SUPPLEMENTARY FILE

RNA isolation and gene expression analysis

Total RNA was isolated from 5-µm–thick formalin-fixed, paraffin-embedded tumor sections using the Ambion RecoverAll kit (Thermo Fisher Scientific) according to the manufacturer's instructions and was quantified using the NanoDrop ND1000 spectrophotometer (Thermo Fisher Scientific). Total RNA was mixed with a 3′ biotinylated capture probe and a 5′ reporter probe tagged with a fluorescent barcode from the custom gene expression code set. Probes and target transcripts were hybridized overnight at 65°C. Resultant samples were run on the NanoString nCounter platform (NanoString Technologies) using the high-sensitivity protocol and were analyzed using a custom code set consisting of a 680-gene panel related to T-cell biology, immune regulation, and cellular markers of tumor-infiltrating lymphocytes and tumor-associated macrophages. The samples were scanned at maximum resolution on the nCounter Digital Analyzer (NanoString Technologies).

Analysis of tumor mutational burden

Single-nucleotide variants

Whole exome sequence (WES) reads were aligned to reference human genome GRCh37; the method is described elsewhere.[1-3] MuTect (v1.1.7-3-gcb7069b) from The Cancer Genome Analysis suite released by Appistry (http://www.appistry.com/) was used to generate somatic single-nucleotide variant (SNV) calls.[4] Those present in the Single Nucleotide Polymorphism Database (dbSNP,v141)[5] but not in the Catalogue of Somatic Mutations in Cancer (COSMIC496v68)[6] were filtered out, as were SNVs with fewer than four mutant reads in tumor samples. Tumor mutational burden (TMB) for a patient was defined as the sum of somatic nonsynonymous SNVs remaining after the filtering process.

Haddad RI, *et al. J Immunother Cancer* 2022; 10:e003026. doi: 10.1136/jitc-2021-003026

Clonality and clonality-weighted TMB

Cellular prevalence for all somatic SNVs was estimated by inputting somatic SNVs with variant allele frequency information obtained using MuTect[4] and Sequenza[7] output allele-specific copy number and cellularity estimation into PyClone software. Mutational clonality was inferred through the clustering process of PyClone.[2,8] Clonality for each sample was determined using the method of McGranahan et al.[9] Clonality-weighted TMB was calculated by multiplying the clonality by TMB as an estimate of the TMB restricted to clonal mutations.

SNV annotation and neoantigen detection

TMB was defined as nonsynonymous mutations in protein coding regions. Of the possible 9-mer peptide sequences with mutated amino acid inside for each nonsynonymous mutation locus, that with the highest binding affinity with the human leukocyte antigen (HLA) alleles from a nonsynonymous mutation locus was selected as the representative antigen for the mutation. HLA-I major loci, A, B, and C, were typed at four-digit resolution by using OptiType (v1.0) as previously described.[2] Further, for output-typed alleles not found in the NetMHC (v3.4) input list, the corresponding supertype was identified for each allele and the supertype-representative allele was used for NetMHC.[2] Neoantigens were defined as representative antigens with an HLA-A or HLA-B binding affinity of <50 nM.

Mutational signature analysis

Mutational signature analysis was performed using the deconstructSigs package (v1.6.0),[10] and exome regions were defined by Sureselect V5 target region (Agilent). Only somatic mutations in exome regions were considered, and trinucleotide counts were normalized by the number of times each trinucleotide context was observed in the exome region. Mutational signatures, as defined in Alexandrov et al,[11] were the target signature set to be screened, and the relationships

between clinical outcomes and these various mutational signatures, including specific nucleotide changes, DNA repair, smoking, neoantigen, tp53, and apolipoprotein B mRNA editing enzyme, were evaluated in samples across the pan-tumor cohort.

Assessment of tumor PD-L1 expression

Tumor programmed death ligand 1 (PD-L1) expression was assessed using PD-L1 IHC 22C3 pharmDx (Agilent) combined positive score (CPS), where CPS is defined as the ratio of the number of all PD-L1–staining cells (tumor cells, lymphocytes, macrophages) to the number of all viable tumor cells, multiplied by 100.

Association between TMB, T-cell–inflamed gene expression profile, and PD-L1 with clinical response

Agreement between p16 immunohistochemistry (IHC)-determined and WES-determined HPV status was established. Clinical response associations were assessed based on objective response rate (ORR) and progression-free survival (PFS) per RECIST v1.1. Associations of ORR, PFS, and overall survival with TMB, neoantigen load (NL), clonality, PD-L1 CPS, and T-cell–inflamed gene expression profile ($Tcell_{inf}$GEP) were assessed in all patients for whom transcriptomic data were available.

Determination of human papillomavirus status by whole exome sequencing

Cancer-associated DNA virus signals were identified from Exome-Seq data using virus reference genomes for189 human papillomavirus (HPV) types from the Human Papillomavirus Episteme[12] combined with the human reference genome hg19/GRCh37 as a chimeric reference genome. All Exome-Seq reads were mapped to this chimeric reference genome using BWA (v0.7.5a-r405) aln[1] with default parameters. Using cohort B as a training set, a cutoff of 20 reads was identified

by maximizing concordance of WES-based HPV with clinical HPV assay; subsequently, patients were classified as positive for a particular virus if ≥20 reads were properly mapped to the reference genome for this virus.

Statistical analysis

For ORR, a responder was defined as a patient with a "complete response" or a "partial response" as determined by RECIST v1.1, assessed by central review. Models include covariate terms for Eastern Cooperative Oncology Group performance status and cohorts. TMB and clonality-weighted TMB were evaluated on a log scale because of the skewness of TMB score. However, NL was evaluated on a square root scale because two patients had a zero count. PD-L1 was also evaluated on a square root scale. Joint modeling that incorporated both the TMB and an inflammatory biomarker ($Tcell_{inf}$GEP or PD-L1) were conducted to establish whether TMB confers additional value after adjustment for the inflammatory biomarker.

Table S1 Baseline characteristics of the patients from cohorts B1 and B2 of the KEYNOTE-012 trial

Characteristic	**Overall[a] N=192**	**WES[b] n=106**
Age, median (range), years	60 (20-84)	61 (25-83)
Male	159 (83)	87 (82)
ECOG PS 1	135 (70)	71 (67)
Metastatic staging (M1)	165 (86)	91 (86)
Prior therapy		
0	34 (18)	21 (20)
1	41 (21)	18 (17)
2	45 (23)	33 (31)
≥3	72 (38)	34 (32)
HPV status: positive p16 IHC	45 (23)	25 (24)
HPV status: positive WES	NA	28 (26)

All values are n (%) unless stated otherwise.

ECOG PS, Eastern Cooperative Oncology Group performance status; HPV, human papillomavirus; IHC, immunohistochemistry; NA, not applicable; WES, whole exome sequencing.

[a]Overall KEYNOTE-012 study cohort (B1 + B2).

[b]Patients with available WES data.

Table S2 HPV status by p16 immunohistochemistry and WES methods

Method	**HPV-negative (p16)**	**HPV-positive (p16)**	**Total (WES)[a]**
HPV-negative (WES)	71	7	78/106 (74%)
HPV-positive (WES)	10	18	28/106 (26%)
Total HPV (p16)[b]	81/106 (76%)	25/106 (24%)	—

HPV, human papillomavirus; WES, whole exome sequencing.

[a]Defined as >20 reads mapping to the HPV genome, all tumor sites.

[b]Oropharynx.

Table S3 *TP53* mutation by HPV status using p16 IHC or WES methods

Mutation	**WES**		**p16 IHC**	
	HPV-negative	**HPV-positive**	**HPV-negative**	**HPV-positive**
*TP53*wt	36	26	40	22
*TP53*mut	42	2	41	3

HPV, human papillomavirus; IHC immunohistochemistry; WES, whole exome sequencing.

REFERENCES

1 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754-60.

2 Cristescu R, Mogg, R, Ayers M, *et a*l. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science* 2018;362:pii: eaar3593.

3 McKenna A, Hanna M, Banks E, *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297-303.

4 Cibulskis K, Lawrence MS, Carter SL, *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013;31:213-9.

5 Sherry ST, Ward M, Sirotkin K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res* 1999;9:677-9.

6 Forbes SA, Tang G, Bindal N, *et al.* COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucl Acids Res* 2010;38:D652-7.

7 Favero F, Joshi T, Marquard AM, *et al.* Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann Oncol* 2015;26:64-70.

8 Roth A, Khattra J, Yap D, *et al.* PyClone: statistical inference of clonal population structure in cancer. *Nat Methods* 2014;11:396-8.

9 McGranahan N, Furness AJ, Rosenthal R, *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* 2016;351:1463-9.

10 Rosenthal R, McGranahan N, Herrero J, *et al.* DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol* 2016;17:31.

11 Alexandrov LB, Nik-Zainal S, Wedge DC, *et al.* Signatures of mutational processes in human cancer. *Nature* 2013;500:415-21.

12 Van Doorslaer K, Tan Q, Xirasagar S, *et al.* The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis. *Nucl Acids Res* 2013;41:D571-8.