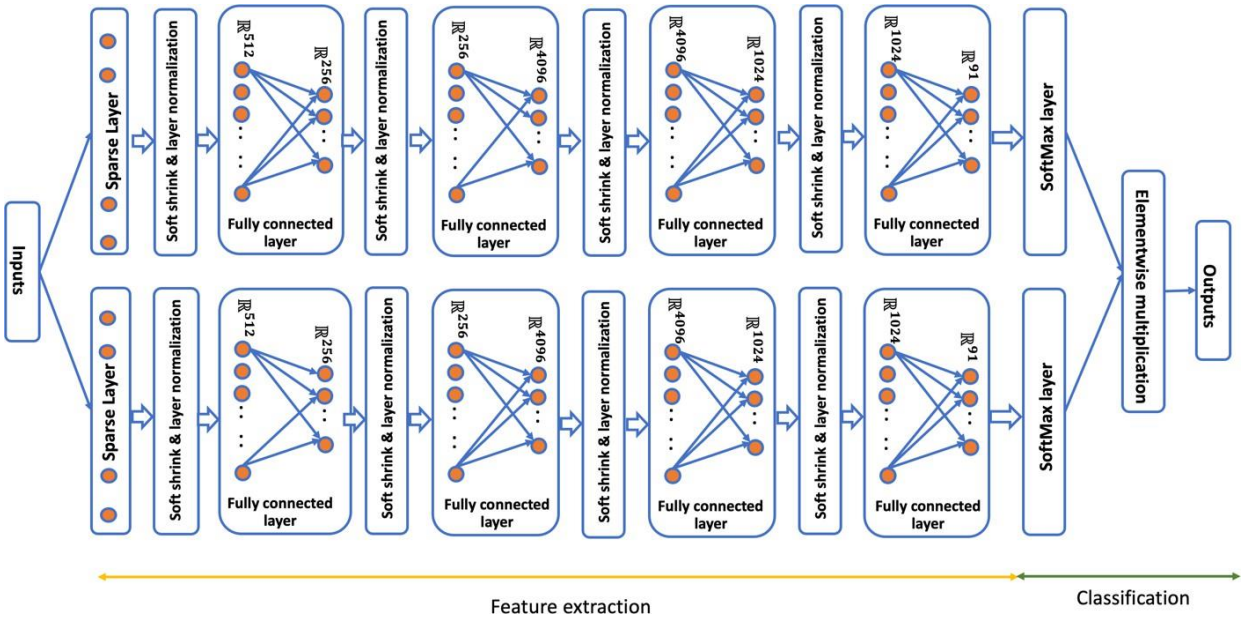
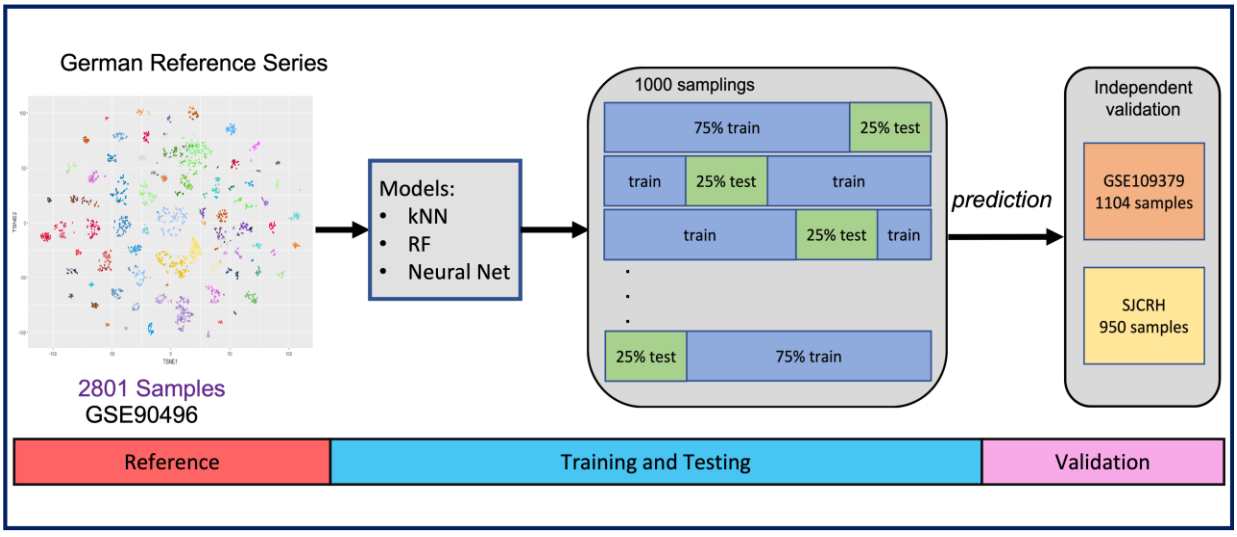
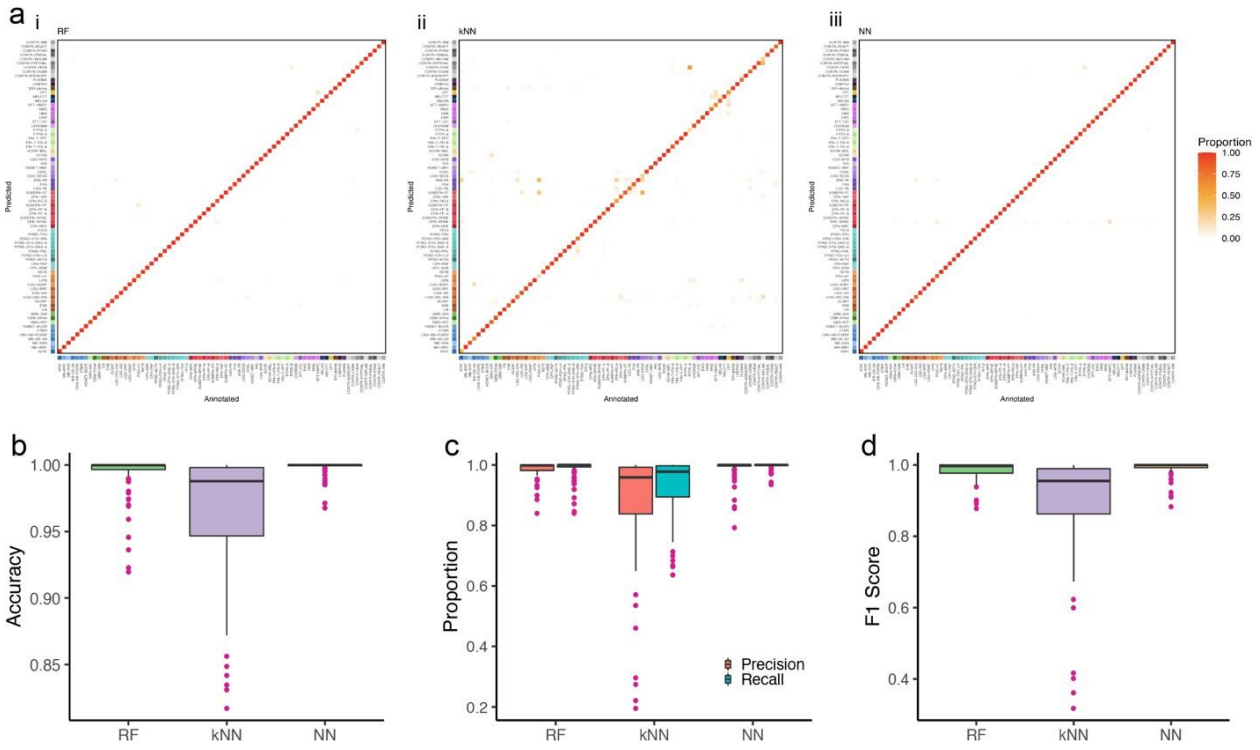


Supplementary Figure 1. Architecture of the multi-layer perceptron neural network. We constructed an 11-layer perceptron neural net. The input dimension is 51,108, composed of probes selected with feature extraction described in the Methods section. The first layer is sparse, while the remaining ten layers are fully connected. The sparse layer maps 139,264 uniformly random sets of 256 features to the space $[0,1]^{512}$. This layer is a forest of random decision stumps that computes cosines of angles of vectors of length 256 drawn with uniform probability without replacement from the 51,108 input probes, which were selected by a LASSO model. The output of this layer is then fed through a standard perceptron.

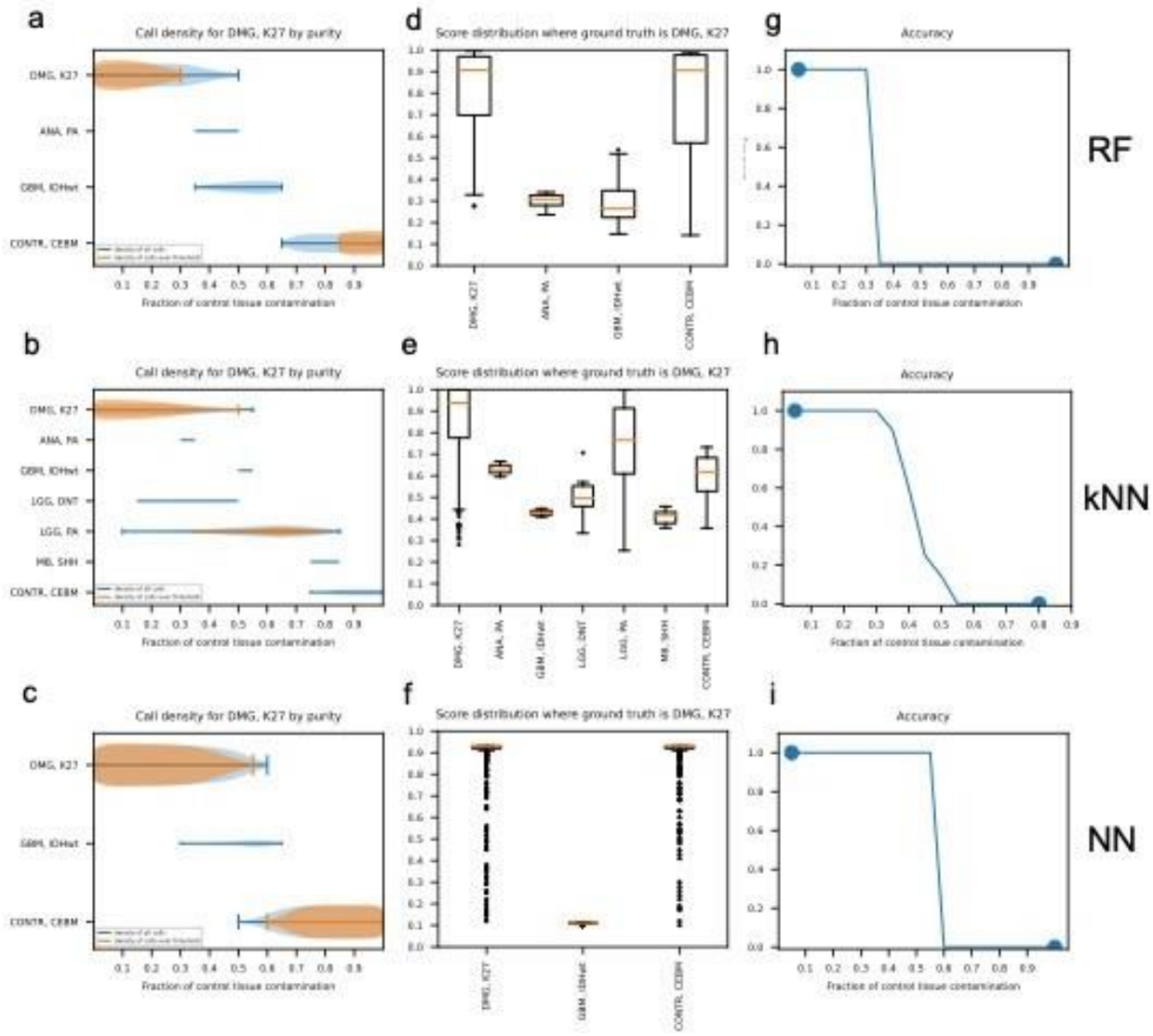


Supplementary Figure 2. Training, testing, and validation scheme of all classifiers. To reduce the overfitting problem when training classifiers on high-dimensional data, all classifiers were cross-validated based

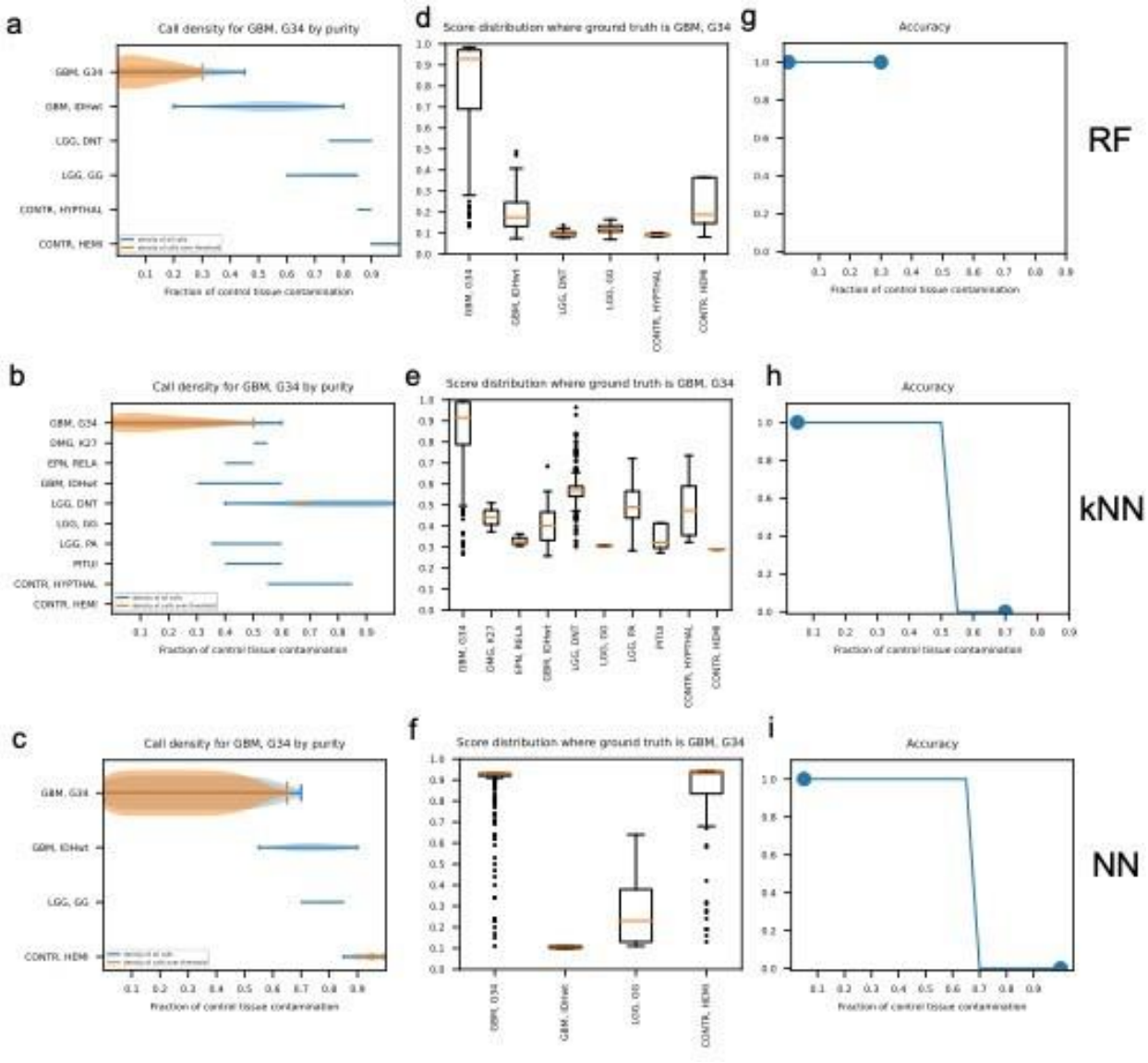
on 1000 leave-out-25% samplings. We randomly selected 75% of the data used to train the classifiers (GSE90496), while the remaining 25% of the data were used for predictions. Stratified random sampling was performed for each methylation class or family to ensure the number of categories remained the same in each iteration. This training and testing process was repeated 1000 times. The final models were validated with two independent data sets: GSE109379 and the St. Jude data set.



Supplementary Figure 3. Leave-out-25% testing results of each methylation family. **(A)** Heat map showing methylation family prediction results after 1000 stratified random samplings ***i*** RF, ***ii*** kNN, and ***iii*** NN classifier incorporating information of n = 2,801 reference tumor samples allocated to 75 methylation class families (GSE90496). Deviations from the bisecting line represent misclassification errors (using the maximum calibrated score for class prediction). Boxplots showing **b** the accuracy, **c** precision and recall, and **d** F1-score for each classifier with outliers.



Supplementary Figure 4. Classification results of RF, kNN, and NN model for high-grade diffused midline glioblastoma with K-27 mutant (DMG, K27) methylation family at different contamination levels. a,b,c Density plots of all calls (blue curve) and calls over the 0.9 clinical threshold (orange curve) at each possible methylation family predicted by RF, kNN, and NN when the ground truth is DMG, K27 at different fractions of control tissue contamination. **d,e,f** Box plots show the score distribution for each methylation family predicted by RF, kNN, and NN models. **g,h,i** Accuracy of each classifier at each purity level.



Supplementary Figure 5. **Classification results of RF, kNN, and NN model for grade 4 glioblastoma, IDH wildtype, H3.3 G34 mutant (GBM, G34) methylation family at different levels of contamination. a, b, c** Density plots of all calls (blue curve) and calls over the 0.9 clinical threshold (orange curve) at each possible methylation family predicted by RF, kNN, and NN when the ground truth is GBM, G34 at different fractions of control tissue contamination. **d, e, f** Box plots show the score distribution for each methylation family predicted by RF, kNN, and NN models. **g, h, i** Accuracy of each classifier at each purity level.

Supplementary Table 1. Control tissues used for the *in silico* mixing experiment

Methylation class/family	Mixing control
IDH-glioma, GBM-IDHwt, CPT, CN, CNS_NB_FOXR2, CPH_ADM, CPH_PAP, EPN_RELA, EPN_YAP, GBM_G34, HGNET_MN1, IHG, LGG_DIG_DIA, LGG_DNT, LGG_GG, LGG_MYB, LGG_SEGA, PGG_nC, PXA, SUBEPN_ST	HEMI
LGG-PA, ANA_PA, MB-SHH, MB-G3/G4, MB- G4/G4, MB-WNT, DMG_K27, EPN_PF_A, EPN_PF_B, HMB, LGG_RGNT, SUBEPN_PF	CBEM
ATRT, DLGNT, EFT_CIC, ETMR, EWS, HGNET_BCOR, MNG	both
ENB, PIN_T_PB_A, PIN_T_PB_B, PIN_T_PPT, PITAD_ACTH, PITAD_FSH_LH, PITAD_PRL, PITAD_STH_DNS_A, PITAD_STH_DNS_B, PITAD_STH_SPA, PITAD_TSH, PTPR_A, PTPR_B, SCHW, SCHW_MEL, MELAN, MELCYT, CHGL, CHORDM, EPN_MPE, EPN_SPINE, LIPN, LYMPHO, PITUI, PLASMA, RETB, SFT_HMPC, SUBEPN_SPINE	none

Supplementary Table 2. Validation results of the silico data using independent samples that are not in the GSE10980 or SJCRH data sets

Case	Ground truth	VAF	kNN_label	kNN_score	RF_label	RF_score	NN_label	NN_score
DMG_case1	H3F3A p.K28M	0.48	DMG, K27	0.85	DMG, K27	0.84	DMG, K27	0.93
DMG_case2	H3F3A p.K28M	0.4	DMG, K27	1	DMG, K27	0.94	DMG, K27	0.93
DMG_case3	H3F3A p.K28M	0.23	DMG, K27	0.9	DMG, K27	0.33	DMG, K27	0.93
AIDH_case1	IDH1 p.R132G	0.38	A IDH	0.68	A IDH	0.87	A IDH	0.94
AIDH_case2	IDH1 p.R132H	0.33	A IDH	0.56	A IDH	0.5	A IDH	0.94
AIDH_case3	IDH1 p.R132H	0.19	LGG, DNT	0.27	CONTR, HEMI	0.08	A IDH	0.93

Description of Additional Supplementary Files:

Supplementary Data 1. Performance metrics at the methylation class level from 3 classifiers (kNN, NN, and RF) for the GSE90496.

Supplementary Data 2. Performance metrics at the methylation family level from 3 classifiers (kNN, NN, and RF) for the GSE90496

Supplementary Data 3. Semi-supervised analysis results for GSE109379 data set.

Supplementary Data 4. Semi-supervised analysis results for SJCRH data set.

Supplementary Data 5. Performance metrics at the methylation class level from 3 classifiers (kNN, NN, and RF) for the validation GSE109379 data set.

Supplementary Data 6. Performance metrics at the methylation family level from 3 classifiers (kNN, NN, and RF) for the validation GSE109379 data set.

Supplementary Data 7. Performance metrics at the methylation class level from 3 classifiers (kNN, NN, and RF) for the validation SJCRH data set.

Supplementary Data 8. Performance metrics at the methylation family level from 3 classifiers (kNN, NN, and RF) for the validation SJCRH data set

Supplementary Data 9. Clinical information for samples from SJCRH.