The Company of Biologists

Comparative genomic analysis of embryonic, lineage-converted and stem cell-derived motor neurons

Justin K. Ichida[1,2,3,§,¶], Kim A. Staats[3,§], Brandi N. Davis-Dusenbery[1,2,*,§], Kendell Clement[1,4], Kate E. Galloway[3], Kimberly N. Babos[3], Yingxiao Shi[3], Esther Y. Son[1,2], Evangelos Kiskinis[1,2], Nicholas Atwater[1,2], Hongcang Gu[4], Andreas Gnirke[4], Alexander Meissner[1,4,5,¶] and Kevin Eggan[1,2,¶]

ABSTRACT

Advances in stem cell science allow the production of different cell types *in vitro* either through the recapitulation of developmental processes, often termed 'directed differentiation', or the forced expression of lineage-specific transcription factors. Although cells produced by both approaches are increasingly used in translational applications, their quantitative similarity to their primary counterparts remains largely unresolved. To investigate the similarity between *in vitro*-derived and primary cell types, we harvested and purified mouse spinal motor neurons and compared them with motor neurons produced by transcription factor-mediated lineage conversion of fibroblasts or directed differentiation of pluripotent stem cells. To enable unbiased analysis of these motor neuron types and their cells of origin, we then subjected them to whole transcriptome and DNA methylome analysis by RNA sequencing (RNA-seq) and reduced representation bisulfite sequencing (RRBS). Despite major differences in methodology, lineage conversion and directed differentiation both produce cells that closely approximate the primary motor neuron state. However, we identify differences in Fas signaling, the Hox code and synaptic gene expression between lineage-converted and directed differentiation motor neurons that affect their utility in translational studies.

KEY WORDS: Motor neuron, Embryonic stem cells, Lineage conversion, iPS cell, Reprogramming, Directed differentiation

INTRODUCTION

Improvements in transcription factor-mediated lineage conversion of cellular identity and differentiation of pluripotent stem cells increasingly allow access to previously inaccessible cell types (Gopalakrishnan et al., 2015; Han et al., 2011; Ichida and Kiskinis, 2015; Xu et al., 2015). In particular, the realization that *Oct4*, *Sox2*, *Klf4* and *Myc* can reprogram easily obtained cell types to pluripotency has fueled new approaches to studying human disease. By reprogramming cells from individuals with genetic forms of disease into induced pluripotent stem cells (iPSCs) or somatic cell types, it has become possible to understand and develop interventions for cell type-specific pathologies that previously could have only been studied using postmortem tissues (Blanchard et al., 2015; Brennand et al., 2011; Ichida et al., 2009, 2014; Ichida and Kiskinis, 2015; Kiskinis et al., 2014; Kramer et al., 2018; Mertens et al., 2015; Pepper et al., 2017; Shi et al., 2018; Son et al., 2011; Takahashi et al., 2007a,b; Toma et al., 2015; Wainger et al., 2015; Wen et al., 2014; Wilkinson et al., 2018; Xu et al., 2015; Zhang et al., 2015; Zhao et al., 2015).

Lineage conversion and directed differentiation possess different advantages and disadvantages. Whereas an iPSC intermediate allows unlimited expansion, cell culture-derived changes to iPSCs or imperfections in directed differentiation strategies could influence how closely iPSC-derived cell types mimic their primary counterparts (Gore et al., 2011; Merkle et al., 2017; Sances et al., 2016; Sandoe and Eggan, 2013). In contrast, lineage conversion is a more streamlined process that may be capable of preserving certain age-dependent gene expression or epigenetic signatures (Mertens et al., 2015). However, there is little cell expansion and the direct, non-developmental nature of the conversion raises questions about its veracity (Xu et al., 2015). Although transcription factor-converted motor neurons resemble primary motor neurons (Abernathy et al., 2017; Briggs et al., 2017; Mazzoni et al., 2013), in-depth transcriptional and DNA methylation analyses are still required to determine their utility in translational applications.

Currently, there is no consensus concerning whether one of these approaches more accurately produces differentiated cell types of interest. There are few comparisons of cells produced by both lineage conversion and embryonic stem cell (ESC)- or iPSC-directed differentiation and their primary counterparts. Here, we have used sequencing to quantitatively compare the gene expression and DNA methylation of cells produced by both of these strategies. To understand where variation might arise, we also profiled cells of origin and the stem cell intermediates.

We selected spinal motor neurons (MNs) as the target cell for this comparison because the developmental biology of this neural subtype is well understood (Dasen et al., 2005; Jessell, 2000; Tsuchida et al., 1994). Strategies for directed differentiation of mouse (ESCs) into motor neurons are well-characterized (Wichterle et al., 2002) and motor neuron-specific expression of the transcription factor Hb9 (*Mnx1*) enables flow-purification (Jessell, 2000). Crucially, *in vitro*-derived motor neurons produced by either directed differentiation or lineage conversion have a motor neuron phenotype, become physiologically active, engraft appropriately into the developing spinal cord and recapitulate sensitivities to neurodegenerative disease stimuli (Di Giorgio et al., 2007; Son et al., 2011; Wichterle et al., 2002).

[1]Harvard Stem Cell Institute, Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA. [2]Howard Hughes Medical Institute, Stanley Center for Psychiatric Research, Cambridge, MA 02142, USA. [3]Department of Stem Cell Biology and Regenerative Medicine, University of Southern California, Los Angeles, CA 90033, USA. [4]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [5]Department of Genome Regulation, Max Planck Institute for Molecular Genetics, Berlin 14195, Germany.
*Present address: Seven Bridges Genomics, Cambridge, MA 02142, USA. §These authors contributed equally to this work

¶Authors for correspondence (ichida@usc.edu; meissner@molgen.mpg.de; eggan@mcb.harvard.edu)

J.K.I., 0000-0002-8827-8087; K.A.S., 0000-0001-8392-7992; K.C., 0000-0003-3808-0811; A.M., 0000-0001-8646-7469; K.E., 0000-0003-4436-8467

DEVELOPMENT

Surprisingly, we find that stem cell-derived and lineage-converted MNs differ from primary MNs to a similar degree. Although the cells from the two methods display differences that affect their relative utility in certain translational applications, the two *in vitro* approaches together express more than 98% of the gene sets enriched in primary MNs. We conclude that the vast majority of the biology of a given cell type can currently be accessed through one or both of these methods.

RESULTS

To comprehensively identify differences between lineage conversion and directed differentiation, we compared cell types produced *in vitro* and *in vivo*, as well as the cellular intermediates used in the *in vitro* processes (Fig. 1A,B). To reduce genetic variation, we derived all cells from *Hb9*::GFP C57Bl/6 mice (Fig. 1A).

Lineage conversion into motor neurons by exogenous Ngn2, Isl1, Lhx3, Ascl1, Brn2, Myt1l and Hb9 occurs over 15 days and includes co-culture with primary mixed glia starting at day 3 (Son et al., 2011). Directed differentiation into motor neurons occurs over 7-10 days and does not incorporate glial co-culture (Di Giorgio et al., 2007; Wichterle et al., 2002) (Fig. 1A). Owing to the kinetic differences between the two differentiation methods, we sought to align the maturation state of MN populations by harvesting each group ~4 days after *Hb9*::GFP activation, which marks the establishment of the MN transcriptional program. First, we isolated mouse embryonic fibroblasts (MEFs) and performed either lineage conversion into induced motor neurons (iMNs) or iPSCs. We then performed directed differentiation of the iPSCs into MNs using an established, morphogen-based method (Wichterle et al., 2002). To control for differences between ESCs and iPSCs, we also derived ESCs from these same animals and differentiated them into MNs. Finally, we analyzed the transcriptomes and DNA methylomes of these cell populations by RNA-seq and RRBS (Gu et al., 2011; Meissner et al., 2008; Ziller et al., 2013), respectively (Fig. 1A).

Because *Hb9*::GFP expression peaks in lineage-converted MN cultures at day 11 post-transduction, we collected iMNs at day 15.

Because transgene expression ensues at day 3 in ESC- and iPSC-derived MN (ESC MN, iPSC MN) cultures, we harvested ESC MNs and iPSC MNs at day 7. Similarly, because *Hb9*::GFP expression fills the spinal cord at E9.5 in transgenic mice, we harvested primary MNs at E13.5 (Fig. 1A). With the caveat that high *Hb9*::GFP transgene levels can appear asynchronously *in vitro* and with slightly different kinetics depending on the differentiation method, this allowed us to 'time stamp' MNs at a particular time of differentiation for comparable analysis. iMNs, ESC/iPSC MNs and embryonic MNs (EMB MNs) at these stages share similar morphological and electrophysiological properties (Son et al., 2011), and their viability in culture is similar, although iMNs and EMB MNs display longer survival than ESC MNs (Fig. S1A). Single cell qRT-PCR analysis determined that ~66-76% of the Hb9$^+$ iMNs co-expressed endogenous *Isl1* and *Lhx3*, consistent with a medial motor columnar identity (Fig. S1B). Wictherle et al. have previously shown that ~70% of Hb9$^+$ ESC MNs and EMB MNs co-express *Isl1* and *Lhx3* (Wichterle et al., 2002). Thus, the subtype composition of the primary and *in vitro* MN preparations is similar.

We performed RNA-seq on two biological replicates of each MN type, stem cell type and MEF condition (Tables S1 and S2). To control for *in vitro* culture effects, we analyzed four MEF culture conditions, including MEFs cultured for 15 days in MN media (N3, Fig. 1B). We obtained an average of 29 (±3 s.e.m.) million mapped 100 bp paired-end reads per sample (Fig. S2A) and biological replicates exhibited tight correlation (Fig. S2B,C) (average Pearson coefficient=0.94±0.03 s.e.m.). MEFs, iMNs and EMB MNs were derived from both male and female embryos to enable sufficient numbers for analysis. However, the ESCs and iPSCs used to generate ESC MNs and iPSC MNs, respectively, were both of a male background, and the expression levels of *Xist* reflected these sex differences between samples (Fig. S3). Therefore, we eliminated X- and Y-chromosome genes from the analyses.

Unsupervised clustering analysis of global gene expression revealed that all MNs grouped together, apart from PSCs and MEFs (Fig. 1C, Fig. S2B). Surprisingly, altering the culture conditions of MEFs changed the levels of more than 7000 genes (Fig. S4A,B,
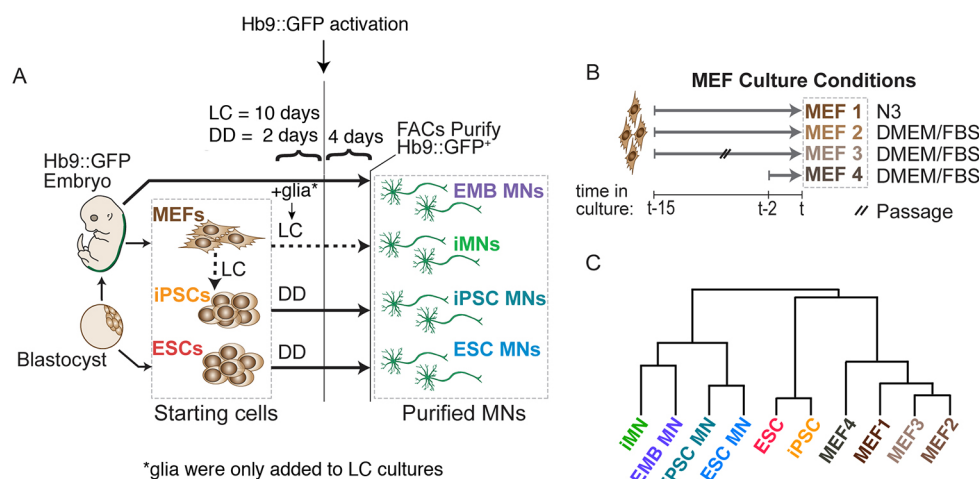

Fig. 1. Experimental design for the *in vitro* and *in vivo* motor neuron comparison. (A) *Hb9*::GFP transgenic embryos (E13.5) were used to isolate embryonic (EMB) MNs and MEFs. The MEFs were then lineage converted (LC) to iPSCs and iMNs. The iPSCs and ESCs were used for directed differentiation (DD) towards iPSC MNs and ESC MNs. Only iMNs were generated in the presence of primary mouse glia. All MN types were purified by fluorescence-activated cell sorting (FACS) using the *Hb9*::GFP reporter. Collected cells were used for transcriptional profiling by RNA-seq and DNA methylation analysis by RRBS. Two biological replicates were analyzed per sample type. (B) MEFs were cultured in fibroblast medium [DMEM+10% fetal bovine serum (FBS) (MEF2-4)] or the neural conversion medium [N3 (MEF1)] for either 2 days (MEF4) or 15 days (MEF1-3). MEF3 samples were passaged once during culture to identify changes that occur due to passaging. (C) Cluster dendrogram of transcriptomes. Two biological replicates were analyzed per sample type.

Table S2). For example, 1611 genes were differentially expressed between MEFs cultured in serum-free N3 medium (MEF1) compared with serum-containing medium (MEF2) (Fig. S4B). In addition, 7283 genes were differentially expressed between MEFs cultured *in vitro* for 2 days compared with the full iMN conversion period of 15 days (MEF2 versus MEF4) (Fig. S4B). Thus, to control for culture-induced changes, we performed all pair-wise analyses relative to MEF1, which had been cultured in MN media.

Consistent with the tight clustering of MN global transcriptional profiles, most genes up- or downregulated in primary EMB MNs versus MEF1 were also appropriately expressed in iMNs, and ESC/ iPSC MNs (Fig. S4A,B). Of all cells analyzed, ESCs and iPSCs were the most similar to each other (0 differentially expressed genes, sum of squared shrunken $\log_2$-fold changes=59.92) (Fig. 2A, Fig. S4A,B). By comparison, gene expression in ESCs and EMB MNs relative to MEF1 were highly divergent (Fig. 2D-F, Fig. S4A,B, Table S2). Although ESC- and iPSC-derived MNs were similar, the differences between them (2122 differentially expressed genes, sum of squared shrunken $\log_2$-fold changes=111.39) were greater than those between ESCs and iPSCs (sum of squared shrunken $\log_2$-fold changes=59.92) (Fig. 2A,B and Fig. S4A,B). MNs derived by directed differentiation or lineage conversion were close in the magnitude of their similarity to EMB MNs (4516, 6202 and 4712 differentially expressed genes and sum of squared shrunken $\log_2$-fold changes=177.66, 209.11 and 177.74 for iMNs, ESC MNs and iPSC MNs relative to embryonic MNs, respectively) (Fig. 2G-I and Fig. S4A,B).

To examine the performance of lineage conversion and directed differentiation on a gene-by-gene basis, we compared the starting cell population and the target cell population (referred to as the 'reference' cell population) to find the number of genes needing upregulation or repression to make the reference cell. To reduce inaccuracies in differential gene expression analyses, we eliminated genes with an average read count of less than three in preparation of the count matrix for DESeq2 and restricted our analyses to genes that qualified in DESeq2 as capable of reliable comparison. In evaluating iPSCs, iMNs, ESC MNs and iPSC MNs, we restricted our analyses to genes that were differentially expressed between the starting cell and reference cell according to DESeq2.

Starting with iPSC reprogramming, we found that 4128 genes needed upregulation and 4253 genes needed repression in MEFs to achieve the ESC state (Fig. 3A). For genes needing upregulation, we reasoned that 50% of the expression observed in the reference cells could serve as a cut-off for successful reprogramming as many genes function normally in their haploinsufficient state (Fig. 3A). Conversely, for genes needing repression, we considered them successfully changed if they were lowered to half the value in the reference cell. We hypothesize that these metrics will be relevant for most genes. However, we also calculated the percentages of genes that were successfully upregulated to within 25% or 10% below the expression level in the reference cell or downregulated to 1.5- or 1.2-fold the reference cell level (Fig. S5A). iPSCs were excellent proxies for ESCs, with appropriate reprogramming of 99%, 93% or 82% (to within 50%, 25% or 10% of ESC levels, respectively) of the 4128 genes needing upregulation and 96%, 87% or 71% (to within 2-, 1.5- or 1.2-fold of ESC levels, respectively) of the 4253 genes needing repression (Fig. 3A and Fig. S5A). In our most stringent analysis, we required genes to not only be induced or repressed to the designated thresholds compared with ESCs, but we also required them not to be over-induced or over-repressed by the same
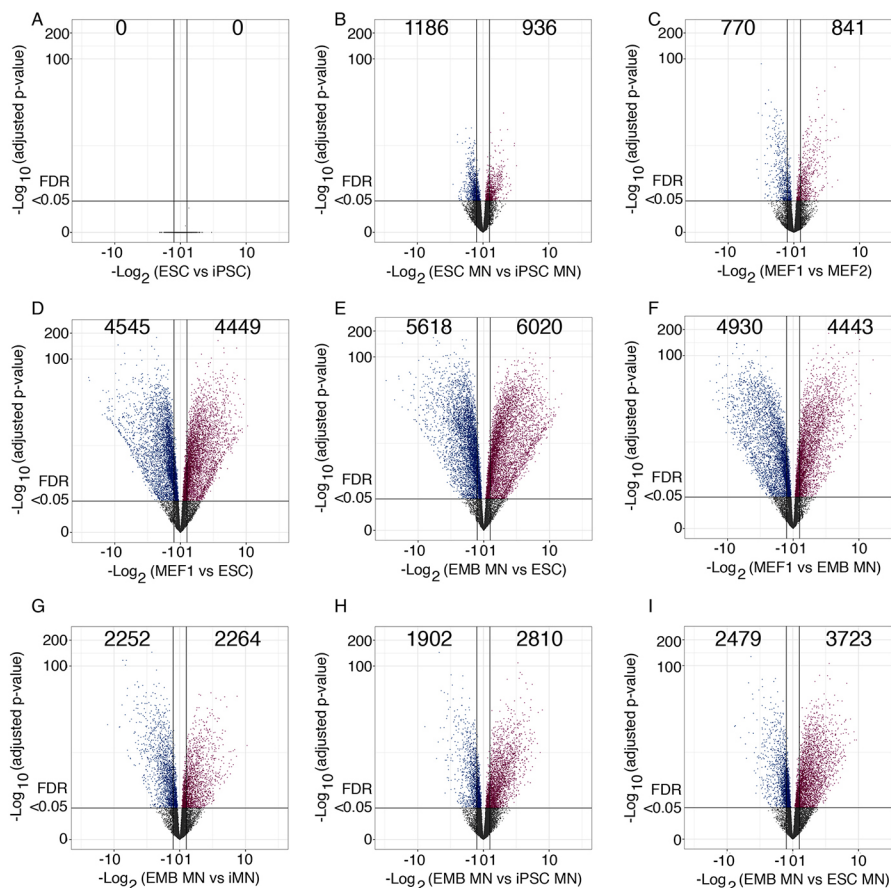

Fig. 2. Transcriptional comparison of *in vitro*- and *in vivo*-derived *Hb9*::GFP+ motor neurons. (A-I) Volcano plots of genes upregulated or downregulated in the indicated pairwise comparisons. Only genes with an average read count of three or greater across all samples are included. The number of significantly upregulated or downregulated genes is shown for each comparison. Genes with a log2-fold-change less than −1 or greater than 1, and FDR<0.05 (DESeq2, default settings) are shown in color. Two biological replicates were analyzed per cell type.
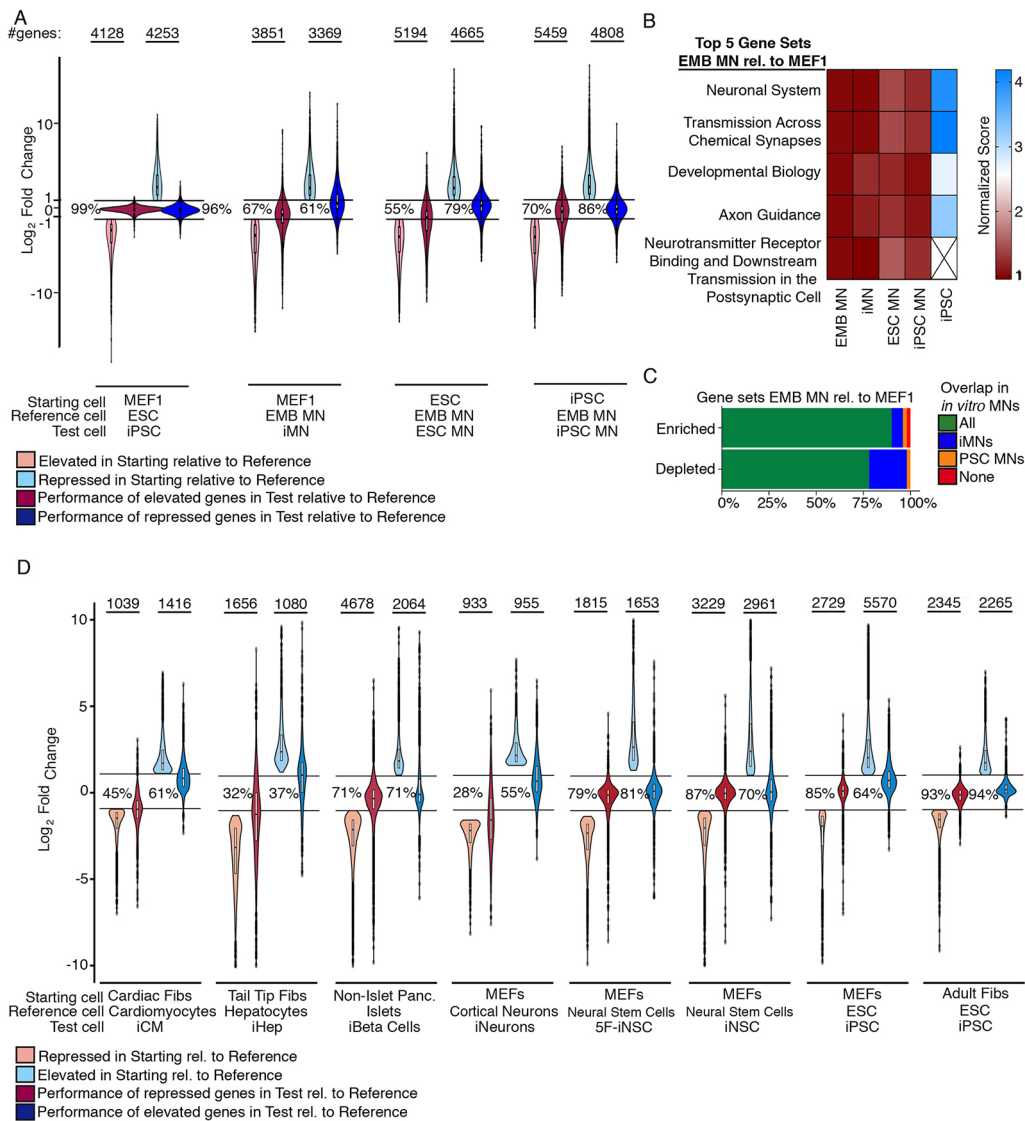
Fig. 3. Transcriptional characterization. (A) Violin plots showing the success of reprogramming. Pink- and light-blue colored violins show fold change of genes detected as differentially expressed (DESeq2) between the reference (target) cell type and the starting cell type. Only differentially expressed genes with starting cell levels of twofold or greater or 0.5-fold or less than in the reference cell were analyzed. The total number of genes that meet the specified criteria and are up- or downregulated between the starting cell and reference cell are indicated at the top of the plot. We define successful regulation of gene expression as upregulated to at least 50% of the level or downregulated to at least half the level expressed by the reference cell type. The percentage of successfully regulated genes is shown for each pair. The 'test' cell is the cell sample that underwent lineage conversion or directed differentiation. Two biological replicates were analyzed per cell type. (B) Heatmap showing enrichment score of the top five functional pathways determined by Gene Ontology enrichment analysis. Pathways enriched among genes differentially expressed in embryonic MN relative to MEFs were selected and the normalized enrichment score (k/K) of these pathways in iMNs, ESC MNs and iPSC MNs is shown. The enrichment of these pathways among genes differentially regulated in iPSCs is shown as a negative control. Two biological replicates were analyzed per cell type. (C) Performance of *in vitro* MNs in the regulation of gene sets enriched in embryonic MNs. Gene sets enriched or depleted among genes significantly regulated in embryonic MNs relative to MEFs were evaluated for regulation in *in vitro* MNs compared with MEFs. Gene sets also enriched in all *in vitro* MNs are shown in green. Gene sets regulated in embryonic MNs but only appropriately regulated in iMNs or IPSC and ESC MNs are shown in blue and orange, respectively. Two biological replicates were analyzed per cell type. (D) Violin plots of gene expression regulation in other *in vitro*-derived cell types compared with natural counterparts. Differentially regulated genes were identified by comparing the gene expression in the reference cell type relative to the starting cell. The performance of these genes in the test cell type is shown as described in more detail in A. The 'test' cell is the cell sample that underwent lineage conversion or directed differentiation. Two (neurons, five-factor neural stem cells), three (cardiomyocytes, β cells, neural stem cells, mouse iPSCs, human iPSCs) or four (hepatocytes) biological replicates were analyzed per sample type.

thresholds. Adding bidirectional cut-offs lowered the percentages of successful genes (Fig. S5A), but may be overly stringent for some genes. When cut-offs were not used and only DESeq2 analysis was applied, no genes were differentially expressed between ESCs and iPSCs (Fig. S5E). Thus, iPSCs faithfully mimic ESCs (Fig. S5A).

When comparing iMNs with embryonic MNs, we found that 67%, 45% or 32% (to within 50%, 25% or 10% of embryonic MN

gene levels, respectively) of the 3851 genes needing upregulation and 61%, 39% or 27% (to within 2-, 1.5- or 1.2-fold of embryonic MN gene levels, respectively) of the 3369 genes needing repression in MEFs were successfully reprogrammed (Fig. 3A and Fig. S5B). When cut-offs were not used and only DESeq2 analysis was applied, 60% of genes needing upregulation and 58% of genes needing repression were correctly expressed in iMNs (Fig. S5E).

Although iMNs were not as close to embryonic MNs as iPSCs were to ESCs, some of the differences we observed may be due to the fact that iMNs were cultured *in vitro* while embryonic MNs were not. Still, at the 50%/twofold cut-offs or with respect to differentially expressed genes, the expression of most genes was appropriately altered during reprogramming. Genes requiring suppression and activation were modulated with similar success (Fig. 3A and Fig. S5B,E).

Compared with MEFs, pluripotent stem cells possessed more differentially expressed genes when compared with embryonic MNs, with ESCs and iPSCs requiring the upregulation of 5194 and 5459 genes and the suppression of 4665 and 4808 genes, respectively (Fig. 3A). Compared with lineage conversion, directed differentiation achieved similar success in generating MNs (Fig. 3A). In ESC MNs, 55%, 35% or 26% of genes were successfully upregulated (to within 50%, 25% or 10% of embryonic MN gene levels, respectively) and 79%, 50% or 37% were appropriately downregulated (to within 2-, 1.5- or 1.2-fold of embryonic MN gene levels, respectively). In iPSC MNs, 70%, 56% or 46% of genes were successfully upregulated (to within 50%, 25% or 10% of embryonic MN gene levels, respectively) and 86%, 72% or 58% were correctly downregulated (to within 2-, 1.5- or 1.2-fold of embryonic MN gene levels, respectively) (Fig. 3A and Fig. S5C,D). When cut-offs were not used and only DESeq2 analysis was applied, 47% and 62% of genes that needed to be upregulated and 63% and 76% of genes that needed to be repressed in ESC MNs and iPSC MNs, respectively, were successfully changed (Fig. S5E). Directed differentiation more frequently changed gene expression correctly when genes needed repression rather than upregulation. Thus, both lineage conversion and directed differentiation produce MNs with ~55-86% of genes being correctly expressed.

Gene ontology analysis of functional pathways by Gene Set Enrichment Analysis (GSEA) of a ranked list of all genes revealed that more than 75% of annotations enriched in primary MNs were also enriched in all *in vitro*-derived MNs (Fig. 3B,C). Notably, most of the remaining functions enriched in primary MNs could be accessed by either ESC/iPSC MNs or iMNs (Fig. 3C). Overall, 98% of gene sets enriched in primary MNs were enriched in either iMNs or PSC MNs (Fig. 3C). We obtained similar results using hypergeometric enrichment of terms among differentially expressed genes (Fig. S6).

We next investigated the activation and repression of ChIP-seq-validated targets of Isl1, a central transcriptional regulator of motor neuron development (Jessell, 2000; Mazzoni et al., 2013). In iMNs, Isl1 correctly activated 834/1120 target genes and suppressed 456/646 targets, indicating that Isl1 activity was largely intact in iMNs (Fig. S7A,B). Transcriptional regulation by Isl1 in ESC MNs was slightly less successful with 747/1120, and similarly successful in iPSC MNs with 864/1120 target genes achieving appropriate activation (Fig. S7A,B). In addition, 16 and 13 Isl1 target genes in ESC MNs and in iPSC MNs, respectively, were incorrectly suppressed. Therefore, most Isl1 target genes are correctly regulated in *in vitro*-derived MNs, providing a rationale for the observation that the transcriptional programs of iMNs, ESC MNs and iPSC MNs closely resemble the natural state.

To extend this analysis, we used the Ingenuity Pathway Analysis Network Builder to curate target genes of the other iMN factors. Most targets identified by this approach were significantly changed in their expression in EMB MNs relative to MEFs, confirming their role in MN biology (Fig. S8A-G). Successful induction of target genes identified using the Ingenuity Pathway Analysis Network Builder was remarkably similar across all *in vitro*-derived MNs

(56%, 48% and 52% for iMNs, ESC MNs and iPSC MNs, respectively). Interestingly, the neural progenitor gene *Nkx6.1* was activated in EMB MNs, ESC MNs and iPSC MNs, but not iMNs (Fig. S8C). This is consistent with previous studies showing that cells do not transition through a neural progenitor state during iMN conversion (Briggs et al., 2017; Son et al., 2011).

To frame the transcriptional analysis of *in vitro*-derived MNs in the context of other *in vitro*-derived cell types, we performed a meta-analysis of existing gene expression data for several other reports of lineage conversion (Fig. 3D). Table S3 includes information on the identity and maturity of the reference cell types. We again required that a gene be upregulated to at least 50% or downregulated to less than half the level in the reference cell type in order to be considered as successfully regulated. We also considered genes that were not altered in the reference cell type relative to the starting cell but induced more than threefold in the *in vitro*-derived cell as aberrantly expressed. Relative to either MEFs or iPSCs, *in vitro*-derived MNs successfully up- and downregulated more than 74% of genes and the number of aberrantly regulated genes was similar (358 for iMNs relative to MEF1 versus 303 for iPSC MNs relative to iPSC). Both mouse and human iPSCs were excellent mimics of the reference cell type (>98% success for the human iPSCs) (Fig. 3D). Lineage-converted induced neural stem cells and *in vivo*-matured induced β cells reflected a similar level of transcriptional similarity as *in vitro*-derived motor neurons to their reference versions (Fig. 3D).

In contrast, most other *in vitro*-derived cells only achieved a 50% or lower rate of success (Fig. 3D). In some cases, factors other than reprogramming accuracy likely exacerbated gene expression differences between primary and *in vitro*-derived cells. For example, the transcription factors A*scl1*, *Brn2* and *Myt1l*, which were used to generate induced neurons, are expressed by many types of neurons, not just the cortical neurons used as the reference cell type. Furthermore, the primary cortical neuron population used for the reference cell transcriptome is highly heterogeneous (Marro et al., 2011) (Table S3). These factors could have increased the number of differentially expressed genes between the test (lineage-converted) cell and reference cells. In the case of induced hepatocytes, the authors compared the reprogrammed cells with primary adult hepatocytes (Huang et al., 2011) (Table S3). Thus, a difference in hepatocyte maturity could have elevated gene expression differences. However, in the case of induced cardiomyocytes, the authors compared them to neonatal primary cardiomyocytes, so maturity was probably not a major contributor to the divergent gene expression (Ieda et al., 2010) (Table S3). Therefore, some of the differences between *in vitro*-derived and primary reference cells may be due to maturity, reference cell diversity or cell culture conditions. Beyond this, however, inaccurate reprogramming would be a major contributing factor. Thus, although directed differentiation and lineage conversion can both generate cells that resemble their natural counterparts, careful investigation of the protocols and the resulting cells is still required. Different transcription factor combinations can significantly change the properties of the resulting cells (Nakagawa et al., 2010), suggesting that additional transcription factors or improved culture conditions could yield better mimics of the reference cell state.

To further investigate the similarity between natural motor neurons and their *in vitro*-derived equivalents, we performed genome-wide methylation analysis by RRBS. Two biological replicates were analyzed across all conditions as well as an additional MEF control transduced with a GFP virus (MEF5) to

control for any changes due to viral transduction (Fig. S9A). In line with the transcriptional analysis, all MN populations clustered together, separately from the pluripotent stem cells and MEFs (Fig. 4A and Fig. S10A).

Most promoters previously detected to be bound by the motor neuron factors Isl1 and Lhx3 by chromatin immunoprecipitation assays (Mazzoni et al., 2013) were de-methylated in all examined cell types, suggesting these regions could be available for regulation

Fig. 4. Integrative epigenetic and transcriptional characterization. (A) Hierarchical clustering of sample methylation using individual CpGs. The distance between sample pairs was computed as 1 minus the correlation between sample pairs, using all CpGs covered in that pair of samples by RRBS. Two biological replicates were analyzed per cell type. (B) Percentage of promoters that are covered by RRBS with Isl1- or Lhx3-binding sites within 10 kb of the transcription start site that are demethylated (methylation≤0.2). Values represent the mean of two biological replicates. (C) Resolution of differentially methylated promoters. Differentially methylated promoters (PMs) are defined as either hypermethylated (HyperMe) (light brown, >10% methylation difference between the starting cell and the reference cell) or hypomethylated (HypoMe) (light purple, ≤10% methylation difference between the starting cell and the reference cell). The resolutions of these hypermethylated and hypomethylated promoters are shown in dark brown and dark purple, respectively, as a comparison between the test and reference methylation levels. The percentage of promoters that have successfully resolved methylation to a level within 10% of the reference level is shown. (D) Expression classification of genes that did not successfully resolve methylation to a level within 10% of the reference level in C. Low expression: genes with approximately three reads or fewer in both starting and reference samples. Unaltered expression: not differentially expressed between starting and reference samples. Successfully altered: differentially expressed between starting and reference levels, and expression level in the test sample is <2-fold or <50% of the level in the reference sample for genes that are higher or lower in the starting sample than the reference sample, respectively. Unsuccessfully altered: differentially expressed between starting and reference levels, and expression level in the test sample is not <2-fold or <50% of the level in the reference sample for genes that are higher or lower in the starting sample than the reference sample, respectively. Me, DNA methylation. (E) Expression (top row) and methylation (bottom row) heatmaps of example genes that are classified as unsuccessful in methylation and expression change in D. Two biological replicates were analyzed per cell type for all experiments.

by MN transcription factors (Fig. 4B). We extended the paradigm of evaluating reprogramming success performed on gene expression to methylation changes (Fig. 4C). We have previously found that defining promoters as hyper- or hypomethylated if they possess greater than a 10% methylation difference between two cell samples can robustly identify methylation differences in heterogeneous populations (Ziller et al., 2013). Similar to gene expression, more than 90% of genes differentially methylated in ESCs relative to MEFs were appropriately regulated in iPSCs (Fig. 4C). Interestingly, all *in vitro*-derived MNs showed a reduced propensity to properly regulate methylation of regions that were hypermethylated in the starting cells compared with the reference cells (dark red bars, Fig. 4C). Each of the *in vitro*-derived motor neurons showed less ability to recapitulate the methylation changes in EMB MNs than gene expression changes. However, all *in vitro*-derived MNs performed similarly in the regulation of both hyper- and hypomethylated genes.

We next examined the expression changes of genes that did not show appropriate methylation regulation. Across all *in vitro*-derived MNs, more than 50% of the examined genes were either expressed at a low level in the reference cell type or showed no gene expression change (grey and black bars, respectively, Fig. 4D). This suggests that the lack of appropriate methylation changes at these regions is unlikely to be relevant for cell type determination. We detected two classes of regulation among the remaining genes: (1) genes showing successful expression changes despite insufficient methylation changes; and (2) genes that were improperly regulated at both the gene expression and methylation levels (Fig. 4D). Seventy-one genes, including *Mmp9*, *Gria1* and *Csmd3* showed a failure to properly regulate at both expression and methylation levels in PSC MNs (Fig. 4E). Whereas 23 genes failed to be appropriately regulated at both expression and methylation levels in iMNs, only two genes showed improper expression and methylation regulation in both iMNs and PSC MNs (Fig. 4E). Thus, in combination, these technologies are able to access the vast majority of gene changes in primary MNs.

We next investigated possible reasons for the differential gene expression between iMNs and PSC MNs. First, we determined whether any differences between motor neuron types might be due to the fact that iMNs were exposed to primary glia during reprogramming. Of the 2336 genes correctly expressed in iMNs but not ESC MNs, 963 (41%) became correctly expressed in ESC MNs when they were differentiated in glia-conditioned medium (Fig. S11A). However, differentiating ESC MNs in glia-conditioned medium also caused aberrant expression of 1271 of the 8954 genes (14%) that both iMNs and ESC MNs derived without glia-conditioned medium expressed correctly (Fig. S11A). Although we cannot rule out the possibility that direct contact with glial cells is partially responsible for inducing the proper motor neuron gene expression program, we conclude that factors secreted by glia induce a similar number of corrective and aberrant gene expression changes and do not account for the genes expressed correctly by iMNs but not ESC MNs.

To determine whether glial contamination in iMN samples could explain the gene expression differences between iMNs and PSC MNs, we examined the expression of microglial, oligodendrocyte and astrocytic genes in the different MN samples. iMN expression of microglia and oligodendrocyte genes was similar to that of MEFs and EMB MNs (Fig. S11B). Although some astrocytic genes such as *Gfap* and *Aldh1l1* were expressed more highly in iMNs than EMB MNs, others, such as *Slc1a2*, were not (Fig. S11B). Therefore, the detection of certain glial markers, such as *Gfap* and *Aldh1l1*, is

likely due to misexpression in iMNs rather than glial contamination. Indeed, MEFs express higher levels of *Gfap* and *Aldh1l1* than pluripotent stem cells, suggesting the misexpression in iMNs could be due to a failure in fully suppressing these genes (Fig. S11B). Therefore, the differences between iMNs and PSC MNs are not caused by large amounts of glial contamination in iMN samples.

To determine whether the gene expression differences between iMNs and PSC MNs are caused by variability in the stoichiometric ratios of the transgenic reprogramming transcription factors in iMNs, we performed single cell qRT-PCR analysis on iMNs. We focused on the ratio between transgenic *Isl1* and *Lhx3* because of their central roles in controlling MN biology (Jessell, 2000) and because we previously found that these were the most critical transgenic factors for iMN formation (Son et al., 2011). Fifty-five percent of iMNs possessed a transgenic *Isl1* to *Lhx3* ratio within twofold of the iMN median ratio (fourfold total range). Seventy percent of cells fell within fourfold of the iMN median ratio (Fig. S11C). Thus, most iMNs have similar levels of transgenic transcription factor expression.

We next investigated the possibility that iMNs and ESC MNs represent different stages of motor neuron maturity. Although during iMN conversion, *Hb9*::GFP reporter activation ensues in postmitotic cells and reaches high levels within 1-2 days, in ESC directed differentiation, *Hb9*::GFP activation occurs in mitotic progenitor cells and reaches high levels in postmitotic cells 2-4 days later (Son et al., 2011; Wichterle et al., 2002). Therefore, it is possible that some ESC/iPSC MNs are at a less mature stage than iMNs at 4 days after reporter activation.

To identify the gene expression pattern of less mature MNs, we performed RNA-seq on EMB MNs collected at E11.5, which represented the mid-point between *Hb9*::GFP transgene activation at E9.5 and our previous samples, which were collected at E13.5. Principle component analysis revealed that iMNs were more similar to E13.5 than E11.5 EMB MNs, whereas ESC/iPSC MNs were more similar to E11.5 MNs along the first principle component (Fig. 5A). Moreover, ESC/iPSC MNs were more similar to E11.5 EMB MNs than iMNs were to E13.5 EMB MNs (Fig. 5A). In addition, in unsupervised hierarchical clustering analysis using a restricted set of genes that differentiate between mature and immature MNs (Ho et al., 2016), ESC MNs and iPSC MNs clearly clustered with E11.5 EMB MNs, while iMNs clustered with E13.5 EMB MNs (Fig. 5B). Differentiating ESC MNs in the presence of glia-conditioned medium did not make ESC MNs more similar to E13.5 MNs (Fig. 5A,B). Thus, differences in MN maturity at the time points we analyzed contribute to the gene expression differences between iMNs and ESC/iPSC MNs.

Among neuronal subtypes, MNs are particularly sensitive to activation of the Fas pathway, which results in rapid MN death (Raoul et al., 2002). Mutations in the *SOD1* gene that cause familial ALS, a motor neuron disease, increase MN degeneration in response to Fas activation. Moreover, activation of the Fas pathway is modulated in part by glial cells within the central nervous system. Thus, this pathway is an important area of investigation for disease studies (Raoul et al., 2002).

The gene encoding the Fas receptor (*Fas*) is demethylated and expressed in iMNs but methylated and silenced in PSC MNs (Fig. 6A). Differentiation of ESC MNs in glia-conditioned medium did not activate *Fas* receptor expression (Fig. S11D) and E13.5 EMB MNs did not express a higher level of *Fas* than E11.5 EMB MNs (Fig. S11E), suggesting that the differences in *Fas* expression do not reflect different stages of motor neuron maturity and are not caused by differential exposure to glia-derived factors.
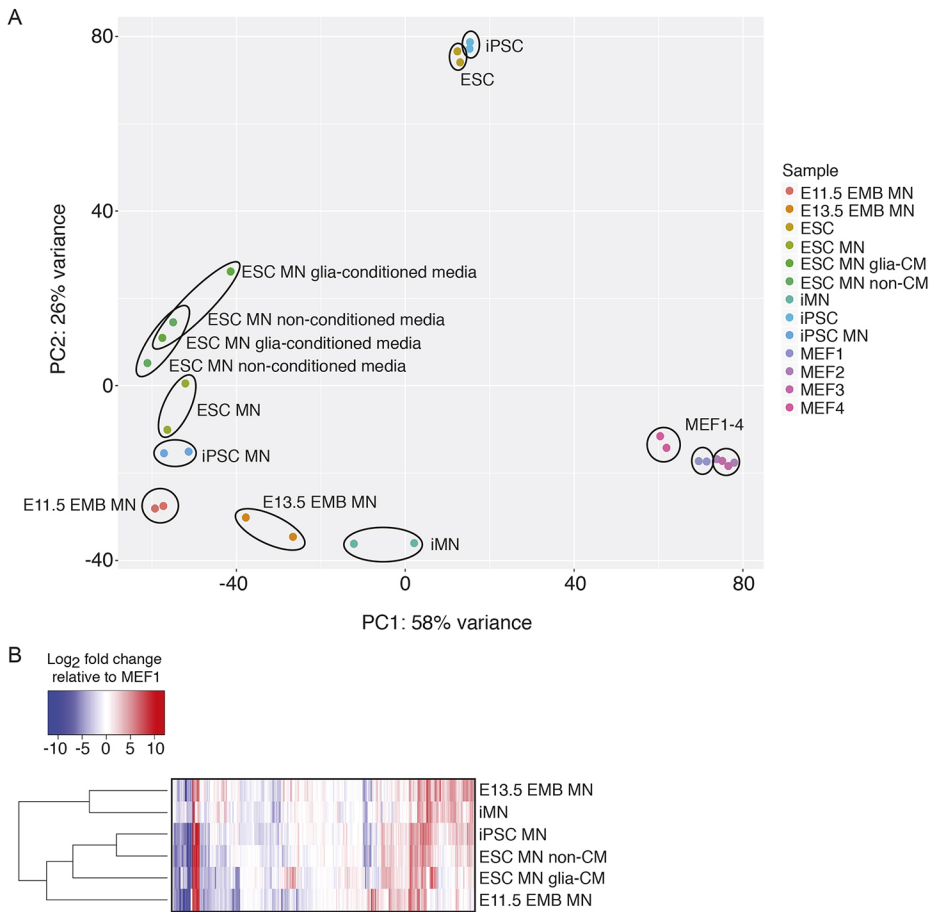
Fig. 5. Differences in maturation state amongst *in vitro* MNs. (A) Principle component analysis of all RNA-seq samples used in this study. CM, conditioned medium. (B) Dendrogram and heat map depicting unsupervised hierarchical clustering of motor neuron samples based upon genes previously identified (Ho et al., 2016) to be associated with MN maturation and aging.

Based on *Fas* expression levels, we hypothesized that iMNs would recapitulate Fas-induced ALS MN death in disease modeling experiments, while PSC MNs would not. Indeed, Fas ligand treatment induced the rapid degeneration of primary motor neurons and iMNs (Fig. 6B), but not ESC MNs (Fig. 6B). Therefore, although the differences in gene expression between *in vitro* motor neuron types are modest, genomic analysis provides valuable insight into their translational utility.

To identify motor neuron components that differ between iMNs and PSC MNs, we performed gene ontology analysis (cellular component, Panther). This revealed a large difference in synapse-related gene expression between iMNs and PSC MNs, with iMNs expressing a much larger number of synapse-related genes (Fig. 6C). Although both iMNs and PSC MNs are capable of forming functional neuromuscular junctions (Shi et al., 2018; Son et al., 2011; Toma et al., 2015), more detailed investigation may reveal differences between iMN and PSC MN neuromuscular junction components and function.

MN subtypes differ in their susceptibility to degenerative stimuli in ALS. Therefore, the ability to access motor neurons with selected subtype identities is particularly significant. The Hox code determines the anterior-posterior identity and target muscles of spinal motor neurons (Dasen et al., 2005). Closer inspection of Hox gene expression revealed that PSC MNs acquired a cervical identity as judged by expression of *Hoxa5*, *Hoxc4*, and *Hoxc5* (Fig. 6D,E), consistent with previous reports (Peljto et al., 2010; Son et al., 2011). Differentiation of ESC MNs in glia-conditioned medium did not increase the expression of thoracic or lumbar genes, including *Hoxc9* and *Hoxc10* (Fig. S12A). In contrast, iMNs and primary

MNs derived from the entire spinal cord exhibited more diverse Hox gene expression (Fig. 6D,E). The iMN profiles were very similar to those of MEFs (Fig. 6D,E) and we wondered whether the Hox code of iMNs was adopted from the starting MEFs. To determine this, we reprogrammed fibroblasts from posterior sections of mouse embryos. MEFs derived from the entire embryo expressed approximately equal levels of the anterior gene *Hoxc6* and the posterior gene *Hoxc9*, whereas posterior MEFs expressed high levels of *Hoxc9* and low levels of *Hoxc6*, indicating that selected harvesting of MEFs could enrich for posterior identity (Fig. 6F). Although MEFs and iMNs from whole embryos expressed similar levels of *Hoxc6* and *Hoxc9*, iMNs generated from posterior MEFs were highly enriched for *Hoxc9* mRNA expression and Hoxc9 protein, indicating they were posterior in character (Fig. 5F, Fig. S12B). Thus, the anterior-posterior identity of iMNs is dictated by the identity of the starting fibroblasts and some cellular pathways are conserved during lineage conversion. Further studies are required to understand how the Hox code may be preserved in iMNs despite the near complete conversion of epigenetic and transcriptional profiles.

DISCUSSION
Our study provides the first in-depth genomic and epigenetic comparison of a mammalian cell type produced by directed differentiation and lineage conversion. Consistent with other studies (Abernathy et al., 2017; Briggs et al., 2017; Mazzoni et al., 2013), the transcriptional and DNA methylation profiles indicate that both *in vitro* approaches generate MNs that closely approximate the natural state. Most Isl1 target genes are appropriately activated or suppressed, providing a rationale for the
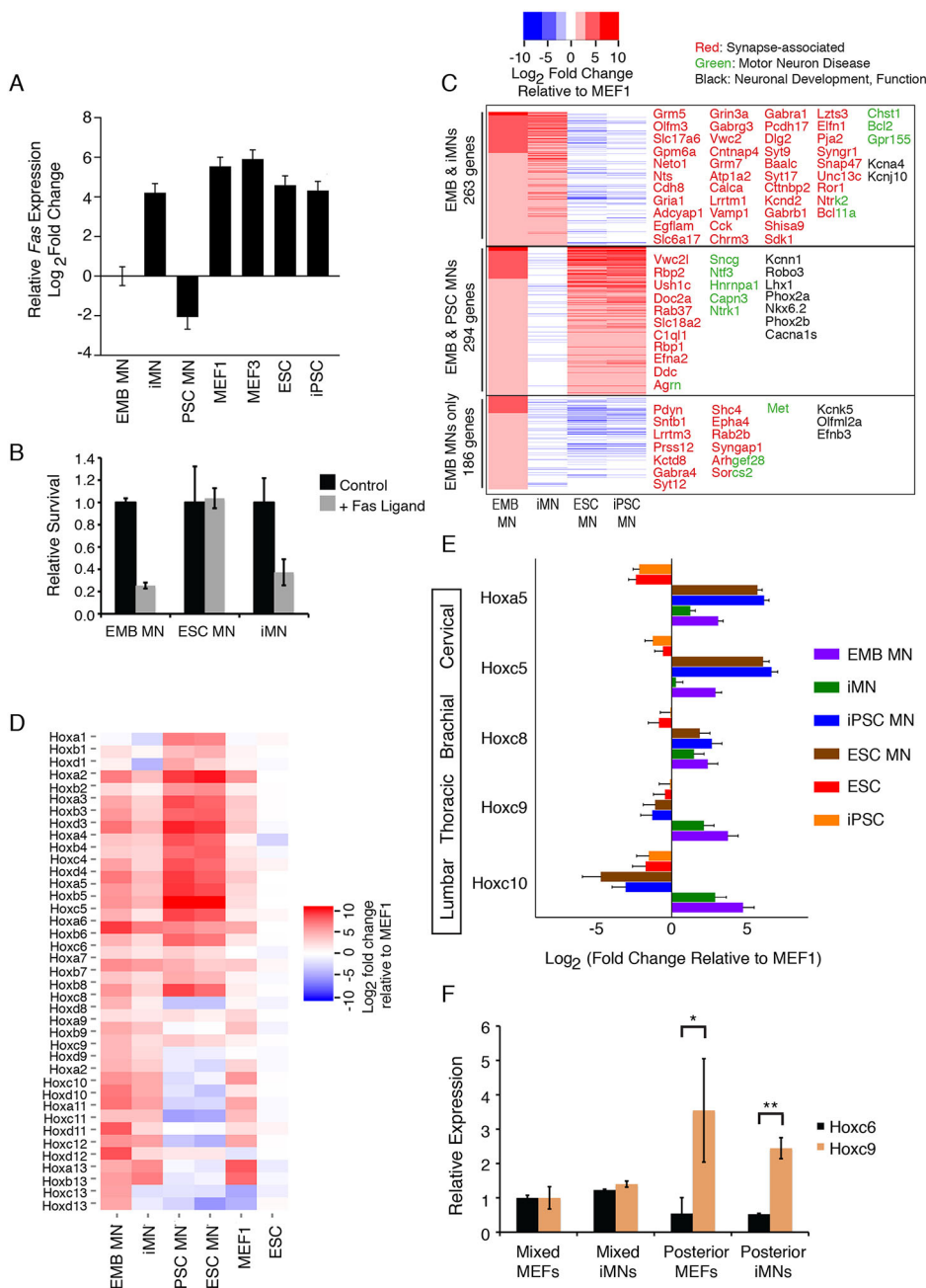
Fig. 6. Functional and etiological analysis of transcriptional differences between *in vitro* motor neuron types. (A) *Fas* gene expression relative to EMB MN, from the RNA-sequencing data. Data are mean±s.d., two biological replicates were analyzed per cell type. (B) Survival of MNs with or without the addition of 10 µg/ml agonistic anti-Fas antibody for 4 days. Anti-Fas antibody was added fresh to the cultures at days 0 and 2 of survival analysis. Data are mean±s.d., $n=2$ (EMB MN), $n=4$ (ESC MN) and $n=3$ (iMN). (C) Differential gene expression analysis between MNs. Expression is shown as $\log_2$ fold change relative to MEF1. Genes within each MN category are color coded according to their functional roles in MN biology. Two biological replicates were analyzed per sample type. (D) Full Hox code gene expression of MNs, MEF1 and ESCs relative to iPSCs. Two biological replicates were analyzed per cell type. (E) Gene expression relative to MEF1 derived from RNA-seq data and positional information of selected Hox genes. Data are mean±s.d. Two biological replicates were analyzed per cell type. (F) qRT-PCR analyses showing relative expression of *Hoxc6* and *Hoxc9* in mixed MEFs (derived from whole E13.5 embryos) or posterior MEFs (derived from posterior half of E13.5 embryos) or iMNs derived from these MEF populations. Each MEF group was derived from five individual embryos per replicate (a mean of three biological replicates±s.e.m.). Two-tailed *t*-test, unpaired. *$P<0.05$, **$P<0.01$.

faithful recapitulation of the MN transcriptional program. The DNA methylation profiles of *in vitro*-derived MNs closely match that of primary MNs and indicate that the methylation acts in concert with transcriptional regulators to establish a genuine MN state.

We made several important discoveries. First, transiting through a lineage-specific progenitor state is not required to produce a terminally differentiated cell with proper transcriptomic and DNA methylation states, as iMNs are highly similar to EMB MNs. This is consistent with a recent study that used single cell RNA-seq to verify that lineage-converted cells transit from ESCs through unnatural intermediate states *en route* to the MN state (Briggs et al., 2017). Although the single cell RNA-seq data did not allow comparison of the *in vitro*-derived and primary neuron transcriptomes at the same depth as our study, they did verify that individual *in vitro*-derived MNs are similar to individual primary MNs, indicating that our findings are not an artifact caused by a

mixture of neurons each expressing different parts of the MN transcriptional program.

Second, the differences between ESC MNs and iPSC MNs are greater than the differences between ESCs and iPSCs. There may be epigenetic abnormalities in iPSCs that do not significantly affect gene expression in the pluripotent state but alter directed differentiation or the resulting iPSC-derived cells. Alternatively, multiple cycles of cell division under the same culture conditions could be an important factor causing ESCs and iPSCs to converge.

Third, there are several non-overlapping aberrances between iMNs and PSC MNs. It may be optimal to use both in combination for disease studies, which would enable access to over 98% of the gene sets that are enriched in primary motor neurons (as determined by gene ontology enrichment analysis) (Fig. 3C). Studying Fas signaling in PSC MNs has been limited by the lack of *Fas* expression in PSC MNs (Fig. 5A). However, we have shown here

that iMNs have *Fas* expression and are sensitive to Fas-mediated degeneration (Fig. 5B), indicating they could enable investigation into this disease mechanism. One cause of differences between iMNs and ESC/iPSC MNs in our study is a difference in MN maturity, with iMNs being more mature than ESC/iPSC MNs at the time points analyzed (Fig. 5A and Fig. S11D). Although glia-conditioned medium did not affect maturation, we cannot rule out the possibility that direct contact with glial cells or signals from surrounding non-reprogrammed/non-differentiated cells modulate maturation. Consistent with our previous study (Son et al., 2011), about 5% of the MEFs gave rise to iMNs in our cultures. The yield of *Hb9*::GFP+ MNs in embryoid body differentiation was about 1-6%. Therefore, the percentage of non-MNs was similar in both paradigms, but differences in the types of cells comprising this fraction could elicit differences in MN maturation or gene expression between iMNs and ESC/iPSC MNs. Examination of later time points in differentiation will be important for more fully assessing the utility of iMNs and ESC/iPSC MNs for disease modeling.

A fourth discovery was that although by reprogramming and differentiation it is possible to generate cells that closely mimic their natural counterparts, not all *in vitro*-derived cell types do so. Improving culture conditions to better mimic the *in vivo* environment would likely increase the similarity of *in vitro* MNs to primary MNs, and we hypothesize this is true for other cell types as well. This may be one of the reasons why iPSCs are so similar to ESCs, whereas other reprogrammed cell types are not as similar to their primary counterparts. Additional signals or optimized timing of signal exposure could also bring about the full utility of these cells.

Finally, lineage-converted cells can inherit properties from their cells of origin, such as anterior-posterior identity, that could profoundly affect their character. Together, these observations argue that in depth evaluation of *in vitro*-derived cells is essential not only for verifying their quality but also for optimizing their use in disease and regenerative studies. To this end, our study demonstrates that transcriptional and DNA methylation maps provide crucial information for ensuring the maximal use of reprogrammed and differentiated cells.

MATERIALS AND METHODS
Generation of induced motor neurons
Mating pairs between 2 and 6 months old were used to generate mouse embryonic fibroblasts. MEFs were derived from *Hb9*::GFP C57Bl/6 E13.5 embryos after removing the head and spinal cord, and were passaged twice before reprogramming to eliminate any contaminating neurons. To generate iMNs, MEFs were transduced twice with *Ascl1*, *Brn2*, *Myt1l*, *Lhx3*, *Ngn2*, *Isl1* and *Mnx1* encoded in the pMXs retroviral backbone. Retrovirus was generated using the Plat-E cell line (Cell Biolabs, RV-101) and harvested 48 and 72 h after transfection. Two days after transduction, primary mouse cortical glia were added to the cultures. On day 3, the medium was changed to N3 medium [DMEM/F12, N2, B27 (Life Technologies), penicillin/streptomycin, 10 ng/ml GDNF, BDNF, CNTF (R&D Systems)], which was replaced every other day for the rest of the reprogramming process.

Derivation of mouse cortical glia
Glia were obtained from P2 ICR mice (mating pairs were 2-6 months old). Cortices were isolated and the meninges was removed in calcium- and magnesium-free Hanks BSS (HBSS) (Life Technologies, 14170-112). Tissue was digested using 0.25% trypsin-EDTA (Life Technologies) at 37°C for 15 min, triturated using a P1000 pipet tip and filtered through a 70 µm cell strainer (VWR, 21008-952). The dissociated cells were centrifuged at 1000× **g** for 5 min and plated on poly-d-lysine-coated dishes (VWR, 62406-044) in glia medium (Modified Eagle's Medium, Life

Technologies, 10370088), 20% glucose, 1× penicillin/streptomycin (Life Technologies, 15140122) and 10% horse serum (VWR, 16777-030). Glia were cultured for 2 weeks before use in experiments.

Generation of induced pluripotent stem cells
Hb9::GFP C57Bl/6 MEFs were transduced with *Oct4*, *Sox2* and *Klf4* encoded in the pMXs retroviral backbone. Two days after transduction, the medium was changed to KSR mESC medium [knockout DMEM, 20% knockout serum replacement, Glutamax, non-essential amino acids, penicillin/streptomycin (all Life Technologies) and LIF (EMD Millipore], which was maintained until colonies were picked on days 20-30. iPSCs were maintained on irradiated MEF feeders in mESC medium containing 15% FBS instead of KSR for at least six passages before use in experiments.

Generation of pluripotent stem cell-derived motor neurons
Motor neurons were generated by directed differentiation using retinoic acid and Smoothened agonist (EMD Millipore) according to Di Giorgio et al. (2007). Briefly, embryoid bodies were formed by dissociating *Hb9*::GFP ESCs or iPSCs on day 1 using 0.25% trypsin/EDTA and re-plated in non-adherent six-well plates at 400,000-800,000 cells/ml in EB medium [DMEM/F12, 10% knockout serum replacement, Glutamax, non-essential amino acids and penicillin/streptomycin (all Life Technologies)]. Retinoic acid (0.1 µM) and 1 µM smoothened agonist were added 48 h after embryoid body formation (on day 3) and replaced every 48 h until motor neuron harvest at day 7.

Culture of pluripotent stem cell-derived motor neurons with glia-conditioned media
EB medium was conditioned by incubating 6 ml of media for 48 h on a confluent 60 mm dish of primary cortical glia that was harvested 2-4 weeks prior to use for conditioning medium. During days 3-7 of directed differentiation, medium conditioned on mixed glia or on a plate without cells (as a negative control) was mixed 1:1 with non-conditioned medium. Retinoic acid (1 µM) and 1 µM smoothened agonist were added to the mixture and this was incubated with the embryoid bodies. The medium was replaced with a fresh 1:1 mixture of conditioned and non-conditioned medium every 48 h until harvest at day 7.

Flow purification of motor neurons
To purify *in vitro*-generated *Hb9*::GFP+ motor neurons, cultures were dissociated using papain/DNase (Worthington) in DMEM/F12 and flow-purified using a Moflo FACS sorter. To isolate primary embryonic motor neurons, spinal cords from E11.5 or E13.5 *Hb9*::GFP embryos were dissected out in PBS and treated with papain/DNase in DMEM/F12 for 15-30 min at 37°C. They were then triturated 10-20 times using a p1000 pipet before FACS sorting. Mating pairs were 2-6 months old at the time of conception.

Single cell qRT-PCR analysis of iMNs
Single *Hb9*::GFP+ iMNs were identified and isolated using an inverted microscope equipped with micromanipulator and micropipette. Cells were collected directly into 5 µl of CellsDirect 2× Buffer (Cells Direct One-step qRT-PCR kit, Thermo). Cells were processed using the manufacturer's protocol for reverse transcription (RT) and specific target amplification (STA). cDNA was synthesized and pre-amplified from single cell lysate. Single cell qPCR was performed using the Fluidigm BioMark HD system on amplified cDNA templates using primers validated in-house to yield efficient PCR amplification and SsoFast EvaGreen supermix.

RNA-seq
RNA was harvested from two biological replicates of each cell type or MEF condition using TRIzol (Invitrogen) according to the manufacturers' directions (Table S1). RNA quality was determined using BioAnalyzer (Aligent). RNA samples with RNA integrity numbers above 7.5 were used for library preparation. In brief, RNA-seq libraries were generated from 100-250 ng total RNA using the Illumina TruSeq RNA kit v.2, according to the manufacturer's directions. Only one library preparation per biological replicate was sequenced. Libraries were sequenced at the Broad Institute's

Genomics Platform on a HiSeq 2500. A total of 20-60 million 100 bp, paired end reads were obtained for each sample. Reads were mapped by STAR alignment (Dobin et al., 2013) against the murine genome build mm10, prior to transcriptomic analysis by DESeq2 (Love et al., 2014). The number of uniquely mapped read pairs is shown in Fig. S2A. Genes with less than an average of three reads and all genes on the X and Y chromosomes were excluded from analysis. After this, 16,077 genes remained for transcriptomic analysis. A genome-wide corrected false discovery rate (FDR) of less than 0.05 calculated by DESeq2 was considered significant. Computations were performed on the Odyssey cluster supported by the FAS Science Division Research Computing Group at Harvard University, and University of Southern California's Center for High-Performance Computing (hpc.usc.edu). For Fig. 3A, DESeq2 was used to identify genes that were differentially expressed between the starting cell and reference cell, and only these genes were used in the analyses. For genes that required upregulation to become expressed similarly to the reference cell, genes were considered successfully reprogrammed if they were upregulated to within the specified cut-off. For genes that required downregulation to become expressed similarly to the reference cell, genes were considered successfully reprogrammed if they were downregulated to within the specified cut-off. To eliminate noise associated with genes expressed at low levels in the reference cell or test cell, DESeq2 was used to identify and eliminate any genes whose expression level or other properties made them unreliable to compare between the reference and test cells. For Fig. 3D, GEOquery and Limma were used in R to analyze the microarray data. A genome-wide corrected false discovery rate (FDR) of less than 0.05 as determined by default settings in Limma was considered significant. Gene ontology analysis in Fig. 6C was performed with Panther geneontology.org/page/go-enrichment-analysis) using default settings. For Figs S10D-G and S11A, the following methods were used. Separately, libraries of the RNA samples from embryonic motor neurons day E11.5, ESC MNs cultured with conditioned media and ESC MNs without conditioned media were prepared using the Amaryllis Nucleics YourSeq Dual (FT & 3′-DGE) RNAseq Library Kit v1.5A in 3′-Digital Gene Expression mode. Sequencing was performed with an Illumina NextSeq 500 with single-read sequencing of 88 bp per read. 8 bp of each read was clipped, and subsequently trimmed for quality and adapter sequence by Trimmomatic (Bolger et al., 2014). Reads from both sample sets were aligned and counted with STAR (v.2.5.3a) (Dobin et al., 2013) on mouse genome build mm10, from UCSC. Genes with an average of less than three reads were removed from the analysis, as well as genes located on the X or Y chromosome or in the mitochondrial genome. Subsequent normalization of the reads occurred with DESeq2 (Love et al., 2014) (including the new samples the experimental design=~batch+celltype). A genome-wide corrected false discovery rate (FDR) of less than 0.05 was considered significant, as computed by DESeq2 as the adjusted P-value. RNA-seq data are available on GEO under accession number GSE89619.

Isl1 target identification from ChIP-seq data
For the empirically determined target list of Isl1, we accessed the ChIP-seq data at GEO accession GSE80321 (Velasco et al., 2017). After obtaining the BED files on GEO, we analyzed the described regions with GREAT (McLean et al., 2010) with mm10 and the settings 'Nearest single gene within a 1000 kb window' and included 'curated regulatory domains'. We continued with all genes that were present across all four provided time points (8 h, 12 h, 24 h and 48 h). Of the 3223 genes identified, 2619 were present in our dataset.

RRBS, data processing and analysis
Global, basepair-resolution measurements of DNA methylation was measured by Reduced Representation Bisulfite Sequencing (RRBS) as described at genomebiology.com/2012/13/10/r92. Briefly, genomic DNA from ESCs was digested using the *Msp*I enzyme, which cuts at C^CGG sites. Bisulfite treatment of DNA fragments was used to convert unmethylated cytosines to uracil, and this change was observed after sequencing and aligning library reads to the reference genome. Libraries were sequenced using the HiSeq 2500 platform, and aligned to an *in silico*

*Msp*I-digested mm9 genome using MAQ (http://maq.sourceforge.net/) and the Picard Pipeline (http://broadinstitute.github.io/picard/faq.html). We used custom Python scripts to call methylation at each CpG, computed as the ratio of methylated reads to total reads at that genomic position. These scripts can be obtained by from the authors on request.

Data accessions
For analysis of reprogramming success in Fig. 3D, publicly available data from the following GEO deposits were used: GSE22292 (Ieda et al., 2010), GSE23635 (Huang et al., 2011), GSE30102 (Marro et al., 2011), GSE12025 (Zhou et al., 2008), GSE30500 (Han et al., 2012), GSE36484 (Thier et al., 2012), GSE16925 (Zhao et al., 2009) and GSE25970 (Bock et al., 2011).

MN survival experiments with Fas ligand
Hb9::GFP+ MNs were FACS purified and plated on a confluent monolayer of primary cortical glia on laminin-coated dishes in 10 ng/ml rock inhibitor Y27632 (Selleck, S1049) in N3 medium+2% FBS. The following day, designated as day 0 of survival, the medium was replaced with N3 medium without rock inhibitor and 2% FBS, *Hb9*::GFP+ neurons were counted, and 10 ng/ml agonistic anti-Fas antibody clone J02 (BD Biosciences, 554255, validation provided by manufacturer) or PBS (vehicle control) was added to the culture medium. At day 2, fresh N3 medium with anti-Fas antibody was added to the neural cultures. At day 4, the number of *Hb9*::GFP+ neurons remaining in each well was counted and the percent survival was calculated using the starting number of *Hb9*::GFP neurons at day 0. Survival of iMNs was performed in biological triplicate, embryonic MN survival was performed in biological duplicate and ESC MN survival was performed in biological quadruplicate. Significance was determined using a two-tailed Student's *t*-test, unpaired.

Hox gene expression in posterior MEFs and iMNs
MEFs were isolated from *Hb9*::GFP+ embryos at E13.5 after removing the heads and spinal cords. Posterior MEFs were isolated from the posterior half of the embryos, using the midpoint between the forelimbs and hindlimbs as the midpoint. Mixed MEFs were isolated from the entire embryo after removing the heads and spinal cords. The posterior MEF and mixed MEF experiments were performed using five individual embryos each in order to control for variation in isolation of the posterior half of the embryos. MEFs were passaged twice to eliminate contaminating neurons and reprogrammed as described above. At day 14 post retroviral transduction, *Hb9*::GFP+ neurons were FACS purified, RNA was isolated by Trizol LS (Life Technologies) according to the manufacturer's instructions and qRT-PCR was performed using iScript Supermix (Bio-Rad, 1708841) on a Viia7 qRT-PCR machine (Life Technologies) using PCR primers. qRT-PCR experiments were performed using the following primers: *Hoxc6* forward – 5′-ACCTTAGGACATAACACACAGA; *Hoxc6* reverse – 5′-TCCAGTT-CCAGGGTCTGGTA; *Hoxc9* forward – 5′-ACCCGGACTACATGTAC-GGC; *Hoxc9* reverse–5′-TGGTACTTGGTGTAGGGGCA; *Gapdh* forward–5′-AGGTCGGTGTGAACGGATTTG; and *Gapdh* reverse–5′-TGTAGACCATGTAGTTGAGGTCA. Immunostaining for Hox gene analysis was performed using the following antibodies: anti-Hoxc6 (Aviva Systems Biology, ARP38484 P050, 1:4000 dilution) and anti-Hoxc9 (Abcam, ab50839, 1:5 dilution). Antibody validation information is available on the manufacturers' websites. Images were taken using an LSM 780 confocal microscope.

Cell lines
Plat-E cells used for generating retrovirus were obtained from Cell Biolabs (catalogue number RV-101). The presence of blasticidin- and puromycin-resistance genes, which were inserted into the cell line to allow antibiotic selection for the presence of the retroviral packaging genes, was confirmed by culturing the Plat-E cells in 10 μg/ml blasticidin and 1 μg/ml puromycin for 1 week. No additional authentication other than that provided by Cell Biolabs was performed. The *Hb9*::GFP embryonic stem cell line was derived from 3.5 days post-coitum embryos from a 2-month-old *Hb9*::GFP C57Bl/6 female. This cell line was validated using PCR genotyping for the *Hb9*::GFP transgene, and functionally verified by confirming GFP activation in motor neurons produced by directed differentiation. The

Hb9::GFP induced pluripotent stem cell line was derived from *Hb9*::GFP C57Bl/6 mouse embryonic fibroblasts that had been generated from an E13.5 embryo. The fibroblasts were reprogrammed by retroviral transduction of *Klf4*, *Sox2* and *Oct4*. This cell line was validated using PCR genotyping for the *Hb9*::GFP transgene, and functionally verified by confirming GFP activation in motor neurons produced by directed differentiation. All cell lines tested negative for mycoplasma before, during and after all experiments.

Animal guidance committees

All experiments involving live vertebrates (MEF, embryonic motor neuron and glial cell isolation) performed at Harvard or the University of Southern California were carried out in compliance with ethical regulations approved by the Harvard University or University of Southern California IACUC committees, respectively.

Methodology and statistics

Sample size was determined by power analysis (www.stat.ubc.ca/~rollin/stats/ssize/n2.html) based on preliminary experiments. One iMN sample was excluded from analysis due to the use of a different sorting gate during FACS sorting. Statistical analyses were performed as described in each figure legend, with verification that the data met the assumptions of the tests, including having similar variance between the groups tested. In addition, R Studio, Cytoscape with ClueGO (Bindea et al., 2009) and GraphPad Prism were used for analyses.

References

Abernathy, D. G., Kim, W. K., McCoy, M. J., Lake, A. M., Ouwenga, R., Lee, S. W., Xing, X., Li, D., Lee, H. J., Heuckeroth, R. O. et al. (2017). MicroRNAs induce a permissive chromatin environment that enables neuronal subtype-specific reprogramming of adult human fibroblasts. *Cell Stem Cell* **21**, 332-348.e339.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.-H., Pagès, F., Trajanoski, Z. and Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091-1093.

Blanchard, J. W., Eade, K. T., Szűcs, A., Lo Sardo, V., Tsunemoto, R. K., Williams, D., Sanna, P. P. and Baldwin, K. K. (2015). Selective conversion of fibroblasts into peripheral sensory neurons. *Nat. Neurosci.* **18**, 25-35.

Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z. D., Ziller, M., Croft, G. F., Amoroso, M. W., Oakley, D. H. et al. (2011). Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144**, 439-452.

Bolger, A. M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120.

Brennand, K. J., Simone, A., Jou, J., Gelboin-Burkhart, C., Tran, N., Sangar, S., Li, Y., Mu, Y., Chen, G., Yu, D. et al. (2011). Modelling schizophrenia using human induced pluripotent stem cells. *Nature* **473**, 221-225.

Briggs, J. A., Li, V. C., Lee, S., Woolf, C. J., Klein, A. and Kirschner, M. W. (2017). Mouse embryonic stem cells can differentiate via multiple paths to the same state. *eLife* **6**, e26945.

Dasen, J. S., Tice, B. C., Brenner-Morton, S. and Jessell, T. M. (2005). A Hox regulatory network establishes motor neuron pool identity and target-muscle connectivity. *Cell* **123**, 477-491.

Di Giorgio, F. P., Carrasco, M. A., Siao, M. C., Maniatis, T. and Eggan, K. (2007). Non-cell autonomous effect of glia on motor neurons in an embryonic stem cell-based ALS model. *Nat. Neurosci.* **10**, 608-614.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21.

Gopalakrishnan, S., Hor, P. and Ichida, J. K. (2015). New approaches for direct conversion of patient fibroblasts into neural cells. *Brain Res.* **1656**, 2-13.

Gore, A., Li, Z., Fung, H.-L., Young, J. E., Agarwal, S., Antosiewicz-Bourget, J., Canto, I., Giorgetti, A., Israel, M. A., Kiskinis, E. et al. (2011). Somatic coding mutations in human induced pluripotent stem cells. *Nature* **471**, 63-67.

Gu, H., Smith, Z. D., Bock, C., Boyle, P., Gnirke, A. and Meissner, A. (2011). Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat. Protoc.* **6**, 468-481.

Han, S. S. W., Williams, L. A. and Eggan, K. C. (2011). Constructing and deconstructing stem cell models of neurological disease. *Neuron* **70**, 626-644.

Han, D. W., Tapia, N., Hermann, A., Hemmer, K., Hoing, S., Araúzo-Bravo, M. J., Zaehres, H., Wu, G., Frank, S., Moritz, S. et al. (2012). Direct reprogramming of fibroblasts into neural stem cells by defined factors. *Cell Stem Cell* **10**, 465-472.

Ho, R., Sances, S., Gowing, G., Amoroso, M. W., O'Rourke, J. G., Sahabian, A., Wichterle, H., Baloh, R. H., Sareen, D. and Svendsen, C. N. (2016). ALS disrupts spinal motor neuron maturation and aging pathways within gene co-expression networks. *Nat. Neurosci.* **19**, 1256-1267.

Huang, P., He, Z., Ji, S., Sun, H., Xiang, D., Liu, C., Hu, Y., Wang, X. and Hui, L. (2011). Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature* **475**, 386-389.

Ichida, J. K. and Kiskinis, E. (2015). Probing disorders of the nervous system using reprogramming approaches. *EMBO J.* **34**, 1456-1477.

Ichida, J. K., Blanchard, J., Lam, K., Son, E. Y., Chung, J. E., Egli, D., Loh, K. M., Carter, A. C., Di Giorgio, F. P., Koszka, K. et al. (2009). A small-molecule inhibitor of Tgf-beta signaling replaces sox2 in reprogramming by inducing nanog. *Cell Stem Cell* **5**, 491-503.

Ichida, J. K., Tcw, J., Williams, L. A., Carter, A. C., Shi, Y., Moura, M. T., Ziller, M., Singh, S., Amabile, G., Bock, C. et al. (2014). Notch inhibition allows oncogene-independent generation of iPS cells. *Nat. Chem. Biol.* **10**, 632-639.

Ieda, M., Fu, J.-D., Delgado-Olguin, P., Vedantham, V., Hayashi, Y., Bruneau, B. G. and Srivastava, D. (2010). Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142**, 375-386.

Jessell, T. M. (2000). Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nat. Rev. Genet.* **1**, 20-29.

Kiskinis, E., Sandoe, J., Williams, L. A., Boulting, G. L., Moccia, R., Wainger, B. J., Han, S., Peng, T., Thams, S., Mikkilineni, S. et al. (2014). Pathways disrupted in human ALS motor neurons identified through genetic correction of mutant SOD1. *Cell Stem Cell* **14**, 781-795.

Kramer, N. J., Haney, M. S., Morgens, D. W., Jovičić, A., Couthouis, J., Li, A., Ousey, J., Ma, R., Bieri, G., Tsui, C. K. et al. (2018). CRISPR-Cas9 screens in human cells and primary neurons identify modifiers of C9ORF72 dipeptide-repeat-protein toxicity. *Nat. Genet.* **50**, 603-612.

Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.

Marro, S., Pang, Z. P., Yang, N., Tsai, M.-C., Qu, K., Chang, H. Y., Südhof, T. C. and Wernig, M. (2011). Direct lineage conversion of terminally differentiated hepatocytes to functional neurons. *Cell Stem Cell* **9**, 374-382.

DEVELOPMENT

Mazzoni, E. O., Mahony, S., Closser, M., Morrison, C. A., Nedelec, S., Williams, D. J., An, D., Gifford, D. K. and Wichterle, H. (2013). Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity. *Nat. Neurosci.* **16**, 1219-1227.

McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M. and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495-501.

Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B. et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766-770.

Merkle, F. T., Ghosh, S., Kamitaki, N., Mitchell, J., Avior, Y., Mello, C., Kashin, S., Mekhoubad, S., Ilic, D., Charlton, M. et al. (2017). Human pluripotent stem cells recurrently acquire and expand dominant negative P53 mutations. *Nature* **545**, 229-233.

Mertens, J., Paquola, A. C. M., Ku, M., Hatch, E., Böhnke, L., Ladjevardi, S., McGrath, S., Campbell, B., Lee, H., Herdy, J. R. et al. (2015). Directly reprogrammed human neurons retain aging-associated transcriptomic signatures and reveal age-related nucleocytoplasmic defects. *Cell Stem Cell* **17**, 705-718.

Nakagawa, M., Takizawa, N., Narita, M., Ichisaka, T. and Yamanaka, S. (2010). Promotion of direct reprogramming by transformation-deficient Myc. *Proc. Natl. Acad. Sci. USA* **107**, 14152-14157.

Peljto, M., Dasen, J. S., Mazzoni, E. O., Jessell, T. M. and Wichterle, H. (2010). Functional diversity of ESC-derived motor neuron subtypes revealed through intraspinal transplantation. *Cell Stem Cell* **7**, 355-366.

Pepper, J.-P., Wang, T. V., Hennes, V., Sun, S. Y. and Ichida, J. K. (2017). Human induced pluripotent stem cell-derived motor neuron transplant for neuromuscular atrophy in a mouse model of sciatic nerve injury. *JAMA Facial Plastic Surgery* **19**, 197-205.

Raoul, C., Estevez, A. G., Nishimune, H., Cleveland, D. W., deLapeyriere, O., Henderson, C. E., Haase, G. and Pettmann, B. (2002). Motoneuron death triggered by a specific pathway downstream of Fas. potentiation by ALS-linked SOD1 mutations. *Neuron* **35**, 1067-1083.

Sances, S., Bruijn, L. I., Chandran, S., Eggan, K., Ho, R., Klim, J. R., Livesey, M. R., Lowry, E., Macklis, J. D., Rushton, D. et al. (2016). Modeling ALS with motor neurons derived from human induced pluripotent stem cells. *Nat. Neurosci.* **19**, 542-553.

Sandoe, J. and Eggan, K. (2013). Opportunities and challenges of pluripotent stem cell neurodegenerative disease models. *Nat. Neurosci.* **16**, 780-789.

Shi, Y., Lin, S., Staats, K. A., Li, Y., Chang, W.-H., Hung, S.-T., Hendricks, E., Linares, G. R., Wang, Y., Son, E. Y. et al. (2018). Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons. *Nat. Med.* **24**, 313-325.

Son, E. Y., Ichida, J. K., Wainger, B. J., Toma, J. S., Rafuse, V. F., Woolf, C. J. and Eggan, K. (2011). Conversion of mouse and human fibroblasts into functional spinal motor neurons. *Cell Stem Cell* **9**, 205-218.

Takahashi, K., Okita, K., Nakagawa, M. and Yamanaka, S. (2007a). Induction of pluripotent stem cells from fibroblast cultures. *Nat. Protoc.* **2**, 3081-3089.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S. (2007b). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-872.

Thier, M., Wörsdörfer, P., Lakes, Y. B., Gorris, R., Herms, S., Opitz, T., Seiferling, D., Quandel, T., Hoffmann, P., Nöthen, M. M. et al. (2012). Direct conversion of fibroblasts into stably expandable neural stem cells. *Cell Stem Cell* **10**, 473-479.

Toma, J. S., Shettar, B. C., Chipman, P. H., Pinto, D. M., Borowska, J. P., Ichida, J. K., Fawcett, J. P., Zhang, Y., Eggan, K. and Rafuse, V. F. (2015). Motoneurons derived from induced pluripotent stem cells develop mature phenotypes typical of endogenous spinal motoneurons. *J. Neurosci.* **35**, 1291-1306.

Tsuchida, T., Ensini, M., Morton, S. B., Baldassare, M., Edlund, T., Jessell, T. M. and Pfaff, S. L. (1994). Topographic organization of embryonic motor neurons defined by expression of LIM homeobox genes. *Cell* **79**, 957-970.

Velasco, S., Ibrahim, M. M., Kakumanu, A., Garipler, G., Aydin, B., Al-Sayegh, M. A., Hirsekorn, A., Abdul-Rahman, F., Satija, R., Ohler, U. et al. (2017). A multi-step transcriptional and chromatin state cascade underlies motor neuron programming from embryonic stem cells. *Cell Stem Cell* **20**, 205-217.e208.

Wainger, B. J., Buttermore, E. D., Oliveira, J. T., Mellin, C., Lee, S., Saber, W. A., Wang, A. J., Ichida, J. K., Chiu, I. M., Barrett, L. et al. (2015). Modeling pain in vitro using nociceptor neurons reprogrammed from fibroblasts. *Nat. Neurosci.* **18**, 17-24.

Wen, X., Tan, W., Westergard, T., Krishnamurthy, K., Markandaiah, S. S., Shi, Y., Lin, S., Shneider, N. A., Monaghan, J., Pandey, U. B. et al. (2014). Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate in vitro and in vivo neuronal death. *Neuron* **84**, 1213-1225.

Wichterle, H., Lieberam, I., Porter, J. A. and Jessell, T. M. (2002). Directed differentiation of embryonic stem cells into motor neurons. *Cell* **110**, 385-397.

Wilkinson, B., Evgrafov, O. V., Zheng, D., Hartel, N., Knowles, J. A., Graham, N. A., Ichida, J. K. and Coba, M. P. (2018). Endogenous cell type-specific disrupted in schizophrenia 1 interactomes reveal protein networks associated with neurodevelopmental disorders. *Biol. Psychiatry* (in press).

Xu, J., Du, Y. and Deng, H. (2015). Direct lineage reprogramming: strategies, mechanisms, and applications. *Cell Stem Cell* **16**, 119-134.

Zhang, K., Donnelly, C. J., Haeusler, A. R., Grima, J. C., Machamer, J. B., Steinwald, P., Daley, E. L., Miller, S. J., Cunningham, K. M., Vidensky, S. et al. (2015). The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* **525**, 56-61.

Zhao, X.-Y., Li, W., Lv, Z., Liu, L., Tong, M., Hai, T., Hao, J., Guo, C.-L., Ma, Q.-W., Wang, L. et al. (2009). iPS cells produce viable mice through tetraploid complementation. *Nature* **461**, 86-90.

Zhao, Z., Sagare, A. P., Ma, Q., Halliday, M. R., Kong, P., Kisler, K., Winkler, E. A., Ramanathan, A., Kanekiyo, T., Bu, G. et al. (2015). Central role for PICALM in amyloid-beta blood-brain barrier transcytosis and clearance. *Nat. Neurosci.* **18**, 978-987.

Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. and Melton, D. A. (2008). In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature* **455**, 627-632.

Ziller, M. J., Gu, H., Müller, F., Donaghey, J., Tsai, L. T.-Y., Kohlbacher, O., De Jager, P. L., Rosen, E. D., Bennett, D. A., Bernstein, B. E. et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**, 477-481.