Supplemental Methods

Nucleic acid extraction and sample assessment

Constitutional DNA was extracted from peripheral blood mononuclear cells using Qiagen's Gentra Puregene Blood Kit. For patients 18 through 36, constitutional DNA was extracted from adjacent normal tissue that was confirmed by a pathologist to not contain tumor cells. To extract tumor DNA and RNA, and constitutional DNA from normal tissue, fresh frozen tissue specimens were homogenized using the Bullet Blender™ Tissue homogenizer in Buffer RLT and stainless steel beads. Samples were passed through Qiagen's QIAshredder spin columns, and subsequently extracted for DNA and RNA using Qiagen's AllPrep Mini Kit. All samples were assessed and quantitated using the Nanodrop, the Invitrogen Qubit, and the Agilent TapeStation.

Library construction and next generation sequencing

Exome libraries were constructed using KAPA Biosystems' Hyper Prep Kit (cat#KK8504) and Agilent's SureSelectXT Human All Exon V5+UTR baits (cat#5190-6221) following the manufacturer's protocols. RNA libraries were constructed using Illumina's TruSeq RNA Library Preparation Kit V2 (cat#RS-122-2001) or Illumina's TruSeq RNA Access Library Prep Kit (cat#RS-301-2001) following the manufacturer's protocols. LIWG libraries were constructed as previously described(Liang et al. 2014) with modifications: 200ng inputs were used, the KAPA Biosystems's Hyper Prep Kit was used for construction, one cycle of PCR was performed prior to size selection and six PCR cycles was performed following size selection for each sample, and size selection was performed using Sage Sciences' Pippin Prep to select library molecules that are 925-1025bp in length. Libraries were clustered onto Illumina V3 flowcells (San Diego, CA) using Illumina's TruSeq PE Cluster Kit V3 (cat#PE-401-3001) and sequenced

for paired 83bp reads using Illumina's TruSeq SBS Kit V3 (cat#FC-401-3002, n=3) on the Illumina HiSeq.

Data analysis

BCL files were converted to FASTQ files using Illumina's BCLConverter tool. FASTQ files were aligned to build 37 of the human reference genome using BWA (Burrows-Wheeler Aligner)(Li and Durbin 2009). For exome data, resulting BAM files were recalibrated using GATK(McKenna et al. 2010) and duplicate pairs marked using Picard. Microalignment was completed to identify reads containing indels (insertion/deletions). Variant calling was performed using Seurat(Christoforides et al. 2013)(quality score>30), MuTect(Cibulskis et al. 2013), and Strelka(Saunders et al. 2012) and calls were annotated using GENCODE version 3 by Ensembl and build 37.1. No quality scores are generated by MuTect and Strelka. Final somatic SNVs were called by at least 2/3 callers. For the oncoprint (Figure 1), mutation burdens for patients with multiple tumors are reflected as the median value across tumors. For patient 33, because constitutional DNA was not available for generation of an exome, a merged constitutional data set comprised of constitutional data from three randomly selected patients (patients 2, 3, and 5) was used for variant calling. Counts of somatic SNV, CNV, and SV load for this patient was not included for overall counts of all ALMs. LIWG data was utilized for copy number and breakpoint detection analyses(Liang et al. 2014). A minimum tumor allele ratio of 0.10 and a minimum quality score (depth) of 20 is required for an SV to be called. For nine patients, sufficient amounts of constitutional DNA was not available for LIWG sequencing—for somatic analyses of unpaired tumor LIWGs, a merged constitutional data set comprised of constitutional LIWGs from three randomly selected patients (patients 2, 3, and 5, similarly used

for patient 33's exome analysis) were thus used for comparisons. Patient 33 data is not included in mutation count metrics because a true matched normal was not available for this patient.

For CNV detection, read depths at every 100 bases across sequenced regions are first determined. Next, normalized log2 fold-changes between tumor and normal are calculated and a smoothing window is applied. In addition, we utilize allele frequencies in the tumor of known heterozygous germline SNPs identified within the normal to both evaluate potential false positives and correct biases. Lastly, we applied a circular binary segmentation (CBS) algorithm to corrected log2 fold changes using the Bioconductor DNAcopy implementation (https://bioconductor.org/packages/ release/bioc/html/DNAcopy.html). For mutation burden metrics, a focal CNV is included if the log2 change is >= |2| with maximum length of 25Mb. RNA reads were aligned to build 37 of the human genome using STAR(Dobin et al. 2013) and differential analysis against a universal RNA control was performed using Cuffdiff (q-value<0.05) and DESeq2(Love et al. 2014) (padjusted<0.05). RNA fusions were detected using TopHat-Fusion(Kim and Salzberg 2011) (quality score>100).

MutSigCV (Mutation Significance covariates)(Lawrence et al. 2013) and IntOgen (Integrative Onco-Genomics; OncodriveFM)(Gonzalez-Perez et al. 2013) were used to identify putative cancer drivers. To remove potential bias introduced from analyzing multiple tumors from the same patient, the union of data derived from tumors from the same patient was used for analysis for patients 25, 29, and 34. Consensus CNVs in both LIWG and exome data were identified using GISTIC2.0 (Genomic Identification of Significant Targets in Cancer; release 2.0.16)(Mermel et al. 2011) (Figure 2B). The q-bound cut-off was set at 0.05 (95% confidence

interval; Benjamini & Hochberg FDR). G-scores, which reflect the amplitude of copy changes compared against the frequency of the event across all samples, were also calculated. Sequenza(Favero et al. 2015) was used to estimate percent tumor cellularities from rank 1 values for each sample. Analysis of mutation signatures was performed using the Mutational Signature Analysis Tool (https://bitbucket.org/jtr4v/analysis-of-mutational-signatures) (Supplemental Fig S2). All somatic SNVs except mutations with multiple alternate alleles were used to capture the mutational signature. The SomaticSignatures R package(Gehring et al. 2015) was used to identify somatic signatures using somatic SNVs called by at least 2 of 3 callers off exome data after: (1) removing ExAC (Exome Aggregation Consortium)(Lek et al. 2016), dbSNP (the NCBI [National Center for Biotechnology Information] Short Genetic Variations database), and NHLBI (National Heart, Lung, and Blood Institute) GO (Grand Opportunity) ESP (Exome Sequencing Project) SNPs with minor allele frequencies > 3%, sans COSMIC (Catalogue of Somatic Mutations in Cancer)(Forbes et al. 2015) SNPs; and (2) identifying the optimal number of signatures that accounts for the greatest amount of variance across samples using an RSS (residual sum of squares) statistic and the expected variance. A final list of 12,802 SNVs was used for analysis and an optimal number of ten signatures was used to identify signatures using the SomaticSignatures tool. ANOVA was used to compare resulting signatures against the somatic cancer signatures described by Alexandrov et al. (Alexandrov et al. 2013). Correlations between ALM signatures and Alexandrov signatures were performed by calculating a cosine similarity value. Analysis of the distribution across samples was also performed using the SomaticSignatures tool.

For gene set analysis of RNA-seq data (Supplemental Fig S4A), an FPKM matrix (where FPKMs of zero or “inf” were imputed across row medians and gene IDs collapsed using the

median) was analyzed using the ssGSEA (single sample gene set enrichment analysis) projection module of GenePattern(Reich et al. 2006) using default parameters and a custom gene ontology gene set file. The raw ssGSEA projection returned values were z-score normalized and clustered using the AutoSOME clustering method with gene sets as rows and samples as columns. Row clustering was set at 500 Ensemble runs and P<=0.05 with median center and sum of squares=1. Column clustering was set at 150 Ensemble runs and P<=0.075 with unit variance adjustment. Results were filtered for confidence values > 40, converted to gct format, and visualized in GENE-E software (http://www.broadinstitute.org/cancer/software/ GENE-E/index.html). For unsupervised clustering analysis (Supplemental Fig S4B), genes demonstrating low variance based on FPKMs (cutoff=30,000) were removed. FPKM data was scaled and standardized across the remaining 421 genes and hierarchical clustering was performed using Spearman's correlation and clusters were joined with complete linkage. Computation of silhouette indices yielded six gene clusters.

For neo-antigen analysis (Supplemental Figure S5), we predicted the number of neo-antigens generated based on somatic mutations. BWA's HLA caller (bwa-0.7.11) is first used to generate six digit HLA calls for the HLA-A, B, and C genes. Mutated protein sequences were generated from somatic non-synonymous mutations, frameshifts, codon insertions, and codon deletions called by at least two of the three somatic variant callers (MuTect, Seurat, Strelka). The IEDB MHC class I binding prediction tool (version 1-2.13) was used to predict binding of mutated peptides to each individual's MHC types. NetMHCpan(Nielsen and Andreatta 2016) was used as the prediction method, with peptide lengths of nine and ten. Neo-antigens with an IC50 < 500 are retained. Analysis of HLA expression was also performed using seq2HLA 2.2(Boegel et al. 2012).

Experimental validations

PCR and Sanger sequencing were used to validate *BRAF*, *NRAS*, and *TERT* point mutations, as well as hotspot *TERT* promoter mutations, on samples with available tumor DNA. Validation of *PAK1* and *TERT* CNV gains were performed using TaqMan Copy Number Assays with real-time PCR using TaqMan MGB (minor groove binder) probe chemistry (ThermoFisher Scientific). Reference genes were selected from copy neutral regions of respective samples.

Cell line quantitative PCR (qPCR)

Cell lines were plated in T25 flasks. 24h after plating cells, dimethyl sulfoxide (DMSO) vehicle, or Telomerase Inhibitor IX (EMD Millipore, Billerica, MA), was diluted in culture medium, then added to cells. Cells were harvested after 72 h of drug treatment, and RNA was extracted using the RNeasy kit (Qiagen, Germantown, MD). Reverse transcription was performed with Quantitect Reverse Transcription kit (Qiagen, Germantown, MD). qPCR was performed using the Kapa Fast qPCR master mix (Kapa Biosystems, Wilmington, MA) and TaqMan probes designed to target TERT and GAPDH (ThermoFisher, Waltham, MA).

References

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL et al. 2013. Signatures of mutational processes in human cancer. *Nature* **500**(7463): 415-421. doi: 410.1038/nature12477. Epub 12013 Aug 12414.

Boegel S, Lower M, Schafer M, Bukur T, de Graaf J, Boisguerin V, Tureci O, Diken M, Castle JC, Sahin U. 2012. HLA typing from RNA-Seq sequence reads. *Genome Med* **4**(12): 102. doi: 110.1186/gm1403. eCollection 2012.

Christoforides A, Carpten JD, Weiss GJ, Demeure MJ, Von Hoff DD, Craig DW. 2013. Identification of somatic mutations in cancer through Bayesian-based analysis of sequenced genome pairs. *BMC Genomics* **14:302.**(doi): 10.1186/1471-2164-1114-1302.

Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G. 2013. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**(3): 213-219. doi: 210.1038/nbt.2514. Epub 2013 Feb 1010.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**(1): 15-21. doi: 10.1093/bioinformatics/bts1635. Epub 2012 Oct 1025.

Favero F, Joshi T, Marquard AM, Birkbak NJ, Krzystanek M, Li Q, Szallasi Z, Eklund AC. 2015. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann Oncol* **26**(1): 64-70. doi: 10.1093/annonc/mdu1479. Epub 2014 Oct 1015.

Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, Ding M, Bamford S, Cole C, Ward S et al. 2015. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* **43**(Database issue): D805-811. doi: 810.1093/nar/gku1075. Epub 2014 Oct 1029.

Gehring JS, Fischer B, Lawrence M, Huber W. 2015. SomaticSignatures: inferring mutational signatures from single-nucleotide variants. *Bioinformatics* **31**(22): 3673-3675. doi: 3610.1093/bioinformatics/btv3408. Epub 2015 Jul 3610.

Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Tamborero D, Schroeder MP, Jene-Sanz A, Santos A, Lopez-Bigas N. 2013. IntOGen-mutations identifies cancer drivers across tumor types. *Nat Methods* **10**(11): 1081-1082. doi: 1010.1038/nmeth.2642. Epub 2013 Sep 1015.

Kim D, Salzberg SL. 2011. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol* **12**(8): R72. doi: 10.1186/gb-2011-1112-1188-r1172.

Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA et al. 2013. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**(7457): 214-218. doi: 210.1038/nature12213. Epub 12013 Jun 12216.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**(7616): 285-291. doi: 210.1038/nature19057.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14): 1754-1760. doi: 1710.1093/bioinformatics/btp1324. Epub 2009 May 1718.

Liang WS, Aldrich J, Tembe W, Kurdoglu A, Cherni I, Phillips L, Reiman R, Baker A, Weiss GJ, Carpten JD et al. 2014. Long insert whole genome sequencing for copy number variant and translocation detection. *Nucleic Acids Res* **42**(2): e8. doi: 10.1093/nar/gkt1865. Epub 2013 Sep 1025.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**(12): 550.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for

analyzing next-generation DNA sequencing data. *Genome Res* **20**(9): 1297-1303. doi: 1210.1101/gr.107524.107110. Epub 102010 Jul 107519.

Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. 2011. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**(4): R41. doi: 10.1186/gb-2011-1112-1184-r1141. Epub 2011 Apr 1128.

Nielsen M, Andreatta M. 2016. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med* **8**(1): 33. doi: 10.1186/s13073-13016-10288-x.

Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. 2006. GenePattern 2.0. *Nat Genet* **38**(5): 500-501.

Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. 2012. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**(14): 1811-1817. doi: 1810.1093/bioinformatics/bts1271. Epub 2012 May 1810.