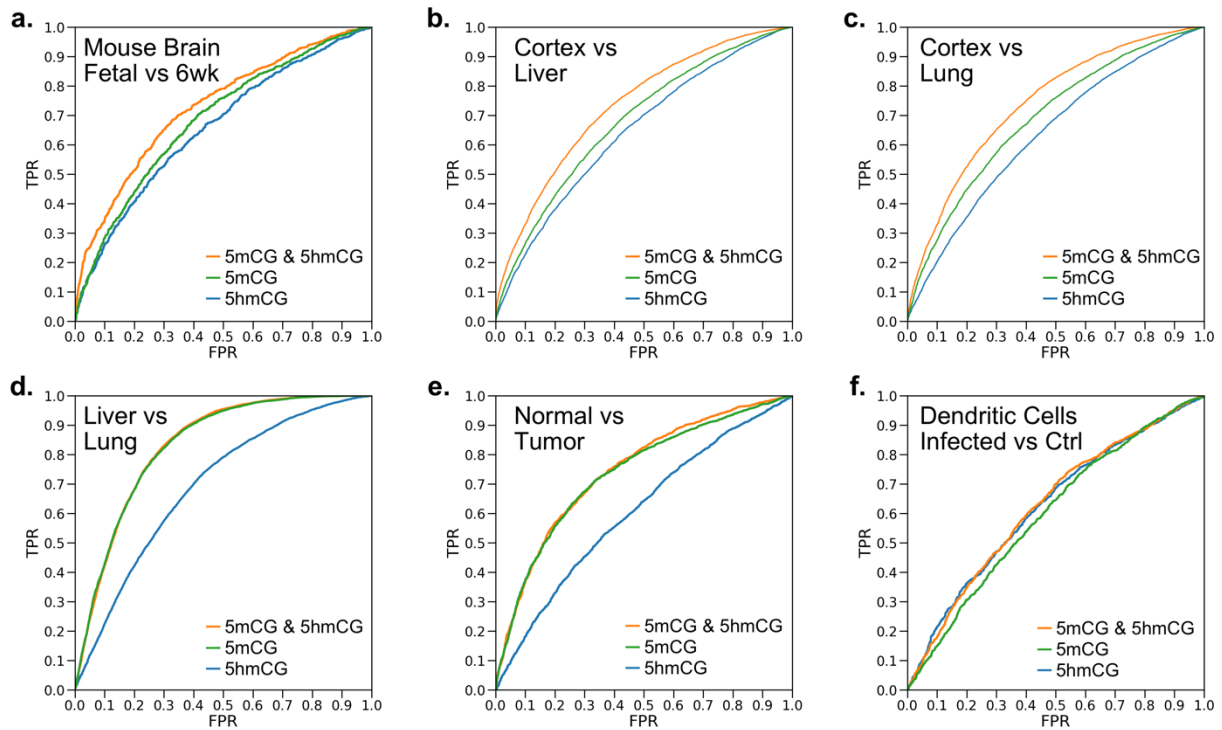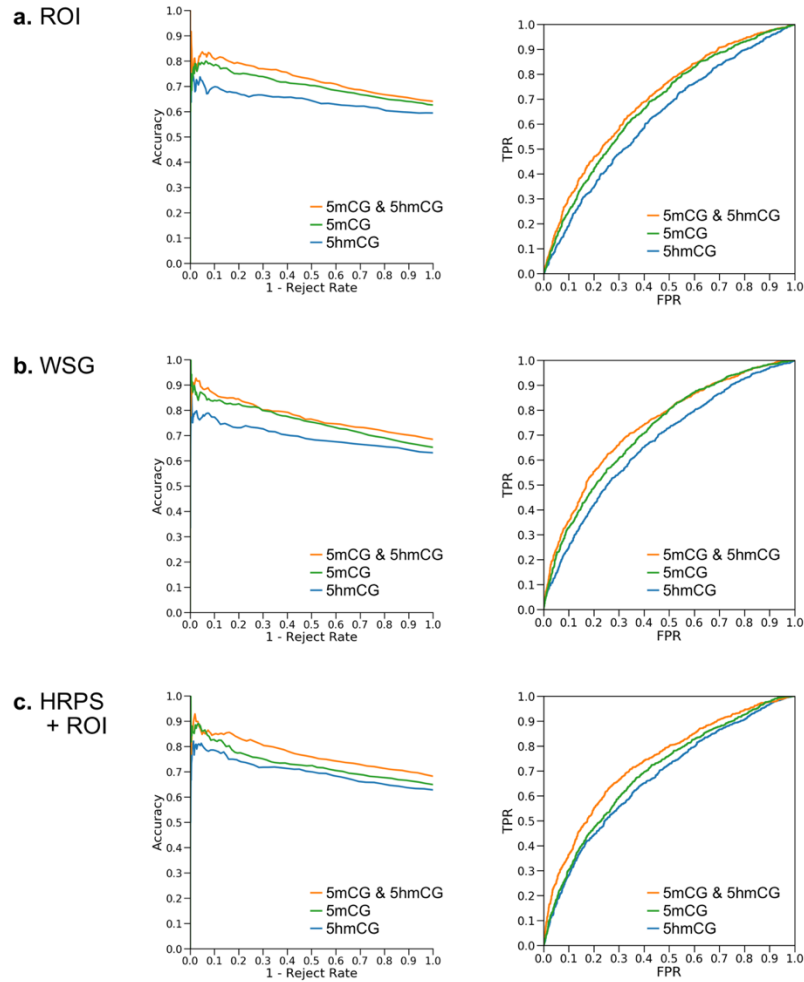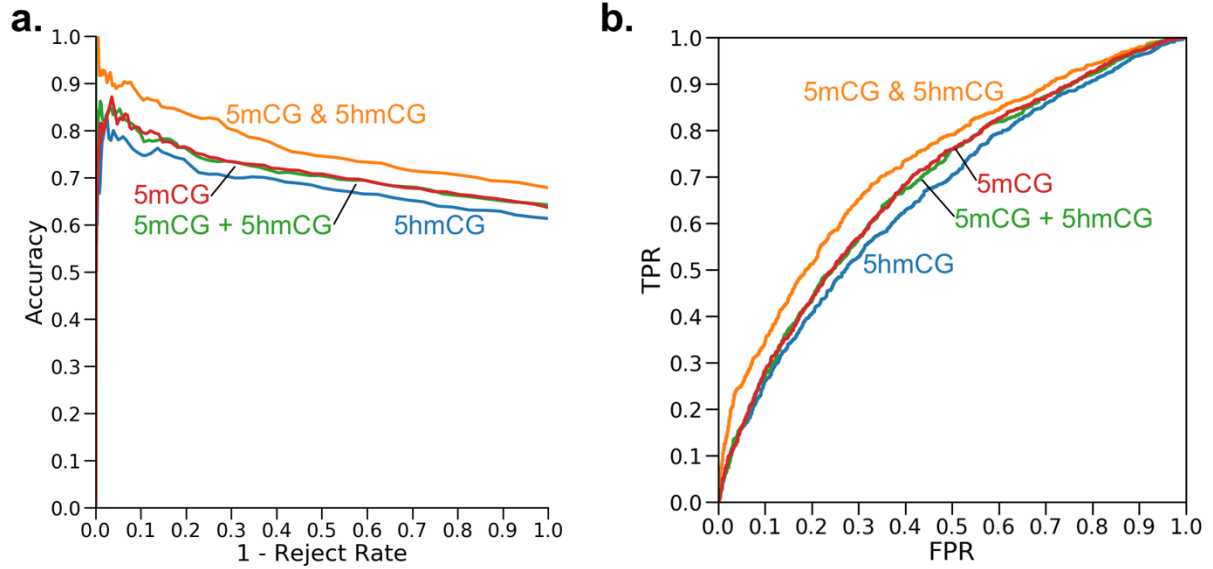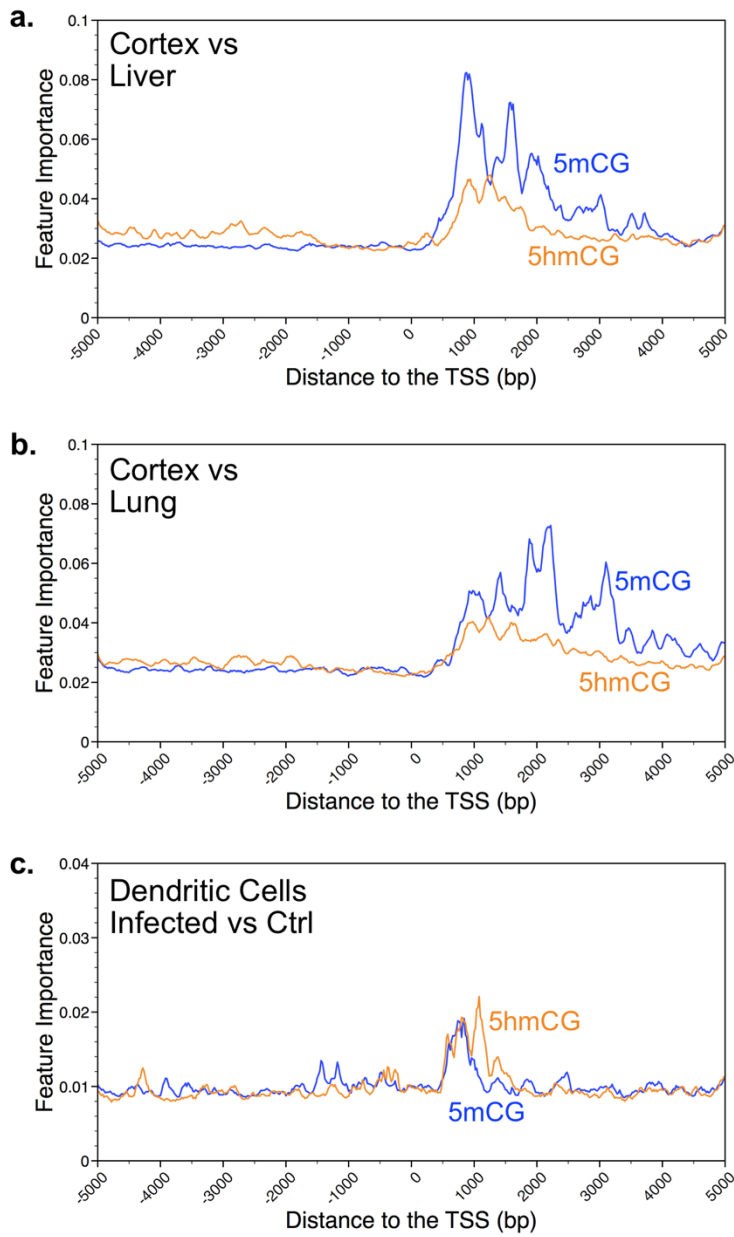# SUPPLEMENTARY FIGURES AND TABLES

Supplementary Figure S1. ROC analysis of ME-Class2 using the HRPS (high-resolution promoter signature) feature model for different 5mCG and 5hmCG datasets. ROC AUC are Mouse brain: 5mCG: 0.688, 5hmCG: 0.661, 5mCG & 5hmCG: 0.724; Human cortex versus liver: 5mCG: 0.679, 5hmCG: 0.644, 5mCG & 5hmCG: 0.732; Human cortex versus lung: 5mCG: 0.688, 5hmCG: 0.634, 5mCG & 5hmCG: 0.739; Dendritic cells: 5mCG: 0.597, 5hmCG: 0.622, 5mCG & 5hmCG: 0.625; Liver vs Lung: 5mCG: 0.829, 5hmCG: 0.692, 5mCG & 5hmCG: 0.830. Normal vs Tumor: 5mCG: 0.736, 5hmCG: 0.605, 5mCG & 5hmCG: 0.745.

Supplementary Figure S2. Performance of ME-Class2 using different feature models for fetal-6wk mouse brain 5mCG and 5hmCG data. ROI is region of interest; WSG is whole-scale gene; and HRPS+ROI is a combination of HRPS (high-resolution promoter signature) and ROI features. ROC AUC are ROI: 5mCG: 0.679, 5hmCG: 0.646, 5mCG & 5hmCG: 0.707; WSG: 5mCG: 0.729, 5hmCG: 0.663, 5mCG & 5hmCG: 0.742; HRPS + ROI: 5mCG: 0.699, 5hmCG: 0.671, 5mCG & 5hmCG: 0.737.

Supplementary Figure S3. Performance of different gene models using ME-Class2 5mCG and 5hmCG data from fetal and 6-week mouse brain as evaluated using (a) accuracy versus 1-reject rate (b) and ROC (receiver operating characteristic) curve analysis. 5mCG & 5hmCG indicates using 5mC and 5hmC data independently in the input feature vector. 5mCG + 5hmCG is the sum of 5mC and 5hmC signals used for the input feature vector. ROC AUC are 5mCG: 0.688, 5hmCG: 0.661, 5mCG & 5hmCG: 0.724, 5mCG + 5hmCG: 0.689.

Supplementary Figure S4. Feature importance for the ME-Class2 random forest classifier for HRPS model using 5mCG and 5hmCG data for human (a) cortex versus liver, (b) cortex versus lung, and (c) infected versus non-infected dendritic cell comparisons.
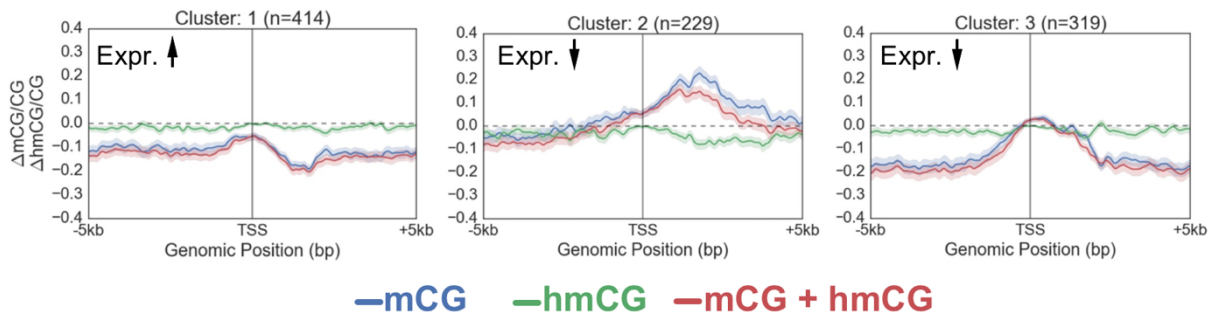
a. Liver-Lung

Cluster: 1 (n=1333)   Expr. ↓

Cluster: 2 (n=276)   Expr. ↑

b. Normal-Tumor

Cluster: 1 (n=414)   Expr. ↑

Cluster: 2 (n=229)   Expr. ↓

Cluster: 3 (n=319)   Expr. ↓

—mCG   —hmCG   —mCG + hmCG
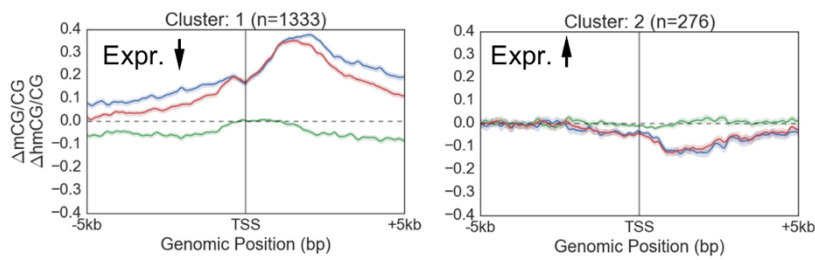
Supplementary Figure S5. Metagene plots for major clusters whose methylation signature is predictive of expression observed in (a) human lung versus liver and (b) paired human lung and liver normal versus tumor comparisons. Differential 5mCG and 5hmCG is lung – liver or tumor – normal. Up arrows indicate genes in the cluster increased in expression in lung relative to normal or tumor versus normal. Down arrows indicate the reverse. Shading indicates the 68% bootstrapped confidence interval.
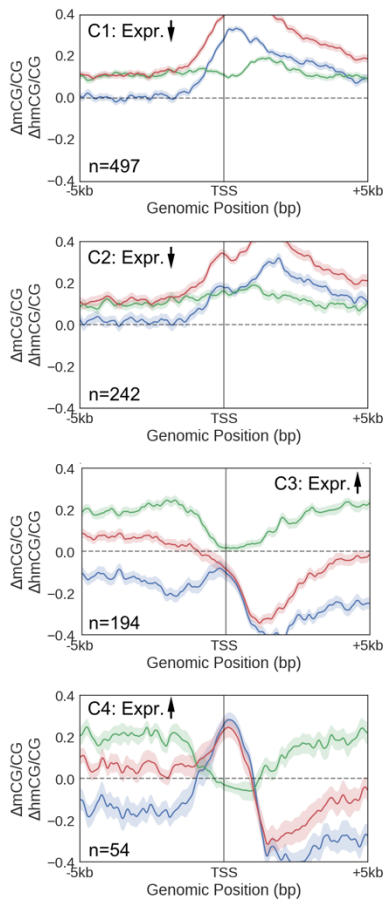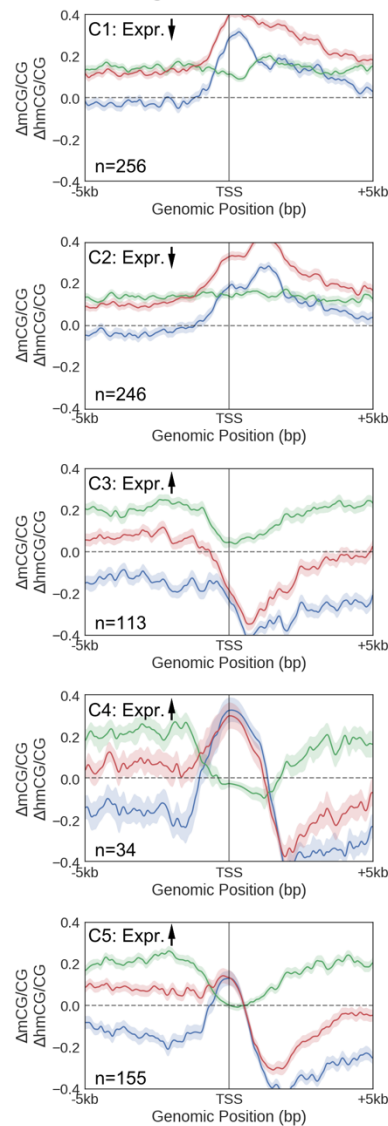
Supplementary Figure S6. Metagene plots for major clusters whose methylation signature is predictive of expression observed in (a) human cortex versus liver and (b) human cortex versus lung comparisons. Differential 5mCG and 5hmCG is brain – liver or brain – lung. Up arrows indicate genes in the cluster increased in expression in cortex relative to lung or cortex relative to lung. Down arrows indicate the reverse. Shading indicates the 68% bootstrapped confidence interval
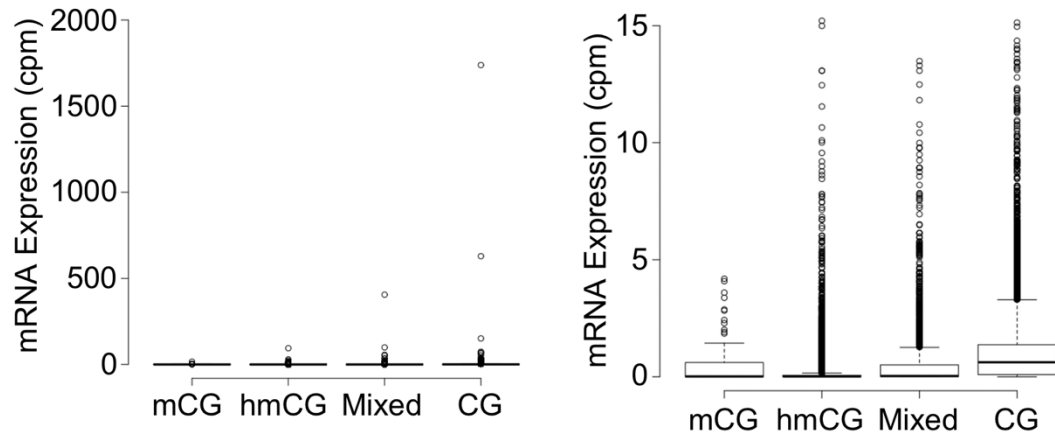
Supplementary Figure S7. Original plot and an in between zoom for plot from Figure 4g. mRNA expression in granule cells of genes whose promoters (defined as [-800bp, +2kb] around the TSS) are greater than 50% marked by mCG, hmCG, a combination of mCG and hmCG, or greater than 50% unmethylated, CG. Outliers have been cropped for clarity.
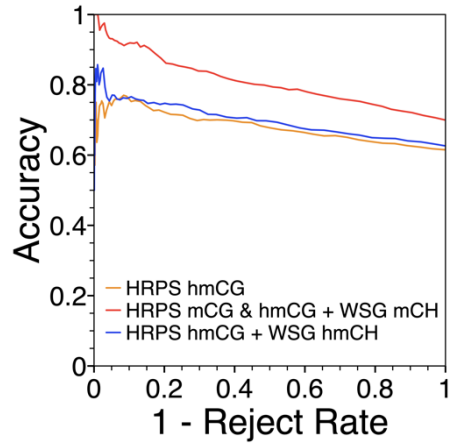
Supplementary Figure S8. Performance of different gene models using ME-Class2 5mC and 5hmC data from fetal and 6-week mouse brain as evaluated using accuracy versus 1-reject rate.

Supplementary Table S1. MLML results, including coverage statistics for all datasets included in the analysis. CpG Coverage is accepted CpGs / total genomic CpGs.

Study	Design			MLML Statistics			
	5hmC assay	Organism	Sample	Overshoot	Conflicts	Accepted CpGs	CpG Coverage
Li et al. *Genome Res* 2016	oxBS-seq	Human	Liver_N1	7483717	45698	23373853	0.828
	oxBS-seq	Human	Liver_N2	7779952	35689	23281047	0.825
	oxBS-seq	Human	Liver_N3	8830879	63528	23395102	0.829
	oxBS-seq	Human	Liver_T1	8706959	56985	23027265	0.816
	oxBS-seq	Human	Liver_T2	9048391	63705	24285424	0.861
	oxBS-seq	Human	Liver_T3	9388948	71465	23380024	0.829
	oxBS-seq	Human	Lung_N1	9487853	64312	24490004	0.868
	oxBS-seq	Human	Lung_N2	8515568	53325	23425058	0.830
	oxBS-seq	Human	Lung_N3	8184630	28879	23159840	0.821
	oxBS-seq	Human	Lung_T1	10236654	94104	24675947	0.874
	oxBS-seq	Human	Lung_T2	9656229	84745	23617849	0.837
	oxBS-seq	Human	Lung_T3	9529502	47862	23186713	0.822
Lister et al. *Science* 2013	TAB-seq	Mouse	Fetal_Brain	250772	1	19634599	0.920
	TAB-seq	Mouse	6wk_Brain	242633	5	19875929	0.931
Pacis et al. *Genome Res* 2017	TAB-seq	Human	Ctrl Dendritic Cells	180799	10	23475529	0.832
	TAB-seq	Human	Inf. Dendritic Cells	177185	3	23505262	0.833
Mellen et al. *PNAS* 2017	oxBS-seq	Mouse	Granule Cells 1	2779858	16289	12733479	0.597
	oxBS-seq	Mouse	Granule Cells 2	4752111	153290	18900922	0.886
	oxBS-seq	Mouse	Granule Cells 3	1310177	295330	6954726	0.326

Supplementary Table S2. Numbers of differentially expressed genes for all datasets included in the analysis.

Organism	Sample 1	Sample 2	Genes (FC<=-2,FC>=2)	Genes (FC<=-2)	Genes (FC>=2)	Genes Down (%)	Genes Up (%)
Human	Liver_N1	Lung_N1	3456	1554	1902	45.0%	55.0%
Human	Liver_N1	Lung_N2	3677	1567	2110	42.6%	57.4%
Human	Liver_N1	Lung_N3	3562	1622	1940	45.5%	54.5%
Human	Liver_N2	Lung_N1	2771	1242	1529	44.8%	55.2%
Human	Liver_N2	Lung_N2	3275	1494	1781	45.6%	54.4%
Human	Liver_N2	Lung_N3	3077	1541	1536	50.1%	49.9%
Human	Liver_N3	Lung_N1	2179	1150	1029	52.8%	47.2%
Human	Liver_N3	Lung_N2	2542	1224	1318	48.2%	51.8%
Human	Liver_N3	Lung_N3	2544	1339	1205	52.6%	47.4%
Human	Cortex	Liver_N1	6370	2949	3421	46.3%	53.7%
Human	Cortex	Liver_N2	6292	2782	3510	44.2%	55.8%
Human	Cortex	Liver_N3	6374	2682	3692	42.1%	57.9%
Human	Cortex	Lung_N1	5955	2439	3516	41.0%	59.0%
Human	Cortex	Lung_N2	5896	2400	3496	40.7%	59.3%
Human	Cortex	Lung_N3	5864	2465	3399	42.0%	58.0%
Human	Liver_N1	Liver_T1	1855	1061	794	57.2%	42.8%
Human	Liver_N2	Liver_T2	2550	1926	624	75.5%	24.5%
Human	Lung_N1	Lung_T1	2196	1270	926	57.8%	42.2%
Human	Lung_N2	Lung_T2	403	250	153	62.0%	38.0%
Human	Lung_N3	Lung_T3	2767	1559	1208	56.3%	43.7%
Mouse	Fetal_Brain	6wk_Brain	3654	2063	1591	56.5%	43.5%
Human	Ctrl Dendritic Cells	Inf. Dendritic Cells	3933	2115	1818	53.8%	46.2%

Supplementary Table S3. Correctly predicted genes using a cutoff of accuracy=90% for infected versus control dendritic cells. Expression fold change is negative for down-regulated genes and positive for up-regulated genes.

	RefSeq	Expression Fold Change	Gene Name
Down-regulated	NM_030806	-3.1	C1orf21
	NM_014858	-9.4	TMCC2
	NM_002051	-4.4	GATA3
	NM_005933	-4.1	MLL
	NM_001759	-2.7	CCND2
	NM_020403	-16.3	PCDH9
	NM_015559	-2.1	SETBP1
	NM_198461	-66.8	LONRF2
	NM_012120	-2.3	CD2AP
	NM_005375	-2.9	MYB
	NM_001259	-6.2	CDK6
	NM_001394	-190.8	DUSP4
	NM_004789	-5.1	LHX2
	NM_006603	-2.4	STAG2
Up-regulated	NM_000913	4.8	OPRL1

Supplementary Table S4. Results from DAVID analysis of genes identified in each cluster in Fig. 4. Enrichment results for brain development and neuronal function related categories are shown. Genes may be found in multiple categories. UP_KEYWORDS are UniprotKB keywords.

Gene Cluster	DAVID Ontology Category	Bonf. Adjusted p-value	Count	Fold Enrichment	RefSeq ID	Gene Name	Gene Description
C1	Brain development, GO:0007420	0.71	6	7.1	NM_013627	Pax6	Paired box 6
					NM_008130	Gli3	GLI-Kruppel family member
					NM_010710	Lhx2	LIM Homeobox protein 2
					NM_007552	Bmi1	Bmi1 polycomb ring finger oncogene
					NM_152229	Nr2e1	Nuclear receptor subfamily 2, group E, member 1
					NM_016701	Nes	Nestin
	Negative regulatation of neuronal differentiation, GO:0045665	0.0092	6	20.1	NM_013627	Pax6	Paired box 6
					NM_008130	Gli3	GLI-Kruppel family member
					NM_152229	Nr2e1	Nuclear receptor subfamily 2, group E, member 1
					NM_008714	Notch1	Notch 1
					NM_008783	Pbx1	Pre B cell leukemia homeobox 1
					NM_007865	Dll1	Delta like canonical Notch ligand 1
	Neuronal stem cell population maintenance, GO:0097150	0.094	4	40.7	NM_009951	Igf2bp1	Insulin-like growth factor 2 mRNA binding protein 1
					NM_008714	Notch1	Notch 1
					NM_007865	Dll1	Delta like canonical Notch ligand 1
					NM_007664	Cdh2	Cadherin 2
C2	Synapse, UP_KEYWORDS	0.59	4	8.8	NM_001009818	Sept11	septin 11
					NM_053167	Trim9	tripartite motif-containing 9
					NM_021604	Agrn	agrin
					NM_001081370	Shank2	SH3 and multiple ankyrin repeat domains 2
C3	Synapse, UP_KEYWORDS	0.007	8	7.3	NM_183023	Rims4	regulating synaptic membrane exocytosis 4
					NM_027712	Dlgap1	DLG associated protein 1
					NM_019697	Kcnd2	potassium voltage-gated channel, Shal-related family, member 2
					NM_022030	Sv2a	synaptic vesicle glycoprotein 2 a
					NM_020252	Nrxn1	neurexin I (Nrxn1), transcript variant 1
					NM_016758	Rgs14	regulator of G-protein signaling 14
					NM_011861	Pacsin1	protein kinase C and casein kinase substrate in neurons 1
					NM_008169	Grin1	glutamate receptor, ionotropic, NMDA1 (zeta 1)