Supplementary Methods

Alternate promoter utilization burden algorithm

Alternate promoter utilization burden was calculated using *proActiv*, an algorithm that estimates promoter activity from short-read RNA-Seq data by mapping and quantifying first intron junctions of the genome. This method has been described extensively and is available as an R package [1]. First, FASTQ files were aligned to GENCODE v19. The Pan-TCGA RNA-Seq was aligned using TopHat2 (v2.0.12), while all remaining RNA-Seq cohorts aligned using STAR v2.6.1. The TCGA STAD cohort was aligned using both TopHat2 v2.0.12 and STAR v2.6.1. Splice-junction files were extracted for input into *proActiv* v0.1.0. Using the calculatePromoterReadCounts function from the *proActiv* package, overlapping first exons of each transcription start site (TSS) were combined to obtain a set of 113,076 promoters across the entire genome. Each promoter was quantified using junction reads aligning into the first introns of the constituting transcripts. Genomic regions were selected using data from our previous H3K4me3 ChIPSeq study on epigenetic promoter alterations that identified specific gain-of-expression ("gain promoters") and loss-of-expression ("loss promoters)" loci which were associated with immune-editing in STAD [2, 3]. Of the 113,076 promoters in the genome, we identified promoters which mapped to these specified regions. We selected 4,519 promoters (3,152 gain promoters and 1,367 loss promoters) which were used to measure alternate promoter utilization burden. Read counts were normalized by library size per sample. Promoter counts were transformed based on weights derived from TCGA gastric cancer samples (STAD). Median normalized promoter read count values of gained promoters were multiplied by 4, and loss promoters were divided by 4 to establish weights. Transformed promoter read counts were assigned a point if the value was ≥ 1. The sum of the gain and loss promoter points was assigned the alternate promoter utilization burden (APB). Correlation between median promoter read counts between STAR and TopHat2 of the TCGA STAD samples was high (Spearman R = 0.98, $p < 0.0001$). Samples were classified into groups: Those in the top quartile of APB were classified as APB_{high}, in the lowest quartile as APB_{low}, while the rest of the samples were classified as APB_{int} (**Fig 1A**). APB across batches/cohorts were normalized prior to categorizing into groups.

Pan-cancer TCGA analysis

Gene expression data and clinical data from the PanCanAtlas were downloaded from Firebrowse [4]. Illumina HiSeq RNA-SeqV2 RSEM normalized gene values were used for correlations of *CD8A, GZMA* and *PRF1* and other genes. All tumor types within the database were included except for tumors of hematological or immune origin (lymphoma, leukemia and thymic carcinoma) as correlation with immune-related outcomes from bulk RNA-Seq data of these tumor types would not be meaningful. “Colorectal cancer” included colon cancer (COAD) and rectal cancer (READ), “kidney cancer” included kidney chromophobe (KICH), kidney renal cell (KIRC) and kidney renal papillary cell carcinoma (KIRP), and “glioma” included glioblastoma multiforme (GBM) and low-grade glioma (LGG). For uterine corpus endometrial carcinoma (UCEC), Illumina Genome Analyzer RNA-Seq V2 RSEM normalized gene expression values from Firebrowse were used, as the dataset was larger compared to the Illumina HiSeq RNA-SeqV2 RSEM dataset (381 *vs*. 176 samples).

Immune correlates

Transcriptomic gene expression Level 3 RSEM-normalized RNASeqV2 data was extracted from the Broad GDAC Firebrowse [4]. STAD TCGA subtypes were downloaded from cBioPortal, Stomach Adenocarcinoma (TCGA PanCancer Atlas), Clinical Data. Differential gene expression analysis in the Pan-Cancer TCGA analysis was performed with p value adjustments using the Bonferroni method, and $-\log_{10}(\text{qval})>2$ were considered significant. The Immunology Database and Analysis Portal (ImmPort) system gene list of 4,627 immunologically related genes was extracted from InnateDB [5]. The list was generated through initial automatic searches of EntrezGene and Gene Ontology using immunology-related keywords and then manually curated by immunology experts. Microsatellite instability (MSI-H) status of samples were extracted from Cortes-Ciriano *et al* [6]. Microsatellite status was derived by combining MSI status from the TCGA consortium, which was performed using a panel of four mononucleotide repeats and three dinucleotide repeats and whole-exome analysis of 7089 samples, using the 0.95 confidence calls. Immune subtypes, other immune signatures and tumor mutational burden (TMB), disease free survival (DFS) and overall survival (OS) including censorship data were

extracted from the Pan-Cancer immune landscape analysis by Thorsson *et al* [7]. To avoid confounding at the pan-cancer level where outcomes may be influenced by intrinsic tissue- or site-specific properties, tumor-type specific analyses were performed for most correlations.

Tumor purity, intratumor heterogeneity, and immune cell fraction data

The consensus tumor purity estimates of the TCGA STAD samples were downloaded from the supplementary data of Ghoshdastider *et al* [8]. Briefly, the consensus tumor purity was caluculated as the mean of normalized purity values obtained from 3 genome-based methods (AbsCN-Seq, PurBayes, and ASCAT) and 1 transcriptome-based method (ESTIMATE). The intratumor heterogeneity scores (from clonality calls of ABSOLUTE) and immune cell fraction estimates (from CIBERSORT) were downloaded from the supplementary data of Thorsson *et al* [7].

Single-Cell RNA-Seq of Gastric Cancer

Sample cohort description

Patients diagnosed with gastric adenocarcinoma and planned for surgical resection at the National University Hospital, Singapore were enrolled after obtaining written informed consent. The study was approved by the local ethics board (DSRB Ref No: 2005/00440). From the fresh resected specimens, four tumor biopsies taken from different and representative quadrants of the tumor were processed for the scRNA-Seq experiment. Additional tumor tissue was also processed for whole-exome sequencing (WES) and bulk whole transcriptome sequencing (bulk RNA-Seq). Histopathological, staging and clinical data from the patients were also available for correlation.

scRNA-Seq Library preparation

Enriched 5’ gene expression libraries were constructed from single cells dissociated from tumor samples. Gel Beads in Emulsion (GEM)s were generated by combining barcoded single cell 5’ gel beads, master mix with cells, and partitioning oil on Chromium Chip A. Reverse transcription occurred within the generated GEM after which the GEMs were dissolved. The 10x barcoded, full-length cDNA was amplified via polymerase chain reaction (PCR), generating sufficient material to construct multiple libraries from the same cells. The enriched libraries were later enzymatically

digested, size selected, and adaptor ligated to be sequenced. 5'gene expression libraries were made similarly from amplified CDNA by enzymatic digestion, size selection and adaptor ligation. Libraries were subsequently sequenced on a Illumina Hiseq sequencer.

Single-cell gene expression quantification and determination of the major cell types:

Unique molecular identified (UMI) count matrix were first generated for each sample by passing the raw data in the cell ranger software. The count matrix was later used to generate a Seurat object which is used for clustering analysis [9, 10]. Quality control (QC) was done by filtering the cells that had unique feature counts over 2,500 or less than 200 and filtering the cells that had >5% mitochondrial counts. After removing the unwanted cells, data normalization followed by feature selection, data scaling and linear dimensional reduction was performed. Highly variable genes were identified using the 'FindVariableGenes' function. The first 20 principal components (PCs) and a resolution of 0.8 were used for clustering using 'FindClusters'. UMap was used for two-dimensional representation of first 20 PCs with 'RunUMAP'. Differential gene expression for identifying markers of a cluster relative to all other clusters or compared to a specific cluster was determined using the 'FindAllMarkers' or 'FindMarkers' functions respectively. Marker genes were compared for each cluster to literature-based markers of cell lineages to assign a cell lineage per cluster. Cell clusters were labelled based on curated and described cell markers [11].

Humanized mouse model

Mice and animal care

NOD-scid Il2rγ^{null} (NSG) mice were purchased from the Jackson Laboratory (stock number 005557). All mice were bred and kept under pathogen-free conditions in Biological Resource Centre, Agency for Science, Technology and Research, Singapore (A*STAR) on controlled 12-hour light-dark cycle. All experiments and procedures were approved by the Institutional Animal Care and Use Committee (IACUC; IACUC# 191440) of A*STAR in accordance with the guidelines of Agri-Food and Veterinary Authority and the National Advisory Committee for Laboratory Animal Research of Singapore.

Generation of humanized-mice

One to three day-old NSG pups were sub-lethally irradiated at 1 Gy and engrafted with 1 x 10^5 human $CD34^+$ cordblood cells (HLA-A24:02; Stemcell Technologies) via intrahepatic injection. Mice were submandibularly bled at 10 weeks post-engraftment to determine the levels of human immune reconstitution via flow cytometry. Mice with more than 10% human immune cell reconstitution (calculated based on the proportion of human CD45 relative to the sum of human and mouse CD45) were included in the study.

Gastric cancer cell-line selection

As the humanized-mouse models were generated from HLA-A24:02 cord-blood cells, gastric cancer cell-lines with HLA-A24:02 subtype in at least one allele were selected. RNA-Seq and WES was performed on 30 gastric cancer cell lines to infer APB and HLA subtype. Cell lines with HLA-A24:02 subtype for at least one allele were selected for testing in the humanized mouse model experiment. To confirm ability of cell-lines to form tumors in immunodeficient mice first, 2 x 10^6 of cells were subcutaneously injected into the shaved right flank of one immune-deficient and monitored for two weeks. Cell-lines of HLA-A24:02 subtype which demonstrated appreciable growth were selected for the study. Cell-lines were assigned APB groups: APB_{high}, APB_{int} and APB_{low}. In total 5 cell lines were selected for the experiment (two APB_{high} (SNU1750 and GSU), one APB_{int} (YCC21) and two APB_{low} cell lines (NCC 59 and SNU16)).

Study design

For each cell-line, 5 humanized-mice and 5 NSG mice were injected with the tumor cells and observed for one month. Mice were sacrificed at the end of one month, necropsies performed, and tumors harvested for further analysis. Parameters that were monitored included differential growth rate (tumor volume) between humanized and NSG mice, tumor size at end of the one month and histopathological analysis for immune cell infiltration, including immunohistochemistry.

Tumorigenesis in mice

For each cell-line, 2 x 10^6 of cells was resuspended in 50µl of sterile saline (NaCl 0.9%; Braun) and subcutaneously injected into the shaved right flank of each NSG and humanized-mouse. The tumor size was monitored twice weekly and the length

and width of tumor was measured using calipers. The tumor volume was calculated using the formula: tumor volume = (length x width^2) x 0.5. Tumor initiation was defined as detection of a tumor of at least 5 mm^3.

Immunohistochemistry (IHC) and Hematoxylin and Eosin (H&E) staining

IHC and H&E staining was performed on the FFPE tissue samples as previously described[12, 13, 14, 15, 16]. Tissue sections (4 μm thick) were labelled with antibodies targeting CD3 and CD8, as listed in the table below. Appropriate positive and negative controls were included. To score antibody-labelled sections, images were captured using an IntelliSite Ultra-Fast Scanner (Philips, Eindhoven, Netherlands). The percentage of cells displaying unequivocal staining of any intensity for CD3 or CD8 were determined by a pathologist blinded to clinicopathological and survival information. Tumor infiltrating lymphocytes (TILs) expressing CD3 or CD8 were identified within the intra-tumoral area defined as lymphocytes within cancer cell nests and in direct contact with tumour cells [16, 17, 18]. Quantification of TILs was determined by the percentage of the intra-tumoral areas occupied by the respective TIL population [16, 17, 18].

Antibody	**Source**	**Clone**	**Labelling Pattern**
CD3	Dako DKO.A045201	polyclonal	cytoplasm
CD8	Leica CD8-4B11-L-CE	4B11	cytoplasm

Bulk RNA-Seq of mouse tumors

RNA-Seq was performed from tumors harvested at the end of mouse experiment using standard pipelines previously described [19]. CIBERSORTx was used to estimate cellular proportions [20]. For reference signature gene matrices, the NSCLC PBMCs scRNAseq dataset provided with the CIBERSORTx suite were used.

Immunotherapy treated clinical cohorts

In a multi-centre, industry-academic collaborative effort, a cohort of immunotherapy treated samples were collected, to assess APB. A majority of the samples were from various ICI clinical trials being conducted by the respective groups. Academic

centres that contributed samples to this study include Samsung Medical Center, South Korea; Fondazione IRCCS Istituto Nazionale dei Tumori of Milan, Italy; Yonsei Cancer Center, South Korea; National University Cancer Institute, Singapore and National Cancer Center, Singapore. Industry collaborators included Roche/Genentech. For samples that had RNA-Seq performed by the respective centres, bioinformatic FASTQ files, along with correlative clinical data were transferred to Singapore for analysis. APB was quantified from RNA-Seq data using the method described earlier. For samples where only formalin-fixed paraffin embedded (FFPE) archival tissue were available, tissue blocks, or slides were shipped to Singapore. A custom-designed NanoString panel was performed on these FFPE samples. The contributing site, tumor-type, ICI treatment and type of transcriptomic analysis performed (RNA-Seq *vs.* NanoString) are listed in **Supplementary Table 5**. All patients were treated with ICI in the metastatic, second-line or beyond setting. A subgroup of the gastric cancer samples analysed in this study was reported as a preliminary analysis previously[3]. The chemotherapy cohort used in the analysis was from the "3G" multi-center international clinical trial [21]. Patients were assigned to one of two chemotherapy treatments: SOX (S-1 (an oral 5-fluoropyramidine pro-drug) and oxaliplatin) or SP (S-1 and cisplatin) based on a pre-determined genomic signature.

PDL1 Immunohistochemistry

PDL1 immunhistochemisty was performed using the Dako PD-L1 IHC 22C3 pharmDx kit (Agilent Technologies). PD-L1 protein expression was determined using CPS, which was the number of PD-L1 staining cells (tumour cells, lymphocytes, macrophages) divided by the total number of viable tumour cells, multiplied by 100.

MSI status

Tumour tissue MSI status was determined by both IHC for MLH1 and MSH2 and PCR analysis of 5 markers with mononucleotide repeats.

EBV status subtypes

EBV status was determined by EBV-encoded small RNA (EBER) in situ hybridization.

TCGA subtype definition

Gastric cancer subtypes defined by TCGA was based on DNA genomic alterations. These groups included EBV, MSI-H, CIN and genome stable tumors, which lack CIN and are heavily enriched in the diffuse histologic subtype. As a proxy for CIN, we stratified EBV negative, MSS tumors into CIN and genome stable based on their TP53 status as described previously [22]. Mutational signature analysis was performed using the deconstructSigs package (v1.6.0) in R.

NanoString analysis

The NanoString nCounter platform has been developed for gene and transcript expression analysis for use with FFPE-derived samples [23]. We have previously demonstrated good correlation between APB in RNA-Seq and NanoString [2, 3]. APB quantification utilizes 4,519 promoters identified in RNA-Seq for calculation. The NanoString nCounter platform allows for measurement of up to maximum of 800 probes or transcripts. From the 4,519 promoters used to calculate the APB algorithm, we identified the top 500 promoters from the pan-TCGA analysis to design alternate promoter probes. These probes were designed to predominantly bind to the unique first exon junctions, allowing for identifying and differentiating between alternate promoter transcripts. The probe design aimed to simulate the *proActiv* algorithm in identifying gain and loss promoter transcripts [1]. After running *in silico* QC on the full set of probes to look for potential probe-probe homology that might cause assay errors, a total of 437 alternate promoter probes (262 gain promoters and 175 lost promoters) were selected. Several immune gene related genes and other cancer related genes and exploratory probes were also included in the final panel design to test correlates. APB was calculated from the NanoString panel using the same parameters as RNA-Seq. Promoter counts were transformed based on weights established from median normalized promoter read count values (of the NanoString cohort). Median values of gain promoters were multiplied by 4, and loss promoters were divided by 4 to establish weights in the gastric cancer NanoString cohort. Transformed promoter read counts were assigned a point if the value was ≥ 1. The sum of the gain and lost promoter points was assigned the alternate promoter utilization burden (APB). Samples were classified into APB groups: Those in the top quartile of APB were classified as APB_{high}, in the lowest quartile as APB_{low}, while the

rest of the samples were classified as APB_{int}. The first NanoString panel was tested on 35 gastrointestinal tumors. Based on the training and probe binding quality from the first panel, a second panel using 169 probes (72 gain and 97 loss), was tested on a second cohort of gastric cancer samples (n = 54). APB was calculated using the same formula.

References

1 Demircioglu D, Cukuroglu E, Kindermans M, Nandi T, Calabrese C, Fonseca NA*, et al.* A Pan-cancer Transcriptome Analysis Reveals Pervasive Regulation through Alternative Promoters. Cell 2019;**178**:1465-77.e17.
2 Qamra A, Xing M, Padmanabhan N, Kwok JJT, Zhang S, Xu C*, et al.* Epigenomic Promoter Alterations Amplify Gene Isoform and Immunogenic Diversity in Gastric Adenocarcinoma. Cancer Discov 2017;**7**:630-51.
3 Sundar R, Huang KK, Qamra A, Kim KM, Kim ST, Kang WK*, et al.* Epigenomic promoter alterations predict for benefit from immune checkpoint inhibition in metastatic gastric cancer. Ann Oncol 2019;**30**:424-30.
4 Deng M, Bragelmann J, Kryukov I, Saraiva-Agostinho N, Perner S. FirebrowseR: an R client to the Broad Institute's Firehose Pipeline. Database : the journal of biological databases and curation 2017;**2017**.
5 Bhattacharya S, Dunn P, Thomas CG, Smith B, Schaefer H, Chen J*, et al.* ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. Scientific data 2018;**5**:180015.
6 Cortes-Ciriano I, Lee S, Park WY, Kim TM, Park PJ. A molecular portrait of microsatellite instability across multiple cancers. Nat Commun 2017;**8**:15180.
7 Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH*, et al.* The Immune Landscape of Cancer. Immunity 2018;**48**:812-30.e14.
8 Ghoshdastider U, Rohatgi N, Mojtabavi Naeini M, Baruah P, Revkov E, Guo YA*, et al.* Pan-cancer analysis of ligand-receptor crosstalk in the tumor microenvironment. Cancer research 2021.
9 Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 2018;**36**:411-20.
10 Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd*, et al.* Comprehensive Integration of Single-Cell Data. Cell 2019;**177**:1888-902.e21.
11 Zhang P, Yang M, Zhang Y, Xiao S, Lai X, Tan A*, et al.* Dissecting the Single-Cell Transcriptome Network Underlying Gastric Premalignant Lesions and Early Gastric Cancer. Cell Rep 2019;**27**:1934-47.e5.
12 Chew V, Chen J, Lee D, Loh E, Lee J, Lim KH*, et al.* Chemokine-driven lymphocyte infiltration: an early intratumoural event determining long-term survival in resectable hepatocellular carcinoma. Gut 2012;**61**:427-38.
13 Yeong J, Lim JCT, Lee B, Li H, Chia N, Ong CCH*, et al.* High Densities of Tumor-Associated Plasma Cells Predict Improved Prognosis in Triple Negative Breast Cancer. Frontiers in immunology 2018;**9**:1209-.
14 Yeong J, Thike AA, Lim JC, Lee B, Li H, Wong SC*, et al.* Higher densities of Foxp3+ regulatory T cells are associated with better prognosis in triple-negative breast cancer. Breast cancer research and treatment 2017;**23**:017-4161.
15 Yeong J, Tan T, Chow ZL, Cheng Q, Lee B, Seet A*, et al.* Multiplex immunohistochemistry/immunofluorescence (mIHC/IF) for PD-L1 testing in triple-negative breast cancer: a translational assay compared with conventional IHC. J Clin Pathol 2020.

16 Yeong J, Lim JCT, Lee B, Li H, Ong CCH, Thike AA*, et al.* Prognostic value of CD8 + PD-1+ immune infiltrates and PDCD1 gene expression in triple negative breast cancer. J Immunother Cancer 2019;**7**:34.
17 Hendry S, Salgado R, Gevaert T, Russell PA, John T, Thapa B*, et al.* Assessing Tumor-infiltrating Lymphocytes in Solid Tumors: A Practical Review for Pathologists and Proposal for a Standardized Method From the International Immunooncology Biomarkers Working Group: Part 1: Assessing the Host Immune Response, TILs in Invasive Breast Carcinoma and Ductal Carcinoma In Situ, Metastatic Tumor Deposits and Areas for Further Research. Advances in anatomic pathology 2017;**24**:235-51.
18 Hendry S, Salgado R, Gevaert T, Russell PA, John T, Thapa B*, et al.* Assessing Tumor-Infiltrating Lymphocytes in Solid Tumors: A Practical Review for Pathologists and Proposal for a Standardized Method from the International Immuno-Oncology Biomarkers Working Group: Part 2: TILs in Melanoma, Gastrointestinal Tract Carcinomas, Non-Small Cell Lung Carcinoma and Mesothelioma, Endometrial and Ovarian Carcinomas, Squamous Cell Carcinoma of the Head and Neck, Genitourinary Carcinomas, and Primary Brain Tumors. Advances in anatomic pathology 2017;**24**:311-35.
19 An O, Song Y, Ke X, So JB, Sundar R, Yang H*, et al.* "3G" Trial: An RNA Editing Signature to Guide Gastric Cancer Chemotherapy. Cancer research 2021;**81**:2788-98.
20 Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F*, et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol 2019;**37**:773-82.
21 Yong WP, Rha SY, Tan IB, Choo SP, Syn NL, Koh V*, et al.* Real-Time Tumor Gene Expression Profiling to Direct Gastric Cancer Chemotherapy: Proof-of-Concept "3G" Trial. Clinical cancer research : an official journal of the American Association for Cancer Research 2018;**24**:5272-81.
22 Kim ST, Cristescu R, Bass AJ, Kim KM, Odegaard JI, Kim K*, et al.* Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. Nat Med 2018;**24**:1449-58.
23 Chen X, Deane NG, Lewis KB, Li J, Zhu J, Washington MK*, et al.* Comparison of Nanostring nCounter(R) Data on FFPE Colon Cancer Samples and Affymetrix Microarray Data on Matched Frozen Tissues. PloS one 2016;**11**:e0153784.