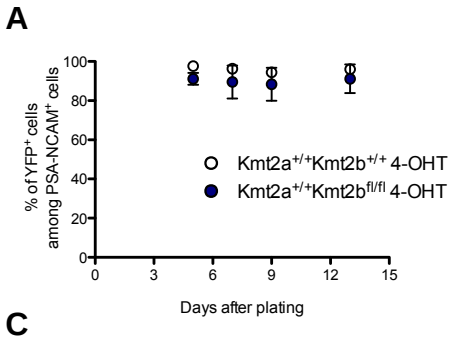
Cell Reports, Volume *25*

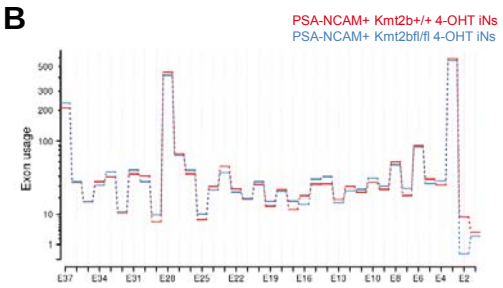
Supplemental Information

KMT2B Is Selectively Required for Neuronal Transdifferentiation, and Its Loss Exposes Dystonia Candidate Genes

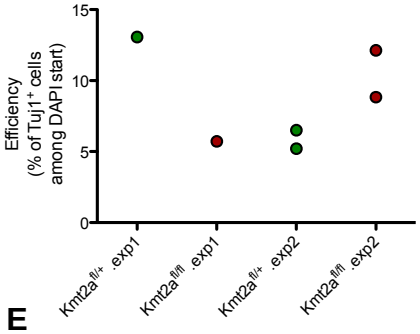
Giulia Barbagiovanni, Pierre-Luc Germain, Michael Zech, Sina Atashpaz, Pietro Lo Riso, Agnieszka D'Antonio-Chronowska, Erika Tenderini, Massimiliano Caiazzo, Sylvia Boesch, Robert Jech, Bernhard Haslinger, Vania Broccoli, Adrian Francis Stewart, Juliane Winkelmann, and Giuseppe Testa

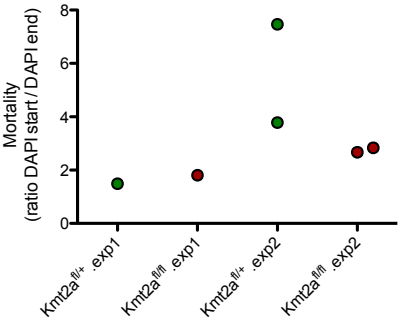
***Kmt2b* cKO (plated 7 days after 4-OHT)**



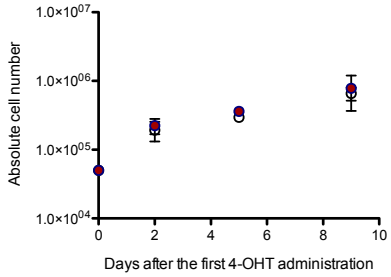


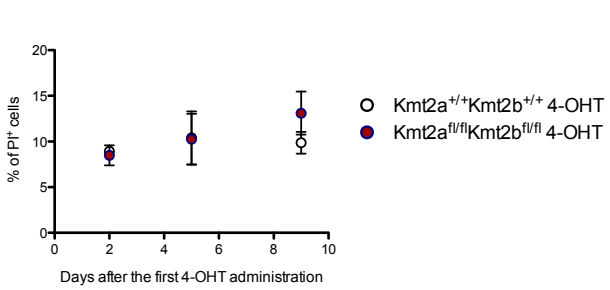
***Kmt2a* cKO: ScanR (plated 2 days after 4-OHT)**



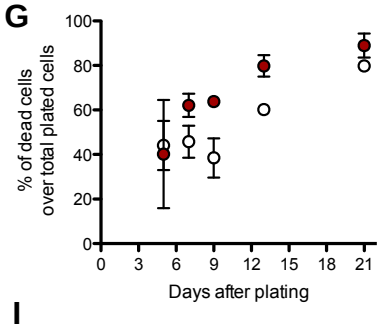


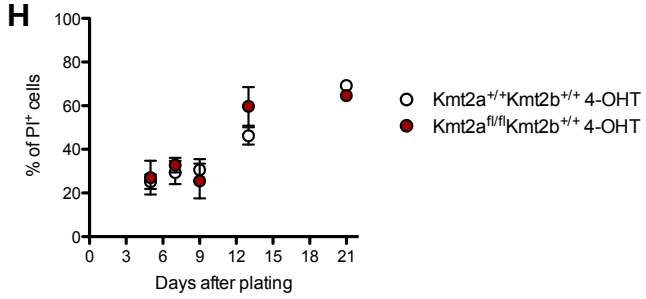
***Kmt2a Kmt2b* cKO before transdifferentiation**



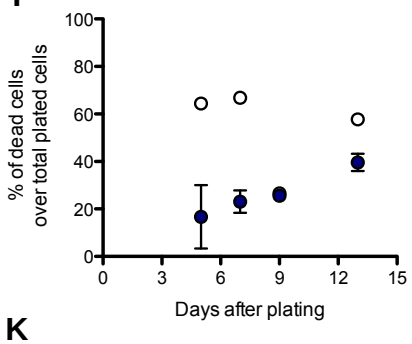


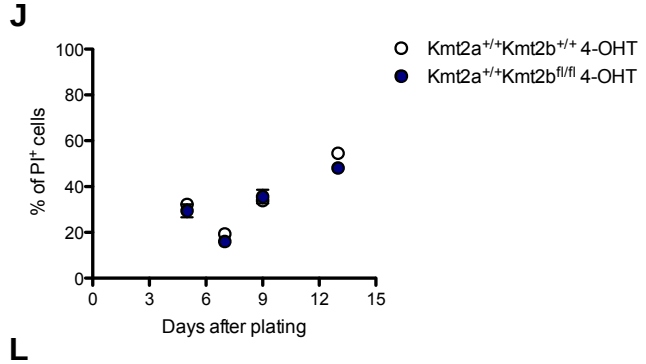
***Kmt2a* cKO: FACS (plated 7 days after 4-OHT)**



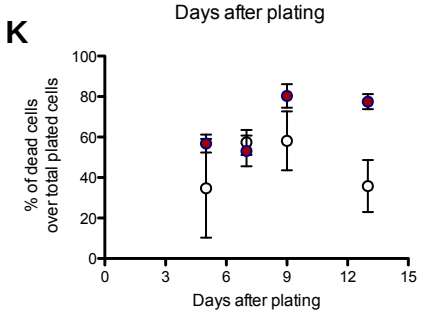


***Kmt2b* cKO FACS (plated 7 days after 4-OHT)**





***Kmt2a Kmt2b* cKO: FACS (plated 7 days after 4-OHT)**



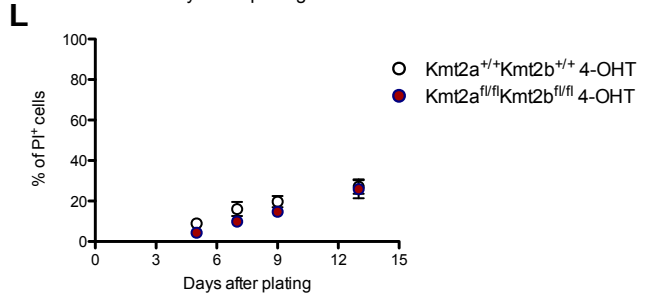


Figure S1. The role of KMT2A and KMT2B during transdifferentiation, related to Figure 2

A, Percentage of YFP$^+$ cells among PSA-NCAM$^+$ cells 5, 7, 9 and 13 days after plating in *Kmt2a*$^{+/+}$*Kmt2b*$^{+/+}$ (n=2) and *Kmt2a*$^{+/+}$*Kmt2b*$^{fl/fl}$ (n=2) 4-OHT transdifferentiating MEFs. **B,** The graph shows *Kmt2b* exon usage in *Kmt2b*$^{fl/fl}$ (in blue) and *Kmt2b*$^{+/+}$ (in red) 4-OHT treated 13-days sorted iNs obtained through RNA-seq analysis. **C,** Efficiency of transdifferentiation in ScanR experiments calculated relating the number of Tuj1$^+$ at 13 days, of the chosen concentration, to the number of DAPI 3 days after plating. **D,** Cell mortality in ScanR experiments calculated as the ratio of DAPI 3 days (DAPI start) and 13 days (DAPI end) after plating. In **C-D**, 2 independent experiments (exp1 and exp2) were performed and cells were plated 2 days after the last 4-OHT administration. **E,** Percentage of dead cells was calculated over the initial number of plated MEFs 2, 5 and 9 days after the first 4-OHT administration. **F**, Mortality rate as percentage of PI$^+$ cells 2, 5 and 9 days after the first 4-OHT administration. **E-F** *Kmt2a*$^{+/+}$ *Kmt2b*$^{+/+}$ n=5; *Kmt2a*$^{fl/fl}$ *Kmt2b*$^{fl/fl}$ n=3. **G, I, K,** Percentage of dead cells was calculated over the initial number of plated MEFs. (G) *Kmt2a*$^{fl/fl}$ n=3; *Kmt2a*$^{+/+}$ n=3 at all time points, but day 21: n=2. (I) *Kmt2a*$^{+/+}$ *Kmt2b*$^{+/+}$ n=1; *Kmt2a*$^{+/+}$ *Kmt2b*$^{fl/fl}$ n=2 despite day 9 n=1. (K) *Kmt2a*$^{+/+}$ *Kmt2b*$^{+/+}$ n=3; *Kmt2a*$^{fl/fl}$ *Kmt2b*$^{fl/fl}$ n=3. **H, J, L,** Mortality rate as percentage of PI$^+$ cells 5, 7, 9, 13 (and 21 for (H)) days after plating, assayed with FACS analysis. (H) *Kmt2a*$^{fl/fl}$ n=3 at all time points, but day 21 in which n=2; *Kmt2a*$^{+/+}$ n=3 at all time points, but day 21 in which n=1. (J) *Kmt2a*$^{+/+}$ *Kmt2b*$^{+/+}$ n=1; *Kmt2a*$^{+/+}$ *Kmt2b*$^{fl/fl}$ n=2. (L) *Kmt2a*$^{+/+}$ *Kmt2b*$^{+/+}$ n=3; *Kmt2a*$^{fl/fl}$ *Kmt2b*$^{fl/fl}$ n=3. **A-L,** Error bars show means +/- SEM. **A**-**B**, **G**-**L**, MEFs were plated for transdifferentiation 7 days after the end of 4-OHT treatment.

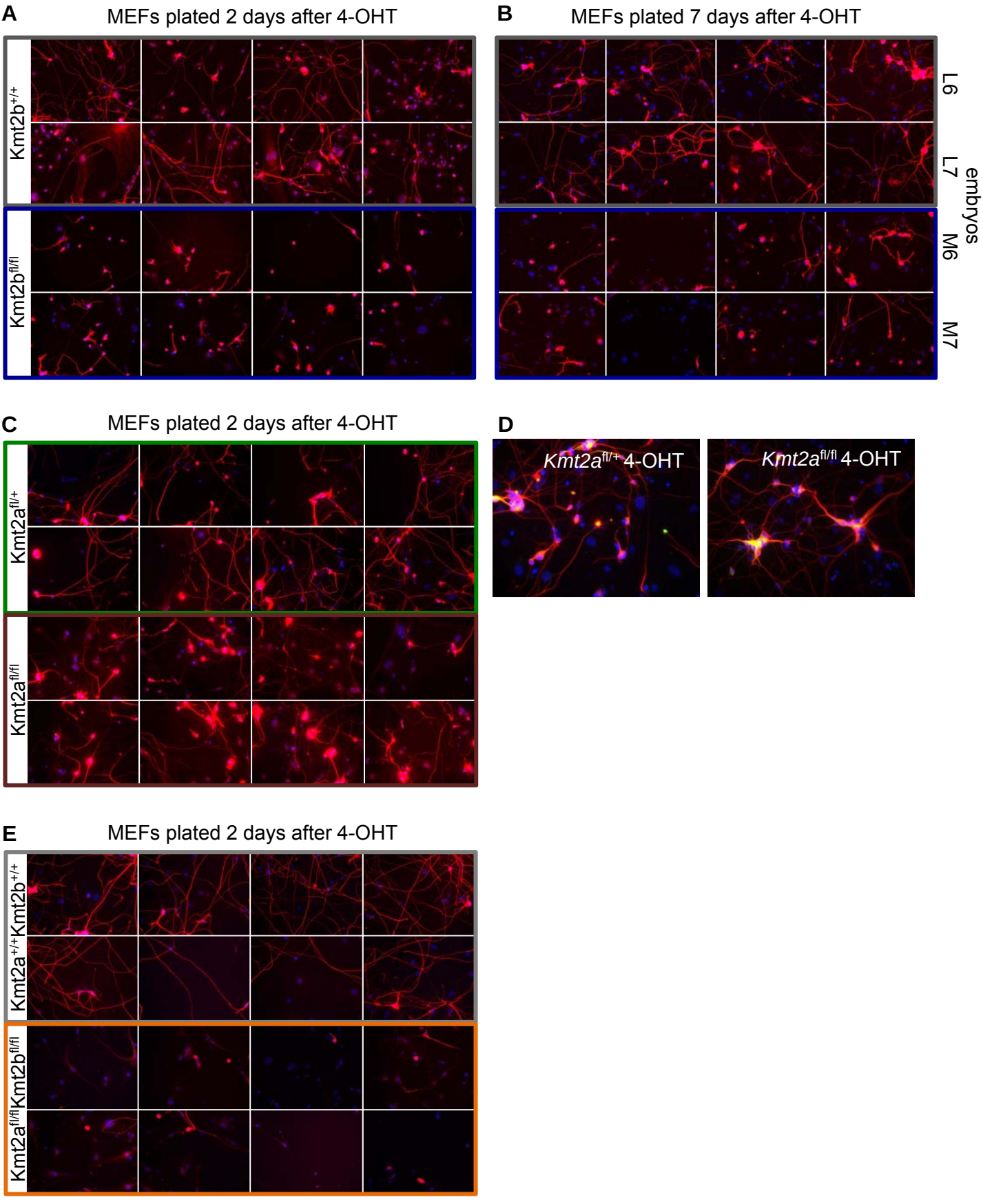

A
MEFs plated 2 days after 4-OHT
Kmt2b+/+
Kmt2bfl/fl
B
MEFs plated 7 days after 4-OHT
L6
L7
M6
M7
embryos
C
MEFs plated 2 days after 4-OHT
Kmt2afl/+
Kmt2afl/fl
D
Kmt2afl/+ 4-OHT
Kmt2afl/fl 4-OHT
E
MEFs plated 2 days after 4-OHT
Kmt2a+/+Kmt2b+/+
Kmt2afl/flKmt2bfl/fl

Figure S2. KMT2B is fundamental for iN maturation while KMT2A is dispensable, related to Figure 2

Representative images of *Kmt2b*$^{fl/fl}$ and *Kmt2b*$^{+/+}$ 4-OHT transdifferentiated MEFs at 13 days, plated either 2 (**A**) or 7 (**B**) days after 4-OHT, of *Kmt2a*$^{fl/fl}$ and *Kmt2a*$^{fl/+}$ (used as control instead of *Kmt2a*$^{+/+}$) 4-OHT transdifferentiated MEFs at 13 days, plated 2 days after 4-OHT (**C**) and of *Kmt2a*$^{+/+}$*Kmt2b*$^{+/+}$ and *Kmt2a*$^{fl/fl}$ *Kmt2b*$^{fl/fl}$ 4-OHT transdifferentiated MEFs at 13 days, plated 2 days after 4-OHT (**E**). The grid of images built for KO and ctrl was the same. Images taken in the same areas of the well are reported for cKOs and control. DAPI is in blue and Tuj1 in red. **D,** Representative images of *Kmt2a*$^{fl/+}$ (used as alternative control to *Kmt2a*$^{+/+}$) (left panel) and $^{fl/fl}$ (right panel) 4-OHT treated iNs (DAPI is in blue, Tuj1 in red, MAP2b in green) at 13 days**.**

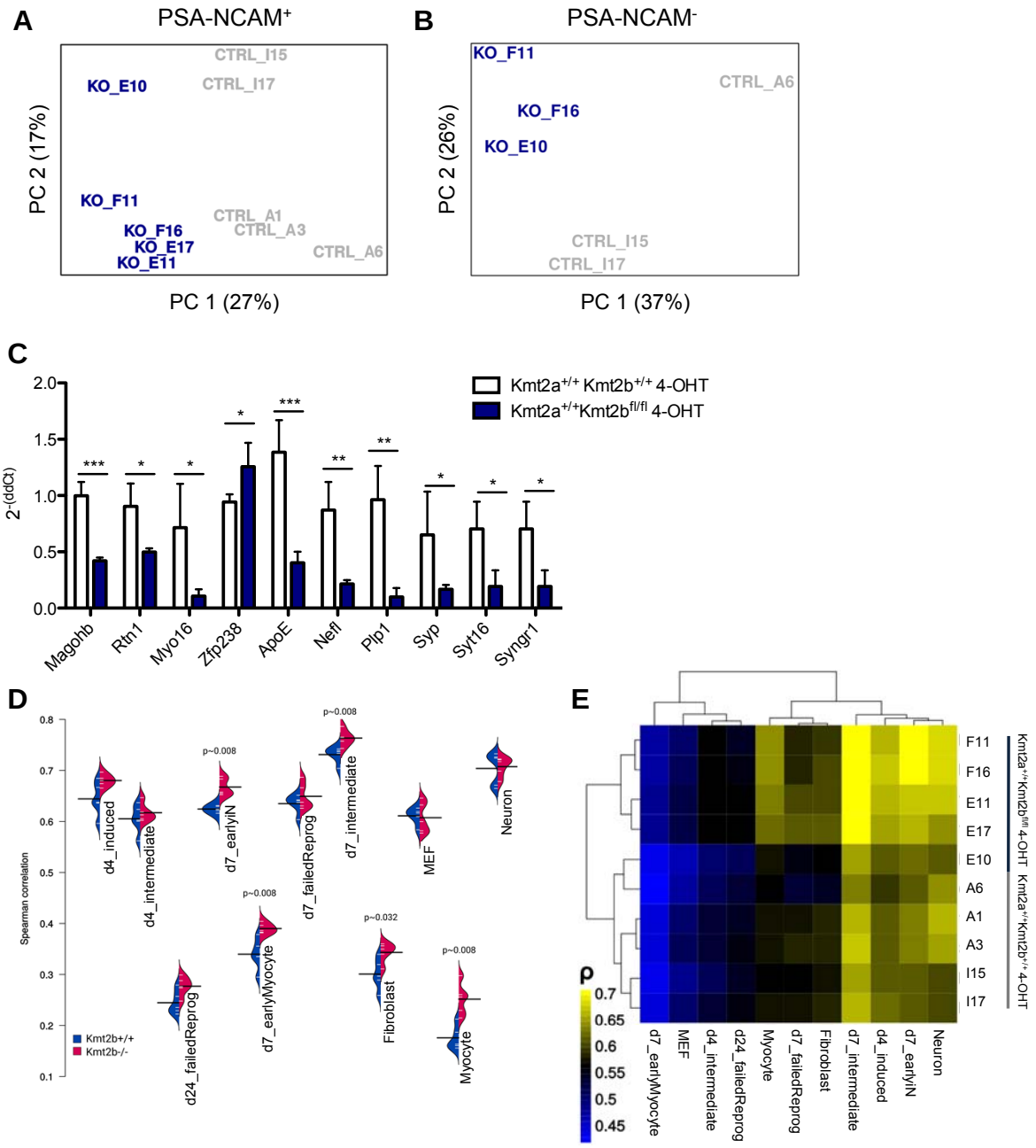


Figure S3. Transcriptomic analysis of PSA-NCAM⁺ cells, related to Figure 3

A, PCA on the log-normalized read counts of *Kmt2a*$^{+/+}$ *Kmt2b*$^{+/+}$ (CTRL_A1, CTRL_A3, CTRL_A6, CTRL_I15, CTRL_I17) and *Kmt2b*$^{-/-}$ (KO_F16, KO_F11, KO_F17, KO_E10, KO_E11) transcriptomes of sorted PSA-NCAM⁺ cells at 13 days. **B,** PCA on the log-normalized read counts of *Kmt2a*$^{+/+}$ *Kmt2b*$^{+/+}$ (CTRL_A6, CTRL_I15, CTRL_I17) and *Kmt2b*$^{-/-}$ (KO_F16, KO_F11, KO_E10) transcriptomes of sorted PSA-NCAM⁻ cells at 13 days. **C**, Validation through RT-qPCR of some of the genes differentially expressed with an impact on the *in vitro* phenotype. Error bars show means and SD. *** p<0.0001; ** p<0.001; *p<0.01. Spearman (**D**) and Pearson (**E**) correlation of the *Kmt2b* cKO and control iNs transcriptomes *vis à vis* the different previously reported fates of single cells undergoing transdifferentiation (Treutlein et al., 2016). A1, A3, A6, I15, I17 and E11, E17, E10, F11, F16 are the names of the embryos from where *Kmt2a*$^{+/+}$*Kmt2b*$^{+/+}$ and *Kmt2a*$^{+/+}$*Kmt2b*$^{fl/fl}$ MEFs were derived, respectively.

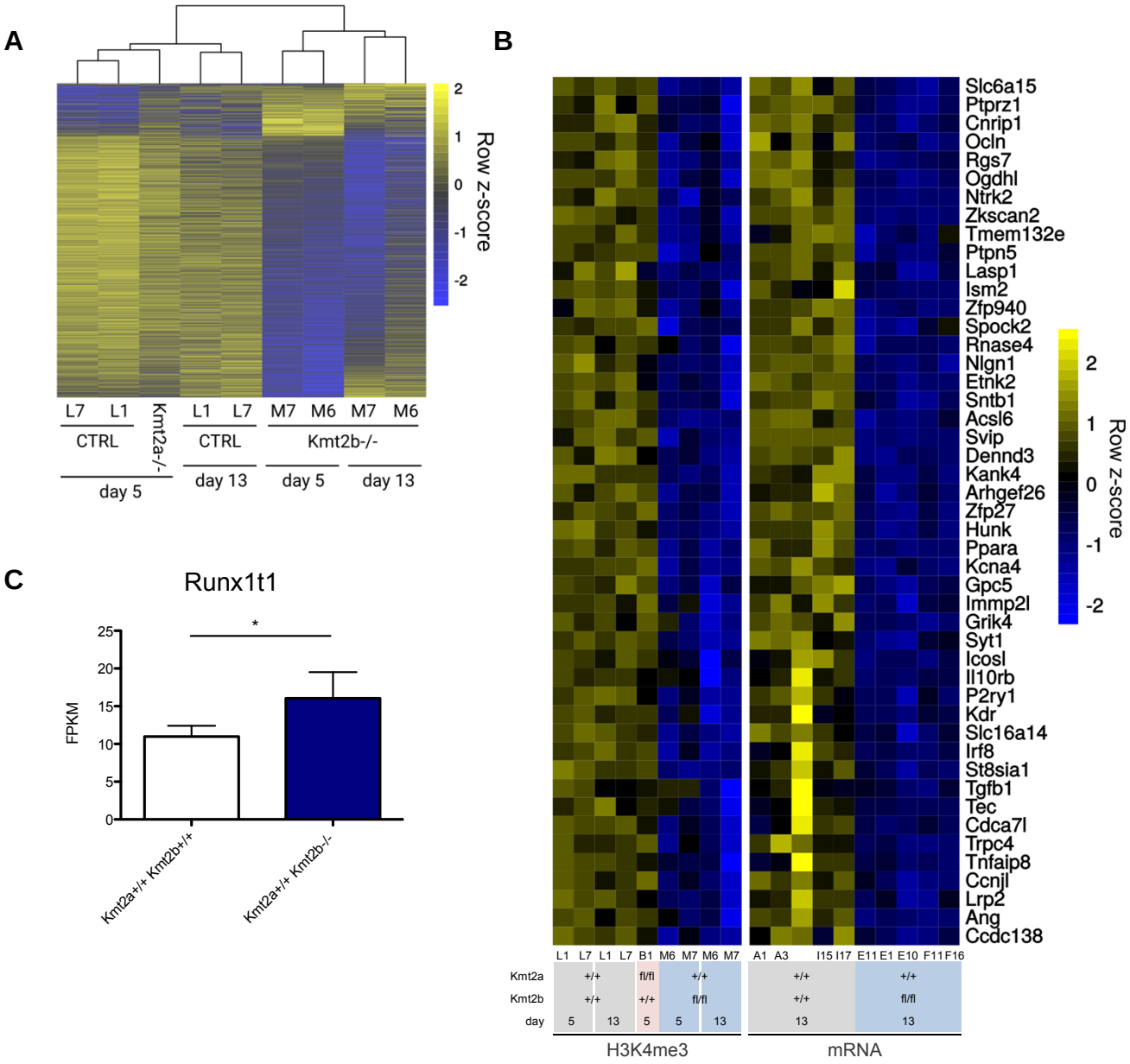


Figure S4. Epigenomic and transcriptomic dissection, related to Figure 6

A, Union of *loci* showing significant changes in H3K4me3 in *Kmt2b* cKO at day 5 and/or 13. **B,** Heatmap of the genes that further lose H3K4 trimethylation in *Kmt2b* cKO samples at 13 days and their expression in RNA-seq at 13 days. **C,** FPKM of *Runx1t1* in our RNA-seq * FDR<0.01**.** Error bars show means and SD. A1, A3, A6, I15, I17, L1, L6, L7 and E11, E17, E10, F11, F16, M6, M7 are the names of the embryos from where *Kmt2a*$^{+/+}$*Kmt2b*$^{+/+}$ and *Kmt2a*$^{+/+}$*Kmt2b*$^{fl/fl}$ MEFs were derived, respectively. B1 is the name of the embryo from where *Kmt2a*$^{fl/fl}$*Kmt2b*$^{+/+}$ MEFs were derived.

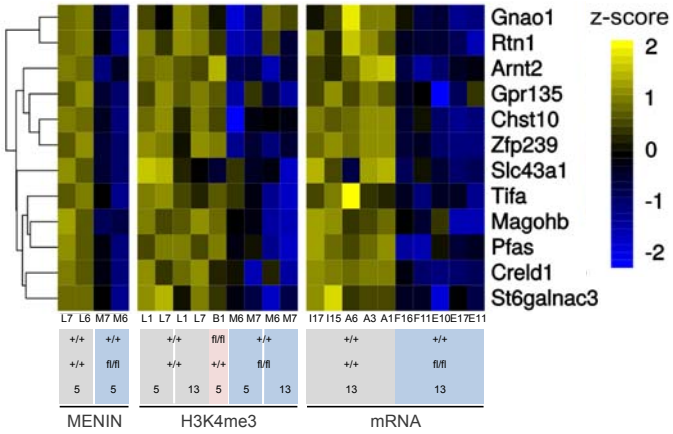


Figure S5. Identification of KMT2B candidate targets through MENIN ChIPseq, related to Figure 6

Heatmap showing the genes both differentially bound by MENIN, differentially expressed (FDR <0.01 and a FC > 0.5) and differentially H3K4me3 in *Kmt2b*$^{-/-}$ with respect to control. A1, A3, A6, I15, I17, L6, L7, L1 and E11, E17, E10, F11, F16, M6, M7 are the names of the embryos from where *Kmt2a*$^{+/+}$*Kmt2b*$^{+/+}$ and *Kmt2a*$^{+/+}$*Kmt2b*$^{fl/fl}$ MEFs were derived, respectively. B1 is the name of the embryo from which *Kmt2a*$^{fl/fl}$ MEFs were derived.

A

Variant calls in 216 *KMT2B*-target genes
across 225 dystonia samples

High-confidence calls
read depth >20x, quality score >30 (Phred scale)

Non-silent variation in canonical transcripts
missense, in-frame indel, loss-of-function (stop-gain/lost, frameshift, splice-site) variants in NCBI RefSeq variant 1 transcripts

Frequency filter

Autosomal-dominant disease model[a] variants not represented in ExAC and in-house control databases[c]		**Autosomal-recessive disease model**[b] variants with minor allele frequency <0.001 in ExAC and in-house control databases[c]	
Predicted damaging missense and in-frame indel variants (CADD score ≥ 20)[d]	**Loss-of-function variants**	**Predicted damaging missense and in-frame indel variants** (CADD score ≥ 10)[d]	**Loss-of-function variants**

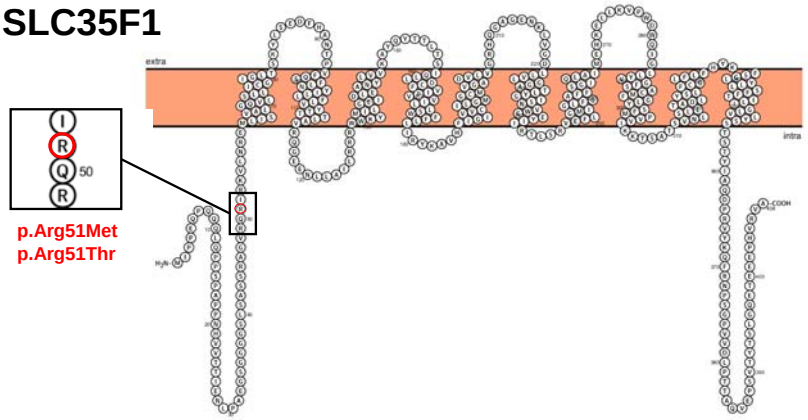
[a]Single heterozygous variant per gene per sample

[b]Homozygous variant or two heterozygous variants per gene per sample

[c]in-house database consisting of roughly 10,000 non-dystonia control exomes

[d]Scores ≥20 predict that variants are among the 1% most deleterious changes that can affect the human genome, scores ≥10 predict that variants are among the 10% most deleterious

B **SLC35F1**



C **SLC40A1**

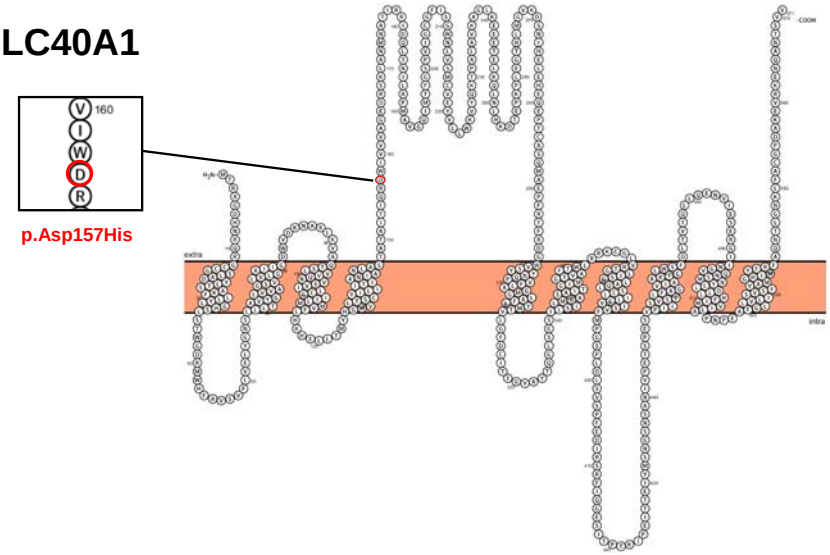


Figure S6. Identification of new putative dystonia-causing variants and genes, related to Table 1

A, Experimental pipeline of variant calling and filtering. **B-C,** Schematics of SLC35F1 and SLC40A1 mutations generated with Protter: interactive protein feature visualization and integration with experimental proteomic data **(Omasits et al., 2014)**.

	Kmt2b fl/fl Vs. +/+		*Kmt2a* fl/fl Vs. fl/+	*Kmt2b* fl/fl *Kmt2a* fl/fl Vs. +/+	
Days after 4-OHT	2	7	2	2	7
N of indipendent embryos per experiment per genotype	1	2	1	1	1
N of indipendent experiments	2	1	2	1	1
N of fields per embryo	100	100	100	100	100
Increased mortality in KO	X	X	1 out of 2	✓	✓
Reduced transdifferentiation efficiency in KO	X	✓	1 out of 2	✓	✓
Reduced neurite length in KO	✓	✓	X	✓	✓

Table S1. Summary of all experiments performed at ScanR, related to Figure 2

	Number of KMT2B targets	pLI ≥ 0.9	Missense Z score ≥ 2	Both
iNs	216	45 (21%)	54 (25%)	26 (12%)
Brain Neurons	127	5 (4%)	11 (9%)	6 (5%)
ESCs	120	9 (7%)	26 (21%)	12 (10%)

Table S4. Summary of the Tables S1 and S2, related to Table 1. χ^2 test: $p < 0{,}005$ iN vs ESC, pLI: 0,00024; iN vs ESC, misZ: 0,58; iN vs ESC, both 0.7.

No. (sex female/male)	225 (133/92)
Age at sampling[a]	50.4 ± 14.6 (18-88)
Age at onset[a]	29.4 ± 15.1 (1-69)
Early symptom onset (<30 years) (%)	48.0
Family history positive (%)	38.7
Dystonia distribution (%)	
Focal	61.8
Segmental	20.0
Generalized	18.2
Dystonia clinical classification[b] (%)	
Isolated	87.9
Combined	5.6
Complex	6.5

Table S5. Clinical characteristics of 225 dystonia-affected probands sequenced by exome, related to Table 1
[a]Mean age in years ± standard deviation (range); [b]according to Albanese et al., 2013 and Marras et al., 2016

N=	Reads[a]	Mapped reads[a]	Percent[a]	Mapped sequence (Gb)[a]	Target bases > 20x[a]	Average coverage[a]
225	129,920,660 ± 18,380,126 (80,589,549-185,236,854)	129,366,209 ± 18,479,892 (79,579,741-184,924,819)	99.55 ± 0.36 (98.74-99.95)	13.14 ± 2.16 (8.04-27.92)	98.49 ± 0.49 (96.47-99.31)	153.54 ± 22.04 (98.46-264.60)

[a]Mean ± standard deviation (range)

Table S6. Summary of statistics for 225 dystonia exomes, related to Table 1

KMT2B	rabbit	Stewart Lab		1:1000	WB
VINCULIN	mouse	Sigma Aldrich	V9131	1:400	WB
TUJ1	mouse	Covance	MMS-435P	1:400	IF
TUJ1	rabbit	Covance	MRB-435P	1:400	IF
VIMENTIN	mouse	Abcam	ab20346	1:200	IF
MAP2B	mouse	BD Biosciences	610460	1:500	IF
H3K4me3	rabbit	Active motif	39159	1ug Ab each 100 ug Cromatin	ChIP
H3K27me3	rabbit	Cell signaling	9733BF	1ug Ab each 25 ug Cromatin	ChIP
MENIN	rabbit	Abcam	ab31902	1ug Ab each 100 ug Cromatin	ChIP
PSA-NCAM	mouse	Miltenyi	130-093-273		FACS

Table S7. List of antibodies used in the study, related to Experimental Procedures

	Forward	**Reverse**
ApoE	CAATTGCGAAGATGAAGGCTC	TAATCCCAGAAGCGGTTCAG
Magohb	CGGGCATAAGGGCAAGTTT	AATTACTGTTGTTGGCATATGTAAGCTT
Myo16	TGTGTCCAGTTGCCTCATAC	CAGGAGAAAATCAGAAAGGTGC
Nefl	GGGTATGAACGAAGCTCTGG	TCTCAGCTCATTCTCCAGTTTG
Plp1	TCAGCCGCAAAACAGACTAG	CACTCCAAAGAAACACAATCCAG
Rtn1	GTGTGGAGCAACTGGAAAAG	CTCCCGAACACAATCCCAG
Syngr1	GGACAACCCTCTGAACGAAG	CCAATCTGGTACCTCTGGAAG
Syp	TGCAGTGGGTCTTTGCCATCTT	ACTTCGATGTTGAGGGCACTCT
Syt16	TCCATACGTCAAAGTCTTCCTG	TGTCTTGCCACCTAATTCCG
Zfp238	TCTCCACTTTGCATCTGTCTC	TGCTGTGGTCTGGAAACTC

Table S8. Sequences of primers used in RT-qPCR, related to Experimental Procedures

Supplemental Experimental Procedures

MEFs derivation

MEFs were derived from embryos at E13.5. After removal of both the head and the liver, cells were enzymatically and mechanically dissociated through the incubation with trypsin at 37° C, 3% O_2, 5% CO_2 for 30 minutes, pipetting every 10 minutes. Dissociated MEFs were seeded in the medium published in the original paper (Vierbuchen et al., 2010) and cultured at 37 °C, 3% O_2 and 5% CO_2.

Control and cKO MEFs at P1-P3 were treated for 5 days with 4-Hydroxytamoxifen (4-OHT), added to the MEF medium in a final concentration of 1000 ng/ml, and then left either 2 or 7 days in medium not supplemented with 4-OHT.

Transdifferentiation was performed according to the protocol published in the original paper (Vierbuchen et al., 2010).

Mice/MEFs were genotyped using the following primers: *YFP*: forward-5'-TGCAGTGCTTCGCCCGCTACC-3' and reverse-5'-CCGTCGCCGATGGGGGTGTTC-3' (Mihailovich et al., 2015); *Cre*: forward-5'-GCCTGCATTACCGGTCGATGCAACGA-3' and reverse-5'-GTGGCAGATGGCGCGGCAAACCATT-3' (Basch et al., 2011); *Kmt2a* forward-5'-GAGGTAAGGAGAGTTTTTGCT-3' and reverse-5'-GGTAACACCTTAAAATGCCCT-3' (Kranz et al., 2010); *Kmt2b*: forward-5'-CGGAGGAAGAGAGCAGTGACG-3' and reverse-5'-GGACAGGAGTCACATCTGCTAGG-3' (Glaser et al., 2006).

Assessment of exon 2 deletion of both *Kmt2a* and *Kmt2b*

To determine the entity of *Kmt2* exon 2 deletion, TaqMan assays were performed extracting DNA, through DNeasy Blood and Tissue kit (Qiagen), the day MEFs were plated for transdifferentiation. In particular, Real Time PCR was performed on 7900HT Fast Real-Time PCR system (Applied Biosystems). Primers were the following:

MLL2EX2: forward-5'-GGTCCCCTAAATCAGGAGTTTCAG-3', reverse-5'-GACCG AAGCGCAGAGC-3', reporter sequence-5'-AAGATGTGGCCCCCAGTTC-3'.

MLL1EX2: forward-5'-GCAGTTCTTAGGTTTTGGCTCAGA-3', reverse-5'-GCTTTAT TGGCCATACCTGAAGGA-3', reporter sequence-5'-CTTCGCACTCTGACTTCTTCA -3'.

Western blot for KMT2B

Proteins were extracted in RIPA buffer (10 mM Tris-HCl pH 8, 1% Triton X-100, 0.1% SDS, 0.1% Sodium Deoxycholate, 140 mM NaCl, 1 mM EDTA plus the protease inhibitor cocktail (PIC) (Sigma Aldrich) and Phenylmethanesulfonyl fluoride (PMSF) (0.5 mM) (Sigma Aldrich) from at least 8 x 10^6 MEFs, the day were plated for transdifferentiation.

At least 70 µg of protein extract was loaded into the NuPAGE Novex 3-8% Tris-Acetate Proteins Gels (ThermoFisher Scientific) and run in the NuPAGE Tris-Acetate SDS Running buffer (20X) (ThermoFisher Scientific). Wet transfer was performed at 4 °C, 30V with the Immobilon-P, 0.45 µm, PVDF membrane (Merck Millipore) overnight. Antibodies used are listed in Table S7.

Vector production

BAM and rtTA vectors were generated through calcium phosphate transfection of human embryonic kidney 293T (HEK293T) cells and ultracentrifugation. The following transfer plasmids, deposited by the Wernig Lab in Addgene, were used (Tet-O-FUW-Brn2, Tet-O-FUW-Ascl1, Tet-O-FUW-Myt1l and UbC-rtTA). BAM vectors were produced using a third generation system (envelope plasmid: pMD2-VSV-G; packaging plasmids: pMDLg/pRRE and pRSV-REV).

Image acquisition and analysis: efficiency and cell mortality calculation

Cells were fixed with 4% paraformaldehyde (PFA), 20 minutes on ice and permeabilized with PBS, 10% FBS, 0.1% Triton X-100, 30 minutes at room temperature (RT). Cells were incubated with primary antibodies (listed in Table S7), overnight at 4°C, and then with secondary antibodies 1 h at RT. Primary and secondary antibodies were diluted in PBS, 10% FBS. DAPI (SIGMA D9542) was added to the secondary antibodies mix. A grid of 100 images per embryo in duplicate was acquired. Through ScanR we covered an area of ~ 14 mm^2 per cell density, in a grid of 100 images equidistant from each other, maintaining the same setting across conditions and time points and we acquired at least 1000 cells at day 3 and 500 cells at the following time point. 3 days after plating the number of cells per image was manually evaluated as number of DAPI in each cell density condition. At 13 days only conditions that at day 3 were in the same number were compared. Cells positive for Tuj1 were classified as iNs and manually quantified.

The number of DAPI at day 3 and day 13, of the chosen cell density, were summed and called DAPI start and DAPI end, respectively. To calculate cell mortality the ratio between DAPI start and DAPI end was calculated, assuming that the cells distribute almost equally in the two duplicate wells (*i.e.,* day 3 and day 13) of each condition. Then, to calculate efficiency, the number of $Tuj1^+$ cells at day 13 was related to the number of DAPI start:

$$\frac{number\ of\ Tuj1^{+}_{day13}}{number\ of\ DAPI_{day3}}\%$$

NeuriteTracer was used to calculate neurite length (Pool et al., 2008). Because this calculation depends on the number of neurons the plugin-retrieved value was divided for the number of $Tuj1^+$ cells. Unpaired t test was performed as statistical analysis.

Cytofluorimetric analysis: efficiency and cell mortality calculation

Before cytofluorimentric analysis cells were counted through the Automated Cell Countess (Thermo Fisher Scientific) 5, 7, 9, 13 and 21 days after they were plated

for transdifferentiation. Cell mortality was reported as the percentage of dead cells with respect to the number of plated cells:

$$\frac{N\ of\ plated\ cells - N\ of\ alive\ cells_{day13}}{N\ of\ plated\ cells}\%$$

Knowing the percentage of PSA-NCAM$^+$ cells and the absolute number of cells at day 13, the absolute number of PSA-NCAM$^+$ iNs was calculated. This value was related to the number of plated cells to calculate the efficiency of transdifferentiation:

$$\frac{number\ of\ PSA - NCAM^+_{day13}}{number\ of\ plated\ cells_{day0}}\%$$

FACS sorting

FACS sorting was executed on 4-OHT treated MEFs left in culture 7 days before transdifferentiation. 10^7 MEFs were plated on matrigel coated dishes both for the RNA-seq and the ChIP-seq on iNs at 13 days. Cells were sorted for PSA-NCAM positivity at MoFlo Astrios (Beckman Coulter) with the Software Summit v6.2.

RNA-seq analysis

RNA-seq reads were aligned with TopHat v2.0.10 (Trapnell et al., 2012), first to mm10 Refseq transcriptome, and genes without a perfect alignment were realigned to genome (*--read-realign-edit-dist=1*), tolerating an edit distance of 3. Quantification was performed on the Refseq transcriptome using Cuffquant v2.2.1, with multi-read correction (-u) (Trapnell et al., 2012). This pipeline was selected because it was one of the highest-performing quantification methods in our recent benchmark (Germain et al., 2016).

For the differential expression analysis, we first excluded genes which had an average read count across samples below 50. We used edgeR v.3.12.1 (which outperformed other methods according to our benchmark) with TMM normalization and the exact test on the estimated fragment counts. We considered differentially expressed with high confidence genes with a FDR below 0.01 and an absolute

log2(foldchange) greater than log2(1.5). Differential exon usage analysis was performed with DEXSeq.

Gene Ontology enrichment analyses

Gene Ontology enrichment analyses were performed using the goseq v.1.22.0 R package for RNAseq results in order to correct for RNA-seq transcript length bias (Young et al., 2010), or with the topGO package for ChIPseq results, in both cases using Fisher's test and excluding genes without annotation. Categories with at least 10 genes and maximum 1500 genes were considered. When the number of enriched categories was large, we focused on the most specific categories by removing any category with enriched children categories. Quilts were generated using treemap R package.

RNA-seq validation

The RNA was retrotranscribed through the SuperScript VILO cDNA Synthesis Kit (Invitrogen, 11754-050).

Real Time Quantitative PCR (RT-qPCR) was performed on 7500 Fast Real-Time PCR system (Applied Biosystems) using Sybr green (Applied Biosystems) as detecting reagent and the standard amplification protocol. Primers used are reported in Table S8. Primers for *Magohb* are described in the paper of Ladopoulus et al (Ladopoulos et al., 2013), while the ones for *Syp* in (Simon et al., 2016).

ChIP-seq

At the desired time point, cells were fixed with PBS 1% formaldehyde. Fixation was quenched with 0,125M glycine. Cells were collected in SDS Buffer (100mM NaCl, 50mM Tris-HCl pH 8.1, 5mM EDTA pH 8, 0.5% SDS), centrifuged, then resuspended in 3 ml of ice-cold IP buffer (2 volumes SDS Buffer: 1 volume Triton Dilution Buffer) (Triton Dilution Buffer: 100mM NaCl, 100mM Tris-HCl pH 8.6, 5mM EDTA pH 8, 5%

Triton X-100) and sonicated through the Digital Sonifier 450 (Branson) (ChIP-seq for Menin: 700-bp DNA fragments; ChIP-seq for H3K4me3 and H3K27me3: 200-bp DNA fragments). Immunoprecipitation was performed in IP buffer using the antibodies listed in Table S7.

The immunoprecipitated product was purified through Protein G dynabeads (ThermoFisher Scientific, 10003D), afterwards washed with low-salt (150mM NaCl, 20mM Tris-HCl pH 8, 2mM EDTA pH 8, SDS 0.1%, 1% Triton X-100) and high-salt wash (500mM NaCl, 20mM Tris-HCl pH 8, 2mM EDTA pH 8, SDS 0.1%, 1% Triton X-100) buffer 3 times with 1 ml of 150 mM Wash Buffer and once with 1 ml of 500 mM Wash Buffer with the use of a Dynamag magnet (ThermoFisher Scientific, catalog number 12321D).

Decrossilinking occurred through the incubation of beads (and the 1% input) with Decrosslinking Buffer (1% SDS, 0.1M $NaHCO_3$) at 65 °C overnight. DNA was purified with the QiaQuick PCR Purification kit (Qiagen) following manufacturer's instructions. Libraries were prepared as previously described (Adamo et al., 2015) and sequenced on a HiSeq 2000 instrument (Illumina) following manufacturer's protocol. Sequencing was performed in single end, 50 bp, with a coverage of 30 millions of reads for inputs, H3K4me1 and menin ChIP-seq and 60 millions for H3K27me3.

ChIP-seq reads were trimmed for potential adapter contamination using scythe 0.981 (min 4 nucleotides) before being aligned to the mm10 genome using bowtie 1.0 (Langmead et al., 2009) with –v 2 –m 1, and peaks were called using MACS 2.0.9 (Langmead et al., 2009; Zhang et al., 2008) with default settings. Although we relied mostly on quantitative analyses (below) to compare across replicates and conditions, when comparing peaks directly we considered peaks as overlapping if they shared at least one nucleotide.

Proximally bound genes were defined as having a peak within a -2.5kb/+1kb window around any of their RefSeq TSS. To identify the putative targets of intergenic sites bound by MENIN but in contact with a TSS only through chromatin conformation, we

relied on Hi-C data from the most similar cell type available, namely neural progenitors. We downloaded already processed, statistically significant interactions from the Gene Expression Omnibus entry GSE68582, extracted a bed file containing each region interacting with a RefSeq TSS, and intersected it (using intersectBed) with our regions of interest to find distal targets.

Peak calling is very sensitive to coverage and technical variation, and quantitative analysis of read distribution yields considerably more robust findings. To find differences across conditions, we therefore worked on the distribution of reads falling within relevant genomic windows. For each protein/mark, the windows were defined by merging the enriched regions across samples (*i.e.,* with BedTools: cat *.bed | sortBed -i - | mergeBed -i -). In this way, all regions enriched in at least one sample were considered for statistical testing, without duplicate genomic regions. The read counts were then compared across conditions with edgeR v.3.12.1 (Robinson et al., 2010), using the total number of mapped reads as library size and the TMM method for normalization. Differentially enriched regions were identified using the classical dispersion model of edgeR (based on a negative binomial model) and the exact test.

For the MENIN ChIP-seq, given the low quality of the peak calling (very low and variable number of peaks across samples, probably owing to a low and highly variable coverage), we performed a peak-call-agnostic differential enrichment analysis using diffReps (Shen et al., 2013), which uses sliding windows to directly identify regions of significant difference between sets of enrichment profiles.

Data representation (RNA-seq and ChIP-seq)

Principal component analyses were performed on the normalized, log-transformed read counts (+1 was added before log-transform to avoid errors on null values). Unless specified otherwise, all heatmaps show row z-scores of log-transformed normalized counts. Heatmaps were produced using the pheatmap R package.

External data (RNA-seq and ChIP-seq)

For external datasets, we used the authors' original peak calls (ChIP-seq) or quantification (RNA-seq) that are available from the respective GEO entries.

Exome analysis

The list of Kmt2b dependent genes was predicted by the intersection between DEGs (at FDR < 0.05) and genes with H3K4me3 reduction in *Kmt2b* cKO transdifferentiating MEFs with respect to controls. To test the hypothesis that the coding regions of *KMT2B*-sensitive genes are an untapped reservoir of dystonia-causing mutations, we queried a large repository of whole-exome sequencing (WES) data from a European-descent dystonia cohort. Accrued from movement disorders centers in Munich (Germany), Innsbruck (Austria), and Prague (Czech Republic), this cohort comprised 225 unrelated probands with various types of dystonic presentations in whom disease had not been explained. The recruitment strategy and clinical characterization process have been described previously (Zech et al., 2017) and demographics as well as phenotypic information of the cohort can be found in Table S5. The cohort was enriched for possible Mendelian disease causation, with 48% of probands having an early dystonia onset, 39% having a positive family history for dystonia, and 18% having a generalized distribution of dystonic symptoms. Dystonia was the sole presenting feature in 88% of the probands, whereas the remainder demonstrated additional movement disorders and/or other neurological/extra-neurological signs.

WES of all 225 dystonia subjects had been performed to an average on-target depth of 154x at the Helmholtz Center in Munich (Germany), through enrichment capture using the SureSelect All Exon system (Agilent Technologies) and subsequent paired-end 100-bp sequencing on HiSeq2000/2500 instruments (Illumina) in an automated manner, as described previously (Zech et al., 2017; Zech et al., 2016). For WES summary statistics, see Table S6. Standard protocols for sequence alignment,

variant annotation, genotype quality control, and variant visual inspection had been applied as detailed elsewhere (Zech et al., 2017; Zech et al., 2016). In the dominant model, variants were retained if they were (1) not referenced in the Exome Aggregation Consortium (ExAC) Browser (Lek et al., 2016) and an in-house control exome database (allowing the interrogation of altogether >141,000 control alleles), and (2) predicted to exert a loss-of-function effect (stop-gain/lost, frameshift, and splice-site alterations) or missense/in-frame indel alterations predicted to be among the 1% most deleterious aberrations in the human genome (Combined Annotation Dependent Depletion [CADD] score ≥20) (Kircher et al., 2014) (Figure S6A). In the recessive model, variants were retained if they were (1) present with a minor allele frequency <0.001 in the ExAC Browser and our in-house control exome database, and (2) predicted to exert a loss-of-function effect or missense/in-frame indel alterations predicted to be among the 10% most deleterious aberrations in the human genome (CADD score ≥10)(Kircher et al., 2014) (Figure S6A). As an additional supportive method for selecting the best possible candidate variants and genes, we obtained gene constraint metrics (probability of being loss-of-function intolerant (pLI) and missense Z scores) from the Exome Aggregation Consortium (ExAC) dataset (Lek et al., 2016) for all 216 *KMT2B*-target genes. ExAC-derived pLI scores ≥0.9 indicate that respective genes are intolerant to loss-of-function variation and likely relevant to haploinsufficient disease (Lek et al., 2016) ExAC-derived missense Z scores ≥2.0 indicate that respective genes are depleted of missense variants as compared with mutational expectation and thus more sensitive to this class of variation (Lek et al., 2016).

The top-scoring candidate variants selected from the dominant and recessive models (in the *KMT2B*-target genes *NOL4*, *SLC35F1* and *SLC40A1*) were verified by PCR amplification and conventional Sanger-sequencing.

References

Adamo, A., Atashpaz, S., Germain, P.L., Zanella, M., D'Agostino, G., Albertin, V., Chenoweth, J., Micale, L., Fusco, C., Unger, C.*, et al.* (2015). 7q11.23 dosage-dependent dysregulation in human pluripotent stem cells affects transcriptional programs in disease-relevant lineages. Nature genetics *47*, 132-141.

Basch, M.L., Ohyama, T., Segil, N., and Groves, A.K. (2011). Canonical Notch signaling is not necessary for prosensory induction in the mouse cochlea: insights from a conditional mutant of RBPjkappa. The Journal of neuroscience : the official journal of the Society for Neuroscience *31*, 8046-8058.

Germain, P.L., Vitriolo, A., Adamo, A., Laise, P., Das, V., and Testa, G. (2016). RNAontheBENCH: computational and empirical resources for benchmarking RNAseq quantification and differential expression methods. Nucleic acids research *44*, 5054-5067.

Glaser, S., Schaft, J., Lubitz, S., Vintersten, K., van der Hoeven, F., Tufteland, K.R., Aasland, R., Anastassiadis, K., Ang, S.L., and Stewart, A.F. (2006). Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. Development *133*, 1423-1432.

Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nature genetics *46*, 310-315.

Kranz, A., Fu, J., Duerschke, K., Weidlich, S., Naumann, R., Stewart, A.F., and Anastassiadis, K. (2010). An improved Flp deleter mouse in C57Bl/6 based on Flpo recombinase. Genesis *48*, 512-520.

Ladopoulos, V., Hofemeister, H., Hoogenkamp, M., Riggs, A.D., Stewart, A.F., and Bonifer, C. (2013). The histone methyltransferase KMT2B is required for RNA polymerase II association and protection from DNA methylation at the MagohB CpG island promoter. Molecular and cellular biology *33*, 1383-1393.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B.*, et al.* (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285-291.

Mihailovich, M., Bremang, M., Spadotto, V., Musiani, D., Vitale, E., Varano, G., Zambelli, F., Mancuso, F.M., Cairns, D.A., Pavesi, G.*, et al.* (2015). miR-17-92 fine-tunes MYC expression and function to ensure optimal B cell lymphoma growth. Nature communications *6*, 8725.

Pool, M., Thiemann, J., Bar-Or, A., and Fournier, A.E. (2008). NeuriteTracer: a novel ImageJ plugin for automated quantification of neurite outgrowth. Journal of neuroscience methods *168*, 134-139.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.

Shen, L., Shao, N.Y., Liu, X., Maze, I., Feng, J., and Nestler, E.J. (2013). diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. PloS one *8*, e65598.

Simon, C.M., Janas, A.M., Lotti, F., Tapia, J.C., Pellizzoni, L., and Mentis, G.Z. (2016). A Stem Cell Model of the Motor Circuit Uncouples Motor Neuron Death from Hyperexcitability Induced by SMN Deficiency. Cell reports *16*, 1416-1430.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols *7*, 562-578.

Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Sudhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. Nature *463*, 1035-1041.

Young, M.D., Wakefield, M.J., Smyth, G.K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. Genome biology *11*, R14.

Zech, M., Boesch, S., Jochim, A., Weber, S., Meindl, T., Schormair, B., Wieland, T., Lunetta, C., Sansone, V., Messner, M.*, et al.* (2017). Clinical exome sequencing in early-onset generalized dystonia and large-scale resequencing follow-up. Movement disorders : official journal of the Movement Disorder Society *32*, 549-559.

Zech, M., Boesch, S., Maier, E.M., Borggraefe, I., Vill, K., Laccone, F., Pilshofer, V., Ceballos-Baumann, A., Alhaddad, B., Berutti, R.*, et al.* (2016). Haploinsufficiency of KMT2B, Encoding the Lysine-Specific Histone Methyltransferase 2B, Results in Early-Onset Generalized Dystonia. American journal of human genetics *99*, 1377-1387.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology *9*, R137.