Differential regulation of lineage commitment in human and mouse primed pluripotent stem cells by NuRD

Ramy Ragheb¹, Sarah Gharbi¹, Julie Cramard¹, Oluwaseun Ogundele¹, Susan Kloet^{2,5}, Thomas Burgold^{1,6}, Michiel Vermeulen^{2,3}, Nicola Reynolds¹ and Brian Hendrich^{1,4}

1. Wellcome– MRC Stem Cell Institute, University of Cambridge, Cambridge CB2 0AW United Kingdom
2. Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University, 6525 GA Nijmegen, The Netherlands
3. Oncode Institute, Radboud University, 6525 GA Nijmegen, The Netherlands
4. Department of Biochemistry, University of Cambridge, Cambridge CB2 1QR United Kingdom
5. Present address: Leiden Genome Technology Center, Department of Human Genetics, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, The Netherlands
6. Present address: Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, United Kingdom

Running Title: NuRD function in primed pluripotency

Keywords: pluripotency, chromatin, lineage commitment, transcriptomics

Corresponding Author: Brian Hendrich, bdh24@cam.ac.uk, @BDH_Lab

24 **Abstract**

25 Differentiation of mammalian pluripotent cells involves large-scale changes in transcription and,
 26 among the molecules that orchestrate these changes, chromatin remodellers are essential to
 27 initiate, establish and maintain a new gene regulatory network. The NuRD complex is a highly
 28 conserved chromatin remodeller which fine-tunes gene expression in embryonic stem cells. While
 29 the function of NuRD in mouse pluripotent cells has been well defined, no study yet has defined
 30 NuRD function in human pluripotent cells. We investigated the structure and function of NuRD in
 31 human induced pluripotent stem cells (hiPSCs). Using immunoprecipitation followed by mass-
 32 spectrometry in hiPSCs and in naive or primed mouse pluripotent stem cells, we find that NuRD
 33 structure and biochemical interactors are generally conserved. Using RNA sequencing, we find that,
 34 whereas in mouse primed stem cells and in mouse naïve ES cells, NuRD is required for an
 35 appropriate level of transcriptional response to differentiation signals, hiPSCs require NuRD to
 36 initiate these responses. This difference indicates that mouse and human cells interpret and
 37 respond to induction of differentiation differently.

38

39 Graphical Abstract

40

41 **NuRD acts like a conductor in an orchestra. A.** In the presence of NuRD (pink blob figure, centre)

42 differentiation occurs in an ordered fashion in both mouse (left) and human (right) ES cells. Gene

43 expression changes in both cell types are tightly controlled with down-regulation of pluripotency

44 genes and up-regulation of lineage appropriate genes. This is akin to a group of musicians producing

45 musical notes in the right order and at the right amplitude to create a coherent piece of music. **B.**

46 Loss of “the conductor” NuRD results in increased transcriptional noise in both systems, indicated

47 here as a low-level blanket of sound in both systems. Consequences of MBD3/NuRD loss differs

48 between human and mouse ES cells. In mouse ES cells, differentiation cues lead to some down-

49 regulation of pluripotency genes and incomplete progression along a lineage appropriate pathway.

50 This is like musicians who know that they should be making music but who lose their way without a

51 conductor’s influence. In human iPS cells the background level of noise without NuRD results in a

52 lack of order to gene expression changes in response to differentiation. The noise from these

53 “musicians” would be truly awful.

54 Introduction

55 The identity of a eukaryotic cell is ultimately determined by its transcriptional output. The
 56 process by which cells transition from one state to another is therefore necessarily subject to tight
 57 transcriptional controls. For example, during development, in the absence of changes in external
 58 cues, transcriptional programs must remain stable for the identity of that cell to be maintained.
 59 Upon changes in external signals, transcription of some genes must be downregulated while that of
 60 others must be increased and this results in a change in cellular identity. The mechanisms which act
 61 either to maintain or change the expression state of a cell therefore underlie the ordered
 62 progression of transitions that occur throughout embryonic development. Failure of regulation of
 63 these gene expression patterns can prevent successful execution of developmental decisions,
 64 leading to developmental abnormalities, tumourigenesis or death. A comprehensive understanding
 65 of how cells control transcription during cell fate decisions is therefore critical for fields where it is
 66 desirable to control or instruct cell fate decisions, such as in regenerative medicine or cancer
 67 biology.

68 The ability of cells to activate or repress transcription relies largely on the conformation of
 69 the chromatin in which these genes reside. A set of chromatin remodelling complexes function to
 70 alter the structure of chromatin at regulatory elements to control gene expression (Hota & Bruneau,
 71 2016). One such complex in particular, the NuRD (**N**ucleosome **R**emodelling and **D**eacetylation)
 72 complex, is important for cells to undergo the changes in identity associated with the exit from
 73 pluripotency (Burgold, Barber et al., 2019, Kaji, Caballero et al., 2006, Reynolds, Latos et al., 2012a).
 74 The NuRD complex is a highly conserved multiprotein chromatin remodeller initially defined as a
 75 transcriptional repressor (Tong, Hassig et al., 1998, Wade, Jones et al., 1998, Xue, Wong et al., 1998,
 76 Zhang, LeRoy et al., 1998). NuRD activity facilitates cell fate transitions in a range of different
 77 organisms and developmental contexts (Signolet & Hendrich, 2015). The complex combines two
 78 enzymatic activities: class I lysine deacetylation, encoded by the Histone Deacetylase (Hdac) 1 and

79 2 proteins, with the Swi/Snf-type ATPase and nucleosome remodelling of the Chromodomain
80 Helicase DNA binding protein 4 (Chd4). This complex also contains histone chaperone proteins
81 Rbbp4 and 7, one of the zinc-finger proteins Gatad2a or Gatad2b, two MTA proteins (Mta1, Mta2,
82 and/or Mta3), Cdk2ap1 and Mbd2 or Mbd3 (Allen, Wade et al., 2013, Kloet, Baymaz et al., 2015,
83 Mohd-Sarip, Teeuwssen et al., 2017). Mbd3 is known to be required for lineage commitment of
84 pluripotent cells and is essential for early mammalian development (Kaji et al., 2006, Kaji, Nichols
85 et al., 2007). Structural and genetic work has found that Mbd3 physically links two biochemical and
86 functional NuRD subcomplexes: a remodelling subcomplex containing Chd4, the Gatad2 protein and
87 Cdk2ap1; and a histone deacetylase subcomplex containing the Hdac, the Rbbps and the Mta
88 proteins (Burgold et al., 2019, Low, Webb et al., 2016, Zhang, Aubert et al., 2016). Mbd3 thus acts
89 as a molecular bridge between these subcomplexes and maintains the structural integrity of NuRD.

90 In mouse ESCs (mESCs) Mbd3/NuRD activity modulates the transcription of pluripotency-
91 associated genes, maintaining expression within a range that allows cells to effectively respond to
92 differentiation signals (Bornelöv, Reynolds et al., 2018, Reynolds et al., 2012a). Despite profound
93 developmental defects, Mbd3 deficiency in mESCs results in only moderate gene expression
94 changes, with the majority of genes changing by less than two-fold. Rather than turning genes on
95 or off, Mbd3/NuRD activity serves to fine-tune gene expression in mESCs (Bornelöv et al., 2018).
96 Although this amounts to many small transcriptional changes, the cumulative effect of this is
97 nevertheless a profound phenotype: the inability of pluripotent cells to undergo lineage
98 commitment. While the function of Mbd3/NuRD in mouse pluripotent cells has been well defined,
99 no study yet has defined NuRD function in human pluripotent cells.

100 Human and mouse ESCs can both be derived from the inner cell mass (ICM) of pre-
101 implantation epiblasts. Yet the cell lines that emerge after culturing differ in transcriptomic,
102 epigenetic, and morphological features (Nichols & Smith, 2009). mESCs show early developmental
103 characteristics such as the expression of pluripotency genes, DNA hypomethylation and the activity

of both X chromosomes in females. Conventional human ESCs (hESCs) are developmentally more advanced and resemble murine post-implantation epiblast or mouse epiblast stem cells (mEpiSCs), and thus are considered to be primed pluripotent (Brons et al., 2007; Tesar et al. 2007). In this study we investigated the function of the NuRD complex in human primed pluripotent stem cells (hiPSCs), and compared this to its function in mEpiSCs. We find that while MBD3/NuRD is required in both systems for cells to properly undergo lineage commitment, the way in which this function is exerted appears different. Whereas in mouse primed stem cells, as in naïve mESCs, NuRD is required for an appropriate level of transcriptional response to differentiation signals, human cells require NuRD activity to initiate these transcriptional responses. This difference in the transcriptional consequences upon loss of an orthologous protein in two different mammalian pluripotent stem cell types indicates that mouse and human cells interpret and/or respond to induction of differentiation differently.

Results

NuRD complex structure is conserved in mouse and human pluripotent stem cells

In order to characterise human NuRD, we used genome editing to insert coding sequence for a 3xFLAG epitope immediately upstream of the stop codon of one endogenous *MBD3* allele in human iPS cells (Fig. S1A, B). An equivalent C-terminally tagged murine endogenous MBD3 protein shows genomic localisation identical to that found for wild type MBD3 protein in mouse ES cells, and supports normal embryonic development in mice (Bornelöv et al., 2018). Biochemical isolation of MBD3/NuRD in MBD3-3xFLAG hiPSCs, or in mEpiSCs containing an identically modified *Mbd3* allele, followed by mass spectrometry identified all known components of NuRD in both systems (Fig 1A, B). A number of interacting proteins were also purified at much lower stoichiometries than was seen for core NuRD components. Comparison of mass spectrometry data between hiPSCs, mEpiSCs and

128 mouse naïve ES cells (using MTA1-3 proteins for NuRD purification: (Burgold et al., 2019)) showed
 129 that most interacting proteins identified in human cells also interact with mouse NuRD (Fig 1C). Two
 130 cell-type specific interactors are VRTN and ZNF423, both of which are not expressed in naïve ES cells,
 131 but are found interacting with NuRD in primed PSCs (mEpiSCs and hiPSCs; Fig 1C). Two nuclear
 132 proteins were identified interacting with human NuRD that were not significantly enriched in the
 133 mouse datasets: PGBD3 and BEND3. PGBD3 is a transpose- derived protein expressed as a fusion
 134 with ERCC6 not present in mice (Newman, Bailey et al., 2008), but previously reported to interact
 135 with NuRD components in human cells (Hein, Hubner et al., 2015). Although not significantly
 136 detected in our mouse NuRD purifications, BEND3 has previously been shown to recruit NuRD to
 137 major satellite repeats in mouse cells (Saksouk, Barth et al., 2014). WDR5, ZNF296 and ZNF462 were
 138 identified interacting with mouse NuRD as described (Burgold et al., 2019, Ee, McCannell et al.,
 139 2017, Kloet, Karemaker et al., 2018), but were beneath our significance cutoff in purifications from
 140 human cells (Fig 1C). We therefore conclude that NuRD structure and biochemical interactors are
 141 generally conserved between mouse and human PSCs.

142 We next asked what function was played by MBD3/NuRD in human PSCs. To this end we used
 143 CRISPR/Cas9-mediated gene targeting to create an *MBD3*-KO iPSC line (Fig. 2A; Fig S1C).
 144 Immunoprecipitation with the remodelling subunit of NuRD, CHD4, allowed for purification of other
 145 complex components in wild type cells, but not in the *MBD3*-KO cells, indicating that human NuRD
 146 does not form without MBD3 (Fig 2B). Interactions between CHD4 and other NuRD components
 147 were restored when an MBD3 transgene was overexpressed in the null cells (“Rescue”, Fig 2A,B),
 148 indicating that the transgenic MBD3 is sufficient for NuRD formation as it is in mESCs (Bornelöv et
 149 al., 2018, Reynolds, Salmon-Divon et al., 2012b).

150 *MBD3*-KO hiPSCs were viable in standard culture conditions (mTESR or E8 (Chen, Gulbranson
 151 et al., 2011)), though unlike wild type cells null cultures showed some degree of spontaneous
 152 differentiation in both culture conditions (Fig 2C). While wild type and Rescue cultures presented as

157 proteins indicated in red. The human data comprise three independent immunoprecipitations from
 158 one preparation of nuclear extract, while the mouse data comprise three independent
 159 immunoprecipitations each from nuclear extract preparations made from two independent cell
 160 lines. C) Relative enrichment of indicated proteins normalised to the bait protein for each
 161 experiment: for mouse, ES cell data (taken from (Burgold et al., 2019); blue) were normalised to two
 162 MTA proteins, while for both the epiSC (purple) and hiPS (red) experiments the data were
 163 normalised to one MBD3 protein. Error bars represent standard deviation of three (hiPSCs), six
 164 (mEpiSCs) or nine (mESCs) replicates. Asterisks indicate situations where the protein enrichment
 165 was not significant in this cell type, but the stoichiometry is displayed for comparison.

166

167 morphologically homogeneous colonies with clear boundaries, mutant cultures were mix of cells
 168 showing a compact, undifferentiated morphology as well as a population of flatter, less dense
 169 colonies with irregular boundaries, reminiscent of differentiated cells (Fig. 2C). This was surprising
 170 since *Mbd3*-KO mESCs are resistant to differentiation (Kaji et al., 2006).

171 To determine whether human NuRD is required for lineage commitment of pluripotent cells,
 172 wild type, mutant and Rescued hiPSCs were induced to differentiate towards a neuroectodermal
 173 fate or a definitive endoderm fate (see Methods). After 20 days of neuroectodermal differentiation,
 174 axon-like extensions were readily identifiable in wild type and Rescue cultures (Fig. 2C). In contrast,
 175 no such appendages were found in *MBD3*-KO cultures, indicating a requirement for MBD3/NuRD
 176 for successful completion of this differentiation process. All three cell lines showed a decrease in
 177 expression of pluripotency markers across both differentiation protocols, indicating that NuRD is
 178 not required for PSC to respond to differentiation signals (Fig. 2D, E). *MBD3*-KO cells failed to
 179 properly induce expression of some lineage-appropriate genes in both differentiation protocols, but
 180 this ability was restored upon rescue with the MBD3 transgene (Fig. 2D, E). While NuRD is therefore
 181 required in human cells to faithfully maintain a self-renewing state, it is also required for appropriate
 182 lineage determination in these two differentiation protocols.

Ragheb et al. Fig. 2

183

184 **Figure 2. Human iPS cells lacking MBD3/NuRD fail to undergo programmed differentiation.** A)
 185 Western blot of nuclear extracts from wild type (WT), *MBD3*-KO and *MBD3*-KO hiPSCs rescued with
 186 an *MBD3*-3xFLAG transgene ("Rescue"). The blot was probed with antibodies indicated at left. The

closed arrowhead indicates native MBD3, while the open arrowhead indicates the MBD3-3xFLAG fusion protein. B) Nuclear extracts from wild type, *MBD3*-KO and Rescued cells was immunoprecipitated with anti-Chd4, western blotted and probed with antibodies indicated at right. Arrowheads as in Panel A. C) Scheme of the differentiation experiment (top) and images of indicated cell cultures at Day 0, 7 or 20 of differentiation. Scale bar indicates 100 μ m. D) Expression of pluripotency (*POU5F1*, *SOX2* and *NANOG*) and lineage specific genes (*FOXG1*, *PAX6* and *NESTIN*) during neural differentiation was measured by qRT-PCR. Y-axis represents expression relative to that in wild type cells at Day 0, while the X-axis represents the time in days. Error bars represent the standard deviation of ≥ 3 biological replicates. E) as in panel D, but for definitive endoderm differentiation protocol. Pluripotency-associated genes (*POU5F1*, *NANOG*, *ZFP42*) on the left, and differentiation-associated genes (*GATA4*, *FOXA2*, *SOX17*) at right.

NuRD mutant hiPS cells are unable to maintain a stable pluripotency state

To determine how NuRD facilitates lineage commitment in hiPSCs, we analysed and compared the transcriptomes of WT, *MBD3*-KO and Rescue cells at 0, 3, 6 and 12 days upon neuroectodermal differentiation. Visualising the data using a multidimensional scaling plot (Ritchie, et al., 2015) separated each sample along the time of differentiation, represented by PC1, and the genotype, represented by PC2 (Fig. 3A). Data from WT and Rescue cells clustered close to each other and followed a similar developmental trajectory, indicating that overexpression of *MBD3* does not dramatically impair early stages of neural differentiation. In contrast, data from *MBD3*-KO cells clustered separately, indicating that they are undergoing aberrant differentiation, consistent with our RT-qPCR data (Fig. 2D). At day 0 NuRD mutant cells occupy a position further along the differentiation trajectory (PC1) than do either WT or Rescue cells, likely resulting from the presence of morphologically differentiated cells within the self-renewing *MBD3*-KO cultures (Fig. 2C).

To try to understand why *MBD3*-KO PSC were unable to stably maintain an undifferentiated state, the transcriptomes of WT and *MBD3*-KO PSCs were compared in self-renewing conditions (Fig. 3B). Null cells showed 823 differentially expressed genes compared to wild type cells (246 up- and 577 down-regulated, FDR 1%). GO term enrichments performed on the set of down-regulated

Ragheb et al. Fig. 3

215

216 **Figure 3. NuRD is required to maintain a stable pluripotency state in hiPSCs** A) MDS plot made
 217 from RNA-seq data of wild type, *MBD3*-KO and Rescued hiPSCs across a neural differentiation
 218 timecourse. Each point represents a biological replicate, and shapes indicate the days of

differentiation. B) Heat map of genes found to be differentially expressed between WT and *MBD3*-KO cells in self-renewing conditions (day 0; FDR 5%). C) GO terms associated with genes significantly activated (open bars) or repressed (filled bars) in *MBD3*-KO cells relative to WT hiPSCs in self-renewing conditions.

genes showed terms related to pluripotency (Fig 3C, Table S1) consistent with a failure to maintain a stable pluripotent state in the absence of MBD3/NuRD. Expression profiles of pluripotency markers (*NANOG*, *TDGF1*, *FOXD3* and *FGF2*) showed a significant down-regulation in mutant cultures when compared to the WT or Rescue cells (Table S1). The 246 up-regulated genes showed enrichment for terms related to the development of different lineages (Fig 3C, Table S1). Given that *MBD3*-KO cells failed to undergo programmed neural or endodermal differentiation despite precociously expressing differentiation markers, we conclude that, like in mESCs (Burgold et al., 2019), human NuRD functions to prevent inappropriate gene expression in undifferentiated pluripotent cells, and this noise reduction function is important for faithful execution of lineage decisions.

Human NuRD activity is required for appropriate transcriptional response to differentiation signals

To better understand how human NuRD facilitates lineage commitment, we next asked how gene expression changes during the differentiation time course differed in *MBD3*-KO cells as compared to WT or Rescue cells. By considering both genotype and the time of differentiation in our differential analysis, we identified genes showing expression changes in at least one cell line compared to the others during differentiation. Clusters of co-expressed genes were identified using K-means clustering (Soukas, Cohen, Socci, & Friedman, 2000), resulting in 6 groups showing similar expression profiles (Figure 4A).

Cluster 1 is composed of 403 genes down-regulated during normal differentiation (Fig. 4A, B; Table S2, S3 and S4). These genes were generally underexpressed in *MBD3*-KO hiPSCs, yet become further down-regulated as cells are subjected to differentiation conditions. This cluster includes pluripotency-associated genes such as *POU5F1*, *NANOG*, *FOXD3*, *TDGF1*, *FGF2*, *ZSCAN10*, *DPPA4* and *PRDM14*, validating and extending our conclusion drawn from data shown in Fig. 3 that *MBD3*-KO hiPSCs have a defect in maintaining the self-renewing state in standard conditions. Transcription factor binding sites enriched within 10Kb of the TSS of genes in this cluster showed significant enrichment of consensus binding sites for general transcription factors associated with activation of pluripotency gene expression (i.e. MYC, ATF, CEBPB; Table 1), consistent with a decrease in expression of pluripotency-associated genes. This analysis additionally identified consensus binding sites for SNAIL and ZEB1 (Table 1), transcription factors associated with epithelial to mesenchymal transition as well as repression of pluripotency gene expression (Jiang, Yan et al., 2018, Moreno-Bueno, Portillo et al., 2008). Human ES cells undergo EMT as part of the differentiation process (Kim, Yi et al., 2014), so this likely results from inappropriate expression of epithelial genes which would normally precede an EMT event.

Genes in clusters 2-4 are predominantly associated with GO terms involved in differentiation (Fig. 4B; Table S2, S3 and S4). Cluster 2 contains genes associated with neuroectodermal differentiation and were induced in wild type and Rescue cells. While Cluster 2 genes (including *PAX6*, *OTX1* and *SOX1*) showed inappropriate expression in *MBD3*-KO cells at time 0 and remained high throughout the differentiation time course, Cluster 3 genes, some of which are associated with neuronal maturation (such as *PLP1*, *SEMA3A*, and *APP*) were expressed at a lower level in mutant cells than in either WT or Rescue cells. Cluster 4 contains genes not induced in either WT or Rescue cells, but which showed inappropriate expression in *MBD3*-KO cells at all time points (e.g. *WNT5A*, *FOXA2*, *PAX7*; Fig 4B, Table S2). Cluster 5 contains only 27 genes which fail to be appropriately silenced during differentiation in *MBD3*-KO cells, but show no significant enrichment with any GO

term, and genes in Cluster 6 show similar expression patterns in mutant and Rescue cells, and are hence unlikely to contribute to the differentiation failure phenotype of *MBD3*-KO cells (Fig. S2). Genes in Clusters 2 and 4 are not associated with any specific TF binding sites, while Cluster 3 genes show enrichment of binding sites for the general transcription factor SRF (Table 1). This lack of evidence for misregulation of a specific transcriptional programme indicates that *MBD3*-KO hiPSCs fail to interpret a range of different differentiation signals, as opposed to just one or two main pathways. The silencing of pluripotency-associated genes (Cluster 1) in *MBD3*-KO cells demonstrates an ability of these cells to respond to the loss of self-renewal signals, yet the failure of genes in Clusters 2 and 5 to appropriately change expression during the differentiation time course indicates that NuRD is required for cells to properly respond to differentiation signals.

MBD3/NuRD controls lineage commitment differently in human and mouse primed PSC

Mbd3/NuRD facilitates exit from the self-renewing state in mouse naïve ESCs (Burgold et al., 2019, Kaji et al., 2006) and thus it was surprising that human primed PSC required MBD3/NuRD to properly maintain the self-renewing state. To determine whether this was a difference between human and mouse cells, or between naïve and primed pluripotent cells we next asked whether NuRD was required to maintain the self-renewing state in mEpiSCs. *Mbd3*^{-/-} mEpiSCs were derived in culture from *Mbd3*^{-/-} mESCs or by transient Cre expression in mEpiSCs derived from *Mbd3*^{Flox/-} mESCs (See Methods). *Mbd3*^{-/-} mEpiSCs derived through either method were indistinguishable, and appeared uniformly undifferentiated (Fig. S3A), indicating that the spontaneous differentiation seen in hiPSCs does not reflect a general requirement for MBD3/NuRD to maintain primed PSC.

To further compare mouse and human primed PSC, gene expression was monitored across a neural differentiation time course in both Floxed and *Mbd3*^{-/-} mEpiSCs by RNA-seq (Fig. 5A; Table S5, S6 and S7). When the data are visualised using a multidimensional scaling plot (Chen, Lun, & Smyth, 2016), each sample separates along PC1, representing time of differentiation, and PC2,

Ragheb et al. Fig. 4

294

295 **Figure 4. NuRD is required for transcriptional responsiveness in hiPSCs.** A) Scheme of the
296 experiment. Wild type (WT), *MBD3*-KO or Rescued cells maintained in mTESR1 media were
297 subjected to neural differentiation. Cultures were sampled at indicated time points for RNA-seq,

298 leading to the identification of 1150 annotated differentially expressed genes (DEG; FDR 1%).
 299 Kmeans clustering of genes by expression pattern led to the heat map shown at right, with six major
 300 gene clusters. B) Mean expression for genes in each cluster is displayed across the differentiation
 301 time course for each cell line. Error bars indicate standard deviations of average expression. The
 302 four most significant GO terms associated with each cluster are plotted as solid bars, while up to
 303 two pathways ($P_{adj} \leq 0.01$) are also plotted in open bars. A full list of GO terms and pathways is
 304 available in Table S4 and S5.

305

306 representing genotype (Fig 5B). Control and *Mbd3*^{-/-} cells occupy the same position along the
 307 differentiation trajectory (PC1), consistent with our observation that mEpiSCs do not require
 308 MBD3/NuRD to maintain a morphologically undifferentiated state.

309 As with the hiPSCs, we considered both the genotype and the time of differentiation in our
 310 analysis and thus identified 699 differentially expressed genes. Co-expressed genes were grouped
 311 using K-means clustering, resulting in 4 clusters showing similar expression profiles (Fig. 5C). In
 312 contrast to the human cells which generally showed a lack of transcriptional response to the
 313 differentiation time course, clusters identified in mouse cells showed transcriptional responses, but
 314 these responses differed from those in wild type cells (Fig. 5D). Cluster 1 genes showed decreased
 315 expression in wild type cells across the time course, but increased expression in *Mbd3*^{-/-} cells. Genes
 316 in clusters 2 and 4 also show increases in both wild type and mutant cells, but Cluster 2 genes
 317 showed a reduced response in mutant cells, whereas cluster 4 genes showed an increased response.
 318 Cluster 3 genes decreased in expression in mutant cells across the time course, but showed no
 319 overall change in wild type cells. In all clusters the genes are associated with very general GO terms,
 320 and are unlikely to represent individual pathways or developmental trajectories (Fig. 5D; Table S6
 321 and S7). These data indicate that in mouse EpiSC Mbd3/NuRD is not strictly required for cells to
 322 respond to differentiation signals as was seen in hPSC, but rather is required for an appropriate level
 323 of response, as it is in naïve mESC (Burgold et al., 2019). Rather than being required for the

Figure 5. NuRD facilitates an appropriate transcriptional response in mEpiSCs. A) An outline of the experiment, as in Figure 4. B) MDS plot of gene expression data collected across the neural differentiation time course, as in Figure 3A. C) Heat map of DEG (FDR 1%) separated into four

328 clusters by K-means clustering. D) Mean expression and most significant GO terms for each cluster
329 as in Figure 4B. A full list of GO terms and pathways is available in Table S6 and S7, and expression
330 plots and GO terms associated with clusters 5 and 6 are shown in Fig. S2.

331

332 transcriptional response to differentiation cues as in the human differentiation course, Mbd3/NuRD
333 functions in mouse primed PSC to facilitate an appropriate transcriptional response to neural
334 induction.

335 Despite the differences in NuRD-dependent gene regulation observed in mouse and human
336 primed PSC described thus far, we asked whether there could be a conserved core set of genes
337 regulated similarly in primed PSC from both species, which might contribute to the shared
338 requirement for NuRD in lineage commitment. Comparing gene expression datasets between the
339 human and mouse experiments (Fig. 6A) identified 153 genes, the orthologues of which were
340 differentially expressed in both human and mouse cells. K-means clustering using expression data
341 from the human cells segregated the genes into three main clusters (Fig. 6; Table S8). Orthologues
342 defined in human clusters showed 3-4 different gene expression profiles in mouse cells (Table S10),
343 generally showing a reduced or inappropriate transcriptional response in the absence of
344 Mbd3/NuRD (Fig. 6). One exception is a subset of thirteen Cluster 2 genes (excluding *Mbd3*) which
345 are rapidly activated in wild type cells but fail to be activated in mutant cells across the time course
346 (human Cluster 2; corresponding mouse subcluster 1: Fig. 6; Table S10). Notably these genes are
347 underexpressed in mutant mEpiSCs in self-renewing conditions, but overexpressed in hiPSCs,
348 indicating that they are not regulated similarly in the two cell types. We therefore conclude that the
349 transcriptional consequences of Mbd3/NuRD loss are different in human versus mouse primed PSC,
350 but in both cases this activity is required for cells to properly undergo lineage commitment.

Ragheb et al. Fig. 6

351

352 **Figure 6. MBD3/NuRD deficiency elicits a different response in human and mouse primed**
353 **pluripotent stem cells.** A) Schematic of the experiment: the Venn diagram shows the overlap of
354 differentially expressed genes identified in human cells, and the identified human orthologues of

those identified in mEpiSCs. K-means clustering of this set of genes in human data led to the formation of three gene clusters. B). GO terms (“Functional annotation”; $P_{\text{adj}} \leq 0.01$) and expression profiles are shown for the human gene clusters. Error bars indicate standard deviations of average expression. K-means clustering has been performed on the mouse gene orthologues of each human cluster resulting in 3 subclusters related to human cluster 1, 3 subclusters related to human cluster 2 and 4 subclusters related to human cluster 3. Heat maps of expression for orthologues in the mouse differentiation experiment are shown at right. There were no significant GO terms associated with Cluster 1 genes.

Discussion

Differentiation of mammalian pluripotent cells involves large-scale changes in transcription, which result in loss of one cell identity and gain of a new, more differentiated identity. Orchestrating these changes in transcription are a large cast of different transcription factors and signalling molecules, but there is also a set of chromatin remodellers whose activity is essential to initiate, establish and maintain a new gene regulatory network (GRN) (Gokbuget & Blelloch, 2019, Hota & Bruneau, 2016). When induced to differentiate, both mouse and human pluripotent stem cells depend on NuRD activity to elicit an appropriate transcriptional response and undergo lineage commitment, but the manner in which NuRD is used to facilitate this response differs. Induction to differentiate elicits changes in transcription from a range of genes in both human and mouse PSC. The absence of MBD3/NuRD activity in human cells results in a subset of these genes failing to respond to the differentiation cues, while in mutant mouse PSC the response is present and widespread, but often muted or inappropriate (Figs 4, 5). NuRD activity is additionally required to maintain the pluripotency GRN of hiPSCs cultured in self-renewing conditions, but neither primed nor naïve mouse PSCs display this requirement (Fig. S3A and (Kaji et al., 2006)). We see no large-scale differences in the biochemical make-up of NuRD between human and mouse primed stem cells, which would be consistent with the human and mouse complexes exerting similar, or identical

381 biochemical functions. The observed differences in the consequences of MBD3 deficiency are
382 therefore likely to result from subtle differences in how NuRD activity is used by the cells to respond
383 to changes in environment.

384 One example of how human and mouse cells respond differently to loss of MBD3/NuRD is in
385 regulation of the *ZEB1/Zeb1* genes. *ZEB1* is overexpressed in self-renewing human PSC and remains
386 high through the neural induction time course, whereas in mouse cells *Zeb1* is underexpressed in
387 self-renewing cells and fails to be activated during the time course (Fig. S4). *ZEB1* has been shown
388 to repress polarity and gap-junction genes associated with an epithelial morphology, promoting an
389 epithelial to mesenchymal transition (EMT) (Aigner, Dampier et al., 2007). This function is required
390 for neural differentiation in vivo and from hESCs in culture (Jiang et al., 2018, Singh, Howell et al.,
391 2016). It is not surprising, then, that Human Cluster 1 genes, which show underexpression
392 throughout the differentiation time course in mutant cells and show enrichment for cell adhesion
393 genes ($p = 2 \times 10^{-3}$; Table S4 and S5), are also enriched for *ZEB1* DNA binding motifs (Fig. 4B; Table 1).
394 In mouse cells, however, *Zeb1* motifs were associated with the cluster of genes highly associated
395 with cell-cell junctions and showing inappropriately high expression levels at all time points (Cluster
396 1: Fig. 5D and Table 1), consistent with aberrantly low expression of the *Zeb1* repressor. While
397 transgenic overexpression of *ZEB1* was reported to increase neural differentiation of hESCs (Jiang
398 et al., 2018), it did not lead to precocious differentiation of self-renewing hESCs, and hence is
399 unlikely to be the principal factor behind the precocious differentiation seen in hiPSCs lacking MBD3.

400 It is possible that differences in transcriptional responses to differentiation in human and
401 mouse cells could be due to the fact that, unlike mouse cells, human PSC are unable to maintain a
402 stable self-renewing state in the absence of MBD3/NuRD, and are, in effect, responding to loss of
403 self-renewal conditions when they have already started to differentiate. One possible, trivial
404 explanation for this difference in the ability of mouse and human PSC to self-renew could be due to
405 differences in the constituents of media used for self-renewal culture. Both mEpiSC culture and

406 hiPSC culture rely on FGF2 and activation of SMAD2/3 through addition of Activin or TGF β (Brons,
407 Smithers et al., 2007, Chen et al., 2011, Tesar, Chenoweth et al., 2007), while naïve mouse ES cells
408 are maintained through LIF signalling and dual inhibition of GSK3 and MEK/ERK (Ying, Wray et al.,
409 2008). One consistent difference between mouse and human PSC culture media is the inclusion in
410 human media of ascorbic acid (Vitamin C). Ascorbic acid has been shown to increase the activity of
411 TET enzymes, which promote the demethylation of 5-methylCytosine in DNA, though this has been
412 shown to promote a more naïve state, rather than promote differentiation (Blaschke, Ebata et al.,
413 2013, Yin, Mao et al., 2013). *Mbd3*-KO mESCs contain a reduced amount of DNA methylation relative
414 to wild type cells (Latos, Helliwell et al., 2012), consistent with them being less able to differentiate.
415 It is therefore unlikely that an increase in TET enzyme activity would be behind the precocious
416 differentiation seen in hPSC cultures. Rather, we suggest that the differences observed between
417 human and mouse PSC in self-renewal or the ability to initiate an appropriate developmental
418 response are most likely due to differences between the two species. As pointed out previously
419 (Takashima, Guo et al., 2014), primates have not evolved the ability to undergo embryonic diapause
420 (Nichols & Smith, 2012), and hence pluripotency may be a less stable state in humans than in mice,
421 and consequently be less tolerant to the loss of a major chromatin remodelling complex such as
422 MBD3/NuRD.

423 The mouse and human NuRD complexes present in primed PSC appear to be biochemically
424 very similar, and our methods identified no notable species-specific interactors or alternate
425 stoichiometries (Fig. 1). NuRD is found at all active enhancers and promoters in both mouse and
426 human cells (Bornelöv et al., 2018, Burgold et al., 2019, de Dieuleveult, Yen et al., 2016, Gunther,
427 Rust et al., 2013, Miller, Ralser et al., 2016, Shimbo, Du et al., 2013), but only a relatively small
428 proportion of these genes changes expression after *MBD3* deletion (Figs. 4A, 5A) (Bornelöv et al.,
429 2018). This is because NuRD acts to fine-tune expression through nucleosome mobilisation, and to
430 cement longer-term gene expression changes through histone deacetylation activity (Bornelöv et

al., 2018, Liang, Brown et al., 2017). NuRD's fine-tuning function also works to ensure cells are able to respond appropriately when stem cells are induced to undergo lineage commitment. Yet the actual series of molecular events through which chromatin remodellers facilitate a cell's ability to respond to differentiation cues remain ill-defined. The rapid development of single molecule and single cell analyses should allow us to now define exactly how chromatin remodellers, signalling molecules and transcription factors all interact at regulatory sequences to allow cells to respond quickly to changes in the local environment.

Materials and Methods

Cell lines and culture conditions

hiPSCs were a generous gift of Prof. Austin Smith (Takashima et al., 2014). Endogenously tagged MBD3-3xFLAG hiPSCs were made using a CRISPR/Cas9 gene editing approach to insert 3xFLAG immediately upstream of the MBD3 stop codon using a guide RNA targeting the sequence 5'-GAGCGAGTGTAGCACAGGTG-3' (Supplemental Fig. 1). *MBD3*-KO cells were generated replacing exons 2 and 3 with a puromycin resistance cassette using CRISPR/Cas9-mediated targeting and guide RNAs targeting the sequences 5'-GGCGGTGGACCAGCCGCGCC-3' and 5'-GTCGCTCTTGACCTTGTTGC-3'. A correctly targeted heterozygous clone was then transiently transfected with Dre recombinase prior to a second round of targeting to generate a homozygous null line (Supplemental Fig. 1). The MBD3 Rescue line was made by transfecting the *MBD3*-KO iPS line with a construct containing a CAG promoter driving expression of full-length MBD3-3xFLAG, followed by an IRES and a hygromycin resistance gene, and a polyA sequence from the human *PGK* gene. Hygromycin resistant cells were expanded and tested for MBD3-3xFLAG expression (e.g. Figure 2A).

454 hiPSCs were cultured in mTESR1 (StemCell Technologies) media or E8 medium (made in house,
455 prepared according to (Chen et al., 2011)) on vitronectin coated plates. hiPSCs were passaged using
456 an enzyme-free passaging reagent (ReleSR, StemCell Technologies) and plated as small clumps.

457 Neuroectoderm differentiation was induced based on (Vallier, Touboul et al., 2009). hiPSCs
458 were plated as clumps (day -1) in chemically defined medium with Polyvinyl Alcohol (CDM-PVA)
459 supplemented with hActivin A (10ng/μl) and FGF2 (12ng/μl) on 0.1% gelatin coated plates pre-
460 treated overnight at 37°C with MEF media (Advanced DMEM-F12, 10% FBS, 2 mM L-glutamine, 1x
461 penicillin/streptomycin). The original composition of CDM is 50% IMDM (Gibco) plus 50% F12
462 Nutrient-MIX (Gibco), supplemented with 4 ug/ml of insulin (Roche), 15 μg/ml transferrin (Roche),
463 450 μM monothioglycerol (Sigma), Chemically Defined lipid concentrate (Invitrogen). The next day
464 (day 0), hiPSCs were cultured in CDM-PVA supplemented with SB431542 (10 μM, Tocris), FGF2 (12
465 ng/ml, R and D Systems) and Noggin (15 ng/ml, Peprotech) for 12 additional days. The cells were
466 harvested using Accutase at 3, 6 and 12 days. The media was changed every day.

467 Definitive endoderm differentiation was induced according to (Yiangou, Grandy et al., 2019).
468 Cells were cultured in CDM-PVA supplemented with 100 ng/ml Activin A (produced in house), 80
469 ng/ml FGF2 (produced in house), 10 ng/ml BMP4 (R&D Systems), 10 μM LY294002 (Promega) and
470 3uM CHIR99021 for one day, with CHIR99021 omitted on the second day. From day three onwards,
471 cells were cultured in RPMI basal medium, supplemented with 100 ng/ml Activin A and 80 ng/ml
472 FGF2 on day 3. From day 4 onwards, RPMI was supplemented with 50 ng/ml Activin A only. The cells
473 were harvested using Accutase at days 2, 4, 6 and 8. The media was changed every day.

474 mEpiSCs were derived from *Mbd3*^{Flox/Δ} mESCs and subsequently transiently transfected with
475 Cre recombinase to create *Mbd3*^{Δ/Δ} cells. *Mbd3*^{Δ/Δ} mEpiSCs were independently derived from
476 *Mbd3*^{Δ/Δ} ES cells. mEpiSC cultures were maintained in N2B27 supplemented with FGF2 (12 ng/μl),
477 Activin A (20 ng/μl), XAV939 (2 mM, Sigma) on fibronectin (15 μg/ml) pre-coated plates. The cells
478 were harvested using Accutase at 2, 4 and 8 days. The media was changed every day. For neural

differentiation cells were plated on laminin-coated plates in N2B27 containing 1 μ M A83-01 (StemMACS).

Gene expression analysis was carried out as described (Burgold et al., 2019). Briefly, total RNA was isolated using RNA mini easy kit (Qiagen) and reverse transcribed using random hexamers and Superscript IV Reverse Transcriptase (Invitrogen). Quantitative PCR was carried out using gene-specific primers and Sybrgreen incorporation, or Taqman reagents on a StepOne or ViiA7 real time PCR system (both Applied Biosystems).

TAQMAN PROBES

GENE	PROBE
<i>βACTIN</i>	Hs01120798_m1
<i>GATA4</i>	Hs00171403_m1
<i>NANOG</i>	Hs04399610_g1
<i>NESTIN</i>	Hs04187831_g1
<i>PAX6</i>	Hs00240871_m1
<i>POU5F1</i>	Hs04260367_gH
<i>ZFP42</i>	Hs01938187_s1

PRIMERS

GENE	FORWARD PRIMER	REVERSE PRIMER	REFERENCE
<i>CRB3</i>	AGGTCAAAGACGCCCCG	TGAAGGCAAAACAGTGCTATTC	
<i>FOXA2</i>	GGGAGCGGTGAAGATGGA	TCATGTTGGCTCACGGCGGCGTA	
<i>FOXG1</i>	TCACGAAGCACTTGTTGAGG	AGGAGGGCGAGAAGAAGAAC	
<i>INADL</i>	GTGATGCCCTTGGAATCAGT	CTGCTCCTCTGTGTCTTCTG	
<i>SOX2</i>	GGACAGTTACGCGCACAT	GCTGGTCATGGAGTTGTACT	
<i>SOX17</i>	GGCGCAGCAGAATCCAGA	CCACGACTTGCCCAGCAT	
<i>ZEB1</i>	CTGACTGTGAAGGTGTACCA	GTACATCCTGCTTCATCTGC	Jiang et al 2018

Nuclear extraction, immunoprecipitation and Proteomics

Nuclear extraction was an immunoprecipitation was performed as described (Burgold et al., 2019). Original western blot images are available on Mendeley Data: XXXX. Antibodies used in this study are indicated below:

497

ANTIBODY	RAISED IN	COMPANY	CATALOGUE NUMBER	DILUTION FOR WESTERN BLOT
α CHD4	Mouse	Abcam	ab70469	1:5000
α GATAD2B	Rabbit	Bethyl Labs	A301-281A	1:2000
α LAMIN B1	Rabbit	Abcam	ab133741	1:10 000
α MBD3	Rabbit	Abcam	ab157464	1:5000
α MTA2	Mouse	Abcam	ab50209	1:5000

498

499 Mass spectrometry was carried out as described (Burgold et al., 2019, Kloet et al., 2018, Smits,
500 Jansen et al., 2013). Briefly, nuclear extract was prepared from a human iPS cell line in which a
501 3xFLAG tag was knocked in to the endogenous *MBD3* locus, or from two independent mouse
502 epiStem cell lines similarly modified as described (Burgold et al., 2019). One preparation of nuclear
503 extract from each cell line was divided into thirds, which were independently processed for
504 proteomic analyses. Proteins associated with 3xFLAG-tagged MBD3 were purified using anti-FLAG
505 sepharose (Sigma) and processed for mass spectrometry as described (Smits et al., 2013). The
506 resulting data were processed as in (Kloet et al., 2018).

507

508 **RNA-seq and analysis.** Sequencing libraries were prepared using the NEXTflex Rapid Directional
509 RNA-seq kit (Illumina) or SMARTer® Stranded Total RNA-Seq Kit v2—Pico Input Mammalian (Takara
510 Bio) and sequenced on the Illumina platform at the CRUK Cambridge Institute Genomics Core facility
511 (Cambridge, UK). Illumina sequence files were converted into FASTQ format. The short sequence
512 reads (75 nucleotides) were aligned to the Human reference genome (hg38;
513 <http://genome.ucsc.edu/>) or to the Mouse reference genome (mm10; <http://genome.ucsc.edu/>)
514 and assigned to genes using BWA (Li & Durbin, 2009). We used the Subread package (R statistical
515 tool; <http://www.r-project.org/>) to count aligned reads. Differentially expressed genes were
516 identified using R package edgeR (Chen, Lun, & Smyth, 2016). We used no fold change filtering and
517 results were corrected for multi-testing by the method of the False Discovery Rate (FDR) at the 1%
518 level. Differentially expressed genes were clustered using the unsupervised classification method of

the Kmeans (Soukas, Cohen, Socci, & Friedman, 2000). Heat maps were done using the pheatmap function (R statistical tool; <http://www.r-project.org/>). Functional annotation enrichment for Gene Ontology (GO) terms were realised using HumanMine [<http://www.humanmine.org>] (Smith, et al., 2012) or MouseMine database [<http://www.mousemine.org>]. Benjamini-Hochberg corrected P values of less than 0.01 were considered significant. GO terms were submitted to REVIGO, a web server that takes long lists of GO terms and summarizes them in categories and clusters of differentially expressed genes by removing redundant entries (Supek, Bošnjak, Škunca, & Šmuc, 2011). We used i-cisTarget tool (Imrichová, Hulselmans, Kalender Atak, Potier, & Aerts, 2015) to look for enrichment in TF position weight matrices and potential binding sites in the regulatory regions of co-expressed genes. i-cisTargetX computes statistical over representation of DNA motifs and ChIP-seq peaks in the non-coding DNA around sets of genes. The enrichment was considered significant when the Normalized Enrichment Score (NES) was higher than 5.

Data Availability

RNA-seq data are available with the Array Express accession number E-MTAB-8753. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD016967. All original western blot images are available at Mendeley Data: DOI: 10.17632/4t99j4c7gx.1.

Acknowledgments

We thank Austin Smith for providing the hiPSC line, Sabine Dietmann, Maike Paramor, Vicki Murray, Peter Humphreys and Sally Lees for technical assistance and advice, Sabine Dietmann, Denis Seyres and Laurence Röder for comments on the manuscript and members of 4DCellFate Consortium for useful discussions and suggestions. Funding to the BH and MV labs was provided through EU FP7 Integrated Project “4DCellFate” (277899). The BH lab further benefitted from

544 grants from the Medical Research Council (MR/R009759/1), the Wellcome Trust (206291/Z/17/Z)
 545 and the Isaac Newton Trust (17.24(aa)), and core funding to the Cambridge Stem Cell Institute
 546 from the Wellcome Trust and Medical Research Council (203151/Z/16/Z). The Vermeulen lab is
 547 part of the OncoCode Institute, which is partly funded by the Dutch Cancer Society (KWF).

548

549 **Author Contributions**

550 NR and BH devised the study; SG, JC, OO, SK, TB, NR and BH generated the data; RR analysed
 551 high throughput sequencing data, SK and MV generated and analysed proteomics data and RR and
 552 BH wrote the manuscript with input from other authors.

553

554 **Conflict of Interest Statement**

555 The authors declare no conflicts of interest.

556 References

- 557 Aigner K, Dampier B, Descovich L, Mikula M, Sultan A, Schreiber M, Mikulits W, Brabletz T, Strand
558 D, Obrist P, Sommergruber W, Schweifer N, Wernitznig A, Beug H, Foisner R, Eger A (2007) The
559 transcription factor ZEB1 (deltaEF1) promotes tumour cell dedifferentiation by repressing master
560 regulators of epithelial polarity. *Oncogene* 26: 6979-88
- 561 Allen HF, Wade PA, Kutateladze TG (2013) The NuRD architecture. *Cell Mol Life Sci* 70: 3513-24
- 562 Blaschke K, Ebata KT, Karimi MM, Zepeda-Martinez JA, Goyal P, Mahapatra S, Tam A, Laird DJ,
563 Hirst M, Rao A, Lorincz MC, Ramalho-Santos M (2013) Vitamin C induces Tet-dependent DNA
564 demethylation and a blastocyst-like state in ES cells. *Nature* 500: 222-6
- 565 Bornelöv S, Reynolds N, Xenophontos M, Gharbi S, Johnstone E, Floyd R, Ralser M, Signolet J, Loos
566 R, Dietmann S, Bertone P, Hendrich B (2018) The Nucleosome Remodeling and Deacetylation
567 Complex Modulates Chromatin Structure at Sites of Active Transcription to Fine-Tune Gene
568 Expression. *Mol Cell* 71: 56-72 e4
- 569 Brons IG, Smithers LE, Trotter MW, Rugg-Gunn P, Sun B, Chuva de Sousa Lopes SM, Howlett SK,
570 Clarkson A, Ahrlund-Richter L, Pedersen RA, Vallier L (2007) Derivation of pluripotent epiblast stem
571 cells from mammalian embryos. *Nature* 448: 191-5
- 572 Burgold T, Barber M, Kloet S, Cramard J, Gharbi S, Floyd R, Kinoshita M, Ralser M, Vermeulen M,
573 Reynolds N, Dietmann S, Hendrich B (2019) The Nucleosome Remodelling and Deacetylation
574 complex suppresses transcriptional noise during lineage commitment. *EMBO J* 38
- 575 Chen G, Gulbranson DR, Hou Z, Bolin JM, Ruotti V, Probasco MD, Smuga-Otto K, Howden SE, Diol
576 NR, Propson NE, Wagner R, Lee GO, Antosiewicz-Bourget J, Teng JM, Thomson JA (2011)
577 Chemically defined conditions for human iPSC derivation and culture. *Nat Methods* 8: 424-9
- 578 de Dieuleveult M, Yen K, Hmitou I, Depaux A, Boussouar F, Bou Dargham D, Jounier S,
579 Humbertclaude H, Ribierre F, Baulard C, Farrell NP, Park B, Keime C, Carriere L, Berlivet S, Gut M,
580 Gut I, Werner M, Deleuze JF, Olaso R et al. (2016) Genome-wide nucleosome specificity and
581 function of chromatin remodellers in ES cells. *Nature* 530: 113-6
- 582 Ee LS, McCannell KN, Tang Y, Fernandes N, Hardy WR, Green MR, Chu F, Fazzio TG (2017) An
583 Embryonic Stem Cell-Specific NuRD Complex Functions through Interaction with WDR5. *Stem Cell*
584 *Reports* 8: 1488-1496
- 585 Gokbuget D, Blelloch R (2019) Epigenetic control of transcriptional regulation in pluripotency and
586 early differentiation. *Development* 146
- 587 Gunther K, Rust M, Leers J, Boettger T, Scharfe M, Jarek M, Bartkuhn M, Renkawitz R (2013)
588 Differential roles for MBD2 and MBD3 at methylated CpG islands, active promoters and binding to
589 exon sequences. *Nucleic Acids Res* 41: 3010-21
- 590 Hein MY, Hubner NC, Poser I, Cox J, Nagaraj N, Toyoda Y, Gak IA, Weisswange I, Mansfeld J,
591 Buchholz F, Hyman AA, Mann M (2015) A human interactome in three quantitative dimensions
592 organized by stoichiometries and abundances. *Cell* 163: 712-23
- 593 Hota SK, Bruneau BG (2016) ATP-dependent chromatin remodeling during mammalian
594 development. *Development* 143: 2882-97
- 595 Jiang Y, Yan L, Xia L, Lu X, Zhu W, Ding D, Du M, Zhang D, Wang H, Hu B (2018) Zinc finger E-box-
596 binding homeobox 1 (ZEB1) is required for neural differentiation of human embryonic stem cells. *J*
597 *Biol Chem* 293: 19317-19329

598 Kaji K, Caballero IM, MacLeod R, Nichols J, Wilson VA, Hendrich B (2006) The NuRD component
599 Mbd3 is required for pluripotency of embryonic stem cells. *Nat Cell Biol* 8: 285-92

600 Kaji K, Nichols J, Hendrich B (2007) Mbd3, a component of the NuRD co-repressor complex, is
601 required for development of pluripotent cells. *Development* 134: 1123-32

602 Kim YS, Yi BR, Kim NH, Choi KC (2014) Role of the epithelial-mesenchymal transition and its effects
603 on embryonic stem cells. *Exp Mol Med* 46: e108

604 Kloet SL, Baymaz HI, Makowski M, Groenewold V, Jansen PW, Berendsen M, Niazi H, Kops GJ,
605 Vermeulen M (2015) Towards elucidating the stability, dynamics and architecture of the
606 nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics.
607 *FEBS J* 282: 1774-85

608 Kloet SL, Karemaker ID, van Voorthuijsen L, Lindeboom RGH, Baltissen MP, Edupuganti RR,
609 Poramba-Liyanage DW, Jansen P, Vermeulen M (2018) NuRD-interacting protein ZFP296 regulates
610 genome-wide NuRD localization and differentiation of mouse embryonic stem cells. *Nat Commun*
611 9: 4588

612 Latos PA, Helliwell C, Mosaku O, Dudzinska DA, Stubbs B, Berdasco M, Esteller M, Hendrich B
613 (2012) NuRD-dependent DNA methylation prevents ES cells from accessing a trophectoderm fate.
614 *Biology Open* 1: 341-352

615 Liang Z, Brown KE, Carroll T, Taylor B, Vidal IF, Hendrich B, Rueda D, Fisher AG, Merckenschlager M
616 (2017) A high-resolution map of transcriptional repression. *Elife* 6

617 Low JK, Webb SR, Silva AP, Saathoff H, Ryan DP, Torrado M, Brofelth M, Parker BL, Shepherd NE,
618 Mackay JP (2016) CHD4 Is a Peripheral Component of the Nucleosome Remodeling and
619 Deacetylase Complex. *J Biol Chem* 291: 15853-66

620 Miller A, Ralser M, Kloet SL, Loos R, Nishinakamura R, Bertone P, Vermeulen M, Hendrich B (2016)
621 Sall4 controls differentiation of pluripotent cells independently of the Nucleosome Remodelling
622 and Deacetylation (NuRD) complex. *Development* 143: 3074-84

623 Mohd-Sarip A, Teeuwssen M, Bot AG, De Herdt MJ, Willems SM, Baatenburg de Jong RJ, Looijenga
624 LHJ, Zatreanu D, Bezstarosti K, van Riet J, Oole E, van Ijcken WFJ, van de Werken HJG, Demmers JA,
625 Fodde R, Verrijzer CP (2017) DOC1-Dependent Recruitment of NURD Reveals Antagonism with
626 SWI/SNF during Epithelial-Mesenchymal Transition in Oral Cancer Cells. *Cell Rep* 20: 61-75

627 Moreno-Bueno G, Portillo F, Cano A (2008) Transcriptional regulation of cell polarity in EMT and
628 cancer. *Oncogene* 27: 6958-69

629 Newman JC, Bailey AD, Fan HY, Pavelitz T, Weiner AM (2008) An abundant evolutionarily
630 conserved CSB-PiggyBac fusion protein expressed in Cockayne syndrome. *PLoS Genet* 4: e1000031

631 Nichols J, Smith A (2009) Naive and primed pluripotent states. *Cell Stem Cell* 4: 487-92

632 Nichols J, Smith A (2012) Pluripotency in the embryo and in culture. *Cold Spring Harb Perspect Biol*
633 4: a008128

634 Reynolds N, Latos P, Hynes-Allen A, Loos R, Leaford D, O'Shaughnessy A, Mosaku O, Signolet J,
635 Brennecke P, Kalkan T, Costello I, Humphreys P, Mansfield W, Nakagawa K, Strouboulis J, Behrens
636 A, Bertone P, Hendrich B (2012a) NuRD suppresses pluripotency gene expression to promote
637 transcriptional heterogeneity and lineage commitment. *Cell Stem Cell* 10: 583-94

638 Reynolds N, Salmon-Divon M, Dvinge H, Hynes-Allen A, Balasooriya G, Leaford D, Behrens A,
639 Bertone P, Hendrich B (2012b) NuRD-mediated deacetylation of H3K27 facilitates recruitment of
640 Polycomb Repressive Complex 2 to direct gene repression. *Embo J* 31: 593-605

641 Saksouk N, Barth TK, Ziegler-Birling C, Olova N, Nowak A, Rey E, Mateos-Langerak J, Urbach S, Reik
642 W, Torres-Padilla ME, Imhof A, Dejardin J, Simboeck E (2014) Redundant mechanisms to form
643 silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. *Mol Cell* 56: 580-
644 94

645 Shimbo T, Du Y, Grimm SA, Dhasarathy A, Mav D, Shah RR, Shi H, Wade PA (2013) MBD3 localizes
646 at promoters, gene bodies and enhancers of active genes. *PLoS Genet* 9: e1004028

647 Signolet J, Hendrich B (2015) The function of chromatin modifiers in lineage commitment and cell
648 fate specification. *FEBS J* 282: 1692-702

649 Singh S, Howell D, Trivedi N, Kessler K, Ong T, Rosmaninho P, Raposo AA, Robinson G, Roussel MF,
650 Castro DS, Solecki DJ (2016) Zeb1 controls neuron differentiation and germinal zone exit by a
651 mesenchymal-epithelial-like transition. *Elife* 5

652 Smits AH, Jansen PW, Poser I, Hyman AA, Vermeulen M (2013) Stoichiometry of chromatin-
653 associated protein complexes revealed by label-free quantitative mass spectrometry-based
654 proteomics. *Nucleic Acids Res* 41: e28

655 Takashima Y, Guo G, Loos R, Nichols J, Ficiz G, Krueger F, Oxley D, Santos F, Clarke J, Mansfield W,
656 Reik W, Bertone P, Smith A (2014) Resetting Transcription Factor Control Circuitry toward Ground-
657 State Pluripotency in Human. *Cell* 158: 1254-69

658 Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, Gardner RL, McKay RD (2007)
659 New cell lines from mouse epiblast share defining features with human embryonic stem cells.
660 *Nature* 448: 196-9

661 Tong JK, Hassig CA, Schnitzler GR, Kingston RE, Schreiber SL (1998) Chromatin deacetylation by an
662 ATP-dependent nucleosome remodelling complex. *Nature* 395: 917-21

663 Vallier L, Touboul T, Chng Z, Brimpari M, Hannan N, Millan E, Smithers LE, Trotter M, Rugg-Gunn P,
664 Weber A, Pedersen RA (2009) Early cell fate decisions of human embryonic stem cells and mouse
665 epiblast stem cells are controlled by the same signalling pathways. *PLoS One* 4: e6082

666 Wade PA, Jones PL, Vermaak D, Wolffe AP (1998) A multiple subunit Mi-2 histone deacetylase
667 from *Xenopus laevis* cofractionates with an associated Snf2 superfamily ATPase. *Curr Biol* 8: 843-
668 846

669 Xue Y, Wong J, Moreno GT, Young MK, Côté J, Wang W (1998) NURD, a novel complex with both
670 ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol Cell* 2: 851-61

671 Yiangou L, Grandy RA, Osnato A, Ortmann D, Sinha S, Vallier L (2019) Cell cycle regulators control
672 mesoderm specification in human pluripotent stem cells. *J Biol Chem*

673 Yin R, Mao SQ, Zhao B, Chong Z, Yang Y, Zhao C, Zhang D, Huang H, Gao J, Li Z, Jiao Y, Li C, Liu S,
674 Wu D, Gu W, Yang YG, Xu GL, Wang H (2013) Ascorbic acid enhances Tet-mediated 5-
675 methylcytosine oxidation and promotes DNA demethylation in mammals. *J Am Chem Soc* 135:
676 10396-403

677 Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, Cohen P, Smith A (2008) The
678 ground state of embryonic stem cell self-renewal. *Nature* 453: 519-23

679 Zhang W, Aubert A, Gomez de Segura JM, Karuppasamy M, Basu S, Murthy AS, Diamante A, Drury
680 TA, Balmer J, Cramard J, Watson AA, Lando D, Lee SF, Palayret M, Kloet SL, Smits AH, Deery MJ,
681 Vermeulen M, Hendrich B, Klenerman D et al. (2016) The Nucleosome Remodeling and
682 Deacetylase Complex NuRD Is Built from Preformed Catalytically Active Sub-modules. *J Mol Biol*
683 428: 2931-42

684 Zhang Y, LeRoy G, Seelig H-P, Lane WS, Reinberg D (1998) The dermatomyositis-specific
685 autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome
686 remodeling activities. *Cell* 95: 279-289
687

Table 1: Transcription factor binding sites associated with gene clusters

Human Cluster	TF	NES score
Cluster 1	MYC	7.99
	ATF4	7.27
	CEBPB	6.57
	SNAI2	5.85
	ZEB1	5.25
Cluster 2	NA	NA
Cluster 3	SRF	6.06
Cluster 4	NA	NA
Cluster 5	NA	NA
Cluster 6	TP53	8.07
	GF11B	6.23

Mouse Cluster	TF	NES score
Cluster 1	Zeb1	6.62
	Rela	5.85
	Smarcc1	5.74
	Fos/Jun	5.57
	Klf1	5.21
	Tead3	5.12
Cluster 2	NA	NA
Cluster 3	NA	NA
Cluster 4	Trp73	5.05

Overlap Cluster	TF	NES score
Cluster 1	RBBP9	5.33
	SF1	5.01
Cluster 2	ZEB1	5.54
	SNAI2	5.43
	RNF114	5.25
	PPARG	5.10
Cluster 3	PSMA6	5.25
	SOX6	5.04

689

690 **Figure S1. Human alleles used in this study.** A) Schematic of the human *MBD3* locus. Filled boxes
 691 represent coding exons, open boxes represent non-coding exons. B). Schematic of the *MBD3-FLAG*
 692 allele. A 3xFLAG epitope was knocked-in immediately upstream of the stop codon in Exon 6 on one
 693 *MBD3* allele. C). Schematic of the *MBD3-KO* alleles. Exons 2 and 3 were replaced by a Puromycin
 694 resistance cassette flanked by Rox sites. After the first round of targeting the Puromycin cassette
 695 was removed by transient Dre expression, followed by a second round of targeting to mutate both
 696 *MBD3* alleles.

697

698

699 **Figure S2. Expression and GO-term plots for Clusters 5 and 6 from Figure 4A.** There were no
700 significant ($P_{adj} \leq 0.01$) GO terms associated with Cluster 5 genes. Error bars indicate standard
701 deviations of average expression.

702

A

Ragheb et al. Fig. S3

B

703

704 **Figure S3. Mouse *Mbd3*-KO epiStem cells.** A) Phase contrast images of wild type or *Mbd3*-KO
705 mEpiSCs in self-renewing conditions. B). Wild type or *Mbd3*-KO cells after 8 days of neural
706 differentiation were stained with antibodies indicated at top or DAPI. The final panels show merged
707 images. Scale bars represent 100µm.

A

B

708

709 **Figure S4. Expression profiles of *ZEB1*, *CRB3* and *INADL* in hiPSCs and mEpiSCs.** A) RT-qPCR of *ZEB1*,
710 *CRB3* and *INADL* in hiPSCs during neuroectodermal differentiation. The y-axis shows the relative
711 expression and the x-axis shows the day of differentiation. The blue line represents the WT while
712 the red line represents *MBD3*-KO hiPSCs (n=3 replicates/timepoint/condition). Error bars indicate
713 the standard deviation. B) Expression profiles of *Zeb1*, *Crb3* and *Inadl* from mEpiSC RNA-seq data.
714 The y-axis represents TPM (Transcript Per Million) while the x-axis and the different shapes
715 represent the days of differentiation (n≥3 replicates/timepoint/condition). Blue indicate the WT and
716 magenta represent *Mbd3*-KO mEpiSC.