List of Supplementary Tables

Supplementary Table 1
Patient information.

Supplementary Table 2
Human patient fibroblasts hypermethylated DMRs. Supplied as BED formatted tab-delimited text file. All co-ordinates in hg19.

Supplementary Table 3
Human patient fibroblasts hypomethylated DMRs. Supplied as BED formatted tab-delimited text file. All co-ordinates in hg19.

Supplementary Table 4
Significantly enriched GO-terms from genes associated with human patient fibroblast hypermethylated DMRs.

Supplementary Table 5
Mouse neural differentiation hypermethylated DMRs. Supplied as BED formatted tab-delimited text file. All co-ordinates in mm10.

Supplementary Table 6
Oligonucleotides used in this study

Supplementary Table 7
Summary of sequencing statistics for RNA-seq. Aligned reads are after multi-mapper removal.

Supplementary Table 8
Summary of sequencing statistics for ChIP-seq. Aligned reads are after duplicate and multi-mapper removal. For ChIP Rx-seq this reports reads aligning to both genomes (hg19/dm6).

Supplementary Table 9
Summary of sequencing statistics for RRBS. Aligned reads counts are following PCR duplicate removal.

Supplementary Table 10
Mouse neural differentiation hypomethylated DMRs. Supplied as BED formatted tab-delimited text file. All co-ordinates in mm10.

Supplementary Table 1: Patient summary

ID	Height (s.d.)	OFC(s.d.)	nucleotide change	protein change	gender	age at measurement	birthweight (s.d.)	neurocognitive impairment	brain MRI	inheritance	country of origin
P1	-5.5	-4.3	c.988T>C	p.W330R	female	1y	-2.78	severe		*de novo*	USA
P2	-4.8	-6.6	c.988T>C	p.W330R	male	1y7m	-1.11	moderate	normal	*de novo*	NZ
P3	-2.8	-5.0	c.997G>A	p.D333N	male	4y6m	-2.33	moderate-severe	normal	*de novo*	Spain

Supplementary Table 4:

GO-terms enriched in genes associated with human patient fibroblast hypermethylated DMRs

P-values are from Benjamini-Hochberg corrected Fisher's exact tests

Fold change was calculated as the % of genes associated with DMR probes that were associated with a specific term over the % genes associated with control probes for that term

Term ID	Adjusted p-value	Fold Change	Term description
GO:0001071	4.07E-61	4.012	nucleic acid binding transcription factor activity
GO:0003677	8.27E-60	4.035	DNA binding
GO:0048513	1.56E-52	2.424	animal organ development
GO:0048856	6.43E-50	1.897	anatomical structure development
GO:0032501	6.48E-41	1.662	multicellular organismal process
GO:0006357	8.28E-40	3.069	regulation of transcription from RNA polymerase II promoter
GO:2000112	2.88E-34	1.91	regulation of cellular macromolecule biosynthetic process
GO:0010468	1.12E-33	1.847	regulation of gene expression
GO:0009889	9.60E-33	1.839	regulation of biosynthetic process
GO:0019438	1.10E-27	1.745	aromatic compound biosynthetic process
GO:1901362	1.39E-27	1.725	organic cyclic compound biosynthetic process
GO:0034645	7.82E-22	1.592	cellular macromolecule biosynthetic process
GO:0010467	3.39E-20	1.551	gene expression
GO:0003006	1.06E-16	2.981	developmental process involved in reproduction
GO:0048522	7.62E-14	1.559	positive regulation of cellular process
GO:0008283	4.65E-11	1.898	cell proliferation
GO:0040011	1.09E-09	1.853	locomotion
GO:0009719	7.10E-08	1.748	response to endogenous stimulus
GO:0048638	1.01E-07	3.213	regulation of developmental growth
GO:0040007	2.15E-07	1.988	growth
GO:0042221	2.57E-07	1.412	response to chemical
GO:0051674	2.97E-07	1.849	localization of cell
GO:0097152	7.75E-07	15.124	mesenchymal cell apoptotic process
GO:2001053	7.75E-07	15.124	regulation of mesenchymal cell apoptotic process
GO:0045844	1.62E-06	6.64	positive regulation of striated muscle tissue development
GO:0071772	2.21E-06	3.855	response to BMP
GO:0007610	3.34E-06	2.204	behavior
GO:0098609	4.36E-06	2.03	cell-cell adhesion
GO:0006935	6.43E-06	2.086	chemotaxis
GO:1905330	7.86E-06	4.365	regulation of morphogenesis of an epithelium
GO:0007224	3.02E-05	4.95	smoothened signaling pathway
GO:0071363	4.62E-05	1.891	cellular response to growth factor stimulus
GO:0009605	0.000109	1.474	response to external stimulus
GO:1905114	0.000229	2.138	cell surface receptor signaling pathway involved in cell-cell signaling
GO:1905207	0.000436	6.534	regulation of cardiocyte differentiation
GO:1905332	0.000436	6.534	positive regulation of morphogenesis of an epithelium
GO:2000677	0.000476	7.642	regulation of transcription regulatory region DNA binding
GO:0198738	0.000594	2.135	cell-cell signaling by wnt
GO:1905209	0.000821	8.47	positive regulation of cardiocyte differentiation
GO:0010817	0.000895	1.96	regulation of hormone levels
GO:0007584	0.001744	3.203	response to nutrient
GO:0000988	0.007606	1.766	transcription factor activity, protein binding
GO:0042573	0.007617	8.25	retinoic acid metabolic process
GO:0098811	0.007805	4.478	transcriptional repressor activity, RNA polymerase II activating transcription factor binding
GO:0015804	0.008626	6.05	neutral amino acid transport
GO:0023061	0.00919	1.837	signal release
GO:0021885	0.009638	3.799	forebrain cell migration

GO:0046632	0.00966	3.22	alpha-beta T cell differentiation
GO:1904029	0.010997	2.809	regulation of cyclin-dependent protein kinase activity
GO:0051046	0.013619	1.675	regulation of secretion
GO:0022412	0.015972	2.036	cellular process involved in reproduction in multicellular organism
GO:0019218	0.017429	2.936	regulation of steroid metabolic process
GO:0048609	0.020188	1.55	multicellular organismal reproductive process
GO:0009914	0.023342	1.895	hormone transport
GO:0023056	0.028532	1.384	positive regulation of signaling
GO:0045182	0.030253	5.021	translation regulator activity
GO:0032656	0.03057	7.26	regulation of interleukin-13 production
GO:0001523	0.030816	3.737	retinoid metabolic process
GO:0003682	0.031807	3.682	chromatin binding
GO:0004497	0.031807	2.893	monooxygenase activity
GO:0023052	0.032996	1.142	signaling
GO:0044700	0.032996	1.143	single organism signaling
GO:0048511	0.034453	2.017	rhythmic process
GO:0006775	0.034541	3.63	fat-soluble vitamin metabolic process
GO:0045619	0.035029	2.375	regulation of lymphocyte differentiation
GO:0034284	0.03559	2.173	response to monosaccharide
GO:0044065	0.03786	6.6	regulation of respiratory system process
GO:0046872	0.039557	1.268	metal ion binding
GO:1905276	0.041906	4.776	regulation of epithelial tube formation
GO:0003299	0.046788	6.05	muscle hypertrophy in response to stress
GO:1905314	0.046788	6.05	semi-lunar valve development
GO:1905939	0.046788	6.05	regulation of gonad development

Supplementary Table 6:

oligonucleotides used in this study

primer name	sequence (5'-3')	purpose
D3a_F1	CCCACTTCCATCACCCCAAT	confirmatory sanger sequencing of human DNMT3A
D3a_R1	TGCCTCATTCAGATGGAGCT	confirmatory sanger sequencing of human DNMT3A
DNMT3A_PWWP_F	GATACGAATTCGAGTACGAGGACGGCCGG	clone human DNMT3A-PWWP into pGEX6p1
DNMT3A_PWWP_R	GTATCCTCGAGCTATGGCTCCAGGCCCTTAGG	clone human DNMT3A-PWWP into pGEX6p1
pW330R_DNMT3a_F	CTGGGTCATGcGGTTCGGAGACG	site-directed mutagenesis of human DNMT3A PWWP domain W330R mutation
pW330R_DNMT3a_R	CGGGTGCCTTCAGCTGCT	site-directed mutagenesis of human DNMT3A PWWP domain W330R mutation
pW297del_F	GGGAAACTGCGGGGCTTC	site-directed mutagenesis of human DNMT3A PWWP domain W297del mutation
pW297del_R	CACCAGCTCCCCAATGCC	site-directed mutagenesis of human DNMT3A PWWP domain W297del mutation
pI310N_DNMT3A_F	CCAGGCCGCAaTGTGTCTTGG	site-directed mutagenesis of human DNMT3A PWWP domain I310N mutation
pI310N_DNMT3A_R	CCACCAGGAGAAGCCCCG	site-directed mutagenesis of human DNMT3A PWWP domain I310N mutation
guide1_PWWP_for	CACCGTGCCTTCAGCTGCTCGGCTC	clone sgRNA into pX461 to knock-in Dnmt3a-W326R with nCas9 into mESCs (targeting sequence blue)
guide1_PWWP_rev	AAACGAGCCGAGCAGCTGAAGGCAC	clone sgRNA into pX461 to knock-in Dnmt3a-W326R with nCas9 into mESCs (targeting sequence blue)
guide2_PWWP_for	CACCGTGGGTCATGTGGTTCGGAGA	clone sgRNA into pX461 to knock-in Dnmt3a-W326R with nCas9 into mESCs (targeting sequence blue)
guide2_PWWP_rev	AAACTCTCCGAACCACATGACCCAC	clone sgRNA into pX461 to knock-in Dnmt3a-W326R with nCas9 into mESCs (targeting sequence blue)

Dnmt3a_ODN1	TCTTCTCTTCCCCGACCTCTTAACCACCAAGTCCAACTCACCACTGAGAACTTGCCATCTC CaAACCtCATGACCCAGCGAGTGCCTTCAGCTGCTCGGCTCCGtCCTGTCATCCACCAA	antisense repair template for Dnmt3a-W326R knock-in into mESCs with nCas9 (substitution, silent PAM-site mutation for guideA and silent blocking mutation for guideB in lower case)
guide3_PWWP_for	CACCGAAGGCACTCGCTGGGTCATG	clone sgRNA in pX458 to knock-in Dnmt3a-W326R with Cas9 (targeting sequence blue)
guide3_PWWP_rev	AAACCATGACCCAGCGAGTGCCTTC	clone sgRNA in pX458 to knock-in Dnmt3a-W326R with Cas9 (targeting sequence blue)
T7_Dnmt3a_guide3_f	TGTAATACGACTCACTATAGGAAGGCACTCGCTGGGTCATG	to generate IVT template for Dnmt3a sgRNAs
universal_reverse	AAAAGCACCGACTCGGTGCC	to generate IVT template for Dnmt3a sgRNAs
Dnmt3a_ODN2	TTCCCCGACCTCTTAACCACCAAGTCCAACTCACCACTGAGAACTTGCCATCTCCGAAtCtC ATGACCCAGCGAGTGCCTTCAGCTGCTCGGCTCCGGCCTGTCATCCACCAAGACACAA	antisense repair template for W326R knock-in into mESCs and mice with Cas9 (substitutions in lower case)
guide4_PWWP_for	CACCGGGGCTTTGGCATTGGAGAGC	clone sgRNA in pX458 to to generate Dnmt3a-W293 deletion with Cas9 in mESCs (targeting sequence blue)
guide4_PWWP_rev	AAACGCTCTCCAATGCCAAAGCCCC	clone sgRNA in pX458 to to generate Dnmt3a-W293 deletion with Cas9 in mESCs (targeting sequence blue)
Dnmt3a_ODN3	TCCACCAAGACACAATTCGGCCTGGCCACCAGGAGAAGCCCCGAAGTTTCCCCACtAGCT CTCCAATGCCAAAGCCCCGGCCATCCTGGAGCCCCAAAGAGCAAAAGTCAATACACAGCA	antisense repair template to generate Dnmt3a-W293 deletion in mESCs with Cas9 (silent PAM site mutation in lower case)
guide5_PWWP_for	CACCGCACAATTCGGCCTGGCCACC	clone sgRNA in pX458 to knock-in I306N with Cas9 into mESCs (targeting sequence blue)
guide5_PWWP_rev	AAACGGTGGCCAGGCCGAATTGTGC	clone sgRNA in pX458 to knock-in I306N with Cas9 into mESCs (targeting sequence blue)
Dnmt3a_ODN4	GAGTGCCTTCAGCTGCTCGGCTCCGGCCTGTCATCCACCAAGACACAtTTCGGCCTGGCC ACCAtGAGAAGCCCCGAAGTTTCCCCCACACCAGCTCTCCAATGCCAAAGCCCCGGCCAT	antisense repair template for I306N knock-in into mESCs with Cas9 (substitution and silent PAM site mutation in lower case)
Dnmt3a_ex7f	ACCCTGCTTCTCCGACTGTG	check cDNA expression of Dnmt3a-W326R
Dnmt3a_ex9r	TTCGTAGATGGCTTTGCGGTA	check cDNA expression of Dnmt3a-W326R
Dnmt3a mm var	CTTTGGCATTGGAGAGCTGG	genotyping Dnmt3a-W326R mouse line
Dnmt3a mm var R	TGGCTATGGCTTCTTCCACT	genotyping Dnmt3a-W326R mouse line
Hoxc13_BS_for	ATTTGTTTTTTAGGGTTAAGGAGTT	primers for bisulfite PCR Hoxc13 (mouse)

Hoxc13_BS_rev	ACAATACACCTAACTATCCCAACCA	primers for bisulfite PCR Hoxc13 (mouse)
Sox1_BS_for	TTTATTGTTGTTTTTAATTTTTTTT	primers for bisulfite PCR Sox1 (mouse)
Sox1_BS_rev	AAACTCCACTTTCTAAATCTAAAAC	primers for bisulfite PCR Sox1 (mouse)
Foxa1_BS_for	GATTTTYGGATAAGTTAGGTAAGGGTTTTTATTG	primers for bisulfite PCR Foxa1 (mouse)
Foxa1_BS_rev	AACCCCRCTATAATCCAAAATCTAAAAAC	primers for bisulfite PCR Foxa1 (mouse)

Supplementary Table 7
Summary of sequencing statistics for RNA-seq. Aligned reads are after multi-mapper removal.

Sample	**Total Reads ($x10^6$)**	**Aligned Reads ($x10^6$)**
Fibroblasts		
Control 1	33.20	21.77
Control 2	42.41	27.65
Patient 1	45.74	29.89
Patient 2	57.86	38.08
NPCs		
WT1	32.44	23.69
WT2	34.89	25.75
WT3	31.24	23.18
Hom1	37.07	27.19
Hom2	29.31	21.68
Hom3	30.19	22.13

Supplementary Table 8
Summary of sequencing statistics for ChIP-seq. Aligned reads are after duplicate and multi-mapper removal. For ChIP Rx-seq this reports reads aligning to both genomes (hg19/dm6).

Sample	**Total Reads ($x10^6$)**	**Aligned Reads ($x10^6$)**
H3K27me3 ChIP-seq		
Control 1 Input	62.67	52.55
Control 2 Input	64.26	53.85
Patient 1 Input	74.13	62.06
Patient 2 Input	52.55	43.85
Control 1 K27me3	64.50	53.66
Control 2 K27me3	53.31	44.46
Patient 1 K27me3	54.17	45.04
Patient 2 K27me3	51.75	42.90
H3K4me3 ChIP-seq		
Control 1 Input	44.49	38.19
Control 2 Input	44.72	38.40
Patient 1 Input	42.67	36.61
Patient 2 Input	42.81	36.61
Control 1 K4me3	37.38	31.80
Control 2 K4me3	36.49	30.94
Patient 1 K4me3	38.03	32.38
Patient 2K4me3	37.51	31.87
H3K36me3 ChIP Rx-seq		
Control 1 Input	71.64	60.28 / 0.62
Control 2 Input	71.95	60.39 / 0.68
Patient 1 Input	74.12	62.37 / 0.63
Patient 2 Input	81.06	67.71 / 0.71
Control 1 K36me3	63.83	53.84 / 0.60
Control 2 K36me3	61.03	51.37 / 0.61
Patient 1 K36me3	64.35	54.20 / 0.55
Patient 2 K36me3	58.91	49.45 / 0.51
H3K27me3 ChIP Rx-seq		
Control 1 Input	39.76	33.78 / 0.30
Patient 1 Input	38.51	32.67 / 0.24
Control 1 K27me3	42.51	35.69 / 0.53
Patient 2 K27me3	45.26	38.01 / 0.46

Supplementary Table 9
Summary of sequencing statistics for RRBS. Aligned reads counts are following PCR duplicate removal.

Sample	Total Reads ($x10^6$)	Aligned Reads ($x10^6$)	Mean CG Coverage	Bisulfite Conversion Rate
Patient blood samples				
Control 3	25.16	11.42	23.66	99.81%
Patient 3	91.95	52.33	43.67	99.64%
NPCs				
WT1	40.17	24.07	42.58	99.68%
WT2	40.03	24.20	42.09	99.68%
WT3	47.01	25.91	46.38	99.69%
Hom1	51.55	24.91	42.99	99.69%
Hom2	45.47	25.19	44.01	99.67%
Hom3	43.62	23.45	42.46	99.69%
Het	46.95	25.60	44.18	99.67%
Mouse cortex				
M1	49.96	26.85	46.36	99.68%
M2	54.13	29.23	51.23	99.67%
M3	51.81	27.75	47.97	99.67%
M4	55.00	29.09	62.88	99.73%
M5	49.59	27.49	49.89	99.69%
M6	53.32	28.29	54.31	99.70%

Supplementary Note

Summary of clinical phenotypes

P1 was referred to the study with severe non-syndromic microcephalic dwarfism. She was delivered by caesarean section at 38 weeks gestation due to intrauterine growth retardation (IUGR) and breech presentation, with a birth weight of 1.9kg (-2.8sd), and length 40.64cm (-4.3sd). Postnatally, growth continued to be severely impaired, such that at 1 year height was 60.5cm (-5.4sd), weight 5.79kg (-4.7sd) and head circumference 41.5cm (-4.1sd). Sparse hair and 11 pairs of ribs were noted, however no significant malformations or dysmorphic features were reported. Short-stature was proportionate, and skeletal survey was non-contributory. A mild leukopenia was noted in early childhood (White blood cells (WBC) 2.4-2.8 x 10^9/L), with impaired responses to some but not all vaccines. However no significant illnesses or immunodeficiency were subsequently reported to age 13. Developmental progress was severely delayed such that by age 13, no speech was evident. Karyotype 46, XX, metabolic investigations, DEB chromosome breakage testing, GH, IGF1 levels all normal.

P2 was born at 36/40 after induction of labour due to being small for gestational age (SGA). Birth weight 2.26kg (-1.1sd), length 47cm (-0.33sd), OFC 32cm (-0.57sd). Postnatally growth restriction became progressively more evident with head circumference 41cm (-6.6sd), height 68.9cm (-4.6sd), and weight 6.64kg (-5.2sd) at 19 months of age. Medically, his only issues were of constipation and recurrent episodes of viral bronchiolitis/wheeze in early childhood. Moderate developmental

delay. He was non-dysmorphic. Bone age, mildly delayed, 15 months at 19 months of chronological age, skeletal survey essentially normal, short broad metacarpals and phalanges noted. Patellae present. WBC 3.1 x 10^9/L. Normal CGH microarray, 46,XY. IGF1, GH, metabolic screen all normal. MRI head, normal.

P3 was referred at age 2 years for investigation of short stature, microcephaly, and severe developmental delay. Normal delivery at 37 weeks of gestation; birth weight 1.9 kg (-2.3 SDS), length 42 cm (-3.5 SDS), and head circumference 31 cm (-1.5 SDS). Postnatal growth was severely delayed. Examination findings: proportional short stature with microcephaly, at age 4.5 years height 93.8 cm (-3.2 SDS) and weight 11 Kg (-2.6 SDS). Facial appearance, strabismus, epicanthic folds, wide forehead, sparse hair. Short broad metacarpals and phalanges were noted and macro-orchidism was apparent (testicular volume 5mL, bilateral). Medical issues noted, included bronchitis and recurrent otitis. At age of 4.5 years, bone age was 5 years. Normal IGF-1 levels (175 ng/ml; normal, 58-211). Moderately raised IGFBP-3 concentrations (5.8 mg/L, normal, 1.0-4.95). The GH stimulation test with L-dopa showed a normal response (GH peak 8.8 ng/mL, normal >7,4 ng/mL). Karyotype 46,XY. Metabolic screening, MLPA, cranial MRI, and microarray CGH, normal.

Supplementary Methods

Exome sequencing

FASTQs were aligned with *BWA-MEM*[1] and variants called in accordance with GATK's best practice guidelines[2,3]. Each trio was analysed under both recessive and *de novo* dominant inheritance models. For patients 1 and 2, rare (<1% or 0.1% allele frequencies in dbSNP and ExAC[4] for recessive and *de novo* inheritance models respectively) and functional variants affecting protein sequence or splice sites were retained according to each inheritance model. Variant filtering was performed similarly in patient 3 except frequency information was obtained from ExAC, 1000 genomes[5] and Genomes and Cincinnati Children's Hospital internal database. *De novo* variants were required to be absent from these databases.

Differentiation of mouse ES cells

For differentiation into embryoid bodies (EB's) 400 cells per 20 µl were seeded into hanging droplets in Dulbecco's modified Eagle's medium (DMEM; Life Technologies) supplemented with 20% FBS (HyClone), 1x MEM non-essential amino acids (Sigma), 2 mM L-Glutamine, 0.1 mM β-mercaptoethanol and 5% penicillin-streptomycin antibiotics. After 3 days embryoid bodies were transferred into ultra-low attachment surface plates (Corning) for further growth. At day 6 EB's were transferred to 0.1% gelatine coated dishes and cultured until day 9. For neural differentiation cells were differentiated as described previously[6]. Briefly, cells were plated in 0.1% gelatine coated dishes in Neurobasal/DMEM/F12 (Life Technologies) supplemented with 1x B27 (Life Technologies), 1 x N2 (Life Technologies) and 5% penicillin-streptomycin antibiotics. Cells were fed daily and harvested after 9 days of differentiation.

Expression and purification of recombinant PWWP protein

DNMT3A-PWWP domain (AA 283-424) was cloned into pGEX6p1 (restriction sites: EcoRI and XhoI). The PWWP-mutations introduced using the Q5® Site-Directed Mutagenesis Kit (NEB). Primers listed in Supplementary Table 6 Recombinant DNMT3A-PWWP was expressed in Rosetta-2 cells (Novagen). Expression was induced with 0.1mM IPTG and cells grown at 20°C overnight. Pelleted cells were lysed in NETN buffer (20 mM Tris pH8, 100 mM NaCl, 1 mM EDTA, 0.5% NP-40) followed by sonication. After centrifugation, the soluble fraction was loaded onto Glutathione Sepharaose 4B beads (GE Healthcare) and incubated for 4h at 4°C. PWWP protein was eluted twice with 40 mM reduced Glutathione (Sigma) in elution buffer (50 mM Tris pH8) at room temperature. Glutathione was removed from the protein eluate using Amicon Ultra 3K devices (Millipore) before loading onto peptide-arrays. Protein concentration determined using Bradford protein assay (Biorad). Expressed proteins separated on 12% SDS-PAGE gels and visualized either with Coomassie Blue R250 or transferred onto nitrocellulose membrane and detected with an antibody against the GST-tag (GST-antibody, GE Healthcare 27-4577-01; 1:5000).

Chromatin immunoprecipitation and sequencing

$2x10^6$ cells were harvested, washed and crosslinked with 1% methanol-free formaldehyde in PBS. Crosslinked cells were lysed directly for 15 min on ice in cold lysis buffer (10 mM EDTA, 50 mM Tris-HCl (pH8), 1% SDS). All buffers used during chromatin capture were freshly supplemented with protease inhibitors (Roche) and 1

mM DTT. For ChIP-Rx[7] crosslinked *Drosophila* chromatin (S2 cells, a kind gift from R. Illingworth) was spiked into samples before sonication (ratio of 20:1 human to *Drosophila* cells). Lysates were sonicated with two 30 sec pulses using a soniprep 150 plus (MSE) followed by 10 cycles (30 s on / 30 s off on 'high' setting at 4°C) on a Bioruptor Plus (Diagenode). Following centrifugation for 10 min at 20,000xg, 4°C, supernatant was supplemented with Triton X-100 (final concentration 1%) and 50 µg/ml BSA and then added to Protein A dynabeads (Invitrogen) pre-coupled with the respective antibody (H3K27me3 antibody (Cell Signaling; C36B11; Cat.no. 9733S); H3K4me3 (Millipore; 07-473); H3K36me3 (Active Motif; #61101)). 10% of input material was retained for use as a reference. The bead-chromatin mixture was incubated under rotation at 4°C overnight. Beads were then washed twice for 10 min at 4°C with each of the following buffers: IP dilution buffer (2 mM EDTA, 150 mM NaCl, 20 mM Tris HCl pH8, 1% Triton X-100), buffer A' (20 mM Tris pH8, 500 mM NaCl, 1 mM EDTA pH8, 1% TritonX-100, 0.1% Na-deoxycholate, 0.1% SDS), buffer B' (20 mM Tris pH8, 1 mM EDTA pH8, 250 mM LiCl, 1% NP-40, 0.1% Na-deoxycholate). Finally, beads were washed three times with 1xTE buffer (1 mM EDTA pH8, 10 mM Tris pH8). Captured chromatin was eluted, followed by reverse crosslinking through addition of Tris HCl pH6.8 (100 mM), NaCl (300 mM) and 20 µg RNAse A at 65°C overnight. Input material was similarly de-crosslinked. Degradation of proteins with 20 µg Proteinase K and incubation at 37°C for 2 hrs. Samples purified with the MinElute PCR purification kit (QIAGEN).

ChIP-seq libraries were prepared using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (E7645) according to the manufacturer instructions using barcoded adapters (NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1 or 2), E7335; E7500S). For input reference libraries 50 ng material were used. Adapter-

ligated DNA was size selected for an insert size of 150 or 200 bp using AMPure XP beads. Libraries were amplified for 7 cycles in the final PCR amplification step. Amplified libraries were purified using AMPure XP beads and eluted in EB. Libraries were pooled in an equimolar ratio.

Analysis: Tracks for data visualisation generated using Homer (*v4.8*)[8]. Aligned BAMs were converted to tag directories setting the fragment length to 180bp and converted to *bigWig* files using Homer's *makeUCSCfile* function after filtering with the *removeOutOfBounds.pl* function. ChIP-seq peaks were called from tag directories using Homer's *findPeaks* function (settings: *-style histone*) and an input chromatin control sample. For the re-analysis of published ChIP-seq[9], such a control was unavailable with peak calling therefore done without. The fragment size used for this data was 200bp.

Histone acid extraction and histone PTM detection by mass spectrometry

In brief $0.5x10^7$ fibroblasts were lysed in TEB buffer (PBS with 0.5% Triton X-100 freshly supplemented with 1xInhibitor Cocktail (Calbiochem Set 1,#539131), 1 mM DTT, 1 mM PMSF, 50 mM sodium fluoride, 1 mM sodium orthovanadate). Acid extraction was done overnight with 0.2N HCl at 4°C. Acid precipitated proteins were resolved by electrophoresis on 4-12% Bis-Tris gels run with MOPS buffer up to 200V and stained with Coomassie blue. Individual gel bands corresponding to histones were excised, propionylated and digested with trypsin as described[10]. After digestion 1-1.5µg of peptides were desalted using STAGE tips[11] and resuspended in 6 µl 0.1% TFA for LC MS/MS. Analyses were performed on an Orbitrap Fusion Lumos coupled to Dionex Ultimate3000RSLCnano UHPLC system fitted with EasySpray 50 cm 2 µm

(Thermo Fischer Scientific). All acquisitions were performed in DIA MSX mode. UHPLC gradient was as follows: 2-32% mobile phase B over 120 min, following 32%-95% B from 120 to 150 min. Mobile phase B was reduced back to 2% at 155 min. Mobile phase A consisted of 0.1% formic acid in water, and mobile phase B was 80% acetonitrile in 0.1% formic acid[12]. Data were processed using Skyline[13] using previously described manually curated spectral library[14]. Relative abundance of modified peptides was calculated as a ratio of peptide precursor intensity to the sum of intensities of all isoforms sharing the same amino acid sequence.

RRBS library preparation and sequencing

Briefly, unmethylated phage λ DNA 0.5ng (NEB) was spiked into each sample to allow assessment of bisulfite conversion efficiency. The methylation-insensitive restriction enzyme MspI was then used to digest the genomic DNA, and digested fragments were ligated to adapters. Adapter-ligated fragments were then repaired before bisulfite conversion with the Qiagen Epitect Fast Bisulfite Conversion kit. Bisulfite-treated adapter-ligated fragments were amplified by 9 cycles of PCR and purified using Agencourt RNAClean XP beads. Libraries were quantified using the Qubit dsDNA HS assay and assessed for size and quality using the Agilent Bioanalyzer DNA HS kit.

Sequencing was performed using the NextSeq 500/550 high-output version 2 kit (75bp paired end reads) on the Illumina NextSeq 550 platform. As instructed for the NuGen RRBS kit, 12bp index reads were generated to sequence the Unique Molecular Identifiers (UMI) in addition to the index present in the adaptors. Libraries were combined into equimolar pools and run on a single flow cell. 10% PhiX control

library was spiked in to facilitate sequencing by generating additional sequence diversity. Library preparation and sequencing performed by WTCRF.

Raw Illumina sequencing output from the NextSeq were converted to paired FASTQ files without demultiplexing using '*bcl2fastq*' and default settings (*v2.17.1.14*). These FASTQ files were then demultiplexed using custom python scripts considering indexes with perfect matches to the sample indexes. The different lanes for each sample were then combined.

Sequencing quality was assessed with FASTQC (*v0.11.4*). Low quality reads and remaining adaptors were removed using *TrimGalore (v0.4.1*, Settings: *--adapter AGATCGGAAGAGC --adapter2 AAATCAAAAAAAC*). NuGen adaptors contain extra diversity bases to facilitate sequencing. These were removed using the '*trimRRBSdiversityAdaptCustomers.py'* Python script provided by NuGen (*v1.11*). The paired end reads were then aligned to the hg19 or mm10 genome using Bismark (*v0.16.3* with Bowtie2 *v2.2.6* with settings: *-N 0 -L 20*)[15,16] before PCR duplicates were identified and removed using the 6bp UMIs present in the index reads and the '*nudup.py*' Python script supplied by NuGen (*v2.3*). Aligned BAM files were processed to report coverage and number of methylated reads for each CpG observed. Forward and reverse strands were combined using Bismark's *methylation extractor* and *bismark2bedgraph* modules with custom Python and AWK scripts. Custom scripts are available from the authors on request.

Infinium® MethylationEPIC BeadChip

Raw *idat* files were processed to beta values with the Bioconducter package *minfi* (*v1.22.1*) using single-sample normal exponential out of band (ss noob) background

correction and dye bias normalization[17,18] together with Bioconductor EPIC array annotations (*IlluminaHumanMethylationEPICmanifest v0.3.0* and *IlluminaHumanMethylationEPICanno.ilm10b2.hg19 v0.6.0*). Probes overlapping known SNPs (MAF>1% and within 5 bp of the probe's CpG target)[19], located on sex chromosomes, or with detection p-value > 0.01 were excluded. Detection p-values were also derived from *idat* files using *minfi*.

RNA sequencing alignment

RNA-seq read quality was analysed using FASTQC (*v0.11.4*, *https://www.bioinformatics.babraham.ac.uk/projects/fastqc*), with low quality reads and adaptors removed using *trim-galore* with default settings (*v0.4.1*). Reads were aligned to hg19 or mm10 genome using bowtie 2 (*v2.3.1*, with settings: *-N 1 -L 20 --no-unal*). Multi-mapping reads excluded using SAMtools (*v1.6,* with settings: *-bq 10*)[20] respectively.

Metagene plots

Scaled metagene plots of Infinium data were generated by defining 40 windows from the Transcription Start Site (TSS) to the Transcription End Site (TES) of each ENSEMBL protein coding transcript (*Ensembl Release 75/GCRh37*: 90,273 transcripts) orientated in the direction of transcription. The methylation level in each window per sample was defined as the mean Beta of the probes located within it. 250bp windows extending to 5Kb upstream of the TSS and 5Kb downstream of the TES were also considered. The mean profile of transcripts annotated to the same

ENSEMBL gene ID was then calculated from these windows before calculating the mean profile per sample.

Comparison of Mouse and human DMRs

NPC-differentiation DMRs were assigned ENSEMBL mouse gene IDs using *ChIPpeakAnno* as described above for Infinium data except using CpGs rather than probes. To permit direct comparison with human DMRs, annotated Illumina EPIC probe locations were used to map human DMR and background control probes onto the hg38 assembly, *ChIPpeakAnno* used to allocate ENSEMBL (Release: 91/mm10) gene IDs and mm10 ENSEMBL mouse gene ID orthologs for the human DMRs/control genes identified from ENSEMBL annotation. Only one-to-one orthologues were assigned. Overlap was then statistically tested by a Fisher's exact test, comparing the resulting gene lists for proportion of mouse-DMR-associated genes found in the human-DMR-associated gene list to the mouse-DMR genes associated with the human background control gene set.

Gene Ontology analysis and Mapping of DMRs and DMVs to genes and GO-terms

DMR and control probes were mapped to their closest annotated TSS with the Bioconductor package ChIPpeakAnno (*v3.10.2*)[21], using ENSEMBL annotated protein coding genes (*Ensembl Release 75/GCRh37*), employing the *annotatePeakInBatch* command, with settings: *FeatureLocForDistance="TSS"* and *output="shortestDistance"* to derive DMR and control Ensembl gene ID lists. Gene lists were mapped to Biological Process and Molecular Function GO-terms using ENSEMBL Biomart. All parental terms were identified using Custom R scripts and

the Bioconductor GO.db package (*v3.4.1*). For each GO term, statistical enrichment for term-associated genes in the DMR list versus control list was tested by Benjami-Hochberg corrected Fisher's exact test, with an FDR rate of 0.05. Terms with < 10 genes in control list were excluded from analysis. GO terms were filtered for semantic similarity (Bioconductor packages *GOSemSim*[22] and *org.Hs.eg.db)*. Similar terms were defined on the basis of Wang Similarity > 0.7[22], using a modified *simplify* function from the *clusterProfiler* Bioconductor package[23]. For each group, the term with the lowest adjusted p-value was retained.

External data sources

ENCODE ChromHMM segmentations for NHLF cells[24] were downloaded from the UCSC browser (hg19).

H3K27me3 ChIP-seq data from mouse NPCs and RNAseq from mESCs and terminally differentiated neurons[9] was downloaded as FASTQ files from the European Nucleotide Archive. Accessions are as follows (TN = terminally differentiated neurons):

ENA SRA ID	GEO ID	Sample
SRS140079	GSM632044	NPC H3K27me3 ChIPseq Rep 1
SRS140078	GSM632043	NPC H3K27me3 ChIPseq Rep 2
SRS140077	GSM632042	NPC H3K27me3 ChIPseq Rep 3
SRS257383	GSM778487	mESC RNAseq Rep A1
SRS257384	GSM778488	mESC RNAseq Rep A2

SRS174590	GSM687304	mESC RNAseq Rep B1
SRS174591	GSM687305	mESC RNAseq Rep B2
SRS257387	GSM778491	TN RNAseq Rep A1
SRS257388	GSM778491	TN RNAseq Rep A2
SRS174592	GSM687306	TN RNAseq Rep B1
SRS174593	GSM687307	TN RNAseq Rep B2

References:

1. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Preprint at* https://arxiv.org/abs/1303.3997v2 (2013).
2. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
3. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-8 (2011).
4. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-91 (2016).
5. Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
6. Pollard, S.M., Benchoua, A. & Lowell, S. Neural stem cells, neurons, and glia. *Methods Enzymol* **418**, 151-69 (2006).
7. Orlando, D.A. *et al.* Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Rep* **9**, 1163-70 (2014).
8. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-89 (2010).
9. Tippmann, S.C. *et al.* Chromatin measurements reveal contributions of synthesis and decay to steady-state mRNA levels. *Mol Syst Biol* **8**, 593 (2012).
10. Sidoli, S. & Garcia, B.A. Characterization of Individual Histone Posttranslational Modifications and Their Combinatorial Patterns by Mass Spectrometry-Based Proteomics Strategies. *Methods Mol Biol* **1528**, 121-148 (2017).
11. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **75**, 663-70 (2003).
12. Egertson, J.D. *et al.* Multiplexed MS/MS for improved data-independent acquisition. *Nat Methods* **10**, 744-6 (2013).
13. MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966-8 (2010).
14. Sidoli, S., Fujiwara, R. & Garcia, B.A. Multiplexed data independent acquisition (MSX-DIA) applied by high resolution mass spectrometry improves quantification quality for the analysis of histone peptides. *Proteomics* **16**, 2095-105 (2016).
15. Krueger, F. & Andrews, S.R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-2 (2011).

16. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-9 (2012).
17. Triche, T.J., Jr., Weisenberger, D.J., Van Den Berg, D., Laird, P.W. & Siegmund, K.D. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res* **41**, e90 (2013).
18. Fortin, J.P., Triche, T.J., Jr. & Hansen, K.D. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics* **33**, 558-560 (2017).
19. Zhou, W., Laird, P.W. & Shen, H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* **45**, e22 (2017).
20. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
21. Zhu, L.J. *et al.* ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237 (2010).
22. Yu, G. *et al.* GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* **26**, 976-8 (2010).
23. Yu, G., Wang, L.G., Han, Y. & He, Q.Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-7 (2012).
24. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-6 (2012).