

Supplementary Figures

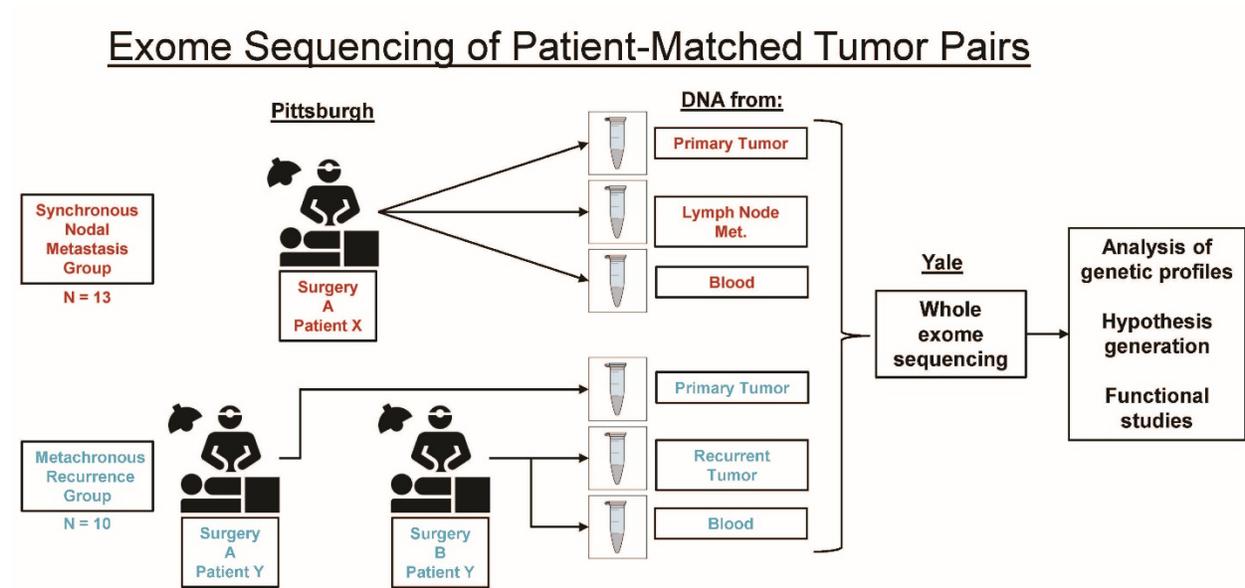


Figure S1. Patient cohorts and study design. To define and interrogate the genetic alterations underlying metastasis and recurrence in HNSCC, we analyzed two groups of patients. The synchronous nodal metastasis group consisted of 13 HNSCC patients that contributed normal tissue (blood), primary tumor tissue, and metastatic lymph node tumor tissue from a single time point. The metachronous recurrence group consisted of 10 HNSCC patients that contributed normal tissue (blood), primary tumor tissue, and recurrent tumor tissue from a later time point following relapse. Genomic DNA was isolated from this fresh frozen tissue and assessed by whole exome sequencing.

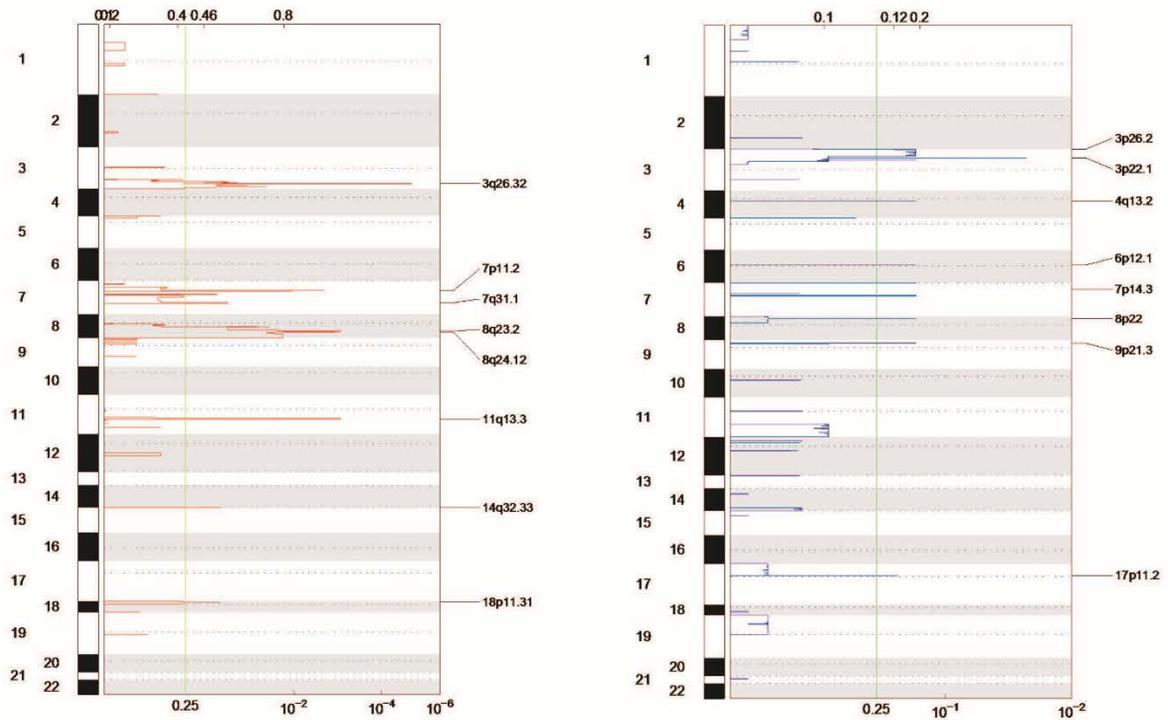


Figure S2. GISTIC2 analysis of index primary tumors. Plots showing significant peaks as identified by GISTIC2. Red, amplifications; blue, deletions; green line, $q=0.25$ signifying threshold for significance. Data from patients: PY-1, PY-3, PY-4, PY-5, PY-6, PY-7, PY-8, PY-10, PY-11, PY-12, PY-13, PY-14, PY-15, PY-16, PY-17, PY-19, PY-20, PY-21, PY-22, PY-23, PY-24, and PY-25.

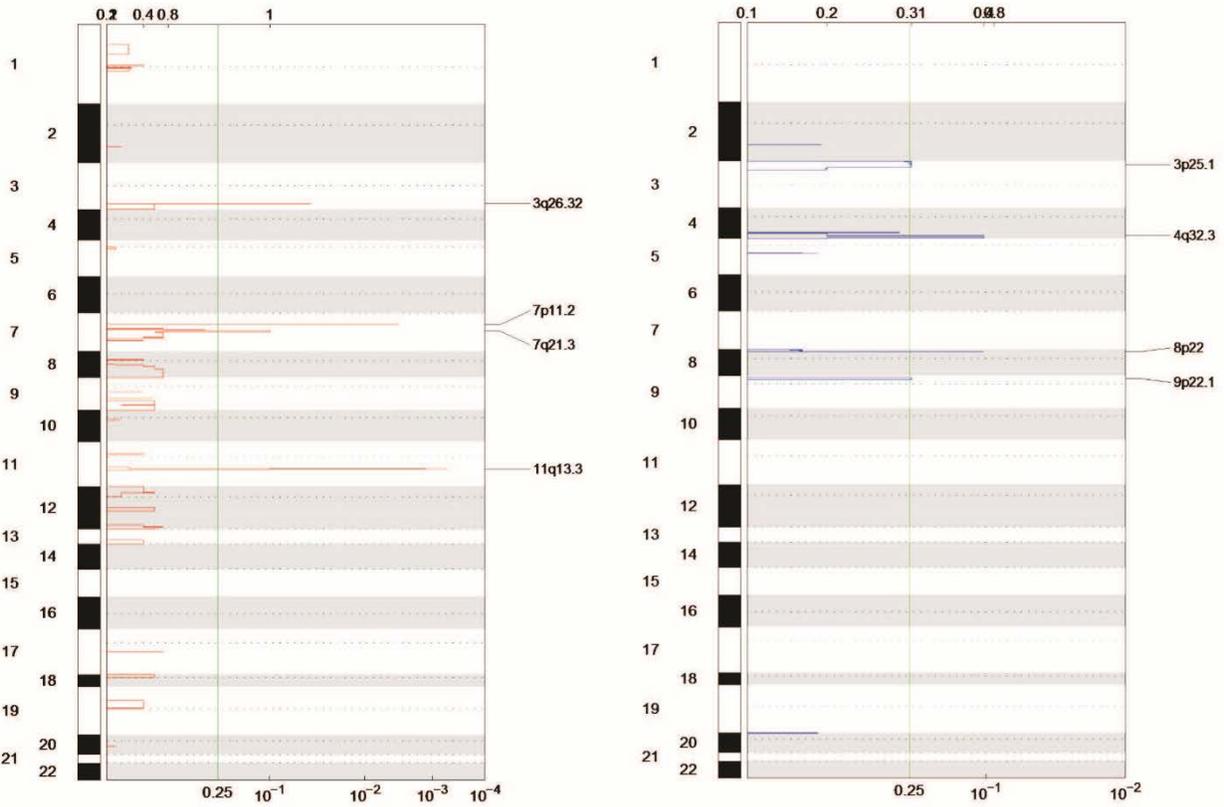


Figure S3. GISTIC2 analysis of synchronous nodal metastases. Plots showing significant peaks as identified by GISTIC2. Red, amplifications; blue, deletions; green line, $q=0.25$ signifying threshold for significance. Data from patients: PY-12, PY-14, PY-15, PY-16, PY-17, PY-20, PY-21, PY-22, PY-23, PY-24, and PY-25.

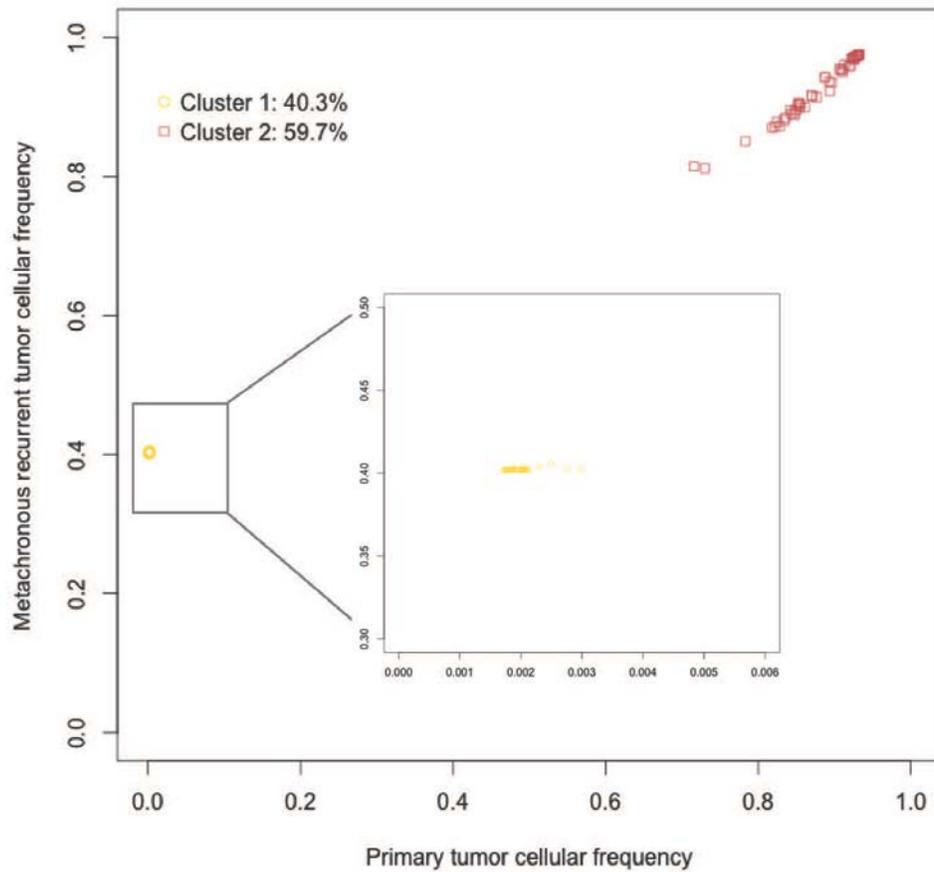


Figure S4. PyClone analysis of PY-4. Scatterplot showing cellular frequencies of SSNVs in primary tumor (x-axis) and metachronous recurrent tumor (y-axis). PyClone analysis reveals 2 separate clusters, with cluster 1 (yellow) at 40% cellular frequency in the metastatic tumor and at low frequencies $< 0.05\%$ in the primary, and cluster 2 (red) at high cellular frequencies in both primary and metastatic tumor. Inset is a magnification to show the distribution of SNVs in cluster 1.

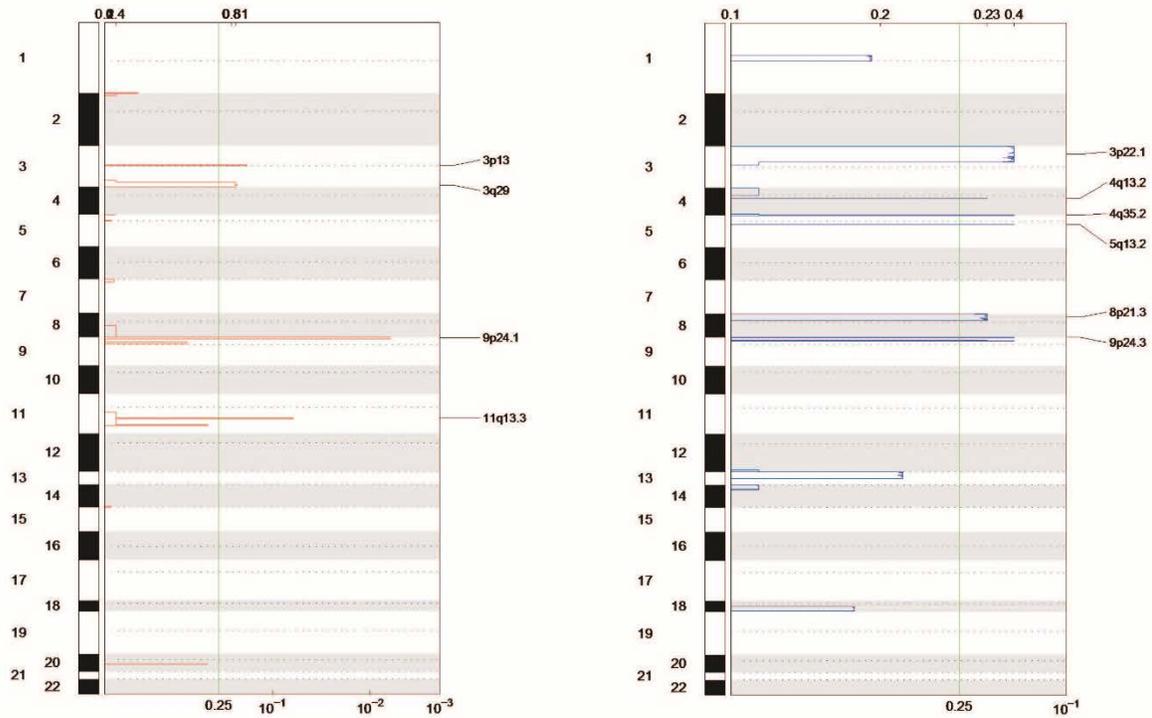


Figure S5. GISTIC2 analysis of metachronous recurrent tumors. Plots showing significant peaks as identified by GISTIC2. Red, amplifications; blue, deletions; green line, $q=0.25$ signifying threshold for significance. Data from patients: PY-1, PY-3, PY-4, PY-5, PY-6, PY-8, PY-9, PY-10, and PY-11.

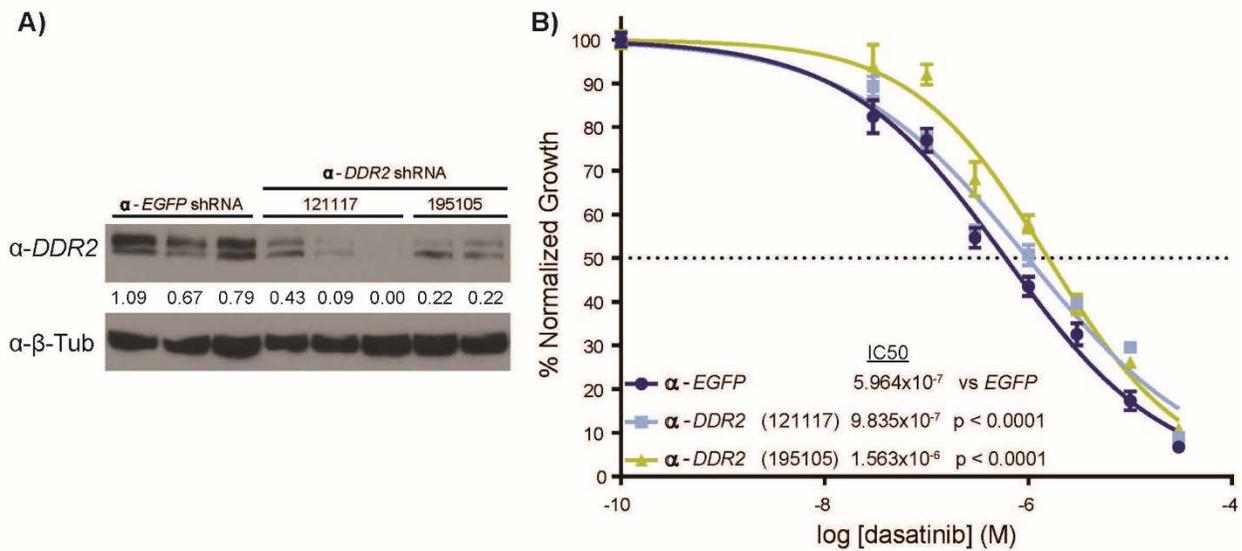


Figure S6. *DDR2* knockdown in BICR 18 modulates sensitivity to dasatinib. BICR 18 cells (HNSCC cells with an endogenous *DDR2*(D590G) mutation), were engineered with lenti-viruses to express one of two different anti-*DDR2* shRNAs, or an anti-*EGFP* shRNA. A representative immunoblot (westerns repeated at least twice) with densitometry normalized to β -tubulin from 8 unique subclones demonstrates the level of *DDR2* knock-down that was achieved. **(A)** BICR 18 cells expressing anti-*DDR2* shRNA (121117) and anti-*DDR2* shRNA (195105) were 1.6x and 2.6x less sensitive, respectively, to dasatinib treatment than BICR 18 cells expressing anti-*EGFP* shRNA. **(B)** Pooled data is presented from at least 8 replicate experiments per group (2-4/subclone). Cells of each subclone were plated in triplicate overnight in a 48-well plate (10×10^3 cells/well) and treated with half-log doses of dasatinib ranging from 30nM to 3 μ M for 48 hours, and assessed by MTT. Growth curves and statistics described in methods.

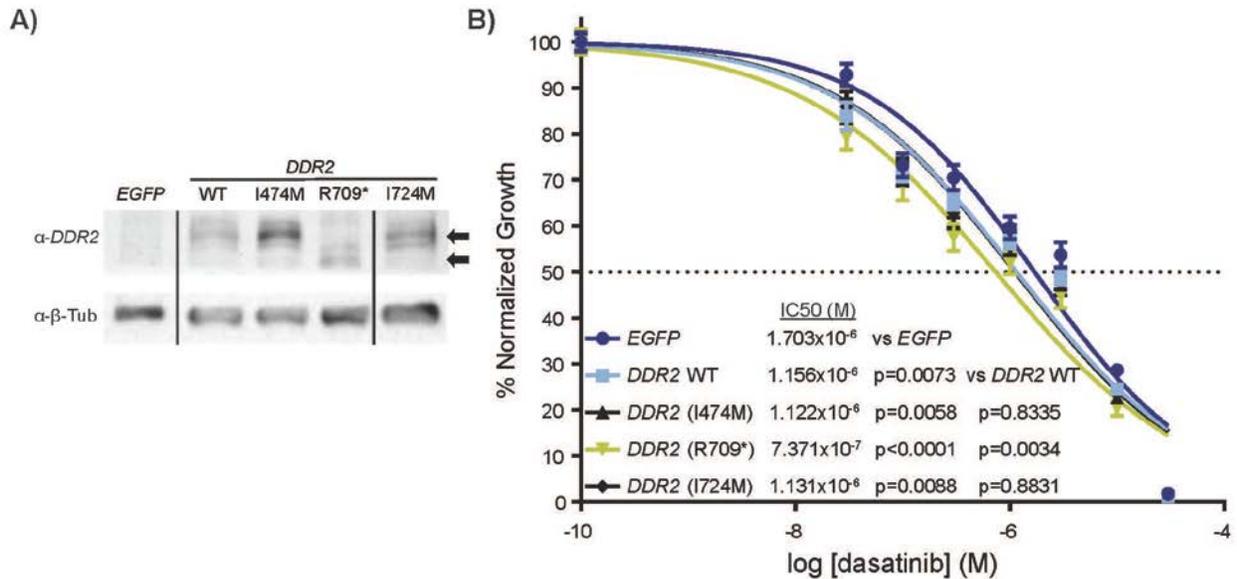


Figure S7. *DDR2* enhances dasatinib sensitivity in the HNSCC cell line, UPCI 15B. UPCI 15B cells (HNSCC cells with WT *DDR2* that was the most dasatinib resistant cell line in figure 5A) were engineered with retroviruses to express *EGFP*, WT *DDR2*, or one of the three mutant forms of *DDR2* that were identified in our cohort. A representative immunoblot (westerns repeated at least twice) demonstrates the expression of the *DDR2* constructs, including a downshifted band corresponding to the truncation mutation R709*, in these engineered cells (A). Lanes separated by black dividing lines were run on the same gel but were noncontiguous. UPCI 15B cells expressing WT *DDR2*, *DDR2*(I474M), or *DDR2*(I724M) were ~1.5x more sensitive to dasatinib treatment than *EGFP* control cells; whereas cells expressing the *DDR2*(R709*) mutant construct were ~2.3x more sensitive to dasatinib than *EGFP* control cells, and ~1.6x more sensitive to dasatinib treatment than UPCI 15B cells engineered with WT *DDR2* constructs (B). Pooled data is presented from 5 experiments composed of both technical and biologic replicates (2 and 3, respectively). Cells were plated in triplicate overnight in a 48-well plate (10×10^3 cells/well) and treated with half-log doses of dasatinib ranging from 30nM to 3 μ M for 48 hours, and assessed by MTT. Growth curves and statistics described in methods.

Supplementary Tables and Data: See attached files

Table S1. Summary of sequencing quality (n=25/group). Standard quality control metrics for exome sequencing averaged for each of the three samples obtained from patients. SLNM: Synchronous Lymph Node Metastasis, MR: Metachronous Recurrence

Table S2. Estimated tumor purity. Tumor purity was estimated from deviation in minor allele frequency of heterozygous SSNVs in segments showing loss of heterozygosity. Mutational data from the primary tumors of patients PY-7, PY-13, PY-19 and the recurrent tumor from PY-9 contributed to the genetic profiles of primary and MR tumors, respectively. But, those 4 patients were excluded from the genetic concordance analyses of matched tumor pairs that assessed transmission of genetic alterations from primary tumors to SLNM or MR tumors. SLNM: Synchronous Lymph Node Metastasis, MR: Metachronous Recurrence, SNV: Single Nucleotide Variant

Table S3. Cancer driver gene list (n=191 genes). This list was composed from the lists of cancer-driving/cancer-associated genes identified by Vogelstein *et al*, the TCGA Pan Cancer Effort, and the TCGA HNSCC cohort.(30, 34, 35)

Table S4. Mutated cancer driver genes in primary tumors (n=22 tumors). List of mutations identified in primary tumors in genes annotated as cancer driver genes as in Table S3.

Table S5. Percentage of genome harboring SCNVs across HNSCC tumors. Total genome length taken as 3.3×10^9 . Percentage of genome with SCNV calculated using same segmentation file and thresholds as in GISTIC analysis.

Table S6. Mutated cancer driver genes in synchronous lymph node metastases (n=11 tumors). List of mutations identified in synchronous lymph node metastases in genes annotated as cancer driver genes as in Table S3.

Table S7. Subclones identified from PyClone analysis. Cellular frequencies listed for primary tumors and matched synchronous lymph node metastases or metachronous tumors. Clonal and subclonal clusters listed as identified from PyClone analysis.

Table S8. Mutated cancer driver genes in metachronous tumors (n=9 tumors). List of mutations identified in metachronous in genes annotated as cancer driver genes as in Table S3.

Table S9. Orthogonal validation of sequencing pipeline. Mutations identified by the current (goldenrod) and previous (blue) sequencing studies of select patient tumors. Green shading indicates identically identified mutations, yellow shading indicates highly similar calls whose residue difference may result from differences due to the use of older versions of the reference sequence in the previous studies.

Table S10. SSNVs identified in HNSCC patient biospecimens.