Constructing a draft Human Fer1L6 sequence

Top(red): predicted sequence of chimpanzee (Pan troglodytes) Fer1L6 (XP_515743)

Bottom(black): human sequence assembled from predicted partial sequence of Fer1L6 (LOC90342) and translated human ESTs. (portions in blue are translated human reference sequence (**not** verified at the RNA level). Vertical lines: | mark approximate exon boundaries.

```
mviskspsvw srvqgsivpe hegahgaedh lgitareaas qklmvpgsta hralsskpqh 61
MVISKSPSVW SRVQGSIVPE HEGAHGAEDH LGITAREAAS QKLMVPGSTA HRALSSKPQH
                           |

fqpragaarl svpggrgrag gvrashrrlp aqppeppppp psmaaapppd sssgsspvsv 121
FQ   !! this matches a     |       low complexity?              V
  |     seq. from Chr. 1!! |                                    |

rvkvfearql mgnnikpvvk vsiagqqhqt rikmgnnpff neiffqnfhe vpakffdeti 181
RVKVFEARQL MGNNIKPVVK VSIAGQQHQT RIKMGNNPFF NEIFFQNFHE VLAKFFDETI
                                              |

liqvvnssam rykaeigrfq tdigfiyhsp ghtllrkwlg lcqpnnpgsg vtgylkvtiy 241
LIQVVNSSAM RYKAEIGRFQ TDIGFIYHSP GHTLLRKWLG LCQPNNPGSG VTGYLKVTIY
    |              |             |

algvgdqali dqkllygtdd tdiqifksav vpmnmaylql fiycaedlhl kkhqsvnpql 301
ALGVGDQALI DQKLLYGTDD TDIQIFKSAV VPINMAYLQL FIYCAEDLHL KKHQSVNPQL
         |                                                 |

eveligeklr thmqtqtdnp iwnqiltfri qlpclssyik frvldcrkkd cpdeigtasl 361
EVELIGEKLR THMQTQTDNP IWNQILTFRI QLPRLSSYIK FRVLDCRKKD CPDEIGTASL
        |                         |              |

slnqisstge eiegvysgfl pcfgpsfltl hggkkapfri qeegacipds vrdglayrgr 421
SLNQISSTGE EIEGVYSGFL PCFGPSFLTL HGGKKAPFRI QEEGACIPDS VRDGLAYRGR
               |                                  |

vflelitqik syqdstikdl shevtriekh qnrqkyglcv iflsctmmpn fkelihfevs 481
VFLELITQIK SYQDSTIKDL SHEVTRIEKH QNRQKYGLCV IFLSCTMMPN FKELIHFEVS
                                 |

ighygnkmdl nykplvsstq yspviydgni yhyvpwyntk pvvavtsnwe dvsfrmncln 541
IGHYGNKMDL NYKPLVSSTP YSPVIYDGNI YHYVPWYNTK PVVAVTSNWE DVSFRMNCLN
                               |

llhftrdrlk anldtpestr npkdpallyq hplpcmtyqp katsldrkrw qlrslllqel 601
LLHFTRDRLK ANLDTLKSTR NPKDPALLYQ  PLPCMTYQP KATSLDRKRW QLRSLLLQEL
         |                         |

aqkakqakpk dmvataedwl yrlnavlpep qmglpdvmiw lvakeqrvay aqvpahsilf 661
AQKAKQAKPK DMVATAEDWL YRLNTVLPEP QMGLPDVMIW LVAKEQRVAY AQVPAHSVLF
                                 |
```

```
spagalhsgr lcgkiqtlfl qvgirgssge draekcpqgs plnsqgspyg acfpgplttw 721
SPAGALHSGR LCGKIQTLFL QVGIRDSSGE DRAEKCPQGS PLNSQGSPYG ACFPGPLTTW

ngvetvprgr egsvavsmlg daksmyllng pqihllaghg gsqypegegq kdvlpahlrv 781
NGVETVPRGR EGSVAVSMLG DAKSMYLLNV PQIHLLAGHG GSQYPEGEGQ KDVLPAHLRV

cmwlgnvtds kdlqllrqgd tavyaemyen qakykdqwgq qglyhcpnfs dvmgnktlpm 841
CMWLGNVTDS KDLQLLRQGD TAVYAEMYEN QAKYKDQWGQ QGLYHCPNFS DVMGNKTLPM

tdfqpplgwh wqdswtvepq rrllldidin ksqvleevye nqgrdtrgaw gpaaipntdv 901
TDFQPPLGWH WQDSWTVEPQ RRLLLDIDIN KSQVLEEVYE NQGRDTRGAW GPAAIPNTDV

ngqpmearen vkcpqgwhfk kdwvvelnha vdskgweygv gippsglpqv wspvektyhs 961
NGQPMEAREN VKCPQGWHFK KDWVVELNHA VDSKGWEYGV GIPPSGLPQV WSPVEKTYHS

crrrrwarvr frnhgelshe qetlsflqlg lakgeeegwe ydtfgskfhl npqpqsrfrr 1021
CRRRRWARVR FRNHGELSHE QETLSFLQLG LAKGEEEGWE YDTFGSKFHL NPQPQSRFRR

rcwrrrlapn kdkgiapifl legslamdlk yhagkeedsk twpwgldrqf rdpqrqdtrp 1081
RCWRRRLAPN KDKGIAPIFL LEGSLAMDLK YHAGKEEDSK TWPWGLDRQF RDPQRQDTRP

pnlpfiyctf nkphyyqlfc yiyqarnlvs nqivtfqapf irvvflnhsq ctqtlrssag 1141
PNLPFIYCTF NKPHYYQLFC YIYQARNLVS NQILTFQGPF IRVVFLNHSQ CTQTLRSSAG

ptwaqtlifq hlllyenpqd tkespplvvl elwqrdfwgk eslwgrsmwp plvwldlqdr 1201
PTWAQTLIFQ HLLLYENPQD TKESPPLVVL ELWQRDFWGK ESLWGRSVWP PMVWLDLQDR

ilppmrwhpl vkelgkeege ilascelilq teklgekqlp ilsvpwknga ytlpksiqpt 1261
ILPPMRWHPL VKELGKEEGE ILASCELILQ TEKLGEKQLP ILSVPWKNGA YTLPKSIQPT

ikrmaieila wglrnmkkas spqllvefge eslrtepird fqtnpnfpes esvlvltvlm 1321
IKRMAIEILA WGLRNMKKAS SPQLLVEFGE ESLRTEPIRD FQTNPNFPES ESVLVLTVLM

pteeayalpl vvkvvdnwaf gqqivtgqan idflqpyfcd pwaqdymhpk lpstalppcl 1381
PTEEAYALPL VVKVVDNWAF GQQTVTGQAN IDFLQPYFCD PWAQDYMHPK LPSTALPPCL

spgltsepsl gvgqallprg flgylyrkfw fksskaedey ehevdwwskl fwatdehksl 1441
SPGLTSEPAL GVGQALLPRG FLGYLYRKFW FKSSKAEDEY EHEVDWWSKL FWATDEHKSL

kykykdyhtl kvyeceleav pafqglqdfc qtfklyqeqp kldspvvgef kglfriypfp 1501
KYKYKDYHTL KVYECELEAV PAFQGLQDFC QTFKLYQEQP KLDSPVVGEF KGLFRIYPFP
```

```
enpeapkppl qflvwpered fpqpclvrvy vvrainlqpq dynglcdpyv ilklgktelg 1561
ENPEAPKPPL QFLVWPERED FPQPCLVRVY MVRAINLQPQ DYNGLCDPYV ILKLGKTELG
                                                 |

nrdmyqpntl dpifgtmfel tcniplekdl eiqlydfdlf spddkigttv idlenrllsg 1621
NRDMYQPNTL DPIFGMMFEL TCNIPLEKDL EIQLYDFDLF SPDDKIGTTV IDLENRLLSG
               |

fgahcglsks ycqsgpfrwr dqmppsylle ryakrkglpp plfspeedav fyngkkfklq 1681
FGAHCGLSKS YCQSGPFRWR DQMPPSYLLE RYAKRKGLPP PLFSPEEDAV FYNGKKFKLQ
             |

sfepktptvh glgpkkerla lyllhtqglv pehvetrtly sdsqpgidqg kvqmwvdifp 1741
SFEPKTPTVH GLGPKKERLA LYLLHTQGLV PEHVETRTLY SHSQPGIDQG KVQMWVDIFP
  |                                                 |

kklgppgpqv ninprkpkry elrciiwkta nvdlvddnls rektsdiyik gwlyglekdm 1801
KKLGPPGPQV NINPRKPKRY ELRCIIWKTA NVDLVDDNLS REKTSDIYIK GWLYGLEKDM
                     |                                   |

qktdihyhsl tgeadfnwrf vftmdylaae rmcvqsqkdy iwsldatsmk fparliiqvw 1861
QKTDIHYHSL TGEADFNWRF IFTMDYLAAE RTCVQSQKDY IWSLDATSMK FPARLIIQVW
                                        |

dndifspddf lgvleldlsd mplparhakq csirmmdadp kwpyfvqykh fslfkkktvt 1921
DNDIFSPDDF LGVLELDLSD MPLPARHAKQ CSIRMMDADP KWPYFIQYKH FSLFKKKTVT
            |

gwwpcqvldg gkwrlsgkvk msleilseke alikpagrgq sepnqyptlh pplrtntsfm 1981
GWWPCQVLDG GKWRLSGKVK MSLEILSEKE ALIKPAGRGQ SEPNQYPTLH PPLRTNTSFT
                  |                                     |

wlrspvqnfc yifwkryrfk liafmvisii almlfnfiys aphylamswi kpqlqlyppi 2041
WLRSPVQNFC YIFWKRYRFK LIAFMVISII ALMLFNFIYS APHYLAMSWI KPQLQLYPPI
                                           |

kifniinsln tsnasssilp tqdpnlkpti dhewklhpgp tnhlsdifpe lpapgd 2096
KIFNIINSLN TSNASSSILP TQDPNLKPTI DHEWKLHPGP TNHLSDIFPE LPAPGD
```

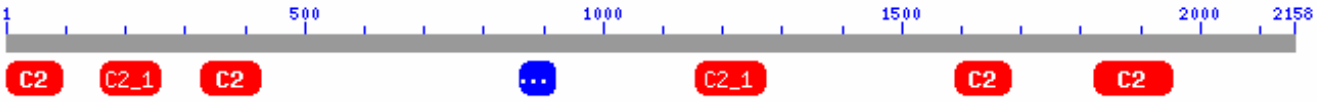
The Fer1L6 protein sequence above corresponds almost exactly in animo acid sequence (with some gaps in each sequence) to protein hCG1810857 (Celera Genomics, accession number EAW71345). The hCG1810857 sequence has a different N-terminal region, much of which is validated by ESTs, which contains an N-terminal C2 domain, similar to dysferlin and myoferlin.

In the composite sequence below, the amino acids in black are from human ESTs; in brown, predicted sequence matched with high homology and no gaps to ESTs in other mammals; in blue, from translated human reference DNA corresponding to the predicted chimpanzee sequence; in green, from hCG1810857 (not validated by ESTs).

human Fer1L6 draft sequence

MLRLVVQSAK IDPPLAPLPR PCMSIDFRDI KKRTRVVEGN DPVWNETLIW HLWNRPLEND
SFLQVTLQDM GSQKKERFIG LATVLLKPLL KQPSEVLFVK DLTLLNHSMK PTDCTVTLQV
AHMSNQDIEK TGAEDHLGIT AREAASQKLM VPGSTAHRAL SSKPQHFQVR VKVFEARQLM
GNNIKPVVKV SIAGQQHQTR IKMGNNPFFN EIFFQNFHEV LAKFFDETIL IQVVNSSAMR
YKAEIGRFQT DIGFIYHSPG HTLLRKWLGL CQPNNPGSGV TGYLKVTIYA LGVGDQALID
QKLLYGTDDT DIQIFKSAVV PINMAYLQLF IYCAEDLHLK KHQSVNPQLE VELIGEKLRT
HMQTQTDNPI WNQILTFRIQ LPRLSSYIKF RVLDCRKKDC PDEIGTASLS LNQISSTGEE
IEGVYSGFLP CFGPSFLTLH GGKKAPFRIQ EEGACIPDSV RDGLAYRGRV FLELITQIKS
YQDSTIKDLS HEVTRIEKHQ NRQKYGLCVI FLSCTMMPNF KELIHFEVSI GHYGNKMDLN
YKPLVSSTPY SPVIYDGNIY HYVPWYNTKP VVAVTSNWED VSFRMNCLNL LHFTRDRLKA
NLDTLKSTRN PKDPALLYQW EKLLRELAED CKRPLPCMTY QPKATSLDRK RWQLRSLLLQ
ELAQKAKQAK PKDMVATAED WLYRLNTVLP EPQMGLPDVM IWLVAKEQRV AYAQVPAHSV
LFSPAGALHS GRLCGKIQTL FLQVGIRDSS GEDRAEKCPQ GSPLNSQGSP YGACFPGPLTTWNG
VETVPRGREG SVAVSMLGDA KSMYLLNVPQ IHLLAGHGGS QYPEGEGQKD VLPAHLRVCM
WLGNVTDSKD LQLLRQGDTA VYAEMYENQA KYKDQWGQQG LYHCPNFSDV MGNKTLPMTD
FQPPLGWHWQ DSWTVEPQRR LLLDIDINKS QVLEEVYENQ GRDTRGAWGP AAIPNTDVNG
QPMEARENVK CPQGWHFKKD WVVELNHAVD SKGWEYGVGI PPSGLPQVWS PVEKTYHSCR
RRRWARVRFR NHGELSHEQE TLSFLQLGLA KGEEEGWEYD TFGSKFHLNP QPQSRFRRRC
WRRRLAPNKD KGIAPIFLLE GSLAMDLKYH AGKEEDSKTW PWGLDRQFRD PQRQDTRPPN
LPFIYCTFNK PHYYQLFCYI YQARNLVSNQ ILTFQGPFIR VVFLNHSQCT QTLRSSAGPT
WAQTLIFQHL LLYENPQDTK ESPPLVVLEL WQRDFWGKES LWGRSVWPPM VWLDLQDRIL
PPMRWHPLVK ELGKEEGEIL ASCELILQTE KLGEKQLPIL SVPWKNGAYT LPKSIQPTIK
RMAIEILAWG LRNMKKASSP QLLVEFGEES LRTEPIRDFQ TNPNFPESES VLVLTVLMPT
EEAYALPLVV KVVDNWAFGQ QTVTGQANID FLQPYFCDPW AQDYMHPKLP STALPPCLSP
GLTSEPALGV GQALLPRGFL GYLYRKFWFK SSKAEDEYEH EVDWWSKLFW ATDEHKSLKY
KYKDYHTLKV YECELEAVPA FQGLQDFCQT FKLYQEQPKL DSPVVGEFKG LFRIYPFPEN
PEAPKPPLQF LVWPEREDFP QPCLVRVYMV RAINLQPQDY NGLCDPYVIL KLGKTELGNR
DMYQPNTLDP IFGMMFELTC NIPLEKDLEI QLYDFDLFSP DDKIGTTVID LENRLLSGFG
AHCGLSKSYC QSGPFRWRDQ MPPSYLLERY AKRKGLPPPL FSPEEDAVFY NGKKFKLQSF
EPKTPTVHGL GPKKERLALY LLHTQGLVPE HVETRTLYSH SQPGIDQGKV QMWVDIFPKK
LGPPGPQVNI NPRKPKRYEL RCIIWKTANV DLVDDNLSRE KTSDIYIKGW LYGLEKDMQK
TDIHYHSLTG EADFNWRFIF TMDYLAAERT CVQSQKDYIW SLDATSMKFP ARLIIQVWDN
DIFSPDDFLG VLELDLSDMP LPARHAKQCS IRMMDADPKW PYFIQYKHFS LFKKKTVTGW
WPCQVLDGGK WRLSGKVKMS LEILSEKEAL IKPAGRGQSE PNQYPTLHPP LRTNTSFTWL
RSPVQNFCYI FWKRYRFKLI AFMVISIIAL MLFNFIYSAP HYLAMSWIKP QLQLYPPIKI
FNIINSLNTS NASSSILPTQ DPNLKPTIDH EWKLHPGPTN HLSDIFPELP APGD

Conserved Domain Map of Draft Human Fer1L6 sequence above



C2 Domains

C2 Domains

Ferlin domain (unknown function)