Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 1366–1369, February 1996
Genetics

Conservation of synteny between the genome of the pufferfish (*Fugu rubripes*) and the region on human chromosome 14 (14q24.3) associated with familial Alzheimer disease (*AD3* locus)

MICHAEL K. TROWER*†, SUZANNE M. ORTON*, IAN J. PURVIS*, PHILIPPE SANSEAU*, JOHN RILEY*, CHRIS CHRISTODOULOU*, DOREEN BURT*, CHEE GEE SEE‡§, GREG ELGAR¶, ROBIN SHERRINGTON‖, EVGENY I. ROGAEV‖, PETER ST. GEORGE-HYSLOP‖, SYDNEY BRENNER¶, AND COLIN W. DYKES*

*Genomics Unit, Glaxo-Wellcome Medicines Research Centre, Stevenage, Hertfordshire, SG1 2NY, England; ‡The Galton Laboratory, Department of Genetics and Biometry, University College London, London NW1 2HE, England; ¶Molecular Genetics Group, Department of Medicine, University of Cambridge Clinical School, Addenbrooks Hospital, Cambridge CB2 2QQ, England; and ‖Centre for Research into Neurodegenerative Diseases, University of Toronto, Toronto, ON Canada M5S 1A8

Contributed by Sydney Brenner, November 1, 1995

ABSTRACT **The genome of the pufferfish (*Fugu rubripes*) (400 Mb) is ≈7.5 times smaller than the human genome, but it has a similar gene repertoire to that of man. If regions of the two genomes exhibited conservation of gene order (i.e., were syntenic), it should be possible to reduce dramatically the effort required for identification of candidate genes in human disease loci by sequencing syntenic regions of the compact *Fugu* genome. We have demonstrated that three genes (dihydrolipoamide succinyltransferase, S31iii125, and S20i15), which are linked to *FOS* in the familial Alzheimer disease locus (*AD3*) on human chromosome 14, have homologues in the *Fugu* genome adjacent to *Fugu cFOS*. The relative gene order of *cFOS*, S31iii125, and S20i15 was the same in both genomes, but in *Fugu* these three genes lay within a 12.4-kb region, compared to >600 kb in the human *AD3* locus. These results demonstrate the conservation of synteny between the genomes of *Fugu* and man and highlight the utility of this approach for sequence-based identification of genes in human disease loci.**

Identification of genes in genomic loci associated with human diseases has been greatly facilitated by the development of techniques such as "exon trapping" (1) and cDNA selection (2, 3). Direct sequencing of disease loci has also been shown to be one of the most effective methods of gene detection, but it requires significant sequencing capacity (4–6). The *Fugu rubripes* genome is 7- to 8-fold smaller than that of human (≈400 Mb compared to ≈3000 Mb), but it is believed to contain a similar complement of genes (7). Thus a typical cosmid clone of genomic DNA might be expected to contain seven to eight *Fugu* genes compared to only one human gene. Therefore, sequencing regions of the *Fugu* genome syntenic with human disease loci should accelerate the identification of candidate genes. We have used this approach to characterize a human chromosomal region (14q24.3) associated with an autosomal dominant, early onset form of Alzheimer disease (*AD3* locus) (8). The results of our investigations are described herein.

MATERIALS AND METHODS

Isolation of *Fugu cFOS* Cosmid. Full-length rat c-*fos* cDNA, amplified by PCR using standard conditions from plasmid pHKGfos with oligonucleotide primers 365F (5′-TGC CAA GCT TGA ATT CAT GAT GTT CTC GGG TTT CAA CG-3′) and 366R (5′-GAA TTC TTC TAG ATC TCT GTA ATG CAC CAG CTC-3′), was used to probe a *Fugu* genomic cosmid library (constructed in lawrist 4 by G.E.) gridded at high density on two 22 × 22 cm filter membranes (total 41,472 clones, equivalent to four *Fugu* haploid genomes). Cosmid DNA, prepared from 20 positively hybridizing clones, was digested with *Eco*RI and Southern blotted. Blots were probed with both full-length rat c-*fos* cDNA and a fragment derived from the 3′ end of the cDNA (amplified with primers 356F (5′-TGG CAG CCC ACC GAC CCG CCT GCA AGA T-3′) and 358R (5′-GTC AGC CTC GGG GTA GGT GAA GAC GAA-3′) that avoids the "leucine zipper" motif. A single *Fugu* cosmid (4D7) was identified that hybridized to both probes.

Cosmid Library Construction. 4D7 cosmid DNA (5 μg) was sheared by sonication, the ends were repaired with T4 DNA polymerase, and the blunt-ended fragments generated were randomly ligated into a phagemid vector, pCRscript (Stratagene Cloning Systems). The ligation mix was transformed into competent cells of *Escherichia coli* XL1-Blue (Stratagene Cloning Systems) using standard procedures. Transformations were plated out on selective medium containing ampicillin and tetracycline, in the presence of 5-bromo-4-chloro-3-indolyl β-D-galactoside and isopropyl β-D-thiogalactoside allowing isolation of recombinant clones. These were picked into a convenient 96-well microtiter plate format for both storage and further processing.

DNA Sequencing, Analysis, and Gene Assembly. The majority of the arrayed subclones isolated were amplified by PCR and then sequenced to a 5-fold redundancy (≈650 sequences), by nonradioactive fluorescent automated cycle sequencing as described by Trower *et al.* (9). DNA sequences were determined after electrophoresis of reactions on an ABI 373A automated DNA sequencer. Sequence homology searches were carried out against the GenBank (Release 88) and Swiss-Prot (Release 31) data bases and a data set composed of transcripts identified in the *AD3* locus (provided by R.H., E.I.R., and P.St.G.-H.), by BLASTX (10), BLASTN (10), and FASTA (11). The detected genes were assembled into contiguous sequences using AUTOASSEMBLER software (Applied Biosystems), and when necessary further sequencing was undertaken, either from the opposite ends of previously sequenced clones or by using custom primers in association with fluorescent dye terminator cycle sequencing for gap closure and for completion of both DNA strands.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: DLST, dihydrolipoamide succinyltransferase; FISH, fluorescence *in situ* hybridization.

†To whom reprint requests should be addressed.
§Present address: Genomics Unit, Glaxo-Wellcome Medicines Research Centre, Stevenage, Hertfordshire, SG1 2NY, England.

**The sequences reported in this paper have been deposited in the GenBank data base (accession nos. U40755–U40761).

Isolation of Human Cosmids. Human S20i15 (p126H2), S31iii125 (I:66-4), and *FOS* (A6-7) cosmids were isolated by hybridization from a flow-sorted human chromosome 14-specific cosmid library using cDNA probes for these genes (12).

"FiberFISH" Analyses. Fluorescence *in situ* hybridization (FISH) to extended DNA fibers was undertaken with the human S20i15, S31iii125, and *FOS* cosmids using the protocol described by Heiskanen *et al.* (13). Cosmid DNAs were nick-translated and labeled for *FOS* with both biotin and digoxigenin, for S31iii125 with biotin, and for S20i15 with digoxigenin. Biotinylated probes were detected by avidin conjugated to fluorescein isothiocyanate, while digoxigenin-labeled probes were detected by mouse anti-digoxgenin in combination with rabbit anti-mouse antibody conjugated to tetramethylrhodamine isothiocyanate.

RESULTS AND DISCUSSION

A conserved gene encoding the transcription factor *FOS*, known to lie in the *AD3* locus (8), was used to screen a *Fugu* genomic cosmid library. After selection of a clone (4D7) that hybridized strongly to the c-*Fos* probe (data not shown), 4D7 cosmid DNA was isolated and sequenced. Eight putative genes, including *cFOS* (Fig. 1), were identified and characterized by DNA sequence analysis (Table 1). Fig. 2 details a map of the *Fugu* cosmid showing the position and order of the putative genes identified. Three of the isolated putative *Fugu* genes, dihydroli-

```
                                                            Intron A
          1                                                    <>           70
rtcfos    MMFSGFNADY EASSSRCSSA SPAGDSLSYY HSPADSFSSM GSPVNTQDFC ADLSVSSANF IPTVTAISTS
mocfos    MMFSGFNADY EASSSRCSSA SPAGDSLSYY HSPADSFSSM GSPVNTQDFC ADLSVSSANF IPTVTAISTS
chcfos    MMYQGFAGEY EAPSSRCSSA SPAGDSLTYY PSPADSFSSM GSPVNSQDFC TDLAVSSANF VPTVTAISTS
hucfos    MMFSGFNADY EASSSRCSSA SPAGDSLSYY HSPADSFSSM GSPVNAQDFC TDLAVSSANF IPTVTAISTS
fucfos    MMFTSFNAEC D-SSSRC-SA SPVGDNL-YY PSPAGSYSSM GSP-QSQDF- TDLTASSASF IPTVTAISTS
Con       MM   F        SSRC SA SP GD L YY  SPA S SSM GSP   QDF   DL  SSA F  PTVTAISTS

                                                                         Intron B
          71                                                                <>   140
rtcfos    PDLQWLVQPT LVSSVAPSQT RAPHPYGLPT P-STGAYARA GVVKTMSGGR AQSIGRRGKV EQLSPEEEEK
mocfos    PDLQWLVQPT LVSSVAPSQT RAPHPYGLPT Q-SAGAYARA GMVKTVSGGR AQSIGRRGKV EQLSPEEEEK
chcfos    PDLQWLVQPT LISSVAPSQN RG-HPYGVPA PAPPAAYSRP AVLKA-PGGR GQSIGRRGKV EQLSPEEEEK
hucfos    PDLQWLVQPA LVSSVAPSQT RAPHPFGVPA P-SAGAYSRA GVVKTMTGGR AQSIGRRGKV EQLSPEEEEK
fucfos    PDLQWMVQP- LISSVAPSHR --AHPYS--- --PSPSYKRT VMRSAASKAH ----GKRSRV EQTTPEEEEK
Con       PDLQW VQP  L SSVAPS      HP             Y R                G R  V EQ  PEEEEK

                                      Intron C
          141                            <>                                      210
rtcfos    RRIRRERNKM AAAKCRNRRR ELTDTLQAET DQLEDEKSAL QTEIANLLKE KEKLEFILAA HRPACKIPND
mocfos    RRIRRERNKM AAAKCRNRRR ELTDTLQAET DQLEDEKSAL QTEIANLLKE KEKLEFILAA HRPACKIPDD
chcfos    RRIRRERNKM AAAKCRNRRR ELTDTLQAET DQLEEEKSAL QAEIANLLKE KEKLEFILAA HRPACKMPEE
hucfos    RRIRRERNKM AAAKCRNRRR ELTDTLQAET DQLEDEKSAL QTEIANLLKE KEKLEFILAA HRPACKIPDD
fucfos    KRIRRERNKQ AAAKCRNRRR ELTDTLQAET DQLEDEKSSL QNDIANLLKE KERLEFILAA HQPICKIPSQ
Con        RIRRERNK  AAAKCRNRRR ELTDTLQAET DQLE EKS L Q  IANLLKE KE LEFILAA H P CK P
                                        *          *         *          *          *

          211                                                                    280
rtcfos    LG-------- --FPEEMSVT -SLDLTGGLP EATT-PESEE AFTLPLLNDP EPK-PSLEPV KNISNMELKA
mocfos    LG-------- --FPEEMSVA -SLDLTGGLP EAST-PESEE AFTLPLLNDP EPK-PSLEPV KSISNVELKA
chcfos    LR-------- --FSEELAAA TALDLGAPSP AAA-----EE AFALPLMTEA PPAVPPKEP- -SGSGLELKA
hucfos    LG-------- --FPEEMSVA -SLDLTGGLP EVAT-PESEE AFTLPLLNDP EPK-PSVEPV KSISSMELKT
fucfos    MDTDFSVVSM SPVHACLSTT VSTQLQTSIP EATTVTSSHS TFTSTSNSIF SGSSDSLLST ATVSNSVVKM
Con                                   L    P             F                      S    K

          281                                                                    350
rtcfos    EPFDDFLFPA SSR---PSGS ETARSVPDVD LSG--SFYAA -DWEPLHSSS LGMGPMVTEL EPLCTPVVTC
mocfos    EPFDDFLFPA SSR---PSGS ETSRSVPDVD LSG--SFYAA -DWEPLHSNS LGMGPMVTEL EPLCTPVVTC
chcfos    EPFDELLFSA GPR------- EASRSVPDMD LPGASSFYAS -DWEPLGAGS GG------EL EPLCTPVVTC
hucfos    EPFDDFLFPA SSR---PSGS ETARSVPDMD LSG--SFYAA -DWEPLHSGS LGMGPMATEL EPLCTPVVTC
fucfos    TDLDSSVLEE SLDLLAKTEA ETARSVPDVN LSN--SLFAA QDWEPLHATI SS-----SDF EPLCTPVVTC
Con          D                  E RSVPD   L    S  A   DWEPL                EPLCTPVVTC

          351                                          400
rtcfos    TPSCTTYTSS FVFTYPEADS FPSCAAAHRK GSSSNEPSSD SLSSPTLLAL    Sim, 77%; Id, 64%
mocfos    TPGCTTYTSS FVFTYPEADS FPSCAAAHRK GSSSNEPSSD SLSSPTLLAL    Sim, 76%; Id, 63%
chcfos    TPCPSTYTST FVFTYPEADA FPSCAAAHRK GSSSNEPSSD SLSSPTLLAL    Sim, 75%; Id, 60%
hucfos    TPSCTAYTSS FVFTYPEADS FPSCAAAHRK GSSSNEPSSD SLSSPTLLAL    Sim, 76%; Id, 63%
fucfos    TPACTTLTSS FVFTFPEAET FPTCGVAHRR RSNSNDQSSD SLSSPTLLAL
Con       TP     TS  FVFT PEA   FP C  AHR   S SN  SSD SLSSPTLLAL
```

FIG. 1. Alignment of the determined deduced protein sequence for *Fugu cFOS* (*fucfos*), with those reported for: rat c-*fos* (rtcfos), mouse c-*Fos* (mocfos), chicken c-*fos* (chcfos), and human *FOS* (hucfos). Protein alignments were undertaken with PILEUP from Version 7.3 of the Wisconsin GCG package (14). Percentage similarities (Sim) and identities (Id) were determined by pairwise alignment of each *cFOS* sequence against the *Fugu cFOS* primary structure using GAP, also from the GCG package. The consensus sequence (Con) shows the positions of conserved residues for all five deduced amino acid sequences; asterisks (*) mark the locations of the five conserved leucine residues, which participate in the formation of the "leucine zipper" structure; hyphens (-) indicate locations where gaps have been inserted into the sequences to maximize homology. The frame and phase of the putative splice junctions for introns A–C (indicated on the figure as < >) in the *Fugu cFOS* genomic DNA sequence were identical to those determined for the human *FOS* genomic sequence. Intron sizes were intron A: human, 753 bp; *Fugu*, 397 bp; intron B: 431 bp, 104 bp; intron C: 112 bp, 103 bp, respectively. The GenBank accession numbers for the *cFOS* sequences described above are X06769 (rat), V00727 (mouse), M37000 (chicken), P01100 (human), and U40757 (*Fugu*).

Table 1. Characterization of putative *Fugu* genes identified on cosmid 4D7

Fugu gene	Peptide length	Homologue	Peptide length	Organism	Accession number	% similarity	% identity	Match length, aa
prospero-like protein	128	*prospero*?	1407	*Drosophila*	M81389	73	60	137
DLST	409	DLST	453	Human	L37418	83	76	422
ATF-like protein	68	ATF3?	222	Human	L19871	82	59	65
Fos-like protein	276	*FOS*	380	Human	P01100	62	49	311
cFOS	376	*FOS*	380	Human	P01100	76	63	399
S31iii125	213	S31iii125	134	Human	L40397	70	55	124
S20i15	355	S20i15	346	Human	L40395	75	64	354
7SL RNA	298	7SL RNA	299	Human	X04249		84	299

The putative *Fugu* coding sequences (following assembly without introns) were run with BLASTX (10) against nonredundant Protein Data Base, Swiss-Prot 31, Spupdate, PIR 44, Genpept 88, and Gpupdate data bases (total of 148, 485 sequences) to identify gene homologues. Exons 1 and 2 of *Fugu* DLST were not detected by these sequence comparisons. Only a partial sequence was obtained for the putative *prospero*-like protein, since the cloned genomic region containing this gene only showed homology to the 3′ end of the *Drosophila prospero* sequence and therefore probably represents a family member. The genes for human S31iii125 (L40397) and S20i15 (L40395) are partial cDNA sequences (12). Lengths of translated human and putative *Fugu* coding sequences are given as numbers of amino acids except for the 7SL RNA genes, which are in numbers of nucleotides. Percent similarities and identities were determined by pairwise alignment of each *Fugu* sequence against the primary structure of its proposed homologue using BESTFIT from Version 7.3 of the Wisconsin GCG package (14). The comparative sequence data suggests that the fos-like protein gene may represent a duplication event in the *Fugu* genome of the *cFOS* gene. Functions for human S31iii125 and S20i15 are postulated, based upon amino acid sequence homologies to canine glycoprotein gp25L (19) and the yeast GCD7 subunit of a GDP–GTP exchange factor (20), respectively. The chromosomal location of DLST and *FOS* have previously been reported to 14q24.2-q24.3 (16) and 14q21-q31 (22), respectively, and were also confirmed to be in the *AD3* locus on 14q24.3, together with S31iii125 and S20i15 by Sherrington *et al.* (12).

poamide succinyltransferase (DLST), S31iii125, and S20i15, were found to be homologues of human transcripts (Table 1) that map to the same 850-kb yeast artificial chromosome (797D11) as *FOS*, in "region A" of the *AD3* locus (12).

To allow a comparison of both gene order and physical distance covered by the linked genes in both genomes, human cosmids for S31iii125, S20i15, and *FOS* were used as probes for FISH analyses. Initially, this was undertaken with human metaphase chromosome spreads, which confirmed that these genes all mapped to human chromosome 14q24.3 (data not shown). High-resolution mapping was achieved by probing the cosmids onto stretched fibers of human DNA using the FiberFISH technique (Fig. 3). The results clearly demonstrated that the order of the three genes was the same in both human and *Fugu* genomes. However, the physical distance covered by these genes was dramatically smaller in the *Fugu* genome, comprising only 12.4 kb, compared to a distance of >600 kb in the human genome. This compression of the *Fugu* genome is even greater than that expected and indicates that local differences in gene density may be superimposed upon the 7- to 8-fold difference predicted from the relative sizes of the two genomes (assuming equal numbers of genes in *Fugu* and man).

The data reported here are consistent with the observation, reported by Brenner *et al.* (7), that intergenic distances are much smaller in the *Fugu* genome compared to human or other mammalian genomes. Intragenic DNA is also reduced in pufferfish genes (24, 25), including the *Fugu* homologue of the Huntington disease gene (26). This was also observed in the genes isolated from the 4D7 cosmid. For example, the 12 introns (C–N) of *Fugu* DLST total 952 bp (mean of 79 bp), compared to 15,482 bp for human DLST (mean of 1290 bp) (27).

These results clearly illustrate the potential of using conserved genes from human genetic disease loci as probes to isolate syntenic *Fugu* genomic clones, which then may be sequenced rapidly to identify candidate genes. Additionally, identification of intron–exon boundaries in *Fugu* genes can be used to predict the genomic structure of the human homologues. This is particularly useful when screening for mutations in candidate genes if genomic DNA is the only template available. Further advantages are that both conserved protein domains of functional importance and conserved noncoding regulatory sequences (28) may also be identified after sequence alignments.

Future work will determine how many of the other genes on cosmid 4D7 are also linked to 14q24.3 and whether synteny in this region extends as far as the gene responsible for the AD3 form of familial early onset Alzheimer disease, reported recently by three of the authors (12).

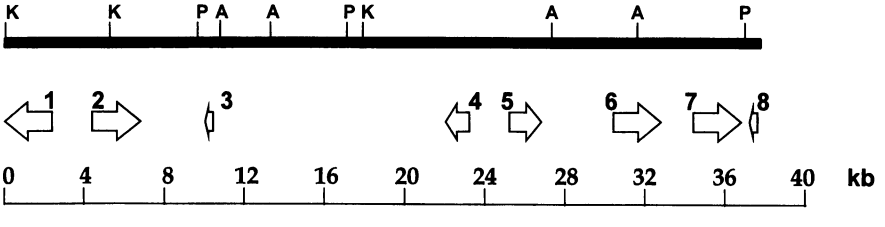

FIG. 2. Map of the *Fugu* genomic DNA segment inserted into cosmid 4D7. Restriction sites are labeled *Kpn* I (K), *Pme* I (P), and *Apa* I (A). The numbers refer to 1, prospero-like protein; 2, dihydrolipoamide succinyltransferase; 3, atf-like protein; 4, fos-like protein; 5, *cFOS*; 6, S31iii125; 7, S20i15; 8, 7SL RNA. The positions of the putative genes were determined by a combination of restriction mapping, long PCR (23), and contiging of DNA sequences. Transcriptional directions of the genes are defined by the arrowed boxes. The 3′ ends of both fos-like protein and atf3-like protein were not detected from analysis of the random DNA sequences obtained. The 4D7 genomic DNA insert lacked the extreme 3′ end of the prospero-like sequence. Note that the distance from the start codon of *cFOS* to the stop codon of S20i15 is 12.4 kb. After comparative DNA analysis by restriction digestion using the above enzymes, Southern blotting, and hybridization of probes generated across the isolated region, no rearrangements were found between the cosmid 4D7 insert DNA and *Fugu* genomic DNA (data not shown).

Table 1. *(Continued)*

Function	Human chromosome
Control of axonal outgrowth (15)	Not known
Oxidative decarboxylation (16)	14q24.3
Transcription factor (17)	Not known
Transcription factor (18)	Not known
Transcription factor (18)	14q24.3
Regulation of protein transport in the endoplasmic reticulum (19)	14q24.3
Subunit of a GDP–GTP exchange factor (20)	14q24.3
RNA component of the signal recognition particle (21)	Not known

We would like to thank Prof. S. Povey, Human Biochemical Genetics Unit, University College London, London, England and the United Kingdom Human Genome Mapping Project (UK HGMP) for the use of the confocal microscope facilities. We also thank others at the Glaxo-Wellcome Medicines Research Centre, Stevenage, England: Andy Lyall, Cary O'Donnell, and Steve Taylor for help with sequence analyses and Gillian Amphlett for statistical advice. We are indebted to Tony Kouzaridies and Christian Hagemeier of the Wellcome Cancer Research Campaign Institute, Cambridge, England, for the kind gift of plasmid pHKGfos and primers 365F (TK65) and 366R (TK118). The human chromosome 14-specific cosmid library was constructed by L. Deaven at the Los Alamos National Laboratory, New Mexico. R.S. is supported by a postdoctoral fellowship from Glaxo Canada. C.G.S. was supported by a grant from the UK HGMP. G.E. is supported by a grant from The Jeantet Foundation. The work in the Department of Medicine, University of Cambridge was partly supported by a grant from Glaxo Research and Development Ltd.

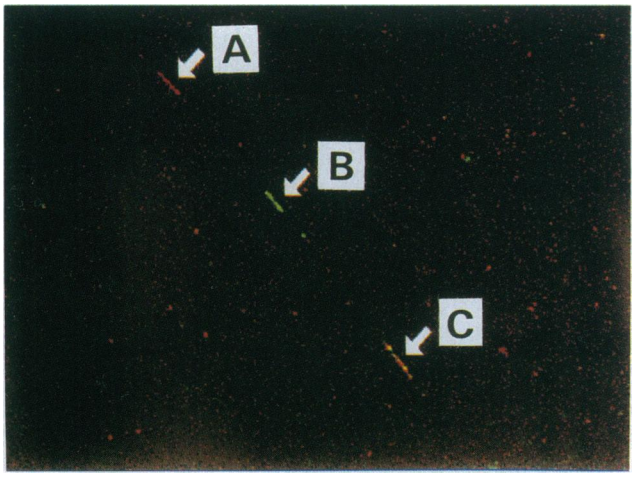

FIG. 3. FiberFISH analysis of human cosmids from the *AD3* locus. Ordering of the human *FOS*, S31iii125, and S20i15 cosmids on chromosome 14q24.3 was performed by FISH to extended human DNA fibers. On the extended DNA fibers, the hybridized S20i15 cosmid (p126H2) shows a red signal (A), the S31iii125 cosmid (I:66-4) shows a green signal (B), and the *FOS* cosmid (A6-7) shows an orange signal (a combination of red and green fluorescence) (C). This pattern was consistent in nine observations. Measurements in triplicate were made in seven of these observations. The calculated mean total distance spanned by S20i15 to *FOS* is 616 kb, with 95% confidence limits of 556 kb to 675 kb.

1. Buckler, A. J., Chang, D. D., Graw, S. L., Brook, J. D., Haber, D. A., Sharp, P. A. & Housman, D. E. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 4005–4009.
2. Lovett, M., Kere, J. & Hinton, L. M. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 9628–9632.
3. Parimoo, S., Patanjali, S. R., Shukla, H., Chaplin, D. D. & Weissman S. M. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 9623–9627.
4. Claverie, J.-M. (1994) *Genomics* **23,** 575–581.
5. Hunkapiller, T., Kaiser, R. J., Koop, B. F. & Hood, L. (1991) *Science* **254,** 59–67.
6. Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N., Green, P., Thierry-Mieg, J., Qiu, L., Dear, S., Coulson, A., Craxton, M., Durbin, R., Berks, M., Metzstein, M., Hawkins, T., Ainscough, R. & Waterson, R. (1992) *Nature (London)* **356,** 37–41.
7. Brenner, S. Elgar, G., Sandford, R., Macrae, A., Venkatesh, B. & Aparicio, S. (1993) *Nature (London)* **366,** 265–268.
8. St. George-Hyslop, P., Haines, J., Rogaev, E., Mortilla, M., Vaula, G., *et al.* (1992) *Nat. Genet.* **2,** 330–334.
9. Trower, M. K., Burt, D., Purvis, I. J., Dykes, C. W. & Christodoulou, C. (1995) *Nucleic Acids Res.* **23,** 2348–2349.
10. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
11. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2444–2448.
12. Sherrington, R., Rogaev, E. I., Liang, Y., Rogaeva, E. A., Levesque, G., *et al.* (1995) *Nature (London)* **375,** 754–760.
13. Heiskanen, M., Karhu, R., Hellsten, E., Peltonen, L., Kallioniemi, O. P. & Palotie, A. (1994) *BioTechniques* **17,** 928–933.
14. Genetics Computer Group (1991) *Program Manual for the GCG Package* (Genetics Computer Group, Madison, WI), Version 7.
15. Vaessin, H., Grell, E., Wolff, E., Bier, E., Jan, L. Y. & Jan, Y. N. (1991) *Cell* **67,** 941–953.
16. Nakano, K., Matuda, S., Sakamoto, T., Takase, C., Nakagawa, S., Ohta, S., Ariyama, T., Inazawa, J., Abe, T., & Miyata, T. (1993) *Biochim. Biophys. Acta* **1216,** 360–368.
17. Chen, B. P. C., Liang, G., Whelan, J. & Hai, T. (1994) *J. Biol. Chem.* **269,** 15819–15826.
18. Setoyama, C., Frunzio, R., Liau, G., Mudryj, M. & De Crombrugghe, B. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 3213–3217.
19. Wada, I., Rindress, D., Cameron, P. H., Ou, W.-J., Doherty, J. J., II, Louvard, D., Bell, A. W., Dignard, D., Thomas, D. Y. & Bergeron, J. J. M. (1991) *J. Biol. Chem.* **266,** 19599–19610.
20. Bushman, J. L., Asuru, A. I., Matts, R. L. & Hinnebusch, A. G. (1993) *Mol. Cell. Biol.* **13,** 1920–1932.
21. Walter, P. & Blobel, G. (1982) *Nature (London)* **299,** 691–698.
22. Barker, P. E., Rabin, M., Watson, M., Breg, W. R., Ruddle, F. H. & Verma, I. M. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 5826–5830.
23. Barnes, W. M. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 2216–2220.
24. Mason, P. J., Stevens, D. J., Luzzatto, L., Brenner, S. & Aparicio, S. (1995) *Genomics* **26,** 587–591.
25. Sandford, R. (1995) Ph.D. thesis (Univ. of Cambridge, Cambridge, U.K.).
26. Baxendale, S., Abdulla, S., Elgar, G., Buck, D., Berks, M., Micklem, G., Durbin, R., Bates, G., Brenner, S., Beck, S. & Lehrach, H. (1995) *Nat. Genet.* **10,** 67–75.
27. Nakano, K., Takase, C., Sakamoto, T., Nakagawa, S., Inazawa, J., Ohta, S. & Matuda, S. (1994) *Eur. J. Biochem.* **224,** 179–189.
28. Aparicio, S., Morrison, A., Gould, A., Gilthorpe, J., Chaudhuri, C., Rigby, P., Krumlauf, R. & Brenner, S. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 1684–1688.