

Cloning and regional localization of the mouse faciogenital dysplasia (*Fgd1*) gene

N.G. Pasteris,¹ B. de Gouyon,² A.B. Cadle,¹ K. Campbell,¹ G.E. Herman,² J.L. Gorski^{1,3}

¹Department of Human Genetics, University of Michigan Medical Center Ann Arbor, Michigan 48109-0688, USA

²Departments of Molecular and Human Genetics and Pediatrics, Baylor College of Medicine, Houston, Texas 77030-3498, USA

³Department of Pediatrics and Communicable Diseases, University of Michigan Medical Center, Ann Arbor, MI 48109-0688, USA

Received: 5 April 1995 / Accepted: 18 May 1995

Faciogenital dysplasia (FGDY, MIM*305400), or Aarskog syndrome, is a rare X-linked recessive multisystemic disease of abnormal human embryonic development. The disease phenotype consists of a number of facial, skeletal, and urogenital malformations. Major manifestations include disproportionate short stature and shortened distal extremities (Porteous and Goudie 1991). Radiographic abnormalities include hypoplastic phalanges, retarded bone maturation, and a variety of vertebral anomalies including cervical spina bifida occulta, odontoid hypoplasia, and segmentation anomalies. Facial features typically consist of hypertelorism, external ear anomalies, and maxillary hypoplasia. Scrotal anomalies, penile hypospadias, and kidney hypoplasia comprise some of the observed urogenital malformations (Gorlin et al. 1990).

Pedigree analyses of families segregating FGDY, and genetic linkage studies in some affected families, confirmed that FGDY was X-linked and mapped to the pericentric region of the human X Chromosome (Chr) (Porteous et al. 1992; Stevenson et al. 1994). FGDY was also mapped to this region on the basis of the observation of a mother and son who both displayed all of the major characteristics of FGDY in association with a reciprocal X:8 chromosome translocation (Bawle et al. 1984). This FGDY-specific translocation breakpoint was mapped by somatic cell and radiation hybrid analyses to a 350-kb interval within Xp11.21 that was flanked by markers ALAS2 and DXS323 (Gorski et al. 1992; Glover et al. 1993). By using a positional cloning strategy, the gene responsible for FGDY, termed FGD1, was recently isolated and characterized (Pasteris et al. 1994).

In this study, the mouse FGD1 homolog, *Fgd1*, was isolated and mapped to the mouse X Chr. A human FGD1 cDNA, clone pFCF3.85 (Pasteris et al. 1994), which contained the entire open reading frame (ORF), was used as a hybridization probe to screen a lambda gt10 library derived from e7.5 mouse embryonic ectoderm. The three mouse cDNA clones with the largest inserts were subcloned into Bluescript KSII(+) and sequenced (Pasteris et al. 1994). To facilitate DNA sequencing, we used oligonucleotides directed against the human FGD1 sequence for sequencing; directed against the mouse cDNAs, 80% of these oligonucleotide primers yielded interpretable sequence data. These results sug-

gested that a high degree of nucleotide sequence conservation existed between FGD1 cDNA clones derived from human and mouse.

The overlapping mouse cDNA clones corresponded to 3080 bp of contiguous nucleotide cDNA sequence and contained a 2880-nucleotide ORF predicted to encode a protein of 960 amino acids, one amino acid shorter than the human FGD1 ORF. Comparison of the mouse and human sequences within the coding region indicated 94.7% identity (96.3% similarity) at the amino acid level (Fig. 1A). The nucleotide sequence of all three mouse cDNAs terminated at an identical *EcoRI* site that corresponded to position 3713 of the human FGD1 and included 97 bp of 3' untranslated sequence. In addition, the largest mouse cDNA contained 104 bp of 5' untranslated sequence (Fig. 1B). Compared with the human FGD1 nucleotide sequence, the 5' untranslated region, the ORF, and the 3' untranslated region of the mouse cDNA were 95%, 92%, and 75% identical respectively. The predicted translation initiation sites of both the human and mouse cDNAs were identical. Together, these results strongly suggested that the composite mouse cDNA represented the mouse homolog of the human FGD1 gene, which was designated *Fgd1*.

An analysis of the derived *Fgd1* and FGD1 protein sequences predicted proteins with Rho/Rac guanine nucleotide exchange factor (GEF) activity. Rho/RacGEFs form a family of cytoplasmic proteins that activate the Ras-like family of Rho and Rac proteins by facilitating the exchange of bound GDP for free GTP (Boguski and McCormick 1993). The *Fgd1* GEF domain (180 residues from aa 375–559) was virtually identical (99.5% identity) to the Rho/RacGEF domain of the human FGD1 sequence (Fig. 1A). In addition, a comparison of the predicted mouse and human FGD1 protein sequences determined that both genes encoded the same putative structural motifs (Fig. 1B). These included a zinc-finger-like structural motif that was similar to but distinct from the zinc-finger regulatory domain of protein kinase C, a pleckstrin homology domain, and a proline-rich amino terminus domain that contained potential Src-homology 3 (SH3) binding sites (Pasteris et al. 1994). Another potential functional motif, an acidic peptide segment directly upstream from the GEF domain, was identified in the human/mouse comparison. A stretch of ten glutamic acid residues in the mouse gene (aa 339–349) was matched by 11 acidic residues in the human FGD1 polypeptide (EEEDDEEEEE). Other Rho/RacGEFs also contained acidic stretches upstream of the GEF domain; for example, rasGRF contained stretches of EDGEVEIE and EDEDSD residues in a similar region (GenBank accession No. P28818).

A

```

1 MHCHRVPGGPGSDPERSAANTFGAAPTACADSDPGALEPGLFVSRGSGT 50
1 MHGHRAPGGAGPSPPEHPATNPPGAAPPACADSDPGASEPGLLARRGSGS 50
51 ALGGPLDPQFVGPSSDALGAPPSRVLPCGSPQOHRALRFSTYHLEGSQP 100
51 ALGGPLDPQFVGPSSDALGAPPSRVLPCGSPQOHRALRFSTYHLEGSQP 100
101 RPLGHQGNRLILVKLSLSLDGQSLSPHPEGPQRLRSDPGPPTETIPGRPSF 150
101 RPLGHQGNRLILVKLSLSLDGQSLSPHPEGPQRLRSDPGPPTETIPGRPSF 150
151 LKRAPGPKPQVPPKPSYLMQPRVLPPEPEIPPPPSRPLPADPRVAKGLVP 200
151 LKRAPGPKPQVPPKPSYLMQPRMPPLEPIPPPSRPLPADPRVAKGLVP 200
201 RAEASTSSAAVSSSLEKFEREPVIVASDRPAPGCPVPPPEPAMLPOQPPQ 250
201 RAEASPSAAVSSSLEKFEREPVIVASDRPVPGPSPPGPEPVMLPQPTSQ 250
251 PTGSQLEGEASRCLFLLAPGPRDGEKVPNRDSDGIDSISSPNSBETCFV 300
251 PVPQLPEGEASRCLFLLAPGPRDGEKVPNRDSDGIDSISSPNSBETCFV 300
301 SDDGPIHSLCPGPPALASMPVALADPHRGSQEVDSLEEE .EEEEEE 349
301 SDDGPPSHSLCPGPPALASVVALADPHRGSQEVDSLEEEDEEEEE 350
350 KEREIPVPMERQESVELIVQKVFHIANELLQTEKAYVSRLLHLLDQVFC 399
351 KDREIPVPMERQESVELIVQKVFHIANELLQTEKAYVSRLLHLLDQVFC 400
400 ARLLLEARNRSFFPADVHGLFSNICSYCHQQLLPELEKRMENWDRY 449
401 ARLLLEARNRSFFPADVHGLFSNICSYCHQQLLPELEKRMENWDRY 450
450 PRIGDILQKLAPFLKMYGEYVKNFDRAVELVNTWTERSTQPKVIIHEVQK 499
451 PRIGDILQKLAPFLKMYGEYVKNFDRAVELVNTWTERSTQPKVIIHEVQK 500
500 EEACRNLTQHRHLEPQRI PRYELLKLDYLLKPHGSPDSKDAQSLEL 549
501 EEACGNLTQHRHLEPQRI PRYELLKLDYLLKPHGSPDSKDAQSLEL 550
550 IATAABHSNAAIRKMERMHKLLKVYELGGGEDIVSPTELKESGHLK 599
551 IATAABHSNAAIRKMERMHKLLKVYELGGGEDIVSPTELKESGHLK 600
600 SAKNGTTQDRYLILFNDRLLYCVPRLLRLLGQKFTVRRARIDVDGMELKSS 649
601 SAKNGTTQDRYLILFNDRLLYCVPRLLRLLGQKFTVRRARIDVDGMELKSS 650
650 NLNMPRTFLVSGKQRSLELQARTEEKKDWQAINSTLLKHEQTLTFPKL 699
651 NLNLPRTFLVSGKQRSLELQARTEEKKDWQAINSTLLKHEQTLTFPKL 700
700 LNSTNRDDEDTPPNSPNVDLGRAPTFIRREKEVTMCMRCQEPFNSITKRR 749
701 LNSTNRDDEDTPPNSPNVDLGRAPTFIRREKEVTMCMRCQEPFNSITKRR 750
750 HHCKACGHVCGKCSFRARLIYDNNRNRVCTDCYVALHGAPGSSPACS 799
751 HHCKACGHVCGKCSFRARLIYDNNRNRVCTDCYVALHGAPGSSPACS 800
800 QHTPQRRSILEKQASVAEENSVCISFLHYMEKGGKWHKAWFVVPENEP 849
801 QHTPQRRSILEKQASVAEENSVCISFLHYMEKGGKWHKAWFVVPENEP 850
850 LVLYTYGAPQDKAQRSLPLIGFVGPPEAGERPDRRHFVKITQSHLSWY 899
851 LVLYTYGAPQDKAQRSLPLIGFVGPPEAGERPDRRHFVKITQSHLSWY 900
900 FSPETEELQRRMVAVLGRAGGDTFCGPTLSEDKEMEETPVAASGATAE 949
901 FSPETEELQRRMVAVLGRAGGDTFCGPTLSEDKEMEETPVAASGATAE 950
950 PPEASQTRDKT 960
951 PPEESQTRDKT 961
    
```

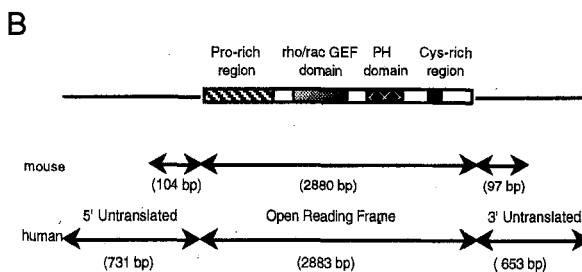


Fig. 1. Comparisons of the mouse and human predicted *Fgd1* protein sequences. (A) The predicted mouse (top) and human (bottom) *Fgd1* protein sequences were aligned by use of the BESTFIT program of the GCG software package (Devereux et al. 1984); vertical lines between amino acids indicate identity, double dots signify conserved amino acids. The mouse *Fgd1* sequences and the human FGD1 sequences are deposited in GenBank under accession numbers U22325 and U11690 respectively. A schematic representation of the *Fgd1* sequence is shown below (B). The ORF region is shown as a boxed segment. Identified structural domains are indicated and include (left to right): a proline-rich N-terminus containing putative Src-homology 3 (SH3) binding sites, a Rho/Rac guanine nucleotide exchange factor (GEF) domain, a pleckstrin homology (PH) domain, and a cysteine-rich, zinc-finger-like domain.

In order to further confirm that *Fgd1* was indeed the mouse homolog of FGD1, we determined the mouse chromosomal location of *Fgd1*. An interspecific backcross analysis was carried out with progeny derived from matings of (B6CBA-A^w/A-Bpa) × *Mus spretus* mice, as previously described by Herman and Walton (1990). This interspecific backcross mapping panel consisted of 126 animals with recombination events spanning the mouse X Chr, including 28 crossovers between *Ags* and *Amg*. Rapid regional mapping was performed by genotyping 14 mice representing crossover events spanning several anchor loci distributed along the entire X Chr (data not shown). These data positioned *Fgd1* between the anchor loci *Ags* and *Amg*, as expected from comparative human and murine maps (Fig. 2 and Table 1).

For finer resolution for loci mapped to the interval between *Ags* and *Amg*, mice were also typed for three microsatellites, *DXMit12*, *34*, and *35*, as well as for cDNA marker *Alas2*. The *Alas2* locus had been previously mapped to the interval between *Ags* and *Amg* by Chapman and associates (1994). The *Alas2* and *Fgd1* probes detected a *Bam*HI variation (RFLV) between *Mus spretus* and C57BL/6J, thus providing a means for scoring the interspecific backcross. The mapping results favored a gene order of: cen-*Ags*-*DXMit35*-(*Alas2*, *DXMit34*)-*Fgd1*-*DXMit12*-*Amg*-tel (Fig. 2, Table 1). As shown in Table 1, a single double recombinant, mouse 199, was detected in these studies and, thus, the relative order for *DXMit12* and *Amg* was based on data from independent, multilocus, interspecific backcrosses (Herman et al. 1994). The *Fgd1* locus was mapped 0.79 cM (± a standard error of 0.79 cM) telomeric to *Alas2* and 6.3 cM (±2.2 cM) centromeric to *DXMit12*. The observation that, like the human FGD1 locus, the mouse *Fgd1* gene was closely linked to *Alas2* further confirmed that the isolated cDNA clones represented the *Fgd1* locus. These results also extended and confirmed that a cluster of loci within the human proximal short arm of the X Chr within region Xp11.21 mapped to the distal region of the mouse X Chr, extending the observations of Blair and colleagues (1994).

Several mouse mutations have been tentatively mapped to the distal portion of the mouse X Chr. These include gyro (*Gy*), hypophosphatemia (*Hyp*), lined (*Li*), and X-linked polydactyly (*Xpl*) (Lyon and Searle, 1989). These mutations are manifested by hypophosphatemia, rickets, inner ear abnormalities, sterility, and polydactyly, features not associated with the FGDY phenotype. However, because the reported cases of FGDY have been associated with FGD1 null alleles only (Pasteris et al. 1994), and since it is not known what phenotype(s) may be associated with more subtle FGD1 mutations, it is not possible to absolutely exclude the possibility that *Fgd1* is responsible for one of these mouse mutations.

An examination of the nucleotide and deduced protein sequences of *Fgd1* and FGD1 revealed a high degree of homology between the two genes. The remarkable fidelity of amino acid conservation of all structural motifs identified in *Fgd1*, including the Rho/RacGEF domain, may signify that the mouse and human signaling cascades are also equally well conserved. Other potential components of this pathway, such as the mouse and human Rac1, Rac2, and Cdc42 proteins, are 100% conserved at the amino acid level (unpublished observation, Pasteris). These results and observations strongly suggest that the mouse will serve as a useful model for studying and characterizing the *Fgd1* development signaling transduction system.

Acknowledgments. We thank Sue Varmuza (University of Toronto) for the mouse e7.5 embryonic ectoderm cDNA library. The work was supported, in part, by the General Clinical Research Center at the University of Michigan and was funded by National Center for Research Resources grant RR-00042. The mapping work was supported by National Institutes of Health grant RO1-HG00131 to G.E. Herman. This work was supported, in part, by the March of Dimes-Birth Defects Foundation basic science grant 1-93-0326 and National Institutes of Health grant NS-30771 to J.L. Gorski.

A

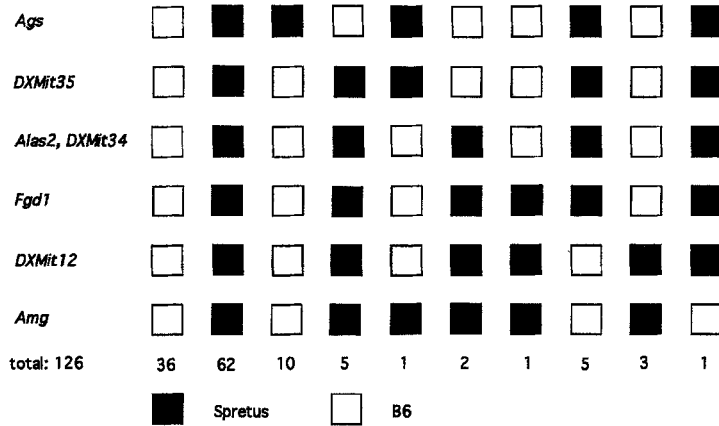


Fig. 2. Genetic mapping *Fgd1* in an interspecific backcross. (A) Analysis of 126 animals at *Ags*, *DXMit35*, *Alas2*, *DXMit34*, *Fgd1*, *DXMit12*, and *Amg*. All 126 N₂ progeny were typed for the loci *Ags*, *DxMit12*, and *Amg*. Only the 28 recombinant progeny were typed for the additional loci *DXMit35*, *Alas2*, *DXMit34*, and *Fgd1*. Black boxes represent *Mus spretus* alleles, and white boxes represent B6CBA-A^{w-J}/A-*Bpa*(domesticus) alleles. A schematic representation of the mouse X Chr is shown to the right of the

Table 1. Haplotype analysis of X Chromosome markers.^a

Mouse	<i>Ages</i>	<i>DXMit35</i>	<i>Alas2</i> , <i>DXMIT34</i>	<i>Fgd1</i>	<i>DXMit12</i>	<i>Amg</i>
102	S	x B	B	B	B	B
119	S	x B	B	B	B	B
124	S	x B	B	B	B	B
147	S	x B	B	B	B	B
220	S	x B	B	B	B	B
233	S	x B	B	B	B	B
235	S	x B	B	B	B	B
247	S	x B	B	B	B	B
253	S	x B	B	B	B	B
268	S	x B	B	B	B	B
161	B	x S	S	S	S	S
216	B	x S	S	S	S	S
243	B	x S	S	S	S	S
278	B	x S	S	S	S	S
280	B	x S	S	S	S	S
199	S	S	x B	B	B	x S
160	B	B	x S	S	S	S
210	B	B	x S	S	S	S
274	B	B	B	x S	S	S
42	S	S	S	S	x B	B
201	S	S	S	S	x B	B
202	S	S	S	S	x B	B
204	S	S	S	S	x B	B
266	S	S	S	S	x B	B
217	B	B	B	B	x S	S
229	B	B	B	B	x S	S
276	B	B	B	B	x S	S
209	S	S	S	S	S	x B

^a S denotes *Mus spretus* allele; B denotes B6CBA-A = /A-*Bpa* allele; × indicates a recombination event.

B

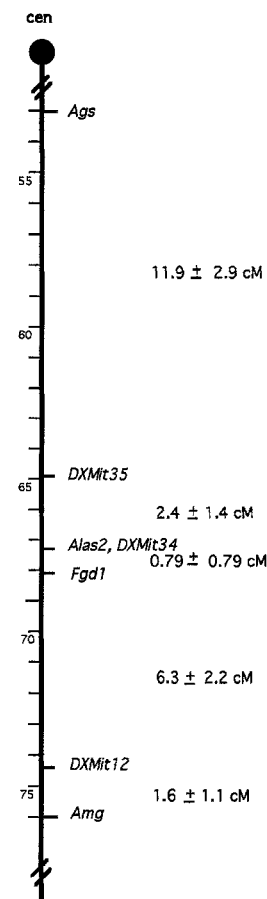


figure (B). The order and position of the anchor loci were previously described (Herman et al. 1994). Calculated distances between markers are shown to the right of the figure. PCR primers for *DXMit12*, *34* and *35* were purchased from Research Genetics and used as recommended by the manufacturer. Genomic DNA restriction digests, electrophoresis, and Southern blotting to Sure Blot membranes (Oncor) were performed as described (Herman and Walton 1990; Angel et al. 1993).

References

Angel, T.A., Faust, C.J., Gonzales, J.C., Kenwick, S., Lewis, R.A., Herman, G.E. (1993). Genetic mapping of the X-linked dominant mutations striated (*Str*) and bare patches (*Bpa*) to a 600-kb region of the mouse X chromosome: implications for mapping human disorders in Xq28. *Mamm. Genome* 4, 171–176.

Bawle, E., Tyrkus, M., Lipman, S., Bozimowski, D. (1984). Aarskog syndrome: full male and female expression associated with an X-autosome translocation. *Am. J. Med. Genet.* 17, 595–602.

Blair, H.J., Reed, V., Lavak, S.H., Boyd, Y. (1994). New insights into the man-mouse comparative map of the X chromosome. *Genomics* 19, 215–220.

Boguski, M.S., McCormick, F. (1993). Proteins regulating Ras and its relatives. *Nature* 366, 643–653.

Chapman, V.M., Keitz, B.T., Bishop, D.F. (1994). Genetic linkage of the erythroid-specific delta-aminolevulinate synthase gene (*Alas2*) to the distal region of the mouse X chromosome. *Mamm. Genome* 5, 714.

Devereux, J., Haerberli, P., Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12, 387–395.

Glover, T.W., Verga, V., Rafael, J., Gorski, J.L., Bawle, E., Higgins, J.V. (1993). Translocation breakpoint in Aarskog syndrome maps to Xp11.21 between ALAS2 and DXS323. *Hum. Mol. Genet.* 10, 1717–1718.

Gorlin, R.J., Cohen, M.M., Levin, L.S. (1990). *Syndromes of the Head and Neck*, 3rd ed. (New York, N.Y. Oxford University Press), pp. 295–297.

Gorski, J.L., Boehnke, M., Reyner, E.L., Burreight, E.N. (1992). A radiation hybrid map of the proximal short arm of the human X chromosome spanning incontinentia pigmenti (IP1) translocation breakpoints. *Genomics* 14, 657–665.

Herman, G.E., Walton, S.J. (1990). Close linkage of the murine locus bare

- patches to the X-linked visual pigment gene: implications for mapping human X-linked dominant chondrodysplasia punctata. *Genomics* 7, 307–312.
- Herman, G.E., Boyd, Y., Chapman, V., Chatterjee, A., Brown, S.D.M. (1994). Mouse X chromosome. *Mamm. Genome* 5 (suppl.), S276–S288.
- Lyon, M.F., Searle, A.G. (1989). *Genetic Variants and Strains of the Laboratory Mouse*, 2nd ed. (New York, N.Y.: Oxford University Press).
- Pasteris, N.G., Cadle A., Logie L.J., Porteous M.E.M., Schwartz, C.E., Stevenson R.E., Glover, T.W., Wilroy, R., Gorski, J.L. (1994). Isolation and characterization of the faciogenital dysplasia (Aarskog-Scott syndrome) gene: a putative Rho/Rac guanine nucleotide exchange factor. *Cell* 79, 669–678.
- Porteous, M.E.M., Goudie, D.R. (1991). Aarskog syndrome. *J. Med. Genet.* 28, 44–47.
- Porteous, M.E.M., Curtis, A., Lindsay, S., Williams, O., Goudie, D., Kamakari, S., Bhattacharya, S.S. (1992). The gene for Aarskog syndrome is located between DXS255 and DXS566 (Xp11.2-Xq13). *Genomics* 14, 298–301.
- Stevenson, R.E., May, M., Arena, J.F., Millar, E.A., Scott, Jr. C.I. (1994). Aarskog-Scott syndrome: confirmation of linkage to the pericentric region of the X chromosome. *Am. J. Med. Genet.* 52, 339–345.