Cell Genomics, Volume *3*

Supplemental information

Cystatin C is glucocorticoid responsive, directs recruitment of Trem2+ macrophages, and predicts failure of cancer immunotherapy

Sam O. Kleeman, Tuba Mansoor Thakir, Breanna Demestichas, Nicholas Mourikis, Dominik Loiero, Miriam Ferrer, Sean Bankier, Yosef J.R.A. Riazat-Kesh, Hassal Lee, Dimitrios Chantzichristos, Claire Regan, Jonathan Preall, Sarthak Sinha, Nicole Rosin, Bryan Yipp, Luiz G.N. de Almeida, Jeff Biernaskie, Antoine Dufour, Pinkus Tober-Lau, Arno Ruusalepp, Johan L.M. Bjorkegren, Markus Ralser, Florian Kurth, Vadim Demichev, Todd Heywood, Qing Gao, Gudmundur Johannsson, Viktor H. Koelzer, Brian R. Walker, Hannah V. Meyer, and Tobias Janowitz

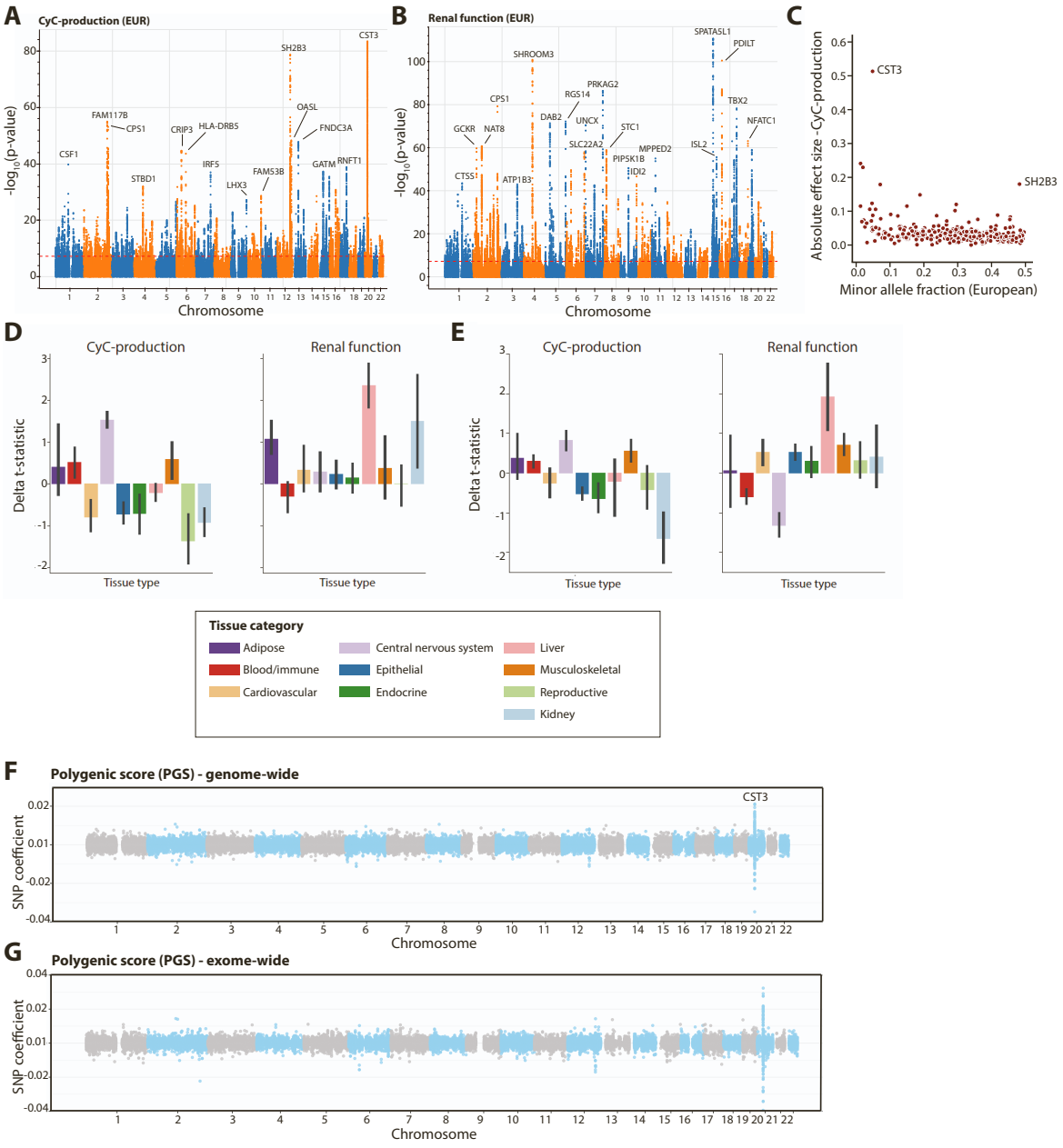

Figure S1 – Derivation of CyC-production polygenic score (PGS), Related to Figure 1. Summary statistics from (A) Cystatin C (CyC)-production and (B) Renal function latent traits in European UKB subjects, displayed a Manhattan plot. Loci with a p-value less than 1e-30 are annotated with gene name from OpenTargets V2G pipeline. (C) Relationship between effect size and minor allele frequency in CyC-production trait, annotated with outlier loci. Partitioned heritability analysis across multiple tissue types derived from (D) gene expression and (E) chromatin accessibility data. Delta t-statistic refers to change in enrichment t-statistic between measured eGFR-CyC summary statistics and latent CyC-production or renal function statistics. Errors bars signify 95% confidence interval. Coefficients for each SNP included in polygenic scores (PGS) generated using (F) HapMap SNPs (n=1,000,000) or (G) HapMap SNPs that can be reliably imputed from exome (n=300,000) sequencing data. CST3 locus on chromosome 20 is annotated.

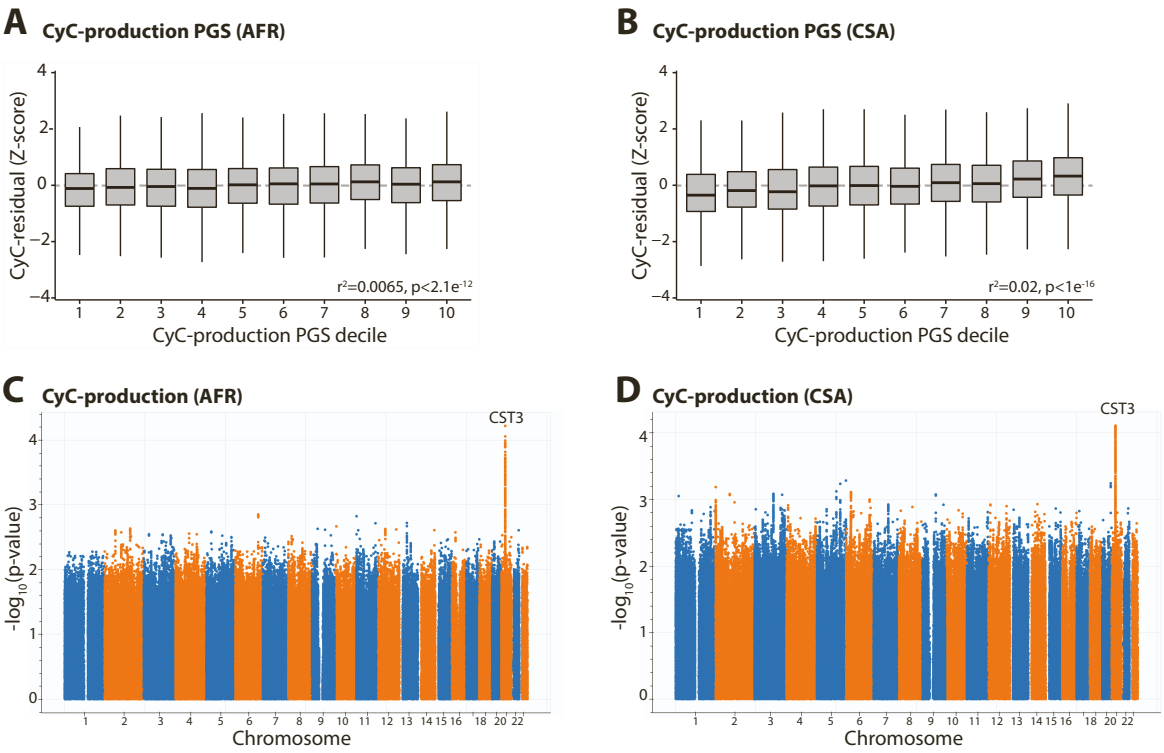

Figure S2 – Trans-ancestral portability of CyC-production polygenic score (PGS), Related to Figure 1. PGS derived in Biobank European training set (n=381,764) applied to subjects of (A) African (AFR, n=8152) and (B) Central and South Asian (CSA, n=9845) genetic ancestry. Boxplots show median (central line) with interquartile range (IQR, box) and extrema (whiskers at 1.5× IQR). Summary statistics from CyC-production latent trait in (C) AFR and (D) CSA genetic ancestry cohorts, derived from GWAS for eGFR-CyC and eGFR-creatinine followed by structural equation modeling. Results displayed as Manhattan plot; no loci reached genome-wide significance in latent trait analysis.

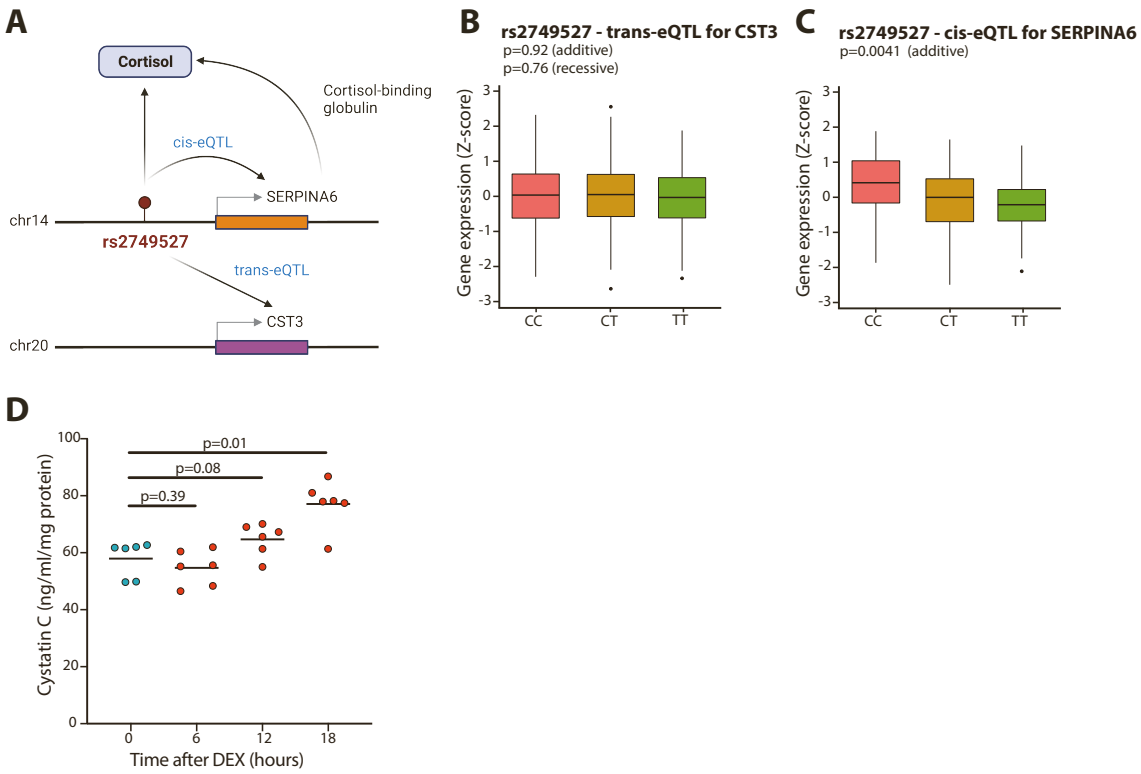

Figure S3 – Supporting data for glucocorticoid-mediated regulation of cystatin C, Related to Figure 3. (A) Role of single genetic instrument, rs2749527, as a trans-eQTL for *CST3* on chromosome 20 (in visceral adipose fat) and as a cis-eQTL for *SERPINA6* on chromosome 14 (in liver), which codes for cortisol-binding globulin, that is also significantly associated with morning plasma cortisol. Association analysis between rs2749527 and both (B) *CST3* gene expression in visceral adipose fat (VAF, n=381) and (C) *SERPINA6* gene expression in liver (n=175) in GTEX cohort (EUR ancestry). The discordance between GTEX and STARNET data is discussed in the text and Figure S5a. P-values for additive and recessive models are shown. Boxplots show median (central line) with interquartile range (IQR, box) and extrema (whiskers at 1.5× IQR). Outliers beyond 1.5× IQR are shown as dots. (D) Extracellular cystatin C concentrations in A549 cells normalized to cellular protein content after 0-, 6-, 12- or 18-hour treatment with 100nM dexamethasone (DEX). Each timepoint comprises 6 biological replicates. P-values refer to two-sided t tests.

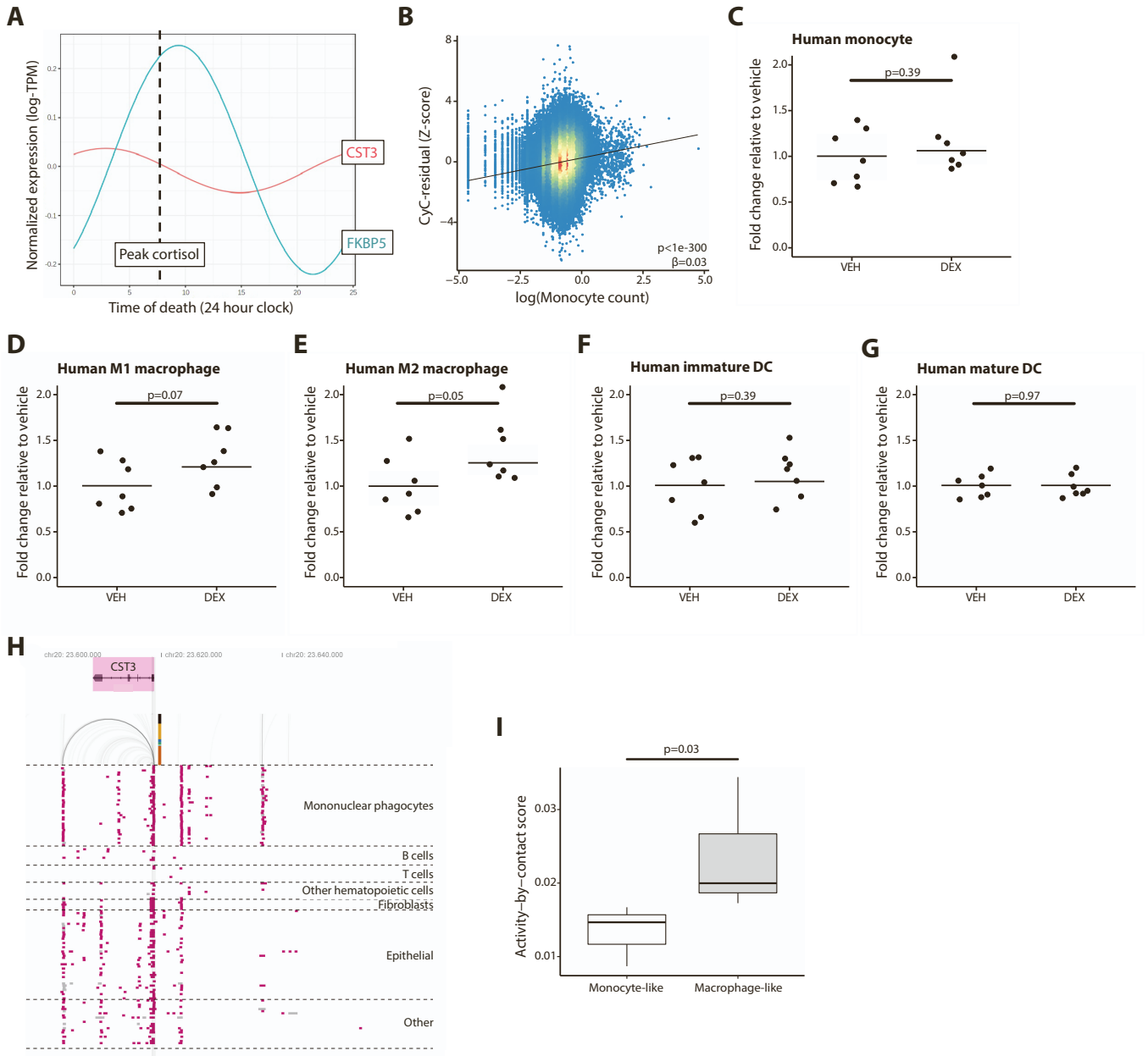

Figure S4 – Supporting data for glucocorticoid-mediated regulation of cystatin C in myeloid cells, Related to Figure 4-5. (A) Diurnal variation in *CST3* and *FKBP5* (canonical glucocorticoid response gene) derived from cosinor regression in GTEX spleen cohort, using time of death for each GTEX donor. (B) Significant positive correlation between logarithm of monocyte count and Z-score cystatin C (CyC)-residual in UKB cohort. Significance refers to multivariate regression including age, sex and body mass index (BMI). Fold change in extracellular CyC concentrations normalized to protein content in (C) primary human monocytes, as well as monocyte-derived (D) M1 and (E) M2 macrophages; (F) immature and (G) mature dendritic cells (DC) treated following 18-hour treatment with dexamethasone. All cell types were treated with 100nM dexamethasone (DEX) or vehicle control (VEH), except mature DCs which were treated with 10nM to minimize cytotoxicity. There are 7 biological replicates per group. (H) Visualization of predicted enhancer elements at *CST3* locus from activity-by-contact (ABC) model[68], showing distal enhancer element acting on cystatin C. Each row corresponds to a cell line-treatment pair and epithelial grouping includes cancer cell lines. (i) ABC model scores for distal enhancer in THP-1 cells - with (macrophage-like) or without (monocyte-like) PMA treatment. P-values refer to two-sided t tests unless otherwise stated.

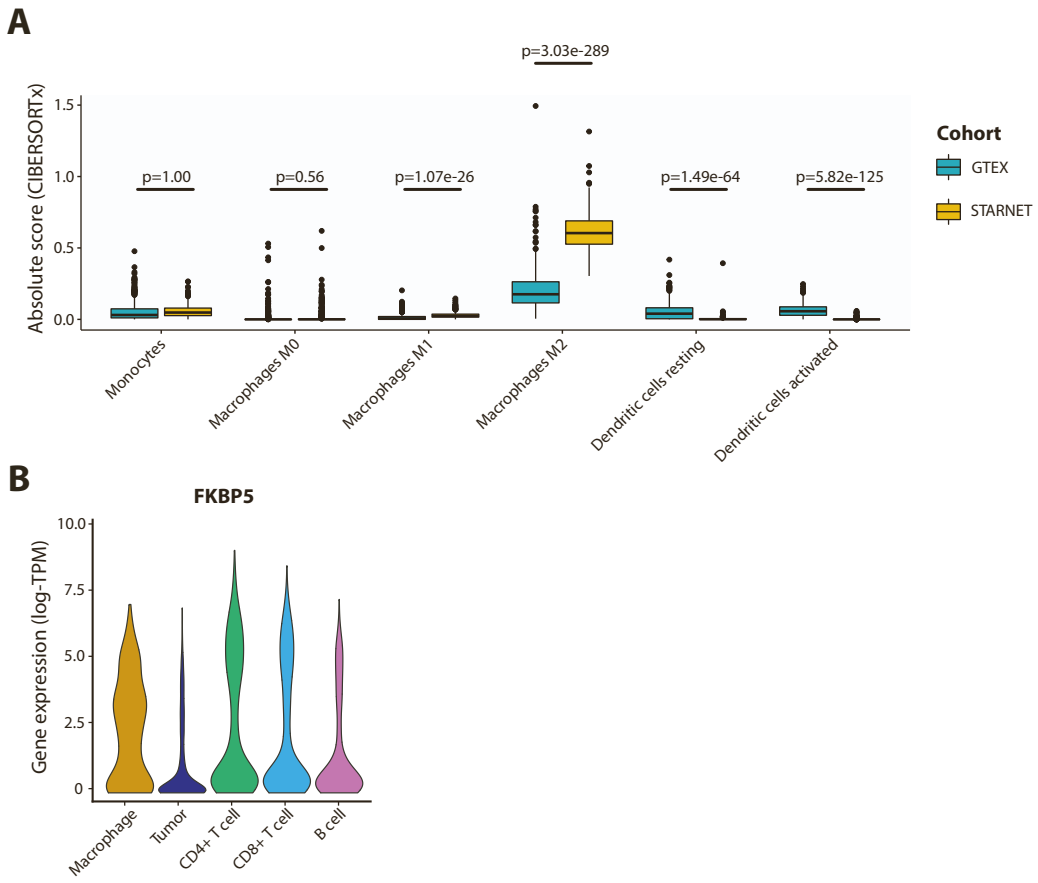

Figure S5 – Supporting data for trans-eQTL analysis, Related to Figure 3. (A) Inferred absolute myeloid cell composition from CIBERSORTx analysis (absolute mode) applied to visceral adipose tissue (VAF) from STARNET and GTEX cohorts. P values refer to t-tests with Bonferroni correction. While units are comparable between cell types and samples, they do not refer to an absolute cell fraction. Boxplots show median (central line) with interquartile range (IQR, box) and extrema (whiskers at 1.5× IQR). Outliers beyond 1.5× IQR are shown as dots. Marker genes used to define M0-like, M1-like (such as *CCL19*) and M2-like (such as *CCL18*) macrophages are described in the CIBERSORT manuscript[100]. (B) Single-cell *FKBP5* (canonical glucocorticoid receptor target) gene expression in each cell cluster in melanoma tumors (n=12) from Jerby-Anon et al[101]. Clusters defined by correlation to reference PBMC data[98], with unclassified cells that exhibit detectable clonal copy number variation classified as tumor. P-values refer to two-sided t tests.

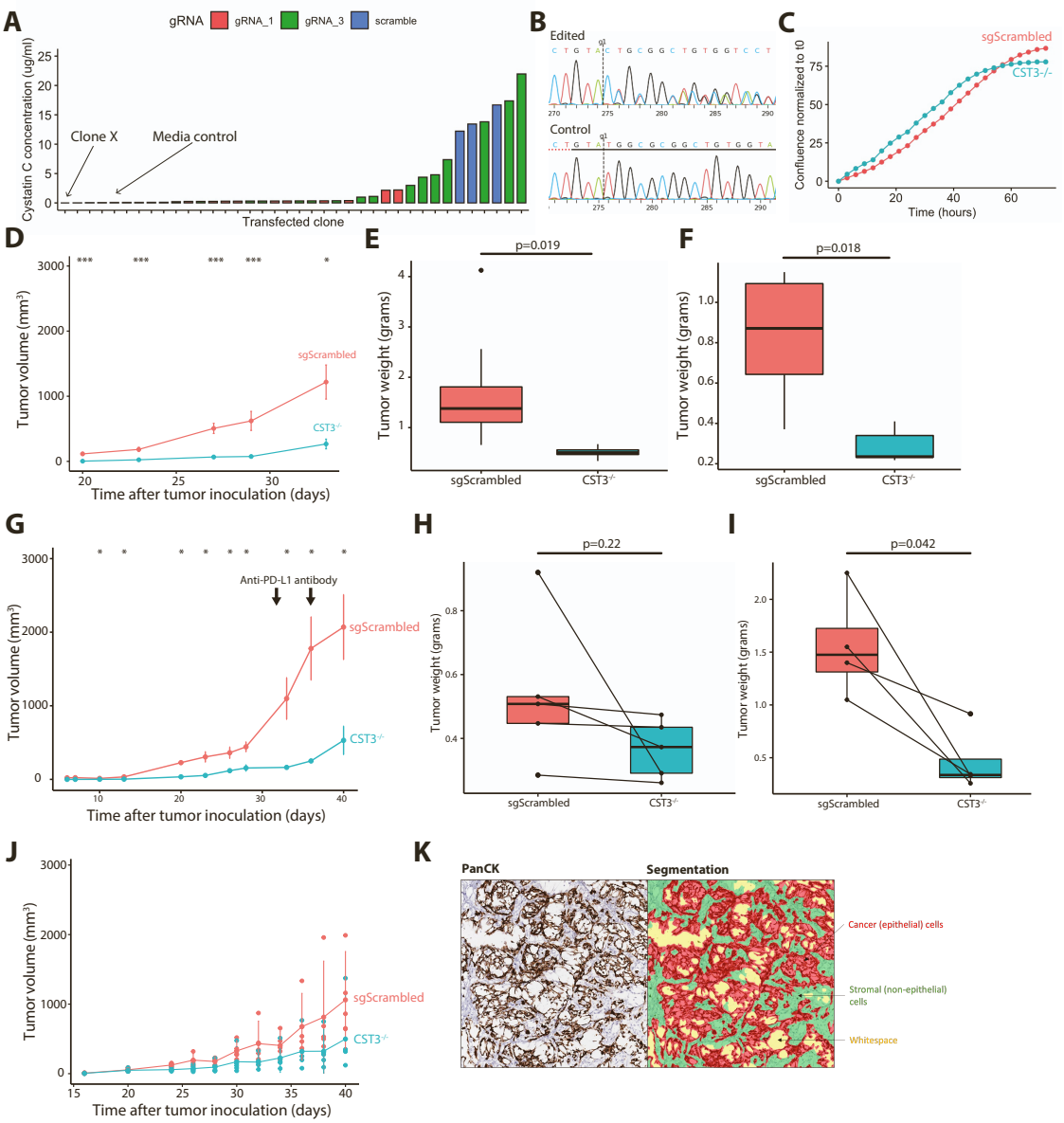

Figure S6 – Supporting data for effects of cancer cell-specific cystatin C knockout *in vivo*, Related to Figure 6. (A) Extracellular cystatin C (CyC) concentrations measured by ELISA in monoclonal cell populations derived from transfection of Mm1 cells with guide RNAs (gRNAs) specific to the *CST3* gene locus (gRNA_1 and gRNA_2, plus sgScrambled control). Clone X was selected on the basis of lowest extracellular CyC concentration. (B) Sanger sequencing trace showing high editing efficiency (>97%) at the predicted binding site for gRNA_1 in clone X. (C) Cell confluence kinetics for sgScrambled and *CST3*$^{-/-}$ (clone X) cells normalized to confluence at the start of the experiment. (D) Tumor growth curves (mean and standard error of the mean) for an independent replication with single-flank sgScrambled (n=10) and *CST3*$^{-/-}$ (n=10) tumors, 200,000 cells inoculated in right flank (cohort B). Endpoint tumor weights for (E) cohort A and (F) cohort B. (G) Tumor growth curves (mean and standard error of the mean) for an independent replication bi-flank paired sgScrambled (n=4) and *CST3*$^{-/-}$ (n=4) tumors, 100,000 cells inoculated in both flanks (cohort D). Mice received two doses of anti-PD-L1 antibody. P-values refer to paired two-sided t tests. Endpoint tumor weights for (H) cohort C and (I) cohort D. P-values refer to paired two-sided t tests. (J) Tumor growth surves for single-flank sgScrambled (n=10) and *CST3*$^{-/-}$ (n=10) tumors, 200,000 cells inoculated in right flank of Rag1-null mice. Error bars reflect confidence intervals, which overlap for all timepoints. (K) Representative PanCK immunohistochemistry annotated with automated segmentation of epithelial and non-epithelial compartments. P-values refer to two-sided t tests unless otherwise stated.

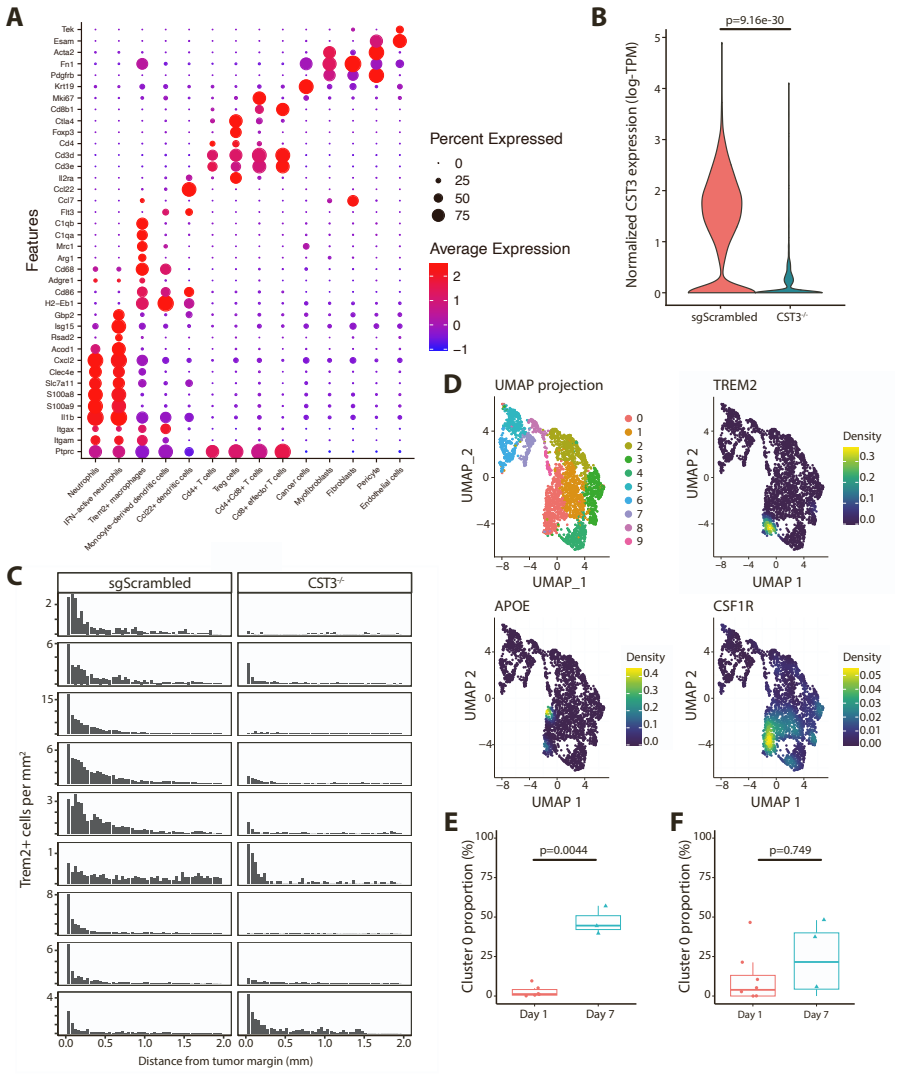

Figure S7 – Supporting data for role of Trem2+ cell population in phenotype of cystatin C-knockout tumors, Related to Figure 6. (A) Dotplot of marker gene expression (marker selection detailed in Table S7) used to define cell cluster identity in scRNA-seq of Mm1 tumors. Dot color signifies mean expression (log-TPM) while dot size signifies the proportion of each cell population that have detectable expression of each gene. (B) Violin plot showing cancer cell compartment-specific *CST3* gene expression in sgScrambled and CST3$^{-/-}$ tumor samples. P-value refers to a pseudobulk comparison. (C) Histogram summarizing number of Trem2+ cells per mm^2 in each 40µm window from the tumor margin in paired sgScrambled and CST3$^{-/-}$ tumor sections, each row corresponds to a mouse (n=9) with tumors inoculated in each flank. (D) UMAP plots of re-clustered CD14+ monocytes (n=10 clusters) from ICU-admitted COVID-19 patients[71] annotated with the *Nebulosa*[80] kernel function (facilitating imputation of cluster gene expression) corresponding to TREM2, APOE and CSF1R gene expression; indicating that cluster 0 comprises detectable Trem2+ monocytes. Proportion of cluster 0 monocytes at day 1 (within 72 hours of admission) and day 7 in ICU-admitted patients with COVID-19 who were treated (E) with or (F) without dexamethasone. P-value is adjusted p-value from linear model of logit-transformed proportions. P-values refer to two-sided t tests unless otherwise stated.

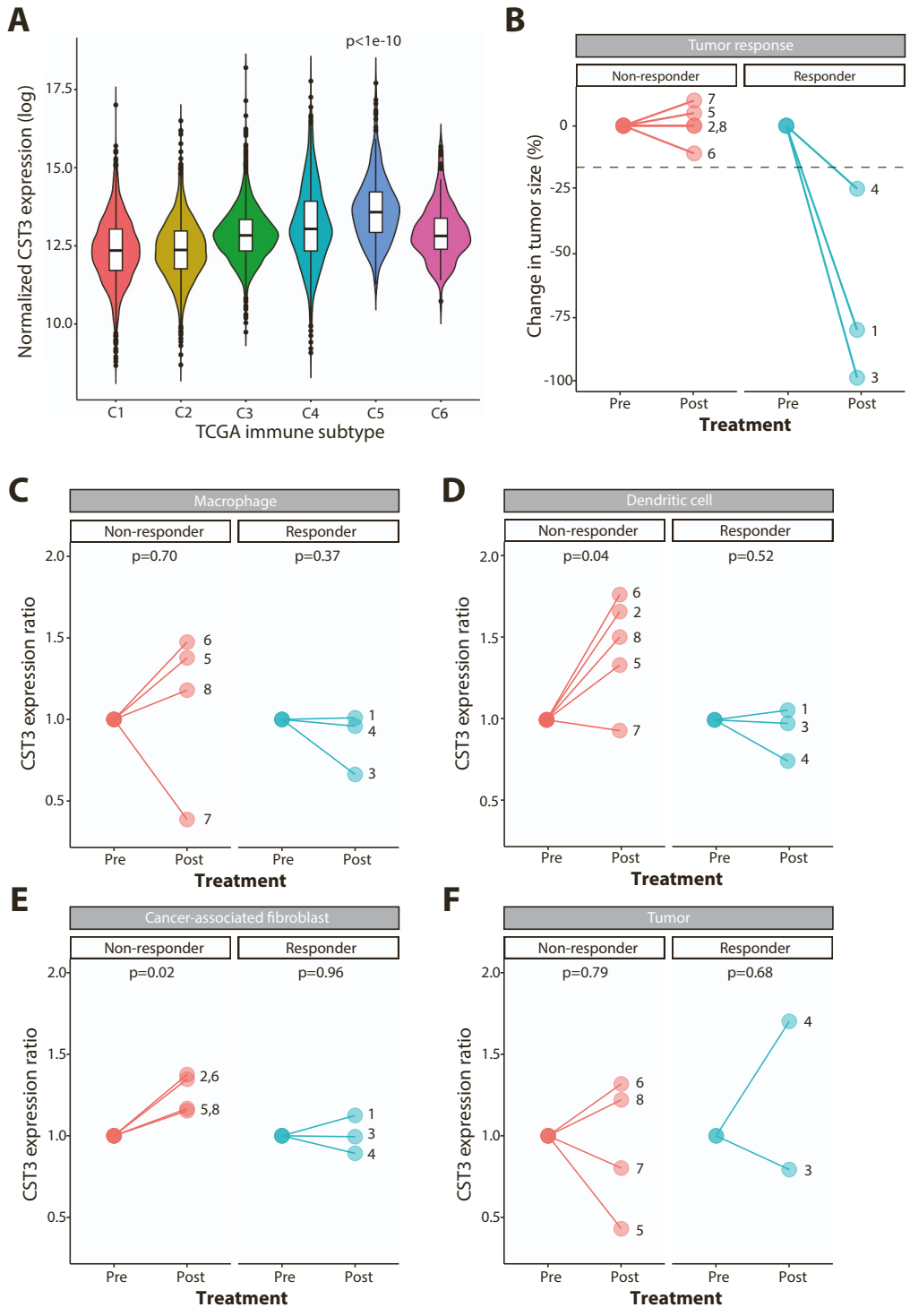

Figure S8 – Role of cystatin C in dynamic resistance to immunotherapy treatment, Related to Figure 6. (A) *CST3* gene expression (log-normalized) in each TCGA immune subtype (Thorssen et al[102]). Subtype 5 corresponds to 'immunological quiet' subtype, characterized by reduced lymphocyte and increased M2 macrophage responses. Significance level refers to one-way ANOVA with post-hoc Tukey's test. (B) Summary of radiological response data for patients with basal cell carcinoma (BCC, n=8) derived from Yost et al[103]. Response is nominally defined as tumor regression >25%. Numbers refer to patient IDs. Ratio of cluster-specific pseudo-bulk *CST3* gene expression (log-TPM) in paired biopsy samples pre/post anti-PD-1 treatment, segregated according to clinical response, for (C) macrophage, (D) dendritic cell, (E) cancer-associated fibroblast (CAF) and (F) tumor subsets. Uncorrected P values refer paired t-tests.

Table S1: Ancestry classification in UK Biobank, related to STAR Methods.

Ancestry code	Ancestry name	Number of subjects
EUR	European	456606
CSA	Central and South Asian	9845
AFR	African	8154
EAS	East Asian	2500
MID	Middle Eastern	843
AMR	Admixed American	707

Table S2: Summary statistics for panIO cohort, related to STAR Methods.

Table S2a: Summary of cohorts comprising pan-immunotherapy (panIO) cohort.

Accession	Cohort size (after QC)	Controlled ac	Data source	Marker papers (Pubmed ID)	Tumor type	Tumor status	Treatment
imvigor210	205	Yes	EGA	29443960	Urothelial	Metastatic	Atezolizumab
phs000452	230	Yes	dbGaP	26359337, 31792460	Melanoma	Metastatic	Pembrolizumab, Nivolumab, Ipilimumab
phs001493	60	Yes	dbGaP	29301960	ccRCC	Metastatic	Nivolumab
phs000980	18	Yes	dbGaP	25765070	NSCLC	Metastatic	Pembrolizumab
phs001041	51	Yes	dbGaP	25409260	Melanoma	Metastatic	Ipilimumab
PRJNA312948	26	No	SRA	26997480	Melanoma	Metastatic	Pembrolizumab, Nivolumab
PRJNA356761	61	No	SRA	29033130	Melanoma	Metastatic	Nivolumab
phs001565	34	Yes	dbGaP	30150660	Urothelial, Melanoma, HNSCC	Metastatic	Various

Table S2: Summary statistics for panIO cohort, related to STAR Methods.

Table S2b: Summary of patient demographics in pan-immunotherapy (panIO) cohort. For details on the individual cancer immunotherapy trial cohorts making up this data see Table S1.

Baseline characteristics	**Subset**	**Statistic**
Age, years	Median	61
	Range	18-89
Gender, n (%)	Female	214 (31)
	Male	471 (69)
Tumor type, n (%)	Melanoma	373 (54)
	Urothelial	221 (32)
	Renal cell carcinoma (RCC)	60 (9)
	Non-small-cell lung cancer (NSCLC)	19 (3)
	Head and neck squamous cell carcinoma (HNSCC)	12 (2)
Treatment, n (%)	Anti-PD-1	320 (47)
	Anti-PD-L1	207 (30)
	Anti-CTLA-4	145 (21)
	Combined PD-1/PD-L1 and CTLA-4	13 (2)
Best overall response (RECIST), n (%)	Complete response (CR)	47 (7)
	Partial response (PR)	128 (19)
	Stable disease (SD)	125 (18)
	Progressive disease (PD)	301 (44)
	Unavailable	84 (12)
Durable clinical benefit, n (%)	Yes	285 (42)
	No	385 (56)
	Unavailable	15 (2)

Table S3: Genome-wide association analysis quality control statistics, related to STAR Methods.

Analysis	Population	Lambda-GC	Attenutation ratio
eGFR-Creatinine	EUR	1.76	0.12
	AFR	1.04	0.67
	CSA	1.03	0.31
eGFR-CyC	EUR	2	0.1
	AFR	1.04	0.52
	CSA	1.04	0.31

Table S5: Primer sequences for RT-qPCR, related to STAR Methods.

Gene	Forward	Reverse
CST3	GCCACATCTGAAAAGGAAAGCA	GCGTCCTGACAGGTGGATTT
PPIA	GCCACCGCCGAGGAAAA	CGACGGCAATGTCGAAGAAC
GUSB	TGCGTAGGGACAAGAACCAC	TGTGAGCGATCACCATCTTCA
RPL19	ACATGGGCATAGGTAAGCGG	CCGGCGCAAAATCCTCATTC
RPL15	CTACAAGGCCAAGCAAGGTTAC	GGACAGGCTTGCCGTAAGTT

Table S6: Guide RNA (gRNA) sequences for CRISPR-knockout, related to STAR Methods.

Guide	Sequence
Non-targeting (sgScrambled)	GCGAGGTATTCGGCTCCGCG
CST3_1	GTACCACAGCCGCGCCATAC
CST3_2	AAAACAAGGCCCGCGAATGT

Table S7: Summary of marker genes (positive and negative) used for scRNA-seq cell type identification, related to STAR Methods.

Cell population	Positive markers	Negative markers
Neutrophils	CD45; Cd11b; Cd11c; Il-1B; S100a9; S100a8; Slc7a11; Clec4e; Cxcl2; Acod1	
IFN-activated neutrophils	CD45; Cd11b; Cd11c; Il-1B; S100a9; S100a8; Slc7a11; Clec4e; Cxcl2; Acod1; viperin(rsad2); Isg15; Gbp2;	
Trem2+ macrophages	CD45; Cd11b; Cd11c; Il-1B; H2-Eb1; CD86; Slc7A11; Adgre1; Clec4e; Cxcl2; CD68; Arg1; CD206; C1qa; C1qb	
Monocyte-derived dendritic cells	CD45; Cd11b; Cd11c; Il-1B; H2-Eb1; CD86; Flt3; CD68	
Ccl22+ dendritic cells	CD45; Il-1B; H2-Eb1; CD86; Slc7A11; CCL7; CCL22; Flt3; CD25	CD11b; CD11c
CD4 T cells	CD45; CD3e; CD3d; CD4	CD8; Foxp3
T regulatory cells	CD45; CD3e; CD3d; CD4; Foxp3; CTLA4	
CD4+ CD8+ T cells	CD45; CD3e; CD3d; CD4; CD8b1; Mki67; CTLA4; CD25	
CD8+ Effector T cells	CD45; CD3e; CD3d; CD8b1	CD4
Cancer cells	Mki67; Krt19	CD45
Myofibroblasts	Pdgfrb; Fn1; Acta2; Esam	CD45
Pericytes	Pdgfrb; Acta2; Fn1	CD45
Fibroblasts	Pdgfrb; Fn1	CD45
Endothelial Cells	Esam; Tek	CD45