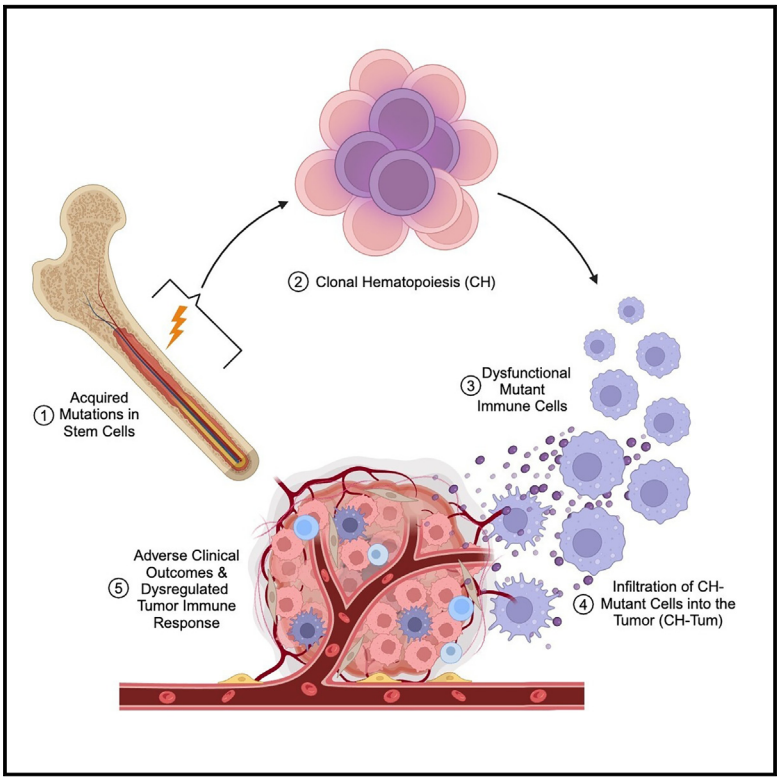
Cell Reports
Medicine

Report

Inflammatory reprogramming of the solid tumor microenvironment by infiltrating clonal hematopoiesis is associated with adverse outcomes

Graphical abstract



Highlights

- Clonal hematopoiesis (CH) is common in patients with cancer and can infiltrate tumors
- Infiltration of CH-mutant cells into the tumor predicts adverse clinical outcomes
- CH associates with dysregulated inflammatory gene expression and immune infiltrates
- In glioblastoma, CH shapes an aggressive, mesenchymal tumor phenotype

Authors

Marco M. Buttigieg, Caitlyn Vlasschaert, Alexander G. Bick, Robert J. Vanner, Michael J. Rauh

Correspondence

robert.vanner@uhn.ca (R.J.V.), rauhm@queensu.ca (M.J.R.)

In brief

Buttigieg et al. demonstrate that the infiltration of somatically mutated leukocytes—clonal hematopoiesis (CH)—into the tumor microenvironment can lead to poor outcomes across solid tumors. By leveraging tumor multi-omics data, they connect tumor-infiltrating CH cells with inflammation, highlighting immune genetics as a key consideration in precision oncology.

Buttigieg et al., 2025, Cell Reports Medicine 6, 101989
March 18, 2025
https://doi.org/10.1016/j.xcrm.2025.101989

CellPress

Cell Reports Medicine

CellPress
OPEN ACCESS

Report

Inflammatory reprogramming of the solid tumor microenvironment by infiltrating clonal hematopoiesis is associated with adverse outcomes

Marco M. Buttigieg,[1] Caitlyn Vlasschaert,[2] Alexander G. Bick,[3,4] Robert J. Vanner,[5,6,*] and Michael J. Rauh[1,2,7,*]
[1]Department of Pathology & Molecular Medicine, Queen's University, Kingston, ON, Canada
[2]Department of Medicine, Queen's University, Kingston, ON, Canada
[3]Vanderbilt Genetics Institute, Vanderbilt University School of Medicine, Nashville, TN, USA
[4]Division of Genetic Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA
[5]Department of Medical Oncology and Hematology, Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada
[6]Institute of Medical Science, Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[7]Lead contact
*Correspondence: robert.vanner@uhn.ca (R.J.V.), rauhm@queensu.ca (M.J.R.)
https://doi.org/10.1016/j.xcrm.2025.101989

SUMMARY

Clonal hematopoiesis (CH)—the expansion of somatically mutated hematopoietic cells—is common in solid cancers. CH is associated with systemic inflammation, but its impact on tumor biology is underexplored. Here, we report the effects of CH on the tumor microenvironment (TME) using 1,550 treatment-naive patient samples from the Clinical Proteomics Tumor Analysis Consortium (CPTAC) cohort. CH is present in 18.3% of patients, with one-third of CH mutations also detectable in tumor-derived DNA from the same individual (CH-Tum), reflecting CH-mutant leukocyte infiltration. Across cancers, the presence of CH-Tum is associated with worse survival outcomes. Molecular analyses reveal an association between CH-Tum and an immune-rich, inflammatory TME that is notably distinct from age-related gene expression changes. These effects are most prominent in glioblastoma, where CH correlates with pronounced macrophage infiltration, inflammation, and an aggressive, mesenchymal phenotype. Our findings demonstrate that CH shapes the TME, with potential applications as a biomarker in precision oncology.

INTRODUCTION

Clonal hematopoiesis (CH) describes the clonal expansion of somatically mutated hematopoietic stem cells (HSCs) with a fitness advantage.[1–3] CH is detected in the blood of 20%–30% of patients with solid cancers, which is considerably higher than in comparably aged, cancer-free individuals.[4,5] CH driver mutations, including inactivation of the most common driver genes *DNMT3A* or *TET2*, are associated with systemic inflammation.[1,6,7] Accordingly, CH is associated with excess morbidity and mortality attributable to numerous chronic diseases[1,8–12]; however, its impact on the tumor microenvironment (TME) and cancer outcomes remains unclear.[13]

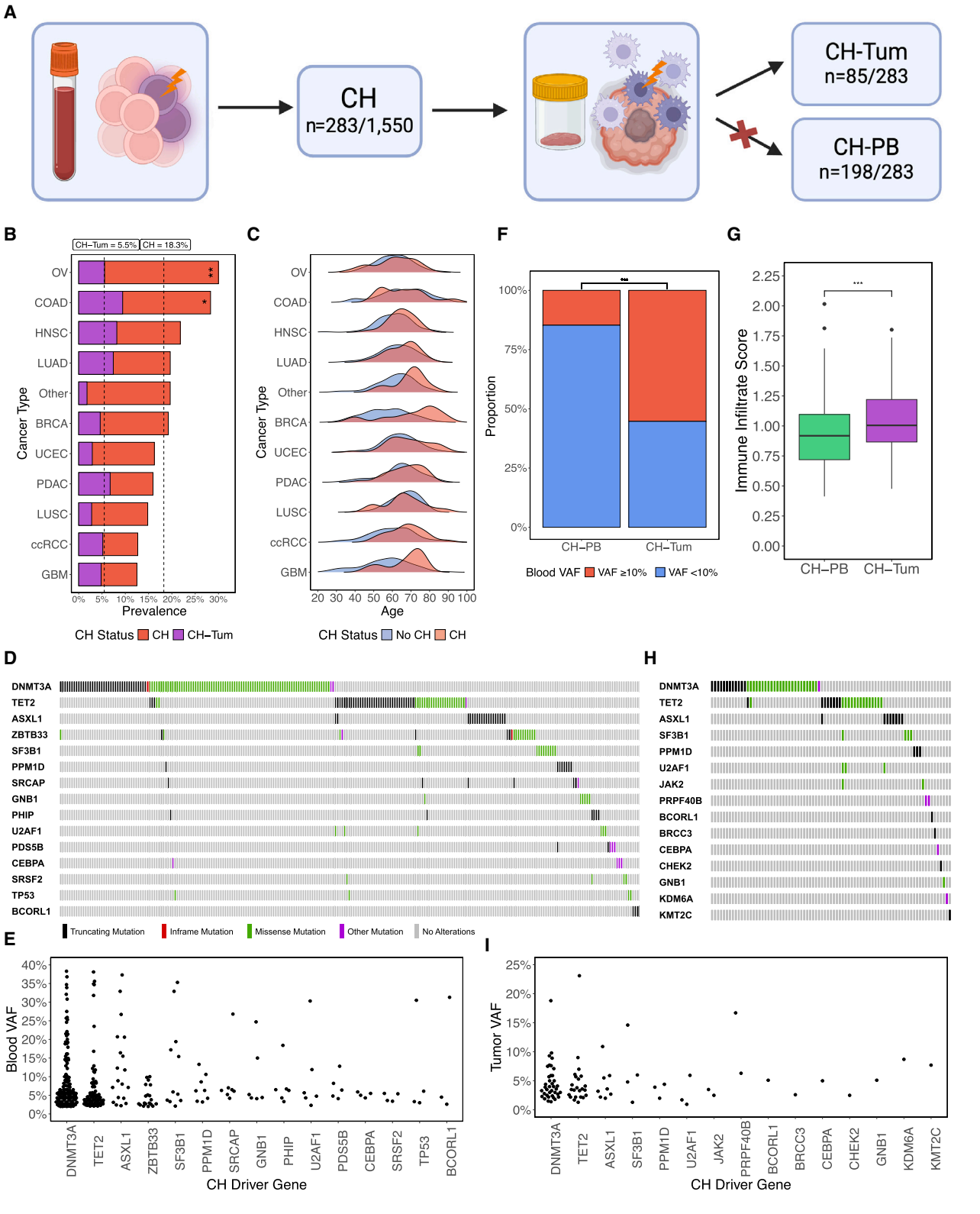
CH has been associated with an increased risk of some solid cancers in several cohorts.[14–20] The responsible mechanisms are unclear, but it is possible that CH accelerates inflammatory tumorigenesis in certain contexts like chronic lung and liver disease.[19,20] Furthermore, associations between CH and survival are heterogeneous and context dependent, with favorable effects reported in patients with metastatic colorectal cancer, no effect in some single-center cohorts, and negative effects in pan-cancer cohorts subject to confounding by disease stage and treatment history.[4,18,21–24] While CH in the human TME may influence inflammation in advanced tumors,[25] the evidence is limited and clinical relevance remains uncertain.

Herein, we present a pan-cancer study of CH as a feature of solid tumor biology, studying 1,550 primary, treatment-naive patients from the Clinical Proteomics Tumor Analysis Consortium (CPTAC). By integrating genomic, transcriptomic, proteomic, and clinical data, we identify the presence of CH mutations in tumor sequencing (CH-Tum) as a critical determinant of the effects of CH in solid cancer, highlighting its potential as a biomarker for precision oncology.

RESULTS

CH mutations are commonly detected in the blood and tumors of treatment-naive patients with cancer

We screened primary, treatment naive patients with cancer in the CPTAC cohort (n = 1,550; Table S1; STAR Methods) for CH by re-analyzing peripheral blood whole-exome sequencing (WES) data using our established protocol described by Vlasschaert et al.[26] Defining CH, also known as CH of indeterminate potential, as the presence of a somatic driver mutation in peripheral blood WES with variant allele fraction (VAF) $\geq$2% (Figure 1A), the prevalence of CH in the CPTAC cohort was 18.3% (n = 283;



(legend on next page)

Figure 1B). CH was significantly associated with patient age across the CPTAC cohort ($p = 1.6 \times 10^{-9}$, rank-sum test; Figures 1C, S1A, and S1B), and in all cancers except colorectal adenocarcinoma (COAD; $p = 0.69$, rank-sum test). There were significant variations in the CH prevalence across different cancer types in CPTAC, even after accounting for patient age and gender. Compared to the cohort average weighted by cancer type, high-grade serous ovarian cancer (OV) and COAD had higher rates of CH (OV odds ratio [OR] = 2.35 [1.41–3.91], COAD OR = 1.61 [1.04–2.48], logistic regression). CH was not associated with tumor stage at the time of diagnosis (OR = 1.26 [0.98–1.62], ordered logistic regression; Figure S1C). The mutational spectrum of CH was consistent with previous reports (Figures 1D and 1E; Table S2). Mutations in *DNMT3A* (37.9% of CH; $n = 132$) and *TET2* (20.7%; $n = 72$) composed the majority of CH drivers, with VAF ranging from 2%–38% (median VAF = 5%). Most patients presented with only one CH mutation (81.3%; $n = 230$), but 53 patients (18.7%) had between 2 and 4 CH drivers (Figure S1D).

CH mutations are detectable in tumor-derived DNA sequencing.[4,5,27] Since CH mutations are of hematopoietic origin, we posited that their identification in tumor sequencing reflects a greater degree of infiltration by CH-mutant leukocytes into the TME. We further hypothesized that this could identify cases where CH most potently impacts tumor biology. Following the identification of putative CH driver mutations in blood sequencing, we defined the detection of the same CH driver mutations in tumor sequencing, after accounting for potential tumor-derived and germline mutations, as CH-Tum (Figure 1A; STAR Methods). The prevalence of CH-Tum was 5.5% ($n = 85$), accounting for 30% of patients with CH (Figure 1B). Mutations in the remaining 70% of patients with CH solely found in peripheral blood are referred to as CH-PB (Figure 1A). Like CH, CH-Tum increased significantly with patient age ($p = 7.9 \times 10^{-5}$, rank-sum test; Figure S1B). Among those with CH, the presence of CH-Tum did not vary by cancer type after accounting for age (Figure 1B). In a multivariate model, we found that the detection of CH-Tum was associated with CH mutation VAF $\geq$10% in the peripheral blood (OR = 9.21 [4.58–18.6], logistic regression; Figure 1F) and degree of immune infiltration in the tumor sample as quantified by CIBERSORTx (OR = 6.69 [1.96–22.9], logistic regression; Figure 1G). Immune infiltrate scoring did not significantly differ based on peripheral blood VAF in CH-Tum or CH-PB (CH-Tum $p = 0.19$, CH-PB $p = 0.2$, rank-sum test; Figure S1E), but when separating between VAF $\geq$10% and <10%, immune infiltrate was significantly higher in CH-Tum relative to CH-PB in both conditions (VAF $\geq$10% $p = 0.02$, VAF <10% $p = 0.002$, rank-sum test; Figure S1F). This suggests that CH-Tum captures both tumor-immune infiltration and CH clone size, rather than acting as a surrogate for large CH clones in the blood. Like CH, CH-Tum was not associated with tumor stage at the time of diagnosis (OR = 1.41 [0.94–2.13], ordered logistic regression; Figure S1G).

The distribution of driver mutations in CH-Tum closely resembled that of CH, with *DNMT3A* (40% of CH-Tum, $n = 38$) and *TET2* (24.2%, $n = 23$) predominating (Figures 1H and 1I; Table S2). There were no discernable patterns of enrichment of any prominent driver genes in CH-Tum compared to CH. Mutation VAF in CH-Tum ranged from 1%–23% and was positively correlated with peripheral blood VAF (median tumor VAF = 4%, median blood VAF = 12%; Spearman's rho = 0.54, $p = 1.9 \times 10^{-8}$, Spearman's rank correlation; Figure S1H).

We also evaluated 8,927 matched blood and tumor samples collected for The Cancer Genome Atlas (TCGA) project ($n = 8,927$; Table S1) from 32 cancer types. CH driver mutations were found in the peripheral blood WES of 780 individuals, yielding a CH prevalence of 8.74% (median VAF = 6.7%; Table S2). Of the 780 TCGA participants with CH, we subsequently identified CH-Tum in 122 (15.6%), representing 1.3% of the entire cohort (median tumor VAF = 5%, median blood VAF = 17%; Table S2).

The prevalence of CH and CH-Tum in TCGA was lower than the CPTAC cohort and prior cancer cohorts profiled with targeted sequencing.[4,5,27] Patients in CPTAC were significantly more likely to have CH, *DNMT3A*-CH, and *TET2*-CH, even after accounting for differences in patient age and gender (CH: OR = 1.93 [1.63–2.27]; *DNMT3A*-CH: OR = 2.45 [1.92–3.12]; *TET2*-CH: OR = 3.74 [2.60–5.38], logistic regression; Figures S1I–S1K). Given the reported coverage gaps in exome capture kits used by the TCGA sequencing sites,[28–30] we compared the depth of coverage in CPTAC and TCGA peripheral blood WES samples (Table S3). Across coding regions of CH driver genes considered in this study, CPTAC samples had a median average depth of 259x, which was significantly higher than the average depth in TCGA samples of 94x ($p < 1 \times 10^{-308}$, rank-sum test; Figure S1L). This was present for *DNMT3A* (CPTAC median coverage = 215x, TCGA median coverage = 77x; $p < 1 \times 10^{-308}$, rank-sum test; Figure S1M) and even more pronounced for *TET2* (CPTAC median coverage = 312x, TCGA median coverage = 45x; $p < 1 \times 10^{-308}$,

Figure 1. Characteristics of CH and CH-Tum in treatment-naive patients with solid cancer

(A) Schematic overview of CH calling and definitions of CH-Tum and CH-PB. Any patient with CH driver mutation(s) in peripheral blood WES was classified as having CH. Of those with CH, if the same mutation(s) were subsequently detected in tumor WES, they were labeled as CH-Tum. If the peripheral blood mutation(s) were not subsequently detected in the tumor sample, they were labeled CH-PB. Created with BioRender.

(B) Prevalence of CH and CH-Tum by cancer type.

(C) Age distributions of patients with and without CH by cancer type.

(D) The mutational spectrum of CH.

(E) Peripheral blood VAFs of CH mutations by driver gene.

(F) Proportions of large (VAF $\geq$10%) and small (VAF 2%–10%) CH clones that make up CH-PB and CH-Tum.

(G) Immune infiltrate scoring by CIBERSORTx in CH-Tum and CH-PB.

(H) The mutational spectrum of CH-Tum.

(I) Tumor VAFs of CH-Tum mutations by driver gene.

*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$

rank-sum test; Figure S1N), with a number of TCGA samples receiving no coverage at all in these critical CH driver genes. Furthermore, although coverage depth did not differ based on CH status in CPTAC, inconsistencies in coverage within TCGA cohort led to the average depth being significantly higher in samples identified to have CH (CPTAC: $p = 0.45$; TCGA: $p = 2 \times 10^{-17}$, rank-sum test; Figures S1O and S1P). Noting these limitations, we largely excluded this cohort from subsequent analyses.

CH-mutant leukocyte infiltration of the TME is associated with adverse clinical outcomes

In Cox proportional hazard regression analyses adjusted for age, gender, cancer type, smoking status, and disease stage (hereafter, "adjusted"), CH status was not associated with overall survival (OS) in the CPTAC cohort in participants where survival data were available (univariate hazard ratio [HR] = 1.07 [0.82–1.40], multivariate HR = 1.12 [0.83–1.50]; Figure 2A; STAR Methods). The presence of CH-Tum, but not CH-PB, was associated with worse OS in the CPTAC cohort (CH-Tum univariate HR = 1.61 [1.07–2.42], CH-Tum multivariate HR = 1.74 [1.13–2.69]; CH-PB univariate HR = 0.88 [0.63–1.23], CH-PB multivariate HR = 0.90 [0.63–1.29]; Figure 2B). Unlike CH-Tum, immune infiltration was not independently associated with OS (univariate HR = 1.29 [0.88–1.90]; Figure S2A), nor did its inclusion in multivariate analysis alter the survival association of CH-Tum (CH-Tum HR = 1.83 [1.18–2.84]) across all cancers. Additionally, there was no significant interaction between CH-Tum and immune infiltrate scoring in an adjusted model ($p_{int} = 0.8$; Figure S2B), suggesting that the effects of CH on survival across cancers are specifically related to CH-Tum and not elevated immune infiltration. To explore the potential cause of this excess mortality, the last known tumor status at follow-up was modeled for CH and CH-Tum in multivariate analysis. CH-Tum, but not CH, was associated with a reduced likelihood of being classified as tumor-free at the latest follow-up (CH OR = 0.67 [0.43–1.08], CH-Tum OR = 0.38 [0.17–0.81]; Figure 2C).

We reasoned that associations between CH-Tum and survival could be captured even in cohorts with low DNA sequencing coverage, or without DNA sequencing, using a gene expression signature associated with CH. After identifying differentially expressed genes in CH-Tum versus no CH (Figure 3A), we developed a 23-gene prognostic signature score (PSS) using lasso penalized Cox regression (Figure 2D; STAR Methods) that was enriched in CH-Tum compared to CH-PB ($p_{adj} = 0.004$, Dunn's test) and no CH ($p_{adj} = 0.004$, Dunn's test; Figure S2C) and was associated with worse OS in adjusted analysis across CPTAC cancers (PSS above median HR = 2.32 [1.55–3.47]; Figure 2E). We then applied the PSS to TCGA cohort to determine whether the transcriptomic changes associated with CH-Tum were present and associated with OS. Across TCGA cancers, including cancer types not represented in CPTAC, the PSS was enriched in samples where CH-Tum was detected, compared to both CH-PB ($p_{adj} = 0.007$, Dunn's test) and no CH ($p_{adj} = 0.013$, Dunn's test; Figure S2D). The PSS was also associated with worse OS in multivariate analysis of TCGA cohort (PSS above median HR = 1.27 [1.14–1.42]; Figure 2F). These findings suggest consistent molecular changes in the TME that occur with CH-Tum across cancers and that these changes associate with poor clinical outcomes.

CH is associated with an inflamed TME

Next, we explored the impact of CH on tumor biology using bulk tumor RNA sequencing and proteomics. Compared to tumors in patients without CH, there were 363 and 301 differentially expressed genes with CH and CH-Tum, respectively, after accounting for gender and cancer type (Figures 3A and S3A). In both CH and CH-Tum, we observed the upregulation of genes related to immunity and tumor-immune interactions including *SPINK1*,[32–34] cytokines *IL1A*, *IL17REL*, and *CXCL6*, and S100 alarmins. *CRP*, encoding a marker of systemic inflammation enriched in CH,[35] was also upregulated. Of these differentially expressed genes, 96 were shared across CH and CH-Tum-positive tumors compared to no CH, with the 64 shared upregulated genes exhibiting significantly greater log fold changes in the CH-Tum condition ($p = 8.5 \times 10^{-9}$, pairwise rank-sum test; Figure S3B). Despite limitations with CH detection in TCGA cohort, multiple differentially expressed genes from CPTAC were also found in CH-positive tumors from TCGA (Figures S3C and S3D). 54 genes were similarly dysregulated with CH-Tum in CPTAC and TCGA, including the upregulation of *CTLA4*, *LILRA2/A5/B2*, *SPINK1*, *TLR2*, and *TLR8*. Functional profiling of the genes upregulated with CH-Tum in both datasets showed an enrichment of pathways related to antigen presentation and interleukin (IL)-6 signaling (Figure 3B).

We then performed gene set enrichment analysis (GSEA) using differential gene expression and mass-spectrometry-based protein abundance analysis of tumor samples in the CPTAC dataset based on CH, CH-Tum, and CH-PB, as well as CH-Tum with *DNMT3A*, *TET2*, and other (non-*DNMT3A/TET2*) driver mutations. In a pan-CPTAC analysis controlled for gender and cancer type, CH-Tum was associated with transcriptomic enrichment of immune signatures in tumors compared to patients without CH (Figure 3C; Table S4). Contrarily, CH-PB was associated with decreased inflammatory gene expression across cancers at the transcriptomic level (Figure 3C). In addition to immune-related signaling, CH-Tum also associated with enrichment of canonical cancer pathways including angiogenesis, epithelial-to-mesenchymal transition, metabolism, and oncogenic signaling that may drive tumor growth and progression, underlying adverse clinical associations (Figure 3D). Since CH has been correlated with survival in patients treated with immune checkpoint inhibitors (ICI),[22] we tested for enrichment of four gene signatures predictive of anti-PD-1 ICI response.[36–38] CH-Tum showed significant enrichment of ICI response signatures at the transcriptomic level (Figure 3E; Table S4). Limited concordance between RNA and proteomics across CPTAC (Figures S3E–S3G) may be related to reduced gene-level proteomics coverage as well as sampling bias, particularly within the uterine corpus endothelial carcinoma (UCEC) cohort, where fewer than half of tumors profiled by RNA sequencing ($n = 223$) also underwent proteomic analysis ($n = 97$). CH and CH-Tum status were not associated with previously reported proteogenomic-defined tumor immunological subtypes (Figure S3H).[39]

To evaluate differences in immune cell composition in the TME, we compared CIBERSORTx-quantified cell fractions in

CellPress
OPEN ACCESS

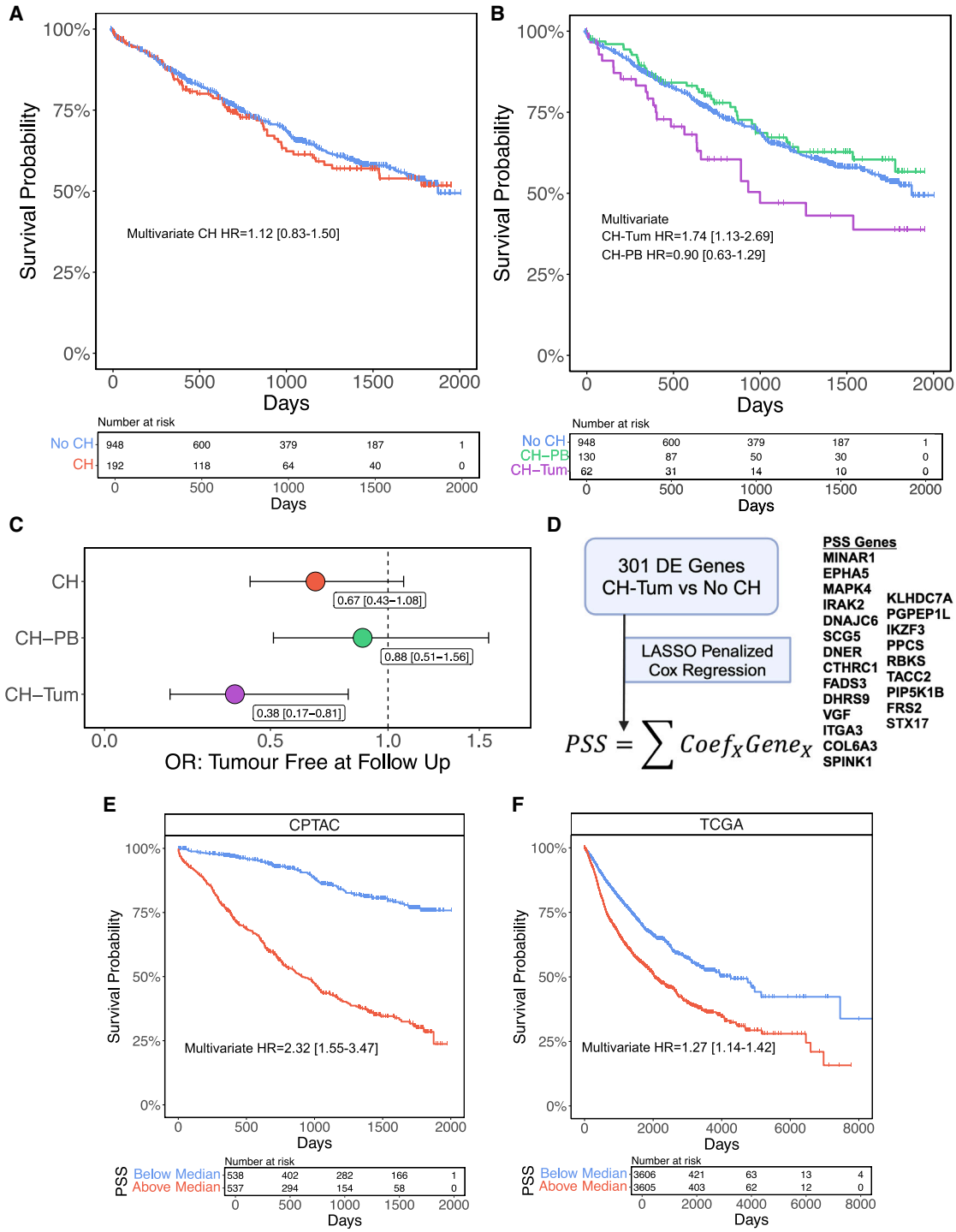


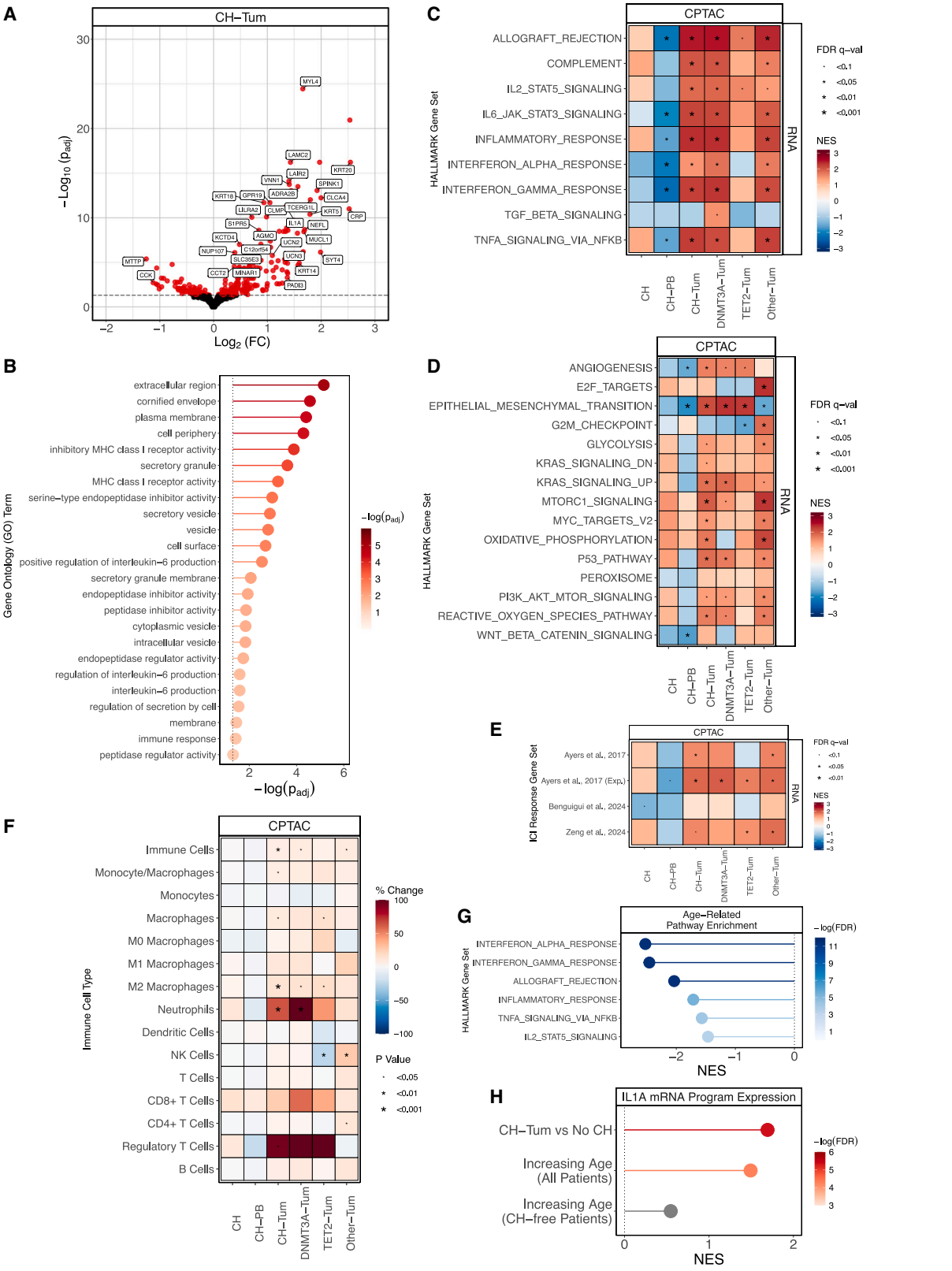
Figure 2. Clinical associations of CH-Tum in patients with solid cancer

(A and B) Kaplan-Meier plots comparing overall survival based on CH status (A) and CH-Tum status (B), with hazard ratios and 95% confidence intervals from multivariate analysis.

(C) Forest plot of odds ratios with 95% confidence intervals from multivariate logistic regression of likelihood of a patient being classified as tumor-free at last known follow-up.

(D) Schematic breakdown for the derivation of the CH-Tum-associated prognostic signature score (PSS) from differential expression data. Expression of PSS genes listed in the left and right columns is associated with poor and favorable prognosis, respectively. Created with BioRender.

(E and F) Kaplan-Meier plots comparing overall survival based on PSS in CPTAC (E) and TCGA (F), with hazard ratios and 95% confidence intervals from multivariate analysis based on a median split.



(legend on next page)

those with and without CH, again adjusted for cancer type, gender, and other cancer-specific factors using linear regression (Figure 3F). There was a 7.67% increase in overall immune infiltrate in CH-Tum across all cancers ($p = 0.008$). This enrichment was largely due to increases in the abundance of myeloid cells including macrophages (+10.3%, $p = 0.01$) and neutrophils (+65.1%, $p = 9 \times 10^{-4}$; Figure 3F). Although regulatory T cells make up a small fraction of the total immune cell population, they were also found to be enriched across cancers with CH-Tum (+95.4%, $p = 0.04$; Figure 3F).

The biological effects of CH and aging are connected by the concept of inflammaging, whereby cumulative systemic inflammation and an age-associated HSC myeloid bias are linked to various pathologic disease processes.[40] To separate the effects of age and CH on gene expression in the TME, we conducted differential gene expression and GSEA using age as the variable of interest, controlled for gender and cancer type, including and excluding patients with CH. Across all patients, aging was associated with reduced inflammatory pathway expression in the TME (Figure 3G). This age-related effect was also observed when restricting the analysis to CH-free patients and is consistent with the well-described concept of immune senescence, which is thought to be related to increased cancer incidence in the aged.[41] Like individuals without CH, tumors from patients with CH-PB showed an age-related reduction in inflammatory gene sets. In contrast, tumors with CH-Tum displayed an enrichment of inflammatory gene expression programs (Figure 3C), again supporting the notion that infiltration of CH-mutant cells is an important driver of CH-associated changes in the TME.

Recently, an age-related increase in monocyte/macrophage-derived IL-1α production during emergency myelopoiesis was associated with accelerated lung tumorigenesis and poor survival.[31] While this increase was attributed to lower *DNMT3A* expression in aged leukocytes, how the effect interacts with or is influenced by CH was not explored. Since tumor expression of *IL1A* and a 7-gene *IL1A* gene expression program was significantly upregulated with CH-Tum in the CPTAC dataset ($\log_2$FC = 1.34, $p_{adj} = 6 \times 10^{-12}$; Figures 3H and S3I), we reasoned that CH could be contributing to the age-related inflammatory myelopoiesis phenotype. We analyzed age-associated differential gene expression changes in all tumors and confirmed that *IL1A* and the *IL1A* gene expression program are proportionally upregulated in tumors with increasing age ($\log_2$FC per SD of age = 0.20, $p_{adj} = 3 \times 10^{-5}$; Figures 3H and S3J). Strikingly, when limiting our analysis to only individuals *without* CH, there was no relationship between tumor *IL1A* expression or the *IL1A* gene expression program and age ($\log_2$FC per SD of age = 0.008, $p_{adj} = 0.34$; Figures 3H and S3J). This suggests that CH-Tum represents a distinct phenomenon from general aging and immune senescence and that CH-mutant monocytes/macrophages could be the predominant drivers of age-related emergency myelopoiesis programs in cancer.

CH shapes tumor-specific phenotypes

After delineating the effects of CH-Tum across cancers, we explored the implications of CH-Tum in different types of cancer, building on previous descriptions of cancer-specific effects.[13] In univariate analysis, CH-Tum was associated with poor survival in glioblastoma (GBM; univariate HR = 3.60 [1.39–9.29]) and clear cell renal cell carcinoma (ccRCC; univariate HR = 3.15 [1.09–9.07]) and trended toward worse survival in pancreatic ductal adenocarcinoma (PDAC; univariate HR = 1.94 [0.93–4.05]; Figures S4A and S4B). However, after applying an adjusted model accounting for age, gender, disease stage, and immune infiltration, there was no significant association between CH-Tum and survival in GBM (adjusted HR = 2.23 [0.63–7.96]) or ccRCC (adjusted HR = 2.14 [0.59–7.83]), while CH-Tum was associated with significantly worse survival in PDAC (adjusted HR = 2.41 [1.13–5.18]; Figure S4A), though the power of these exploratory analyses was limited by sample size.

There was notable heterogeneity between CH-associated immune-related signatures across cancer types (Figure 4A). Cancers like COAD, GBM, and UCEC were marked by a CH-Tum-associated pro-inflammatory phenotype, while breast cancer (BRCA), OV, and PDAC exhibited reduced inflammation (Figure 4A). CH-Tum in COAD and GBM was associated with elevated immune infiltration (COAD: +22.6%, $p = 0.02$; GBM: +36.7%, $p = 0.004$; Figure S4C). Cancers span transcriptionally and biologically distinct molecular subtypes. We identified a strong trend toward CH-Tum being associated with the aggressive, mesenchymal GBM subtype (OR = 6.55 [0.70–61.1], logistic regression; Figure 4B).[42] The mesenchymal GBM subtype is known to be particularly rich with immune infiltration compared to the other subtypes, which we observed in this cohort (mesenchymal vs. classical $p_{adj} = 1 \times 10^{-9}$; vs. IDH-mutant $p_{adj} = 6 \times 10^{-4}$; vs. proneural $p_{adj} = 9 \times 10^{-7}$, Dunn's test; Figure S4D). In addition to the trend toward more mesenchymal tumors, we also noted an enrichment of a glioma stem cell expression program[43] via GSEA in CH-Tum in GBM (Figure S4E).

Given the pronounced immune dysregulation observed with CH-Tum in GBM, we used mesenchymal GBM as a case study to further explore how CH-mutant leukocytes may alter the TME and shape tumorigenesis within a single cancer molecular

Figure 3. Molecular associations of CH-Tum in the TME

(A) Volcano plot depicting differential tumor gene expression in CH-Tum across the CPTAC pan-cancer cohort, adjusted for cancer type and gender. Labeled genes were differentially expressed in both CH and CH-Tum.

(B) Significantly enriched Gene Ontology (GO) terms among the differentially expressed genes in CH-Tum common in CPTAC and TCGA.

(C–E) GSEA of select immune-related Hallmark gene sets (C), select tumor-related Hallmark gene sets (D), and anti-PD-1 ICI predicted response gene sets (E) in RNA by CH status.

(F) Enrichment of select LM22 immune cell subpopulations in the tumor by CH status. Cells with percent change $\geq$+100% are colored as +100%.

(G) Lollipop plot of significantly downregulated immune-related Hallmark gene sets in tumor RNA with increasing age.

(H) Lollipop plot for GSEA enrichment of Park et al.'s IL-1A mRNA expression program[31] in tumors with CH-Tum, increasing age across all patients, and increasing age in patients without CH.

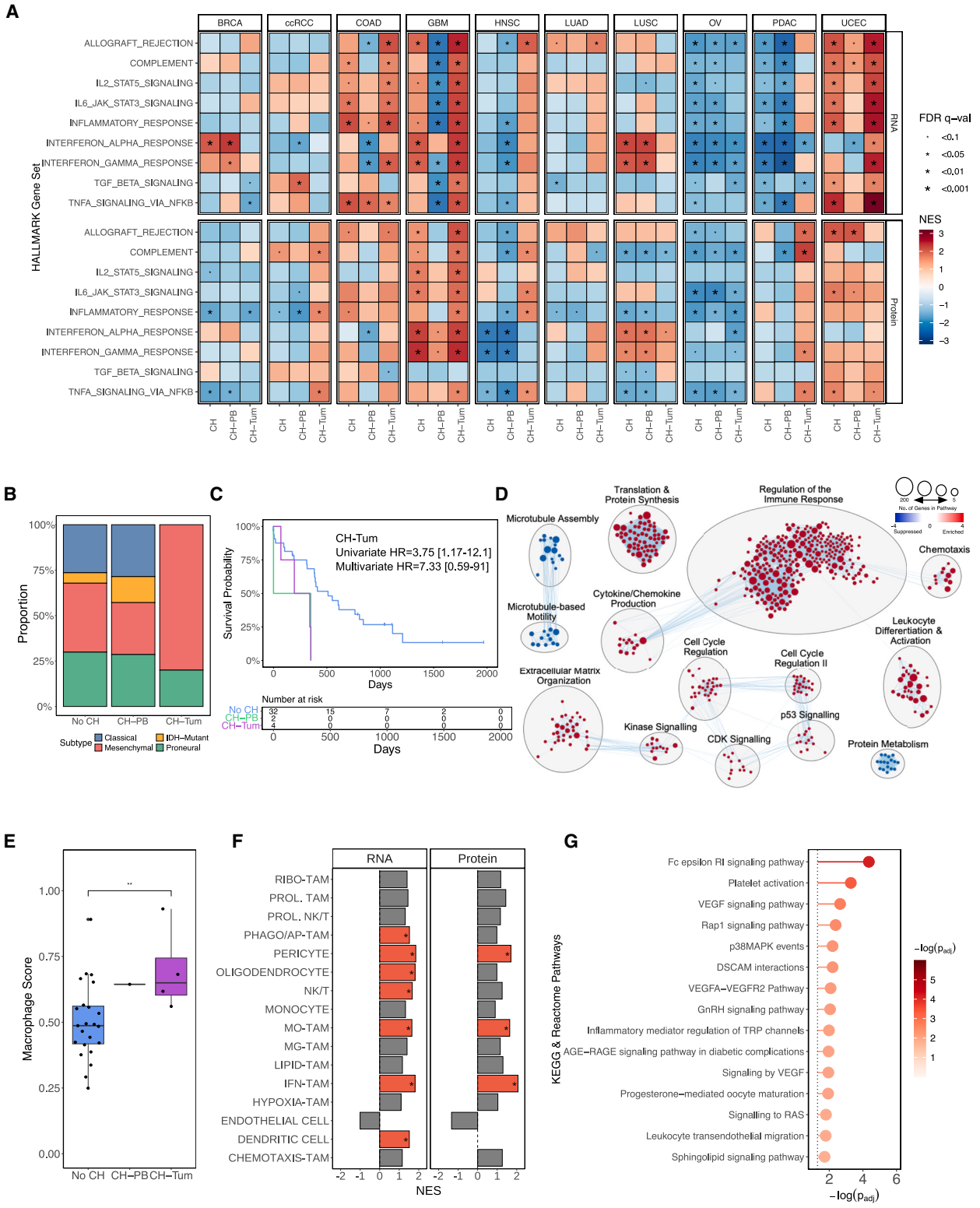


Figure 4. Cancer-specific clinical and molecular associations of CH-Tum in solid tumors

(A) Cancer-specific GSEA of select immune-related Hallmark gene sets in RNA and protein by CH status.

(B) Proportion of GBM molecular subtypes among CH-Tum, CH-PB, and no CH.

(legend continued on next page)

subtype. As such, we re-evaluated survival and inflammatory associations of CH in the context of mesenchymal GBM. Cox proportional hazard regressions of CH status demonstrated that CH-Tum had a significant effect on survival in mesenchymal GBM using a univariate model and trended toward poor outcomes in adjusted analysis despite underpowering in this small sample subset (univariate HR = 3.75 [1.17–12.1], multivariate HR = 7.33 [0.59–91]; Figure 4C).

Transcriptomic profiling of mesenchymal GBM revealed that CH-Tum associates with a pro-inflammatory phenotype with enrichment of epithelial-mesenchymal transition, angiogenesis, and other tumorigenic signaling pathways (Figure 4D; Table S4). Differential expression analyses were controlled for *NF1* mutation status, which is frequently mutated in mesenchymal GBM and linked with inflammation and immune cell recruitment.[42] Similarly, CIBERSORTx showed that CH was associated with prominent increases in macrophage abundance in the TME (+47.4%, $p = 0.009$; Figure 4E). Using previously derived GBM immune cell phenotypes,[44] CH-Tum was found to enrich monocyte-derived tumor-associated macrophages (TAMs), IFN-stimulated TAMs, and pericytes among other cell types (Figure 4F; Table S4). In all GBM and mesenchymal subtype GBM, samples with CH-Tum showed enrichment of gene signatures predictive of ICI response (Figure S4F; Table S4). The exacerbated mesenchymal phenotype of GBM with CH-Tum was also seen at the phosphoproteomic level, with overactivity of tumor-promoting kinases mitogen-activated protein kinase (MAPK)11/13, PRKCA/I, and others contributing to enrichment of immune migration, angiogenesis, and activation of tumorigenic signaling pathways including VEGF and RAS (Figures 4G and S4G). These data suggest that within a seemingly uniform GBM molecular subtype, CH-mutant leukocytes in the TME influence immune infiltration and inflammatory programs to shape an aggressive tumor phenotype.

DISCUSSION

Here, we present a pan-cancer multi-omics analysis of CH in solid tumors as a resource to accelerate the development of CH as a biomarker in precision oncology. We extend previous studies of CH in solid tumors by defining a clinically relevant CH classification—CH-Tum—representing the infiltration of CH-mutant leukocytes into the tumor and their detection via bulk tumor DNA sequencing. With CH-Tum, we demonstrate that CH can influence hallmark cancer properties of immune infiltration, inflammation, angiogenesis, and mitogenic signaling. By leveraging our validated CH identification pipelines in the context of the rich clinical and proteogenomic resources provided by CPTAC, we show that CH has potential as a biomarker that captures multiple molecular characteristics of a tumor, with predictive and prognostic implications.

CH-Tum has been previously identified in tumor sequencing reads as a potential contaminant confounding molecular tumor diagnostics.[4,5,27] Rather than simply acting as diagnostic interference, the presence of CH-Tum has clinical and biological significance in solid cancers, associating with adverse clinical outcomes. Our finding that these effects are exclusive to CH-Tum, and not CH as a whole, provides evidence for a CH-driven effect on cancer outcomes by demonstrating dose dependency, while our treatment-naive study cohort also circumvents treatment-related confounding effects present in prior studies of CH in solid tumors.[4,25] Detection of CH-Tum is feasible for clinical implementation with minor modification of existing paired tumor-blood sequencing practices. Although presented as a binary variable in this study due to technical limitations of WES, CH-Tum (versus CH-PB) likely exists on a spectrum, and future work will be required to delineate appropriate VAF thresholds that define the most clinically relevant levels of CH-mutant leukocyte infiltration in tumors. The relevance of tissue invasion by CH-mutant leukocytes may also extend beyond solid cancers; CH mutations were recently identified in atherosclerotic lesions,[45,46] and future studies are required to determine whether the extent of mutant cell infiltration correlates with disease severity, providing a novel framework for understanding CH-related pathophysiology.

Through the characterization of differential gene expression and protein abundance in tumors with and without CH, we find CH and especially CH-Tum to be associated with immune infiltration, dysregulated inflammation, and activation of tumorigenic pathways in the TME. Myeloid-biased differentiation of mutant HSCs and the production of functionally altered myeloid cells occur with many CH driver mutations[47] and may underlie the elevated presence of macrophages and neutrophils in the TME with CH-Tum. The presence of CH-mutant leukocytes in the TME may also facilitate further recruitment and infiltration of immune cells. CH-mutant leukocytes have potential to alter tumor biology and the tumor-immune interface through excessive release of pro-inflammatory cytokines that interact with wild-type immune cells, tumor cells, and stroma.[6,7,48,49] Cancer-specific variation in response to these inflammatory interactions may underlie the heterogeneous effects of CH observed in different cancer types.

Our results demonstrate the potential for CH, or CH-Tum, as a biomarker that may identify patients with distinct inflammatory tumor biology and worse prognosis. Furthermore, CH or a variant thereof may have predictive potential if it identifies individuals whose tumors are primed for an immunotherapy response. Recently, we have shown that CH drives anti-tumor immune cell phenotypes with ICI to augment immunotherapy response in a colorectal cancer model.[50] Alternatively, given the association between CH and incident inflammation-related cancers,[14–20] CH may identify patients for anti-inflammatory

(C) Kaplan-Meier plot comparing overall survival in patients with mesenchymal GBM with CH-Tum, CH-PB, and no CH.
(D) Cytoscape network diagram representing GSEA results of significantly enriched (false discovery rate [FDR]< 0.1) EnrichmentMap gene sets in RNA for CH-Tum in mesenchymal GBM.
(E) LM22 macrophage scoring in CH-Tum, CH-PB, and no CH for mesenchymal GBM; $^{**}p < 0.01$.
(F) RNA and protein GSEA enrichment for GBM immune cell phenotypes for CH-Tum in mesenchymal GBM; highlighted bars indicate significance at FDR < 0.1.
(G) Top 15 significantly enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome pathways based on functional profiling of phosphoproteomic-enriched kinases with CH-Tum in mesenchymal GBM.

prophylaxis. The impact of CH and its potential for biomarker development is highlighted in GBM, where CH-Tum is linked with inflamed, mesenchymal tumors. Mesenchymal glioma cell differentiation is reportedly induced by inflammatory macrophages in the TME,[42,51,52] implicating CH-associated macrophage dysfunction as a potential driver of this process. A CH-related shift in the tumor-immune microenvironment has been recently demonstrated in a preclinical model of anaplastic thyroid cancer, which is similarly characterized by pronounced myeloid infiltration and inflammation.[53]

CH has considerable potential as a biomarker in precision oncology beyond its known connection to therapy-related myeloid cancers.[13] Viewing tumor biology in the context of CH also offers an alternative perspective for biological discovery in cancer research. Still, the impact of CH is nuanced, and altered molecular pathways are likely to be cancer specific and potentially driver mutation specific, warranting further investigation. In our survey of CH and its consequences in solid cancer, we focus on somatic alterations in the immune cells, rather than in the tumor, to advance our understanding of cancer. We envision a future where CH and immunogenomics can be integrated with molecular tumor diagnostics to unlock the full potential of personalized cancer care.

Limitations of the study

Limitations of our study include tumor sample purity and sampling bias influencing CH-Tum detection in tumor sequencing and sample size limitations for cancer- and gene-specific analyses. ICI response signatures used in our analyses have been validated across numerous tumor types; however, in several cancers, including GBM, immunotherapy is yet to demonstrate significant clinical benefits. The lack of single-cell analysis precludes the identification of the source of dysregulated gene expression pathways in patients with CH. As large-scale cancer datasets are established in the future, the availability of these data will allow for more thorough examinations of CH and its role in the TME.

RESOURCE AVAILABILITY

Lead contact

Inquiries regarding data and code access should be directed to the lead contact, Michael J. Rauh (rauhm@queensu.ca).

Materials availability

No new materials or reagents were generated for this study.

Data and code availability

- Clinical, genomic, and transcriptomic data from the CPTAC and TCGA cohorts were used to conduct this study and can be accessed via the NIH NCI genomic data commons portal (dbGaP: phs000892.v.6.p1, phs001287.v.18.p6, phs000178.v.11.p8). Processed proteomics data were acquired via the CPTAC software package described by Li et al.[54] Additional patient and sample metadata were collected from several CPTAC and TCGA marker papers.[39,54–67] Additional supplemental data produced by this study are available online. (Zenodo: https://doi.org/10.5281/zenodo.13737566).
- This study does not report any original code. Scripts used to conduct study analyses are available online (Zenodo: https://doi.org/10.5281/zenodo.13737566). For CH variant filtering scripts, see the study by Vlasschaert et al.[26]
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We would like to acknowledge the CPTAC and TCGA programs for providing the data used in this study and the Center for Advanced Computing at Queen's University for providing computing resources. We also thank Mark Minden, Stanley Ng, and Andy Zeng for their feedback and Amy McNaughton and Kathrin Tyryshkin for their technical support.

This work was supported by funding from the CIHR (202010PJT-451137), OICR (CPTRG-056), NFRF-E (NFRFE-2019-01575), MOHCCN (Clinician Scientist Award 3256-04), LLSC, and BTFC (Elevation Grant). M.M.B. was supported by a CIHR CGS-M award and an Ontario Graduate Scholarship.

AUTHOR CONTRIBUTIONS

Conceptualization, all authors; formal analysis, M.M.B., R.J.V., and M.J.R.; funding acquisition, R.J.V. and M.J.R.; methodology, all authors; supervision, R.J.V. and M.J.R.; visualization, M.M.B.; writing – original draft, M.M.B.; writing – review and editing, all authors.

DECLARATION OF INTERESTS

A.G.B. is on the scientific advisory board of TenSixteen Bio.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL & STUDY PARTICIPANT DETAILS
  - Patients
- METHOD DETAILS
  - Genomic data processing & clonal hematopoiesis variant calling
  - Survival analysis
  - RNA-seq data processing
  - Differential gene expression analysis
  - Penalized Cox regression and development of CH-Tum prognostic signature score
  - Deconvolution of immune cell subpopulations
  - Differential protein & phosphoprotein abundance analysis
  - Gene set enrichment analysis
  - Kinase-substrate enrichment analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2025.101989.

Received: September 16, 2024
Revised: December 4, 2024
Accepted: February 4, 2025
Published: March 3, 2025

REFERENCES

1. Jaiswal, S., and Ebert, B.L. (2019). Clonal hematopoiesis in human aging and disease. Science *366*, eaan4673. https://doi.org/10.1126/science.aan4673.
2. Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burtt, N., Chavez, A., et al. (2014). Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. N. Engl. J. Med. *371*, 2488–2498. https://doi.org/10.1056/NEJMoa1408617.

3. Genovese, G., Kähler, A.K., Handsaker, R.E., Lindberg, J., Rose, S.A., Bakhoum, S.F., Chambert, K., Mick, E., Neale, B.M., Fromer, M., et al. (2014). Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. N. Engl. J. Med. Overseas. Ed. *371*, 2477–2487. https://doi.org/10.1056/NEJMoa1409405.

4. Coombs, C.C., Zehir, A., Devlin, S.M., Kishtagari, A., Syed, A., Jonsson, P., Hyman, D.M., Solit, D.B., Robson, M.E., Baselga, J., et al. (2017). Therapy-Related Clonal Hematopoiesis in Patients with Non-hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes. Cell Stem Cell *21*, 374–382.e4. https://doi.org/10.1016/j.stem.2017.07.010.

5. Bolton, K.L., Ptashkin, R.N., Gao, T., Braunstein, L., Devlin, S.M., Kelly, D., Patel, M., Berthon, A., Syed, A., Yabe, M., et al. (2020). Cancer therapy shapes the fitness landscape of clonal hematopoiesis. Nat. Genet. *52*, 1219–1226. https://doi.org/10.1038/s41588-020-00710-0.

6. Cook, E.K., Izukawa, T., Young, S., Rosen, G., Jamali, M., Zhang, L., Johnson, D., Bain, E., Hilland, J., Ferrone, C.K., et al. (2019). Comorbid and inflammatory characteristics of genetic subtypes of clonal hematopoiesis. Blood Adv. *3*, 2482–2486. https://doi.org/10.1182/bloodadvances.2018024729.

7. Cook, E.K., Luo, M., and Rauh, M.J. (2020). Clonal hematopoiesis and inflammation: Partners in leukemogenesis and comorbidity. Exp. Hematol. *83*, 85–94. https://doi.org/10.1016/j.exphem.2020.01.011.

8. Jaiswal, S. (2020). Clonal hematopoiesis and nonhematologic disorders. Blood *136*, 1606–1614. https://doi.org/10.1182/blood.2019000989.

9. Jaiswal, S., Natarajan, P., Silver, A.J., Gibson, C.J., Bick, A.G., Shvartz, E., McConkey, M., Gupta, N., Gabriel, S., Ardissino, D., et al. (2017). Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. N. Engl. J. Med. *377*, 111–121. https://doi.org/10.1056/NEJMoa1701719.

10. Yu, Z., Filder, T.P., Ruan, Y., Vlasschaert, C., Nakao, T., Uddin, M.M., Mack, T., Niroula, A., Heimlich, J.B., Zekavat, S.M., et al. (2023). Genetic modification of inflammation and clonal hematopoiesis-associated cardiovascular risk. J. Clin. Investig. *133*, e168597. https://doi.org/10.1172/JCI168597.

11. Vlasschaert, C., McNaughton, A.J.M., Chong, M., Cook, E.K., Hopman, W., Kestenbaum, B., Robinson-Cohen, C., Garland, J., Moran, S.M., Paré, G., et al. (2022). Association of Clonal Hematopoiesis of Indeterminate Potential with Worse Kidney Function and Anemia in Two Cohorts of Patients with Advanced Chronic Kidney Disease. JASN (J. Am. Soc. Nephrol.) *33*, 985–995. https://doi.org/10.1681/ASN.2021060774.

12. Vlasschaert, C., Robinson-Cohen, C., Chen, J., Akwo, E., Parker, A.C., Silver, S.A., Bhatraju, P.K., Poisner, H., Cao, S., Jiang, M., et al. (2024). Clonal hematopoiesis of indeterminate potential is associated with acute kidney injury. Nat. Med. *30*, 810–817. https://doi.org/10.1038/s41591-024-02854-6.

13. Buttigieg, M.M., and Rauh, M.J. (2023). Clonal Hematopoiesis: Updates and Implications at the Solid Tumor-Immune Interface. JCO Precis. Oncol. *7*, e2300132. https://doi.org/10.1200/PO.23.00132.

14. Kar, S.P., Quiros, P.M., Gu, M., Jiang, T., Mitchell, J., Langdon, R., Iyer, V., Barcena, C., Vijayabaskar, M.S., Fabre, M.A., et al. (2022). Genome-wide analyses of 200,453 individuals yield new insights into the causes and consequences of clonal hematopoiesis. Nat. Genet. *54*, 1155–1166. https://doi.org/10.1038/s41588-022-01121-z.

15. Kessler, M.D., Damask, A., O'Keeffe, S., Banerjee, N., Li, D., Watanabe, K., Marketta, A., Van Meter, M., Semrau, S., Horowitz, J., et al. (2022). Common and rare variant associations with clonal haematopoiesis phenotypes. Nature *612*, 301–309. https://doi.org/10.1038/s41586-022-05448-9.

16. Tian, R., Wiley, B., Liu, J., Zong, X., Truong, B., Zhao, S., Uddin, M.M., Niroula, A., Miller, C.A., Mukherjee, S., et al. (2023). Clonal Hematopoiesis and Risk of Incident Lung Cancer. JCO, JCO *41*, 1423–1433. https://doi.org/10.1200/JCO.22.00857.

17. Cheng, C., Hong, W., Li, Y., Xiao, X., McKay, J., Han, Y., Byun, J., Peng, B., Albanes, D., Lam, S., et al. (2023). Mosaic Chromosomal Alterations Are Associated With Increased Lung Cancer Risk: Insight From the INTEGRAL-ILCCO Cohort Analysis. J. Thorac. Oncol. *18*, 1003–1016. https://doi.org/10.1016/j.jtho.2023.05.001.

18. Singh, J., Li, N., Ashrafi, E., Thao, L.T.P., Curtis, D.J., Wood, E.M., and McQuilten, Z.K. (2024). Clonal hematopoiesis of indeterminate potential as a prognostic factor: A systematic review and meta-analysis. Blood Adv. *8*, 3771–3784. https://doi.org/10.1182/bloodadvances.2024013228.

19. Vlasschaert, C., Buttigieg, M., Pershad, Y., Lanktree, M., Aldrich, M.C., Rauh, M.J., and Bick, A.G. (2024). Clonal hematopoiesis of indeterminate potential-associated non-small cell lung cancer risk is potentiated by small particulate matter air pollution among non-smokers: a novel somatic variant–environment interaction. Preprint at: medRxiv, 2024.01.17.24301439. https://doi.org/10.1101/2024.01.17.24301439.

20. Marchetti, A., Pelusi, S., Marella, A., Malvestiti, F., Ricchiuti, A., Ronzoni, L., Lionetti, M., Moretti, V., Bugianesi, E., Miele, L., et al. (2024). Impact of clonal hematopoiesis of indeterminate potential on hepatocellular carcinoma in individuals with steatotic liver disease. Hepatology *80*, 816–827. https://doi.org/10.1097/HEP.0000000000000839.

21. Arends, C.M., Dimitriou, S., Stahler, A., Hablesreiter, R., Strzelecka, P.M., Stein, C.M., Tilgner, M., Saiki, R., Ogawa, S., Bullinger, L., et al. (2022). Clonal hematopoiesis is associated with improved survival in patients with metastatic colorectal cancer from the FIRE-3 trial. Blood *139*, 1593–1597. https://doi.org/10.1182/blood.2021014108.

22. Hsiehchen, D., Sfreddo, H.J., Zhao, K., Han, C.Y., and Morris, L.G.T. (2022). Clonal hematopoiesis and differential outcomes after immune checkpoint blockade. Cancer Cell *40*, 1071–1072. https://doi.org/10.1016/j.ccell.2022.08.024.

23. Morganti, S., Gibson, C.J., Jin, Q., Santos, K., Patel, A., Wilson, A., Merrill, M., Vincuilla, J., Stokes, S., Lipsyc-Sharf, M., et al. (2024). Prevalence, Dynamics, and Prognostic Role of Clonal Hematopoiesis of Indeterminate Potential in Patients With Breast Cancer. J. Clin. Oncol. *42*, 3666–3679, JCO2301071. https://doi.org/10.1200/JCO.23.01071.

24. Diplas, B.H., Ptashkin, R., Chou, J.F., Sabwa, S., Foote, M.B., Rousseau, B., Argilés, G., White, J.R., Stewart, C.M., Bolton, K., et al. (2023). Clinical Importance of Clonal Hematopoiesis in Metastatic Gastrointestinal Tract Cancers. JAMA Netw. Open *6*, e2254221. https://doi.org/10.1001/jamanetworkopen.2022.54221.

25. Fairchild, L., Whalen, J., D'Aco, K., Wu, J., Gustafson, C.B., Solovieff, N., Su, F., Leary, R.J., Campbell, C.D., and Balbin, O.A. (2023). Clonal hematopoiesis detection in patients with cancer using cell-free DNA sequencing. Sci. Transl. Med. *15*, eabm8729. https://doi.org/10.1126/scitranslmed.abm8729.

26. Vlasschaert, C., Mack, T., Heimlich, J.B., Niroula, A., Uddin, M.M., Weinstock, J., Sharber, B., Silver, A.J., Xu, Y., Savona, M., et al. (2023). A practical approach to curate clonal hematopoiesis of indeterminate potential in human genetic data sets. Blood *141*, 2214–2223. https://doi.org/10.1182/blood.2022018825.

27. Ptashkin, R.N., Mandelker, D.L., Coombs, C.C., Bolton, K., Yelskaya, Z., Hyman, D.M., Solit, D.B., Baselga, J., Arcila, M.E., Ladanyi, M., et al. (2018). Prevalence of Clonal Hematopoiesis Mutations in Tumor-Only Clinical Genomic Profiling of Solid Tumors. JAMA Oncol. *4*, 1589–1593. https://doi.org/10.1001/jamaoncol.2018.2297.

28. Wang, V.G., Kim, H., and Chuang, J.H. (2018). Whole-exome sequencing capture kit biases yield false negative mutation calls in TCGA cohorts. PLoS One *13*, e0204912. https://doi.org/10.1371/journal.pone.0204912.

29. Rasnic, R., Brandes, N., Zuk, O., and Linial, M. (2019). Substantial batch effects in TCGA exome sequences undermine pan-cancer analysis of germline variants. BMC Cancer *19*, 783. https://doi.org/10.1186/s12885-019-5994-5.

30. Choi, J.-H., Hong, S.-E., and Woo, H.G. (2017). Pan-cancer analysis of systematic batch effects on somatic sequence variations. BMC Bioinf. *18*, 211. https://doi.org/10.1186/s12859-017-1627-7.

31. Park, M.D., Le Berichel, J., Hamon, P., Wilk, C.M., Belabed, M., Yatim, N., Saffon, A., Boumelha, J., Falcomatà, C., Tepper, A., et al. (2024). Hematopoietic aging promotes cancer by fueling IL-1α–driven emergency

myelopoiesis. Science *386*, eadn0327. https://doi.org/10.1126/science.adn0327.

32. Lin, T.-C. (2021). Functional Roles of SPINK1 in Cancers. Int. J. Mol. Sci. *22*, 3814. https://doi.org/10.3390/ijms22083814.
33. Mehner, C., Miller, E., Hockla, A., Coban, M., Weroha, S.J., Radisky, D.C., and Radisky, E.S. (2020). Targeting an autocrine IL-6–SPINK1 signaling axis to suppress metastatic spread in ovarian clear cell carcinoma. Oncogene *39*, 6606–6618. https://doi.org/10.1038/s41388-020-01451-4.
34. Räsänen, K., Lehtinen, E., Nokelainen, K., Kuopio, T., Hautala, L., Itkonen, O., Stenman, U.-H., and Koistinen, H. (2016). Interleukin-6 increases expression of serine protease inhibitor Kazal type 1 through STAT3 in colorectal adenocarcinoma. Mol. Carcinog. *55*, 2010–2023. https://doi.org/10.1002/mc.22447.
35. Busque, L., Sun, M., Buscarlet, M., Ayachi, S., Feroz Zada, Y., Provost, S., Bourgoin, V., Mollica, L., Meisel, M., Hinterleitner, R., et al. (2020). High-sensitivity C-reactive protein is associated with clonal hematopoiesis of indeterminate potential. Blood Adv. *4*, 2430–2438. https://doi.org/10.1182/bloodadvances.2019000770.
36. Ayers, M., Lunceford, J., Nebozhyn, M., Murphy, E., Loboda, A., Kaufman, D.R., Albright, A., Cheng, J.D., Kang, S.P., Shankaran, V., et al. (2017). IFN-γ–related mRNA profile predicts clinical response to PD-1 blockade. J. Clin. Investig. *127*, 2930–2940. https://doi.org/10.1172/JCI91190.
37. Benguigui, M., Cooper, T.J., Kalkar, P., Schif-Zuck, S., Halaban, R., Bacchiocchi, A., Kamer, I., Deo, A., Manobla, B., Menachem, R., et al. (2024). Interferon-stimulated neutrophils as a predictor of immunotherapy response. Cancer Cell *42*, 253–265.e12. https://doi.org/10.1016/j.ccell.2023.12.005.
38. Zeng, S., Chen, L., Tian, J., Liu, Z., Liu, X., Tang, H., Wu, H., and Liu, C. (2024). Integrative analysis of pan-cancer single-cell data reveals a tumor ecosystem subtype predicting immunotherapy response. npj Precis. Oncol. *8*, 205–213. https://doi.org/10.1038/s41698-024-00703-w.
39. Petralia, F., Ma, W., Yaron, T.M., Caruso, F.P., Tignor, N., Wang, J.M., Charytonowicz, D., Johnson, J.L., Huntsman, E.M., Marino, G.B., et al. (2024). Pan-cancer proteogenomics characterization of tumor immunity. Cell *187*, 1255–1277.e27. https://doi.org/10.1016/j.cell.2024.01.027.
40. Franceschi, C., Garagnani, P., Parini, P., Giuliani, C., and Santoro, A. (2018). Inflammaging: a new immune–metabolic viewpoint for age-related diseases. Nat. Rev. Endocrinol. *14*, 576–590. https://doi.org/10.1038/s41574-018-0059-4.
41. Fu, Z., Xu, H., Yue, L., Zheng, W., Pan, L., Gao, F., and Liu, X. (2023). Immunosenescence and cancer: Opportunities and challenges. Medicine *102*, e36045. https://doi.org/10.1097/MD.0000000000036045.
42. Wang, Q., Hu, B., Hu, X., Kim, H., Squatrito, M., Scarpace, L., deCarvalho, A.C., Lyu, S., Li, P., Li, Y., et al. (2017). Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. Cancer Cell *32*, 42–56.e6. https://doi.org/10.1016/j.ccell.2017.06.003.
43. Chen, B., Zhou, X., Yang, L., Zhou, H., Meng, M., Wu, H., Liu, Z., Zhang, L., and Li, C. (2022). Glioma stem cell signature predicts the prognosis and the response to tumor treating fields treatment. CNS Neurosci. Ther. *28*, 2148–2162. https://doi.org/10.1111/cns.13956.
44. Wang, W., Li, T., Cheng, Y., Li, F., Qi, S., Mao, M., Wu, J., Liu, Q., Zhang, X., Li, X., et al. (2024). Identification of hypoxic macrophages in glioblastoma with therapeutic potential for vasculature normalization. Cancer Cell *42*, 815–832.e12. https://doi.org/10.1016/j.ccell.2024.03.013.
45. Büttner, P., Böttner, J., Krohn, K., Baber, R., Platzbecker, U., Cross, M., Desch, S., Thiele, H., Steiner, S., Scheinert, D., et al. (2023). Clonal Hematopoiesis Mutations Are Present in Atherosclerotic Lesions in Peripheral Artery Disease. Int. J. Mol. Sci. *24*, 3962. https://doi.org/10.3390/ijms24043962.
46. Heimlich, J.B., Raddatz, M.A., Wells, J., Vlasschaert, C., Olson, S., Threadcraft, M., Foster, K., Boateng, E., Umbarger, K., Su, Y.R., et al. (2024). Invasive Assessment of Coronary Artery Disease in Clonal Hematopoiesis of Indeterminate Potential. Circ. Genom. Precis. Med. *17*, e004415. https://doi.org/10.1161/CIRCGEN.123.004415.
47. Ostrander, E.L., Kramer, A.C., Mallaney, C., Celik, H., Koh, W.K., Fairchild, J., Haussler, E., Zhang, C.R.C., and Challen, G.A. (2020). Divergent Effects of Dnmt3a and Tet2 Mutations on Hematopoietic Progenitor Cell Fitness. Stem Cell Rep. *14*, 551–560. https://doi.org/10.1016/j.stemcr.2020.02.011.
48. Trowbridge, J.J., and Starczynowski, D.T. (2021). Innate immune pathways and inflammation in hematopoietic aging, clonal hematopoiesis, and MDS. J. Exp. Med. *218*, e20201544. https://doi.org/10.1084/jem.20201544.
49. Cobo, I., Tanaka, T., Glass, C.K., and Yeang, C. (2022). Clonal hematopoiesis driven by DNMT3A and TET2 mutations: role in monocyte and macrophage biology and atherosclerotic cardiovascular disease. Curr. Opin. Hematol. *29*, 1–7. https://doi.org/10.1097/MOH.0000000000000688.
50. Vanner, R.J., Bansal, S., Buttigieg, M.M., Zeng, A.G.X., Rondeau, V., Chan, D.Y., Chan-Seng-Yue, M., Jin, L., McLeod, J., Donato, E., et al. (2024). Hematopoietic Tet2 inactivation enhances the response to checkpoint blockade immunotherapy. Preprint at: bioRxiv, 2024.09.09.612140. https://doi.org/10.1101/2024.09.09.612140.
51. Bhat, K.P.L., Balasubramaniyan, V., Vaillant, B., Ezhilarasan, R., Hummelink, K., Hollingsworth, F., Wani, K., Heathcock, L., James, J.D., Goodman, L.D., et al. (2013). Mesenchymal Differentiation Mediated by NF-κB Promotes Radiation Resistance in Glioblastoma. Cancer Cell *24*, 331–346. https://doi.org/10.1016/j.ccr.2013.08.001.
52. Hara, T., Chanoch-Myers, R., Mathewson, N.D., Myskiw, C., Atta, L., Bussema, L., Eichhorn, S.W., Greenwald, A.C., Kinker, G.S., Rodman, C., et al. (2021). Interactions between cancer cells and immune cells drive transitions to mesenchymal-like states in glioblastoma. Cancer Cell *39*, 779–792.e11. https://doi.org/10.1016/j.ccell.2021.05.002.
53. Tiedje, V., Vela, P.S., Yang, J.L., Untch, B.R., Boucai, L., Stonestrom, A.J., Costa, A.B., Expósito, S.F., Srivastava, A., Kerpelev, M., et al. (2024). Targetable treatment resistance in thyroid cancer with clonal hematopoiesis. Preprint at: bioRxiv, 2024.10.10.617685. https://doi.org/10.1101/2024.10.10.617685.
54. Li, Y., Dou, Y., Da Veiga Leprevost, F., Geffen, Y., Calinawan, A.P., Aguet, F., Akiyama, Y., Anand, S., Birger, C., Cao, S., et al. (2023). Proteogenomic data and resources for pan-cancer analysis. Cancer Cell *41*, 1397–1406. https://doi.org/10.1016/j.ccell.2023.06.009.
55. Wang, L.-B., Karpova, A., Gritsenko, M.A., Kyle, J.E., Cao, S., Li, Y., Rykunov, D., Colaprico, A., Rothstein, J.H., Hong, R., et al. (2021). Proteogenomic and metabolomic characterization of human glioblastoma. Cancer Cell *39*, 509–528.e20. https://doi.org/10.1016/j.ccell.2021.01.006.
56. Vasaikar, S., Huang, C., Wang, X., Petyuk, V.A., Savage, S.R., Wen, B., Dou, Y., Zhang, Y., Shi, Z., Arshad, O.A., et al. (2019). Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. Cell *177*, 1035–1049.e19. https://doi.org/10.1016/j.cell.2019.03.030.
57. Clark, D.J., Dhanasekaran, S.M., Petralia, F., Pan, J., Song, X., Hu, Y., da Veiga Leprevost, F., Reva, B., Lih, T.-S.M., Chang, H.-Y., et al. (2019). Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. Cell *179*, 964–983.e31. https://doi.org/10.1016/j.cell.2019.10.007.
58. Huang, C., Chen, L., Savage, S.R., Eguez, R.V., Dou, Y., Li, Y., da Veiga Leprevost, F., Jaehnig, E.J., Lei, J.T., Wen, B., et al. (2021). Proteogenomic insights into the biology and treatment of HPV-negative head and neck squamous cell carcinoma. Cancer Cell *39*, 361–379.e16. https://doi.org/10.1016/j.ccell.2020.12.007.
59. Krug, K., Jaehnig, E.J., Satpathy, S., Blumenberg, L., Karpova, A., Anurag, M., Miles, G., Mertins, P., Geffen, Y., Tang, L.C., et al. (2020). Proteogenomic Landscape of Breast Cancer Tumorigenesis and Targeted Therapy. Cell *183*, 1436–1456.e31. https://doi.org/10.1016/j.cell.2020.10.036.
60. Satpathy, S., Krug, K., Jean Beltran, P.M., Savage, S.R., Petralia, F., Kumar-Sinha, C., Dou, Y., Reva, B., Kane, M.H., Avanessian, S.C., et al. (2021). A proteogenomic portrait of lung squamous cell carcinoma. Cell *184*, 4348–4371.e40. https://doi.org/10.1016/j.cell.2021.07.016.

61. Gillette, M.A., Satpathy, S., Cao, S., Dhanasekaran, S.M., Vasaikar, S.V., Krug, K., Petralia, F., Li, Y., Liang, W.-W., Reva, B., et al. (2020). Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma. Cell *182*, 200–225.e35. https://doi.org/10.1016/j.cell.2020.06.013.

62. Cao, L., Huang, C., Cui Zhou, D., Hu, Y., Lih, T.M., Savage, S.R., Krug, K., Clark, D.J., Schnaubelt, M., Chen, L., et al. (2021). Proteogenomic characterization of pancreatic ductal adenocarcinoma. Cell *184*, 5031–5052.e26. https://doi.org/10.1016/j.cell.2021.08.023.

63. Dou, Y., Kawaler, E.A., Cui Zhou, D., Gritsenko, M.A., Huang, C., Blumenberg, L., Karpova, A., Petyuk, V.A., Savage, S.R., Satpathy, S., et al. (2020). Proteogenomic Characterization of Endometrial Carcinoma. Cell *180*, 729–748.e26. https://doi.org/10.1016/j.cell.2020.01.026.

64. Zhang, H., Liu, T., Zhang, Z., Payne, S.H., Zhang, B., McDermott, J.E., Zhou, J.-Y., Petyuk, V.A., Chen, L., Ray, D., et al. (2016). Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. Cell *166*, 755–765. https://doi.org/10.1016/j.cell.2016.05.069.

65. Cancer Genome Atlas Research Network; Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nat. Genet. *45*, 1113–1120. https://doi.org/10.1038/ng.2764.

66. Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., et al. (2018). An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell *173*, 400–416.e11. https://doi.org/10.1016/j.cell.2018.02.052.

67. Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.-H., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A., et al. (2018). The Immune Landscape of Cancer. Immunity *48*, 812–830.e14. https://doi.org/10.1016/j.immuni.2018.03.023.

68. Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods *12*, 453–457. https://doi.org/10.1038/nmeth.3337.

69. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425. https://doi.org/10.1016/j.cels.2015.12.004.

70. Hornbeck, P.V., Kornhauser, J.M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., and Sullivan, M. (2012). PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res. *40*, D261–D270. https://doi.org/10.1093/nar/gkr1122.

71. Linding, R., Jensen, L.J., Pasculescu, A., Olhovsky, M., Colwill, K., Bork, P., Yaffe, M.B., and Pawson, T. (2008). NetworKIN: a resource for exploring cellular phosphorylation networks. Nucleic Acids Res. *36*, D695–D699. https://doi.org/10.1093/nar/gkm902.

72. Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. *28*, 27–30.

73. Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., et al. (2011). Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res. *39*, D691–D697. https://doi.org/10.1093/nar/gkq1018.

74. Van der Auwera, G.A., and O'Connor, B.D. (2020). Genomics in the Cloud (O'Reilly Media Inc.).

75. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164. https://doi.org/10.1093/nar/gkq603.

76. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative Genomics Viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

77. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discov. *2*, 401–404. https://doi.org/10.1158/2159-8290.CD-12-0095.

78. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. Sci. Signal. *6*, pl1. https://doi.org/10.1126/scisignal.2004088.

79. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

80. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

81. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527. https://doi.org/10.1038/nbt.3519.

82. Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047–3048. https://doi.org/10.1093/bioinformatics/btw354.

83. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

84. Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. *4*, 1521. https://doi.org/10.12688/f1000research.7563.2.

85. Ignatiadis, N., Klaus, B., Zaugg, J.B., and Huber, W. (2016). Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. Nat. Methods *13*, 577–580. https://doi.org/10.1038/nmeth.3885.

86. Stephens, M. (2017). False discovery rates: a new deal. Biostatistics *18*, 275–294. https://doi.org/10.1093/biostatistics/kxw041.

87. Kolberg, L., Raudvere, U., Kuzmin, I., Adler, P., Vilo, J., and Peterson, H. (2023). g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). Nucleic Acids Res. *51*, W207–W212. https://doi.org/10.1093/nar/gkad347.

88. Newman, A.M., Steen, C.B., Liu, C.L., Gentles, A.J., Chaudhuri, A.A., Scherer, F., Khodadoust, M.S., Esfahani, M.S., Luca, B.A., Steiner, D., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat. Biotechnol. *37*, 773–782. https://doi.org/10.1038/s41587-019-0114-2.

89. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47. https://doi.org/10.1093/nar/gkv007.

90. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550. https://doi.org/10.1073/pnas.0506580102.

91. Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., Wadi, L., Meyer, M., Wong, J., Xu, C., et al. (2019). Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. Nat. Protoc. *14*, 482–517. https://doi.org/10.1038/s41596-018-0103-9.

92. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A Software

Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. *13*, 2498–2504. https://doi.org/10.1101/gr.1239303.

93. Wiredja, D.D., Koyutürk, M., and Chance, M.R. (2017). The KSEA App: a web-based tool for kinase activity inference from quantitative phosphoproteomics. Bioinformatics *33*, 3489–3491. https://doi.org/10.1093/bioinformatics/btx415.
94. te Grotenhuis, M., Pelzer, B., Eisinga, R., Nieuwenhuis, R., Schmidt-Catran, A., and Konig, R. (2017). When size matters: advantages of weighted effect coding in observational studies. Int. J. Public Health *62*, 163–167. https://doi.org/10.1007/s00038-016-0901-1.
95. Miller, C.A., Walker, J.R., Jensen, T.L., Hooper, W.F., Fulton, R.S., Painter, J.S., Sekeres, M.A., Ley, T.J., Spencer, D.H., Goll, J.B., and Walter, M.J. (2022). Failure to Detect Mutations in U2AF1 due to Changes in the GRCh38 Reference Sequence. J. Mol. Diagn. *24*, 219–223. https://doi.org/10.1016/j.jmoldx.2021.10.013.
96. Bowler, T.G., Pradhan, K., Kong, Y., Bartenstein, M., Morrone, K.A., Sridharan, A., Kessel, R.M., Shastri, A., Giricz, O., Bhagat, T.D., et al. (2019). Misidentification of MLL3 and other mutations in cancer due to highly homologous genomic regions. Leuk. Lymphoma *60*, 3132–3137. https://doi.org/10.1080/10428194.2019.1630620.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
CPTAC & TCGA Clonal Hematopoiesis Variant Calls	This paper	Table S2
CPTAC & TCGA Genomics & Transcriptomics Data (dbGaP: phs000892.v6.p1, phs001287.v18.p6, phs000178.v11.p8)	NIH NCI Genomic Data Commons	https://portal.gdc.cancer.gov/ https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000892.v6.p1 https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001287.v18.p6 https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000178.v11.p8
CPTAC Proteomics & Phosphoproteomics Data	Li et al.[54]	https://doi.org/10.1016/j.ccell.2023.06.009
CPTAC & TCGA Patient Clinical Information & Sample Metadata	NIH NCI Genomic Data Commons; Petralia et al.[39]; Li et al.[54]; Wang et al.[55]; Vasaikar et al.[56]; Clark et al.[57]; Huang et al.[58]; Krug et al.[59]; Satpathy et al.[60]; Gillette et al.[61]; Cao et al.[62]; Dou et al.[63]; Zhang et al.[64]; Weinstein et al.[65]; Liu et al.[66]; Thorsson et al.[67]	https://portal.gdc.cancer.gov/ https://doi.org/10.1016/j.cell.2024.01.027 https://doi.org/10.1016/j.ccell.2023.06.009 https://doi.org/10.1016/j.ccell.2021.01.006 https://doi.org/10.1016/j.cell.2019.03.030 https://doi.org/10.1016/j.cell.2019.10.007 https://doi.org/10.1016/j.cell.2020.10.036 https://doi.org/10.1016/j.cell.2021.07.016 https://doi.org/10.1016/j.cell.2020.06.013 https://doi.org/10.1016/j.cell.2021.08.023 https://doi.org/10.1016/j.cell.2020.01.026 https://doi.org/10.1016/j.cell.2016.05.069 https://doi.org/10.1038/ng.2764 https://doi.org/10.1016/j.cell.2018.02.052 https://doi.org/10.1016/j.immuni.2018.03.023
cDNA Reference Transcriptome (Release 109)	Ensembl	https://ftp.ensembl.org/pub/
LM22 Gene Signature Matrix	Newman et al.[68]	https://doi.org/10.1038/nmeth.3337
HALLMARK Gene Sets	Liberzon et al.[69]	https://doi.org/10.1016/j.cels.2015.12.004 https://www.gsea-msigdb.org/gsea/msigdb/human/collections.jsp#H
Predicted Anti-PD-1 ICI Response Gene Sets	Ayers et al.[36]; Benguigui et al.[37]; Zeng et al.[38]	https://doi.org/10.1172/jci91190 https://doi.org/10.1016/j.ccell.2023.12.005 https://doi.org/10.1038/s41698-024-00703-w
IL1A mRNA Expression Program Gene Set	Park et al.[31]	https://doi.org/10.1126/science.adn0327
Glioma Stem Cell Gene Sets	Chen et al.[43]	https://doi.org/10.1111/cns.13956
Glioblastoma TAM Phenotype Gene Sets	Wang et al.[44]	https://doi.org/10.1016/j.ccell.2024.03.013
EnrichmentMap Gene Sets	Bader Lab	https://baderlab.org/GeneSets
PhosphoSitePlus	Hornbeck et al.[70]	https://doi.org/10.1093/nar/gkr1122 https://www.phosphosite.org/homeAction.action
NetworKIN	Linding et al.[71]	https://doi.org/10.1093/nar/gkm902 https://networkin.info/
Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathways	Kanehisa & Goto[72]	https://doi.org/10.1093/nar/28.1.27 https://www.genome.jp/kegg/
Reactome Pathways	Croft et al.[73]	https://doi.org/10.1093/nar/gkq1018 https://reactome.org/
Software and algorithms		
Genome Analysis Toolkit (GATK)-Mutect2 & DepthOfCoverage (v4.2.4.0)	Van der Auwera and O'Connor[74]	https://gatk.broadinstitute.org/hc/en-us/articles/4413079262875-Mutect2 https://gatk.broadinstitute.org/hc/en-us/articles/4413072451611-DepthOfCoverage-BETA

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Pileup region custom script	Vlasschaert et al.[26]	https://doi.org/10.1182/blood.2022018825 https://github.com/weinstockj/pileup_region
ANNOVAR	Wang et al.[75]	https://doi.org/10.1093/nar/gkq603 https://annovar.openbioinformatics.org/en/latest/
Integrated Genomics Viewer (v2.16.1)	Robinson et al.[76]	https://doi.org/10.1038/nbt.1754 https://igv.org/
cBioPortal Oncoprinter (v6.0.5)	Cerami et al.[77]; Gao et al.[78]	https://doi.org/10.1158/2159-8290.cd-12-0095 https://doi.org/10.1126/scisignal.2004088 https://www.cbioportal.org/oncoprinter
survival (v3.5-7)	CRAN	https://cran.r-project.org/web/packages/survival/index.html
survminer (v0.4.9)	CRAN	https://cran.r-project.org/web/packages/survminer/index.html
samtools (v1.17)	Li et al.[79]	https://doi.org/10.1093/bioinformatics/btp352 http://www.htslib.org/
fastp (v0.23.1)	Chen et al.[80]	https://doi.org/10.1093/bioinformatics/bty560 https://github.com/OpenGene/fastp
kallisto (v0.46.1)	Bray et al.[81]	https://doi.org/10.1038/nbt.3519 https://pachterlab.github.io/kallisto/
MultiQC (v1.13)	Ewels et al.[82]	https://doi.org/10.1093/bioinformatics/btw354 https://multiqc.info/
DESeq2 (v1.26.0)	Love et al.[83]	https://doi.org/10.1186/s13059-014-0550-8 https://bioconductor.org/packages/release/bioc/html/DESeq2.html
tximport (v1.28.0)	Soneson et al.[84]	https://doi.org/10.12688/f1000research.7563.2 https://bioconductor.org/packages/release/bioc/html/tximport.html
IHW (v1.14.0)	Ignatiadis et al.[85]	https://doi.org/10.1038/nmeth.3885 https://bioconductor.org/packages/release/bioc/html/IHW.html
ashr (v2.2-63)	Stephens[86]	https://doi.org/10.1093/biostatistics/kxw041 https://cran.r-project.org/web/packages/ashr/index.html
EnhancedVolcano (v1.20.0)	Bioconductor	https://bioconductor.org/packages/release/bioc/html/EnhancedVolcano.html
ggvenn (v0.1.10)	CRAN	https://cran.r-project.org/web/packages/ggvenn/index.html
g:Profiler g:GOSt	Kolberg et al.[87]	https://doi.org/10.1093/nar/gkad347 https://biit.cs.ut.ee/gprofiler/gost
glmnet (v4.1-8)	CRAN	https://cran.r-project.org/web/packages/glmnet/index.html
CIBERSORTx	Newman et al.[68,88]	https://doi.org/10.1038/nmeth.3337 https://doi.org/10.1038/s41587-019-0114-2 https://cibersortx.stanford.edu/
limma (v3.58.1)	Ritchie et al.[89]	https://doi.org/10.1093/nar/gkv007 https://bioconductor.org/packages/release/bioc/html/limma.html
GSEA (v4.3.2)	Subramanian et al.[90]	https://doi.org/10.1073/pnas.0506580102 https://www.gsea-msigdb.org/gsea/index.jsp
Enrichment Map Pipeline Collection (EnrichmentMap v3.3.6, clusterMaker2 v2.3.4, AutoAnnotate v1.4.1)	Reimand et al.[91]	https://doi.org/10.1038/s41596-018-0103-9 https://apps.cytoscape.org/apps/enrichmentmappipelinecollection
Cytoscape (v3.10.1)	Shannon et al.[92]	https://doi.org/10.1101/gr.1239303 https://cytoscape.org/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
KSEA App (v1.0)	Wiredja et al.[93]	https://doi.org/10.1093/bioinformatics/btx415 https://casecpb.shinyapps.io/ksea/
R (v4.3.1)	The R Project	https://www.r-project.org/
ggplot2 (v3.5.0)	CRAN	https://cran.r-project.org/web/packages/ggplot2/index.html
ggridges (v0.5.6)	CRAN	https://cran.r-project.org/web/packages/ggridges/index.html
ggbeeswarm (v0.7.2)	CRAN	https://cran.r-project.org/web/packages/ggbeeswarm/index.html
wec (v0.4-1)	te Grotenhuis et al.[94]	https://doi.org/10.1007/s00038-016-0901-1 https://cran.r-project.org/web/packages/wec/index.html
MASS (v7.3–60.0.1)	CRAN	https://cran.r-project.org/web/packages/MASS/index.html
glmmTMB (v1.1.9)	CRAN	https://cran.r-project.org/web/packages/glmmTMB/index.html
Analysis Scripts	This paper	https://doi.org/10.5281/zenodo.13737566

EXPERIMENTAL MODEL & STUDY PARTICIPANT DETAILS

Patients

Human research protocol was approved by the Queen's University Health Sciences Research Ethics Board (PATH-29-22). Informed consent was obtained by the participating CPTAC and TCGA partner projects, with access to controlled raw genomic and transcriptomic data granted by the NCI data access committee (dbGaP Project #33927).

This study was conducted using cancer patient data available from the CPTAC and TCGA cohorts (Table S1). The cancers represented in CPTAC (n = 1,550) include breast cancer (BRCA; n = 130), clear cell renal cell carcinoma (ccRCC; n = 214), colorectal adenocarcinoma (COAD; n = 106), glioblastoma (GBM; n = 104), head and neck squamous cell carcinoma (HNSC; n = 110), lung adenocarcinoma (LUAD; n = 229), lung squamous cell carcinoma (LUSC; n = 108), high grade serous ovarian cancer (OV; n = 90), pancreatic ductal adenocarcinoma (PDAC; n = 163), and uterine corpus endometrial carcinoma (UCEC; n = 240), as well as a small number of other cancers from various sites (Other; n = 56). TCGA (n = 8,927) included 32 different solid tumor types. All patients were treatment naive at the time of sample collection, with primary tumor samples collected from both local (stage I-III, n = 1,263) and metastatic (stage IV, n = 218) cancers. All patients with available peripheral blood whole exome sequencing (WES) were included in CH variant calling. Only primary, treatment-naïve tumor samples in patients with metadata for all relevant covariates were included in transcriptomic, proteomic, and phosphoproteomic analyses.

METHOD DETAILS

Genomic data processing & clonal hematopoiesis variant calling

Peripheral blood and tumor WES data aligned to the GRCh38.d1.vd1 reference genome was acquired from the CPTAC project and was processed via the Genome Analysis Toolkit (GATK)-Mutect2 v4.2.4.0.0 algorithm to identify potential CH driver mutations.[74] Depth of coverage data in CH driver genes was collected using GATK-DepthOfCoverage v4.2.4.0.0[74] There are known challenges in variant calling related to a false duplication of *U2AF1* that precluded the use of GATK-Mutect2 for variant calling in this gene.[95] To circumvent this, a custom pileup region script was applied as in Vlasschaert et al.[26]

Functional annotation of variants was performed with ANNOVAR using the Ensembl database.[75] The annotated variants were then subject to R-based filtering scripts and manual review as adapted from Vlasschaert et al.[26] to curate the final list of CH driver mutations in the study cohort. For sequencing-based filtering, we required depth of coverage $\geq$20x, alternative allele reads $\geq$5, and paired-end forward and reverse reads (F1R2/F2R1) $\geq$1. Insertion/deletion variants occurring at homopolymer sites $\geq$3 base pairs long were removed. Remaining variants were filtered against a curated list of CH driver mutations. Consistent with standard definitions of CH, VAF $\geq$2% was required. To eliminate germline variants, variants that had a binomial $p > 0.01$ of an equal number of normal and alternative allele reads (i.e., 50% VAF). Any remaining recurring artifacts were removed if they appeared more frequently than established CH hotspots like *DNMT3A* R882H. During CH calling, three patients exhibited JAK2 V617F mutations with VAF >50% (C3*N*-03021; TCGA-26-5135, TCGA-DJ-A2Q2). These clones may represent a CH mutation with co-occurring loss of heterozygosity at the same locus, but due to the potential of polycythemia vera in these cases, these patients were excluded from subsequent analyses.

KMT2C (formerly *MLL3*) suffers from collapsed duplications in the reference genome, creating challenges for alignment algorithms to appropriately map reads to their origin in the genome.[96] As a result of this, true variant calls in *KMT2C* are obscured by a high proportion of false positive variants that arise from elsewhere in the genome. Bam files were manually inspected at the *KMT2C* locus using Integrated Genomics Viewer v2.16.1 to identify exons impacted by poor mapping quality – indicative of the collapsed duplication.[76] Exons 7, 8, 14–20, and 23–25 were found to have evidence of non-specific genome mapping, and were therefore blacklisted due to poor confidence in variant calling. The remaining exons were retained for variant filtering as described above.

Variant filtering and curation to identify CH-Tum was conducted in all tumor samples from patients with CH mutations found in their peripheral blood sample. Somatic variants in each tumor sample were parsed for the same CH mutation(s) that presented in the peripheral blood. Mutations that presented in both the tumor and blood with VAF $\geq$40% and/or a tumor to blood VAF ratio $\geq$2, representing likely germline variants and tumor variants, respectively, were identified as false positive CH calls and removed from the final list of CH mutations.

Oncoplots representing CH and CH-Tum mutations were created for the 15 most frequently mutated genes using cBioPortal Oncoprinter v6.0.5.[77,78]

Survival analysis

Survival data was available for ccRCC, GBM, HNSC, LUAD, LUSC, PDAC, UCEC, and 'Other' cancer types. Differences in overall survival based on CH, CH-Tum, and CH-PB status were evaluated using univariate and multivariate Cox proportional hazard models provided by the survival v3.5-7 R package, with Kaplan-Meier curves illustrated with the survminer v0.4.9 R package. Follow up time was calculated based on time from pathologic diagnosis to death or last known follow up. The modeled covariates in survival analysis included patient age (years), gender, smoking (smoker versus non-smoker), and disease stage (local versus metastatic). Cancer type was controlled for in pan-cancer analyses. CIBERSORTx immune infiltrate scoring (described below) was also included as a continuous variable where stated. In the mesenchymal glioblastoma subset, *NF1* mutation status was also included as a covariate. Modeling of last known tumor status at follow up excluded patients with metastatic disease (including GBM) at the time of enrollment. Significant results were indicated by 95% confidence intervals that did not overlap with HR = 1 and $p < 0.05$.

RNA-seq data processing

Aligned bulk tumor RNA-seq bam files were de-aligned into raw, unstranded, paired-end fastq files with samtools v1.17.[79] Quality filtering and adapter trimming of fastq files was performed using fastp v0.23.1, where sequencing reads with base Phred score $\geq$15, and $\leq$40% bases unqualified were retained in the fastq files.[80] Reads in the filtered paired-end fastq files were pseudo-aligned and transcript expression was quantified using kallisto v0.46.1 based on release 109 of the Ensembl cDNA reference for GRCh38.[81] Reports were collected in MultiQC v1.13 for review of sequencing and quantification quality.[82]

Differential gene expression analysis

Differential gene expression analysis was performed using DESeq2 v1.26.0 to identify variations in the transcriptional landscape of the TME based on CH, CH-Tum, CH-PB, and *DNMT3A*-Tum, *TET2*-Tum and Other-Tum (non-*DNMT3A*/*TET2*-Tum).[83] Additional differential expression data for *DNMT3A* and *TET2* (CH-Tum and CH-PB) are deposited on Zenodo. Transcript quantifications generated by kallisto were aggregated to the gene level and imported into DESeq2 with tximport v1.28.0.[84] Genes with low abundance were filtered out if they showed expression counts <10 in *n* samples, where *n* represents the number of samples with CH. Non-coding genes were also removed from the analysis. Analyses was conducted with the following covariates: PAM50 subtype for BRCA, microsatellite instability status for COAD, and *NF1* mutation status for mesenchymal GBM, as well as patient gender where appropriate. HPV-positive HNSC and non-clear cell renal cell carcinomas were excluded. Cancer type was controlled for in pan-cancer analyses. Differential gene expression for aging was conducted across all CPTAC participants agnostic of CH status, then in all participants without CH. Age was modeled as a continuous variable after centering and scaling, with gender and cancer type included as covariates. Genes with low abundance were filtered out if they showed expression counts <10 in 150 of the samples.

Independent hypothesis weighting methods from the IHW v1.14.0 package were used to derive multiple testing adjusted *p* values.[85] Adjusted $p < 0.05$ was required for a gene to be considered differentially expressed. Generated log fold change values were subject to ashr v2.2-63 effect size shrinkage for the purposes of ranking and downstream analysis via GSEA.[86] Volcano plots displaying DGE were generated using the EnhancedVolcano v1.20.0 R package. Venn diagrams of differentially expressed genes in CPTAC and TCGA were created using the ggvenn v0.1.10 R package. Overlapping differentially expressed genes in CH-Tum were subject to functional profiling of Gene Ontology (GO) terms using g:Profiler g:GOSt with default settings.[87]

Penalized Cox regression and development of CH-Tum prognostic signature score

The CH-Tum-associated PSS was developed using a lasso penalized Cox regression model implemented with the glmnet v4.1-8 with 10-fold cross validation to identify the optimal lambda value. TPM normalized counts for the 301 genes differentially expressed ($p_{adj} < 0.05$) in CH-Tum were analyzed for their association with OS in CPTAC. 23 genes were selected by the model for their association with prognosis:

Poor Prognosis (Positive Coefficient): *ITGA3, DHRS9, DNAJC6, VGF, IRAK2, MAPK4, EPHA5, COL6A3, SPINK1, CTHRC1, SCG5, MINAR1, DNER, FADS3*.

Good Prognosis (Negative Coefficient): *PIP5K1B, PPCS, STX17, TACC2, IKZF3, FRS2, RBKS, KLHDC7A, PGPEP1L*.

Coefficients from the penalized model can be viewed in the deposited data, and were used to generate the PSS, which is calculated as a weighted sum of TPM gene counts as follows:

$$\mathrm{PSS} = \sum \mathrm{CoefficientX}^{*}\mathrm{GeneX}$$

When applying the PSS in the TCGA dataset, we noted neuroendocrine tumors (NETs; pheochromocytoma and paraganglioma [TCGA-PCPG], pancreatic NETs [misclassified samples included in TCGA-PDAC]) specifically to have very high PSS, in excess of 3 standard deviations above the TCGA mean. As NETs generally have favorable prognosis compared to many other cancers and were not represented in the CPTAC dataset where the PSS was developed, these samples (n = 186) were excluded from analysis. Covariates in multivariate survival analysis for TCGA were the same as those applied in CPTAC, however, smoking data was not widely available for TCGA and thus not included in the model.

Deconvolution of immune cell subpopulations

Normalized RNA expression counts were obtained from DESeq2 and loaded into the CIBERSORTx cell fractions module for deconvolution.[68,88] Deconvolution was based on the LM22 immune gene signature matrix with B mode batch correction and disabled quantile normalization, and the analysis was run in absolute mode with 100 permutations.

Abundances of certain LM22 populations were aggregated to create derivative cell populations as listed below:

$$\mathrm{Monocyte/Macrophages} = \mathrm{Monocytes} + \mathrm{M0\ Macrophages} + \mathrm{M1\ Macrophages} + \mathrm{M2\ Macrophages}$$

$$\mathrm{Macrophages} = \mathrm{M0\ Macrophages} + \mathrm{M1\ Macrophages} + \mathrm{M2\ Macrophages}$$

$$\mathrm{B\ Cells} = \mathrm{Naive\ B\ Cells} + \mathrm{Memory\ B\ Cells} + \mathrm{Plasma\ Cells}$$

$$\mathrm{T\ Cells} = \mathrm{CD8^{+}\ T\ Cells} + \mathrm{Naive\ CD4^{+}\ T\ Cells} + \mathrm{Resting\ Memory\ CD4^{+}\ T}$$

$$\mathrm{Cells} + \mathrm{Activated\ Memory\ CD4^{+}\ T\ Cells} + \mathrm{Follicular\ Helper\ T\ Cells} +$$

$$\mathrm{Regulatory\ T\ Cells} + \gamma\delta\ \mathrm{T\ Cells}$$

$$\mathrm{CD4^{+}\ T\ Cells} = \mathrm{Naive\ CD4^{+}\ T\ Cells} + \mathrm{Resting\ Memory\ CD4^{+}\ T}$$

$$\mathrm{Cells} + \mathrm{Activated\ Memory\ CD4^{+}\ T\ Cells} + \mathrm{Follicular\ Helper\ T\ Cells} +$$

$$\mathrm{Regulatory\ T\ Cells}$$

$$\mathrm{NK\ Cells} = \mathrm{Resting\ NK\ Cells} + \mathrm{Activated\ NK\ Cells}$$

$$\mathrm{Dendritic\ Cells} = \mathrm{Resting\ Dendritic\ Cells} + \mathrm{Activated\ Dendritic\ Cells}$$

Absolute scoring of cell type abundances by CIBERSORTx is described elsewhere,[68,88] and represents the median expression of cell type-specific genes relative to median expression of all genes in the sample. Samples with deconvolution $p \geq 0.05$ were deemed unreliable for further analysis. Cell type abundances were compared for CH, CH-Tum, CH-PB, *DNMT3A*-Tum, *TET2*-Tum, and Other-Tum status with linear regression using a Tweedie distribution, adjusting for the same covariates used in differential expression analysis described above.

Differential protein & phosphoprotein abundance analysis

Proteomics and phosphoproteomics data pre-processing is extensively described in Li et al.[54] Differential protein and phosphoprotein analysis for CH, CH-Tum, CH-PB, and *DNMT3A*-Tum, *TET2*-Tum, and Other-Tum status was conducted using limma v3.58.1 with the same covariates described above, with adjusted p values derived using the Benjamini-Hochberg.[89] Proteins or phosphoproteins with low abundance were filtered out if they held an NA value in n samples, where n represents the number of samples with CH. Adjusted $p < 0.05$ was required for a protein to be considered differentially expressed.

Gene set enrichment analysis

Enrichment of biologically relevant pathways at the transcriptomic and proteomic level was evaluated with GSEA v4.3.2.[29,30,91] Analysis was performed using GSEAPreranked with 1000 permutations and weighted enrichment statistics. Rank scores of features input to GSEA was calculated as -log(p)*sign(log[FC]). All comparisons made were subject to GSEA for the HALLMARK gene sets and four gene sets predictive of response to anti-PD-1 ICI therapies.[36–38,69] Comparisons for CH-Tum and aging were also subject to GSEA using a 7 gene *IL1A* mRNA expression signature (*IL1A*, *BCL2L1*, *SPP1*, *CLEC4N*, *CDKN1A*, *MORRBID*, *IL1RN*).[31] Results for CH-Tum in GBM were additionally subject to GSEA using glioma stem cell signatures, the extended EnrichmentMap gene sets, and GBM TAM phenotype gene sets.[43,44,91] FDR <0.1 was required for a gene set to be considered significantly enriched.

GSEA output from the EnrichmentMap gene sets was loaded into Cytoscape v3.10.1 using EnrichmentMap v3.3.6.[91,92] Gene sets were clustered using clusterMaker2 v2.3.4 and AutoAnnotate v1.4.1, with the 15 largest clusters retained in the figure. Gene set nodes were clustered based on similarity, with edges representing higher set overlap. Cluster labels were manually generated to summarize the included gene sets.

Kinase-substrate enrichment analysis

The fold change and *p* values for each phosphoprotein calculated by limma were input into the KSEA App v1.0 to infer relative activity of kinases.[93] Kinase-substrate enrichment analysis (KSEA) was run using the PhosphoSitePlus and NetworKIN databases following the default app settings.[70,71] Kinases were considered to be significantly enriched with an FDR <0.1. Positively enriched kinases were then passed into g:Profiler g:GOSt for functional profiling, using the KEGG and Reactome pathway databases and default web settings.[72,73,87]

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were conducted using R v4.3.1. Figures were produced using ggplot2 v3.5.0, ggridges v0.5.6, and ggbeeswarm v0.7.2, along with the other R packages stated above. The graphical abstract and Figure 1A and 2D were created with BioRender.com. Statistical tests employed for analysis include the Wilcoxon rank-sum test, Spearman's rank-order correlation, Kruskal-Wallis test with post hoc Dunn's test, and generalized linear models incorporating binomial and Tweedie distributions. The wec v0.4-1 R package was used to generate weighted mean contrasts for logistic regression modeling of CH and CH-Tum ORs by cancer type.[94] The polr function of the MASS v7.3–60.0.1 R package was used to conduct ordered logistic regression for cancer stage analyses. The glmmTMB v1.1.9 was used to model CIBERSORTx fractions using linear regression with the Tweedie distribution.