

# 1 **DUX4 regulates oocyte to embryo transition in human**

2

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

10

11 <sup>1</sup>Karolinska Institutet, Department of Biosciences and Nutrition, Huddinge, Sweden

12 <sup>2</sup>University of Helsinki, Department of Obstetrics and Gynecology, Helsinki, Finland

13 <sup>3</sup>Helsinki University Hospital, Department of Obstetrics and Gynecology, Helsinki,  
14 Finland

15 <sup>4</sup>University of Helsinki, Institute of Biotechnology, Helsinki, Finland

16 <sup>5</sup>Karolinska Institutet, Bioinformatics and Expression Analysis Core Facility,  
17 Huddinge, Sweden

18 <sup>6</sup>RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

19 <sup>7</sup>Competence Centre for Health Technologies, Tartu, Estonia

20 <sup>8</sup>Stem Cells and Metabolism Research Program, University of Helsinki, Finland

21 <sup>9</sup>Folkhälsan Research Center, Helsinki, Finland

22 <sup>10</sup>Institut d'Optique Graduate School, CNRS - Université de Bordeaux, Talence,  
23 France

24 <sup>11</sup>Structural Bioinformatics Laboratory, Biochemistry, Faculty of Science and  
25 Engineering, Åbo Akademi University, Turku, Finland

26 <sup>12</sup>Oulu University Hospital, Oulu, Finland

27 <sup>13</sup>RIKEN Preventive Medicine and Diagnosis Innovation Program, Wako, Japan

28 <sup>14</sup>Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan

29 <sup>15</sup>IFOM, The FIRC Institute of Molecular Oncology, Milan, Italy

30 <sup>16</sup>University of Basel, Department of Biomedicine, Basel, Switzerland

31

32

33 <sup>#</sup>Current affiliation

34 <sup>\*</sup>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.**

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

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

activation of the retroelements and unannotated genomic regions. Our protein interaction data demonstrated that DUX4 interacts with transcriptional mediators, chromatin modifiers, and ubiquitinases expressed in the human oocytes and early embryos. We conclude that DUX4 regulates different aspects of the OET in humans by affecting mRNA degradation, transcriptional regulation and chromatin structure.

## Results

### *DUX4 is transiently expressed immediately after fertilization*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

We next integrated STRT RNA-seq and ATAC-seq data from EmGFP +/- sorted TetOn-Dux4 hESC. DUX4 expression lead to consistent modifications in chromatin accessibility across all studied clones (Fig 7a). The transcriptionally upregulated TFEs correlated with more accessible (gained) genomic regions in the DUX4 induced cells while the correlation was less obvious for the downregulated TFEs and less accessible (lost) chromatin (Fig 7a). We found that the gained chromatin regions correlated with upregulated TFEs and lost chromatin regions correlated with downregulated TFEs (Fig 7b), likely implying that the transcriptional downregulation induced by DUX4 expression is faster than the nucleosome-mediated closing of the chromatin (Fig 7a and b). We then asked how DUX4 expression modified openness of the different regions of the chromatin. In general, DUX4 rapidly caused chromatin remodelling, especially chromatin opening, far from transcription start sites (TSS), demonstrating that the TSSs seem to be less targeted by the DUX4 expression than other genomic regions (Fig 7c). We next focused on the ATAC-gained chromatin sites. These chromatin regions were remarkably enriched with DUX4 binding sites compared with unchanged regions ( $P < 2.2 \times 10^{-16}$ ). Out of the ATAC-gained chromatin sites, 48.9% overlapped with ERVL-MaLR elements and they were significantly enriched for the DUX4 binding sites compared with non ERVL-MaLR overlapping sites (55.8%  $P < 2.2 \times 10^{-16}$ ) (Fig. 7d). The ATAC-gained ERVL-MaLR regions remarkably overlapped

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

## Discussion

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

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

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

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

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

## Online methods

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

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

### *Bulk RNA-sequencing using STRT method*

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

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

#### *Preprocess of raw STRT RNAseq reads*

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

#### *Downstream STRT RNA-sequencing data analysis*

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

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

#### *Human ESC culture*

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

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

#### *Plasmid construction*

The full-length DUX4 (NM\_001293798.2) was synthesized and cloned between the SalI and BamHI sites of the pB-tight-hMAFA-ires-EmGFP-pA-PGK-Puro vector (a kind gift from Diego Balboa, Biomedicum Stem Cell Centre) at GenScript (Genscript, NJ, USA).

#### *Doxycycline-inducible DUX4 expressing human ESCs*

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

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

#### *cDNA cloning of unannotated DUX4 targets*

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

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

The sequence for the human DUX4 (Q9UBX2) protein was obtained from the UniProt database (The UniProt Consortium<sup>30</sup>), whereas other sequences were

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

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

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

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

### *Cell Culture and Affinity Purification*

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

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

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  $\mu\text{m}$  x 2 cm, 3  $\mu\text{m}$ , 100 Å; Thermo Scientific) and  
543 analytical column (Acclaim PepMap RSLC, 65  $\mu\text{m}$  x 15 cm, 2  $\mu\text{m}$ , 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  $\mu\text{l}$  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<sup>2</sup> 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  $\pm 15$  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

## *Identification of statistical confidence of interactions*

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

## *Overrepresentation Analysis*

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

## *Interaction network*

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

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

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

# *Fluorescence associated cell sorting*

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

# *ATAC-sequencing library preparation and data analysis*

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

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

### *ATAC-seq data analysis*

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

visualized with plotHeatmap from deepTools<sup>47</sup>. All the scripts and command line options can be provided upon request.

### *Immunocytochemistry of the human ESC*

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

### *Immunocytochemistry of the human embryos*

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

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

### *Imaging of the fixed human embryos*

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

### *Confocal microscopy image analysis*

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

### *Culture and microinjection of human embryos*

Human triploid zygotes were warmed using Gems Warming Set (Genea Biomedx) and cultured in G-TL medium (Vitrolife) in 6 %O<sub>2</sub> and 6 % CO<sub>2</sub> at 37°C. 12  $\mu$ l of either 20  $\mu$ M scrambled control siRNA (AM4611, Thermo Fisher Scientific) or

DUX4-targeting siRNA (cat. 4457308, Thermo Fisher Scientific) diluted in nucleotide-free H<sub>2</sub>O were mixed with total of 500 ng of GAP-GFP mRNA and centrifuged at maximum speed at 4°C for 10 min. The embryos were microinjected using Eppendorf microinjector and placed in G-TL medium in Geri dish for 3D time-lapse imaging (Geri incubator, Genea Biomedx, Australia).

### *Human embryo live imaging*

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

### *Ethical approvals*

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

### *Acknowledgements*

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

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

## Contributions

SV, SK and JK conceived and coordinated the study. YM, TRB, MV, MSJ, TT, SK and JK supervised the work in each contributing laboratory. Every author participated in either planning or conducting respective experiments and analyzing or interpreting the data. SV, MY, LG, VR, TA, MT, MSJ, JK wrote the manuscript. All authors 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<sub>2</sub> 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.

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

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

(e) Sequence comparison of residues forming inter-homeodomain contacts.

(f) Root-mean-squared fluctuations (RMSF) of the C  $\alpha$  atoms of the X-ray structure of DUX4 with (red curve) and without (blue curve) bound DNA during a 100 ns MD simulation. HD1 (blue), linker (magenta) and HD2 (gold).

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

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

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

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

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

(a) Schematic of the experimental set up.

(b) Human embryos immunostained with DUX4 antibody (green) 24 h after microinjection with either control siRNA (left panel) or DUX4 targeting siRNA (right panel). Nuclei counterstained with DAPI (blue). Overlay of a single representative z plane and the corresponding z-planes shown for DUX4 staining, nucleus and bright field channels on the right side of each overlay. Scale bar 50  $\mu$ m.

(c) Scatter plot of the expression levels of TFes across the siCTRL and siDUX4 embryos. Red dots represent significantly upregulated TFes by siDUX4 and grey crosses represent TFes, which showed no significant change. (d-f) TFes upregulated (red dots) or showing no significant change (grey crosses) by siDUX4 in the human OET transcriptome as in<sup>3</sup>. Comparisons in d: oocyte to 4-cell, e: zygote to 4-cell, and f: 4-cell to 8-cell. The dotted line marks the cell division effect on cellular RNA content. *P*-values were calculated with Fisher's exact test for the frequency of the siDUX4-upregulated TFes of the TFes normally downregulated during respective stages.

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

**Figure 5. Transcriptome and ChIP-seq analysis on the TetOn-DUX4 hESCs.**

(a) Schematic of the experimental set up.

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

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

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

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

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

**(g)** DUX4 ChIP-seq intensity<sup>26</sup> around the peaks of reads within the upregulated TFEs (blue) and all the detected TFEs (green).

**(h)** De novo motif enrichment analysis of the DUX4-induced TFEs. Top: the most significantly enriched motif ( $P = 1e-961$ ). Bottom: the best-matched known motif (DUX4 ChIP-seq of myoblasts: GSE75791<sup>29</sup>; matched score = 0.92).

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

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

# **Figure 6. Novel DUX4 targets.**

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

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

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

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

**(a)** Heatmap of the 4,686 ATAC-sequencing reads across all samples at ATAC-seq peaks that overlap with differentially regulated TFE reads. Counts for each peak were standardized across each sample (mean=0, sd=1). Samples and peaks were then clustered using hierarchical clustering. The Separate heatmap of the ATAC-seq shows if the changes in the heatmap are significant (red: ATAC-reg gained; blue: ATAC-reg lost; grey: ATAC-reg non-significant). TFE-Reg heatmap shows if the overlapping TFE site is upregulated or downregulated (red: upregulated; blue: downregulated).

**(b)** Quasi random plot showing the results of the differential peak analysis on the ATAC-seq and STRT-RNA-seq. Each point is an ATAC-seq peak. Analyses were carried out on peaks that were repeated at least three times. Red: ATAC-seq gained; blue: ATAC-seq lost; grey: ATAC-seq non-significant. Y-axis: The log fold change of the ATAC-seq reads in the DUX4 expressing versus the control samples. X-axis: The ATAC-seq peaks overlapping either with the down-regulated TFEs or upregulated TFEs.

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

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

866 (e) ATAC-reg intensity of human early embryo<sup>30</sup> around the gained, non-significant  
867 (NS), and lost ATAC-reg peaks by DUX4 induction which overlap with ERVL-  
868 MaLR elements.

869

870

# References

- 1 Jukam, D., Shariati, S. A. M. & Skotheim, J. M. Zygotic Genome Activation in Vertebrates. *Developmental cell* **42**, 316-332, doi:10.1016/j.devcel.2017.07.026 (2017).
- 2 Braude, P., Bolton, V. & Moore, S. Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature* **332**, 459-461, doi:10.1038/332459a0 (1988).
- 3 Tohonen, V. *et al.* Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. *Nature communications* **6**, 8207, doi:10.1038/ncomms9207 (2015).
- 4 Blakeley, P. *et al.* Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development* **142**, 3613, doi:10.1242/dev.131235 (2015).
- 5 Galan, A. *et al.* Functional genomics of 5- to 8-cell stage human embryos by blastomere single-cell cDNA analysis. *PloS one* **5**, e13615, doi:10.1371/journal.pone.0013615 (2010).
- 6 Stirparo, G. G. *et al.* Integrated analysis of single-cell embryo data yields a unified transcriptome signature for the human pre-implantation epiblast. *Development* **145**, doi:10.1242/dev.158501 (2018).
- 7 Boroviak, T. *et al.* Single cell transcriptome analysis of human, marmoset and mouse embryos reveals common and divergent features of preimplantation development. *Development* **145**, doi:10.1242/dev.167833 (2018).
- 8 De Iaco, A. *et al.* DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat Genet* **49**, 941-945, doi:10.1038/ng.3858 (2017).
- 9 Hendrickson, P. G. *et al.* Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERV1/HERV1 retrotransposons. *Nature genetics* **49**, 925-934, doi:10.1038/ng.3844 (2017).
- 10 Tohonen, V. *et al.* Transcriptional activation of early human development suggests DUX4 as an embryonic regulator. doi:<https://doi.org/10.1101/123208> (2017).
- 11 Yan, L. *et al.* Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nature structural & molecular biology* **20**, 1131-1139, doi:10.1038/nsmb.2660 (2013).
- 12 Dang, Y. *et al.* Tracing the expression of circular RNAs in human pre-implantation embryos. *Genome biology* **17**, 130, doi:10.1186/s13059-016-0991-3 (2016).
- 13 Xue, Z. *et al.* Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500**, 593-597, doi:10.1038/nature12364 (2013).
- 14 Whiddon, J. L., Langford, A. T., Wong, C. J., Zhong, J. W. & Tapscott, S. J. Conservation and innovation in the DUX4-family gene network. *Nat Genet* **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. *FL1000Research* **5**, 2748,  
944 doi:10.12688/fl1000research.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., Tohonon, V., Linnarsson, S. & Kere, J. SAMstr: 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., Dunder, 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).

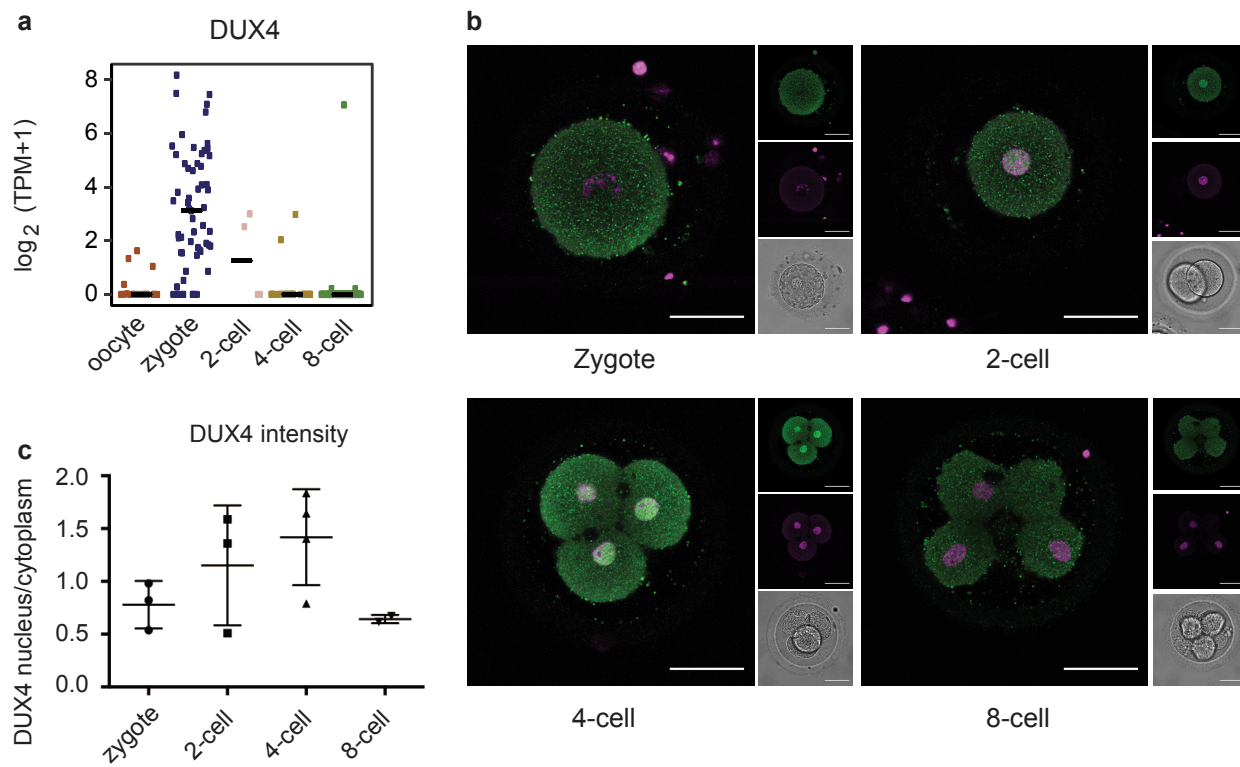
1077 70 Wang, J., Vasaikar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: a  
1078 more comprehensive, powerful, flexible and interactive gene set enrichment  
1079 analysis toolkit. *Nucleic acids research* **45**, W130-W137,  
1080 doi:10.1093/nar/gkx356 (2017).

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).

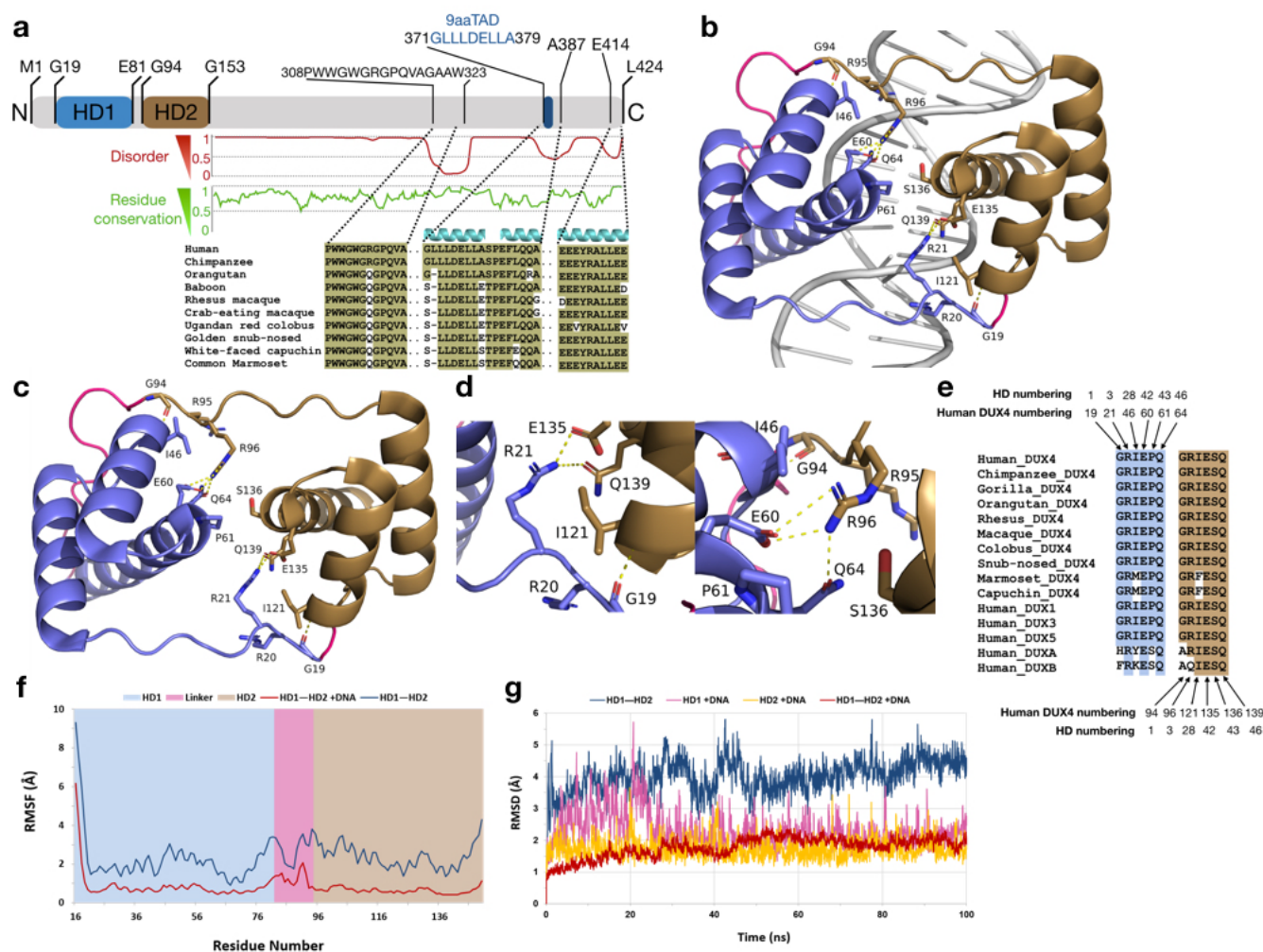
1084 72 Giurgiu, M. *et al.* CORUM: the comprehensive resource of mammalian  
1085 protein complexes-2019. *Nucleic acids research* **47**, D559-D563,  
1086 doi:10.1093/nar/gky973 (2019).

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).

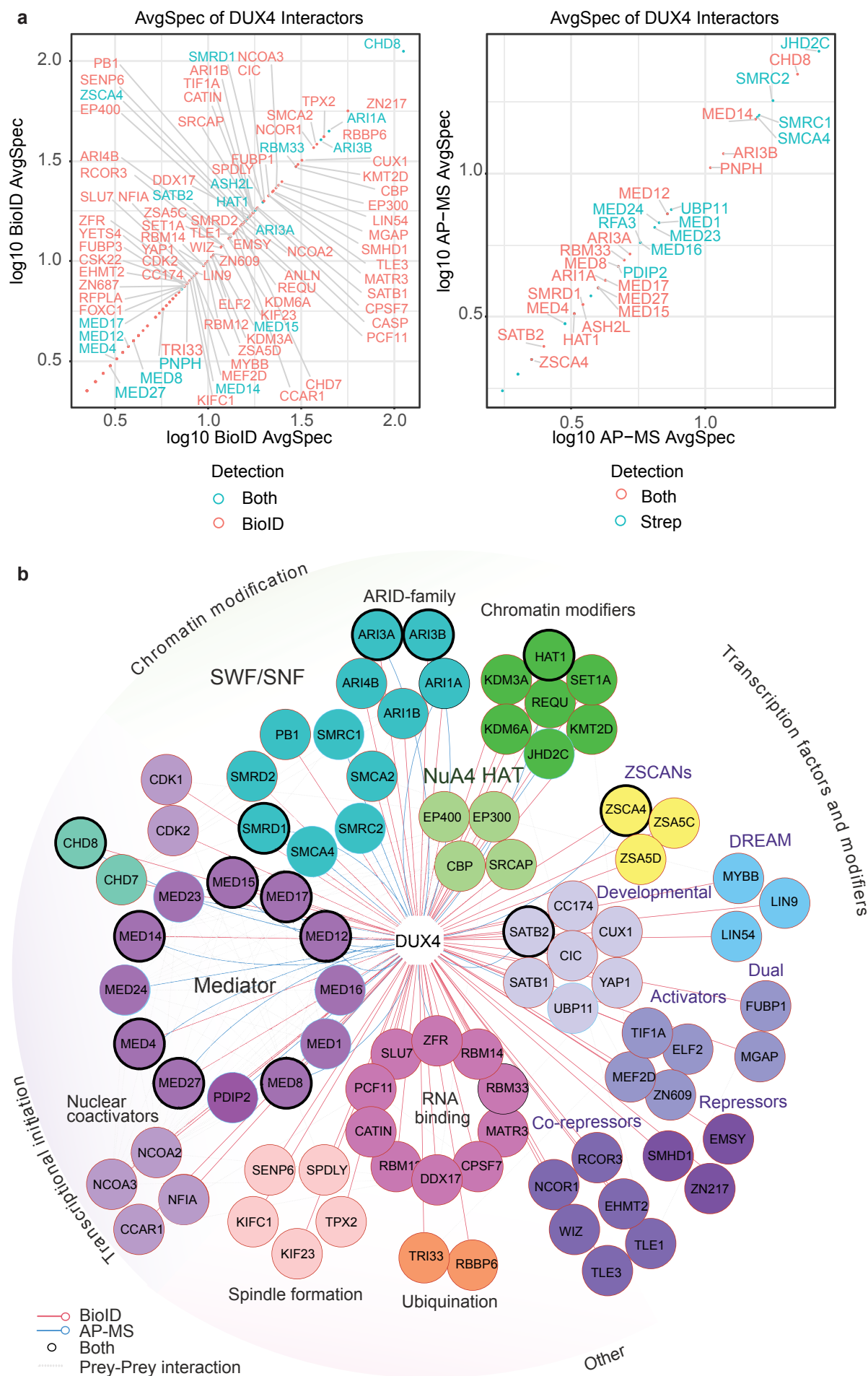
1091 74 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*  
1092 **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).  
1093



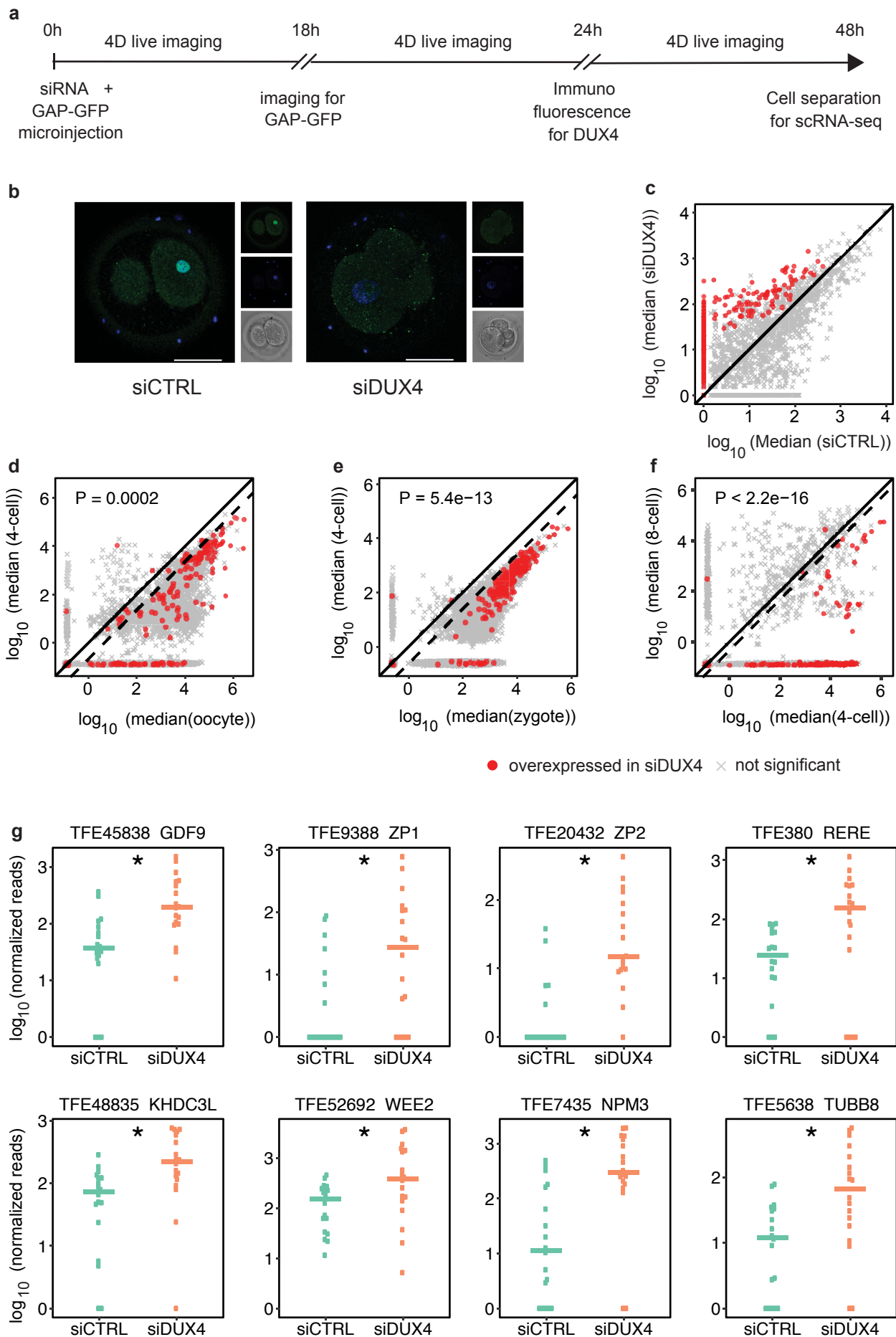
**Fig. 2**



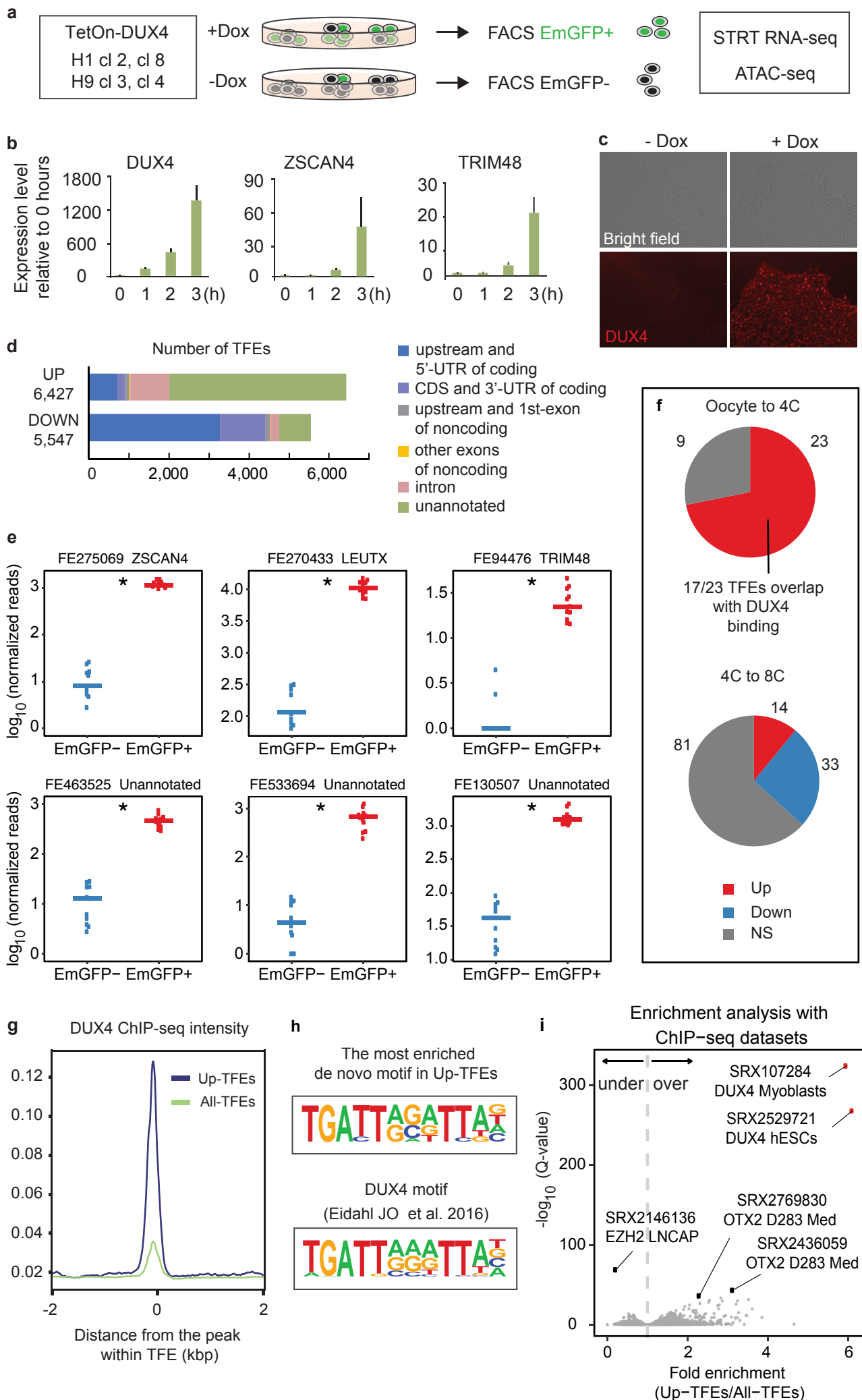
**Fig.3**



**Fig. 4**



## Fig. 5



## Fig. 6

**a**

KHDC1 pseudo gene 1

ENA accession LR694084, LR694085

TFE; Töhönen et al. 2015<sup>3</sup>

TFE TetOn-DUX4

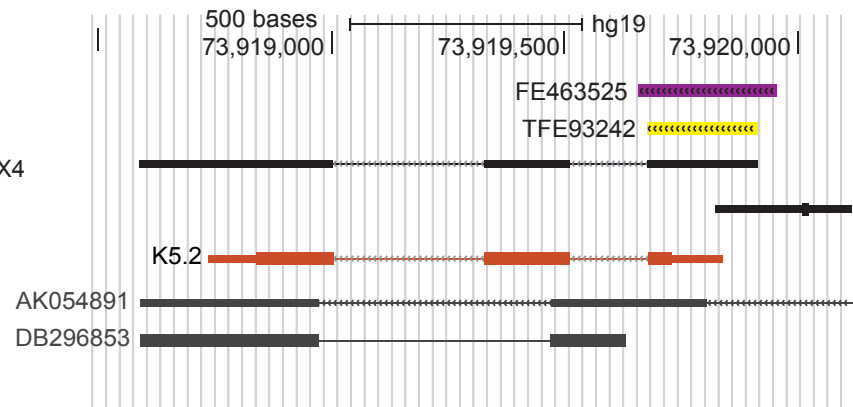
Transcripts by assemble on TetON-DUX4

DUX4 ChIP\*

cDNA clone

Human mRNA Genbank

Human ESTs



**b**

RING-finger type E3 ubiquitin ligase

ENA accession LR694082, LR694083

TFE; Töhönen et al. 2015<sup>3</sup>

TFE TetOn-DUX4

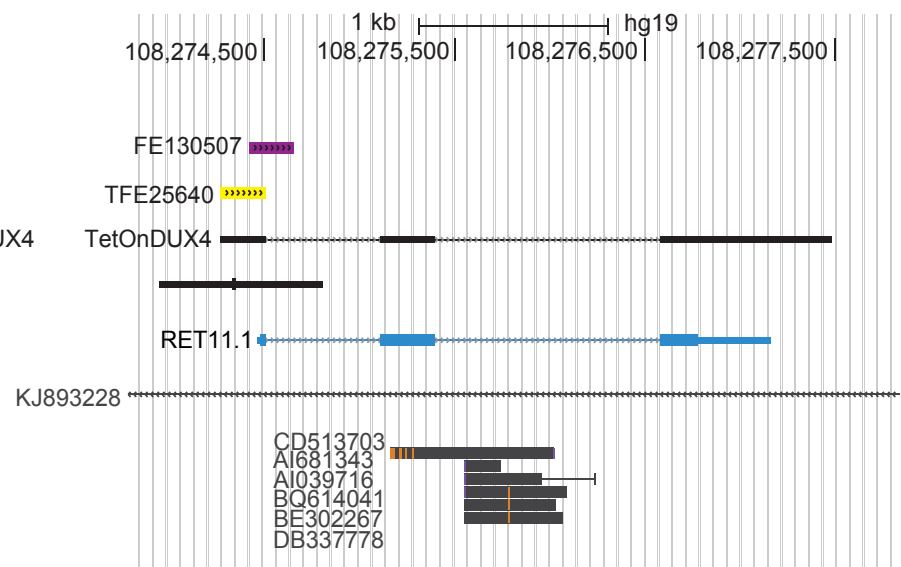
Transcripts assemble TetON-DUX4

DUX4 ChIP\*

cDNA clone

Human mRNA Genbank

Human ESTs



**c**

RING-finger domain protein

ENA accession:

LR694086, LR694087 (Ring10.2)

LR694088, LR694089 (Ring4.2)

TFE; Töhönen et al. 2015<sup>3</sup>

TFE TetOn-DUX4

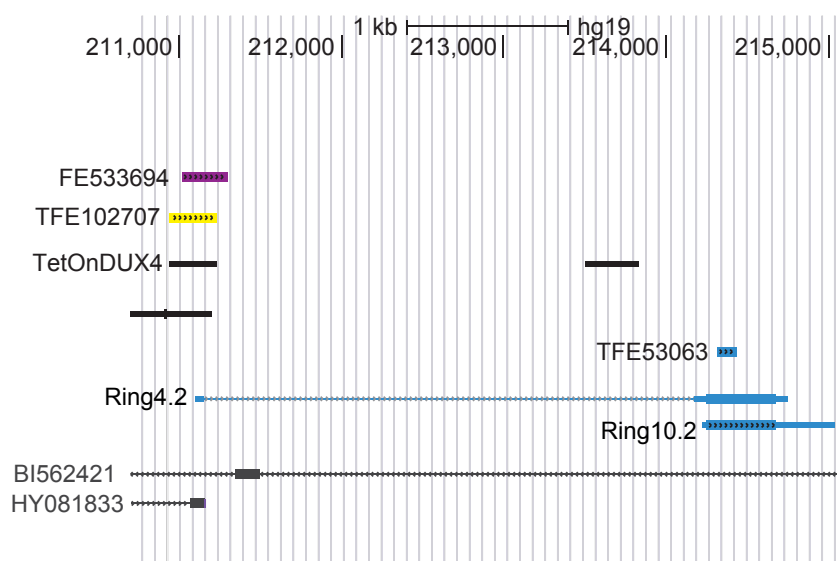
Transcripts assemble TetON-DUX4

DUX4 ChIP\*

cDNA clones

Human mRNA Genbank

Human ESTs



**Fig. 7**

