Comparative genomic analysis of oral versus laryngeal and pharyngeal cancer


DOI 10.1016/j.oraloncology.2018.04.006

Publication date 2018

Document Version Other version

Published in Oral Oncology

License Article 25fa Dutch Copyright Act

Supplementary Figure 1. Cluster analysis shows similarity of subsites between datasets. (A) Clustering on genes$_{SPM}$. The input value for each gene was the percentage of samples with a somatic point mutation in that gene, for each of the four datasets. Subsites of different datasets cluster together. (B) Clustering on genes$_{CNA}$. The input value for each gene was the percentage of samples with a copy number aberration of that gene, for each of the four datasets. OSCC$_{NKI}$ and OSCC$_{TCGA}$ cluster together, but L/P-SCC$_{NKI}$ and L/P-SCC$_{TCGA}$ do not.

The distance measure used for clustering was Euclidean distance.
Supplementary Figure 2. Mutational profiles of subsites correlate strongly between datasets. (A) Genes$_{SPM}$ and the percentage of samples with a somatic point mutation in each gene, in the NKI and TCGA dataset. Each circle and triangle represents a gene, in oral squamous cell carcinomas (OSCC) and laryngeal and pharyngeal squamous cell carcinomas (L/P-SCC), respectively. (B) Genes$_{CNA}$ and the percentage of samples with a copy number aberration in each gene, in the NKI and TCGA dataset.

Spearman’s rank correlation coefficient is reported.
Supplementary Figure 3. Absolute and relative frequency of transitions and transversion in TCGA samples. (A) Each transition and transversion is more frequent in L/P-SCC than OSCC. The frequency of each transition and transversion was calculated per sample. Boxplots show the distribution of these frequencies per subsite. (B) Boxplot of the proportion of each TiTv relative to the total number of TiTv's. Proportion of each TiTv relative to the total number TiTv's was calculated per sample. Boxplots show the distribution of these proportions per subsite.
Supplementary Figure 4. Association between mutation signature and genomic scar signature scores associated with HR deficiency. On the x-axis the absence or presence of COSMIC signature 3, associated with HR deficiency, in each TCGA tumor sample. This signature is based on somatic point mutations. On the y-axis the HR deficiency associated genomic scar signature scores, derived from copy number data. Significant associations exist between these measurements of HR deficiency, that were derived from different mutation types.

P-values from Wilcoxon rank-sum test.
Supplementary Figure 5. Proportion of samples with a somatic point mutation in the HR/FA genes in OSCC and L/P-SCC. (A) In TCGA data OSCC has a lower proportion of HR/FA mutated samples. (B) In NKI data the proportion of samples with an HR/FA mutation do not differ between OSCC and L/P-SCCC.

P-values from Fisher’s exact test.
Supplementary Figure 6. Overall survival Kaplan-Meier curves of TCGA data split by clinical variables. Top row shows overall survival in TCGA Oral Squamous Cell Carcinoma (OSCC) patients, split according to (A) pT stage, (B) pN stage and (C) age. Bottom row shows overall survival in TCGA ovarian cancer patients split according to (D) disease stage and (E) age. pT and pN stage were not available in the ovarian cancer dataset, hence disease stage was used. Age is shown as a split, but was also significant as a continuous variable in univariable Cox models in the OSCC ($P = 0.004$) and ovarian cancer ($P < 0.001$) datasets.