nature biotechnology

ANALYSIS

https://doi.org/10.1038/s41587-019-0259-z

Integrated genomic profiling expands clinical options for patients with cancer

Nike Beaubier[1,2], Martin Bontrager[1,2], Robert Huether[1,2], Catherine Igartua[1,2], Denise Lau[1,2], Robert Tell[1,2], Alexandria M. Bobe[1], Stephen Bush[1], Alan L. Chang[1], Derick C. Hoskinson[1], Aly A. Khan[1], Emily Kudalkar[1], Benjamin D. Leibowitz[1], Ariane Lozachmeur[1], Jackson Michuda[1], Jerod Parsons[1], Jason F. Perera[1], Ameen Salahudeen[1], Kaanan P. Shah[1], Timothy Taxter[1], Wei Zhu[1] and Kevin P. White[1*]

Genomic analysis of paired tumor–normal samples and clinical data can be used to match patients to cancer therapies or clinical trials. We analyzed 500 patient samples across diverse tumor types using the Tempus xT platform by DNA-seq, RNA-seq and immunological biomarkers. The use of a tumor and germline dataset led to substantial improvements in mutation identification and a reduction in false-positive rates. RNA-seq enhanced gene fusion detection and cancer type classifications. With DNA-seq alone, 29.6% of patients matched to precision therapies supported by high levels of evidence or by well-powered studies. This proportion increased to 43.4% with the addition of RNA-seq and immunotherapy biomarker results. Combining these data with clinical criteria, 76.8% of patients were matched to at least one relevant clinical trial on the basis of biomarkers measured by the xT assay. These results indicate that extensive molecular profiling combined with clinical data identifies personalized therapies and clinical trials for a large proportion of patients with cancer and that paired tumor–normal plus transcriptome sequencing outperforms tumor-only DNA panel testing.

Genomic analysis of tumors is rapidly becoming routine clinical practice. Estimates of the proportion of patients whose testing changes their trajectory of care vary from approximately 10% to more than 50%, depending on tumor type and clinical setting[1–4]. Growing evidence suggests that patients who receive personalized therapy have better outcomes[5]. For example, the use of matching scores based on the number of genomic aberrations and therapeutic associations for each patient demonstrated that high matching scores are independently associated with a greater frequency of stable disease, longer time to treatment failure and greater overall survival[4]. Improved progression-free survival rates were observed in 43% of next-generation sequencing (NGS)-tested patients who received a genome-guided therapy, as compared to only 5.3% of patients who did not[3]. The IMPACT trial, which tested advanced-stage tumors from 3,743 patients and matched approximately 19% of patients on the basis of their tumor biology, reported a 16.2% objective response rate for patients with matched treatments versus a 5.2% rate with non-matched treatments. Additionally, the rate of 3-year overall survival for patients treated with a molecularly matched therapy was more than twice that of non-matched patients (15% versus 7%)[6].

The Tempus xT assay tests matched tumor and normal samples and generates DNA alteration data for single nucleotide variants (SNVs), insertions and deletions (indels) and copy number variants (CNVs) for approximately 600 genes, plus chromosomal rearrangements for 21 genes[7]. Tempus xT also provides whole-transcriptome RNA sequencing (RNA-seq), which detects clinically validated fusion transcripts and provides research use only (RUO) information regarding dysregulated genes and cancer type predictions for tumors of unknown origin (TUOs). xT immuno-oncology (IO) assays include clinical immunohistochemistry (IHC) testing for DNA mismatch repair (MMR) deficiency and PD-1/PD-L1 status, clinical genomic determination of microsatellite instability (MSI) status, measurement of tumor mutational burden (TMB), neoantigen predictions and RUO gene expression analyses of the tumor microenvironment (TME)[7–11]. We have previously described laboratory and analytic processes, including additional machine learning approaches for integrating genomic and imaging data generated during the course of cancer diagnosis and treatment[7,12].

Integrating DNA- and RNA-seq data into the analyses of patient samples to produce clinically validated patient reports requires advanced bioinformatics and data analytics. Properly interpreting results also requires clinical context, including patient data from physician notes and tests. Together, the genomic sequencing, computational algorithms, and software built to communicate patient clinical profiles and molecular test results to physicians are referred to as the Tempus platform. Here we examined the results of applying the Tempus platform to a cohort of 500 randomly selected patients with common tumor types, rare tumors and TUOs. We present clinical and RUO analyses from tumor–normal matched Tempus xT DNA- and RNA-seq testing on the DNA mutational spectra, whole-transcriptome profiling, chromosomal rearrangement detection, and the immunogenomic landscape based on immunotherapy biomarkers in the xT 500 cohort across multiple tumor types. Molecular insights derived from these analyses were used to match patients with evidence-based therapies and clinical trials. Lastly, we present a comparison of the Tempus platform to tumor-only sequencing, examining germline versus somatic variant detection and therapeutic matching.

[1]Tempus Labs, Inc., Chicago, IL, USA. [2]These authors contributed equally: Nike Beaubier, Martin Bontrager, Robert Huether, Catherine Igartua, Denise Lau, Robert Tell. *e-mail: kevin@tempus.com

Results

xT 500 analysis cohort. We selected a cohort of 500 paired tumor–normal samples sequenced with the xT assay in 2017 or 2018, which we refer to as the xT 500 cohort. The cases included were randomly selected from the Tempus de-identified database of structured genomic and clinical data. To be eligible for inclusion in the cohort, each case required complete data elements for tumor–normal DNA-seq and clinical data from abstracted medical records. After filtering for eligibility, a set of patients was sampled via a pseudo-random number generator. Patients were assigned to one of eight cancer types on the basis of pathologic diagnosis, with 50 patients per category of brain, breast, colorectal, lung, ovarian, endometrial, pancreatic and prostate cancer types. Additionally, 50 tumors from a combined set of rare malignancies and 50 TUOs were included, for a total of 500 patients. The xT 500 cohort was roughly balanced between male and female patients ($n=212$ and $n=288$, respectively), given the tumor types included. The specimens sequenced were primarily from patients with advanced-stage cancer, with 7.8% from patients with recurrent metastatic disease (Supplementary Table 1).

Genomic analyses of the xT 500 cohort. We examined the xT 500 mutational spectra in comparison to broad patterns of genomic alteration observed in large-scale studies across cancer types (see Methods). First, we identified alterations by gene (Fig. 1a and Supplementary Fig. 1) and found that the most commonly mutated genes were well-known for driver mutations, including *TP53*, *KRAS*, *PIK3CA*, *CDKN2A*, *PTEN*, *ARID1A*, *APC*, *ERBB2* (*HER2*), *EGFR*, *IDH1* and *CDKN2B*. As expected, homozygous deletions were commonly observed in the tumor suppressor genes *CDKN2A*, *CDKN2B* and *PTEN*. Mutational spectrum data were compared to a published pan-cancer analysis using the Memorial Sloan Kettering Cancer Center (MSKCC) IMPACT panel (Fig. 1b)[13]. The same commonly mutated genes were observed at similar relative frequencies in the two datasets, indicating that the xT 500 mutational spectra are representative of tumors sequenced in this previous large-scale study.

As part of the xT assay, each sample underwent whole-transcriptome profiling by RNA-seq. Cancer type was predicted using a random forest classifier trained on an internal gene expression reference database with labels from 33 cancer types (see Methods). Cancers were correctly classified for 100% of breast, 98% of prostate, 96% of brain and ovarian, 94% of colorectal, 92% of pancreatic, 88% of lung and 58% of endometrial tumor samples (Fig. 2). In 60% of misclassifications, the second prediction was matched to the cancer diagnosis. In 36% of misclassifications, gynecologic cancers (ovarian and endometrial) accounted for the difference and the classification was influenced by low tumor purity, as in the case of misclassified endometrial cancers ($P=0.02$; Supplementary Fig. 2a,b). Another notable trend was the misclassification of lung, rare and unknown origin squamous cell carcinomas (SCCs) as head and neck SCC due to a shared SCC signature[14]. A total of 11.1% of misclassifications were affected by signal contributions from background tissue transcriptome profiles, as in the case of misclassified metastatic samples ($P=0.09$; Supplementary Fig. 2c). Because the classifier was accurate for most tumor types, most TUOs could be matched to the appropriate tissue of origin.

We also evaluated oncogenic gene fusions[15]. Fusions were detected by DNA-seq of 21 common gene rearrangement targets and by whole-transcriptome RNA-seq analysis (see Methods)[7]. Of the fusions detected, 26 were identified by both DNA- and RNA-seq, 2 were detected by DNA-seq alone, and 4 were detected by RNA-seq alone (Supplementary Fig. 3a). Within the four RNA-seq-detected fusions, two were potentially detectable but were not detected by DNA-seq, and two were not detected because they were not represented in the breakpoints assayed in DNA-seq, illustrating the value of an unbiased whole-transcriptome approach for fusion identification. The predicted structures of these two fusions were further examined, revealing in-frame fusions with intact tyrosine kinase domains, such as *RET* and *NTRK3*, which are therapeutically targetable (Supplementary Fig. 3b,c).

We next characterized the immunogenomic landscape of the xT 500 cohort. TMB is a key biomarker of immunotherapy response[8,16]. In the cohort, TMB ranged from 0 to 54.2 mutations per megabase of DNA across cancer types, with a median of 2.09 (see Methods; Fig. 3a). These xT assay-derived TMB values were highly correlated with whole-exome TMB (see Methods; Supplementary Fig. 4). We identified a hypermutated tumor population across cancer types with significantly higher median TMB. These TMB-high samples included cancers previously associated with a low TMB, such as glioblastoma. Consistent with previous reports, TMB was highly correlated with neoantigen load ($R=0.931$, $P=8.20\times10^{-199}$), which is an estimate of the number of somatic mutations presented to the immune system (see Methods; Fig. 3b). The TMB-high population also contained all MSI-high (MSI-H) samples (see Methods). The remaining TMB-high samples were associated with mutational signatures related to smoking, UV exposure and APOBEC-mediated mutagenesis (see Methods; Supplementary Fig. 5).

We then assessed the relationship between tumor immunogenicity and the levels of immune infiltration and activation. Cytotoxic immune activity levels measured by the cytolytic index (CYT) were significantly higher in hypermutated populations ($P=1.95\times10^{-4}$ for TMB-high, $P=2.50\times10^{-2}$ for MSI-H; Fig. 3c) (see Methods)[17,18]. Additionally, an estimation of the immune cell composition using an RNA deconvolution model (see Methods) revealed that inflammatory cells, such as $CD8^+$ T cells and M1 macrophages, were significantly higher in TMB-high samples ($P=4.9\times10^{-4}$ and $P=1.4\times10^{-7}$, respectively), whereas frequencies of non-inflammatory immune cells, such as monocytes, were significantly higher in TMB-low samples ($P=2.0\times10^{-4}$) (Fig. 3d).

Increased immune pressure from infiltrating immune cells can lead tumors to express higher levels of immune checkpoint molecules, such as PD-L1 (*CD274*). Accordingly, RNA-seq determined that PD-L1 expression was significantly higher in the immune-infiltrated TMB-high tumors ($P=6.69\times10^{-4}$; Fig. 3e). *CD274* expression was also highly correlated with the expression of its binding partner *PDCD1* (PD-1) on immune cells ($R=0.59$, $P\leq2.2\times10^{-16}$), as well as T cell lineage-specific markers, such as *CD3E* ($R=0.63$, $P\leq2.2\times10^{-16}$; Fig. 3f). Furthermore, samples that stained positive for PD-L1 via Tempus' clinically validated IHC tests clustered with higher *CD274* RNA expression levels, suggesting that *CD274* expression may be an indicator of PD-L1 protein levels.

Finally, a 28-gene interferon (IFN)-γ-related signature was used to determine whether any patients lacking classically defined immunotherapy biomarkers exhibited traits of immunologically active tumors (see Methods)[19]. We found that tumor samples could be broadly categorized as immunologically active or silent tumors. Our results support this stratification, with the immunologically active population enriched for samples that were TMB-high, MSI-H or PD-L1 IHC-positive (Fig. 3g). Patients within this immunologically active cluster who lack traditional immunotherapy biomarkers represent an interesting population that might benefit from immunotherapy. Overall, the IFN-γ signature scores were significantly different between patients on the basis of their immunotherapy biomarker status ($P=3.77\times10^{-4}$; Fig. 3h). In particular, TMB-high, MSI-H or PD-L1 IHC-positive and TMB-high tumors expressed higher levels of IFN-γ-related genes than tumors lacking these biomarkers ($P\leq0.05$).

Identification of therapeutic options and clinical trials. We investigated the extent to which molecular profiling aids in identifying therapies and clinical trials. First, we identified the proportion of patients matched to therapies within each therapeutic evidence tier (see Methods). Evidence tiers contain somatic biomarker

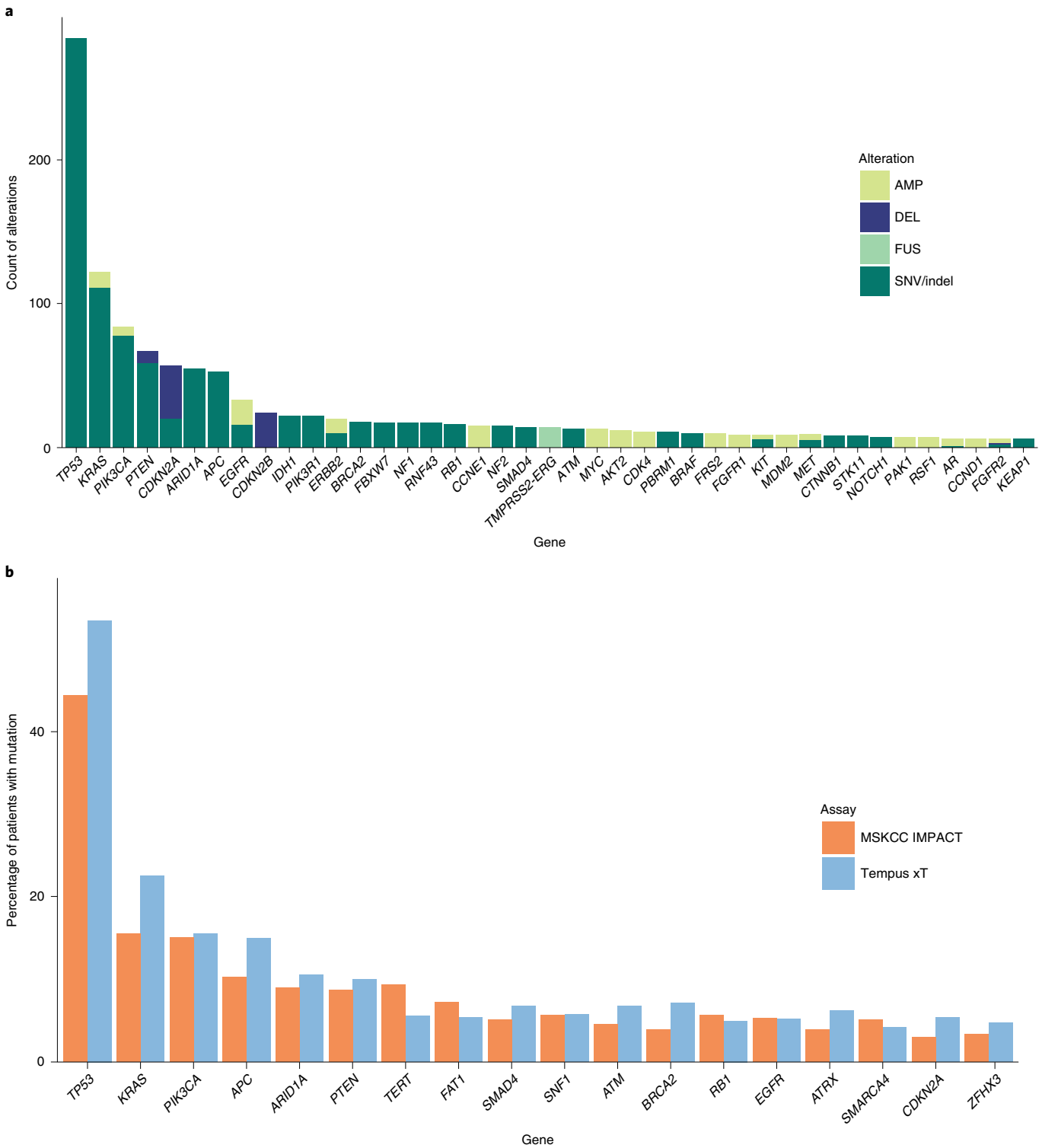


Fig. 1 | Mutational spectrum of the xT 500 cohort. a, Alteration types among the xT 500 cohort for the most commonly mutated genes. Alterations including single nucleotide variants and indels (SNV/indel), fusions (FUS), and a subset of CNVs, amplifications (AMP) and deletions (DEL) were grouped. Those appearing in at least five patients (1% prevalence) are plotted. **b**, Comparison of the Tempus xT SNV/indel detection assay against the MSKCC IMPACT study plotted by the prevalence of altered genes commonly known to be hallmarks of cancer.

information related to therapeutic response and/or resistance and are divided into four categories: tier I level A (IA), tier I level B (IB), tier II level C (IIC) and tier II level D (IID). Tiers are ranked according to the strength of biomarker evidence, ranging from consensus clinical guidelines to case reports and preclinical evidence[20]. Across all cancer types, 91.4% of patients were matched to a therapeutic option on the basis of all evidence levels for response to therapy, and 22% of patients were matched to a therapeutic option on the basis of all levels of evidence for resistance to therapy (Fig. 4a). Together, response and resistance matching based on the highest evidence tiers (IA and IB) accounted for 29.6% of cases, whereas 62% of cases were matched for either response or resistance

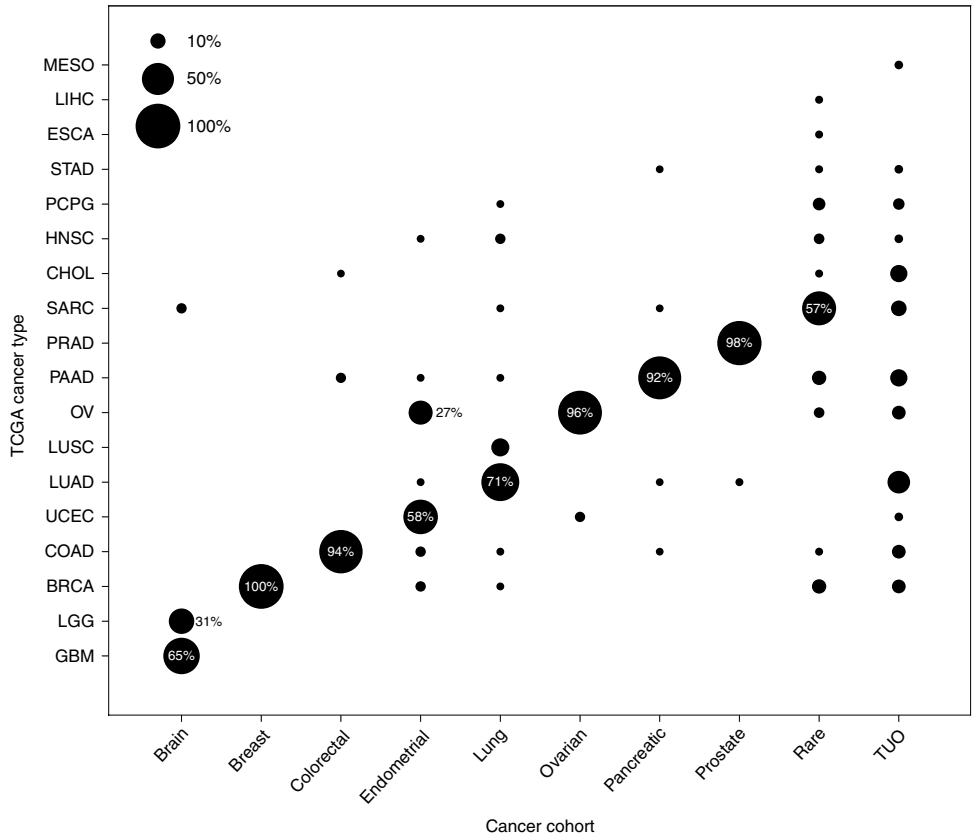


Fig. 2 | Predicted TCGA cancer types for samples within each xT 500 cohort cancer type. Cancer type predictions are based on a random forest model trained on a Tempus internal reference dataset. Bubble size corresponds to the percentage of samples from the cohort predicted to have a given TCGA cancer type. Bubbles representing ≥25% of samples are labeled. BRCA: breast invasive carcinoma, COAD: colon adenocarcinoma, CHOL: cholangiocarcinoma, ESCA: esophageal carcinoma, GBM: glioblastoma multiforme, HNSC: head and neck squamous cell carcinoma, LGG: brain lower grade glioma, LIHC: liver hepatocellular carcinoma, LUAD: lung adenocarcinoma, LUSC: lung squamous cell carcinoma, MESO: mesothelioma, OV: ovarian serous cystadenocarcinoma, PAAD: pancreatic adenocarcinoma, PCPG: pheochromocytoma and paraganglioma, PRAD: prostate adenocarcinoma, SARC: sarcoma, STAD: stomach adenocarcinoma, UCEC: uterine corpus endometrial carcinoma.

on the basis of evidence of lower clinical utility (IIC and IID) (Fig. 4a and Supplementary Table 2). The tiers of evidence-based therapies matched to patients varied significantly by cancer type (Fig. 4b). For example, 56% of matches for patients with lung cancer were made using tier IA evidence for *EGFR*, *KRAS* and *MET*, as well as targets that have emerged more recently, such as *BRAF* and *ERBB2* (HER2). Additionally, 58% of matches for patients with colorectal cancer used tier IA evidence, most of which were based on resistance to therapy due to *KRAS* mutations. In contrast, patients with pancreatic cancer were matched exclusively according to lower-tier evidence on the basis of resistance to anti-EGFR therapy due to *KRAS* mutations.

We next determined the contribution of each molecular assay component to therapy matching. First, therapeutic matches based on clinically actionable CNVs, SNVs and indels were examined (Supplementary Fig. 6a,b). CNVs accounted for tiers IA and IB evidence-based matching of 29 patients (160 patients across all levels of evidence). SNVs and indels accounted for tiers IA and IB evidence-based matching of 124 patients (429 patients across all levels of evidence). Although most patients exhibited a mutation within a gene of clinical relevance, the context of these mutations within tumor type and evidence level was considered to fully assess their clinical utility (Supplementary Fig. 6c). Some of the most commonly mutated genes had low-level evidence or evidence related to resistance. For instance, *TP53* had tier IIC evidence and drugs in clinical trials, whereas *KRAS* had tier IA evidence in two cancer types for resistance to anti-EGFR therapy. Many of the less common mutations had tier IA evidence for targeted therapies across a variety of cancer types. A notable example was PARP inhibitors for *BRCA1*- and *BRCA2*-mutated breast and ovarian cancers, which are currently in clinical trials and are used off-label in other cancer types harboring *BRCA* mutations, such as prostate and pancreatic cancers.

Therapeutic options were also matched to clinically relevant gene fusions detected via DNA- and RNA-seq (Supplementary Fig. 6d). These fusions were clear drivers of cancer, part of consensus therapeutic guidelines, and identified with high sensitivity by the Tempus xT assay[7]. Therapeutic options for fusions occurred in 29 patients (5.8%) of the xT 500 cohort, indicating that comprehensive fusion identification for all patients leads to therapeutic matching for a modest but clinically important subset of patients. Similarly to previous reports, most fusion events detected were *TMPRSS2-ERG* in prostate cancer[21]. Although no clear clinical interventions are associated with this fusion, *TMPRSS2-ERG* fusions were given a tier IID evidence level owing to early evidence regarding therapeutic response[22,23]. Of the 12 non-prostate cancer fusions, one was rated as evidence tier IA, one was rated as IIC and 10 were rated as IID.

We next examined the potential for therapy matching using the expression profiles of clinically relevant genes selected on the basis of their relevance to disease diagnosis, prognosis and/or possible therapeutic intervention (see Methods). Over- or under-expression calls were reported in 133 patients (28.1%) for 16 genes, with therapeutic evidence based on clinical, case or preclinical studies (Fig. 4c, Supplementary Fig. 7 and Supplementary Table 3). Metastatic tumors were equally likely to have at least one reportable expression

call as non-metastatic tumors. The most commonly reported over-expressed gene was *NRG1*, which was observed in 35 cases (7.3% of samples) across the cohort. *NRG1* has been shown to play a biological role with treatment implications across cancer types[24–27]. Over-expression of *NRG1* has been associated with primary cetuximab resistance in colon cancer cell lines in the absence of RAS pathway mutations[28,29], primary resistance to trastuzumab or lapatinib in *ERBB2* (HER2)-amplified breast cancer cells[30,31], and response to monoclonal HER2-directed antibodies in lung[32] and ovarian cancers[25].

We investigated the percentage of patients eligible for immunotherapy based on the xT IO biomarkers assayed in the cohort. There were 52 patients (10.4%) identified as potential candidates for immunotherapy on the basis of TMB, MSI status and PD-L1 IHC results (Fig. 4d). The number of MSI-H and TMB-high cases varied among cancer types, with 22 patients (4.4%) positive for both biomarkers. A PD-L1-positive IHC result alone was measured in 15 patients (3%) and occurred most frequently in patients with lung cancer. TMB-high status alone was measured in 13 patients (2.6%), who primarily had lung and or breast cancer. Lastly, a PD-L1-positive IHC result and TMB-high status were rarely observed simultaneously (0.4%).

As noted above, therapy matches based on level IA or IB evidence for SNVs, CNVs and fusions alone were observed for 29.6% of patients. With the addition of more comprehensive molecular profiling that included gene expression and immunotherapy biomarker information, we observed an increase in the proportion of patients with therapy matches to 43.4% of the xT 500 cohort (217 patients) (Fig. 4e and Supplementary Table 2)[33]. This percentage increased even further, to 93.6% (468 patients), when matches from level IIC and IID evidence and preclinical RNA-based evidence were included.

Additionally, 1,966 clinical trial matches were reported for the xT 500 cohort. On the basis of molecular and clinical data (see Methods), at least one clinical trial option was reported for 96.2% of the cohort (481 patients). Examples of the criteria used to match patients to a clinical trial option are shown in Supplementary Table 4. At least one biomarker-based clinical trial match was made for 76.8% of the cohort according to a gene variant on the patients' xT report. Of the patients who were not matched to a biomarker-based clinical trial, 19.4% were matched to at least one disease-based clinical trial from clinical data alone. The frequency of biomarker-based clinical trial matches outnumbered disease-based matches and varied by cancer diagnosis (Fig. 4f). For example, patients with gynecologic or pancreatic cancers typically received biomarker-based clinical trial matches, whereas patients with rare cancers had an almost equal ratio of biomarker-based to disease-based trial matching. The differences observed between biomarker- and disease-based trial matching were most likely due to the frequency of targetable alterations and the heterogeneity of these cancer types.

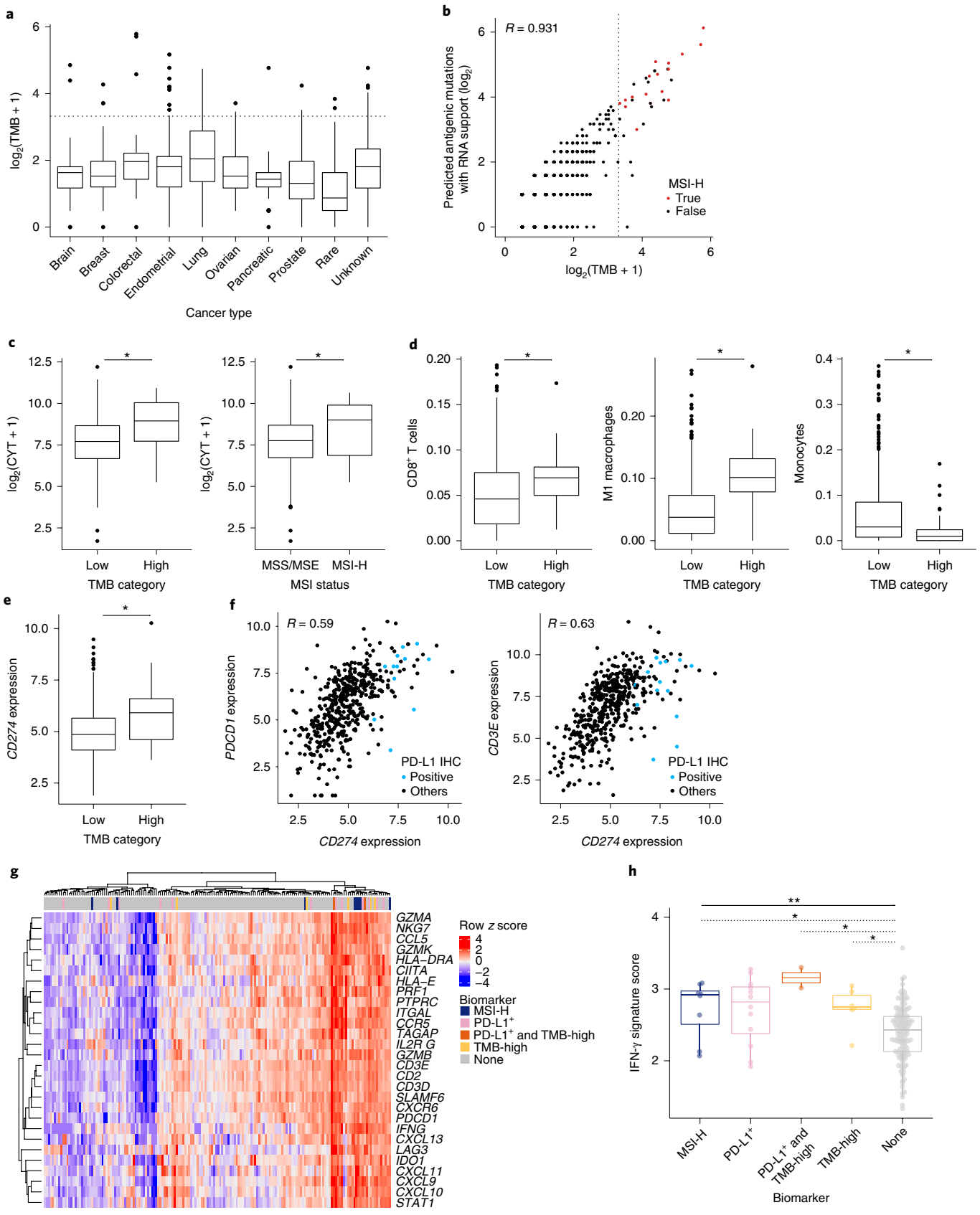
Comparison of the full Tempus platform with tumor-only analyses. Most commercial oncology assays only test tumor samples. Because paired tumor–normal samples were sequenced within the xT 500 cohort, we were able to examine the effect of germline sequencing on the accuracy of somatic mutation calling. We randomly selected 50 cases from the cohort with a range of TMB profiles and re-evaluated them using a tumor-only analytical pipeline. We identified 8,557 coding variants after filtering with a publicly available population database (see Methods)[34]. By further filtering with an internally developed list of technical artifacts, an internal pool of normal samples, and classification criteria (see Methods), the number of variants was reduced to 642 while still retaining all true somatic alterations (72.3%) (Fig. 5a).

Within the 642 filtered tumor–only variants, 27.7% of variants were classified as somatic false positives—that is, true germline variants or artifacts. The use of tumor–normal sequencing data allowed for these false-positive variants to be filtered and more accurately classified as germline variants (Fig. 5a,b and Supplementary Table 5). When we further separated the dataset by classification criteria, 1.10% of germline variants were classified as pathogenic and, thus, potentially clinically actionable. One such example involved a *BRCA* mutation with somatic loss of heterozygosity in colorectal cancer. In tumor types where *BRCA* mutations are not common, such as colon cancer, *BRCA* mutations with loss of heterozygosity would trigger a recommendation for PARP inhibitor therapy. In cases without loss of heterozygosity, genetic counseling would be recommended instead. The ability to differentiate between these cases is enhanced by the more accurate classification of somatic versus germline variants via tumor–normal sequencing.

To assess the effect of tumor-only testing on therapeutic matching, we evaluated which therapies would be offered to the 50 patients in two scenarios: a tumor-only test versus a full xT test (matched tumor–normal DNA-seq plus RNA-seq and IO analyses). We found that divergent therapies would have been reported for 8 of the 50 patients (16%) if they had received a tumor-only test alone rather than a full xT test (Supplementary Table 6). Of these eight patients, four had different hypothetical treatment matches owing to information obtained via RNA-seq or to the tumor having somatic mutations with low clonality, which are difficult to detect in a tumor-only test. One tumor-only prostate cancer DNA-seq result did not show any contraindication to the anti-androgen therapy the patient was receiving; but RNA-seq included in the full xT test showed androgen receptor over-expression, indicating possible resistance.

Fig. 3 | Immunogenomic landscape of solid tumors in the xT 500 cohort. a, TMB is shown by cancer type ($n=50$ per group) and plotted on a $\log_2$(TMB+1) scale. **b**, Neoantigen predictions generated for each patient ($n=493$) are plotted against TMB (Pearson's $R=0.931$, $P=8.20\times10^{-199}$). The Pearson correlation coefficient is shown, and the dotted lines in **a** and **b** denote the threshold for defining TMB-high samples. **c**, Distribution of CYT by TMB category (TMB-high, $n=35$, TMB-low, $n=439$) and MSI status (MSI-H, $n=20$, MSS/MSE, $n=454$). Statistical significance was determined using a two-sided Wilcoxon rank-sum test (TMB, $P=1.95\times10^{-4}$; MSI, $P=2.50\times10^{-2}$) **d**, The composition of immune cell infiltrate was estimated using a support vector regression model. The distribution of the estimated percentages of selected immune cell types is shown by TMB category. Statistical significance was determined using a two-sided Wilcoxon rank-sum test, and *P* values were adjusted for multiple testing using the Benjamini-Hochberg method ($CD8^+$ T cells, $P=4.9\times10^{-4}$; M1 macrophages, $P=1.4\times10^{-7}$; monocyte, $P=2.0\times10^{-4}$). **e**, Expression of *CD274* (PD-L1) in $\log_2$(TPM+1) by TMB category. Statistical significance was determined using a two-sided Wilcoxon rank-sum test ($P=6.69\times10^{-4}$). **f**, Expression of *CD274* as compared to expression of *PDCD1* (PD-1) and *CD3E* ($n=474$). Samples that stained positive for PD-L1 by IHC are shown in blue. Other samples either stained negative for PD-L1 by IHC or were not tested. The Pearson correlation coefficients are shown (*PDCD1*, $P<2.2\times10^{-16}$; *CD3E*, $P<2.2\times10^{-16}$). **g**, Heat map of the 28 IFN-γ-related genes in the cohort. Patients with RNA data and complete biomarker testing were categorized into the following groups: PD-L1-positive IHC ($n=15$); TMB-high ($n=6$); PD-L1-positive IHC and TMB-high ($n=2$); MSI-high ($n=8$); and none ($n=150$). **h**, IFN-γ score as calculated by the arithmetic mean of the 28 genes compared between patient groups. Statistical significance was determined using the Kruskal–Wallis test ($P=3.77\times10^{-4}$) followed by Dunn's test for multiple comparisons (MSI-H, $P=0.024$; PD-L1-positive IHC, $P=0.066$; PD-L1-positive IHC and TMB-high, $P=0.037$; TMB-high, $P=0.029$). * $P\leq0.05$, ** $P<0.01$. In the boxplots, the upper and lower hinges represent the first and third quartile. The whiskers extend to the most extreme value within 1.5 times the interquartile range on either end of the distribution. The center line represents the median.

The other three patients had divergent therapy matches owing to the tumor-only test reporting a germline mutation as somatic. These patients potentially would not have received genetic counseling with a tumor-only test. Lastly, we compared therapies matched for all DNA variants detected by the tumor-only dataset to therapies matched by a patient-facing website, My Cancer Genome (MCG).



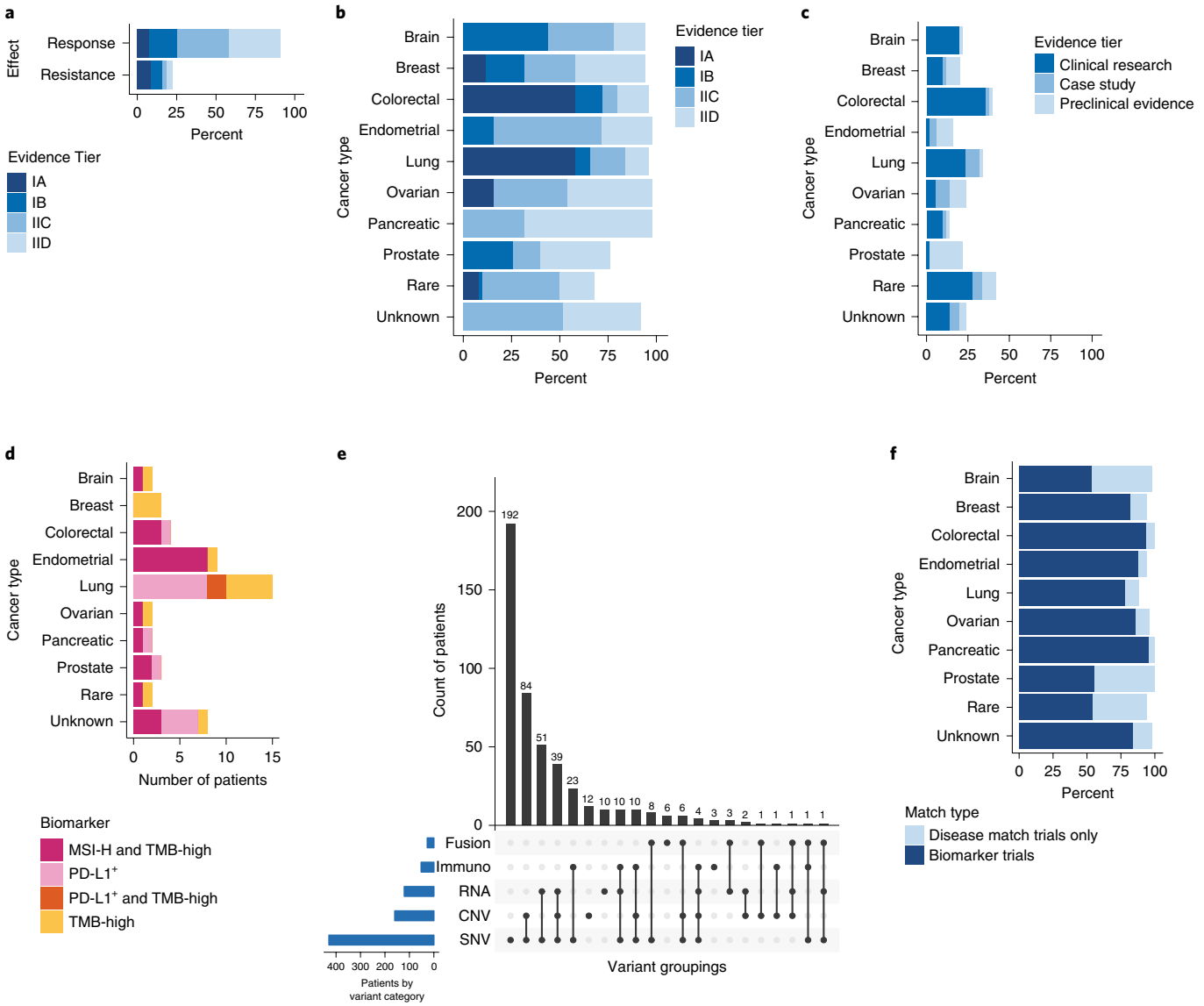


Fig. 4 | Evidence-based therapy and clinical trial matching. a, Distribution of samples with matched evidence for response or resistance to a therapeutic option. **b**, Distribution of samples with matched therapeutic evidence tiers by cancer type. **c**, Distribution of samples with reportable gene expression calls and therapeutic evidence by cancer type. **d**, Number of patients positive for immunotherapy-relevant biomarkers. **e**, UpSetR plot of the number of patients for whom therapeutic options were reported on the basis of five variant categories: SNVs/indels, CNVs, RNA-seq gene expression calls, immunotherapy biomarkers and fusions. The horizontal blue bars on the left of the matrix represent the total number of patients who had at least one matched therapy reported on the basis of each variant category. The black vertical bars represent the total counts of patients with matched therapies reported on the basis of groupings of variant categories. The matrix with connected filled circles represents the groupings of variant categories that correspond to the counts for the black vertical bars. **f**, Counts of the most common biomarker-based trial matches versus disease-based trial matches for each cancer type.

A total of 43 cases were matched to therapies via the full xT test, whereas only 5 cases were matched to therapies via MCG.

Discussion

We examined the molecular and clinical insights gained from extensive genomic profiling, including matched tumor and germline DNA-seq and whole-transcriptome RNA-seq. Comparison between genomic alterations in the xT 500 cohort and previously published clinical NGS data indicated that our cohort is representative of the mutational spectra observed within and across tumor types[35–37]. The xT tumor–normal sequencing pipeline robustly classifies true somatic versus germline variants and eliminates the 27.7% somatic false-positive rate observed in the tumor-only analysis. Erroneous identification of germline variants as somatic mutations can negatively affect patient care. For example, germline variants in genes that can be mutated in somatic cells or in the germline, such as *BRCA*, would be classified as somatic in a tumor-only analysis, missing the opportunity to provide germline findings with genetic counseling and cancer screening recommendations[38].

Whole-transcriptome profiling is another key attribute of the Tempus platform. RNA expression data are currently RUO; however, in the future, oncologists may use RNA findings in conjunction with clinical, pathologic, radiologic and CAP/CLIA-validated molecular test data for the assessment of patients who have failed multiple lines of therapy. A total of 28.1% of patients with RNA expression calls were matched to some level of evidence-based therapy in a tissue-agnostic fashion. For example, National Comprehensive Cancer Network (NCCN) guidelines for breast, gastric and lung cancers

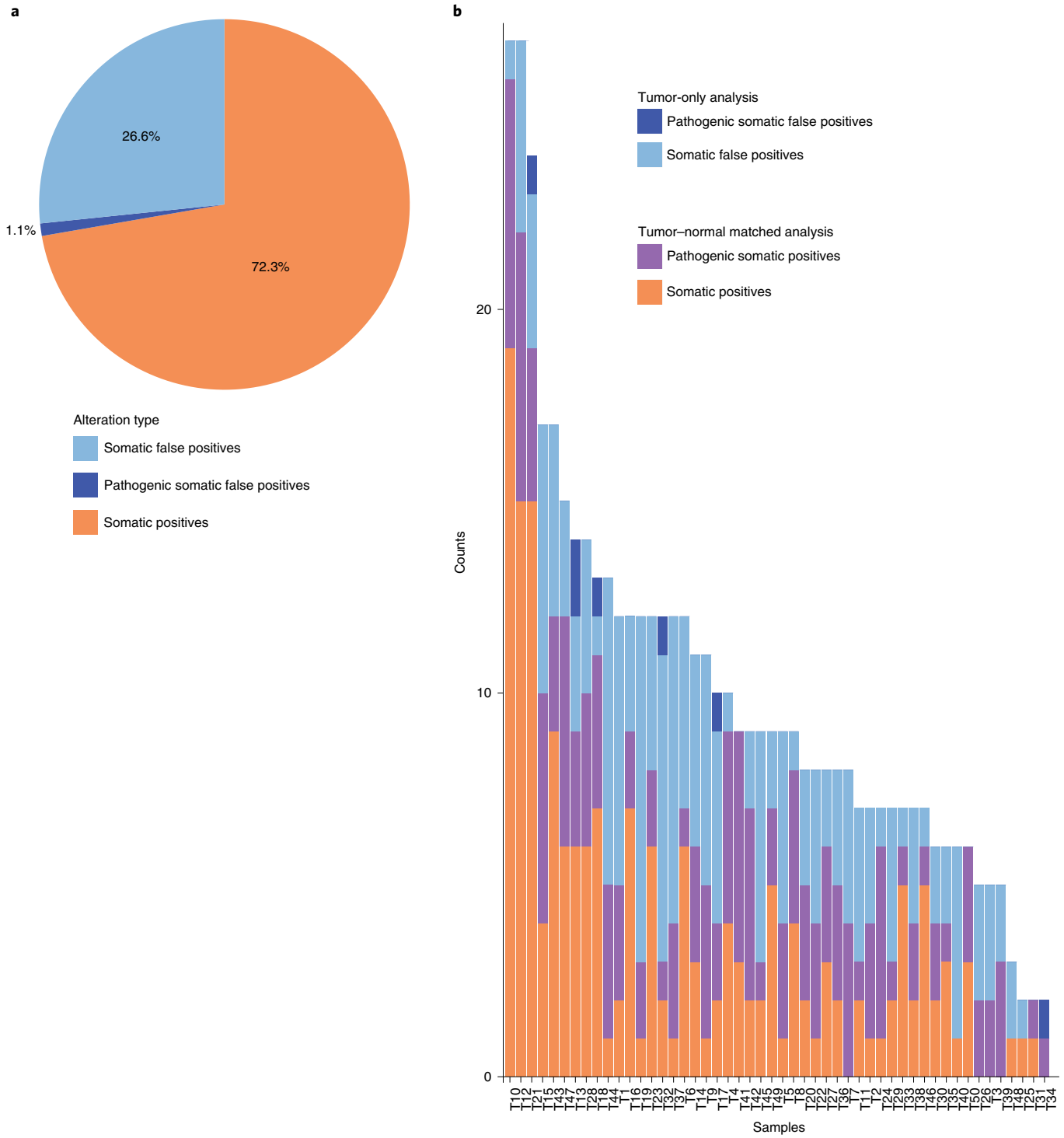


Fig. 5 | Tumor-only versus tumor–normal analyses. a, Percentage of somatic false positives (germline in origin) detected in the tumor-only analyses of 50 patient samples. **b**, Germline detection in tumor–normal matched DNA sequencing versus tumor-only sequencing. Tumor–normal counts of true somatic variants (orange and purple bars) are included for reference.

recommend FDA-approved drugs targeting HER2 overexpression. Patients with HER2 overexpression in other cancer types may also benefit from anti-HER2 therapies[39]. Because HER2 evaluation by IHC is not standard practice for most cancer types, these patients cannot be identified without a comprehensive profiling method. In addition, RNA expression data provide insight into tumor type, which helps to refine diagnoses for TUOs and determine chemotherapy regimens.

Likewise, immunotherapy RNA-seq data analyses identified patients with and without traditional biomarkers who might benefit from immunotherapy. Immunotherapy has provided lasting results for some previously untreatable cancers[40]. However, the need for effective biomarkers of immunotherapy response is clear considering the low proportion of patients who experience clinical benefit, the associated adverse events and the cost of treatment. With the growing use of immunotherapy, it is becoming increasingly important to

measure TME attributes that signal potential responsiveness in patients[41]. PD-L1 IHC and MSI status are currently used as complementary or companion diagnostics for many cancer indications. Furthermore, TMB is an emergent biomarker associated with clinical benefit from checkpoint blockade and is being tested as a companion diagnostic[8,16,42]. Approximately 10% of patients in the xT 500 cohort were positive for at least one of these IO biomarkers and, consequently, could be candidates for immunotherapy. Additional context about the immunologic phenotype of tumors was derived from RUO transcriptome analysis. Patients lacking IO biomarkers still grouped into immunologically active clusters, suggesting that these biomarkers may not fully capture information about immunotherapy-conducive TMEs. Thus, further studies to identify other biomarkers will increase the understanding of TMEs and may help identify additional patients who would benefit from immunotherapy.

Overall, the integration of molecular data and structured clinical data resulted in precision therapy matches with tier IA or IB levels of evidence for 43.4% of the xT 500 cohort. A precision medicine option across all tiers and levels of therapeutic evidence was reported for 93.6% of the cohort. Identification of both therapeutic response and resistance across all evidence levels provides valuable information for physicians that could influence the prescription or timing of therapies. Integrating molecular data with clinical data also allows clinical trial matching for the most vulnerable patient populations. For example, although pancreatic cancer has few well-established therapeutic options, we were able to identify biomarker-based clinical trial options for 94% of patients with pancreatic cancer.

More broadly, our results indicate that the overall population of patients with tumor types lacking viable options may benefit from molecular testing that matches patients to therapies and trials for which they otherwise would not have been considered. According to the American Cancer Society, only 27% of patients in the United States are provided with the option to enroll in a local clinical trial[43]. Furthermore, only an estimated 3–8% of patients enroll in clinical trials nationwide[44,45]. The use of molecular testing and structured clinical data allowed us to provide 96.2% of the patients in our cohort with at least one clinical trial option. The fact that most patients were matched to biomarker-based trials (76.8%) likely reflects both the large number of clinical trials that are biomarker dependent and the extensive genomic profiling performed on the xT 500 cohort.

Additionally, the value of comprehensive, multimodality testing is demonstrated by the Tempus platform's ability to find rare events with treatment implications in this cohort. For example, to meet NCCN guidelines for 'broad' molecular profiling of lung adenocarcinomas, testing would only have to include *EGFR*, *ALK*, *ROS1*, *BRAF*, *KRAS*, *MET*, *RET*, *NTRK* and *ERBB2* (*HER2*), and IHC analysis of PD-L1 expression. In one patient with lung adenocarcinoma, none of the above targetable genes contained an alteration, including *CD274* (PD-L1), for which staining was negative by IHC. If testing had stopped there, the patient would not have been eligible for targeted immunotherapy. However, the xT assay revealed a pathogenic germline mutation in the MMR gene *MSH3*, with somatic loss of heterozygosity. *MSH3* deficiency does not cause the MSI-H phenotype and is not included in standard IHC panels for MMR deficiency. Additionally, an elevated TMB of 19.6 mutations per megabase was observed, suggesting that the tumor had an increased probability of immunotherapy response[8]. These findings would likely motivate an oncologist to consider using immunotherapy.

In summary, our results indicate that paired tumor–normal DNA-seq and RNA profiling of patient cancer biopsies yields high match rates to targeted therapies and clinical trials. This study determines clinically relevant insights via comprehensive genomic analysis of a de-identified dataset derived from patients with cancer nationwide. Our results demonstrate the value of harnessing tumor–normal genomic sequencing, gene expression profiling, genomic rearrangement detection and immunotherapy biomarker prediction to address emergent clinical indications. These results also illustrate the value of integrating and contextualizing clinical and molecular data to provide physicians with distilled information regarding their patients' disease and potentially actionable characteristics. These insights help to maximize personalized therapeutic options for a broader proportion of patients with cancer, which cannot be attained through limited tumor-only DNA-seq panels alone.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41587-019-0259-z.

Received: 17 December 2018; Accepted: 5 August 2019;
Published online: 30 September 2019

References

1. Fernandes, G. et al. Next-generation sequencing-based genomic profiling: ostering innovation in cancer care? *Clinics* **72**, 588–594 (2017).
2. Radovich, M. et al. Clinical benefit of a precision medicine based approach for guiding treatment of refractory cancers. *Oncotarget* **7**, 56491–56500 (2016).
3. Dhir, M. et al. Impact of genomic profiling on the treatment and outcomes of patients with advanced gastrointestinal malignancies. *Cancer Med.* **6**, 195–206 (2017).
4. Wheler, J. J. et al. Cancer therapy directed by comprehensive genomic profiling: a single center study. *Cancer Res.* **76**, 3690–3701 (2016).
5. Gong, J. et al. Value-based genomics. *Oncotarget* **9**, 15792–15815 (2018).
6. The ASCO Post. 2018 ASCO: IMPACT trial matches treatment to genetic changes in the tumor to improve survival across multiple cancer types. *The ASCO Post* http://www.ascopost.com/News/58897 (2 June 2018).
7. Beaubier, N. et al. Clinical validation of the tempus xT next-generation sequencing targeted oncology assay. *Oncotarget* **10**, 2384–2396 (2019).
8. Goodman, A. M. et al. Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. *Mol. Cancer Ther.* **16**, 2598–2608 (2017).
9. Le, D. T. et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **357**, 409–413 (2017).
10. Miller, A. et al. High somatic mutation and neoantigen burden are correlated with decreased progression-free survival in multiple myeloma. *Blood Cancer J.* **7**, e612 (2017).
11. Desrichard, A., Snyder, A. & Chan, T. A. Cancer neoantigens and applications for immunotherapy. *Clin. Cancer Res.* https://doi.org/10.1158/1078-0432.CCR-14-3175 (2016).
12. Reiman, D. et al. Integrating RNA expression and visual features for immune infiltrate prediction. *Biocomputing* **2019**, 284–295 (2018).
13. Zehir, A. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).
14. Newton, Y. et al. TumorMap: exploring the molecular similarities of cancer samples in an interactive portal. *Cancer Res.* **77**, e111–e114 (2017).
15. Solomon, B., Varella-Garcia, M. & Camidge, D. R. ALK gene rearrangements: a new therapeutic target in a molecularly defined subset of non-small cell lung cancer. *J. Thorac. Oncol.* **4**, 1450–1454 (2009).
16. Chae, Y. K. et al. Association of tumor mutational burden with DNA repair mutations and response to anti-PD-1/PD-L1 therapy in non-small cell lung cancer. *Clin. Lung Cancer* https://doi.org/10.1016/J.CLLC.2018.09.008 (2018).
17. Rooney, M. S. et al. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).
18. Roufas, C. et al. The expression and prognostic impact of immune cytolytic activity-related markers in human malignancies: a comprehensive meta-analysis. *Front. Oncol.* **8**, 27 (2018).
19. Ayers, M. et al. IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Invest.* **127**, 2930–2940 (2017).
20. Li, M. M. et al. Standards and guidelines for the interpretation and reporting of sequence variants in cancer. *J. Mol. Diagnostics* **19**, 4–23 (2017).
21. Wang, Z. et al. Significance of the *TMPRSS2:ERG* gene fusion in prostate cancer. *Mol. Med. Rep.* **16**, 5450–5458 (2017).
22. Chatterjee, P. et al. The *TMPRSS2-ERG* gene fusion blocks XRCC4-mediated nonhomologous end-joining repair and radiosensitizes prostate cancer cells to PARP inhibition. *Mol. Cancer Ther.* **14**, 1896–1906 (2015).
23. Tomlins, S. A. et al. Recurrent fusion of *TMPRSS2* and *ETS* transcription factor genes in prostate cancer. *Science* **310**, 644–648 (2005).

24. Hegde, G. V. et al. Blocking NRG1 and other ligand-mediated Her4 signaling enhances the magnitude and duration of the chemotherapeutic response of non-small cell lung cancer. *Sci. Transl. Med.* **5**, 171ra18 (2013).
25. Sheng, Q. et al. An activated ErbB3/NRG1 autocrine loop supports in vivo proliferation in ovarian cancer cells. *Cancer Cell* **17**, 298–310 (2010).
26. Han, M.-E. et al. Overexpression of NRG1 promotes progression of gastric cancer by regulating the self-renewal of cancer stem cells. *J. Gastroenterol.* **50**, 645–656 (2015).
27. Yun, S. et al. Clinical significance of overexpression of NRG1 and its receptors, HER3 and HER4, in gastric cancer patients. *Gastric Cancer* **21**, 225–236 (2018).
28. Luraghi, P. et al. A molecularly annotated model of patient-derived colon cancer stem-like cells to assess genetic and nongenetic mechanisms of resistance to anti-EGFR therapy. *Clin. Cancer Res.* **24**, 807–820 (2018).
29. Yonesaka, K. et al. Activation of ERBB2 signaling causes resistance to the EGFR-directed therapeutic antibody cetuximab. *Sci. Transl. Med.* **3**, 99ra86 (2011).
30. Yang, L. et al. NRG1-dependent activation of HER3 induces primary resistance to trastuzumab in HER2-overexpressing breast cancer cells. *Int. J. Oncol.* **51**, 1553–1562 (2017).
31. Wilson, T. R., Lee, D. Y., Berry, L., Shames, D. S. & Settleman, J. Neuregulin-1-mediated autocrine signaling underlies sensitivity to HER2 kinase inhibitors in a subset of human cancers. *Cancer Cell* **20**, 158–172 (2011).
32. Mendell, J. et al. Clinical translation and validation of a predictive biomarker for patritumab, an anti-human epidermal growth factor receptor 3 (HER3) monoclonal antibody, in patients with advanced non-small cell lung cancer. *EBioMedicine* **2**, 264–271 (2015).
33. Conway, J. R., Lex, A., Gehlenborg, N. & Hancock, J. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940 (2017).
34. Teer, J. K. et al. Evaluating somatic tumor mutation detection without matched normal samples. *Hum. Genomics* **11**, 22 (2017).
35. Lawrence, M. S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
36. AACR Project GENIE Consortium. AACR Project GENIE: powering precision medicine through an international consortium. *Cancer Discov.* **7**, 818–831 (2017).
37. Hartmaier, R. J. et al. High-throughput genomic profiling of adult solid tumors reveals novel insights into cancer pathogenesis. *Cancer Res.* **77**, 2464–2475 (2017).
38. Maxwell, K. N. et al. *BRCA* locus-specific loss of heterozygosity in germline *BRCA1* and *BRCA2* carriers. *Nat. Commun.* **8**, 319 (2017).
39. Yan, M. et al. HER2 expression status in diverse cancers: review of results from 37,992 patients. *Cancer Metastasis Rev.* **34**, 157–164 (2015).
40. Darvin, P., Toor, S. M., Sasidharan Nair, V. & Elkord, E. Immune checkpoint inhibitors: recent progress and potential biomarkers. *Exp. Mol. Med.* **50**, 165 (2018).
41. Lau, D., Bobe, A. M. & Khan, A. A. RNA sequencing of the tumor microenvironment in precision cancer immunotherapy. *Trends Cancer* **5**, 149–156 (2019).
42. Cristescu, R. et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade–based immunotherapy. *Science* **362**, eaar3593 (2018).
43. Allen, J. et al. Barriers to patient enrollment in therapeutic clinical trials for cancer: a landscape report. *J. Oncol. Navig. Surviv.* **9** (2018).
44. Unger, J. M., Vaidya, R., Hershman, D. L., Minasian, L. M. & Fleury, M. E. Systematic review and meta-analysis of the magnitude of structural, clinical, and physician and patient barriers to cancer clinical trial participation. *J. Natl Cancer Inst.* **111**, 245–255 (2019).
45. Institute of Medicine. *Transforming Clinical Research in the United States: Challenges and Opportunities: Workshop Summary* (National Academies Press, 2010); https://doi.org/10.17226/12900

Acknowledgements

We are thankful to the operations, product, engineering and clinical data teams at Tempus Labs, including but not limited to U. Pipic, C. Schwalbach, S. Hynes, K. Stenglein, L. Sachse, A. Hoyer, S. Carsanaro, H, Lefkofsky, R. Chang, M. Barber, R. Pe Benito, R. Star, H. Whipple and D. King. We thank the pathology and lab teams for sample processing and data collection. We are grateful to M. Salazar for managing the work required for this manuscript. We thank G. Palmer and A. Schwarzbach for review of the manuscript, M. Kase and A. Hoffman-Peterson for proofreading, and A. Sheals and B. Santacaterina for help with figure aesthetics and assembly. We thank E. Lefkofsky for his support and discussions.

Author contributions

N.B., M.B., R.H., C.I., R.T. and D.L. led data analysis and interpretation, and wrote sections of the manuscript. N.B. and T.T. performed the pathologic review of the cohort and wrote sections of the manuscript. C.I., J.M., B.D.L., K.P.S., T.T. and N.B. contributed to gene expression and cancer type predictor analyses and figures. D.L., A.L.C., J.F.P., A.L. and A.A.K. contributed to immune analyses and figures. R.T., S.B., J.P. and W.Z. contributed to mutational and genomic rearrangement analyses and figures. R.H., R.T., D.C.H., N.B., A.S. and M.B. contributed to tumor-only and tumor–normal analyses and figures. R.H., N.B., E.K. and M.B. contributed to therapeutic evidence and clinical trial matching analyses. A.M.B. provided critical review of drafts and figures, wrote sections of the manuscript and reviewed the final manuscript. K.P.W. oversaw manuscript preparation, provided scientific direction, wrote sections of the manuscript and reviewed the final manuscript.

Competing interests

Authors are employees of Tempus Labs, Inc.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41587-019-0259-z.

Correspondence and requests for materials should be addressed to K.P.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Mutational spectrum analyses. After random selection from the de-identified Tempus database, patients were grouped by pre-specified cancer type and filtered for variants classified as clinically relevant. Data from patient xT clinical reports given to oncologists were exclusively used for analyses, with some patients having multiple issued reports. The gene set was then filtered for genes having more than five variants across the entire cohort to select for recurrently mutated genes. The collated set of patients was clustered by mutational similarity across SNVs, indels, fusions, amplifications and homozygous deletions. Subsequently, alteration prevalence for SNVs and indels from the MSKCC IMPACT cohort was extracted from the MSKCC cBioportal (http://www.cbioportal.org/study?id=msk_impact_2017#summary) to compare Tempus xT assay variant calls against publicly available variant data for solid tumors. After selecting only for genes on both panels, variants with a minimum of 2.5% prevalence within their respective cohort were plotted (Fig. 1).

Detection of gene rearrangements from DNA-seq. Gene rearrangements were detected and analyzed by a separate parallel process optimized for the detection of structural alterations. After demultiplexing, tumor FASTQ files were aligned against the human reference genome using BWA (0.7.1)[46]. Reads were sorted, and duplicates were marked with SAMBlaster (0.1.2.4)[47]. Discordant and split reads were further identified and separated by this process. These data were then read into LUMPY (0.2.1.3) for structural variant detection[48]. A Variant Call Format (VCF) file was generated and then parsed by the Tempus fusion VCF parser. The data were pushed to the Tempus bioinformatics database, where structural alterations were grouped by type, recurrence and presence, and displayed through the Tempus quality-control application. Known and novel fusions highlighted by the application were selected by the variant science team for loading into a patient report.

Gene expression data collection and normalization. RNA-seq gene expression data were generated from formalin-fixed, paraffin-embedded tumor samples using an exome-capture-based RNA-seq protocol. After sequencing quality control, the final gene expression sample size was 474 samples. In brief, RNA-seq data were aligned to GRCh38 using STAR (2.4.0.1)[49] and expression quantification per gene was computed with FeatureCounts (1.4.6)[50]. Raw read counts were then normalized to correct for G+C content and gene length using full-quantile normalization and adjusted for sequencing depth via the size factor method. Normalized gene expression data for cancer types were $\log_{10}$ transformed and used for all subsequent analyses.

Detection of gene rearrangements from RNA-seq. Gene rearrangements in RNA were analyzed by a workflow that quantifies gene-level expression and chimeric transcripts through non-canonical exon-exon junctions mapped using split or discordant read pairs. Subsequent to expression quantification, reads were mapped across exon–exon boundaries to unannotated splice junctions, and evidence was computed for potential chimeric gene products. If sufficient evidence was present for the chimeric transcript, a rearrangement was called as detected.

Tempus gene expression reference database. Gene expression data generated at Tempus Labs were combined with publicly available cancer and normal gene expression datasets to create the Tempus reference database. Expression-calling analyses included the cancer expression data from The Cancer Genome Atlas (TCGA) and the normal expression data from the Genotype–Tissue Expression (GTEx) project[51,52]. Raw data from these publicly available datasets were downloaded from the Genomic Data Commons (GDC) or Sequence Read Archive (SRA) and processed with the Tempus RNA-seq pipeline. In total, 4,703 Tempus, 4,865 TCGA and 6,541 GTEx samples were processed and included as part of the larger Tempus reference database for this analysis. After processing, these datasets were corrected to account for batch effect differences between sequencing protocols across institutions (for example, formalin-fixed, paraffin-embedded versus fresh tissue, poly(A) versus exome capture). To account for these differences, we calculated per-gene sizing factors on $\log_{10}$ normalized counts by subsampling TCGA and Tempus samples 100 times. A linear transformation from the sizing factors calculated on TCGA samples was applied to TCGA and GTEx samples to ensure that genes had equivalent means and variances across studies.

Gene expression calling. For each patient in each cancer type (brain, $n=49$; breast, $n=50$; colorectal, $n=50$; lung, $n=48$; ovarian, $n=49$; endometrial, $n=48$; pancreas, $n=50$; prostate, $n=46$; rare, $n=46$; TUO, $n=38$), we compared the expression of key cancer genes to the Tempus reference database to determine overexpression or underexpression. A maximum of 43 genes were evaluated based on the specific cancer type of the sample. Genes associated with immunotherapy are reported as relative expression calls in the immunotherapy portion of the Tempus platform and were excluded from this analysis. To make an expression call, each patient's expression percentile was calculated relative to four distributions: all cancer samples from TCGA, all normal samples from GTEx, specific cancer-matched samples from TCGA, and specific tissue-matched normal samples from GTEx. For example, the tumor expression for each patient with breast cancer was compared to that of all cancer samples, all normal samples, all breast cancer samples and all normal breast tissue samples. Distribution thresholds specific to each gene and cancer type were optimized using literature curation and statistical analysis to reflect over- or underexpressing cancer subtypes. Thresholds at the time of xT reporting were applied to determine gene expression calls and varied slightly across the dataset as thresholds and genes reported have evolved over time.

Cancer type prediction. A random forest cancer type prediction model was trained on normalized gene expression data from 4,703 Tempus samples spanning 33 cancer types, as defined in TCGA. The 500 samples in the xT cohort were excluded from the training dataset. The model was generated using scikit-learn RandomForestClassifer (0.20.0). Hyperparameter tuning on the training data using three-fold cross-validation on 1,000 trees identified a minimum split size of two and a maximum depth of 50 as the best performing parameters with a cross-validation classification accuracy of 81%.

Tumor mutational burden. TMB was calculated by dividing the number of non-synonymous mutations by the megabase size of the panel (2.4 Mb). All non-silent somatic coding mutations, including missense, indel and stop-loss variants with coverage greater than 100× and an allelic fraction greater than 5%, were counted as non-synonymous mutations. Hypermutated tumors were considered TMB-high if they had a TMB >9 mutations per megabase. This threshold was established by testing for the enrichment of tumors with orthogonally defined hypermutation (MSI-H) in the larger Tempus clinical database. A hypergeometric test was performed in increments of 0.5 mutations per megabase from 5 to 15 mutations per megabase. Greater than 9 mutations per megabase was found to be significantly enriched ($P=4.23\times10^{-31}$) for orthogonally defined hypermutated tumors.

Whole-exome tumor mutational burden comparison. TMB for gene panels ranging from 100 to 5,000 genes were simulated from TCGA variant data using the 8,507 samples available on UCSC Xena (http://xena.ucsc.edu/)[53]. For each gene panel size tested, genes were randomly selected for inclusion in the simulated panel, TMB was calculated as described above, and the Pearson correlation between the simulated panel TMB and the whole-exome TMB was determined. Five simulations were performed for each panel size. The correlation between panel TMB and whole-exome TMB was experimentally validated using samples sequenced with both the xT panel and the Tempus whole-exome panel. Whole-exome TMB was calculated as described above, except with a coverage threshold of 35× and an allelic fraction threshold of 10%.

Human leukocyte antigen class I typing. Human leukocyte antigen (HLA) class I typing for each patient was performed using Optitype (1.3.1) on DNA-seq data[54]. Normal samples were used as the default reference for matched tumor–normal samples. Tumor-only-determined HLA type was used when the normal sample did not meet internal HLA coverage thresholds.

Neoantigen prediction. Neoantigen prediction was performed on all non-silent mutations identified by the xT pipeline, including indels, SNVs and frameshifts. For each mutation, the binding affinities for all possible 8- to 11-amino-acid peptides containing that mutation were predicted using MHCflurry (0.9.1)[41]. For alleles with insufficient training data to generate an allele-specific MHCflurry model, binding affinities were predicted for the nearest-neighbor HLA allele as assessed by amino acid homology. A mutation was determined to be antigenic if any resulting peptide was predicted to bind to any of the patient's HLA alleles using a 500 nM affinity threshold. RNA support was calculated for each variant using varlens (0.0.4, https://github.com/openvax/varlens). Predicted neoantigens were determined to have RNA support if at least one read supporting the variant allele could be detected in the RNA-seq data.

Microsatellite instability status. The Tempus xT panel included probes for 43 microsatellites that are frequently unstable in tumors with MMR deficiencies. The MSI classification algorithm used reads mapping to these frequently unstable regions to classify tumors into three categories: microsatellite instability high (MSI-H), microsatellite stable (MSS) or microsatellite equivocal (MSE). This assay can be performed with paired tumor–normal samples or tumor-only samples. Both algorithms return the probability of the patient being MSI-H, which is then translated into an MSI status of MSI-H, MSS or MSE. All loci with sufficient coverage were tested for instability, as measured by changes in the distribution of the number of repeat units in the tumor reads as compared to the normal reads using the Kolmogorov-Smirnov test. If $P \leq 0.05$, the locus was considered unstable. The proportion of unstable loci was fed into a logistic regression classifier trained on tumor samples with clinically determined MSI statuses.

Cytolytic index. CYT was calculated as the geometric mean of the normalized RNA counts of granzyme A (*GZMA*) and perforin-1 (*PRF1*)[17].

Immune infiltration estimation. The relative proportion of immune subtypes was estimated using a support vector regression (SVR) model, which includes an L2 regularizer and an epsilon insensitive loss function, similar to that of

Newman et al[55]. The SVR was implemented in Python using the nuSVR function in the SVM library of scikit-learn (0.18), with the LM22 reference matrix downloaded from the supplement of Newman et al[55].

IFN-γ gene signature score. Twenty-eight IFN-γ pathway-related genes were used as the basis for an IFN-γ gene signature score[19]. Hierarchical clustering was performed on the basis of Euclidean distance using the R package ComplexHeatmap (1.17.1), and the heat map was annotated with PD-L1-positive IHC staining, TMB-high status and/or MSI-H status. IFN-γ score was calculated using the arithmetic mean of the 28 genes.

Somatic signatures. Thirty previously described somatic signatures of mutational processes[56,57] were estimated using non-negative least-squares regression as implemented in the deconstructSigs package (1.8.0)[58]. Mutations in this analysis included all discovered somatic SNVs, independent of their pathogenicity. Somatic signature estimates were calculated for all TMB-high samples with at least 50 detected somatic mutations. For visualization, the contributions of signatures 2 and 13 were summed for the APOBEC signature, signatures 4 and 29 were summed for the tobacco signature, and signatures 6, 15, 20 and 26 were summed for the DNA MMR deficiency signature. Additional signatures visualized included signature 1 for age, signature 3 for homologous recombination deficiency, signature 7 for UV and signature 11 for alkylating agent. All other signatures were not plotted, given their unknown etiology and/or limited contribution to the mutational spectra of the patients analyzed.

Knowledge database and evidence-based therapy matching. To determine therapeutic actionability for sequenced patients, Tempus maintains an internal knowledge database (KDB) with structured data regarding drug-gene interactions and precision medicine findings reported in the oncology, pathology and basic science literature. The KDB of therapeutic and prognostic evidence, which includes therapeutic response and resistance information, is compiled from a combination of external sources, including, but not exclusive to, NCCN, CIViC[59] and DGIdb[60], and is maintained with constant annotation by Tempus experts. Clinical actionability entries in the KDB are structured by both the disease to which the evidence applies and the level or strength of the evidence. Therapeutic actionability entries are binned into tiers of somatic evidence by patient disease matches as established by the American Society of Clinical Oncology, the Association for Molecular Pathology and the College of American Pathologists working group[20].

Evidence-based therapies are grouped by their level of evidence strength into tiers IA, IB, IIC and IID. Briefly, tier IA evidence includes biomarkers that follow consensus guidelines and match disease type. Tier IB evidence includes biomarkers that follow clinical research and match disease indication. Tier IIC evidence biomarkers follow the off-indication use of consensus guidelines or clinical research, or either on- or off-indication patient case studies. Tier IID evidence biomarkers follow preclinical evidence regardless of disease indication matching. Patients from the xT 500 cohort were matched to actionability entries by gene, specific variant, diagnosis and level of evidence.

Alteration classification. Somatic alterations were interpreted on the basis of a collection of internally weighted criteria composed of knowledge from known evolutionary models, functional data, clinical data, hotspot regions within genes, internal and external somatic databases, primary literature and other features of somatic drivers[61,62]. The criteria included features of an internally derived heuristic algorithm that groups alterations into one of four categories: pathogenic, variants of unknown significance (VUS), benign or reportable. Pathogenic variants were defined as driver events or tumor-prognostic signals. Benign variants were defined as alterations with evidence indicating a neutral state in the population and were removed from reporting. VUSs were regarded as passenger events. Reportable variants were considered to be diagnostic, offering therapeutic guidance, or associated with disease but not key driver events. Gene amplifications, deletions and translocations were reported on the basis of the features of known gene fusions, relevant breakpoints, biological relevance and therapeutic actionability. Germline pathogenic and VUS alterations identified in the tumor–normal matched samples were reported as secondary findings for consenting patients. These include a subset of genes recommended by the American College of Medical Genetics[63] and genes associated with cancer predisposition or drug resistance.

Tumor-only analyses. For the tumor-only analyses, germline variants from 50 patient samples within the xT 500 cohort were computationally identified and removed using an internal algorithm that considered copy number, tumor purity and sequencing depth. Further filtering was performed on the observed frequency in a population database (positions with variant allele frequency (VAF) ≥1% in the ExAC non-TCGA cohort)[64]. The algorithm was designed to be conservative when calling germline variants in therapeutic genes to minimize removal of true somatic pathogenic alterations that occur within the general population. To remove potential artifacts and biases within our cohort, alterations observed in an internal pool of 50 unmatched normal samples were removed. The remaining variants were assumed to be somatic variants with VAF ≥5% and coverage ≥90%.

The alteration classification rules were applied, and evidence-based therapies were assigned to each patient. Using matched normal sequencing data, we were able to identify true germline variants and evaluate contamination. The 50 patient cases were reviewed by two Tempus pathologists and an oncologist to determine which patients would have significantly different therapeutic matches on the basis of the tumor-only analysis instead of the full Tempus test, including tumor–normal matched DNA-seq, RNA-seq and IO analysis. For this direct comparison, data from clinical reports were evaluated. Relevant variants and therapies from the full Tempus test reflect present-day therapy matches. As an additional comparison using the Tempus tumor-only DNA variant results, we manually searched the public resource http://www.mycancergenome.org/ (7 November 2018) for returned therapies based on these DNA variants.

Clinical trial reporting. Clinical trial options were identified by associating a patient's actionable variants and structured clinical data with an internally curated database of clinical trials largely procured from http://ClinicalTrials.gov. Criteria considered for clinical trial reporting based on the patient information available at the time included, but was not limited to, molecular alterations, diagnosis, age, previous treatment, medical history, stage of cancer and distance from point-of-care. Biomarker-based clinical trials were defined as those that required specific molecular alterations to qualify, whereas disease-based clinical trials did not have such a requirement. All reported clinical trials were checked for recruitment status at the time of xT report generation.

Statistical analysis. All statistical analyses were conducted in R (3.4.4). Statistical significance was determined by a two-sided Wilcoxon rank-sum test or a Kruskal–Wallis test, as indicated in the figure legends, with $P < 0.05$ considered significant. P values were adjusted for multiple testing using the Benjamini-Hochberg method. Relationships between variables were assessed by Pearson correlation. In the data presented as boxplots, the upper and lower hinges represent the first and third quartile. The whiskers extend to the most extreme value within 1.5 times the interquartile range on either end of the distribution. The center line represents the median. The exact sample sizes (n) used to calculate each statistic are listed in the figure legends.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

VCF files, RNA count files and associated deidentified clinical data that support these findings will be available through Vivli (ID T19.01).

References

46. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
47. Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503–2505 (2014).
48. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014).
49. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
50. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
51. Lonsdale, J. et al. The Genotype–Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
52. Peng, L. et al. Large-scale RNA-Seq transcriptome analysis of 4043 cancers and 548 normal tissue controls across 12 TCGA cancer types. *Sci. Rep.* **5**, 13413 (2015).
53. Goldman, M. et al. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. Preprint at https://doi.org/10.1101/326470 (2019).
54. Szolek, A. et al. OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics* **30**, 3310–3316 (2014).
55. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
56. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
57. Forbes, S. A. et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45**, D777–D783 (2017).
58. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17**, 31 (2016).

59. Griffith, M. et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat. Genet.* **49**, 170–174 (2017).
60. Finan, C. et al. The druggable genome and support for target identification and validation in drug development. *Sci. Transl. Med.* **9**, eaag1166 (2017).
61. Madhavan, S. et al. ClinGen Cancer Somatic Working Group—standardizing and democratizing access to cancer molecular diagnostic data to drive translational research. *Pac. Symp. Biocomput.* **23**, 247–258 (2018).
62. Dienstmann, R. et al. Standardized decision support in next generation sequencing reports of somatic cancer variants. *Mol. Oncol.* **8**, 859–873 (2014).
63. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics And Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
64. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).

nature research

Corresponding author(s): Kevin White, Ph.D.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☒	☐	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Patient DNA and RNA sequencing data was generated via formalin fixed paraffin embedded tissue, saliva, or blood samples using the Illumina HiSeq 4000. Data was stored and downloaded for analyses from the Tempus de-identified database. Raw data from publicly available datasets, TCGA and GTEx, were downloaded via the GDC or SRA.
Data analysis	For genomic rearrangement DNA analysis, reads were align to the human reference genome using BWA (v0.7.1) and sorted and marked with SAMBlaster (v0.1.2.4). LUMPY (v0.2.1.3) was used for structural variant detection. For genomic rearrangement RNA analysis, reads were aligned to GRCh38 using STAR (v2.4.0.1) and expression quantitation per gene was computed using FeatureCounts (v1.4.6). For cancer type prediction, TCGA data were processed via the Tempus RNA-seq pipeline and analyzed along with Tempus xT 500 cohort data using a Scikit-learn (v0.20.0) random forest model. Optitype (v1.3.1) was used for HLA class 1 typing. Neoantigen prediction utilized MHCflurry (v0.9.1) to predict peptide binding affinities to HLA. Hierarchical clustering for interferon gamma signature analysis was performed based on Euclidean distance using the R package ComplexHeatmap (v1.17.1). Somatic signatures of mutational processes were estimated using non-negative least-squares regression as implemented in the DeconstructSigs package (v1.8.0). In estimating the relative proportion of immune subtypes, the SVR was implemented in python using the nuSVR function in the SVM library of scikit-learn (v0.18). Neoantigen RNA support data was calculated for each variant using Varlens (v0.0.4). The tumor-only pipeline was run using proprietary code. Analyses were performed using R (v3.4.4) with plots generated using ggplot2 (v3.0). Additional information about statistical tests performed can be found in the methods section.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

VCF files, RNA count files, and associated de-identified clinical data that support these findings will be available through Vivli (ID T19.01).

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We randomly selected a cohort of 500 cancer patients who had clinical tumor and matched germline specimens sequenced with the xT panel at CAP/CLIA-certified Tempus Labs, Inc. In order to be eligible for inclusion in the cohort, each case was required to have complete data elements for tumor-normal matched DNA sequencing, and clinical data. Subsequent to filtering for eligibility, a set of patients were randomly sampled via a pseudo-random number generator to bin 50 patients representing ten cancer categories based on pathologic diagnosis, with 50 patients per brain, breast, colorectal, lung, ovarian, endometrial, pancreatic, prostate, rare, and unknown cancer type. Only solid tumor types were included in these analyses. Hematological malignancies were excluded.
Data exclusions	
Replication	There was no experimental replication completed within this study.
Randomization	
Blinding	N/A

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging