Supplementary Materials: Transcriptomic Analysis for Prognostic Value in Head and Neck Squamous Cell Carcinoma

Supplementary Table 1:
Overexpression of 10 genes with Cox's hazard ratio > 1.5.

Supplementary Table 2:
Overexpression of 10 genes with Cox's hazard ratio < 0.6.

Table 1: The top 10 genes overexpressed with poor prognosis in HNSCC (ranked by adjusted P value)

Gene ID	Gene Description	Kaplan-Meier survival		Univariate		Multivariate	
		FDR P value	Bonferroni P value	HR*	95% CI	HR*	95% CI
DKK1	dickkopf WNT signaling pathway inhibitor 1	3.8×10^{-6}	0.001	2.266	1.666-3.082	2.135	1.559-2.924
CAMK2N1	calcium/calmodulin-dependent protein kinase II inhibitor 1	1.5×10^{-5}	0.002	2.101	1.572-2.809	2.007	1.490-2.704
STC2	stanniocalcin 2	1.5×10^{-5}	0.004	2.147	1.578-2.921	2.075	1.515-2.843
PGK1	phosphoglycerate kinase 1	2.4×10^{-5}	0.006	2.127	1.563-2.895	2.046	1.498-2.795
SURF4	surfeit 4	6.2×10^{-5}	0.006	2.055	1.531-2.757	2.089	1.543-2.829
USP10	ubiquitin specific peptidase 10	7.9×10^{-5}	0.012	2.083	1.532-2.834	2.119	1.551-2.895
NDFIP1	Nedd4 family interacting protein 1	1.1×10^{-4}	0.017	2.031	1.502-2.746	2.027	1.483-2.771
FOXA2	forkhead box A2	1.6×10^{-4}	0.018	1.976	1.479-2.640	1.914	1.426-2.569
STIP1	stress-induced-phosphoprotein 1	1.8×10^{-4}	0.029	1.958	1.463-2.621	1.957	1.451-2.640
DKC1	dyskeratosis congenita 1, dyskerin	2.8×10^{-4}	0.042	2.046	1.490-2.808	1.837	1.332-2.534

Selection criteria:
Kaplan-Meier Bonferroni-adjusted $P < 0.05$
Cox's univariate and multivariate $HR > 1.5$
(* Cox's model: $P < 0.001$)

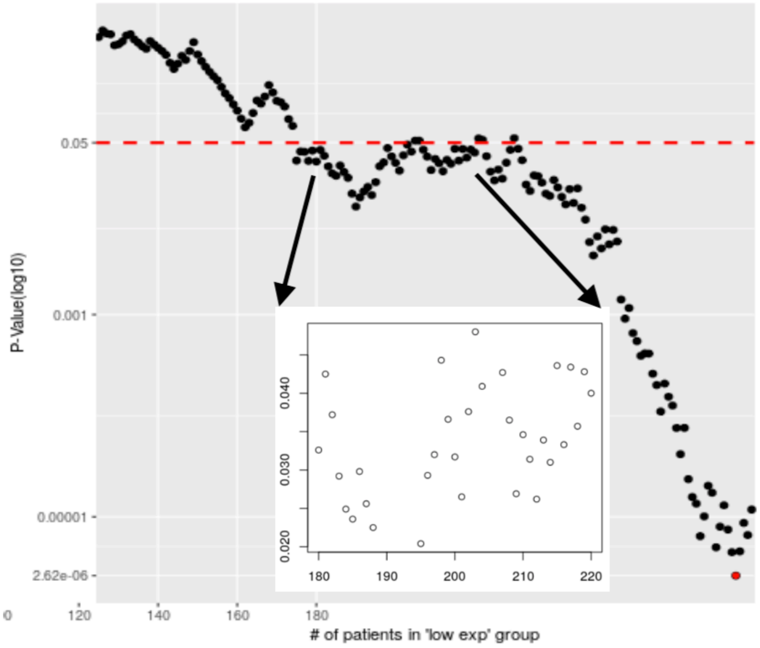
Table 2: The other top 10 genes overexpressed with better prognosis in HNSCC (ranked by adjusted P value)

Gene ID	Gene Description	Kaplan-Meier survival		Univariate		Multivariate	
		FDR P value	Bonferroni P value	HR*	95% CI	HR*	95% CI
ZNF557	zinc finger protein 557	4.7×10^{-6}	0.001	0.465	0.348-0.619	0.499	0.372-0.669
ZNF266	zinc finger protein 266	5.2×10^{-6}	0.001	0.474	0.355-0.632	0.453	0.338-0.607
IL19	interleukin 19	6.5×10^{-6}	0.002	0.472	0.351-0.635	0.459	0.340-0.619
MYO1H	myosin 1H	1.4×10^{-5}	0.003	0.468	0.347-0.632	0.467	0.344-0.634
FCGBP	Fc fragment of IgG binding protein	4.8×10^{-5}	0.008	0.484	0.359-0.653	0.496	0.366-0.674
EVPLL	envoplakin-like protein	7.5×10^{-5}	0.013	0.490	0.363-0.661	0.494	0.364-0.672
PNMA5	paraneoplastic antigen like 5	3.0×10^{-4}	0.017	0.499	0.371-0.671	0.481	0.357-0.650
IQCN	IQ motif containing N	1.5×10^{-4}	0.020	0.500	0.371-0.673	0.483	0.356-0.654
NPB	neuropeptide B	2.6×10^{-4}	0.027	0.460	0.328-0.646	0.457	0.324-0.646
CALML5	calmodulin like 5	2.0×10^{-4}	0.039	0.510	0.379-0.686	0.493	0.364-0.667

Selection criteria:
Kaplan-Meier Bonferroni-adjusted P value < 0.05
Cox's univariate and multivariate HR < 0.6
(* Cox's model: $P < 0.001$)

Supplementary figure S1:

The gene NDFIP1, one of our 20 preliminary candidates, has a P value (around 0.05) at 50% quantile cutoff, achieving a P value of 2.62×10^{-6} at 70% quantile cutoff.After the FDR correction, NDFIP1 still has P value of 1.07×10^{-4}. However, NDFIP1 could not pass the validation by using GSE65858 cohort.



Supplementary Figure S1. Under cutoff-finding procedure of Kaplan-Meier analysis, the P-value plot of gene "NDFIP1" shows: (1) 70% of P values is < 0.05; (2) the median-cut zone (zoom-in and revealed in inset box) has a "W"-like distribution; (3) sliding-window cutoff selection could find it's optimized P values (far less than 0.001) while a median cut might yield P value $>= 0.05$. (x-axis: grouping by person number; y-axis: P value in log10 transformed)