2. dpFeature

We also compare with dpFeature. By applying dpFeature to gene expression profiles, we identified 13775 genes and 13362 associated with adjusted P-vales less than 0.01 for human and mouse, respectively. Thus, as in the case of bimodal genes approach, dpFeature does not have enough ability to screen reasonable (restricted) number of genes. In order to further investigate biological reliability of genes selected by dpFeature, top ranked 200 genes (those with smaller P-values) were selected intentionally. Since there are 76 genes selected commonly between human and mouse, dpFeature could identify common genes between two species. As for common genes with those selected by PCA based unsupervised FE, there are 51 and 44 genes commonly selected with those PCA based unsupervised FE for human and mouse, respectively. Thus, as for the consistency with PCA based unsupervised FE, dpFeature is better than highly variable genes approach and bimodal genes approach.

We also upload these genes to Enrichr in order to validate selected genes biologically.

Table S8 Enrichment analysis by Enrichr, "MGI Mammalian Phenotype 2017", of 200 genes selected by dpFeature for mouse. Top 5 ranked terms.

Term	Overlap	P-value	Adjusted P-value
MP:0001262_decreased_body_weight	34/1189	2.71E-08	1.84E-05
MP:0002083_premature_death	25/834	9.89E-07	2.68E-04
MP:0001463_abnormal_spatial_learning	12/162	8.71E-08	3.94E-05
MP:0011086_postnatal_lethality,_incomplete_penetrance	18/563	1.52E-05	2.95E-03
MP:0000807_abnormal_hippocampus_morphology	10/86	1.46E-08	1.84E-05

Table S8 shows the results for "MGI Mammalian Phenotype 2017", of 200 genes selected by dpFeature for mouse. Among top ranked five terms, only the fifth term is related to brain.

Table S9 Enrichment analysis by Enrichr, "Allen Brain Atlas down", of 200 genes selected by dpFeature for human. Top 5 ranked terms

Term	Overlap	P-value	Adjusted P-value
Primary somatosensory area, trunk, layer 1	18/300	1.42E-09	2.76E-06
Main olfactory bulb, glomerular layer	15/300	3.75E-07	1.46E-04
layer 1 of FCx	15/300	3.75E-07	1.46E-04
Primary motor area, Layer 1	14/300	2.10E-06	3.71E-04
Entorhinal area, lateral part, layer 1	15/300	3.75E-07	1.46E-04

Table S10 Enrichment analysis by Enrichr, "Allen Brain Atlas down", of 200 genes selected by dpFeature for mouse. Top 5 ranked terms.

Term	Overlap	P-value	Adjusted P-value
layer 1 of FCx	19/300	1.95E-10	3.65E-07
layer 1 of ERCx	17/300	9.73E-09	9.12E-06
Primary motor area	15/300	3.75E-07	2.34E-04
dorsal pallium/isocortex	14/300	1.09E-05	1.20E-03
Cortical amygdalar area, posterior part, lateral zone, layer 1	14/300	2.10E-06	6.55E-04

Tables S9 and S10 show the results for "Allen Brain Atlas down", of 200 genes selected by dpFeature for human and mouse, respectively. Although there are many significant terms, these are far less

significant than those for genes selected by PCA based unsupervised FE (Tables 2 and 3).

Table S11 Enrichment analysis by Enrichr, "GTEx Tissue Sample Gene Expression Profiles down", of 200 genes selected by dpFeature gene for human. Top 5 ranked terms.

Term	Overlap	P-value	Adjusted P-value
GTEX-WH7G-0002-SM-4LVN9_blood_male_40-49_years	71/3070	1.51E-12	3.99E-09
GTEX-UPIC-0002-SM-3NMDC_blood_female_20-29_years	69/3033	8.04E-12	1.06E-08
GTEX-XUYS-0005-SM-47JZ2_blood_male_50-59_years	69/3115	2.78E-11	2.44E-08
GTEX-XUJ4-0004-SM-4BOQE_blood_female_60-69_years	70/3348	2.63E-10	1.29E-07
GTEX-SNOS-0003-SM-3NMAO_blood_male_40-49_years	67/3123	2.57E-10	1.29E-07

Table S12 Enrichment analysis by Enrichr, "GTEx Tissue Sample Gene Expression Profiles down", of 200 genes selected by dpFeature genes for mouse. Top 5 ranked terms

Term	Overlap	P-value	Adjusted P-value
GTEX-X5EB-0526-SM-46MVP_pancreas_male_40-49_years	31/1139	3.59E-07	7.22E-04
GTEX-R55E-0726-SM-48FCZ_testis_male_20-29_years	22/647	5.92E-07	7.22E-04
GTEX-WFON-0001-SM-3P61W_blood_male_40-49_years	52/2647	9.14E-07	7.26E-04
GTEX-WH7G-0002-SM-4LVN9_blood_male_40-49_years	57/3070	1.43E-06	7.26E-04
GTEX-S7SE-0005-SM-2XCEA_blood_male_50-59_years	55/2919	1.49E-06	7.26E-04

Tables S11 and S12 show the results for "GTEx Tissue Sample Gene Expression Profiles down", of 200 genes selected by dpFeature for human and mouse, respectively. Although there are many significant terms, these are no terms related to brains while while almost all terms are related to brains when PCA based unsupervised FE was use (Tables 4 and 5). Thus, it is obvious that genes selected by dpFeature are biologically less reliable than genes selected by PCA based unsupervised FE.

Table S13 Enrichment of selected genes by dpFeature in Embryonic brain by "Jensen TISSUES" by Enrichr.

Term	Overlap	P-value	Adjusted P-value
Human			
Embryonic_brain	110/4936	3.32E-20	1.14E-18
Mouse			
Embryonic_brain	122/4936	3.32E-20	1.14E-18

On the other hand, since Embryonic brain by "Jensen TISSUES" is highly significant (Table S13), dpFature is, at least, better than highly variable genes. In contrast to this, genes selected by dpFeatire are unlikely regulated by limited number of TFs.

Table S14 TF enrichment enriched in "ENCODE and ChEA Consensus TFs from ChIP-X" by Enrichr for human and mouse genes selected by dpFature. Bold TFs are common. Genes in red are also commonly selected between human and mouse in Table 7.

human	ATF2, BRCA1, **CEBPD**, CHD1, CREB1, E2F4, **FOXM1**, KAT2A, **KLF4**, **MYC**, NANOG, **NELFE**, NFI, NFYA, PML, RELA, REST, SALL4, SIN3A, SOX2, TAF7, **TCF3**, TRIM28, **ZMIZ1**

mouse	**CEBPD**, **FOXM1**, **KLF4**, **MYC**, **NELFE**, **TCF3**, **ZMIZ1**

As can be seen in Table S14, identified TFs are smaller than TFs selected by PCA based unsupervised FE (Table 7) or bimodal genes approach (Table S7).