Supplementary materials

Protein intrinsic disorder and induced pluripotent stem cells

Bin Xue[a,b], Christopher J. Oldfield[b], Ya-Yue Van[c], A. Keith Dunker[b,§], Vladimir N. Uversky[a,b,d,§]

[a]Department of Molecular Medicine, College of Medicine, University of South Florida, Tampa, Florida 33612, USA;

[b]Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN 46202, USA;

[c]Molecular Kinetics Inc. Indianapolis, IN 46254, USA;

[d]Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia;

[§]Corresponding authors

Supplementary materials:

Suppl. Fig. 1 Sequence alignment for proteins in four families: (a) Sox; (b) Myc; (c) Klf; (d) Dppa.

```
Sox1    ---MYSMMMETDLHSPGGAQAPTNLSGP-------------------AGARGGGGGGGGG    38
Sox3    ----MYSLLETELKNPVGPPTPAAGTGVPAAPGAAGKSGANPAGGANAGNGGSGGANGGG   56
Sox2    ----MYNMMETELKPPGPQQAS--------------------------GGGGGGGNATAA    30
Sox15   MALTSSSQAETWSLHPRASTASLP--------------------------LGPQEQEAGG    34
                 **    *    :.                             *      ..

Sox1    GGGGGTKANQDRVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKVMSEAEKR    98
Sox3    GGGGGGGSDQDRVKRPMNAFMVWSRGQRRKMALENPKMHNSEISKRLGADWKLLTDAEKR   116
Sox2    ATGGNQKNSPDRVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKR    90
Sox15   SPGASGGLPLEKVKRPMNAFMVWSSVQRRQMAQQNPKMHNSEISKRLGAQWKLLGDEEKR    94
        . *..     ::************  ***:** :***************:**:: : ***

Sox1    PFIDEAKRLRALHMKEHPDYKYRPRRKTKTLLK-KDKYSLAGGLLAAGAGGGGAAVAMGV   157
Sox3    PFIDEAKRLRAVHMKEYPDYKYRPRRKTKTLLK-KDKYSLPGGLPPPGAAAAAAAAAAAA   175
Sox2    PFIDEAKRLRALHMKEHPDYKYRPRRKTKTLMK-KDKYTLPGGLLAPGGNSMASGVGVGA   150
Sox15   PFVEEAKRLRARHLRDYPDYKYRPRRKSKNSSTGSVPFSQEGGGLACGGSHWGPGYTTTQ   154
        **::******* *::::**********:*.  . .  ::  **  . *.   ...

Sox1    GVGVGAAAVGQRLESPGGAAGGGYAHVNGWANGAYPGSVAAAAAAAAMMQEAQLAYGQHP   217
Sox3    AASS-PVGVGQRLDT--------YTHVNGWANGAYS------------LVQEQLGYAQPP   214
Sox2    GLGG---GLNQRMDS--------YAHMNGWSNGSYS------------MMQEQLGYPQHP   186
Sox15   GSRG--------------------------------------------------------   158
        .

Sox1    GAGGRHPHAHPAHPHPHHPHAHPHNPQPMHRYDMGALQYSPIS--NSQGYMSASPSGYGG    275
Sox3    SMSSPPP---------------PPALPQMHRYDMAGLQYSPMMPPGAQSYMNAAAAAAAA   259
Sox2    GLN----------------AHGAAQMQPMHRYVVSALQYNSMT--SSQTYMNGSP-----   223
Sox15   ------------------------------------FGYQPPN-----------------   165
                                            : *..

Sox1    IPYG-AAAAAAAAAGGAHQNSAVAAAAAAAAASSGALGALGSLVKSEPSGSP---PAP--   329
Sox3    SGYGGMAPSAAAAAAAAAYGQQPATAAAAAAAAAAMSLGPMGSVVKSEPSSPP---PAIR-   315
Sox2    ---------------------TYSMSYSQQGTPGMALGSMGSVVKSEASSSP---PVVTS   259
Sox15   ----------------------YSTAYLPGSYTSSHCRPEAPLPCTFPQSDPRLQG----   199
                               : :    . ..    . ..:  : ... *

Sox1    -AHSRAPCPG-DLREMISMYLPAGEGGDPAAAAAAAAQSRLHSLPQHYQG--AGAGVNGT   385
Sox3    -SHSQRACLG-DLRDMISMYLPPG--GDAADAASPLPGGRLHGVHQHYQG--AGTAVNGT   369
Sox2    SSHSRAPCQAGDLRDMISMYLPGA------EVPEPAAPSRLHMAQHYQSGPVPGTAKYGT   313
Sox15   -----------ELRPSFSPYLSPD-----------------------SSTPYNTSLAGAP   225
                   :**  :* **.                          .     :   ..

Sox1    VPLTHI   391
Sox3    VPLTHI   375
Sox2    LPLSHM   319
Sox15   MPVTHL   231
        :*::*:
```

Fig. 1 (a)

```
l-Myc      ------------------MDFDSYQHYFYDYDCGEDFYR-------STAPSEDIWKKFEL    35
n-Myc      MPSCTASTMPGMICKNPDLEFDSLQPCFYPDEDDFYFGGPD-----STPPGEDIWKKFEL    55
c-Myc      -----MPLNVNFTNRNYDLDYDSVQPYFICDEEENFYHQQQQSELQPPAPSEDIWKKFEL    55
                             :::** *  *   :    :         ...*.*********

l-Myc      VPSPPTSPPWGSGPGAVDPASGINPGEPWPGGGAG--DEAESRGHSKAWGRN--------    85
n-Myc      LPTPPLSPSRAFPEHSPEPSNWATEMLLPEADLWGNPAEEDAFGLGGLGGLT--------   107
c-Myc      LPTPPLSPSRRSGLCSPSYVAVATSFSPREDDDGGGGNFSTADQLEMMTELLGGDMVNQS   115
           :*:** **.      : .     .       .  *      :

l-Myc      ----------YASIIRRDCMWSGFSARERLERVVSDRLAPGAPR----------------   119
n-Myc      ----------PNPVILQDCMWSGFSAREKLERAVNEKLQHGHGPPGVSSACSAPGVGASS   157
c-Myc      FICDPDDETFIKNIIIQDCMWSGFSAAAKLVSEKLASYQAARKDSTSLSPARG-------   168
                        :* :*********  :*          .

l-Myc      --------------------------------------------------GNPPKAPATP   129
n-Myc      PGGRALGGSSSASHTGATLPTDLSHPAAECVDPAVVFPFPVNKRESASVPAAPTSAPATS   217
c-Myc      ---------HSVCSTSSLYLQDLTAAASECIDPSVVFP-------------YPLNDSSSP   206
                                                               * . .::.

l-Myc      DGTPSLEASNPAPATQ--------CQLGEPKTQACSGSESPSDS---------EGEEIDV   172
n-Myc      AAVTSVSVPATAPVAAPARAGGRPASSGEAKALSTSGEDTLSDSDDEDDEEEDEEEEIDV   277
c-Myc      KSCTSSDSTAFSPSSDSLLSS--ESSPRASPEPLVLHEETPPTTSSDSEEEQEDEEEIDV   264
            . .* . .  :* :         ..   .       .:: . :         : *****

l-Myc      VTVEKRRSLDIRKPVTITVRADPLDPCMKHFHISIHQQQHNYAARFPPESCSQEGDPEPG   232
n-Myc      VTVEKRRSSSNNK--AVTTFTITVRPKTSALGLGRAQPGELILKRCVP-IHQQHNYAAPS   334
c-Myc      VSVEKRQTPAKRSESGSSPSRGHSKPPHSPLVLKRCHVSTHQHNYAAPPSTRKDYPAAKR   324
           *:****::   ..    :       *  . : :   :          *    :.  .

l-Myc      PQEEAPEIEAPKEKEEEEEEEEEEEIVSPPPVGSEAPQSCHPKPVSSDTEDVTKRKNHNF   292
n-Myc      PYVES-EDAPPQKKIKSEASPRPLKSVVPAKAKSLSPR-------NSDSEDSERRRNHNI   386
c-Myc      AKLDSGRVLKQISNNRKCSSPR-----------------------SSDTEENDKRRTHNV   361
           .  :: .     .: ..  . .                       .**:*:  :*:.**.

l-Myc      LERKRRNDLRSRFLALRDQVPTLASCSKAPKVVILSKALEYLQALVGAEKKMATEKRQLR   352
n-Myc      LERQRRNDLRSSFLTLRDHVPELVKNEKAAKVVILKKATEYVHALQANEHQLLLEKEKLQ   446
c-Myc      LERQRRNELKRSFFALRDQIPELENNEKAPKVVILKKATAYILSIQADEHKLTSEKDLLR   421
           ***:***:*:  *::***::* * . .**.*****.**  *: :: . *:::  **  *:

l-Myc      CRQQQLQKRIAYLSGY--   368
n-Myc      ARQQQLLKKIEHARTC--   462
c-Myc      KRREQLKHKLEQLRNSGA   439
            *::** :::
```

Fig. 1(b)

```
Klf2      ----------------MALSEPILPSFATFAS----------------------------   16
Klf4      -------MRQPPGESDMAVSDALLPSFSTFASGPAGREKTLRPAGAPTNRWREELSHMKR   53
Klf5      MPTRVLTMSARLGPLPQPPAAQDEPVFAQLKPVLGAANPARDAALFSGDDLKHAHHHPPA   60
                           . :    * *: : .

Klf2      --------------------------PCERGLQERWPRNEPEAGG-TDEDLNNVLDFILS   49
Klf4      LPPLPGRPYDLAATVATDLESGGAGAACSSNNPALLARRETEEFN-DLLDLDFILSNSLT   112
Klf5      PPPAAG--------------------PRLPSEELVQTRCEMEKYLTPQLPPVPIISEHKK   100
                                    .   .     .* * *           ::.   .

Klf2      MGLDGLGAENPPEPPPQPPPPAFYYPE-------------PGAPPPYSIPADSLGTELLR   96
Klf4      HQESVAATVTTSASASSSSSPASSGPASAPSTCSFSYPIRAGGDPGVAASNTGGLLYSR   172
Klf5      YRRDSASVVDQFFTDTEGIPYSINMNVFLPDITHLRTGLYKSQRPCVTQIKTEPVTIFSH   160
             .  ..     . ..  . :                   .  *  :           :

Klf2      PD---LDPPQGPALHGRFLLAPPGRLVKAEPPEVDGGGYGCAPGLAHGPRGLKLEGAPGA   153
Klf4      ES---APPPTAPFNLADINDVSPSGGFVAELLRPELDPVYIPPQQPQPPGGGLMGKFVLK   229
Klf5      QSESTAPPPPPAPTQALPEFTSIFSSHQTTAPPQEVNNIFIKQELPIPDLHLSVPSQQGH   220
           .     **  .   .    ..      :     : .        .       :

Klf2      TGACMRGPAGRPPPPPDTPPLSPDG---------------------PLRIPASGPRNPFP   192
Klf4      ASLTTPGSEYSSPSVISVSKGSPDGSHPVVVAPYSGGPPRMCPKIKQEAVPSCTVSRSLE   289
Klf5      LYQLLNTPDLDMPSSTNQTAVMDTLNVSMAGLNPHPSAVPQTSMKQFQGMPPCTYTMPSQ   280
                .    *.  . .                              :*..    .

Klf2      PPFGPGPSFGGPGPALHYGPPAPGAFGLFEDAAAAAAALGLAPPATRGLLTPP-------   245
Klf4      AHLSAGPQLSNGHRPNTHDFPLGRQLPTRTTPTLSPEELLNSRDCHPGLPLPPGFHPHPG   349
Klf5      FLPQQATYFPPSPPSSEPGSPDRQAEMLQNLTPPPSYAATIASKLAIHNPNLP-------   333
               .. :     .   . *          .. ..     :          *

Klf2      -------------------------SSPLELLEAKPKRGRRSWPRKRAATHTCSYTNCGK   280
Klf4      PNYPPFLPDQMQSQVPSLHYQELMPPGSCLPEEPKPKRGRRSWPRKRTATHTCDYAGCGK   409
Klf5      ----------------------ATLPVNSPTLPPVRYNRRSNPDLEKRRIHFCDYNGCTK   371
                                   .       .   . * .   ::   * *.* .* *

Klf2      TYTKSSHLKAHLRTHTGEKPYHCNWEGCGWKFARSDELTRHYRKHTGHRPFQCHLCDRAF   340
Klf4      TYTKSSHLKAHLRTHTGEKPYHCDWDGCGWKFARSDELTRHYRKHTGHRPFQCQKCDRAF   469
Klf5      VYTKSSHLKAHLRTHTGEKPYKCTWEGCDWRFARSDELTRHYRKHTGAKPFQCMVCQRSF   431
          .********************:* *:**.*:**************** :****  *:*:*

Klf2      SRSDHLALHMKRHM-   354
Klf4      SRSDHLALHMKRHF-   483
Klf5      SRSDHLALHMKRHQN   446
          *************
```

Fig. 1(c)

```
Dppa2   MSYFGLETFNENQSEENLDEESVILTLVPFKEEEEPNTDYATQSNVSSSTLDHTPPARSL     60
Dppa4   -----METAGDKKWSAEEPKEEVELQMSSQPSTAPAKAKATGKKQKKSETDNGCKPKEGK     55
Dppa3   -------------------MEEPSEKVDPMKDPETPQKKDEEDALDDTDVLQPETLVKVM     41
Dppa5   ------------------------MMVTLVTRKDIPPWVKVPEDLKDPEVFQVQS---LV     33
                                  :        .      .   .... :

Dppa2   VRHAGIKHPTR-TIPSTCPPPSLPPIRDVSRNTLREWCRYHNLSTDGKKVEVYLRLRRHS    119
Dppa4   PQDTETPGQTRRKVPIPPIPEYLPPVNLIHRDVLRAWCQKKRVSSKGQKLDAYKRLLARA    115
Dppa3   KKLTLNPGVKR----------------SARRRSLRNRIAAVPVEN---KSEKIRREVQSA     82
Dppa5   LKYLFGPQGSR--------------------------------------MSHIEQVSQAM     55
         :       .*                                       .   :

Dppa2   YSKQE---CYIPNTSREARMKQGPKKSKIVFRGIG-----------PPSGCQRKKEESGV    165
Dppa4   FPEQMLELRNVPDSAKDARLKTAHKKMKTEPGEESEVTVPLEMVPVPEEQIPALIDPPML    175
Dppa3   FPKRR-------------------------------------------------------     87
Dppa5   FELKN-------------------------------------------------------     60
        :  :

Dppa2   LEILTSPKESTFAAWARIAMRAAQSMSKNRCPLPSNVEAFLPQATGSRWCVVHGRQLPAD    225
Dppa4   YEEVSTTVVTTPATEAVLASWARIASNAKKYEAVPADASSSSEVKGEMWCVVHGTSLPGN    235
Dppa3   ---VRTLLSVLKDPIAKMRRLVRIEQRQKRLEGNEFERDSEP-----FRCLCT-------    132
Dppa5   ----------LESPEELIEVFIYGSQNNKIRAKWMLQSMAER------------------     92
                     .   :          :

Dppa2   KKGWVRLQFLAGQTWVPDTPQRMNFLFLLPACIIPEPGVEDNLLCPECVHSNKKILRNFK    285
Dppa4   SRGWVRLQFHAGQAWVPDKKGKAIALFLLPACTFPPPHLEDNMLCPKCVHKNKILTKSLE    295
Dppa3   --------FCHYQRWDPSENAKIGKN----------------------------------    150
Dppa5   --------YHLRQQKGVLKLEESMKTLELGQCIE--------------------------    118
                :   *        .

Dppa2   IRSRAKKNALPPNMPP   301
Dppa4   G---------------   296
Dppa3   ----------------
Dppa5   ----------------
```

Fig.1(d)

Suppl. Fig. 2 Sequence alignment between Klf2 and Klf4.

```
Klf2    ---------MALSEPILPSFATFASPCERGLQERWPRNEPE-------------------    32
Klf4    MRQPPGESDMAVSDALLPSFSTFASGPAGREKTLRPAGAPTNRWREELSHMKRLPPLPGR    60
                 **:*:.:****:****      :   * . *

Klf2    -----------------------------AGGTDEDLNNVLDFILSMG------------    51
Klf4    PYDLAATVATDLESGGAGAACSSNNPALLARRETEEFNDLLDLDFILSNSLTHQESVAAT   120
                                     *    *::*::**: : :.

Klf2    ----LDGLGAENPPEPPPQPPP-----------------------PAFYYPEPGAPPPYS    84
Klf4    VTTSASASSSSSPASSGPASAPSTCSFSYPIRAGGDPGVAASNTGGGLLYSRESAPPPTA   180
             .. .:..*... * ..*                        .: *.. .**** :

Klf2    ----------IPADSLGTELLRPDLD----PPQGP-----ALHGRFLLAP----PGR---   118
Klf4    PFNLADINDVSPSGGFVAELLRPELDPVYIPPQQPQPPGGGLMGKFVLKASLTTPGSEYS   240
                   *:..: :*****:**    *** *     .* *:*:* .    **

Klf2    ---LVKAEPPEVDGGGYGCAPGLAHGPRGLKLEGAPGATGACMRG---------------   160
Klf4    SPSVISVSKGSPDGSHPVVVAPYSGGPPRMCPKIKQEAVPSCTVSRSLEAHLSAGPQLSN   300
           ::...  . **.    ..  : **  :  :    *. :*  .

Klf2    ---------PAGRPPPPPDTPPLSPDGPLRIPASGPRNPFPPPFGPGPSFGGPGPALHYG   211
Klf4    GHRPNTHDFPLGRQLPTRTTPTLSPEELLNSRDCHPGLPLPPGFHPHP---GP----NYP   353
                 * **  *.  **.***:  *.   . *  *:** * * *   **    :*

Klf2    PPAPGAFGLFEDAAAAAAALGLAPPATRGLLTPPSSPLELLEAKPKRGRRSWPRKRAATH   271
Klf4    PFLP------------DQMQSQVPSLHYQELMPPGS-CLPEEPKPKRGRRSWPRKRTATH   400
        *  *                . .*.     * **.*     *.*************:***

Klf2    TCSYTNCGKTYTKSSHLKAHLRTHTGEKPYHCNWEGCGWKFARSDELTRHYRKHTGHRPF   331
Klf4    TCDYAGCGKTYTKSSHLKAHLRTHTGEKPYHCDWDGCGWKFARSDELTRHYRKHTGHRPF   460
        **.*:.**************************:*:*************************

Klf2    QCHLCDRAFSRSDHLALHMKRHM   354
Klf4    QCQKCDRAFSRSDHLALHMKRHF   483
        **: ******************:
```

Suppl. Fig. 3 Sequence alignment for subgroups of Dppa proteins. (a) Dppa2 and Dppa4; (b) Dppa3 and Dppa5.

(a)

```
Dppa2   MSYFGLETFNENQSEENLDEESVILTLVPFKEEEEPNTDYATQSNVSSSTLDHTPPARSL
Dppa4   -----METAGDKKWSAEEPKEEVELQMSSQPSTAPAKAKATGKKQKKSETDNGCKPKEGK
             :** .::: . :  :*.* * : .  .   .::. : :.: .*.* :   * ..

Dppa2   VRHAGIKHPTR-TIPSTCPPPSLPPIRDVSRNTLREWCRYHNLSTDGKKVEVYLRLRRHS
Dppa4   PQDTETPGQTRRKVPIPPIPEYLPPVNLIHRDVLRAWCQKKRVSSKGQKLDAYKRLLARA
         :.:     ** .:* .  *  ***:. : *:.** **: :.:*:.*:*::.* **  ::

Dppa2   YSKQE---CYIPNTSREARMKQGPKKSKIVFRGIG-----------PPSGCQRKKEESGV
Dppa4   FPEQMLELRNVPDSAKDARLKTAHKKMKTEPGEESEVTVPLEMVPVPEEQIPALIDPPML
        :.:*      :*:::::**:* . ** *      .           * .      : . :

Dppa2   LEILTSPKESTFAAWARIAMRAAQSMSKNRCPLPSNVEAFLPQATGSRWCVVHGRQLPAD
Dppa4   YEEVSTTVVTTPATEAVLASWARIASNAKKYEAVPADASSSSEVKGEMWCVVHGTSLPGN
         * :::.  :* *: * :*  *  : . ::    .   :  .:..*. ****** .**.:

Dppa2   KKGWVRLQFLAGQTWVPDTPQRMNFLFLLPACIIPEPGVEDNLLCPECVHSNKKILRNFK
Dppa4   SRGWVRLQFHAGQAWVPDKKGKAIALFLLPACTFPPPHLEDNMLCPKCVHKNKILTKSLE
        .:******* ***:****.  :   ******* :* * :***:***:***.** : :.::

Dppa2   IRSRAKKNALPPNMPP
Dppa4   G---------------
```

(b)

```
Dppa3   MEEPSEKVDPMKDPETPQKKDEEDALDDTDVLQPETLVKVMKKLTLNPGVKRSARRRSLR
Dppa5   -----MMVTLVTRKDIPPWVKVPEDLKDPEVFQVQSLV---LKYLFGP----QGSRMSHI
               *  :.  : *   .  : *.*.:*:* ::**    *  :.*    .. * *

Dppa3   NRIAAVPVENKSEKIRREVQSAFPKRRVRTLLSVLKDPIAKMRRLVRIEQRQKRLEGNEF
Dppa5   EQVSQAMFELKNLESPEELIEVFIYG-SQNNKIRAKWMLQSMAERYHLRQQKGVLKLEES
        :::: . .* *. :  .*: ..*     :.     *  : .* .  ::.*::  *: :*

Dppa3   ERDSEPFRCLCTFCHYQRWDPSENAKIGKN
Dppa5   MKTLELGQCIE-------------------
         :  *  :*:
```

Suppl. Fig. 4 Secondary structure prediction on Dppa5 KH (AA26-AA86) domain by Jpred

```
Dppa5-KH   VFQVQSLVLKYLFGPQGSRMSHIEQVSQAMFELKNLESPEELIEVFIYGSQNNKIRAKWML
Jpred      EEEHHHHHHHHH---------HHHHHHH-HEEEE-------EEEEEEE----HHHHHHHHH
```

Suppl. Fig. 5 Sequence alignment between Ecat1 and Dppa5

```
Ecat1   MASLKRFQTLVPLDHKQGTLFEIIGEPKLPKWFHVECLE-DPKRLYVEPRLLEIMFGKDG    59
Dppa5   ------------------MMVTLVTRKDIPPWVKVPEDLKDPEVFQVQSLVLKYLFGPQG
                           :. :: . .:* *.:*     **: : *:. :*: :** :*

Ecat1   EHIPHLESMLHTLIHVN-VWGPERRAEIWIFGPPPFRRDVDRMLTDLAHYCRMKLMEIEA   118
Dppa5   SRMSHIEQVSQAMFELKNLESPEELIEVFIYGSQ-------------NNKIRAKWMLQSM
        .::.*:*.: ::::.:: : .**.  *::*:*.               :  * * *  .

Ecat1   LEAGVERRRMAAHKAATQPAPVKVREAAPRPASVKVPETATQPAPVKVREAAPQPAPVQE   178
Dppa5   AERYHLRQ---------------------------------QKGVLKLEESMKTLELGQC
         *    *:                                 * . :*:.*:       *

Ecat1   VREAAPQQASVQEEVREAATEQAPVQEVREAATEQAPVQEVSEAATEQAPVQEVNEAATE   238
Dppa5   IE----------------------------------------------------------
        :.

Ecat1   QASVQAVREAATRPAPGKVRKAATQPAPVQVCQEATQLAPVKVREAATQPASGKVREAAT   298
Dppa5   ------------------------------------------------------------

Ecat1   QLAPVKVRKAATQLAPVKVHEAATQPAPGKVSDAATQSASVQVREAATQLSPVEATDTSQ   358
Dppa5   ------------------------------------------------------------

Ecat1   LAQVKADEAFAQHTSGEAHQVANGQSPIEVCETATGQHSLDVSRALSQKCPEVFEWETQS   418
Dppa5   ------------------------------------------------------------

Ecat1   CLDGSYVIVQPPRDAWESFIIL   440
Dppa5   ----------------------
```

Suppl. Fig. 6 Sequence alignment between Klf4 and Nanog.

```
Klf4    MRQPPGESDMAVSDALLPSFSTFASGPAGREKTLRPAGAPTNRWREELSHMKRLPPLPGR    60
Nanog   -------------------MSVGLPGPHSLPSSEEASNSGNASSMPAVFHPENYSCLQGS    41
                           :*.  .** .  .: ..:.: .      : * :. . * *

Klf4    PYDLAATVATDLESGGAGAACSSNNPALLARRETEEFNDLLDLDFILSNSLTHQESVAAT   120
Nanog   ATEMLCTEAASPRPSSEDLPLQGSPDSSTSPKQKLSSPEADKGPEEEENKVLARKQKMRT   101
        . :: .* *:. .... . . ...  :  : ::. .  :  .     .*.:  ::.   *

Klf4    VTTSASASSSSSPASSGPASAPSTCSFSYPIRAGGDPGVAASNTGGGLLYSRESAPPPTA   180
Nanog   VFS-----------------QAQLCALKDRFQKQK-------------YLSLQQMQELSS   131
        * :                  .. *::.  ::                  * :.    ::

Klf4    PFNLADINDVSPSGGFVAELLRPELDPVYIPPQQPQPPGGGLMGKFVLKASLTTPGSEYS   240
Nanog   ILNLSYKQVKTWFQNQRMKCKRWQKN-------QWLKTSNGLIQKGSAPVEYPSIHCSYP   184
         :**:  :  :   .   :  * : :       *   ...**: *    .. .:  ..*.

Klf4    SPSVISVSKGSPDGSHPVVVAPYSGGPPRMCPKIKQEAVPSCTVSRSLEAHLSAGPQLSN   300
Nanog   QGYLVNASGSLSMWGSQTWTNPTWSSQTWTNPTWNNQTWTNPTWSSQAWTAQSWNGQPWN   244
        .  ::..* . .  .  . . *  .. .   *. :::: .. * * .  :  * . *  *

Klf4    GHRPNTHDFPLGRQLPTRTTPTLSPEELLNSRDCHPGLPLPPGFHPHPGPNYPPFLPDQM   360
Nanog   -------AAPLHNFGEDFLQPYVQLQQNFSASDLEVNLEATRESHAHFS----------T   287
                  ** .       * :. :: :.: * . .*  .   *.* .

Klf4    QSQVPSLHYQELMPPGSCLPEEPKPKRGRRSWPRKRTATHTCDYAGCGKTYTKSSHLKAH   420
Nanog   PQALELFLNYSVTPPGEI------------------------------------------   305
         . :  :   .: ***.

Klf4    LRTHTGEKPYHCDWDGCGWKFARSDELTRHYRKHTGHRPFQCQKCDRAFSRSDHLALHMK   480
Nanog   ------------------------------------------------------------


Klf4    RHF   483
Nanog   ---
```

Suppl. Fig. 7 Matched segment after sequence alignment between: (a) Klf4 and Nanog; (b) Lin28 and c-Myc; (c) Esrrb and Klf4; (d) Esrrb and c-Myc. In all the cases when the original names were followed by "r", the sequences were reversed.

(a)

```
Klf4    303   RPNTHDFPLGRQLPTRTTPTLSPEELLNS   331
Nanog    54   RPSSEDLPL-QGSPDSST---SPKQKLSS    78
              **...*:** .  *  .*   **.. *.*

Klf4    136   SGPASAPST-CSF--SYPIRAGGDPGVAASNT   164
Nanog   171   SAPVEYPSIHCSYPQGYLVNASGSLSMWGSQT   202
              *.*.  **  **:  .* : *.*. .: .*.*


Klf4     273   KIKQEAVPSCTVSRSLE--AHLSAGPQLS   299
Nanogr   150   KMRQNQFWTKVQKYSLNLISSLEQMQQLS   122
               *::*.   .   . **.  . *    ***

Klf4     329   LLNSRDCHPG   337
Nanogr    93   LVKNEEEEPG    84
               *:...: .**

Klf4     335   HPGLPLPPGFHPHPGPNYP   353
Nanogr   259   YPQL-FDEGFNHLPAANWP   242
               .* * :  **.  *. *:*

Klf4     177   PPTAPFNLADIN   188
Nanogr   302   PPTVSYNLFLEL   291
               ***. :**
```

(b)

```
Lin28     57   SMTARAGVALDPPVDVF    73
c-Myc    181   DLTAAASECIDPSV-VF   196
               .:** *.  :** * **

c-Myc    294   PLVLKRCHV   302
Lin28r   205   PL-LAPSHI   218
               ** * *.*:

c-Myc    323   KRAKLDSGRVLKQISNNRKCSSPR   346
Lin28r   180   QQAKLPCSAVMHNISQCFHCKKPQ   157
               ..***  . *::.**.  :*..*.

c-Myc    224   LLSSESSPRAS   234
Lin28r    38   LLQPEDAARAA    28
               **. *.. **.

c-Myc    354   DKRRTHNVLERQ---RRNELK-RSFFALRDQIPELENNEKA   390
Lin28r   100   SKKFTFEVAEGEKLSRFGEMHLKSQHVFVDVPPDLAVGARA    60
               .*: * .* * .   * .*:: :* ..: *  *:*. . :*
```

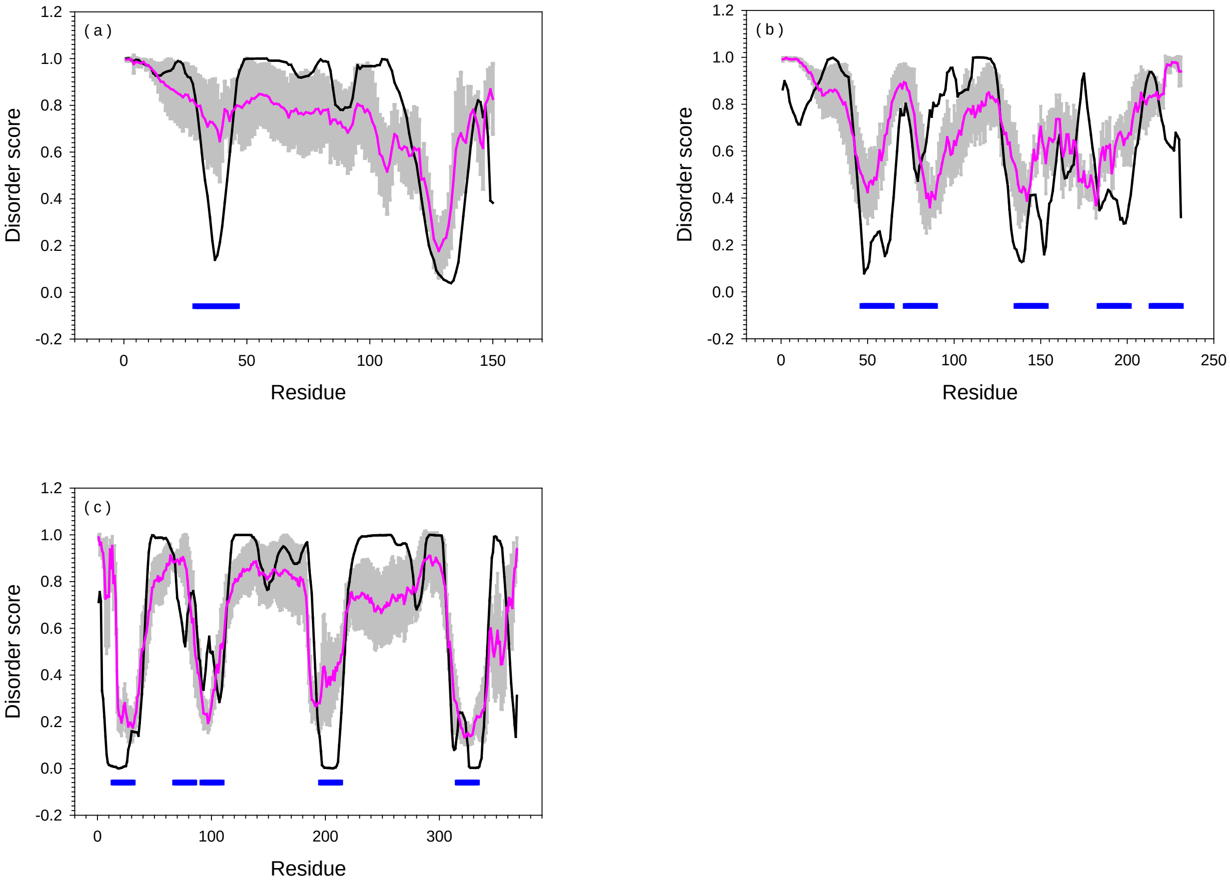
(c)

```
Esrrb   212   IVSNLLGVEQDKLY   225
Klf4    196   FVAELLRPELDPVY   209
              :*..**  * * **

Esrrb   262   PGFPSLTLGDQM   273
Klf4    250   PNYPPF-LPDQM   360
              *.:* : * ***

Esrrb   289   VYRSLPYDDK   298
Klf4r   315   LQRGLPFDHT   306
              : *.**:*.

Esrrb    21   SSPS-SGIDALS    31
Klf4r   194   GSPSVDNIDALN   183
              .*** ..****.

Esrrb   231   DIPEGDIKALTTLCELADREL   249
Klf4r    63   DYPRGPLPPLRKMHSLEERWR    43
              *:*.* :  *  .  *.:*

Esrrb   192   ENSPYLNLPISPPAKKPLTKIVSNLLG   218
Klf4r   189   DNIDALNFPATPP---PASE-RSYLLG   167
              :*   **:*  **   * ..  * ***
```

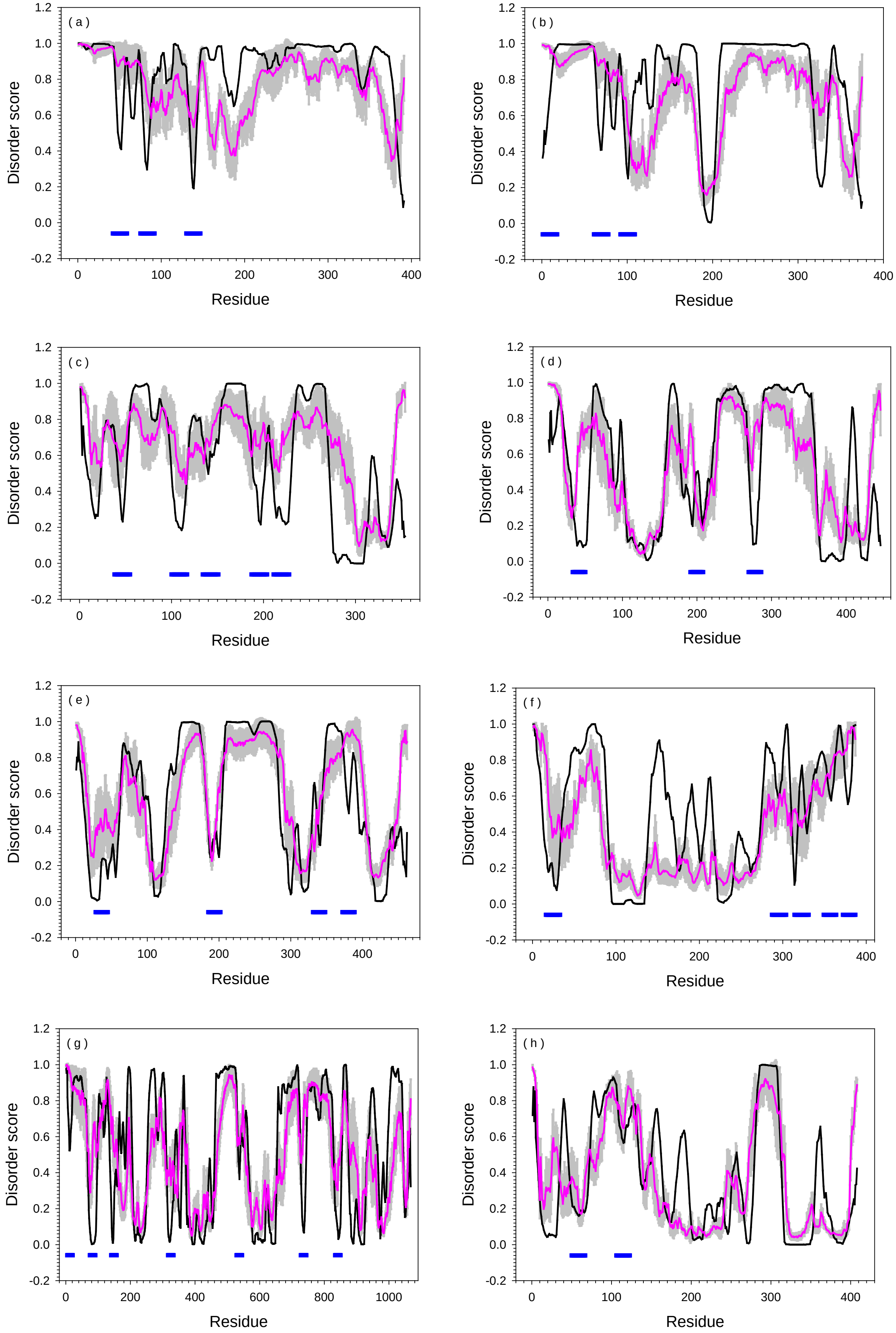
Suppl. Fig. 7 (d)

```
Esrrb    266   SLTLGDQMSLLQSAWMEILILGIVYRSLPYD-DKLAYAEDYIMDE   309
c-Myc     93   NFSTADQLEMMT----ELLGGDMVNQSFICDPDDETFIKNIIIQD   133
                :. .**: ::.    ***  .:*  *:  * *: .: ..:*: :

Esrrb    345   KALALANSDSMYIENLEA   362
c-Myc    167   RGHSVCSTSSLYLQDLTA   184
               :. .:....*:**:.* *

Esrrb    391   KLLWTLPLLRQTAAKAVQ   408
c-Myc    261   EIDVVSVEKRQTPAKRSE   278
               .: :.    *** **  .


Esrrb     15   FIKTEPSS-PSSGIDALSHHSPS-GSSDASG    43
c-Mycr   240   LVLPEPSARPSSESSLLSDSSPSFATSDSST   210
               :: .***. ***. . **. *** ..**.*.

Esrrb    365   KLQDLLHEA------LQDYELSQRHEEPR------RAGKL   392
c-Mycr   435   RLQELKHKLQERRKRLLDKESTLKHEDAQISLIYATAKKL   396
               :**:* *.       * * * . :**:         * **

Esrrb    407   VQH---FYSVKLQGK   418
c-Mycr   381   IQDRLAFFSRKLENR   367
               :*.   *:* **..:

Esrrb    123   CKAFFKRTIQGN   134
c-Mycr   378   RLAFFSRKLENR   367
                 *** * *..

Esrrb    270   GDQMSLLQSAWMEILILGIVYRS---LPYDDKLAY   301
c-Mycr   409   DAQISLIYATAKKLIVVKPAKENNELEPIQDRLAF   375
                 *:**: ..  .::::  . :     *:.*:**:

Esrrb     55   GLDSPPMFAGAGLGGNPCRK    74
c-Mycr   224   LSDSSPSFATSD--SSTCSK   207
                 ** * **...  .. * *
```

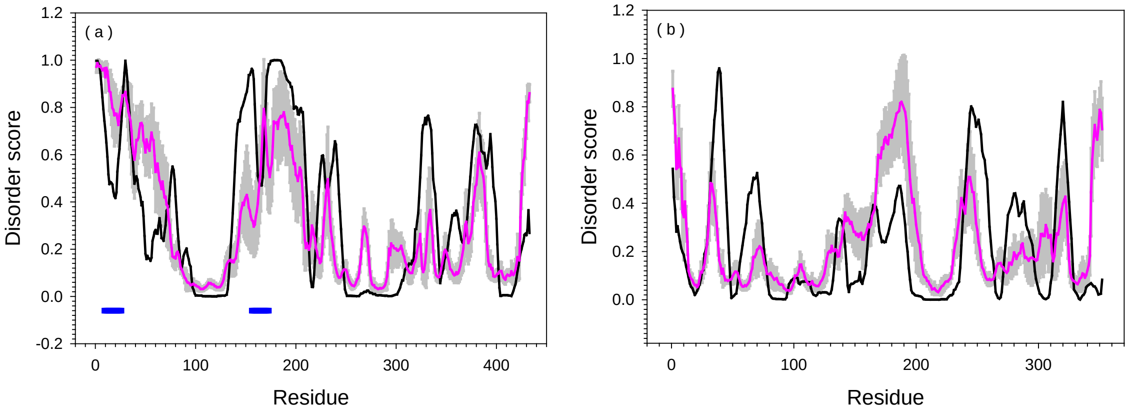
Suppl. Fig. 8 Disorder prediction for proteins in quadrant IV of CH-CDF plot: (a) Dppa3; (b) Sox15; and (c) l-Myc.



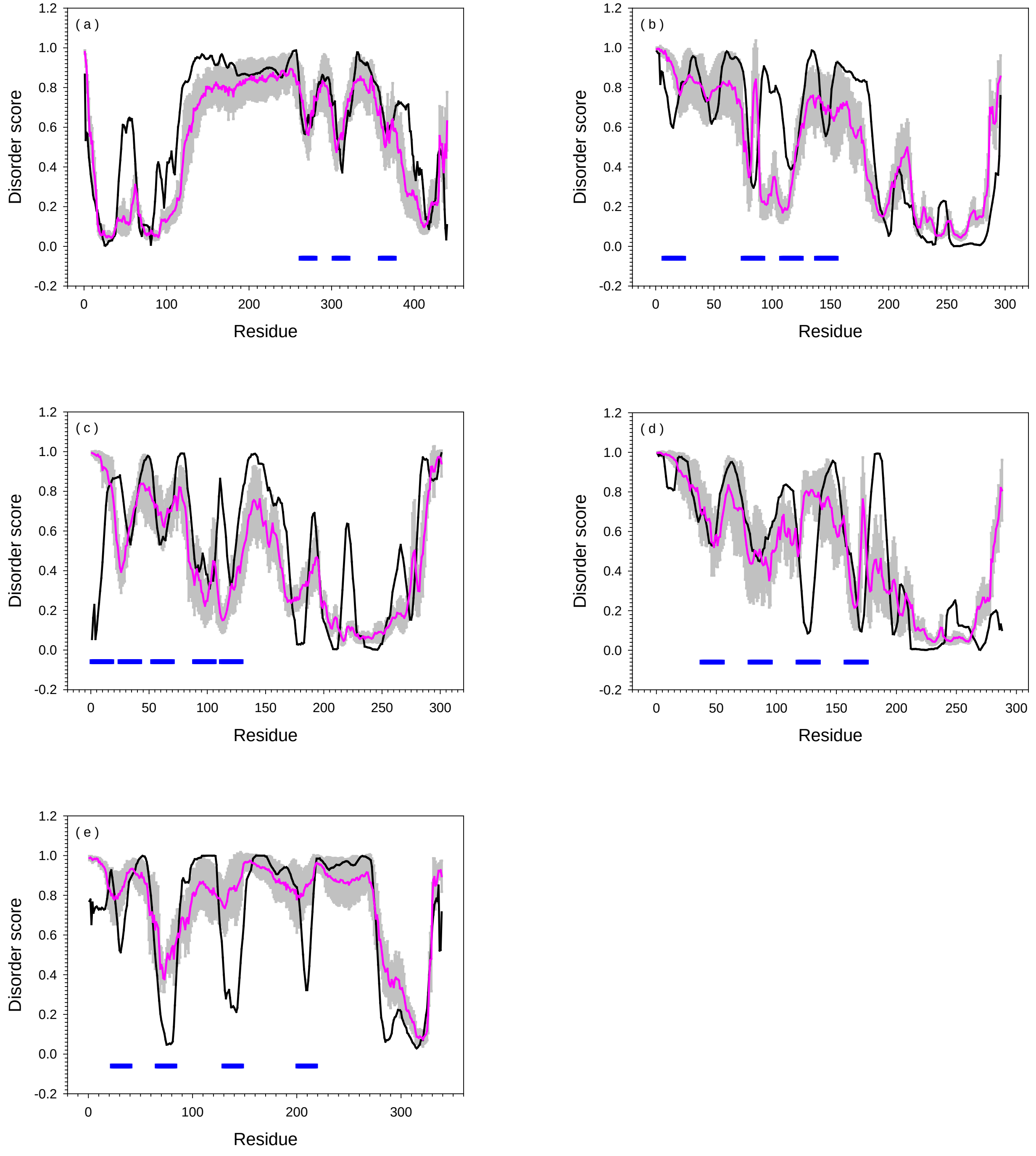
Suppl. Fig. 9 Disorder prediction for effecter proteins in Quadrant III of CH-CDF plot. (a) Sox1; (b) Sox3; (c) Klf2; (d) Klf5; (e)n-Myc; (f) P53; (g)Sall4; (h) BMP4



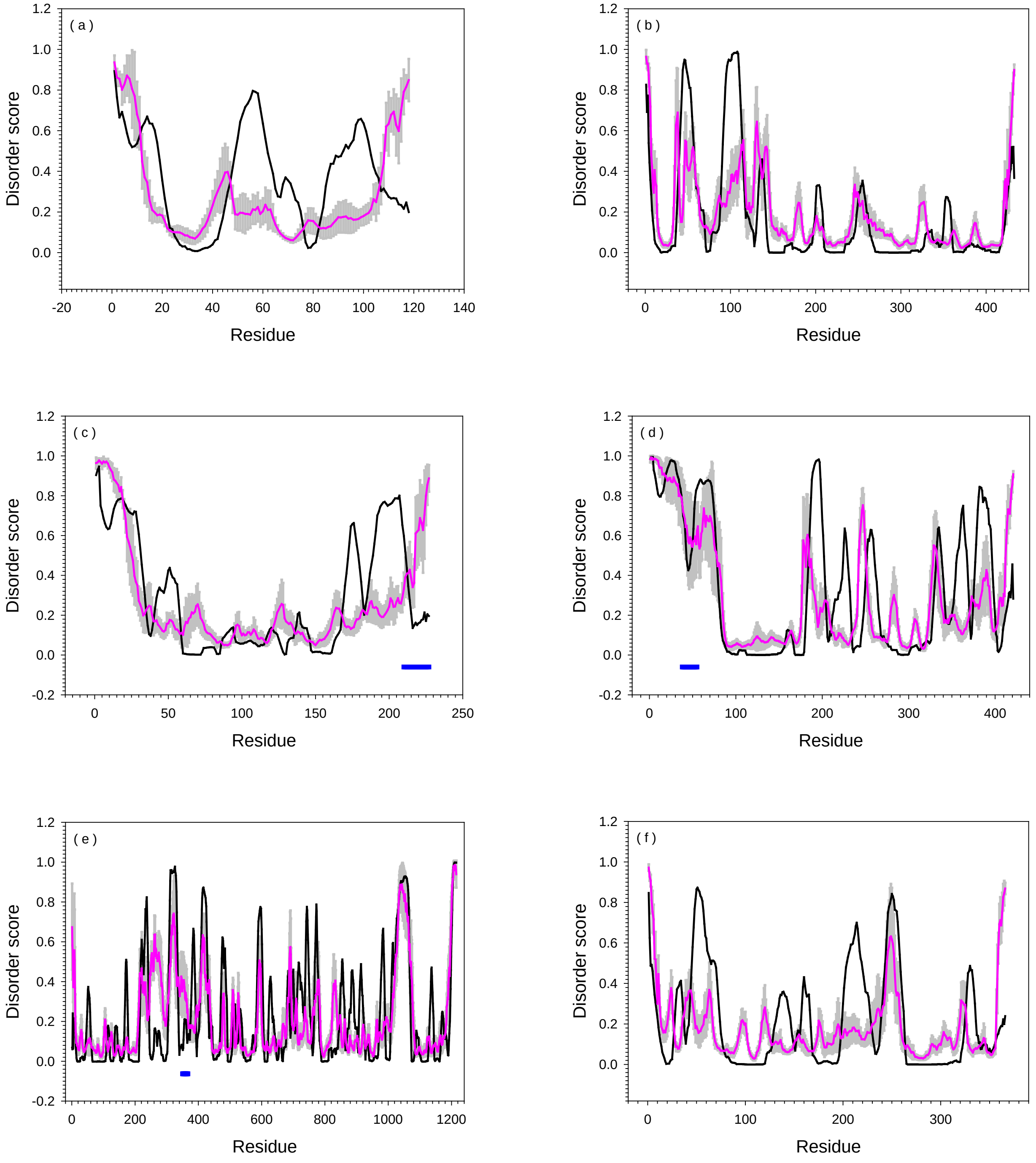
Suppl. Fig. 10 Disorder prediction for effector proteins in quadrant II of CH-CDF plot. (a) Esrrb and (b) Wnt3a. (a) AA31-AA46: LCR; AA100-AA171: Zinc finger; AA245-AA403: HOLI. (b) AA44-AA352: WNT1.

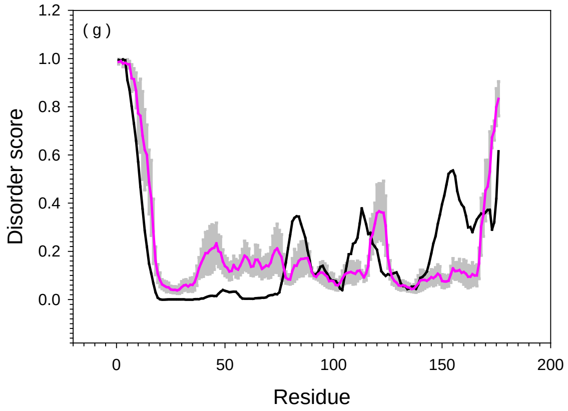


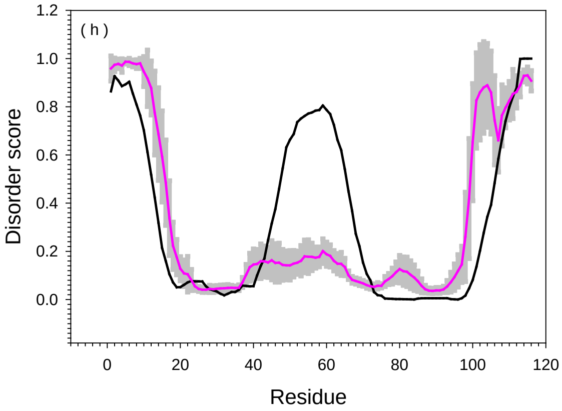
Suppl. Fig. 11 Disorder prediction of Bio-marker proteins in quadrant III of CH-CDF plot. (a) Ecat1; (b)Dppa4; (c) Dppa2; (d) Rex1; and (e) Utf1.

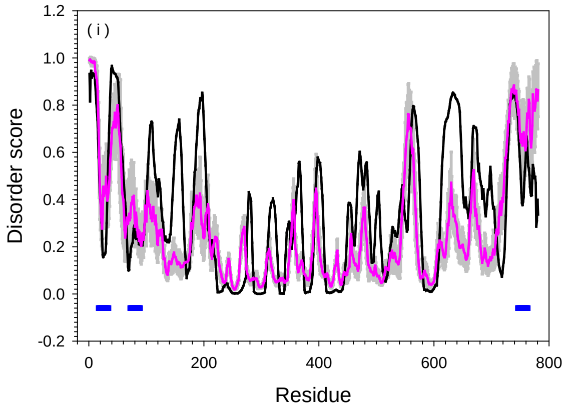


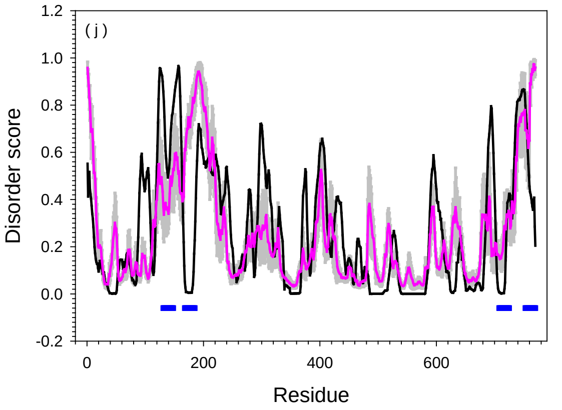
Suppl. Fig. 12 Disorder prediction of Bio-marker proteins in quadrant II of CH-CDF plot. (a) Dppa5; (b) Fbxo15; (c) Eras; (d) Dnmt31; (e) Ecat8; (f) Gdf3; (g) Fthl17; (h) Tcl1; (i) β-catenin; (j) Stat3; and (k) Grb2.

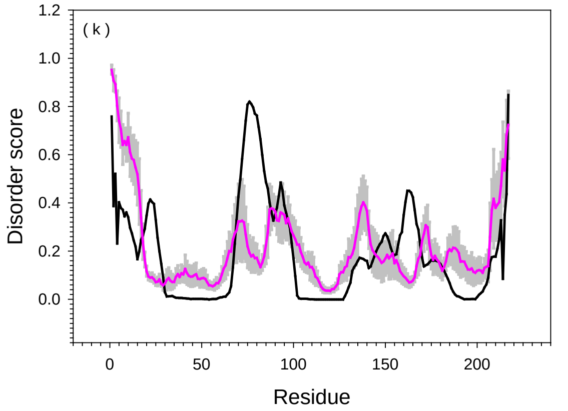



(g)
Disorder score
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
Residue


(h)
Disorder score
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
120
Residue


(i)
Disorder score
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
200
400
600
800
Residue


(j)
Disorder score
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
200
400
600
Residue


(k)
Disorder score
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
Residue