locksley, R.M., Blankespoor, C.M., Wang, Z.E., Miller, W., Rubin, E.M., and Frazer, K.A. 2000. Identification of a coordinate regulator of interleukins 4, 13 and 5 by cross-species sequence comparisons. *Science* **288**: 136–140.
- Marra, M.A., Kucaba, T.A., Dietrich, N.L., Green, E.D., Brownstein, B., Wilson, R.K., McDonald, K.M., Hillier, L.W., McPherson, J.D., and Waterston, R.H. 1997. High throughput fingerprint analysis of large-insert clones. *Genome Res.* **7**: 1072–1084.
- Mural, R.J., Adams, M.D., Myers, E.W., Smith, H.O., Miklos, G.L., Wides, R., Halpern, A., Li, P.W., Sutton, G.G., Nadeau, J., et al. 2002. A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. *Science* **296**: 1661–1671.
- Nambu, J.R., Lewis, J.O., Wharton Jr., K.A., and Crews, S.T. 1991. The *Drosophila* single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. *Cell* **67**: 1157–1167.
- Osoegawa, K., Tateno, M., Woon, P.Y., Frengen, E., Mammoser, A.G., Catanese, J.J., Hayashizaki, Y., and de Jong, P.J. 2000. Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res.* **10**: 116–128.
- Schwartz, S., Kent, W.J., Smit, A., Zhang, Z., Baertsch, R., Hardison, R.C., Haussler, D., and Miller, W. 2003. Human–mouse alignments with BLASTZ. *Genome Res.* **13**: 103–107.
- Thomas, J.W., Prasad, A.B., Summers, T.J., Lee-Lin, S.Q., Maduro, V.V., Idol, J.R., Ryan, J.F., Thomas, P.J., McDowell, J.C., and Green, E.D. 2002. Parallel construction of orthologous sequence-ready clone contig maps in multiple species. *Genome Res.* **12**: 1277–1285.
- Thomas, J.W., Touchman, J.W., Blakesley, R.W., Bouffard, G.G., Beckstrom-Sternberg, S.M., Margulies, E.H., Blanchette, M., Siepel, A.C., Thomas, P.J., McDowell, J.C., et al. 2003. Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* **424**: 788–793.
- Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520–562.
- Yamaki, A., Tochigi, J., Kudoh, J., Minoshima, S., Shimizu, N., and Shimizu, Y. 2001. Molecular mechanisms of human single-minded 2 (SIM2) gene expression: Identification of a promoter site in the SIM2 genomic sequence. *Gene* **270**: 265–275.

## WEB SITE REFERENCES

- <http://bio.cse.psu.edu/genome/hummus/>; Whole Genome Human/Mouse Homology Web site.
- <http://bacpac.chori.org/>; BACPAC Resources Center Home Page (Children's Hospital Oakland Research Institute).

Received September 10, 2003; accepted in revised form December 22, 2003.