Mbd3 and deterministic reprogramming

Paul Bertone, Brian Hendrich and José C.R. Silva

Wellcome Trust – Medical Research Council Stem Cell Institute
University of Cambridge, Gleeson Building, Tennis Court Road, Cambridge CB2 1QR, UK
Correspondence: pb407@cam.ac.uk (PB), bdh24@cam.ac.uk (BH), jcs64@cam.ac.uk (JCRS)

Embryonic development requires the activity of the Nucleosome Remodeling and Deacetylase (NuRD) complex. NuRD functionality can be ablated by rendering cells devoid of methyl-CpG-binding domain protein 3 (Mbd3), a critical component that confers stability to the complex. Previous studies noted that Mbd3+/–– embryonic stem (ES) cells misregulate a subset of pluripotency-associated genes, and subsequently fail to engage in cell differentiation into embryonic lineages when self-renewal requisites (e.g. LIF) are withdrawn from culture media. Components of the NuRD complex have been shown to interact with Oct4 and Nanog, two important transcription factors operative in the production of iPS cells. Thus, elucidating the role of Mbd3/NuRD in the reprogramming process is of relevance to the field.

Rais et al. reported the remarkable observation that suppressing formation of the NuRD complex by depleting Mbd3 promotes near 100% induction of cells reprogrammed to pluripotency. This was an important result, as typical iPS cell conversion rates are extremely low; such an increase in reprogramming efficiency would represent a considerable advance in the production of iPS cells for research and therapeutic applications. However, concurrent and independent work from our labs obtained contrasting results, where a profound reduction in reprogramming efficiency was observed from cells where Mbd3 had been ablated.

We sought to understand this discrepancy, in part by analyzing the data provided by Rais et al. The study employs cells containing a single functional allele of Mbd3 that can be conditionally deleted (Mbd3+/-), and that are also transgenic for several constructs inserted into the genome by random integration: a doxycycline (DOX)-inducible polycistronic reprogramming cassette, a promoter-driven Oct4-GFP reporter, and a constitutive mCherry reporter used for quantification of colony sizes and single-cell deposition by flow sorting. Performance of Mbd3+/- cells in reprogramming assays was described in Rais et al. relative to Mbd3+/- counterparts.

A gene expression dataset was produced for the study, where Mbd3+/- and Mbd3+/- mouse embryonic fibroblasts (MEFs) were profiled on Affymetrix arrays during a reprogramming timecourse. Comprehensive analysis of this experiment is precluded by its design: only four time points are represented, one of those is inconsistent between experiment and control samples (taken at 8 vs 11 days), and no replicates were provided. Nevertheless, differential expression analysis reveals highly similar outcomes in each condition (Fig. 1A), with little divergence among pluripotency genes during induction of Mbd3 heterozygous and wild-type cells (Fig. 1B).

Figure 1. A) Microarray data from Rais et al. indicate similar gene expression patterns from Mbd3+/- and Mbd3+/- cells during a timecourse of doxycycline-induced reprogramming (Z-score transformed log2 expression). B) Pluripotency factors show little variation between Mbd3+/- and Mbd3+/- lines. Mbd3 is expressed at comparable levels in both conditions (bottom row).
Figure 2. A) Genes exhibiting the most discrepant changes between Mbd3<sup>1/+</sup> and Mbd3<sup>−/−</sup> cells profiled over the timecourse. B) Probe-level intensity data from the complete Mbd3 probe set (upper data tracks) and reduced probe set excluding probes to exon 1 and UTRs remaining in the knockout allele (lower data tracks; see Methods). Mbd3 locus and transcript isoforms are depicted below (antisense orientation). C) Comparable expression of Mbd3 and Nanog in Mbd3<sup>1/+</sup> and Mbd3<sup>−/−</sup> cells over the time series (top); Mbd3 is transcribed in heterozygous and Mbd3<sup>−/−</sup> MEFs at 85% and 66% wild-type levels, respectively (bottom). Continued expression in "null" cells may be due to reprogramming of MEF lines after Cre excision was performed, thereby selecting for iPS cell colonies where reversion to pluripotency was facilitated by the presence of an intact allele.

The greatest fold-change differences between experiment and control cells over the time series arise from a discordant set of genes devoid of canonical pluripotency regulators (Fig. 2A), suggesting the dominating effect to be due to biological variation expected from distinct and independently derived cell lines. However, the degree of such variability is impossible to assess in the absence of experimental replication.

Rais et al. evaluated the potential of Mbd3 depletion primarily in the Mbd3<sup>3/4</sup> heterozygous background without deleting the remaining allele. This was predicated on the notion that Mbd3 displays hypomorphic expression, based on the authors’ estimate of protein abundance in Mbd3<sup>3/4</sup> cells at 20% that of wild-type levels. Markedly different results were obtained in two independent studies by our groups<sup>8,11</sup>, where near wild-type Mbd3 protein abundance was measured from cells of this genotype regardless of the culture conditions used.

Analysis of the microarray data from Rais et al. shows Mbd3 transcript levels in heterozygous cells to be 85% relative to Mbd3<sup>1/+</sup> controls, consistent with the behavior previously observed. Although few experiments in Rais et al. involve cells in which the floxed allele had been deleted to assess reprogramming efficiency in null (Mbd3<sup>−/−</sup>) conditions, this was performed and expression data from those cells were included in the dataset. Mbd3<sup>−/−</sup> MEFs profiled in the study express Mbd3 transcript at 66% wild-type levels and 78% relative to Mbd3<sup>3/4</sup> cells (Fig. 2B,C), calling into question the effective depletion of Mbd3 protein and impairment of NuRD function as a causal factor contributing to the reported increase in reprogramming efficiency.

Much of the Rais et al. study makes use of a reporter of Pou5f1 (Oct4) expression, consisting of the complete endogenous Oct4 regulatory sequence linked to GFP<sup>12</sup>. Analysis of the ChIP-seq data provided by Rais et al. allows inspection of the promoter fragment used to regulate GFP expression in the reporter lines. The Oct4 promoter region contains several well-characterized elements<sup>13</sup>, notably the proximal and distal enhancers (PE and DE). Their functions have been previously defined using genomic Oct4 fragment (GOF)-18, an 18 kb intact sequence, and derivatives where regions encompassing each enhancer have been deleted (ΔPE and ΔDE)<sup>14</sup>. Sequencing reads corresponding to the reporter transgene DNA map to the endogenous Oct4 locus in the reference genome at high copy number (Fig. 3A). Alignments from Mbd3<sup>3/4</sup> cells are contiguous and span the entire promoter region. In contrast, a gap in read coverage is present in Mbd3<sup>−/−</sup> cells corresponding to the segment deleted in the ΔPE construct (Fig. 3B). The intact GOF-18 construct is solely described in Rais et al. and indicated schematically in Extended Data Figure 3a (top). The full Oct4 promoter is illustrated with PE and DE elements included, implying that all cells received this plasmid. In contrast, it is evident that Mbd3<sup>3/4</sup> and Mbd3<sup>−/−</sup> control cells harbor different variants of GOF-18 reporter constructs.
Figure 3. Sequencing data from whole-cell extract (WCE) genomic DNA libraries reveals numerous transgene copies relative to genomic background (A), with Mbd3<sup>−/−</sup> (blue) and Mbd3<sup>+/+</sup> (red) Oct4-GFP reporter lines harboring intact GOF-18 and GOF-18 ∆PE constructs, respectively (B). Proximal and distal enhancer regions of the Oct4 promoter are denoted, together with sequence conservation and Nanog binding site occupancy from an independent dataset<sup>28</sup>. Scales indicate read count (left) and transgene copy range estimated at 1 kb intervals (right).
The reprogramming system described in Rais et al. employed a polycistronic reprogramming cassette (STEMCCA) as well as a constitutive mCherry reporter used for quantification of colony sizes and single-cell deposition by flow sorting. To verify that GOF-18 ∆PE Mbd3+/− ChIP-seq data originated from the Oct4-GFP reporter cells used throughout the study, we identified sequencing reads corresponding to mCherry and parts of the STEMCCA construct design, including the internal ribosomal entry site (IRES) and 2A peptide sequences linking each reprogramming factor (Fig. 4).

Analysis and commentary

To properly evaluate the role of the NuRD complex by Mbd3 depletion, both copies of the gene must be ablated. Rais et al. report hypomorphic expression from Mbd3+/− cells estimated at 20% wild-type levels, thereby justifying the use of a heterozygous cell line to represent a functional Mbd3 mutant. That assessment disagrees with our experience and the authors' microarray data, where robust Mbd3 expression is apparent in both Mbd3+/− and Mbd3+/− lines. The latter finding may have arisen from incomplete Cre excision and/or positive selection of reprogramming-competent cells with an intact Mbd3 allele. This suggests that differences in reprogramming kinetics are unlikely to be related to Mbd3 depletion, and indeed transcriptional states are comparable between the experiment and control cells profiled in the study.

Nonetheless, substantial improvements in reprogramming efficiency are described in Rais et al. Dramatic enhancement of pluripotency induction is reported from assays in which GFP was used as a readout for imaging and flow cytometry. Sequencing data from the study reveal that Mbd3+/− cells were transfected with an Oct4-GFP reporter based on the GO/F-18 ∆PE construct, whereas Mbd3+/− cells harbor an intact GO/F-18 promoter fragment. Oct4 is expressed in a wider repertoire of tissues and cell types than embryonic stem cells and reporters based on the intact GO/F-18 construct display similarly broad activity. The PE is the most highly conserved region of the Oct4 promoter in eutheria and also drives transcription in post-implantation embryos. Deleting the PE confines expression to naïve pluripotent cells, and thus a construct lacking the PE effectuates a much more stringent reporter of authentic reprogramming outcomes.

Differential application of a promiscuous test reporter and a considerably weaker control compromises the study design and undermines the conclusions drawn. An invalid experimental setup is imposed where no combination of Oct4-GFP reporter lines can be legitimately compared, as the two constructs have been applied in a mutually exclusive fashion to the experiment and control groups (Fig. 5). This applies to all ES-derived and iPS-derived MEFs where Oct4-GFP+ selection or quantification was used to establish differential reprogramming efficiency. No scientific motivation for comparative evaluation of alternate Oct4-GFP reporters is described in Rais et al., and use of the ∆PE variant is not declared. Thus the paper is lacking a key methodological disclosure essential for accurate interpretation of the results.
Table 1. Exhibits from Rais et al. presenting data based on Oct4-GFP quantification to assess reprogramming efficiency.

<table>
<thead>
<tr>
<th>Figure</th>
<th>Assay</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1f</td>
<td>Primordial germ cell derivation</td>
<td>p.66: “Single cell isolated Mbd3fl/fl Oct4-GFP E8.5 PGCs from chimaeric mice were proficient in forming EG cell colonies and lines (.95% efficiency), whereas PGCs isolated from chimaeras that were generated by micro-injecting Mbd3fl/fl or Mbd3+/- cells carrying an exogenous Mbd3 transgene reprogrammed at less than 10% efficiency (Fig. 1f).”</td>
</tr>
<tr>
<td>Extended Data 3c</td>
<td>Oct4-GFP expression from reprogrammed cells and intermediates</td>
<td>p.66: “Notably, 95% of Mbd3fl/fl and Mbd3+/- cells were Oct4-GFP” at day 10, whereas only levels up to 18% were observed in control Mbd3fl/fl fibroblasts (Fig. 2a).”</td>
</tr>
<tr>
<td>Extended Data 3d</td>
<td>Oct4-GFP expression from reprogrammed cells and intermediates</td>
<td></td>
</tr>
<tr>
<td>Extended Data 3f</td>
<td>Oct4-GFP expression from reprogrammed cells and intermediates</td>
<td></td>
</tr>
<tr>
<td>2a</td>
<td>Reprogramming efficiency from Mbd3+/-, Mbd3fl/fl and Mbd3+/- cells</td>
<td></td>
</tr>
<tr>
<td>2b</td>
<td>Matrix of reprogramming outcomes (secondary IPS cells)</td>
<td>p.66: “Single cell sorting of secondary mCherry Mbd3+/- mouse embryonic fibroblasts (MEFs) and subsequent reprogramming in 2i/LIF plus doxycycline conditions reproducibly yielded 100% IPS cell derivation efficiency by day 8. Wild-type cells reprogrammed under identical conditions, no more than 20% of clones reactivated Oct4-GFP (Fig. 2b).”</td>
</tr>
<tr>
<td>Extended Data 3e</td>
<td>Oct4-GFP expression from reprogrammed cells and intermediates</td>
<td>p.67: “High single-cell reprogramming efficiency rates were obtained from a variety of adult progenitor and terminally differentiated cells (Fig. 2d and Extended Data Fig. 3e,f).”</td>
</tr>
<tr>
<td>2e</td>
<td>Imaging of Oct4-GFP+ vs mCherry Mbd3+/- and Mbd3fl/fl colonies</td>
<td>p.67: “By day 6 after doxycycline induction, .98% of Mbd3+/- clonal populations reactivated the Oct4-GFP pluripotency marker, whereas only up to 20% efficiency was detected in control samples reprogrammed in identical growth conditions (Fig. 2e, f).”</td>
</tr>
<tr>
<td>2f</td>
<td>Secondary reprogramming assay comparing Oct4-GFP+ Mbd3+/- and Mbd3fl/fl colonies (top) and cells (bottom)</td>
<td>p.67: “By day 6, approximately 85% of cells within each individual Mbd3 clonal population became Oct4-GFP+ cells, whereas &lt;2% of cells within successfully reprogrammed Mbd3+/- clones turned on the Oct4-GFP marker (bottom panel in Fig. 2f).”</td>
</tr>
<tr>
<td>Extended Data 4a</td>
<td>Flow sorting of Oct4-GFP+ Mbd3+/- and Mbd3fl/fl cells</td>
<td>p.67: “Detection of Oct4-GFP by flow cytometry on polyclonal populations demonstrated similar IPS cell reprogramming kinetics (Extended Data Fig. 4a).”</td>
</tr>
<tr>
<td>Extended Data 4b</td>
<td>Time course of lentiviral transduction</td>
<td>p.67: “Re-infection with lentiviruses encoding Mbd3, but not Mbd2, before day 5 of reprogramming had a profound inhibitory effect on IPS cell generation from Mbd3fl/fl MEFs, whereas re-infection after day 5 had a diminished effect (Extended Data Fig. 4b).”</td>
</tr>
<tr>
<td>3b</td>
<td>Secondary reprogramming assay</td>
<td>p.67: “After the depletion of Mbd3 expression, we were not able to isolate stable, partially reprogrammed cells that did not reactivate Oct4-GFP or Nanog-GFP and could be stably expanded in vitro, as typically can be obtained from OSKM-transduced wild-type somatic cells (Fig. 3b).”</td>
</tr>
<tr>
<td>3c</td>
<td>Flow sorting of Oct4-GFP+ cells</td>
<td>p.67: “Notably, by introducing Mbd3 siRNA, all clones markedly turned on Oct4-GFP or Nanog-GFP pluripotency markers after continued OSKM expression in 2i/LIF (Fig. 3c).” [only Oct4-GFP data are shown]</td>
</tr>
<tr>
<td>5e</td>
<td>Oct4-GFP+ cells in reprogramming assay</td>
<td>p.70: “Mbd3 mutants with a compromised ability to interact with OSKM reprogramming factors directly (Extended Data Fig. 9d) were deficient in reducing reprogramming of Mbd3+/- somatic cells, supporting the notion that direct OSKM-Mbd3 interactions are important for inhibiting IPS cell formation (Fig. 5e).”</td>
</tr>
</tbody>
</table>

The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY 4.0 International license.
Figure 5. Data from equivalent ChIP-seq profiles indicate a 1:1 correspondence between the intact GOF-18 promoter construct applied to Mbd3+/− cells (left) versus the GOF-18 ΔPE variant present in Mbd3+/− control cells (right).

The line of investigation presented in Rais et al. heavily relies on Oct4-GFP expression as a proxy for the reversion of somatic cells to pluripotency. Differences in reporter activity arising from the tandem use of intact GOF-18 and GOF-18 ΔPE constructs may have adversely affected a significant number of assays and conclusions presented in the study (Table 1). The trend depicted in Figure 2f of Rais et al. provides an illustrative example, where Mbd3+/− cells appear to revert to pluripotency at an accelerated rate relative to controls. Expression data from the study do not support that finding, which may have been construed on the basis of GFP output alone. Mbd3+/− cells, where Oct4-GFP is driven by the ΔPE reporter, would be expected to yield profoundly reduced fluorescence signal relative to a variant based on the full promoter sequence.

During reprogramming, partially reverted intermediates are inherently produced en route to iPS cell colony formation. GFP expression from Mbd3+/− cells is unrestricted in these transitional states and is nonspecific for ground state pluripotency. This shortcoming is exacerbated by the authors’ use of serum replacement factors (e.g. KSR) in culture media, which abolishes specificity for naïve pluripotent cells conferred by inhibition of glycogen synthase kinase 3 (GSK3) and mitogen-activated protein kinase pathways. Numerous Oct4-GFP transgene insertions present in Mbd3+/− (up to 24) and Mbd3+/+ (up to 38) cells were uncharacterized with respect to the regulatory context of integration sites, potentially leading to spurious GFP activation unrelated to complete reprogramming state or the expression of endogenous Oct4.

Appropriate controls were not implemented for the reprogramming system utilized in Rais et al. Mbd3+/− cells did not constitute the parental line of the Mbd3+/− cells acquired for the study; genetic modifications were delivered separately to Mbd3+/− and Mbd3+/+ primary and secondary donor cells; random transgene integrations were not assessed in any condition; and cell lines were independently derived. Reprogramming experiments were performed in permissive conditions and cells were transformed with excessive Oct4-GFP transgene copies such that fluorescence activation is likely to be misregulated. All of these factors contribute to considerable experimental variation and impair the determination of biological significance. Assays in which incompatible fluorescence reporters are directly compared cannot be considered valid.

Assessment of Mbd3/NuRD function in reprogramming must be conducted with validated Mbd3-null cells, compatible and equivalent genetic modifications in test and control conditions, rigorous evaluation of authentic pluripotent cells and reprogramming outcomes, and matched cell lines from an isogenic parental background. Mbd3+/− cells are not sufficient to assess the impact of Mbd3 depletion, as cells of this genotype feature near wild-type transcript levels and protein abundance. In the absence of independent verification and in light of the deficiencies outlined above, results presented in Rais et al. describing 100% reprogramming efficiency based on the use of Mbd3+/− cells must be questioned as a potential artifact of the authors’ experimental system.
Concluding remarks

We brought this matter to the attention of the authors, and upon receiving an unsatisfactory explanation for the disparities found, ultimately raised the issue with *Nature*. The editors declined to publish our exchange as a contribution to the Communications Arising section, and instead encouraged the authors to post a comment to the *Nature* website21. The comment makes readers aware of a difference in Oct4-GFP reporter usage, but the significance of this issue and its implications for the study as a whole are diminished. We therefore issue this letter as an expression of concern to investigators who would follow this work.

Acknowledgements

We are grateful to Austin Smith and Wolfgang Huber for helpful discussions and advice.

Methods

Microarray data analysis

Affymetrix Mouse Gene 1.0 ST array data were obtained from GEO22 record GSE453528 and processed with the oligo Bioconductor package23. Microarray data were normalized with the robust multi-array average (RMA) method23. Transcript clusters were mapped to mouse gene annotation based on release 7B of Ensembl25.

Mbd3 transcript expression

Microarray probesets targeting Mbd3 were originally assigned a value of 1 in the crosshyb_type field of the Affymetrix design files, indicating each probe in the Mbd3 transcript cluster (10370824) to be unique with respect to other putatively transcribed sequences targeted by the array. No additional perfect matches were found to any other mouse transcript annotated in Ensembl release 7B, consistent with the assessment of cross-hybridization potential carried out at design time. Heterozygous knockout (Mbd3+/−) cells had been targeted such that exons 2–7 were replaced with the β-geo selection marker, leaving exon 1 and UTR sequences intact10. To discount residual contribution from the non-functional allele, sense-orientation probe sequences were mapped to the reverse complement of Mbd3 genomic DNA, and probes corresponding to exon 1 (84510, 233909, 995596, 1042262), 5’ UTR (314091, 646154, 26469) and 3’ UTR (1028146, 391255, 585086, 333495) were deleted from the pd.mogene.1.0.st.v1 annotation database26 prior to normalization. Expression levels were estimated as described above from the remaining 20 of 31 original probes spanning the Mbd3 locus. Probe-level data were plotted with the GenomeGraphs Bioconductor package27.

ChIP-seq data analysis

Illumina sequencing data deposited under accessions SRP0287188 and SRX00054528 were obtained from the Sequence Read Archive29 and aligned to the mouse genome GRCm38 (mm10) using BWA30, allowing permissive treatment of low-quality base calls (−1 2 5 −q 2 0). For conservative copy number estimation, duplicate reads likely arising from PCR amplification were removed with Picard31, and suboptimal alignments (−q 10) filtered with SAMtools32. Focal gains corresponding to transgene insertions were estimated from genomic DNA (WCE, whole-cell extract) samples, accounting for G/C content33 and assuming ploidy = 2 over windows of 1–10 kb. Read density was computed with F-Seq34 and visualized in the Integrative Genomics Viewer35.

References


