DUX4 regulates oocyte to embryo transition in human

Sanna Vuoristo[1,2#*], Christel Hydén-Granskog[3], Masahito Yoshihara[1], Lisa Gawriyski[4], Anastassius Damdimopoulos[5], Shruti Bhagat[1,6], Kosuke Hashimoto[6], Kaarel Krjutškov[1,7,8], Sini Ezer[8,9], Priit Paluoja[7], Karolina Lundin[2], Pauliina Paloviita[2], Gaëlle Recher[10], Vipin Ranga[11], Tomi Airenne[11], Mahlet Tamirat[11], Eeva-Mari Jouhilahti[8], Timo Otonkoski[8], Juha S. Tapanainen[2,3,12], Hideya Kawaji[6,13,14], Yasuhiro Murakawa[6,15], Thomas R. Bürglin[16], Markku Varjosalo[4], Mark S. Johnson[11], Timo Tuuri[2, 3], Shintaro Katayama[1,9*] and Juha Kere[1,8,9 *]

[1]Karolinska Institutet, Department of Biosciences and Nutrition, Huddinge, Sweden

[2]University of Helsinki, Department of Obstetrics and Gynecology, Helsinki, Finland

[3]Helsinki University Hospital, Department of Obstetrics and Gynecology, Helsinki, Finland

[4]University of Helsinki, Institute of Biotechnology, Helsinki, Finland

[5]Karolinska Institutet, Bioinformatics and Expression Analysis Core Facility, Huddinge, Sweden

[6]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

[7]Competence Centre for Health Technologies, Tartu, Estonia

[8] Stem Cells and Metabolism Research Program, University of Helsinki, Finland

[9]Folkhälsan Research Center, Helsinki, Finland

22 [10]Institut d'Optique Graduate School, CNRS - Université de Bordeaux, Talence,
23 France

24 [11]Structural Bioinformatics Laboratory, Biochemistry, Faculty of Science and
25 Engineering, Åbo Akademi University, Turku, Finland

26 [12]Oulu University Hospital, Oulu, Finland

27 [13]RIKEN Preventive Medicine and Diagnosis Innovation Program, Wako, Japan

28 [14]Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan

29 [15]IFOM, The FIRC Institute of Molecular Oncology, Milan, Italy

30 [16]University of Basel, Department of Biomedicine, Basel, Switzerland

31

32

33 [#]Current affiliation

34 [*]Corresponding author

35

36

37

38

39

Abstract

In oocyte to embryo transition, the fertilized oocyte undergoes final maturation and the embryo genome is gradually activated during the first three cell divisions. How this transition is coordinated and which factors drive the processes in humans is largely unknown. Here we studied the role of the double homeodomain transcription factor DUX4 in regulating the human oocyte to embryo transition. DUX4 knockdown zygotes show delayed transcriptome reprogramming during the first three days after fertilization. Our combined experimental approaches allowed integrated analysis on the transcriptome, chromatin, and proteome data in human embryos or a DUX4 expressing human embryonic stem cell model. We conclude that DUX4 is a pioneering factor that regulates human oocyte to embryo transition through regulating oocyte mRNA degradation, as well as direct binding and activation of minor genome activation genes, and genomic repeat elements.

63 Mammalian pre-implantation development involves a series of coordinated processes,
64 starting with oocyte to embryo transition (OET). OET is a major developmental
65 reprogramming event from the oocyte to a totipotent embryo, involving fundamental
66 changes in the epigenetic landscapes, degradation of maternal mRNAs, and
67 embryonic genome activation (EGA)[1]. In humans, the major EGA takes place by the
68 8-cell stage[2-4]. Minor EGA genes are upregulated in the human 4-cell embryos where
69 they subsequently induce genes upregulated at the major EGA[3]. Until now, most
70 studies focusing on human EGA have concentrated on the genes that are expressed in
71 the cleavage stage (2-cell, 4-cell, and 8-cell) embryos[3,5] and set the stage for the
72 forthcoming lineage commitment[6,7]. How OET and EGA are orchestrated in the
73 human embryos and which genes act as pioneers remain poorly understood.

74

75 The conserved double homeodomain transcription factor DUX4 is expressed in early
76 human embryos[8-10]. It represents a plausible candidate to regulate the OET in humans,
77 given its capacity to activate EGA-related genes and the genomic repeat elements[8,9].
78 In this work we have used a combination of methods to investigate DUX4 in the
79 course of early human development. Our data highlight DUX4 mRNA upregulation
80 already in the zygotes followed by down-regulation within the next cell division.
81 Abundant cytoplasmic and nuclear DUX4 protein was apparent only during the first
82 two days of development. Our in-depth characterisation of DUX4 in the human
83 embryos suggested that it is not required for survival of the human embryos during
84 the first three days of development but that DUX4 regulates OET in the human.
85 siRNA silencing of DUX4 in zygotes delayed oocyte mRNA degradation. Our
86 comprehensive analysis of transcriptome and chromatin data in human embryos or
87 DUX4 expressing cell lines suggested that DUX4 opens up chromatin through direct

88 activation of the retroelements and unannotated genomic regions. Our protein
89 interaction data demonstrated that DUX4 interacts with transcriptional mediators,
90 chromatin modifiers, and ubiquitinases expressed in the human oocytes and early
91 embryos. We conclude that DUX4 regulates different aspects of the OET in humans
92 by affecting mRNA degradation, transcriptional regulation and chromatin structure.
93
94
95
96
97
98
99
100
101

102 **Results**

103 *DUX4 is transiently expressed immediately after fertilization*

104 We first measured DUX4 mRNA levels in human MII oocytes (N=20), zygotes
105 (N=59), and cleavage embryos (2-cell, N=4; 4-cell, N=14; 8-cell, N=15). We found
106 significant DUX4 upregulation in the zygotes, while few transcripts were found in the
107 MII oocytes or the cleavage embryos[10-13] (Fig. 1a). The expression of the DUX is
108 evolutionary conserved as shown by stage-specific expression in mouse, bovine, and
109 non-human primates[14]. We stained embryos with a monoclonal antibody targeting the
110 DUX4 protein and detected cytoplasmic DUX4 in the zygotes and all cleavage
111 embryos although less in 8-cell embryos (Fig. 1b). DUX4 protein was abundantly
112 present in the nuclei of the 2- and 4-cell embryos whereas nuclei of the 8-cell
113 embryos were mostly negative (Fig. 1b). In a single early 8-cell stage embryo there
114 was high variability in the nuclear DUX4 staining, consistent with a snapshot of on-
115 going degradation. We quantified nuclear DUX4 intensities in 3D and normalized
116 them to the cytoplasmic DUX4 intensities and found variable but increasing nuclear
117 signal from the zygotes up to the 4-cell embryos, while nuclear DUX4 was low or
118 absent in the 8-cell embryos (Fig 1c). These results demonstrated that DUX4
119 transcription takes place immediately after fertilisation and is followed by
120 cytoplasmic and nuclear localisation of the DUX4 protein during the first two days of
121 human embryo development.

122

123 *DUX4 protein may form a relatively stable structure even when not bound to DNA*

124 Given predominant DUX4 protein presence in the cytoplasm of the embryos as well
125 as stage-specific nuclear localization, we analysed the structural features of DUX4

126 (Fig. 2). Human DUX4 comprises two homeodomains (HD1: residues 19-81; and
127 HD2: 94-153; UniProt numbering), an intrinsically disordered region (IDR: 154-370)
128 and a C-terminal domain (residues 371-424; Fig. 2a). A stretch of hydrophobic
129 residues (308-323) within IDR is conserved in primates and could have a regulatory
130 role by interacting with other transcription co-regulators or masking the C-terminal
131 domain from solvent when no interacting partners are present. The C-terminal domain
132 is predicted to have two structurally ordered and evolutionarily conserved regions:
133 residues 371-387 and residues 414-423 (Fig. 2a). As a secondary structure the C-
134 terminal domain is predicted to contain three alpha helices and may form a stable fold
135 similar to that revealed by NMR for the Pax8 paired box domain (PDB: 2K27;[15]). We
136 found a nine amino acid transactivation domain (9aaTAD) located at the C-terminal
137 domain (371-379), also recently reported by Mitsuhashi et al.[16]. This motif is also
138 present in the PRD-like homeoprotein LEUTX[17] and might be recognized by other
139 proteins involved in the regulation of transcription, similar to the 9aaTAD motif of
140 the MLL that interacts with the cAMP-response element binding protein, a
141 transcriptional co-activator[18].

142

143 In the crystal structure of DUX4 (PDB: 6E8C;[19]) HD1 and HD2 bind DNA in a
144 symmetric manner (Fig. 2b). Residue R21, located at the N-terminal loop of HD1,
145 interacts with the residues I121, E135 and Q139 of HD2 and, equivalently, R96
146 located at the N-terminal arm of the HD2 interacts with the residues I46, E60 and Q64
147 of HD1 (Fig. 2b, c, d). Moreover, the main-chain carbonyl groups of G19 of HD1 and
148 G94 of HD2 respectively form a hydrogen bond with the main-chain nitrogen atom of
149 I121 of HD2 and I46 of HD1. We next asked whether these residues are conserved
150 within the DUX family of homeoproteins and within primates (Fig. 2e): based on our

151 sequence alignment, residues equivalent to G19, R21, I46, E60, Q64, G94, R96, I121,
152 E135 and Q139 of DUX4 are highly conserved. Interestingly, the residues G19, R21,
153 I46, E60, Q64, G94, R96, I121, E135 and Q139 of DUX4 (Fig. 2c,d) are not directly
154 involved in DNA binding, which prompted us to speculate that these residues could
155 be important for locking HD1 and HD2 together as a unit before DNA binding, a
156 hypothesis, which we further tested using molecular dynamics (MD) simulations.

157

158 Based on MD simulations and the RMSD among snapshots made every 10 ps (Fig.
159 2f), the DUX4 HD2-DNA complex appeared the most stable complex (average
160 RMSD over backbone atoms of 1.73 Å), followed by DUX4-DNA (1.77 Å) and
161 DUX4 HD1-DNA complex (2.3 Å). The DNA-free DUX4 HD1-HD2 structure (4.0
162 Å) was the least stabile yet the interactions between HD1 and HD2 were maintained
163 over the 100 ns simulation. Indeed, ionic interactions between R96 of HD2 and E50
164 and E60 of HD1 seem to be fundamental for the stability of the double HD structure
165 of DUX4: the electrostatic interactions/hydrogen bonds between R96 and E60 were
166 present during 92% of the simulation time. Additional stabilizing interactions between
167 the two HDs also take place between R21 (HD1) and E135 (HD2), and R22 (HD1)
168 and E125 (HD2). While these charged interactions hold the two HDs together, the
169 intermediate linker loop imparts flexibility, which could be vital to accommodate
170 DNA once DUX4 enters the nucleus and locates its binding motif. Even with bound
171 DNA, the linker loop fluctuates more relative to HD1 and HD2, as observed from the
172 RMSF values for the CA atoms of residues of the DUX4-DNA structure (Fig. 2g).

173

174 *DUX4 protein interacts with transcriptional mediators and chromatin modifiers*

175 Abundant nuclear and cytoplasmic DUX4 in the human embryos and the modelled
176 stable structure of the DUX4 when not bound to DNA suggested that DUX4 might
177 have functions beyond DNA binding. To study this, we applied a recently published
178 MAC-tag method to identify DUX4 protein-to-protein interactions[20,21]. We identified
179 altogether 158 BioID and 43 AP-MS high-confidence DUX4 interactions, out of
180 which 19 appeared in both datasets (FDR < 0.05, corresponding to > 0.73 SAINT
181 Score). Single BioID interactions and AP-MS interactions together with the
182 interactions that appeared in both data sets, based on the scored frequency of
183 interaction with DUX4, are shown in Fig 3a. We concentrated on the DUX4
184 interacting proteins that scored above the median value (Fig 3b). Overrepresentation
185 Enrichment Analysis (ORA) of protein pathway markers (Reactome, KEGG) showed
186 significant enrichment ($p < 0.05$, FDR < 0.01) of markers linked to generic
187 transcription and 'RNA Polymerase II Transcription', 'Chromatin organization' and
188 'Chromatin modifying enzymes'. Comparing our list of genes to protein complex
189 databases such as ComplexPortal and Corum using Fisher's Exact Test yielded
190 significant overrepresentation of several variants of the SWI/SNF ATP-dependent
191 chromatin remodelling complex, Core mediator complex, NSL histone
192 acetyltransferase complex, SRCAP histone exchanging complex and the NuA4
193 histone acetyltransferase complex ($p < 0.05$, FDR 0.01) (Fig. 3b). There were 28
194 DUX4 interacting proteins classified as RNA binding (GO:0003723) and 19 out of
195 these were linked to spliceosome and pre-mRNA-splicing. In the protein complex
196 database Corum, DUX4 interactors were significantly overrepresented in the
197 Spliceosome, with 10 interactors comprising 7% of the whole complex. In addition,
198 we found six DUX4 interacting proteins (ZSCAN4, ZSCAN5C, ZSCAN5D, RFPLA,

199 RFPLB, RNF8, PTOV1, and MB3LB) that have not appeared in the analyses of other
200 transcription factors. As the protein interaction assay was run in a non-embryonic cell
201 line (HEK), we next studied which of the identified DUX4 interacting proteins are
202 expressed by human oocytes or embryos[3,11]. Importantly, the vast majority of the
203 genes coding for the DUX4 interacting proteins were expressed in oocytes (maternal
204 genes), embryos, or both. These results suggested that DUX4 could potentially
205 regulate maternal and embryonic proteins in the cytoplasm and nucleus during the
206 OET.

207

208 *Knock-down of DUX4 in human zygotes leads to dysregulation of the maternal*
209 *transcriptome*

210 We next asked how DUX4 regulates the OET and early human embryo development.
211 We microinjected either DUX4 targeting siRNA (siDUX4) or control siRNA
212 (siCTRL) together with GAP-GFP mRNA to triploid human zygotes and followed
213 their development for 48 h after the microinjections, until the third day of
214 development (Fig 4a). 18 h after microinjection, GAP-GFP protein was expressed in
215 all embryos, confirming successful microinjections. Staining for the DUX4 protein
216 was very faint or absent in the siDUX4 embryos but strongly positive in the siCTRL
217 embryos 24 h after microinjection, showing that the DUX4 targeting siRNA
218 efficiently down-regulated DUX4 (Fig 4b). The cells from the embryos were
219 collected 48 h after microinjections for single-cell-tagged reverse transcription RNA
220 sequencing (STRT RNA-seq[22,23]), which detects the transcript far 5'-ends (TFEs).
221 siDUX4 embryos did not arrest during the experiment by live imaging follow-up, but
222 their transcriptome was dysregulated in comparison to that of the siCTRL embryos. A

223 number of transcripts downregulated in the siCTRL appeared enriched in the siDUX4
224 embryos (Fig. 4c). In order to study how the siDUX4 enriched transcripts typically
225 behave during the first three days of development, we next compared siDUX4 and
226 siCTRL embryos to our published gene expression data set[3,10] on human MII oocytes,
227 zygotes, and cleavage cells. These analyses further confirmed that a large number of
228 TFEs that remained enriched in the siDUX4 embryos were typically downregulated
229 between the oocytes and the 4-cell embryos (Fig 4d), the zygotes and the 4-cell
230 embryos (Fig 4e), and the 4-cell and 8-cell embryos (Fig 4f), indicating delayed
231 degradation of the maternal transcripts. Gene expression enrichment analysis using
232 TopAnat[24] for the altogether 91 genes enriched in siDUX4 embryos resulted in terms
233 such as 'female germ cell' and 'oocyte', in agreement with non-degraded maternal
234 transcripts. As shown for the siDUX4 enriched gene set, a number of well-known
235 maternal genes such as *GDF9*, *ZP1*, *ZP2*, *ZP3*, *KHDC3L*, *WEE2*, *NPM3*, *TUBB8*, and
236 *RERE* failed to downregulate (Fig. 4g), demonstrating that the OET remained
237 incomplete after DUX4 abolishment in human zygotes.

238

239 *DUX4 directly activates the minor EGA genes and a number of unannotated regions*

240 We next generated two TetOn-DUX4 human embryonic stem cell (hESC) lines
241 expressing DUX4-ires-EmGFP under doxycycline, and studied the effects of DUX4
242 on the activity and accessibility of the genome (Fig 5a). *DUX4* mRNA (Fig 5b) and
243 protein (Fig 5c) expression promptly followed doxycycline induction and the known
244 DUX4 target genes *ZSCAN4* and *TRIM48* followed the induction with slight delay
245 (Fig 5b). We performed bulk RNA-seq using STRT RNA-seq methods and the data
246 analysis of the sorted EmGFP+ DUX4 expressing cells showed roughly equal

247 numbers of up- and downregulated TFEs (Fig 5d). Notably, the majority of the
248 upregulated TFEs were mapped to unannotated genomic regions, whereas the
249 downregulated TFEs were mapped to protein coding regions (Fig 5d). The known
250 target genes of DUX4, i.e., *ZSCAN4*, *LEUTX* and *TRIM48* were significantly
251 upregulated in the EmGFP+ cells (Fig 5e). Downregulated protein-coding TFEs
252 included a number of ribosomal genes and genes maintaining pluripotency. This is in
253 agreement with previous findings showing that DUX4 downregulates some
254 pluripotency markers[25]. We integrated the data from our RNA-seq and published
255 DUX4 ChIP-seq analysis[26,27] and found that out of the 32 minor EGA genes induced
256 in 4-cell embryos[3], 23 were induced in the EmGFP+ cells and 17 out of these
257 overlapped with DUX4 binding sites. This suggested that DUX4 can induce the
258 majority of the minor EGA-related genes (Fig 5f). We also identified three previously
259 unannotated DUX4 targets *KHDC1* pseudogene 1 (FE463525; Fig 5e and 6a), RING-
260 finger type E3 ubiquitin ligase (FE533694; Fig 5e and 6b), and RING-finger domain
261 protein 4.2 (FE130507; Fig 5e and 6c) that were induced in 4-cell embryos[3], were
262 upregulated by DUX4, and overlapped with DUX4 binding sites. We cloned novel
263 DUX4 target transcripts from a cDNA pool of human day 4 embryos, confirming
264 their presence in early human cleavage embryos (Fig 6). On the other hand, out of the
265 129 major EGA genes upregulated in 8-cell human embryos[3], 14 were upregulated by
266 DUX4 and interestingly, 33 were downregulated (Fig 5f). These data suggested that
267 DUX4 could upregulate some major EGA genes, but that most of them are likely
268 activated by the minor EGA genes (such as LEUTX[28]), and that DUX4 might also
269 negatively regulate major EGA genes. DUX4-induced TFEs were highly enriched
270 with DUX4 binding sites[26,27] (Fig 5g), and the most highly enriched motif in the
271 DUX4-induced TFEs was similar to the known DUX4 motif[29] (Fig 5h). Furthermore,

272 these TFEs were remarkably overrepresented with the DUX4 binding sites among
273 hundreds of transcription factors (Fig 5i).

274

275 *Ectopic DUX4 expression causes chromatin opening at ERVL-MaLR elements similar*
276 *to 2-cell human embryos*

277 We next integrated STRT RNA-seq and ATAC-seq data from EmGFP +/- sorted
278 TetOn-Dux4 hESC. DUX4 expression lead to consistent modifications in chromatin
279 accessibility across all studied clones (Fig 7a). The transcriptionally upregulated TFEs
280 correlated with more accessible (gained) genomic regions in the DUX4 induced cells
281 while the correlation was less obvious for the downregulated TFEs and less accessible
282 (lost) chromatin (Fig 7a). We found that the gained chromatin regions correlated with
283 upregulated TFEs and lost chromatin regions correlated with downregulated TFEs
284 (Fig 7b), likely implying that the transcriptional downregulation induced by DUX4
285 expression is faster than the nucleosome-mediated closing of the chromatin (Fig 7a
286 and b). We then asked how DUX4 expression modified openness of the different
287 regions of the chromatin. In general, DUX4 rapidly caused chromatin remodelling,
288 especially chromatin opening, far from transcription start sites (TSS), demonstrating
289 that the TSSs seem to be less targeted by the DUX4 expression than other genomic
290 regions (Fig 7c). We next focused on the ATAC-gained chromatin sites. These
291 chromatin regions were remarkably enriched with DUX4 binding sites compared with
292 unchanged regions ($P<2.2e-16$). Out of the ATAC-gained chromatin sites, 48.9%
293 overlapped with ERVL-MaLR elements and they were significantly enriched for the
294 DUX4 binding sites compared with non ERVL-MaLR overlapping sites (55.8% $P <$
295 $2.2e-16$) (Fig. 7d). The ATAC-gained ERVL-MaLR regions remarkably overlapped

296 with the open chromatin regions found in 2-cell human embryos[30] (Fig. 7e). Out of
297 the DUX4 induced gained chromatin regions that overlapped with those of the 2-cell
298 embryos, upregulated by DUX4 induction, and overlapped with DUX4 binding sites,
299 76.7% were unannotated. Only few protein-coding genes, for instance *ZSCAN4* and
300 the transcriptional and chromatin regulators *KDM5B (JARID1B)* and *ZNF296* were
301 included. These results show that DUX4 directly binds ERVL-MaLR elements and
302 converts the chromatin landscape of the hESCs towards that of the human 2-cell
303 embryos. These data also suggest that DUX4 largely functions through yet
304 unannotated genomic regions.

305

306 **Discussion**

307 The OET, including fertilization and activation of the oocyte to totipotent blastomeres
308 and subsequent EGA, gradually sets the stage for embryo development[31-33]. How the
309 OET is orchestrated in human and which factors are the main drivers are still poorly
310 known. One of the potential candidates driving the OET in human embryos is DUX4.
311 Our data here show that DUX4 transcripts appear immediately after fertilisation and
312 are downregulated rapidly following the first and the second cleavage divisions. The
313 origin of *DUX4* transcripts is still unclear. It could be one of the maternal dormant
314 RNAs, supported by the fact that DUX4 was absent in the majority of the oocytes but
315 was significantly induced in the zygotes. Dormant maternal RNAs are stored in the
316 oocytes as deadenylated transcripts and they are polyadenylated and translated only
317 after resumption of meiosis or after fertilisation[34]. The increasing nuclear DUX4
318 protein intensity from zygotes to 4-cell embryos and its disappearance in the nuclei of
319 8-cell embryos suggested that DUX4 can modify transcriptome and chromatin of the
320 embryos already before the genome activation takes place. Detailed mechanisms of
321 DUX4 protein degradation in the 8-cell embryos remain to be further investigated;
322 however, DUX4 upregulated and interacted with several protein ubiquitinases, such
323 as TRIM48, a well-known DUX4 target gene. We also identified two previously
324 unannotated and possibly embryo-specific putative RING-finger type E3 ubiquitin
325 ligases that were expressed in early human embryos and induced by DUX4. The
326 putative ubiquitinases RFPLA (RFPL4A) and RFPLB (RFPL4B) regulate protein
327 degradation in germ cells[35,36]. Another ubiquitin-ligase, RNF114, was recently shown
328 to be essential for the OET in the mouse[37]. Taken together, our data suggested that
329 DUX4 induces expression of ubiquitin ligases and also interacts with ubiquitinases,

330 possibly regulating the presence of DUX4 itself as well as general proteome during
331 OET.

332

333 Knock-down of DUX4 in the human zygotes did not cause mitotic arrest during the 2-
334 day experiment, in agreement with recent findings on Dux in mouse embryos where a
335 minority of embryos may proceed until blastocyst stage[8,38,39]. In the mouse Dux-/-
336 embryos, around 30% of the EGA transcripts that should be upregulated were
337 downregulated[39], while in the human DUX4 knock-down embryos, many of the
338 maternal, normally downregulated genes remained unchanged, suggesting that
339 Dux/DUX4 alone is sufficient for neither the OET nor the EGA. Another candidate
340 gene regulating OET is *PLAG1*[40]. De novo PLAG1 binding site is found in the EGA
341 genes in the human embryos, and the phenotype of Plag1+/- mice lacking the
342 maternal *Plag1* allele show enriched expression of maternal transcripts at the 2-cell
343 stage, when major EGA occurs in the mouse. The question remains how DUX4,
344 together with other factors such as PLAG1, coordinates regulation of the maternal and
345 EGA transcripts in human and which yet unnamed genes might be involved in the
346 OET in human.

347

348 Ectopic expression of DUX4 in the hESC caused opening of the chromatin regions
349 outside of TSSs, largely at ERVL-MaLR elements. Dux binding at *Mervl* loci drives
350 chromatin reorganisation at *Mervl* loci in the mouse 2-cell embryo-like cells, and
351 chromatin organisation during early mouse development is a consequence of the
352 *Mervl* integration[41]. Human 2-cell-like cells have not been established by now, but
353 importantly, in our experiments binding of the DUX4 at ERVL-MaLR elements[14]

354 could modify chromatin towards embryo-like stage even in the hESCs. Long terminal
355 repeat elements abundantly present in the genome have been suggested as key
356 elements contributing to the OET, when major epigenetic and chromatin changes take
357 place[42]. Our integrated analysis on chromatin openness and transcriptional regulation,
358 together with the DUX4 ChIP-seq data[26,27] demonstrated that DUX4 regulates several
359 transcripts and the corresponding genomic loci coding for chromatin modifiers and
360 epigenetic regulators, as also suggested by Liu et al.[43]. Taken together, our data
361 indicate that DUX4 alone is not a sufficient inducer for the first three days of human
362 embryo development but that it regulates the OET by regulating maternal RNA
363 degradation, EGA genes, and repetitive elements, all of which have been shown to be
364 crucial for the successful OET in organisms other than human. In addition to
365 regulating genetic elements by DNA-binding, DUX4 may regulate the proteome by
366 inducing ubiquitination pathway genes during the OET in human.

367

368

369

370

371

372 **Online methods**

373

374 *Human pre-implantation embryos for single cell RNA-sequencing using STRT method*

375 We analysed single cell RNA-sequencing data from Töhönen et al.[3] for MII oocytes
376 (N=20), zygotes (N=59), 2-cell (N=4), 4-cell (N=15) and 8-cell (N=14) embryos.
377 For the DUX4 knockdown experiment 18 siCTRL cells (N=2) and 18 siDUX4 cells
378 (N=3) were analysed. The embryos were incubated in Ca^{2+}/Mg^{2+} -free culture
379 medium (Biopsy Medium, Origio) at 37°C heated stage for separation of the cells.
380 Individuals cells were briefly rinsed in Ca^{2+}/Mg^{2+} -free PBS and placed directly in
381 lysis buffer (5mM Tris-HCl, pH 7.0 (LifeTechnologies); 5mM DTT (Thermo
382 Scientific), 0.02 % Tx100 (Fisher Scientific); 0.5 U/µl Ribolock RNAse inhibitor
383 (ThermoFisher)). The library was prepared according to the published protocol[3,22,44].
384 The amplified libraries were sequenced on the Illumina HiSeq 2000 instrument.

385

386 *Bulk RNA-sequencing using STRT method*

387 Total RNA was isolated from the FAC-sorted DUX4-TetOn hESCs using the
388 RNAqueous Total RNA Isolation Kit (AM1912; ThermoFisherScientific). 20ng of
389 total RNA from each sample was used for library preparations. The libraries were
390 prepared using the STRT method as above, with minor modifications. Briefly, RNA
391 samples were placed in a 48-well plate in which a universal primer, template-
392 switching oligos, and a well-specific 8-bp barcode sequence (for sample
393 identification) were added to each well[23,45]. The synthesized cDNAs from the samples
394 were then pooled into one library and amplified by single-primer PCR with the

395 universal primer sequence. The resulting amplified library was then sequenced using
396 the Illumina NextSeq 500 instrument.

397

398 *Preprocess of raw STRT RNAseq reads*

399 The sequenced STRT raw reads were processed by STRTprep[23], v3dev branch
400 91a62d2 commit at https://github.com/shka/STRTprep/tree/v3dev. The processed
401 nonredundant reads were aligned to hg19 human reference genome sequences, ERCC
402 spike-in sequences and human ribosomal DNA unit (GenBank: U13369) with RefSeq
403 transcript alignments as a guide of exon junctions. For gene-based statistics, uniquely
404 mapped reads within (i) the 5'-UTR or the proximal upstream (up to 500 bp) of the
405 RefSeq protein coding genes, and (ii) within the first 50 bp of spike-in sequences,
406 were counted. For TFE-based statistics, the mapped reads were assembled according
407 to the alignments, and uniquely mapped reads within the first exons of the assembled
408 transcripts were counted, as described in Töhönen et al 2015[3].

409

410 *Downstream STRT RNA-sequencing data analysis*

411 Differentially expressed genes and TFEs required significantly different tendency on
412 the expression levels between two groups (q-value < 0.05), and significantly larger
413 variation than the technical variation (adjusted p-value < 0.05 by BH correction). The
414 former tendency was tested by the R package SAMstrt v0.99.0[45], and the latter
415 variation (fluctuation) was estimated based on gene-to-spikein (or TFE-to-spikein)
416 ratios in the squared coefficient of variation, described in Supplementary Text 1 of
417 Krjutskov et al. 2016[23]. The minimum value but non-zero was added to all the

418 normalized read counts and then the counts were divided by the minimum value so
419 that the logarithm of zero counts become zero. Enrichment analysis of anatomical
420 terms for the list of upregulated genes by siDUX4 was performed using the TopAnat[24]
421 (https://bgee.org/?page=top_anat). All human genes in the Bgee database were used
422 as background. STRT data of human early embryo were obtained from Töhönen et al.
423 2015 and 2017[3,10] and were overlapped with TFEs using the intersectBed function
424 from BEDTools[46] (v2.27.1). DUX4 ChIP-seq data was obtained from GSE33838[26]
425 and scores around the FEs were calculated with computeMatrix and visualized with
426 plotProfile from deepTools[47] (v3.1.3). Motif enrichment was analyzed using the
427 command findMotifsGenome.pl from HOMER[48] (v4.10.3) with the option "-size -
428 300,100". Enrichment analysis with publicly available ChIP-seq datasets was
429 conducted with ChIP-Atlas[49] (http://chip-atlas.org). A total of 7,216 human
430 transcription factor ChIP-seq datasets which had more than 500 peaks were analyzed.
431 Fold enrichment was calculated as (the number of ChIP-seq peaks overlapping with
432 upregulated TFEs / the number of upregulated TFEs) / (the number of ChIP-seq peaks
433 overlapping with all TFEs / the number of all TFEs). *P*-values were calculated with
434 Fisher's exact test and *Q*-values were calculated with the Benjamini & Hochberg
435 method. After excluding the TFEs annotated on ribosomal DNA, 6,425 upregulated
436 TFEs were used as foreground and 109,624 all the detected TFEs were used as
437 background both in the motif and ChIP-seq enrichment analysis.

438

439 *Human ESC culture*

440 hESC lines H1 (WA01) and H9 (WA09) were purchased from WiCell. The hESCs
441 were maintained on Geltrex-coated tissue culture dishes in Essential 8 culture medium

442 and passaged every three to five days by incubation with 0.5 mM EDTA (all from
443 Thermo Fisher Scientific).

444

445 *Plasmid construction*

446 The full-length DUX4 (NM_001293798.2) was synthesized and cloned between the
447 SalI and BamHI sites of the pB-tight-hMAFA-ires-EmGFP-pA-PGK-Puro vector (a
448 kind gift from Diego Balboa, Biomedicum Stem Cell Centre) at GenScript (Genscript,
449 NJ, USA).

450

451 *Doxycycline-inducible DUX4 expressing human ESCs*

452 The hESCs were incubated with StemPro Accutase (Thermo Fisher Scientific) until
453 the edges of the colonies started to curl up. The Accutase was aspirated and the cells
454 were gently detached in cold 5% FBS (Thermo Fisher Scientific) -PBS (Corning) and
455 counted. One million cells were centrifuged at 800rpm for 5 min and the pellet was
456 transferred into 120 µl of R-buffer containing 1 µg of pB-tight-DUX4-ires-EmGFP-
457 pA-PGK-Puro, 0.5 µg of pBASE and 0.5 µg of rtTA-M2-IN plasmids. 100 µl of the
458 cell-plasmid suspension was electroporated with two pulses of 1100 V, 20 ms pulse
459 width, using Neon Transfection system. The electroporated cells were plated on
460 Geltrex-coated dishes in Essential 8 medium with 10 µM ROCK inhibitor Y27632
461 (Peprotech). Fresh Essential 8 medium without ROCK-inhibitor was changed to the
462 cells on the day following the electroporation. The cells were selected with
463 Puromycin at 0.3 µg/mL. The TetOn-DUX4 hESC clones were picked manually on
464 Geltrex-coated 96-well plates, expanded and selected again with Puromycin.

465 Appearance of the EmGFP reporter protein was tested using Doxycycline at
466 concentrations ranging from 0.2 μg/ml to 1.0 μg/ml and detected using EVOS FL Cell
467 imaging system (Thermo Fisher Scientific). When indicated for the experiments
468 presented in this paper, the EmGFP+ DUX4 expressing hESC clones had been treated
469 with 1μg/ml of Doxicycline for 1, 2, 3, (qPCR) or 4 hours prior to downstream
470 analyses.

471

472 *cDNA cloning of unannotated DUX4 targets*

473 Single human 4-cell embryo cDNA library was prepared according to the protocol by
474 Tang et al.[50] and used for cloning of the putative transcripts. The transcripts were
475 amplified using Phusion High-Fidelity DNA polymerase (New England Biolabs)
476 according to manufacturer's instructions. Predicted KHDC1 pseudo gene 1, putative
477 RING-finger type E3 ubiquitin ligase and putative RING-finger domain protein
478 encoding genes were amplified using touchdown PCR: 98˚C for 30 s; 24 cycles of
479 98˚C for 10 s, annealing for 30 s, temperature decreasing from 63˚C to 56˚C, 1 ˚C/3
480 cycles, 72˚C for 30 s; 16 cycles of 98 ˚C for 10 s, 55˚C for 30 s, 72˚C for 30 s; final
481 extension 72˚C for 10 min. All PCR products were cloned into pCR4Blunt-TOPO
482 vector using Zero Blunt TOPO PCR Cloning kit (Invitrogen) and sequences were
483 verified by Sanger sequencing (Eurofins Genomics). Clone sequences are available
484 from the ENA browser at http://www.ebi.ac.uk/ena/data/view/LR694082-LR694089.

485

486 *Bioinformatics analysis and molecular dynamics simulations of the DUX4 protein*

487 The sequence for the human DUX4 (Q9UBX2) protein was obtained from the
488 UniProt database (The UniProt Consortium[30]), whereas other sequences were

489 retrieved from the non-redundant database of NCBI using blastp[51] and with human
490 DUX4 as the query sequence. Multiple sequence alignment was carried out using
491 MAFFT[52]. Secondary structures, solvent accessibility and disordered regions were
492 predicted using POLYVIEW-2D[53], SABLE[54], SCRATCH[55] and RaptorX-Property[56].
493 The 9aaTAD web server[57] was used to predict 9aaTAD motifs. The 2.12 Å resolution
494 crystal structure of the DUX4 HD1-linker-HD2 fragment bound to DNA[19] (PDB:
495 6E8C) was obtained from the Protein Data Bank[58]. PyMOL (version 1.6; Schrödinger
496 LLC) and Bodil[59] were used to analyze inter-HD interactions. Based on the DUX4
497 structure, molecular dynamics (MD) simulations, over all atoms, were used to explore
498 the dynamic states of the (1) double HD complex with (HD1-HD2 + DNA) and (2)
499 without (HD1-HD2) bound DNA and the individual HDs with bound DNA: (3) HD1
500 + DNA and (4) HD2 + DNA. MD simulations of these four structures were carried
501 out with the AMBER package[60] (version 18) using the ff14SB[61] (for protein) and
502 OL15[62] (for DNA) force fields. The structures were solvated with explicit TIP3P
503 water molecules[63] within an octahedral box ensuring a 12 Å distance between the
504 boundaries of the simulation box and solute atoms. Sodium counter ions were added
505 to neutralize the system and additional Na^+/Cl^- ions were added to bring the salt
506 concentration to 150 mM. Periodic boundary conditions were employed and the
507 particle-mesh Ewald algorithm[64] was applied to electrostatic interactions with a
508 distance cutoff of 9 Å.

509

510 Prior to the production simulation, 5000 cycles of steepest descent and conjugate
511 gradient energy minimization were carried out on each system, initiated by
512 introducing a 25 kcal mol^{-1} $Å^{-2}$ restraint on solute atoms that was gradually reduced
513 to 0 kcal mol^{-1} $Å^{-2}$ over the total minimization. The systems were then heated from

514 100 K to 300 K during 100 ps with a 10 kcal mol^{-1} $Å^{-2}$ restraint on solute atoms,
515 followed by a 900 ps equilibration at constant pressure while systematically reducing
516 the restraint to 0.1 kcal mol^{-1} $Å^{-2}$. The equilibration protocol was finalized with a
517 restraint-free 5 ns simulation. The production simulation was performed for 100 ns at
518 constant temperature (300 K) and pressure (1 bar), which was maintained using the
519 Berendsen algorithm[65] with 5 ps coupling constant. Trajectories were saved every 10
520 ps and the resulting structural snapshots were analyzed further by calculating the root-
521 mean-square deviations (RMSD; over backbone atoms) and root-mean-square
522 fluctuations (RMSF; over C α atoms), as well monitoring hydrogen bond interactions
523 using the programs CPPTRAJ[66] and VMD[67].

524

525 *Affinity purification of protein complexes, mass spectrometry and data analysis*

526 *Cell Culture and Affinity Purification*

527

528 *Cloning of DUX4 to MAC-tag Gateway® destination vector*

529 DUX4 was first amplified in two-step PCR reaction from pB-tight-DUX4-ires-
530 EmGFP-pA-PGK-Puro and cloned to Gateway compatible entry clone using Gateway
531 BP Clonase II (Invitrogen) according to manufacturer's instructions. The entry clone
532 was further cloned to Gateway compatible destination vectors containing the C-
533 terminal and N-terminal tags as described[21]. Transfection and selection of the T-Rex
534 293 cells (Invitrogen, Life Technologies, R78007, cultured in manufacturer's
535 recommended conditions) and affinity purification of the final product was done as
536 previously[21].

537

538 *Liquid Chromatography-Mass Spectrometry*

539 Analysis was performed on a Q-Exactive mass spectrometer with an EASY-nLC 1000
540 system via an electrospray ionization sprayer (Thermo Fisher Scientific), using
541 Xcalibur version 3.0.63. Peptides were eluted from the sample with a C18 precolumn
542 (Acclaim PepMap 100, 75 µm x 2 cm, 3 µm, 100 Å; Thermo Scientific) and
543 analytical column (Acclaim PepMap RSLC, 65 µm x 15 cm, 2 µm, 100 Å; Thermo
544 Scientific), using a 60 minute buffer gradient ranging from 5 to 35% Buffer B, then a
545 5 min gradient from 35 to 80% Buffer B and 10 minute gradient from 80 to 100%
546 Buffer B (0.1% formic acid in 98% acetonitrile and 2% HPLC grade water). 4 µl of
547 peptide sample was loaded by a cooled autosampler. Data-dependent FTMS
548 acquisition was in positive ion mode for 80 min. A full scan (200-2000 m/z) was
549 performed with a resolution of 70,000 followed by top10 CID-MS^2 ion trap scans with
550 a resolution of 17,500. Dynamic exclusion was set for 30 s. Database search was
551 performed with Proteome Discoverer 1.4 (Thermo Scientific) using the SEQUEST
552 search engine on the Reviewed human proteome in UniProtKB/SwissProt databases
553 (http://www.uniprot.org, downloaded Nov. 2018). Trypsin was selected as the
554 cleavage enzyme and maximum of 2 missed cleavages were permitted, precursor
555 mass tolerance at ±15 ppm and fragment mass tolerance at 0.05 Da.
556 Carbamidomethylation of cysteine was defined as a static modification. Oxidation of
557 methionine and biotinylation of lysine and N-termini were set as variable
558 modifications. All reported data were based on high-confidence peptides assigned in
559 Proteome Discoverer (FDR < 0.01).

560

561 *Identification of statistical confidence of interactions*

562 Significance Analysis of INTeractome (SAINT[68]) -express version 3.6.3 and
563 Contaminant Repository for Affinity Purification (CRAPome,
564 http://www.crapome.org) were used to discover statistically significant interactions
565 from the AP-MS data[69]. The DUX4 LC-MS data was ran alongside a large dataset of
566 other transcription factors, as well as a large GFP control set. Final results represent
567 proteins with a SAINT score higher than 0.73, and present in all four replicates.

568

569 *Overrepresentation Analysis*

570 Overrepresentation analysis of statistically significant interactions in Gene Ontology
571 and Reactome was done in WebGestalt[70], and overrepresentation of prey proteins in
572 ComplexPortal[71] (https://www.ebi.ac.uk/complexportal) and CORUM[72]
573 (https://mips.helmholtz-muenchen.de/corum/) was done using Fisher's exact test and
574 multiple testing correction in an in-house R-script.

575

576 *Interaction network*

577 Protein interaction networks were constructed from filtered SAINT data that was
578 imported to Cytoscape 3.6.0. Known prey-prey interactions were obtained from the
579 iRef database (http://irefindex.org).

580

581 *RNA isolation, reverse transcription and quantitative real-time quantitative PCR*

582 Total RNA was isolated by NucleoSpin RNA kit (Macherey Nagel). 1µg of RNA was
583 reverse transcribed by MMLV-RTase with oligodT, dNTPs, and Ribolock in MMLV-
584 RTase buffer (Thermo Fisher Scientific). 5X HOT FirelPol qPCR Mix (Solis
585 Biodyne) was used to measure relative mRNA levels with Lightcycler (Roche). Δ Δ
586 CT method was followed to quantify relative gene expression where *CYCLOPHILIN*
587 *G* was used as endogenous control. Relative expression of each gene was normalized
588 to the expression without doxycycline treatment.

589

590 *Fluorescence associated cell sorting*

591 TetOn-DUX4 hESCs were treated with TrypLE for 5 min and suspended into cold
592 FACS buffer (5% FBS-PBS). Single cell suspension was filtered through 40µm Cell
593 strainers and centrifuged at 800 rpm for 5min. The cell pellets were suspended in cold
594 FACS buffer and placed on ice. EmGFP- and EmGFP+ cells were separated to FACS
595 buffer by Sony SH800Z Cell Sorter with blue laser (488) and 100 µm nozzle.

596

597 *ATAC-sequencing library preparation and data analysis*

598 In principle the ATAC-sequencing libraries were prepared as in[73]. $5x10^4$ EmGFP-
599 negative and EmGFP-positive TetOn-hESCs for four biological samples; TetOn-
600 DUX4 in H1 clone 2, H1 clone 8, H9 clone 3 and H9 clone 4, were centrifuged at
601 500g for 5 min. The pellets were washed in cold 1X PBS by centrifugation at 500g for
602 5min. Each cell pellet was lysed in 50 µl of cold lysis buffer (10 mM Tris-HCl, pH
603 7.4; 10 mM NaCl, 3 mM $MgCl_2$, and 0.1% IGEPAL CA-630) and centrifuged at 500g

604 at 4°C for 10 min. The pellet was then resuspended in the transposase reaction mix
605 (2.5 µl of transposase in TD buffer) and incubated at 37°C for 30min. The reactions
606 were purified through columns and eluated in 20 µl. After addition of the barcode
607 oligos the DNA samples were amplified for 12 cycles (98°C for 10 seconds, 63°C for
608 30 seconds and 72°C for 60 seconds) in Phusion PCR master mix (Thermo Fisher
609 Scientific). The PCR products were purified through the columns and eluted in 20 µl.

610

611 *ATAC-seq data analysis*

612 Bcl files were converted and demultiplexed to fastq using the bcl2fastq program.
613 STAR[74] was used to index the human reference genome (hg19), obtained from
614 UCSC, and align the resulting fastq files. The resulting bam files with the mapped
615 reads were then converted to tag directories with subsequent peaks calling using the
616 HOMER suit of programs[48]. HOMER was also employed for counting the reads in the
617 identified peak regions. The raw tag counts from the peaks were then imported to
618 R/Bioconductor and differential peak analysis was performed using the edgeR
619 package and its general linear models pipeline. Peaks with an FDR adjusted p value
620 under 0.05 were termed significant. Plotting was done in R using packages Complex
621 heatmap, ggplot2 and ggbeeswarm. RepeatMasker table downloaded from UCSC
622 (http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/rmsk.txt.gz) was
623 converted to BED format and then intersected with the ATAC-seq peaks using the
624 intersectBed from BEDTools[46] to determine the peaks overlapped with ERVL-MaLR
625 elements. ATAC-seq data of human early embryo were obtained from GSE101571[30],
626 and scores around the ATAC-seq peaks were calculated with computeMatrix and

627 visualized with plotHeatmap from deepTools[47]. All the scripts and command line
628 options can be provided upon request.

629

630 *Immunocytochemistry of the human ESC*

631 The cells were fixed with 3.8% PFA, washed three times, permeabilised in 0.5% (v/v)
632 Tx100-PBS for 7 min, and washed with washing buffer (0.1% (v/v) Tween20-PBS).
633 The samples were incubated with ProteinBlock (Thermo Fisher Scientifi) at room
634 temperature for 10 min to prevent unspecific binding of primary antibody. Primary
635 antibody (rabbit MAb anti DUX4, clone E5-5, 1:400; Abcam) was diluted as
636 indicated in washing buffer and incubated at 4°C overnight. After washings,
637 fluorescence-conjugated secondary antibody (anti rabbit 594, A-21207; Thermo
638 Fisher Scientific) was diluted 1:1000 in washing buffer and incubated at room
639 temperature for 20 min. Nuclei were counterstained with DAPI 1:1000 in washing
640 buffer. The images were captured with Evos FL Cell Imaging system with 10X and
641 20X Plan Achromatic objectives.

642

643 *Immunocytochemistry of the human embryos*

644 The embryos were fixed in 3.8 % PFA at room temperature for 15min, washed three
645 times in the washing buffer (above), and permeabilised in 0.5 % Tx100-PBS at room
646 temperature for 15 min. Unspecific primary antibody binding was blocked as above.
647 DUX4 (as above) was incubated at 4°C overnight. The embryos were washed and
648 incubated in the secondary antibody (anti-rabbit 488, A-21202; Thermo Fisher
649 Scientific) diluted 1:500 in washing buffer (as above) at room temperature for 2

650 hours. After washings, nuclei were counterstained with DAPI 1:500 in washing
651 buffer.

652

653 *Imaging of the fixed human embryos*

654 The embryos were imaged in washing buffer on Ibidi 8-well μ slides with Leica TCS
655 SP8 confocal laser scanning microscope (Leica Microsystems, Mannheim, Germany)
656 using Leica HC PL APO CS2 40X/1.10NA and Leica HC PL APO CS2 63X/1.20NA
657 water objectives.

658

659 *Confocal microscopy image analysis*

660 Confocal images were processed using Fiji (http://fiji.sc). For the data presented in
661 the Fig 1b, images were smoothened by Gaussian filter (radius=1 pixel kernel). For
662 the quantification of the DUX4 intensity in the nucleus (Fig 1c), the DAPI channel
663 was denoised using rolling ball (radius=100). The images were smoothened in 3D
664 using Gaussian filter (radius=2 pixel kernel) and cell nuclei were segmented. The
665 segmented regions were used to measure average pixel intensity per nucleus in each
666 cell in the DUX4 channel. DUX4 intensity in the nucleus was normalized to intensity
667 of the corresponding cytoplasmic DUX4 staining in the single representative plane.

668

669 *Culture and microinjection of human embryos*

670 Human triploid zygotes were warmed using Gems Warming Set (Genea Biomedx)
671 and cultured in G-TL medium (Vitrolife) in 6 %O_2 and 6 % CO_2 at 37°C. 12 µl of
672 either 20 µM scrambled control siRNA (AM4611, Thermo Fisher Scientific) or

673 DUX4-targeting siRNA (cat. 4457308, Thermo Fisher Scientific) diluted in
674 nucleotide-free H_2O were mixed with total of 500 ng of GAP-GFP mRNA and
675 centrifuged at maximum speed at 4°C for 10 min. The embryos were microinjected
676 using Eppendorf microinjector and placed in G-TL medium in Geri dish for 3D time-
677 lapse imaging (Geri incubator, Genea Biomedx, Australia).

678

679 *Human embryo live imaging*

680 Imaging of the human triploid embryos was initiated immediately after
681 microinjections (Geri incubator). Images were captured in 3D every 15 minutes until
682 the embryos were removed for fluorescence staining or termination of the experiment.

683

684 *Ethical approvals*

685 Collection and experiments on human oocytes and embryos were approved by the
686 Helsinki University Hospital ethical committee, diary numbers 308/13/03/03/2015
687 and HUS/1069/2016. Human surplus oocytes, zygotes, and embryos were donated by
688 couples that had undergone infertility treatments at Helsinki University Hospital
689 Reproduction Medicine Unit. The donations were done with an informed consent and
690 patients understood that donating oocytes, zygotes, or embryos is voluntary.

691

692 *Acknowledgements*

693 We are grateful for all the couples that donated their surplus oocytes, zygotes, or
694 embryos for this project. We thank the IVF nurses at Helsinki University Hospital

695 Reproductive Medicine Unit for recruiting the couples for oocyte and embryo
696 donation program, Dr. Diego Balboa-Alonso for the pB-ires-EmGFP-pA-PGK-Puro
697 plasmid and discussions of the data and Dr. Jere Weltner for helpful discussions. We
698 acknowledge Biomedicum Imaging Unit (Helsinki), Functional Genomics Unit
699 (Helsinki), Biomedicum Flow Cytometry Unit (Helsinki), Bioinformatics and
700 Expression Analysis Core Facility (Stockholm) for skilled technical assistance. We
701 thank Dr. Jukka Lehtonen for scientific IT support (Biocenter Finland) and CSC IT
702 Center for Science for supercomputing. We thank prof. Outi Hovatta for inspiration
703 and introducing the senior author to the field of early embryogenesis. This study is
704 based on collaboration between the groups of the last seven authors. This project was
705 supported by Jane and Aatos Erkko foundation, Sigrid Jusélius foundation, Finnish
706 Academy, and Helsinki University Hospital funds. SV was supported by Jane and
707 Aatos Erkko foundation. MY was supported by the Scandinavia-Japan Sasakawa
708 Foundation, the Japan Eye Bank Association, the Astellas Foundation for Research on
709 Metabolic Disorders, and the Japan Society for the Promotion of Science Overseas
710 Research Fellowships. MT was supported by Joe, Pentti and Tor Memorial
711 Foundation, Graduate School of Åbo Akademi University. VR was supported by
712 Foundation of Åbo Akademi University and Magnus Ehrnrooth Foundation.

713

714 Contributions

715 SV, SK and JK conceived and coordinated the study. YM, TRB, MV, MSJ, TT, SK
716 and JK supervised the work in each contributing laboratory. Every author participated
717 in either planning or conducting respective experiments and analyzing or interpreting
718 the data. SV, MY, LG, VR, TA, MT, MSJ, JK wrote the manuscript. All authors
719 approved the final version of the manuscript.

720

721 *Competing interests*

722 The authors declare no competing interests.

723

724

725 *Figure Legends*

726 **Figure 1. DUX4 expression in human embryos.**

727 **(a)** Log_2 RPM of DUX4 mRNA reads in human oocytes (N=20), zygotes (N=59), 2-
728 cell (N=4), 4-cell (N=14), and 8-cell (N=15) embryos.

729 **(b)** Representative confocal images of zygotes (N=3), 2-cell (N=3), 4-cell (N=4), and
730 8-cell (N=2) human embryos stained with monoclonal DUX4 antibody E5-5 (green).
731 Nuclei counterstained with DAPI (magenta).

732 **(c)** 3D quantification of DUX4 intensity in the nuclei of the human embryos,
733 normalised to cytoplasmic DUX4 staining (single plane) with standard deviation.

734

735 **Figure 2. Structural features of DUX4.**

736 **(a)** Domain structure of full-length DUX4: N- and C-terminal amino acid residues, as
737 well as boundary residues for the homeodomains HD1 and HD2, three predicted
738 ordered C-terminal regions (disorder value < 0.5; red curve) and 9aaTAD motif
739 (blue). Conservation of residues in primates versus the human sequence C-terminal to
740 residue G153 (green curve) and sequence alignment and predicted secondary
741 structures (alpha helices) of the ordered regions. Residue numbering from UniProt ID
742 Q9UBX2.

743 **(b)** Crystal structure (PDB:6E8C; Lee et al., 2018) of DUX4 HD1 (blue) and HD2
744 (gold) in complex with the consensus DNA motif "TAATCTAATCA" (grey).
745 Disordered/linker regions, magenta; residues forming inter-homeodomain contacts
746 drawn as sticks.

747 **(c)** Structure of DUX4 HD1 (blue) and 2 (gold) without DNA. The coordinates of the
748 bound DNA of the X-ray structure shown in (b) were deleted from the coordinate file
749 of DUX4 structure. Coloring as in (b).

750 **(d)** View focused on the inter-homeodomain interactions shown in (c). Hydrogen
751 bonds, yellow dash lines; coloring as in (b).

752 **(e)** Sequence comparison of residues forming inter-homeodomain contacts.
753 **(f)** Root-mean-squared fluctuations (RMSF) of the Cα atoms of the X-ray structure
754 of DUX4 with (red curve) and without (blue curve) bound DNA during a 100 ns MD
755 simulation. HD1 (blue), linker (magenta) and HD2 (gold).

756 **(g)** Root-mean-squared deviation (RMSD) of the backbone atoms of the X-ray
757 structure of DUX4 with and without bound DNA during a 100 ns MD simulation.
758 Homedomains with (HD1-HD2 +DNA) and without (HD1-HD2) bound DNA and
759 single homedomains (HD1+DNA and HD2 +DNA) with bound DNA were used in
760 the simulations.

761

762 **Figure 3. The protein – protein interaction network of DUX4.**

763 **(a)** High confidence protein-protein interactions detected by AP-MS (n=24) and
764 BioID (n=139) -methods (SAINT score > 0.74). Average spectral count of the
765 interaction filtered to show interactions larger than the median (AP-MS=4.125,
766 BioID=7.5).

767 **(b)** DUX4 interactome, filtered to show spectral counts larger than median. BioID -
768 interactions shown in red lines and AP-MS -interactions in blue, if protein appeared in
769 both data sets it is outlined in bold black. Known prey-prey interactions shown in grey
770 (iREF).

771

772 **Figure 4. Knockdown of DUX4 in human embryos.**

773 **(a)** Schematic of the experimental set up.

774 **(b)** Human embryos immunostained with DUX4 antibody (green) 24 h after
775 microinjection with either control siRNA (left panel) or DUX4 targeting siRNA (right
776 panel). Nuclei counterstained with DAPI (blue). Overlay of a single representative z
777 plane and the corresponding z-planes shown for DUX4 staining, nucleus and bright
778 field channels on the right side of each overlay. Scale bar 50 μm.
779 **(c)** Scatter plot of the expression levels of TFEs across the siCTRL and siDUX4
780 embryos. Red dots represent significantly upregulated TFEs by siDUX4 and grey
781 crosses represent TFEs, which showed no significant change.
782 **(d-f)** TFEs upregulated (red dots) or showing no significant change (grey crosses) by
783 siDUX4 in the human OET transcriptome as in[3]. Comparisons in d: oocyte to 4-cell,
784 e: zygote to 4-cell, and f: 4-cell to 8-cell. The dotted line marks the cell division effect
785 on cellular RNA content. *P*-values were calculated with Fisher's exact test for the
786 frequency of the siDUX4-upregulated TFEs of the TFEs normally downregulated
787 during respective stages.

788 **(g)** Expression levels of the oocyte-specific genes in siCTRL and siDUX4 embryos.
789 Asterisks represent statistical significance (q-value < 0.05). Horizontal lines represent
790 the median values in each group.

791

792 **Figure 5. Transcriptome and ChIP-seq analysis on the TetOn-DUX4 hESCs**.
793 **(a)** Schematic of the experimental set up.

794 **(b)** Doxicycline induction of TetOn-DUX4 hESCs induces expression of DUX4 and
795 DUX4 target genes, ZSCAN4 and TRIM48. Shown for H1-TetOn-DUX4 clone 2. All
796 clones selected for experiments followed the same trend. X-axis indicates time (h)
797 incubated in doxycycline. Relative mRNA expression levels were normalized to the
798 non-induced cells.

799 **(c)** 4-hour doxycycline induction upregulates DUX4 protein expression in the
800 nucleus, shown for H1-TetOn-DUX4 clone 2.

801 **(d)** Proportion of the upregulated and downregulated TFEs based on the genome
802 annotation.

803 **(e)** Expression level of putative DUX4 target genes. Asterisks represent statistical
804 significance (q-value < 0.05). Horizontal lines represent the median values in each
805 group.

806 **(f)** Proportion of the upregulated (Up), downregulated (Down), and non-significantly
807 changed (NS) TFEs by DUX4 induction among the minor (Oocyte to 4-cell embryo)
808 and major (4- to 8-cell embryo) EGA genes. One TFE out of the 129 major EGA
809 genes annotated on unassigned chromosome (ChrUn) and was excluded from the
810 analysis.

811 **(g)** DUX4 ChIP-seq intensity[26] around the peaks of reads within the upregulated TFEs
812 (blue) and all the detected TFEs (green).

813 **(h)** De novo motif enrichment analysis of the DUX4-induced TFEs. Top: the most
814 significantly enriched motif (P = 1e-961). Bottom: the best-matched known motif
815 (DUX4 ChIP-seq of myoblasts: GSE75791[29]; matched score = 0.92).

816 **(i)** Enrichment analysis of the DUX4-induced TFEs with publicly available ChIP-seq
817 datasets. A total of 7,216 ChIP-seq data for transcription factors are shown. ChIP-seq

818 data for DUX4 are shown in red. Dots on the left side of the dashed line are
819 underrepresented, whereas dots on the right side are overrepresented.

820

821 **Figure 6. Novel DUX4 targets.**

822 **(a)** Predicted *KHDC1* pseudogene 1 (clone K5.2), at chromosome 6 (73,918,461-
823 73,920,115) was expressed by the human 4-cell embryos (FE463525) and upregulated
824 in the TetOn -DUX4- hESCs (TFE93242). TFEs overlapped with DUX4 binding sites
825 (DUX4 ChIP). cDNA clone K5.2 (thick orange regions indicate exons and grey thin
826 regions indicate introns) corresponded to the *KHDC1* pseudogene 1 transcript
827 assembly in TetOn-DUX4 cells. Transcript assemblies (mRNA Genbank and human
828 ESTs), including unspliced, are shown.

829 **(b)** Putative RING-finger type E3 ubiquitin ligase at chromosome 2 (108,273,771-
830 108,277,850) was expressed by the human 4-cell embryos (FE130507) and it was
831 upregulated in the TetOn-DUX4 hESCs (TFE25640). DUX4 ChIP-seq peak
832 overlapped with the TFEs. RET11.1 was cloned from human 4-cell embryo (clone
833 RET11.1). Thick blue regions indicate exons and thin grey regions indicate introns.
834 Transcript assemblies (mRNA Genbank and human ESTs), including unspliced, are
835 shown.

836 **(c**) Putative RING-finger domain protein at chromosome 8 (210,701-215,100) was
837 expressed by the human 4-cell embryos (TFE533694) and induced by TetOn-DUX4
838 hESCs (TFE102707). ChIP-seq overlapped with the TFEs. Two cDNA clones, Ring
839 4.2 and Ring 10.22, were expressed in the human 4-cell embyos. Thick blue regions
840 indicate exons and grey thin regions indicate introns. Transcript assemblies (mRNA
841 Genbank and human ESTs), including unspliced, are shown.

842

843 **Figure 7. Integrated analysis of the DUX4 induced changes in the chromatin and**
844 **transcriptome of the hESCs**.

845 **(a)** Heatmap of the 4,686 ATAC-sequencing reads across all samples at ATAC-seq
846 peaks that overlap with differentially regulated TFE reads. Counts for each peak were
847 standardized across each sample (mean=0, sd=1). Samples and peaks were then
848 clustered using hierarchical clustering. The Separate heatmap of the ATAC-seq shows
849 if the changes in the heatmap are significant (red: ATAC-reg gained; blue: ATAC-reg
850 lost; grey: ATAC-reg non-significant). TFE-Reg heatmap shows if the overlapping
851 TFE site is upregulated or downregulated (red: upregulated; blue: downregulated).
852 **(b)** Quasi random plot showing the results of the differential peak analysis on the
853 ATAC-seq and STRT-RNA-seq. Each point is an ATAC-seq peak. Analyses were
854 carried out on peaks that were repeated at least three times. Red: ATAC-seq gained;
855 blue: ATAC-seq lost; grey: ATAC-seq non-significant. Y-axis: The log fold change
856 of the ATAC-seq reads in the DUX4 expressing versus the control samples. X-axis:
857 The ATAC-seq peaks overlapping either with the down-regulated TFEs or
858 upregulated TFEs.

859 **(c)** Density plot showing distribution of the ATAC-seq peaks relative to the TSS of
860 genes separated by how the peak is regulated by the DUX4-expression. red: ATAC-
861 reg gained; blue: ATAC-reg lost; grey: ATAC-reg non-significant.

862 **(d)** Proportion of the peaks overlapped with ERVL-MaLR elements in the gained,
863 non-significant and lost ATAC-reg peaks (pink) by DUX4 induction. Inset pie charts
864 inidicate the proportion of the ATAC-gained peaks overlapping with DUX4 binding
865 sites (green).

866 **(e)** ATAC-reg intensity of human early embryo[30] around the gained, non-significant
867 (NS), and lost ATAC-reg peaks by DUX4 induction which overlap with ERVL-
868 MaLR elements.
869
870

871 *References*

872 1 Jukam, D., Shariati, S. A. M. & Skotheim, J. M. Zygotic Genome Activation
873 in Vertebrates. *Developmental cell* **42**, 316-332,
874 doi:10.1016/j.devcel.2017.07.026 (2017).

875 2 Braude, P., Bolton, V. & Moore, S. Human gene expression first occurs
876 between the four- and eight-cell stages of preimplantation development.
877 *Nature* **332**, 459-461, doi:10.1038/332459a0 (1988).

878 3 Tohonen, V. *et al.* Novel PRD-like homeodomain transcription factors and
879 retrotransposon elements in early human development. *Nature*
880 *communications* **6**, 8207, doi:10.1038/ncomms9207 (2015).

881 4 Blakeley, P. *et al.* Defining the three cell lineages of the human blastocyst by
882 single-cell RNA-seq. *Development* **142**, 3613, doi:10.1242/dev.131235
883 (2015).

884 5 Galan, A. *et al.* Functional genomics of 5- to 8-cell stage human embryos by
885 blastomere single-cell cDNA analysis. *PloS one* **5**, e13615,
886 doi:10.1371/journal.pone.0013615 (2010).

887 6 Stirparo, G. G. *et al.* Integrated analysis of single-cell embryo data yields a
888 unified transcriptome signature for the human pre-implantation epiblast.
889 *Development* **145**, doi:10.1242/dev.158501 (2018).

890 7 Boroviak, T. *et al.* Single cell transcriptome analysis of human, marmoset and
891 mouse embryos reveals common and divergent features of preimplantation
892 development. *Development* **145**, doi:10.1242/dev.167833 (2018).

893 8 De Iaco, A. *et al.* DUX-family transcription factors regulate zygotic genome
894 activation in placental mammals. *Nat Genet* **49**, 941-945, doi:10.1038/ng.3858
895 (2017).

896 9 Hendrickson, P. G. *et al.* Conserved roles of mouse DUX and human DUX4 in
897 activating cleavage-stage genes and MERVL/HERVL retrotransposons.
898 *Nature genetics* **49**, 925-934, doi:10.1038/ng.3844 (2017).

899 10 Tohonen, V. *et al.* Transcriptional activation of early human development
900 suggests DUX4 as an embryonic regulator. doi:https://doi.org/10.1101/123208
901 (2017).

902 11 Yan, L. *et al.* Single-cell RNA-Seq profiling of human preimplantation
903 embryos and embryonic stem cells. *Nature structural & molecular biology* **20**,
904 1131-1139, doi:10.1038/nsmb.2660 (2013).

905 12 Dang, Y. *et al.* Tracing the expression of circular RNAs in human pre-
906 implantation embryos. *Genome biology* **17**, 130, doi:10.1186/s13059-016-
907 0991-3 (2016).

908 13 Xue, Z. *et al.* Genetic programs in human and mouse early embryos revealed
909 by single-cell RNA sequencing. *Nature* **500**, 593-597,
910 doi:10.1038/nature12364 (2013).

911 14 Whiddon, J. L., Langford, A. T., Wong, C. J., Zhong, J. W. & Tapscott, S. J.
912 Conservation and innovation in the DUX4-family gene network. *Nat Genet*
913 **49**, 935-940, doi:10.1038/ng.3846 (2017).

914 15 Codutti, L. *et al.* The solution structure of DNA-free Pax-8 paired box domain
915 accounts for redox regulation of transcriptional activity in the pax protein
916 family. *The Journal of biological chemistry* **283**, 33321-33328,
917 doi:10.1074/jbc.M805717200 (2008).

918 16 Mitsuhashi, H. *et al.* Functional domains of the FSHD-associated DUX4
919 protein. *Biology open* **7**, doi:10.1242/bio.033977 (2018).

920 17 Katayama, S. *et al.* Phylogenetic and mutational analyses of human LEUTX, a
921 homeobox gene implicated in embryogenesis. *Scientific reports* **8**, 17421,
922 doi:10.1038/s41598-018-35547-5 (2018).

923 18 Goto, N. K., Zor, T., Martinez-Yamout, M., Dyson, H. J. & Wright, P. E.
924 Cooperativity in transcription factor binding to the coactivator CREB-binding
925 protein (CBP). The mixed lineage leukemia protein (MLL) activation domain
926 binds to an allosteric site on the KIX domain. *The Journal of biological*
927 *chemistry* **277**, 43168-43174, doi:10.1074/jbc.M207660200 (2002).

928 19 Lee, J. K. *et al.* Crystal Structure of the Double Homeodomain of DUX4 in
929 Complex with DNA. *Cell reports* **25**, 2955-2962 e2953,
930 doi:10.1016/j.celrep.2018.11.060 (2018).

931 20 Varjosalo, M. *et al.* Interlaboratory reproducibility of large-scale human
932 protein-complex analysis by standardized AP-MS. *Nature methods* **10**, 307-
933 314, doi:10.1038/nmeth.2400 (2013).

934 21 Liu, X. *et al.* An AP-MS- and BioID-compatible MAC-tag enables
935 comprehensive mapping of protein interactions and subcellular localizations.
936 *Nature communications* **9**, 1188, doi:10.1038/s41467-018-03523-2 (2018).

937 22 Islam, S. *et al.* Highly multiplexed and strand-specific single-cell RNA 5' end
938 sequencing. *Nature protocols* **7**, 813-828, doi:10.1038/nprot.2012.022 (2012).

939 23 Krjutskov, K. *et al.* Single-cell transcriptome analysis of endometrial tissue.
940 *Human reproduction* **31**, 844-853, doi:10.1093/humrep/dew008 (2016).

941 24 Komljenovic, A., Roux, J., Wollbrett, J., Robinson-Rechavi, M. & Bastian, F.
942 B. BgeeDB, an R package for retrieval of curated expression datasets and for
943 gene list expression localization enrichment tests. *F1000Research* **5**, 2748,
944 doi:10.12688/f1000research.9973.2 (2016).

945 25 Dandapat, A., Hartweck, L. M., Bosnakovski, D. & Kyba, M. Expression of
946 the human FSHD-linked DUX4 gene induces neurogenesis during
947 differentiation of murine embryonic stem cells. *Stem cells and development*
948 **22**, 2440-2448, doi:10.1089/scd.2012.0643 (2013).

949 26 Geng, L. N. *et al.* DUX4 activates germline genes, retroelements, and immune
950 mediators: implications for facioscapulohumeral dystrophy. *Developmental*
951 *cell* **22**, 38-51, doi:10.1016/j.devcel.2011.11.013 (2012).

952 27 Young, J. M. *et al.* DUX4 binding to retroelements creates promoters that are
953 active in FSHD muscle and testis. *PLoS genetics* **9**, e1003947,
954 doi:10.1371/journal.pgen.1003947 (2013).

955 28 Jouhilahti, E. M. *et al.* The human PRD-like homeobox gene LEUTX has a
956 central role in embryo genome activation. *Development* **143**, 3459-3469,
957 doi:10.1242/dev.134510 (2016).

958 29 Eidahl, J. O. *et al.* Mouse Dux is myotoxic and shares partial functional
959 homology with its human paralog DUX4. *Human molecular genetics* **25**,
960 4577-4589, doi:10.1093/hmg/ddw287 (2016).

961 30 Wu, J. *et al.* Chromatin analysis in human early development reveals
962 epigenetic transition during ZGA. *Nature* **557**, 256-260, doi:10.1038/s41586-
963 018-0080-8 (2018).

964 31 Tadros, W. & Lipshitz, H. D. The maternal-to-zygotic transition: a play in two
965 acts. *Development* **136**, 3033-3042, doi:10.1242/dev.033183 (2009).

966 32 Walser, C. B. & Lipshitz, H. D. Transcript clearance during the maternal-to-
967 zygotic transition. *Current opinion in genetics & development* **21**, 431-443,
968 doi:10.1016/j.gde.2011.03.003 (2011).

969 33 Lee, M. T., Bonneau, A. R. & Giraldez, A. J. Zygotic genome activation
970 during the maternal-to-zygotic transition. *Annual review of cell and*
971 *developmental biology* **30**, 581-613, doi:10.1146/annurev-cellbio-100913-
972 013027 (2014).

973 34 Winata, C. L. *et al.* Cytoplasmic polyadenylation-mediated translational
974 control of maternal mRNAs directs maternal-to-zygotic transition.
975 *Development* **145**, doi:10.1242/dev.159566 (2018).

976 35 Suzumori, N., Burns, K. H., Yan, W. & Matzuk, M. M. RFPL4 interacts with
977 oocyte proteins of the ubiquitin-proteasome degradation pathway.
978 *Proceedings of the National Academy of Sciences of the United States of*
979 *America* **100**, 550-555, doi:10.1073/pnas.0234474100 (2003).

980 36 Rajkovic, A., Lee, J. H., Yan, C. & Matzuk, M. M. The ret finger protein-like
981 4 gene, Rfpl4, encodes a putative E3 ubiquitin-protein ligase expressed in
982 adult germ cells. *Mechanisms of development* **112**, 173-177 (2002).

983 37 Yang, Y. *et al.* The E3 ubiquitin ligase RNF114 and TAB1 degradation are
984 required for maternal-to-zygotic transition. *EMBO reports* **18**, 205-216,
985 doi:10.15252/embr.201642573 (2017).

986 38 Chen, Z. & Zhang, Y. Loss of DUX causes minor defects in zygotic genome
987 activation and is compatible with mouse development. *Nat Genet* **51**, 947-951,
988 doi:10.1038/s41588-019-0418-7 (2019).

989 39 De Iaco, A., Verp S., Offner S., Trono D. DUX is a non-essential synchronizer
990 of zygotic genome activation. doi:https://doi.org/10.1101/569434 (2019).

991 40 Madissoon, E. *et al.* Pleomorphic Adenoma Gene 1 Is Needed For Timely
992 Zygotic Genome Activation and Early Embryo Development. *Scientific*
993 *reports* **9**, 8411, doi:10.1038/s41598-019-44882-0 (2019).

994 41 Kruse K., D. N., Enriquez-Gasca R., Gaume X., Torres-Padilla M-E.,
995 Vaquerizas JM. Transposable elements drive reorganisation of 3D chromatin
996 during early embryogenesis. doi:https://doi.org/10.1101/523712 (2019).

997 42 Franke, V. *et al.* Long terminal repeats power evolution of genes and gene
998 expression programs in mammalian oocytes and zygotes. *Genome research*
999 **27**, 1384-1394, doi:10.1101/gr.216150.116 (2017).

1000 43 Liu, L. *et al.* An integrated chromatin accessibility and transcriptome
1001 landscape of human pre-implantation embryos. *Nature communications* **10**,
1002 364, doi:10.1038/s41467-018-08244-0 (2019).

1003 44 Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular
1004 identifiers. *Nature methods* **11**, 163-166, doi:10.1038/nmeth.2772 (2014).

1005 45 Katayama, S., Tohonen, V., Linnarsson, S. & Kere, J. SAMstrt: statistical test
1006 for differential expression in single-cell transcriptome with spike-in
1007 normalization. *Bioinformatics* **29**, 2943-2945,
1008 doi:10.1093/bioinformatics/btt511 (2013).

1009 46 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for
1010 comparing genomic features. *Bioinformatics* **26**, 841-842,
1011 doi:10.1093/bioinformatics/btq033 (2010).

1012 47 Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a
1013 flexible platform for exploring deep-sequencing data. *Nucleic acids research*
1014 **42**, W187-191, doi:10.1093/nar/gku365 (2014).

1015 48 Heinz, S. *et al.* Simple combinations of lineage-determining transcription
1016 factors prime cis-regulatory elements required for macrophage and B cell
1017 identities. *Molecular cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004
1018 (2010).

1019 49 Oki, S. *et al.* ChIP-Atlas: a data-mining suite powered by full integration of
1020 public ChIP-seq data. *EMBO reports* **19**, doi:10.15252/embr.201846255
1021 (2018).

1022 50 Tang, F. *et al.* RNA-Seq analysis to capture the transcriptome landscape of a
1023 single cell. *Nature protocols* **5**, 516-535, doi:10.1038/nprot.2009.236 (2010).

1024 51 Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic acids*
1025 *research* **36**, W5-9, doi:10.1093/nar/gkn201 (2008).

1026 52 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software
1027 version 7: improvements in performance and usability. *Molecular biology and*
1028 *evolution* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).

1029 53 Porollo, A. A., Adamczak, R. & Meller, J. POLYVIEW: a flexible
1030 visualization tool for structural and functional annotations of proteins.
1031 *Bioinformatics* **20**, 2460-2462, doi:10.1093/bioinformatics/bth248 (2004).

1032 54 Adamczak, R., Porollo, A. & Meller, J. Accurate prediction of solvent
1033 accessibility using neural networks-based regression. *Proteins* **56**, 753-767,
1034 doi:10.1002/prot.20176 (2004).

1035 55 Magnan, C. N. & Baldi, P. SSpro/ACCpro 5: almost perfect prediction of
1036 protein secondary structure and relative solvent accessibility using profiles,
1037 machine learning and structural similarity. *Bioinformatics* **30**, 2592-2597,
1038 doi:10.1093/bioinformatics/btu352 (2014).

1039 56 Wang, S., Li, W., Liu, S. & Xu, J. RaptorX-Property: a web server for protein
1040 structure property prediction. *Nucleic acids research* **44**, W430-435,
1041 doi:10.1093/nar/gkw306 (2016).

1042 57 Piskacek, S. *et al.* Nine-amino-acid transactivation domain: establishment and
1043 prediction utilities. *Genomics* **89**, 756-768, doi:10.1016/j.ygeno.2007.02.003
1044 (2007).

1045 58 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic acids research* **28**, 235-
1046 242, doi:10.1093/nar/28.1.235 (2000).

1047 59 Lehtonen, J. V. *et al.* BODIL: a molecular modeling environment for
1048 structure-function analysis and drug design. *Journal of computer-aided*
1049 *molecular design* **18**, 401-419 (2004).

1050 60 Case, D. *et al.* AMBER 2018. (2018).

1051 61 Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and
1052 Backbone Parameters from ff99SB. *Journal of chemical theory and*
1053 *computation* **11**, 3696-3713, doi:10.1021/acs.jctc.5b00255 (2015).

1054 62 Zgarbova, M. *et al.* Refinement of the Sugar-Phosphate Backbone Torsion
1055 Beta for AMBER Force Fields Improves the Description of Z- and B-DNA.
1056 *Journal of chemical theory and computation* **11**, 5723-5736,
1057 doi:10.1021/acs.jctc.5b00716 (2015).

1058 63 Jorgensen, W., Chandrasekhar, J. & Madura, J. Comparison of simple
1059 potential functions for simulating liquid water. *The Journal of chemical*
1060 *physics* **79**, doi: https://doi.org/10.1063/1.445869 (1983).

1061 64 Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of*
1062 *chemical physics* **103**, 8577-8593, doi:10.1063/1.470117 (1995).

1063 65 Berendsen, H. J. C., Postma, J. P., van Gunsteren, A., DiNola, A. & J.R., H.
1064 Molecular dynamics with coupling to an external bath. *The Journal of*
1065 *chemical physics* **81**, doi:https://doi.org/10.1063/1.448118 (1984).

1066 66 Roe, D. R. & Cheatham, T. E., 3rd. PTRAJ and CPPTRAJ: Software for
1067 Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of*
1068 *chemical theory and computation* **9**, 3084-3095, doi:10.1021/ct400341p
1069 (2013).

1070 67 Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics.
1071 *Journal of molecular graphics* **14**, 33-38, 27-38 (1996).

1072 68 Choi, H. *et al.* SAINT: probabilistic scoring of affinity purification-mass
1073 spectrometry data. *Nature methods* **8**, 70-73, doi:10.1038/nmeth.1541 (2011).

1074 69 Mellacheruvu, D. *et al.* The CRAPome: a contaminant repository for affinity
1075 purification-mass spectrometry data. *Nature methods* **10**, 730-736,
1076 doi:10.1038/nmeth.2557 (2013).

1077 70 Wang, J., Vasaikar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: a
1078 more comprehensive, powerful, flexible and interactive gene set enrichment
1079 analysis toolkit. *Nucleic acids research* **45**, W130-W137,
1080 doi:10.1093/nar/gkx356 (2017).

1081 71 Meldal, B. H. *et al.* The complex portal--an encyclopaedia of macromolecular
1082 complexes. *Nucleic acids research* **43**, D479-484, doi:10.1093/nar/gku975
1083 (2015).

1084 72 Giurgiu, M. *et al.* CORUM: the comprehensive resource of mammalian
1085 protein complexes-2019. *Nucleic acids research* **47**, D559-D563,
1086 doi:10.1093/nar/gky973 (2019).

1087 73 Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A
1088 Method for Assaying Chromatin Accessibility Genome-Wide. *Current*
1089 *protocols in molecular biology* **109**, 21 29 21-29,
1090 doi:10.1002/0471142727.mb2129s109 (2015).

1091 74 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*
1092 **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

1093

Fig.1

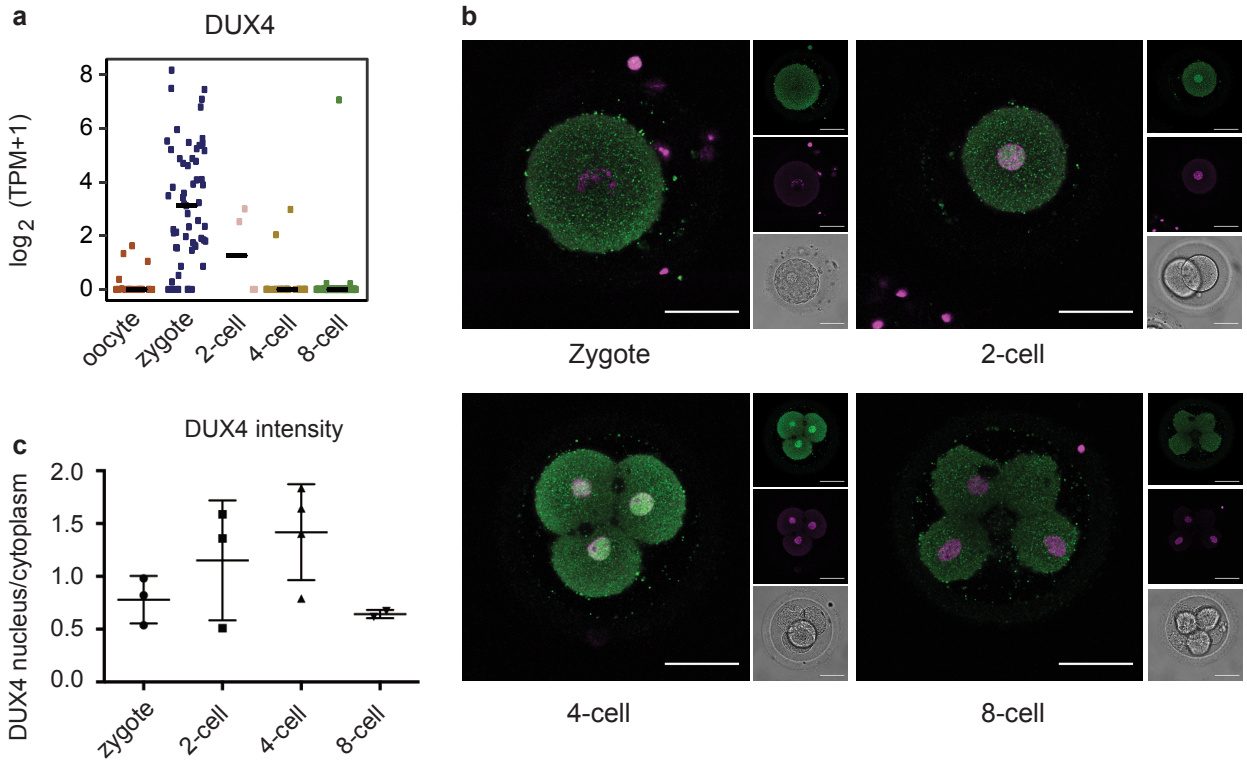

a
DUX4
log2 (TPM+1)
8
6
4
2
0
oocyte
zygote
2-cell
4-cell
8-cell
b
Zygote
2-cell
4-cell
8-cell
c
DUX4 intensity
DUX4 nucleus/cytoplasm
2.0
1.5
1.0
0.5
0.0
zygote
2-cell
4-cell
8-cell

Fig. 2

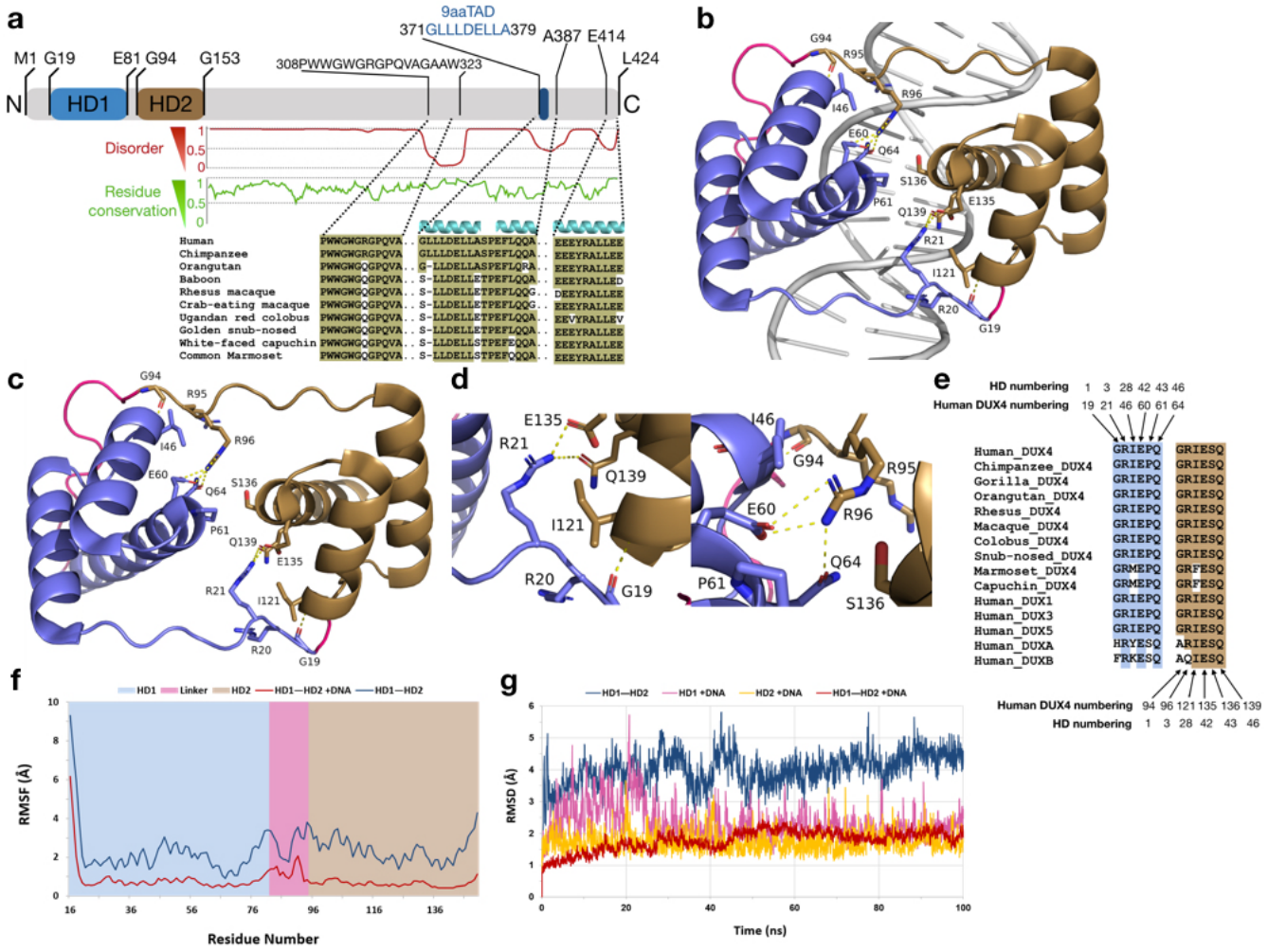

a
9aaTAD
371GLLLDELLA379
A387 E414
M1 G19
E81 G94
G153
308PWWGWGRGPQVAGAAW323
L424
N
HD1
HD2
C
Disorder
Residue conservation
Human
Chimpanzee
Orangutan
Baboon
Rhesus macaque
Crab-eating macaque
Ugandan red colobus
Golden snub-nosed
White-faced capuchin
Common Marmoset
b
c
d
e
HD numbering
Human DUX4 numbering
Human_DUX4
Chimpanzee_DUX4
Gorilla_DUX4
Orangutan_DUX4
Rhesus_DUX4
Macaque_DUX4
Colobus_DUX4
Snub-nosed_DUX4
Marmoset_DUX4
Capuchin_DUX4
Human_DUX1
Human_DUX3
Human_DUX5
Human_DUXA
Human_DUXB
f
HD1
Linker
HD2
HD1—HD2 +DNA
HD1—HD2
RMSF (Å)
Residue Number
g
HD1—HD2
HD1 +DNA
HD2 +DNA
HD1—HD2 +DNA
RMSD (Å)
Time (ns)

Fig.3

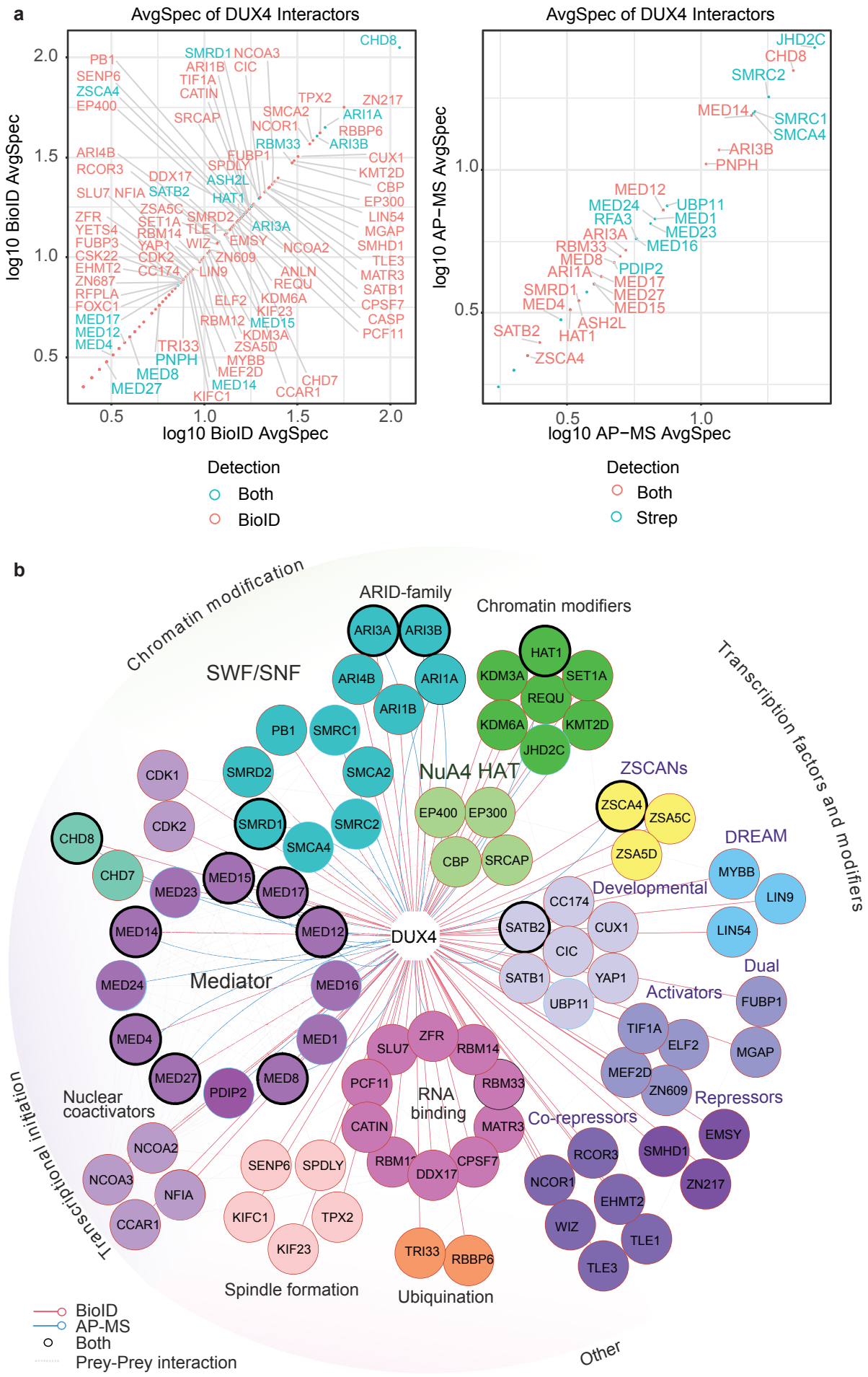

a
AvgSpec of DUX4 Interactors
log10 BioID AvgSpec
log10 BioID AvgSpec
Detection
Both
BioID
AvgSpec of DUX4 Interactors
log10 AP−MS AvgSpec
log10 AP−MS AvgSpec
Detection
Both
Strep
b
Chromatin modification
ARID-family
Chromatin modifiers
SWF/SNF
Transcription factors and modifiers
NuA4 HAT
ZSCANs
DREAM
Developmental
Dual
Activators
Repressors
Co-repressors
Mediator
Nuclear coactivators
RNA binding
Transcriptional initiation
Spindle formation
Ubiquination
Other
DUX4
BioID
AP-MS
Both
Prey-Prey interaction

Fig. 4

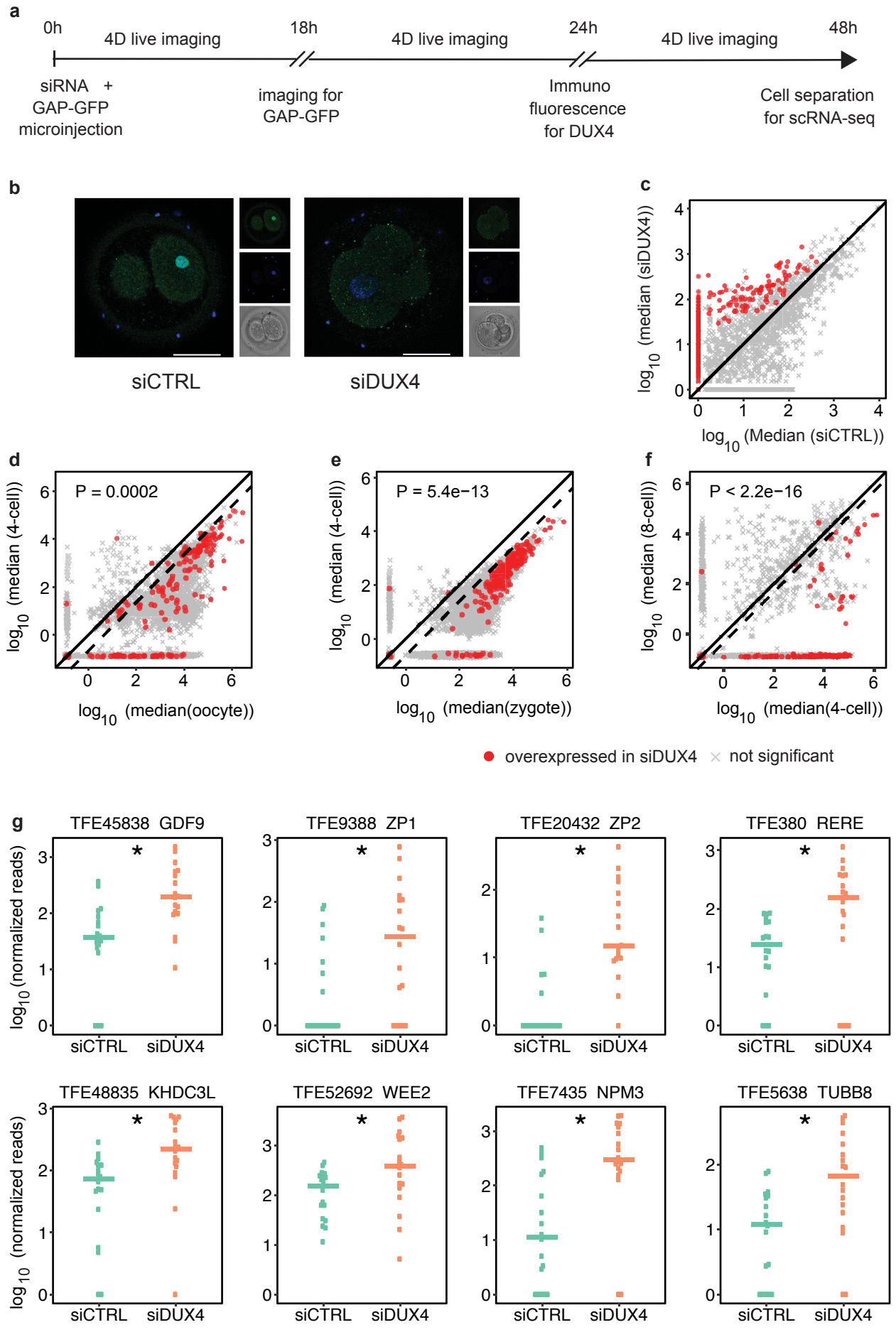

a
0h
4D live imaging
18h
4D live imaging
24h
4D live imaging
48h
siRNA + GAP-GFP microinjection
imaging for GAP-GFP
Immuno fluorescence for DUX4
Cell separation for scRNA-seq
b
siCTRL
siDUX4
c
log10 (median (siDUX4))
log10 (Median (siCTRL))
d
P = 0.0002
log10 (median (4-cell))
log10 (median(oocyte))
e
P = 5.4e−13
log10 (median (4-cell))
log10 (median(zygote))
f
P < 2.2e−16
log10 (median (8-cell))
log10 (median(4-cell))
overexpressed in siDUX4
not significant
g
TFE45838 GDF9
TFE9388 ZP1
TFE20432 ZP2
TFE380 RERE
TFE48835 KHDC3L
TFE52692 WEE2
TFE7435 NPM3
TFE5638 TUBB8
log10 (normalized reads)
siCTRL
siDUX4

Fig. 5

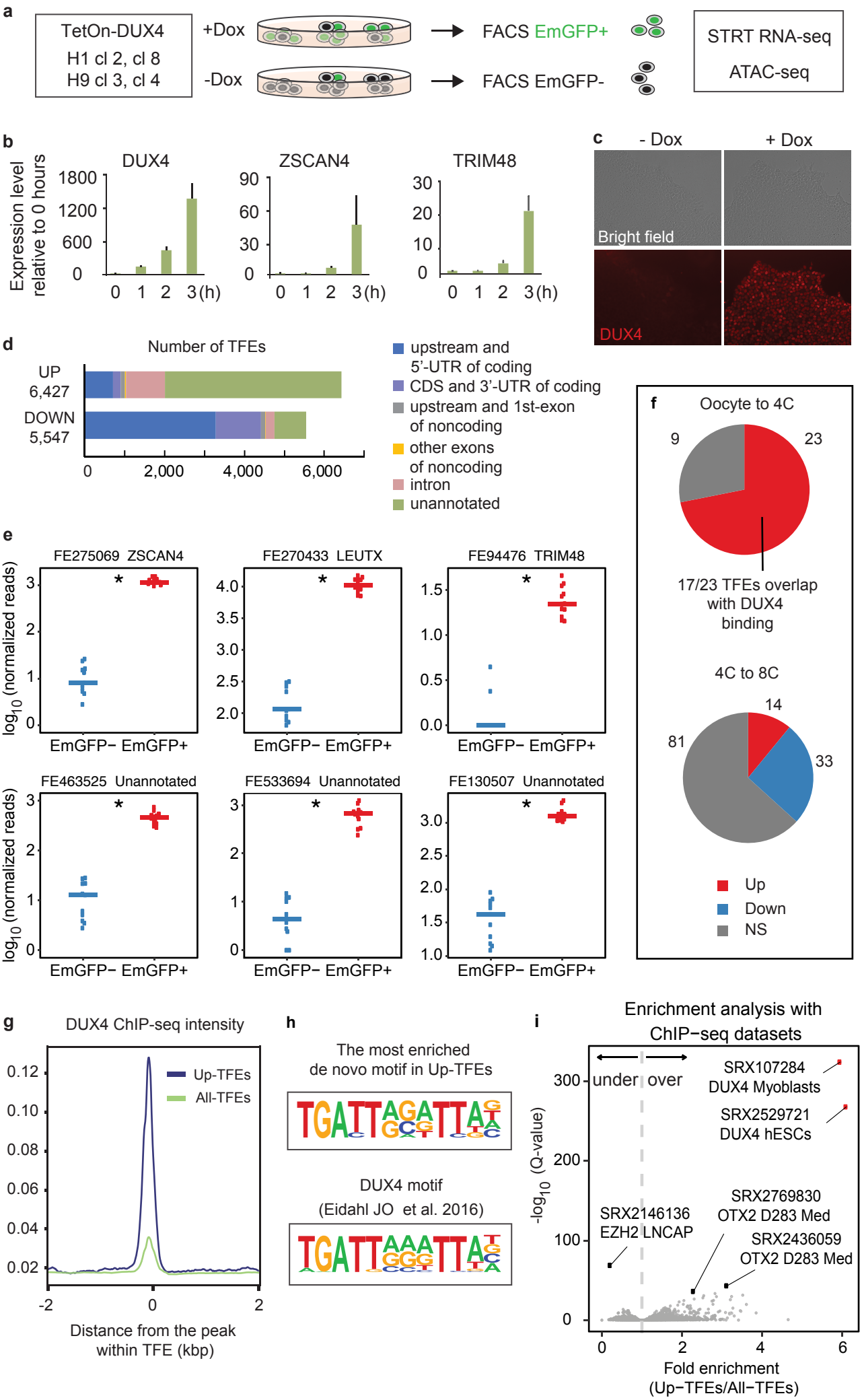
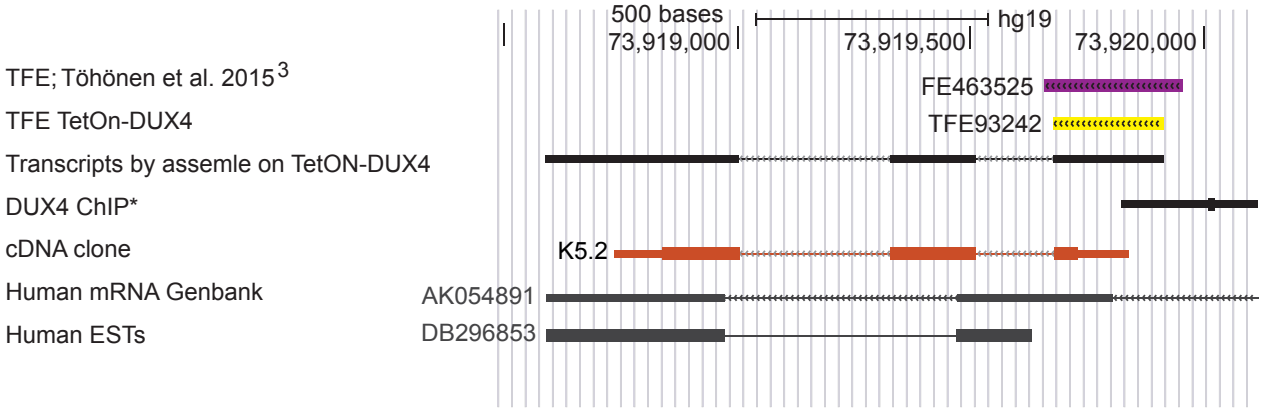

a
TetOn-DUX4
H1 cl 2, cl 8
H9 cl 3, cl 4
+Dox
-Dox
FACS EmGFP+
FACS EmGFP-
STRT RNA-seq
ATAC-seq
b
Expression level relative to 0 hours
DUX4
ZSCAN4
TRIM48
0 1 2 3(h)
c
- Dox
+ Dox
Bright field
DUX4
d
Number of TFEs
UP 6,427
DOWN 5,547
0 2,000 4,000 6,000
upstream and 5'-UTR of coding
CDS and 3'-UTR of coding
upstream and 1st-exon of noncoding
other exons of noncoding
intron
unannotated
e
FE275069 ZSCAN4
FE270433 LEUTX
FE94476 TRIM48
FE463525 Unannotated
FE533694 Unannotated
FE130507 Unannotated
log10 (normalized reads)
EmGFP− EmGFP+
f
Oocyte to 4C
9
23
17/23 TFEs overlap with DUX4 binding
4C to 8C
14
81
33
Up
Down
NS
g
DUX4 ChIP-seq intensity
Up-TFEs
All-TFEs
Distance from the peak within TFE (kbp)
h
The most enriched de novo motif in Up-TFEs
DUX4 motif
(Eidahl JO et al. 2016)
i
Enrichment analysis with ChIP−seq datasets
under
over
SRX107284 DUX4 Myoblasts
SRX2529721 DUX4 hESCs
SRX2769830 OTX2 D283 Med
SRX2146136 EZH2 LNCAP
SRX2436059 OTX2 D283 Med
-log10 (Q-value)
Fold enrichment (Up−TFEs/All−TFEs)

Fig. 6

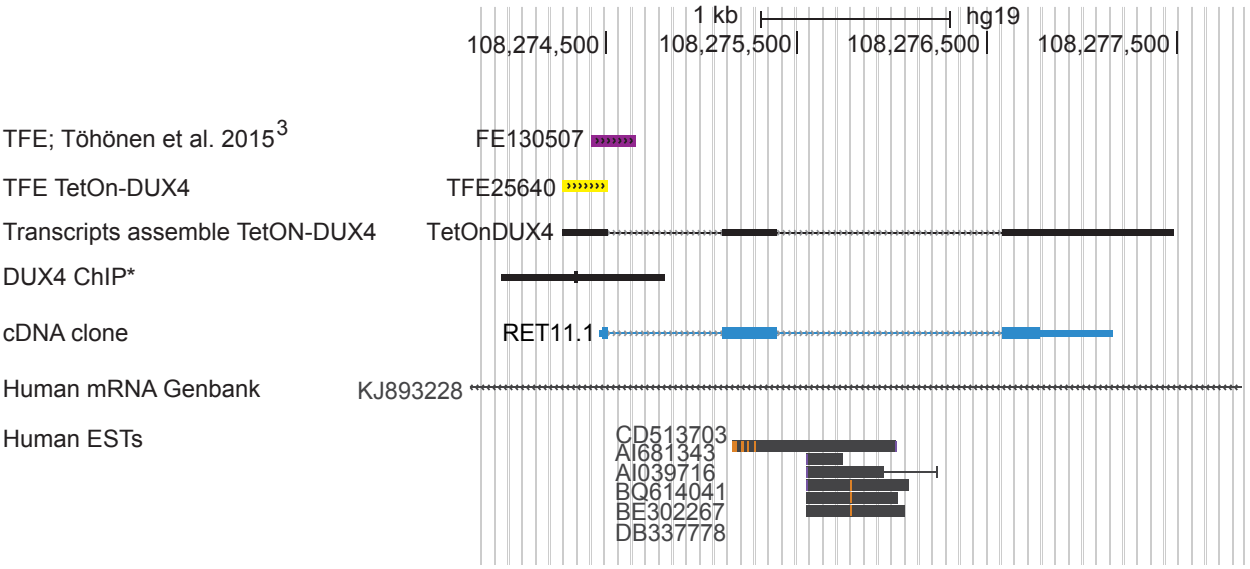
a

KHDC1 pseudo gene 1
ENA accession LR694084, LR694085



b

RING-finger type E3 ubiquitin ligase
ENA accession LR694082, LR694083



c

RING-finger domain protein
ENA accession:
LR694086, LR694087 (Ring10.2)
LR694088, LR694089 (Ring4.2)

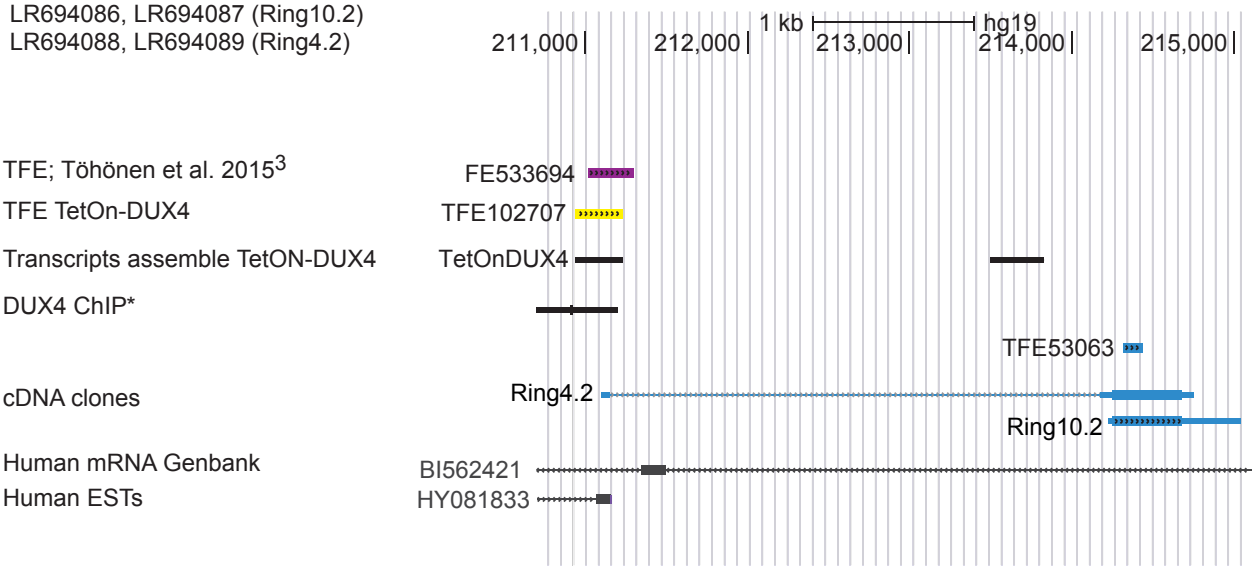
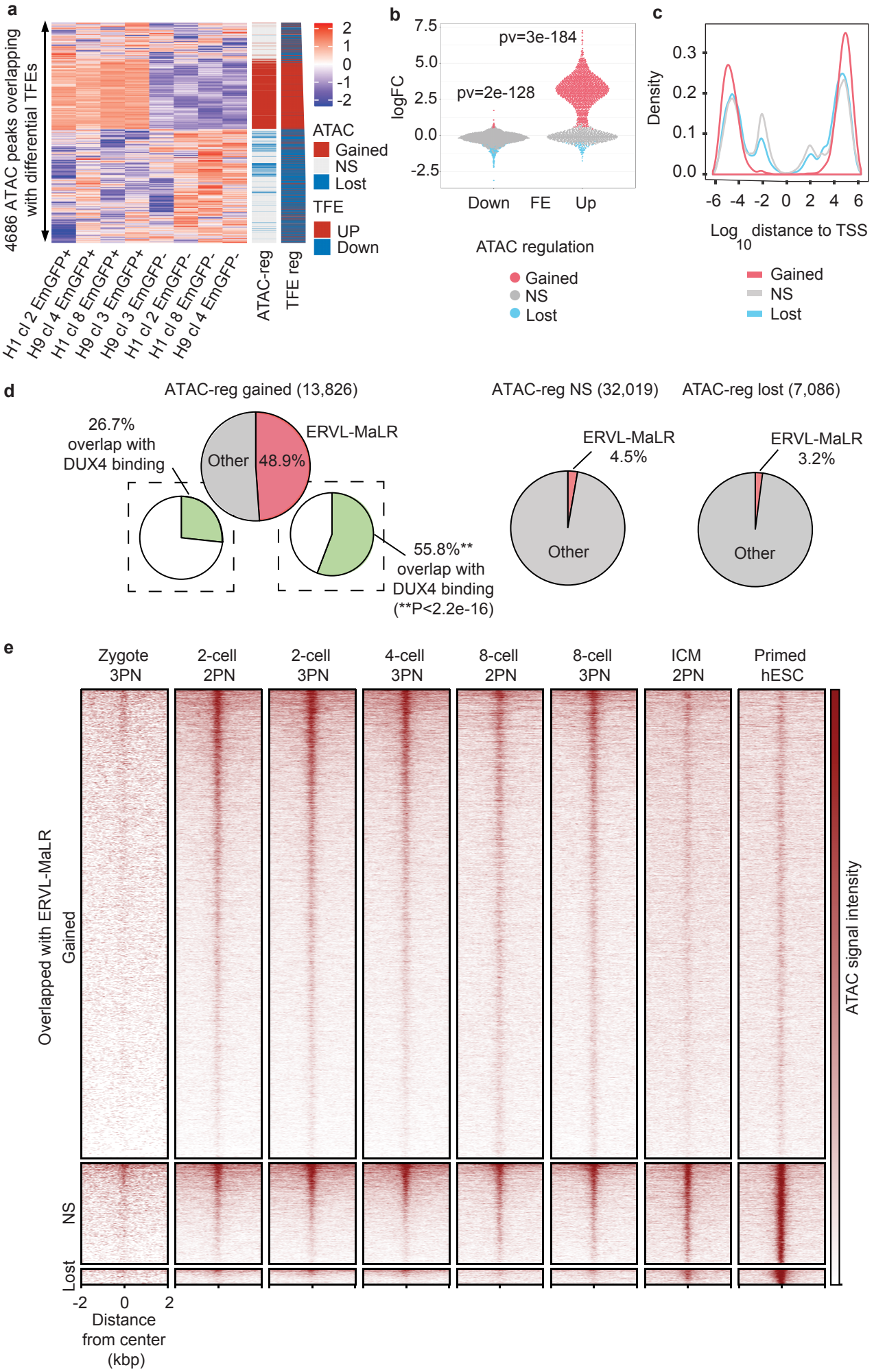


Fig. 7


a
4686 ATAC peaks overlapping with differential TFEs
H1 cl 2 EmGFP+
H9 cl 4 EmGFP+
H1 cl 8 EmGFP+
H9 cl 3 EmGFP+
H9 cl 3 EmGFP-
H1 cl 2 EmGFP-
H1 cl 8 EmGFP-
H9 cl 4 EmGFP-
ATAC-reg
TFE reg
2
1
0
-1
-2
ATAC
Gained
NS
Lost
TFE
UP
Down
b
logFC
7.5
5.0
2.5
0.0
-2.5
pv=3e-184
pv=2e-128
Down
FE
Up
ATAC regulation
Gained
NS
Lost
c
Density
0.3
0.2
0.1
0.0
-6 -4 -2 0 2 4 6
Log10 distance to TSS
Gained
NS
Lost
d
ATAC-reg gained (13,826)
ATAC-reg NS (32,019)
ATAC-reg lost (7,086)
26.7% overlap with DUX4 binding
Other
48.9%
ERVL-MaLR
55.8%** overlap with DUX4 binding (**P<2.2e-16)
ERVL-MaLR 4.5%
Other
ERVL-MaLR 3.2%
Other
e
Zygote 3PN
2-cell 2PN
2-cell 3PN
4-cell 3PN
8-cell 2PN
8-cell 3PN
ICM 2PN
Primed hESC
Overlapped with ERVL-MaLR
Gained
NS
Lost
ATAC signal intensity
-2 0 2
Distance from center (kbp)