

Supplementary Methods

BAC tiling path clones not derived from the fingerprint map were chosen for sequencing based on positive RPCI-11 and RPCI-13 library screening results using overgo¹ probes from STS markers, unique BAC-end sequence or gene sequences. Candidate clones were end-sequenced prior to inclusion in the tiling path. BAC clone DNA was sheared with nebulizers and used to construct random shotgun libraries, with 2-4 kb inserts, in a modified pUC18 vector using the double adaptor method². Random subclones were end-sequenced using ABI dye-terminator chemistry and were base-called, assembled and visualized using the Phred-Phrap-Consed package (<http://www.phrap.org/phredphrapconsed.html>). BAC clones were finished using custom primers, multiple insert-length shatter libraries, transposon insertion and sequence derived from PCR products. We assessed BAC clone order, integrity and coverage using available genetic and radiation hybrid map markers including the deCODE³, Genethon⁴, Marshfield⁵, GeneMap99⁶ and Whitehead Institute YAC maps⁷. Markers were placed on the genomic sequence using a combination of ePCR (<http://www.ncbi.nlm.nih.gov/sutils/e-pcr/>)⁸ and BLAST (see Supplementary Figure 1).

We used a combination of three approaches to assign ncRNA loci to the finished chromosome 3 sequence. First, we constructed a manually curated database of experimentally confirmed or expressed and bioinformatically confirmed ncRNA sequences including those from FANTOM^{9,10}, FLJ/H-InV-FL-cDNA^{11,12} and others totaling 17,372 entries. We aligned these sequences to the NCBI Build 34 chromosome 3 sequence using BLAST and SIM4¹³. Sequences showing alignment at > 95% identity over 95% of their length were retained as candidates. Second, we used *in silico* prediction

software including Rfam¹⁴, snoscan¹⁵, SRPscan¹⁶, tRNAscan-SE¹⁷, and FISHER¹⁸ to identify further candidates. Finally, we mined the human Unigene (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>) and EST (<http://www.ncbi.nlm.nih.gov/dbEST/index.html>) databases for non-protein-coding transcripts that appeared in a least four EST libraries and aligned these to the chromosome using SIM4. Those candidates aligning with >85% identity over >80% of their length, and whose locus is devoid of GENSCAN (<http://genes.mit.edu/GENSCAN.html>) protein predictions within 10 kb were retained.

Comparative analysis used alignment of sequences from the chimpanzee¹⁹, dog (<http://www.ncbi.nlm.nih.gov/genome/guide/dog/>), mouse²⁰, rat²¹, chicken²², zebrafish (<http://www.ncbi.nlm.nih.gov/projects/genome/guide/zebrafish/>) and Fugu²³.

REFERENCES

1. Ross, M. T., LaBrie, S., McPherson, J. & Stanton, V. P. (eds.) *Screening Large-Insert Libraries by Hybridization* (John Wiley and Sons, New York, 1999).
2. Andersson, B., Wentland, M. A., Ricafrente, J. Y., Liu, W. & Gibbs, R. A. A "double adaptor" method for improved shotgun library construction. *Anal Biochem* **236**, 107-13 (1996).
3. Kong, A. et al. A high-resolution recombination map of the human genome. *Nat Genet* **31**, 241-7 (2002).
4. Dib, C. et al. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**, 152-4 (1996).
5. Broman, K. W., Murray, J. C., Sheffield, V. C., White, R. L. & Weber, J. L. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am J Hum Genet* **63**, 861-9 (1998).
6. Deloukas, P. et al. A physical map of 30,000 human genes. *Science* **282**, 744-6 (1998).
7. Hudson, T. J. et al. An STS-based map of the human genome. *Science* **270**, 1945-54 (1995).
8. Schuler, G. D. Sequence mapping by electronic PCR. *Genome Res* **7**, 541-50 (1997).

9. Okazaki, Y. et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**, 563-73 (2002).
10. Numata, K. et al. Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection. *Genome Res* **13**, 1301-6 (2003).
11. Ota, T. et al. Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat Genet* **36**, 40-5 (2004).
12. Imanishi, T. et al. Integrative annotation of 21,037 human genes validated by full-length cDNA clones. *PLoS Biol* **2**, e162 (2004).
13. Florea, L., Hartzell, G., Zhang, Z., Rubin, G. M. & Miller, W. A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res* **8**, 967-74 (1998).
14. Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A. & Eddy, S. R. Rfam: an RNA family database. *Nucleic Acids Res* **31**, 439-41 (2003).
15. Lowe, T. M. & Eddy, S. R. A computational screen for methylation guide snoRNAs in yeast. *Science* **283**, 1168-71 (1999).
16. Regalia, M., Rosenblad, M. A. & Samuelsson, T. Prediction of signal recognition particle RNA genes. *Nucleic Acids Res* **30**, 3368-77 (2002).
17. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**, 955-64 (1997).
18. Edvardsson, S. et al. A search for H/ACA snoRNAs in yeast using MFE secondary structure prediction. *Bioinformatics* **19**, 865-73 (2003).
19. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69-87 (2005).
20. Waterston, R. H. et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-62 (2002).
21. Gibbs, R. A. et al. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493-521 (2004).
22. Hillier, L. W. et al. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695-716 (2004).
23. Ovcharenko, I., Stubbs, L. & Loots, G. G. Interpreting mammalian evolution using Fugu genome comparisons. *Genomics* **84**, 890-5 (2004).