

Table S2: Gene-based vs. FAIME-based classification of HNSCC tumor vs. non-tumor control samples. *Oncogenic FAIME Features of HNSCC (Methods)* were derived from datasets A, B and C (**Figure 3**), and tested in datasets D and E for its ability to discriminate between tumor vs. non-tumor control tissue. As shown below, FAIME-based classification required less features than Gene-based classification, and obtained equivalent or better accuracy scores. Note that the accuracies reported in the top rows under "Gene-based analysis" are extracted from the original manuscripts cited in the Reference row.

Dataset ID	A	B	C	D	E
Reference	GSE6631 [61]	GSE2379 [62]	E-MEXP-44 [63]	E-MEXP-44 [63]	JCO2010 [47]
Gene-based analysis					
Computational method	H-clustering	PAM; cross-validation	PCA; H-clustering	PCA; H-clustering	Spectral clustering
# probe features	42	121	338	338	1000
Accuracy for tumor	100%	94%	100%	100%	99%
Accuracy for non-tumor control	59%	100%	74%	74%	86%
<i>Overall Accuracy</i>	71%	95%	90%	90%	96%
<hr/>					
FAIME analysis					
# mechanism features				57	57
Accuracy for tumor				100%	97%
Accuracy for non-tumor control				100%	86%
	Training sets				
<i>Overall Accuracy</i>				100%	96%

Legend: PAM: Partitioning Around Medoids clustering; PCA: Principal component analysis; H-clustering: Hierarchical clustering.