A.

n.s.

B.

Positive expression =
Markers expressed at > 50%
of the mean value

Positive expression =
Markers expressed at > 25%
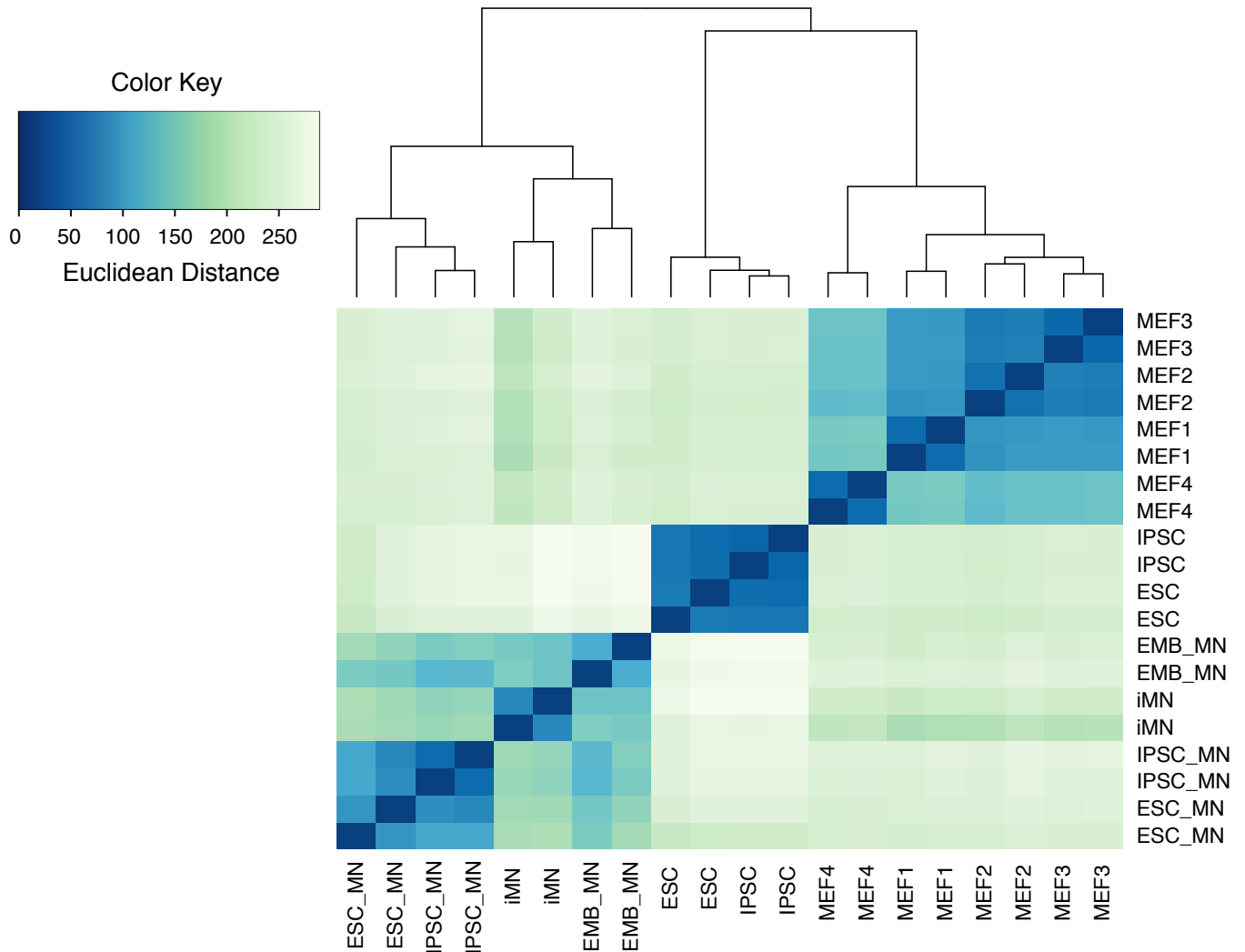of the mean value

Figure S1. Comparison of motor neuron survival and subtype properties. A) Survival of primary and *in vitro*-derived motor neurons over a four-day period on a monolayer of primary cortical glia in medium with 10ng/ml GDNF, BDNF, and CNTF. Values were calculated from two (embryo MN and iMN) or four (ESC MN) biological replicates. Error bars denote standard deviation. Statistical significance was calculated using one-way ANOVA. * - $p < .05$. B) Single cell qRT-PCR analysis of Hb9+ iMNs expressing endogenous *Isl1*, *Lhx3*, or both *Isl1* and *Lhx3*. A total of 62 Hb9+ iMNs were analyzed from two independent lineage conversions. The graph show the fraction of iMNs expressing endogenous *Isl1* or *Lhx3* if those genes were required to be expressed at a level no less than 50% (left graph) or 25% (right graph) of the mean value of that gene amongst the iMNs analyzed.

A.

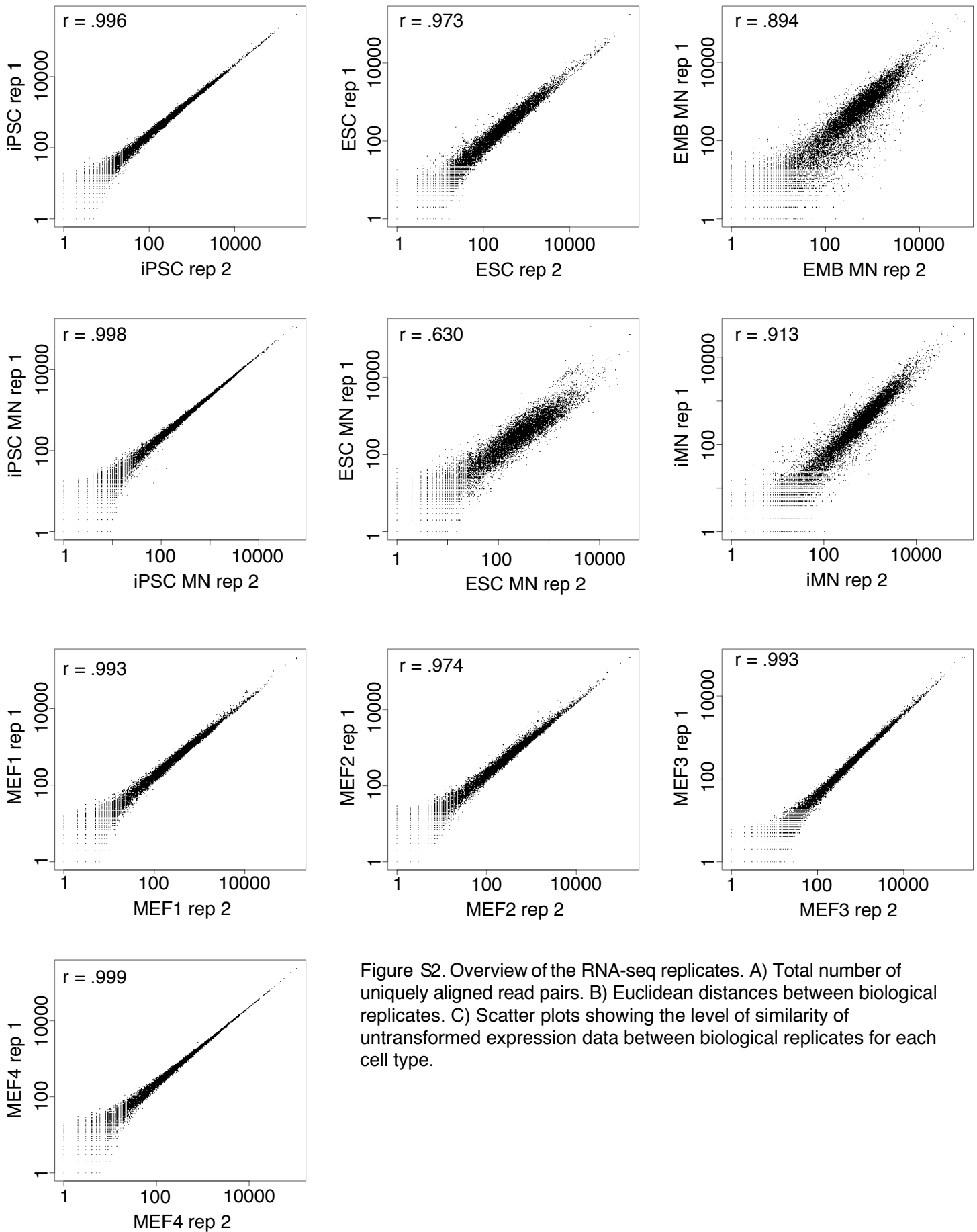| Sample	Read Pairs	Sample	Read Pairs
EMB_MN rep1	16339620	IPSC rep1	14333103
EMB_MN rep2	24687703	IPSC rep2	30910218
ESC_MN rep1	7885038	MEF2 rep1	11670915
ESC_MN rep2	11366424	MEF2 rep2	21967205
ESC rep1	13069853	MEF1 rep1	9890309
ESC rep2	8287515	MEF1 rep2	16746981
iMN rep1	19416554	MEF3 rep1	21728914
iMN rep2	10854935	MEF3 rep2	8503474
IPSC_MN rep1	10787937	MEF4 rep1	9881073
IPSC_MN rep2	23232916	MEF4 rep2	22401592

B.

C.

Figure S2. Overview of the RNA-seq replicates. A) Total number of uniquely aligned read pairs. B) Euclidean distances between biological replicates. C) Scatter plots showing the level of similarity of untransformed expression data between biological replicates for each cell type.
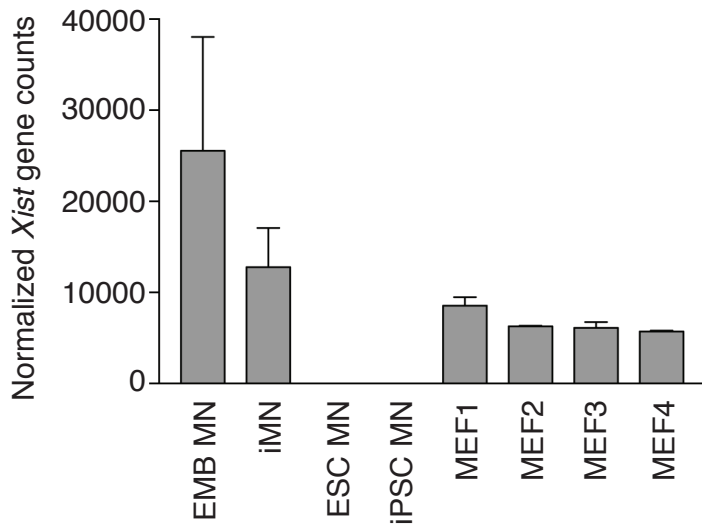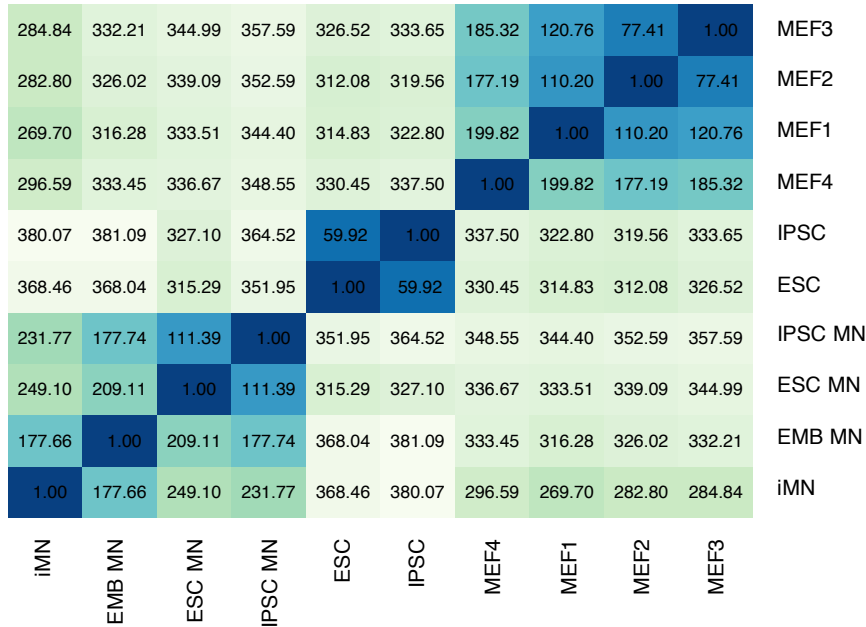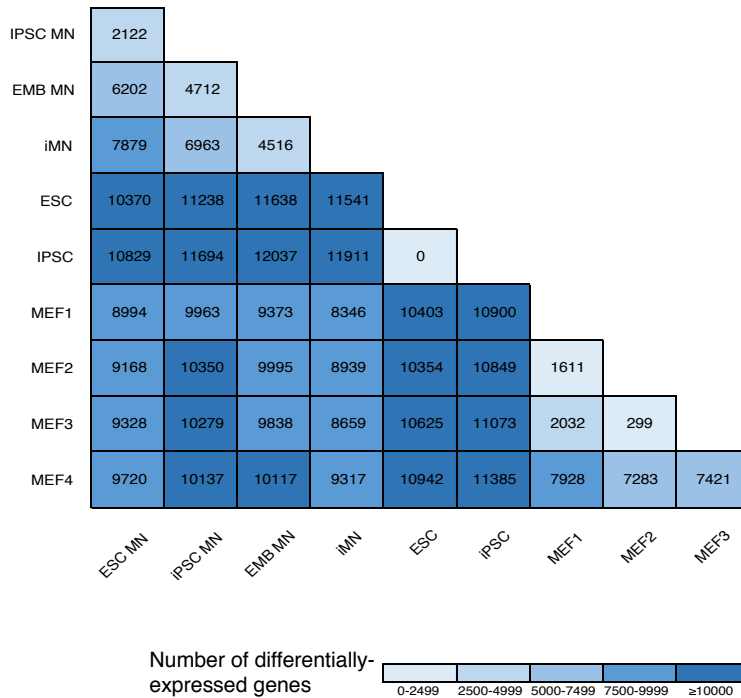
Figure S3. Transcription of *Xist* in motor neuron and MEF samples. Values are the normalized gene counts from DESeq2, and are obtained by the RNA sequencing of two biological replicates per cell type. Mean +/- standard deviation.

A.

284.84	332.21	344.99	357.59	326.52	333.65	185.32	120.76	77.41	1.00	MEF3
282.80	326.02	339.09	352.59	312.08	319.56	177.19	110.20	1.00	77.41	MEF2
269.70	316.28	333.51	344.40	314.83	322.80	199.82	1.00	110.20	120.76	MEF1
296.59	333.45	336.67	348.55	330.45	337.50	1.00	199.82	177.19	185.32	MEF4
380.07	381.09	327.10	364.52	59.92	1.00	337.50	322.80	319.56	333.65	IPSC
368.46	368.04	315.29	351.95	1.00	59.92	330.45	314.83	312.08	326.52	ESC
231.77	177.74	111.39	1.00	351.95	364.52	348.55	344.40	352.59	357.59	IPSC MN
249.10	209.11	1.00	111.39	315.29	327.10	336.67	333.51	339.09	344.99	ESC MN
177.66	1.00	209.11	177.74	368.04	381.09	333.45	316.28	326.02	332.21	EMB MN
1.00	177.66	249.10	231.77	368.46	380.07	296.59	269.70	282.80	284.84	iMN

Columns: iMN, EMB MN, ESC MN, IPSC MN, ESC, IPSC, MEF4, MEF1, MEF2, MEF3

B.

	ESC MN	iPSC MN	EMB MN	iMN	ESC	iPSC	MEF1	MEF2	MEF3
IPSC MN	2122								
EMB MN	6202	4712							
iMN	7879	6963	4516						
ESC	10370	11238	11638	11541					
IPSC	10829	11694	12037	11911	0				
MEF1	8994	9963	9373	8346	10403	10900			
MEF2	9168	10350	9995	8939	10354	10849	1611		
MEF3	9328	10279	9838	8659	10625	11073	2032	299	
MEF4	9720	10137	10117	9317	10942	11385	7928	7283	7421

Number of differentially-expressed genes

0-2499   2500-4999   5000-7499   7500-9999   ≥10000

C.

Number of genes increased in each MN vs MEF1

EMB MNs, ESC MNs, iMNs, iPSC MNs

499, 493, 68, 507, 31, 85, 389, 41, 114, 484, 2387, 641, 464, 770, 217

Number of genes decreased in each MN vs MEF1

EMB MNs, ESC MNs, iMNs, iPSC MNs

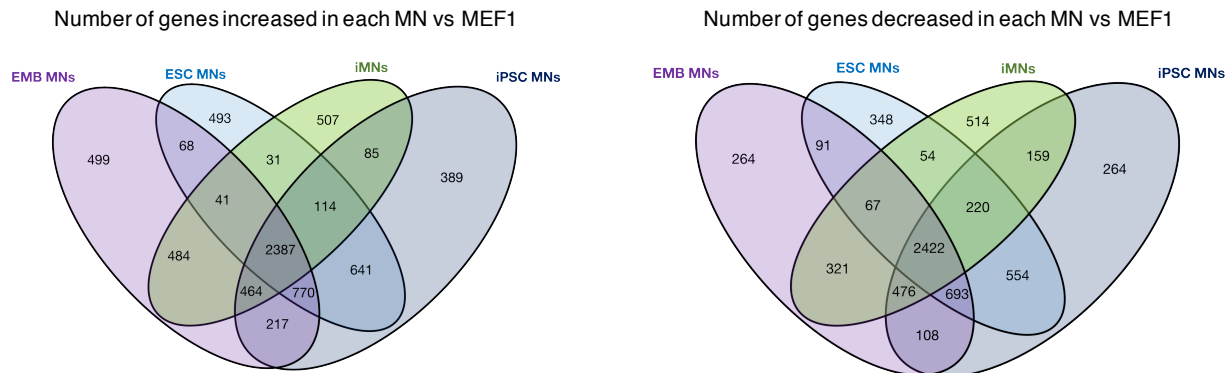264, 348, 91, 514, 54, 159, 264, 67, 220, 321, 2422, 554, 476, 693, 108

Figure S4. Global analysis of transcriptional data. A) Distance matrix between groups by the sum of squared shrunken log2 fold changes. B) Total number of differentially expressed genes between each pairwise comparison. C). Overlap of genes significantly increased or decreased relative to MEF1 between motor neuron types. Two biological replicates were analyzed per sample type.
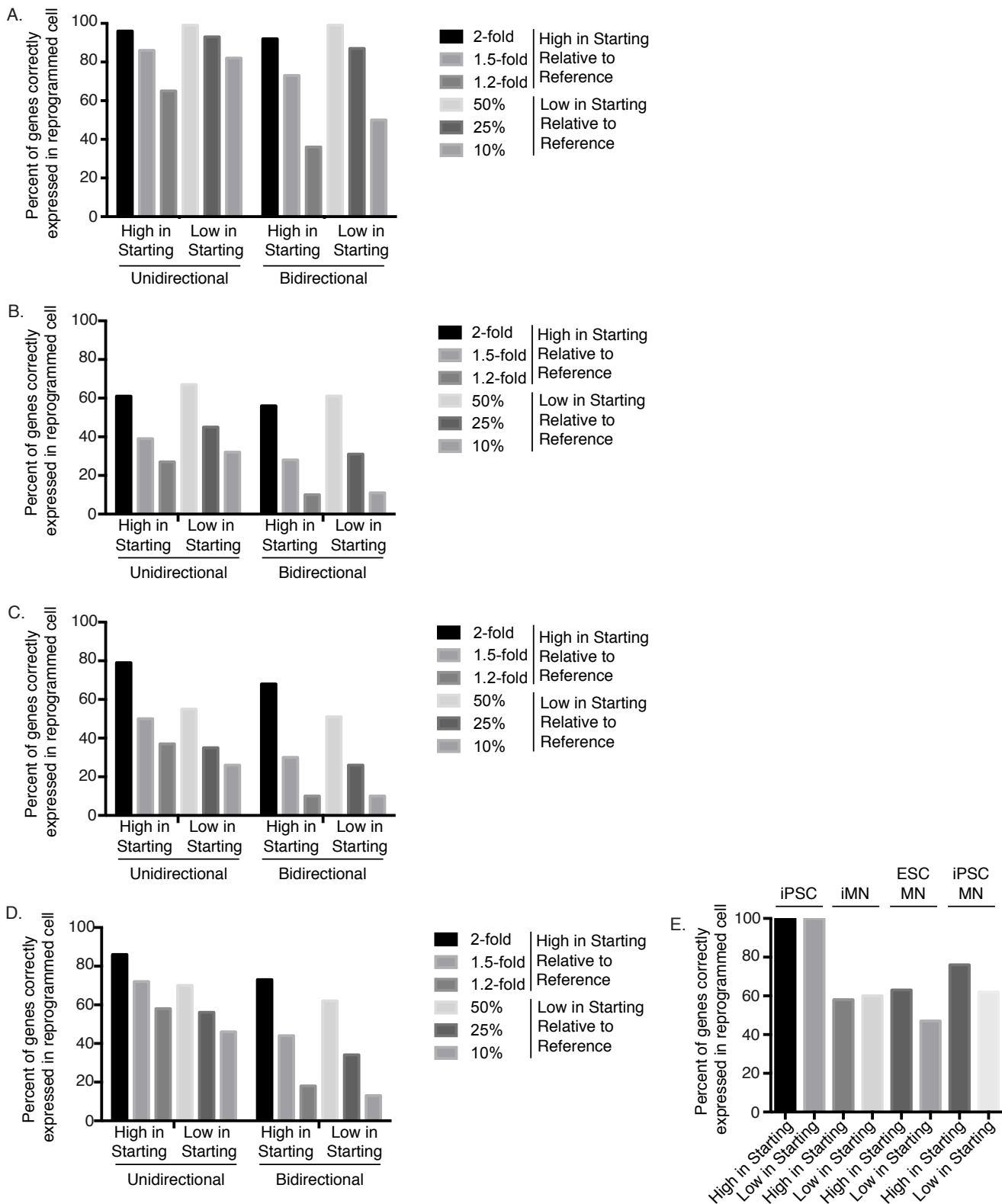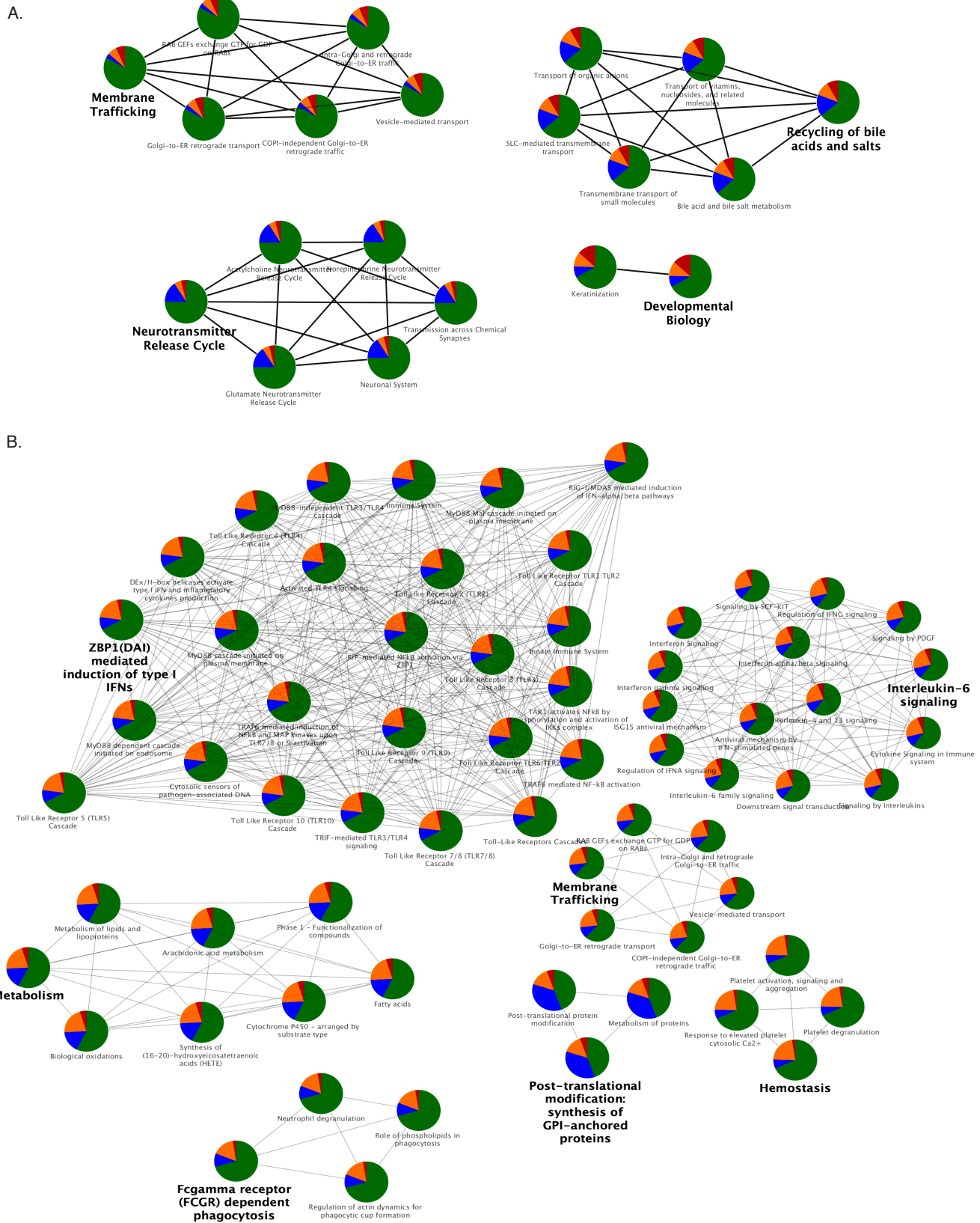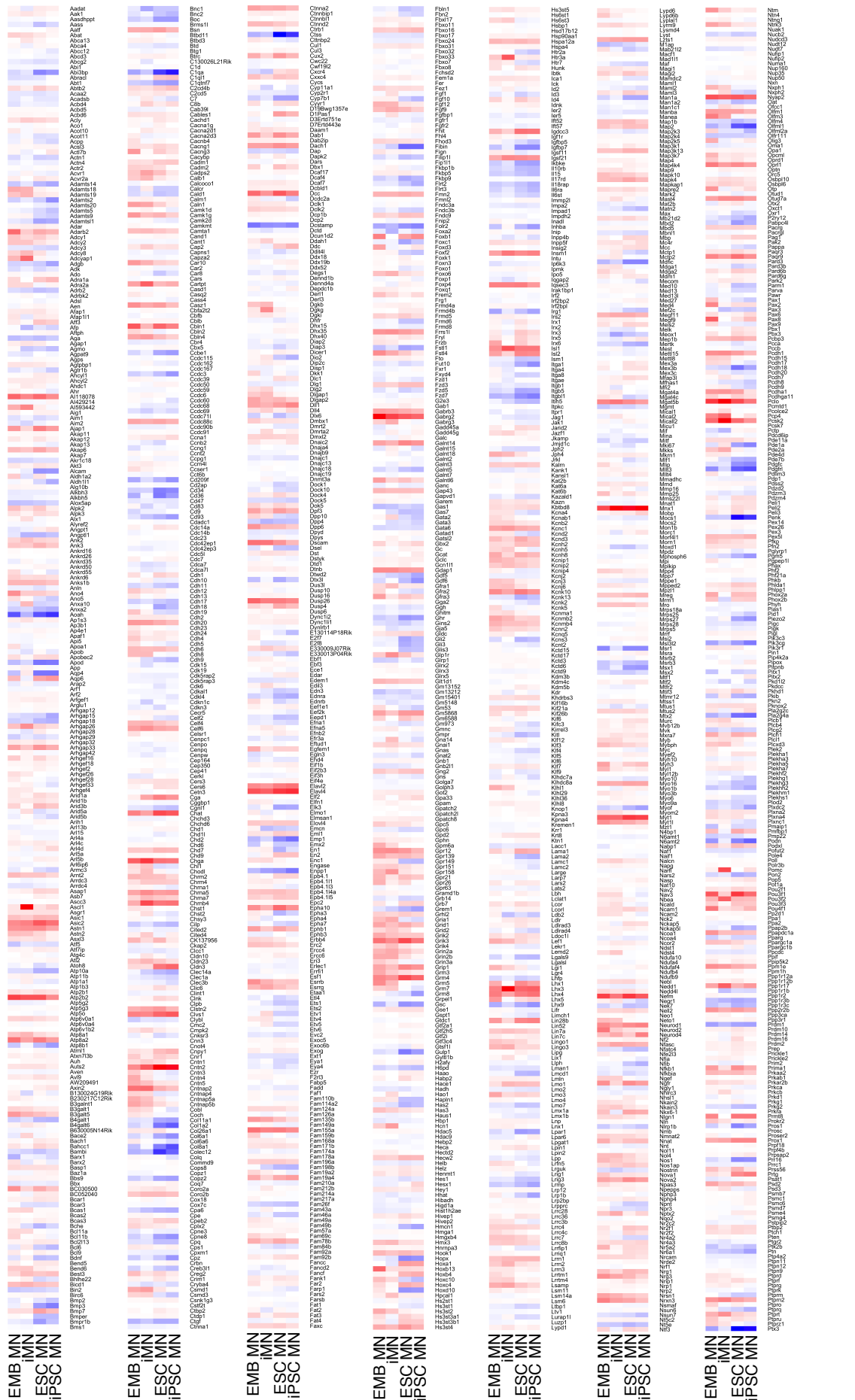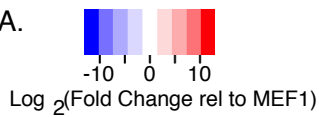
Figure S5. Percent of genes correctly expressed after lineage conversion or directed differentiation. The percent of genes that needed to be repressed or activated to become correctly expressed in (A) lineage conversion into iPSCs, (B) lineage conversion into iMNs, (C) directed differentiation into ESC MNs, or (D) directed differentiation into iPSC MNs. "High in Starting Relative to Reference" represents genes that were expressed higher in the starting cell than the reference cell and needed to be repressed during reprogramming. "Low in Starting Relative to Reference" represents genes that were expressed lower in the starting cell than the reference cell and needed to be increased during reprogramming. "Unidirectional" pertains to analyses in which genes were considered correctly expressed after lineage conversion or directed differentiation as long as they changed to a level greater than 50%, 25%, or 10% below the reference cell level for genes that needed to be increased, or to a level less than 2-, 1.5-, or 1.2-fold of the reference cell level for genes that needed to be repressed. "Bidirectional" pertains to the same analyses as "unidirectional" but in addition, genes that needed to be repressed could not be repressed to a level less than 50%, 25%, or 10% below the reference cell and genes that needed to be activated could not be over-activated to more than 2-, 1.5, or 1.2-fold the reference cell level. (E) Percent of genes in iPSCs, iMNs, ESC MNs, and iPSC MNs that were differentially expressed between the starting and reference cells but are not differentially expressed between the test and reference cells according to DESeq2. Two biological replicates were analyzed per sample type.
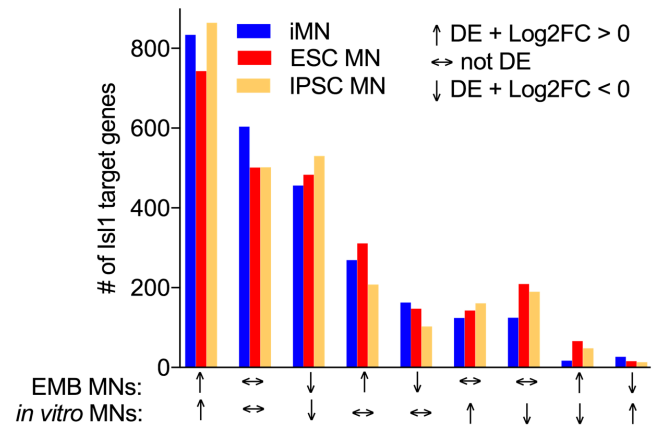
Figure S6. Functional analysis of genes differentially expressed between motor neurons and MEFs. A) Functional analysis of genes upregulated or B) downregulated in MNs vs MEFS. ClueGO ontology enrichment of Reactome terms was performed on all genes upregulated or downregulated (as determined by DESeq2) in embryonic MNs relative to MEF1 . Enriched terms (q <=0.05 after Benjamini-Hochberg multiple testing correction) are displayed as nodes with edges representing the kappa statistic of association determined by clueGO. The resulting network layout was performed using the Organic layout algorithm in Cytoscape. Representative functions of each node cluster are indicated. Enrichment of Reactome terms among significantly upregulated genes in iMNs, iPSC MNs, and ESC MNs was also performed. Nodes of the EMB MN network are colored based on co-enrichment in indicated groups where "PSC MN" refers to enrichment in either iPS MN or ES MN. The colors indicate that the term was enriched among genes up- or downregulated in all MNs (green), EMB MNs and iMNs (blue), EMB MNs and PSC MNs (orange), or EMB MNs only (red) relative to MEF1. Two biological replicates were analyzed per sample type.

A.

Log $_2$(Fold Change rel to MEF1)

A. (continued)

B.

Figure S7. Transcriptomic analysis of Isl1 target genes in primary and *in vitro*-derived motor neurons. A) Gene expression analysis (derived from RNA-seq data) of known ISL1 target genes identified by Chip-Seq analysis (Mazzoni *et al.*, 2013).
B) Number of Isl1 target genes differentially expressed in EMB MNs or *in vitro* MNs in comparison to MEF1 (DESeq2) and the directionality of the gene expression change relative to MEF1. DESeq2 was used to restrict the analyses to genes that could be reliably compared between samples. Two biological replicates were analyzed per cell type. DE = differentially expressed, FC = fold change.
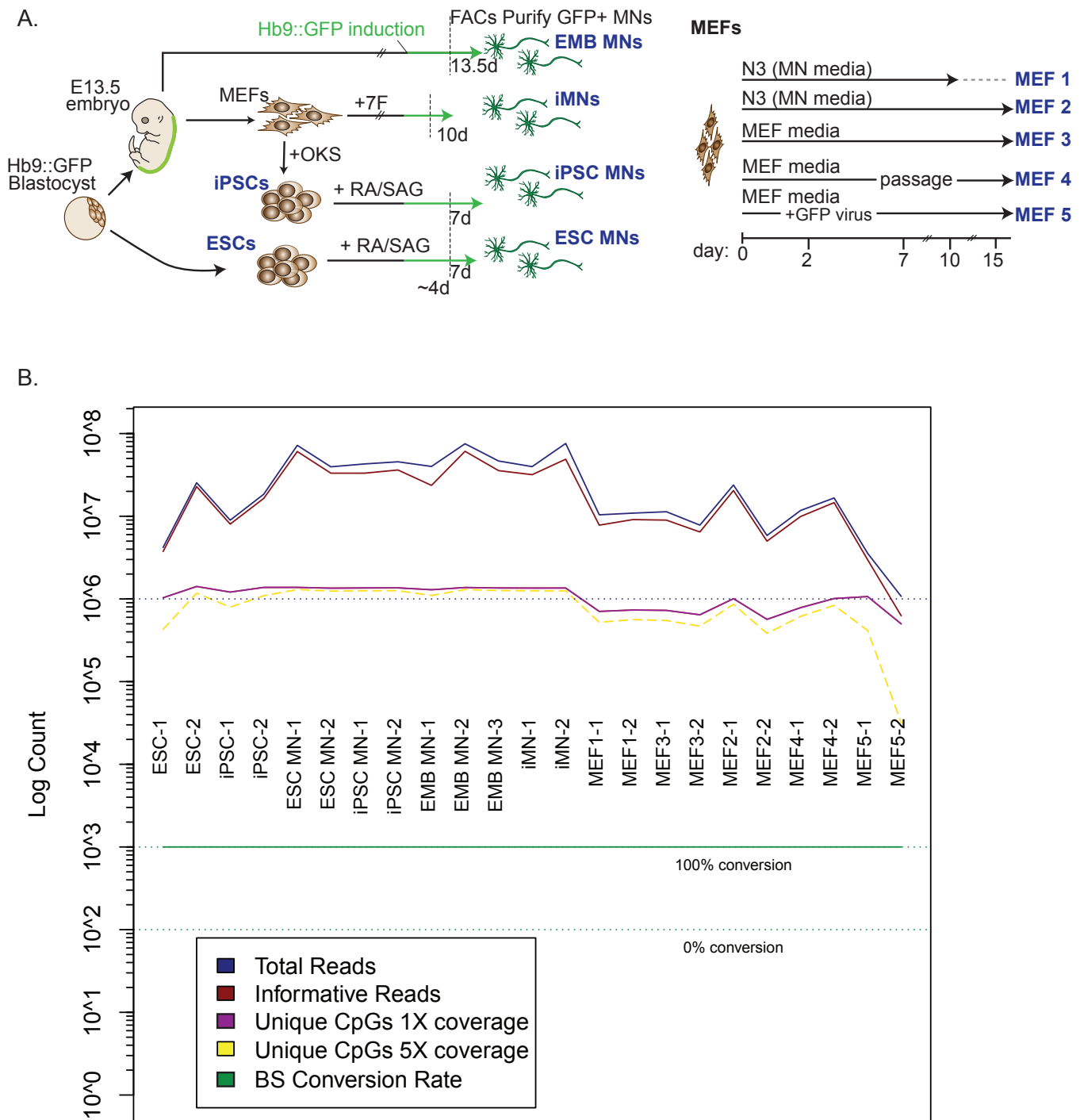
Figure S8. Transcriptomic analysis of iMN factor target genes identified using the Ingenuity Pathway Analysis network builder. (A, C-G) Gene expression analysis (from RNA-seq data) of (A) Isl1, (C) Ngn2, (D) Lhx3, (E) Hb9, (F) Ascl1, (G) Brn2 target genes. B) Number of Isl1 target genes differentially expressed in the EMB MNs and *in vitro* MNs in comparison to MEF1 (DESeq2) and the directionality of the gene expression change relative to MEF1. DESeq2 was used to restrict the analyses to genes that could be reliably compared between samples. Two biological replicates were analyzed per cell type.

Figure S9. Overview of Methylation Analysis. A. Schematic of samples analyzed by reduced representation bisulfite sequencing (RRBS). B. Quality control metrics for all samples. At least 1M reads were obtained for each sample. Informative reads have at least 1 CpG covered within the sequencing read. 100% bisulfate (BS) conversion was acheived for all samples. Two biological replicates were analyzed per sample type.
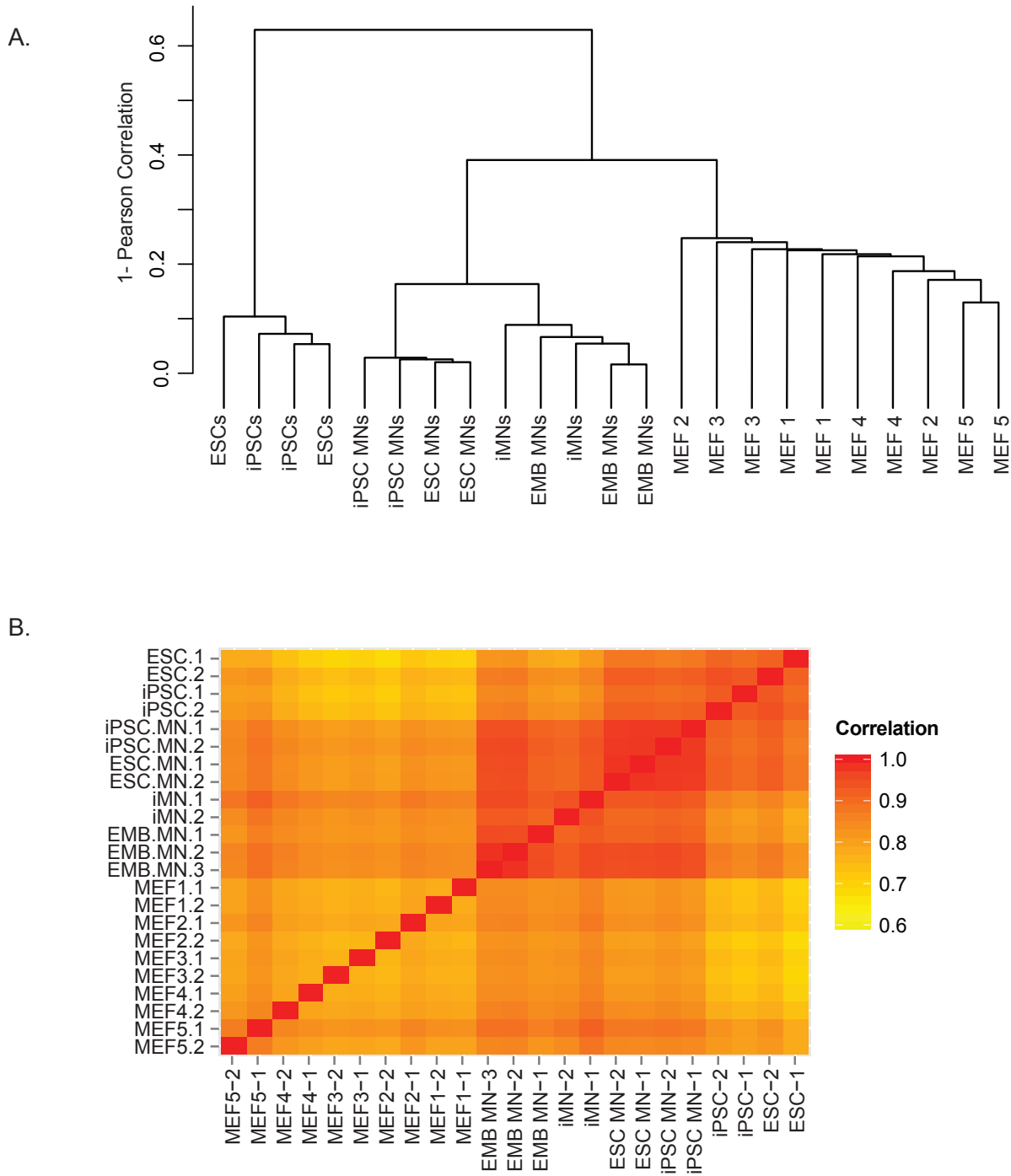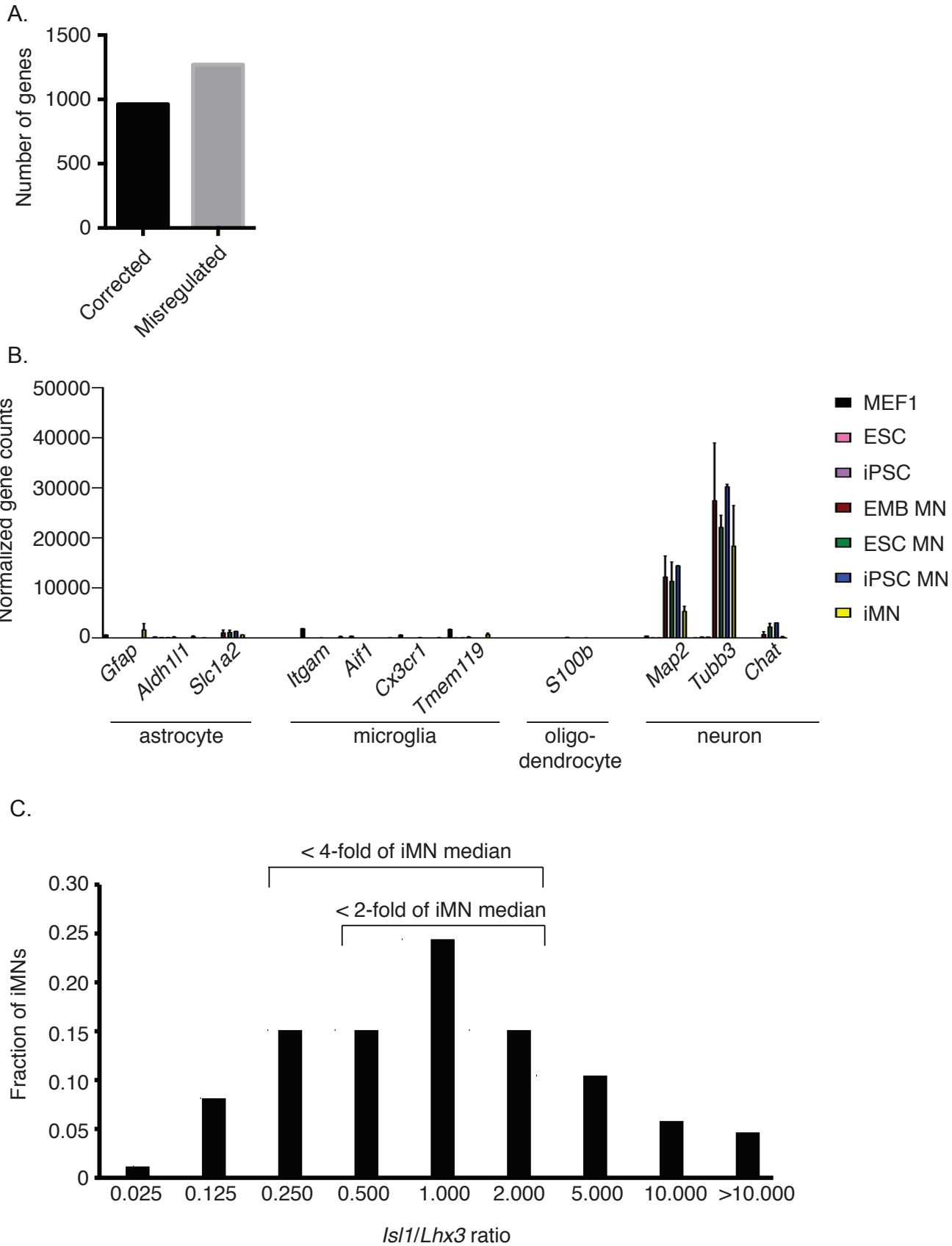
A.

B.

Figure S10. Global overview of methylation analysis. A). Dendogram of biological replicates, constructed by pairwise complete linkage of all the methylation values at all CpGs. B. Pearson correlation of biological replicates based on methylation of all CpGs. Two biological replicates were analyzed per sample type in A and B (all biological replicates are shown for B).
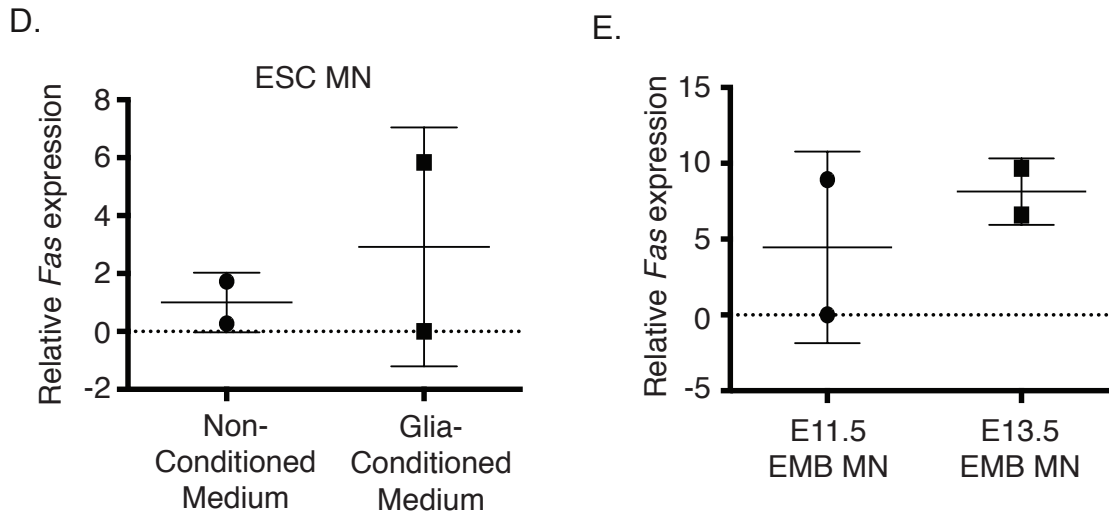
A.

B.

C.

D.

ESC MN

E.

Figure S11. Effects of glial co-culture, reprogramming factor stoichiometry, and maturation on in vitro-derived motor neurons.
A) Number of genes that are correctly expressed in iMNs that change from incorrect to correct, or change from correct to incorrect expression in ESC MNs if they are differentiated in glia-conditioned medium. B) mRNA expression levels of glia-specific genes in different samples as a quantitative measure of glial contamination in iMN samples. Mean +/- s.d. C) Fraction of iMNs expressing the specified *Isl1/Lhx3* ratios. A total of 62 Hb9+ iMNs were analyzed from two independent lineage conversions. D) Relative mRNA expression of the Fas receptor in ESC MNs differentiated with or without glia-conditioned medium. Mean of two biological replicates +/- s.d. E) Relative mRNA expression of the Fas receptor in E11.5 or E13.5 EMB MNs. Mean of two biological replicates +/- s.d. All mRNA levels were determined by RNA-seq analysis. Two biological replicates were analyzed per sample type in all experiments.
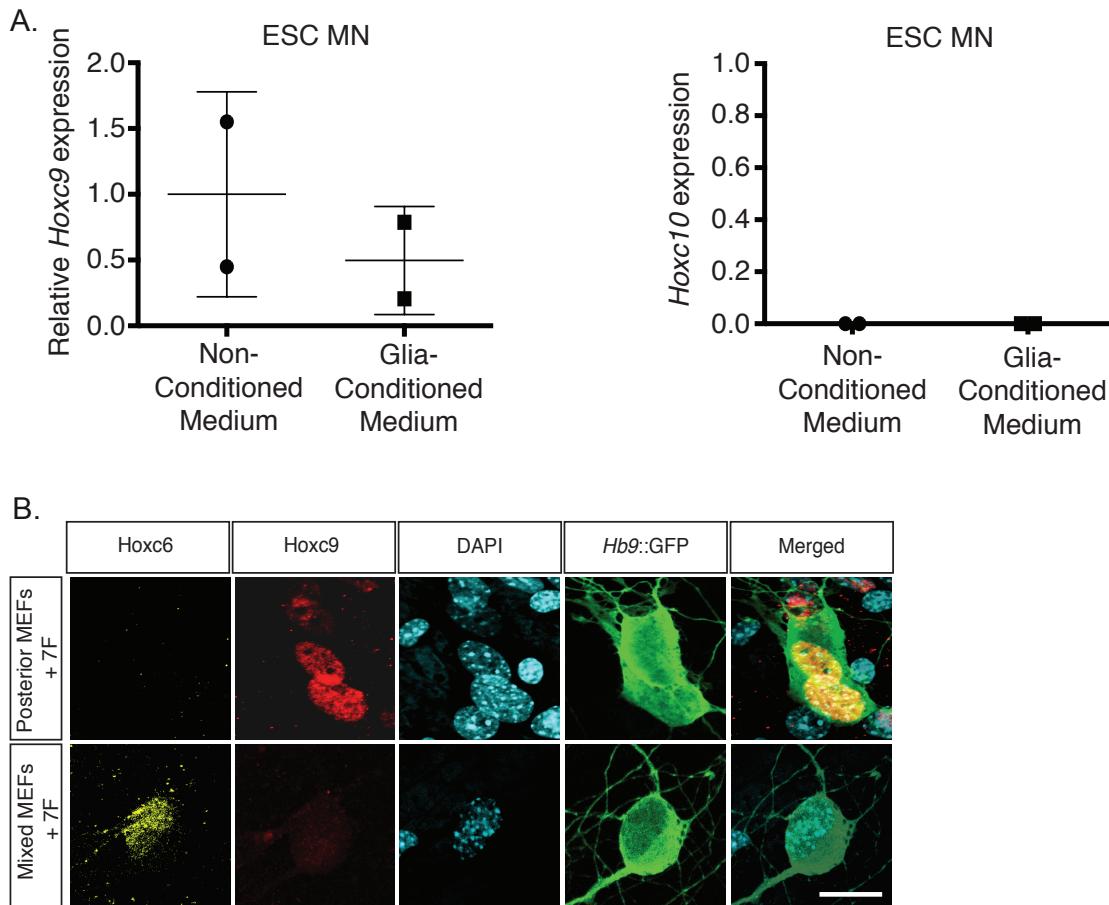
A.

B.

Figure S12. *Hox* gene expression in *in vitro*-derived motor neurons.
A) *Hoxc9* and *Hoxc10* mRNA expression in ESC MNs differentiated in the presence or absence of glia-conditioned medium. Mean +/- s.d. of two biological replicates. mRNA expression was determined by RNA-Seq analysis. B) Immunocytochemical analysis of Hox gene levels in iMNs generated from either posterior MEFs or mixed MEFs. iMN conversions were fixed at day 14 post-transduction and immunostained. Images are representative of three biological replicates. Scale bar - 10 μm.

Table S1. Names and descriptions of the samples used to perform RNA-seq and reduced representation bisulfite sequencing analysis. The genetic background, cell type, biological replicate number, and culturing or FACS-sorting information is included for each sample.

Click here to Download Tables S1

Table S2. DESeq2 normalized gene counts for all samples simultaneously.

Click here to Download Tables S2

Study	Target Cell Type	Starting Cell Type	Reference Cell Type	Maturity of Reference Cell	Conversion Method	Species
Ieda et al 2010, GSE22292	4-week Cardiomyocyte	(Neonatal) Cardiac Fibroblast	Cardiomyocyte	Neonatal	Lineage Conversion	Mus musculus
Huang et al 2010, GSE23635	Hepatocyte	P19Arf-null (Adult) Tail Tip Fibroblast	Hepatocyte	Adult, cultured *in vitro* 6 days	Lineage Conversion	Mus musculus
Zhou et al 2008, GSE12025	Pdx1-GFP+ Beta Cell	(Adult) Non-Islet Pancreatic Cell	Pdx1-GFP+ Beta Cell	Adult	Lineage Conversion (*in vivo*)	Mus musculus
Marro et al 2011, GSE30102	Tau-GFP+ Neuron	(Embryonic) Fibroblast	Tau-GFP+ Cortical Neuron	Neonatal	Lineage Conversion	Mus musculus
Han et al 2012, GSE30500	5-factor Neural Stem Cell	(Embryonic) Fibroblast	Neural Stem Cell	Embryonic	Lineage Conversion	Mus musculus
Thier et al 2012, GSE36484	Neural Stem Cell	(Embryonic) Fibroblast	Neural Stem Cell	Embryonic	Lineage Conversion	Mus musculus
Zhao et al 2009, GSE16925	Induced Pluripotent Stem Cell	(Embryonic) Fibroblast	Embryonic Stem Cell	Embryonic	Lineage Conversion	Mus musculus
Bock et al 2011, GSE25970	Induced Pluripotent Stem Cell	(Adult) Fibroblast	Embryonic Stem Cell	Embryonic	Lineage Conversion	Homo sapiens

Table S3. Identity and maturity level of target, starting, and reference cell types analyzed in Figure 3D. GEO accession numbers for each dataset are included in the first column. Two (neurons, 5-factor neural stem cells), three (cardiomyocytes, beta cells, neural stem cells, mouse iPSCs, human iPSCs), or four (hepatocytes) biological replicates were examined per sample type.