



## Evolutionarily Conserved Noncoding DNA in the Human Genome: How Much and What For?

Miriam H. Meisler

*Genome Res.* 2001 11: 1617-1618

Access the most recent version at doi:[10.1101/gr.211401](https://doi.org/10.1101/gr.211401)

---

**References** This article cites 15 articles, 11 of which can be accessed free at:  
<http://genome.cshlp.org/content/11/10/1617.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

# Evolutionarily Conserved Noncoding DNA in the Human Genome: How Much and What For?

Miriam H. Meisler

Department of Human Genetics, University of Michigan, Ann Arbor, Michigan 48109-0618, USA

**T**he noncoding component of the human genome is receiving increased attention from biologists because of its predicted role in regulation of transcription, DNA replication, chromosome pairing, and chromosome condensation. Finding the functional elements within this 97% of the human genome presents major intellectual and experimental challenges. By comparing genomic DNA sequence from diverse species, functional elements may be recognized on the basis of their evolutionary conservation. In this issue, Frazer et al. (2001) describe the large-scale identification of conserved noncoding elements from human chromosome 21 using oligonucleotide array technology. In a two-way comparison between mouse and human, they found that the amount of conserved noncoding sequence was roughly equal to the coding sequence in this region. One-half of the human/mouse conserved noncoding sequence was also conserved in a third mammal, the dog. This work produced a catalog of potential functional elements for chromosome 21 that will be valuable to future studies of gene regulation and chromosome mechanics. In addition, a method for identification of conserved sequences prior to genome sequencing is shown.

## Box 1. Proliferating Acronyms

CNS	Conserved Noncoding Sequence
CSB	Conserved Sequence Block
ECNS	Evolutionarily Conserved Noncoding Sequence
NCS	Noncoding Conserved Sequences
NIE	conserved sequences Not in Identified Exons

## Background

Evolutionary comparisons have been performed since the earliest days of cloning and sequencing of mammalian DNA (Table 1). One of the first examples was the identification of short, highly conserved noncoding regions in the cloned human and mouse im-

**E-MAIL** [meislerm@umich.edu](mailto:meislerm@umich.edu); **FAX (734) 763-9691**.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.211401>.

munoglobulin genes by detection of heteroduplex molecules in the electron microscope (Ravetch et al. 1980). During the 1980s the sequences of several human and mouse promoters were compared, to identify determinants of tissue specificity such as the pancreas-specificity of the amylase gene (Gumucio et al. 1988). Extensive multispecies sequence comparison of the primate globin cluster contributed to the identification of conserved enhancers that influence transcription from a distance (Gumucio et al. 1996). The first large-scale study compared 100 kb of human and mouse DNA containing the T-cell receptor gene family (Hood et al. 1995). The noncoding regions of this gene cluster proved to have an unusually high level of sequence conservation. In a more typical 100-kb segment from chromosome 2p13, 1% of the sequence was accounted for by conserved elements of length > 80 bp with sequence identity > 75% (Jang et al. 1999).

In an important demonstration of the function of a conserved noncoding segment, Loots et al. (2000) carried out multi-species sequence comparison of a 1 Mb region containing an interleukin gene cluster. Deletion of a conserved noncoding element of 401 bp was shown to change interleukin expression in T cells of transgenic mice. The chromosome 21 analysis by Frazer et al. in this issue provides the most extensive human/mouse comparison available to date (Table 1).

A genome-wide alignment of human and mouse sequence is becoming available from the public genome project. Whole genome shotgun sequence data comprising 2.5–3X coverage of the mouse genome has been aligned with the assembled human draft sequence; the alignment can be viewed at <http://genome.cse.ucsc.edu> and <http://www.ensembl.org>. As more finished mouse sequence is added, this resource will identify candidate regulatory elements for a large proportion of the human genome.

## Large-Scale Detection of Conserved Sequences in Nonsequenced Genomes Using Oligonucleotide Arrays

The resource used by Frazer et al. for their cross-species comparison was an oligonucleotide array containing four 25mers for each

nucleotide in 16.6 Mb of nonrepetitive DNA from human chromosome 21 (Frazer et al. 2001). BAC contigs of the orthologous regions of the mouse and dog genomes were constructed and DNA fragments from the contigs were hybridized to the arrays. The sequence of conserved elements from the mouse and dog DNA could be determined from the hybridization pattern, as in “re-sequencing” of human DNA (Hacia et al. 1999). In the human–mouse comparison, 3400 conserved elements ranging in length from 30 bp to > 1 kb were identified, corresponding to 1.6% of the tested sequence. Only 44% of the conserved elements corresponded to exons of identified genes in the region, indicating that the amount of conserved noncoding DNA is approximately equal to the amount of exonic DNA in this region of chromosome 21. Interestingly, one-half of the conserved elements were located in intergenic regions more than 10 kb distant from the nearest known gene. The complete catalog of conserved noncoding sequences on Chromosome 21 is provided for followup analysis.

In comparisons between two mammalian species, it is difficult to estimate the proportion of conserved sequence resulting solely from common origin in the absence of active selection for function. To address this issue, 2.6 Mb of orthologous dog DNA was hybridized with the oligo arrays. Only one-half of the human–mouse noncoding elements were conserved in the dog sequence. This important result indicates that it will be worthwhile to extend comparisons beyond two species before initiating functional tests of putative regulatory elements.

## Conservation of Transcriptional Regulatory Elements in Genomes of More Distant Species

Comparisons of human DNA with pre-mammalian vertebrate genomes have a lower false-positive rate than mammalian comparisons because of the greater time available for accumulation of neutral mutations in non-functional sequences. Many biological and developmental processes are conserved among vertebrates. Conservation of function has been experimentally demonstrated for a

**Table 1. Examples of Comparative Analysis of Human Genomic Sequence**

Year	Length	Species	Gene	Reference
1980	30 kb	H/M	Immunoglobulin heavy chain	Ravetch et al. 1980
1992	50 kb	H/primate	Globin genes	Gumucio et al. 1996
1994	100 kb	H/M	T cell receptor complex	Hood et al. 1995
1999	100 kb	H/M	Human 2p13	Jang et al. 1999
2000	200 kb	H/M/D	Human 5q31	Dubchak et al. 2000
2000	1 Mb	H/M	Interleukin gene cluster on 5q31	Loots et al. 2000
2001	16 Mb	H/M	Human Chr 21	Frazer et al. 2001
2001	2 Mb	H/M/D	Human Chr 21	Frazer et al. 2001

H, human; M, mouse; D, dog.

small number of transcriptional regulators from fish and fly that contain conserved non-coding sequences (Table 2). The lack of long-range linkage conservation in fish may limit this approach to small-scale, gene-by-gene comparisons. If the large-scale array approach of Frazer et al. (2001) could be extended to chicken and other amniotes, the resulting conserved elements would be of great interest.

### Assignment of Function to Conserved Noncoding Sequences

Subsequent to the identification of putative regulatory elements by sequence comparison, the confirmation of biological function will depend upon experimental assays. Expression in transgenic mice can provide definitive tests of function for putative transcription

regulatory elements (e.g., Table 2). Unfortunately, the cost and effort involved will limit this approach to a relatively small number of genes. Expression in differentiated cell lines can increase the number of constructs that can be studied for individual genes. Large-scale analysis of expression patterns of mammalian genes will define sets of co-regulated genes that may share conserved noncoding sequence elements, as observed in recent studies in yeast (Gasch et al. 2000). Large-scale sequence comparisons of co-regulated genes can be combined with experimental analysis of protein binding sites to define targets for transcription factor interaction, as shown for muscle-specific factors (Wasserman and Fickett 1999). There are likely to be > 1000 transcription factors in the human genome. Analysis of conserved noncoding se-

quences promises to be an important tool in elucidating the regulatory circuits involved in the immensely complex process of human development and differentiation.

### REFERENCES

- Aparicio, S., Morrison, A., Gould, A., Gilthorpe, J., Chaudhuri, C., Rigby, P., Krumlau, R., and Brenner, S. 1995. *Proc. Natl. Acad. Sci.* **92**: 1684–1688.
- Dubchak, I., Brudno, M., Loots, G.G., Pachter, L., Mayor, C., Rubin, E.M., and Frazer, K.A. 2000. *Genome Res.* **10**: 1304–1306.
- Frazer, K.A., Sheehan, J.B., Stokowski, R.P., Chen, X., Hosseini, R., Cheng, J-F., Fodor, S.P.A., Cox, D.R., and Patil, N. 2001. *Genome Res* **11**: 1651–1659.
- Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. 2000. *Mol. Biol. Cell* **11**: 4241–4257.
- Gumucio, D.L., Shelton, D.A., Zhu, W., Millinoff, D., Gray, T., Bock, J.H., Slightom, J.L., and Goodman, M. 1996. *Mol. Phylogenet. Evol.* **5**: 18–32.
- Gumucio, D.L., Wiebauer, K., Caldwell, R.M., Samuelson, L.C., and Meisler, M.H. 1988. *Mol. Cell. Biol.* **8**: 1197–1205.
- Hacia, J.G., Fan, J.B., Ryder, O., Jin, L., Edgemon, K., Ghandour, G., Mayer, R.A., Sun, B., Hsie, L., Robbins, C.M., et al. 1999. *Nat. Genet.* **22**: 164–167.
- Hood, L., Rowen, L., and Koop, B.F. 1995. *Ann. N.Y. Acad. Sci.* **758**: 390–412.
- Jang, W., Hua, A., Spilson, S.V., Miller, W., Roe, B.A., and Meisler, M.H. 1999. *Genome Res.* **9**: 53–61.
- Loots, G.G., Locksley, R.M., Blankespoor, C.M., Wang, Z.E., Miller, W., Rubin, E.M., and Frazer, K.A. 2000. *Science* **288**: 136–140.
- Müller, F., Chang, B-E., Albert, S., Fischer, N., Tora, L., and Strähle, U. 1999. *Development* **126**: 2103–2116.
- Ravetch, J.V., Kirsch, I.R., and Leder, P. 1980. *Proc. Natl. Acad. Sci.* **77**: 6734–6738.
- Venkatesh, B., Si-Hoe, S.L., Murphy, D., and Brenner, S. 1997. *Proc. Natl. Acad. Sci.* **94**: 12462–12466.
- Wasserman, W.W. and Fickett, J.W. 1998. *J. Mol. Biol.* **278**: 167–181.
- Xu, P.X., Zhang, X., Heaney, S., Yoon, A., Michelson, A.M., and Maas, R.L. 1999. *Development* **126**: 383–395.

**Table 2. Transcriptional Regulatory Elements from Fish and Fly with Conserved Function in Mammals**

Genome	Gene	Functional test	Reference
<i>Fugu</i>	<i>Hoxb4</i>	Rhombomere expression in transgenic mice	Aparicio et al. 1995
<i>Fugu</i>	<i>Oxytocin</i>	Neuron-specific expression in transgenic rat	Venkatesh et al. 1997
<i>Drosophila</i>	<i>Pax6</i>	Eye-specific expression in transgenic mice	Xu et al. 1999
<i>Danio</i>	<i>Shh</i>	Expression in notochord of transgenic mice	Müller et al. 1999