Supplemental Material

Supplemental Materials and Methods

DNA Microarray. The RNA and DNA from each specimen were simultaneously extracted using the TRIzol method (Invitrogen, Carlsbad, CA). To increase DNA purity, we modified the DNA extraction protocol to include the use of a "back extraction buffer" (4 M guanidine thiocyanate, 50 mM sodium citrate, and 1 M Tris, pH 8.0). RNA was further purified with the use of an RNeasy mini kit (Qiagen, Valencia, CA) per Affymetrix (Santa Clara, CA) recommendations. For expression array analysis, 1.0 to 2.5 µg of total RNA was used to generate biotin-labeled cRNA using the GeneChip Expression 3'-Amplification Reagents Kit (Affymetrix) per manufacturer's protocol. The cRNA was hybridized to an Affymetrix U133 2.0 Plus GeneChip arrays and scanned using an Affymetrix GeneChip arrays Scanner 3000 7G in the Fred Hutchinson Cancer Research Center's Genomics Shared Resources per Affymetrix protocols. At least one clinically normal tissue sample from a control subject was processed in tandem with every seven to eight tumor tissue samples from OSCC cases.

Validation of Gene Expression of *LAMC2*, *COL4A1*, *COL1A1*, and *PADI1* by qRT-PCR. From the 167 OSCC cases, a subset of 30 was randomly chosen from amongst a group of 40 in which RNA were plentiful and dilutions were readily available. Among the 45 controls, 41 had plentiful RNA available, from which 30 samples were chosen at random. Each sample containing 7.5 ng purified total RNA was assayed in triplicate in 10 µl reaction volumes using the QuantiTect SYBR Green RT-PCR kit (Qiagen, Valencia, CA) and bioinformatically validated QuantiTect primers (Qiagen, Valencia, CA) on a 7900HT Sequence Detection System (ABI, Foster City, CA). The cycling conditions were as follows: 30 minutes at 50° C, 15 minutes at 95° C, and 40 cycles of 15 seconds at 94° C, 30 seconds at 55° C, and 30 seconds at

72° C. For *COL1A1* (NM_000088), a 118-bp amplicon spanning exons 1 and 2 was amplified. For *COL4A1* (NM_001845), a 119-bp amplicon spanning exons 6, 7, 8, and 9 was amplified. For *LAMC2* (NM_005562), a 74-bp amplicon spanning exons 18 and 19 was amplified. For *PADI1* (NM_013358), an 80-bp amplicon spanning exons 3, 4, and 5 was amplified. We used *ACTB* as the reference gene and amplified a 146-bp amplicon that spans exons 3 and 4. Ten point standard curves were generated using Universal Human Reference RNA (Stratagene, La Jolla, CA) for all genes, except *PADI1* in which Normal Adjacent Esophagus Total RNA (Ambion, Austin, TX) was used. The linear correlation coefficient (R^2) was 0.99 or greater for all runs. The mean threshold cycle (Ct) values were calculated from the triplicate Ct values. Mean Ct values were further normalized in relation to the mean Ct value of the *ACTB* gene.