Diversity and divergence of the glioma-infiltrating T-cell receptor repertoire

Jennifer S. Sims[a], Boris Grinshpun[b], Yaping Feng[b,c], Timothy H. Ung[a,d], Justin A. Neira[a], Jorge L. Samanamud[a], Peter Canoll[e], Yufeng Shen[b,f,g,1], Peter A. Sims[b,g,h,1], and Jeffrey N. Bruce[a,1]

[a]Department of Neurological Surgery, Columbia University Medical Center, New York, NY 10032; [b]Department of Systems Biology, Columbia University Medical Center, New York, NY 10032; [c]Waksman Institute of Microbiology Genomics Core Facility, Rutgers, The State University of New Jersey, Piscataway, NJ 08854; [d]Department of Neurosurgery, University of Colorado School of Medicine, Aurora, CO 80045; [e]Department of Pathology and Cell Biology, Columbia University Medical Center, New York, NY 10032; [f]Department of Biomedical Informatics, Columbia University Medical Center, New York, NY 10032; [g]JP Sulzberger Columbia Genome Center, Columbia University Medical Center, New York, NY 10032; and [h]Department of Biochemistry & Molecular Biophysics, Columbia University Medical Center, New York, NY 10032

Edited by Lawrence Steinman, Stanford University School of Medicine, Stanford, CA, and approved April 26, 2016 (received for review January 20, 2016)

Although immune signaling has emerged as a defining feature of the glioma microenvironment, how the underlying structure of the glioma-infiltrating T-cell population differs from that of the blood from which it originates has been difficult to measure directly in patients. High-throughput sequencing of T-cell receptor (TCR) repertoires (TCRseq) provides a population-wide statistical description of how T cells respond to disease. We have defined immunophenotypes of whole repertoires based on TCRseq of the α- and β-chains from glioma tissue, nonneoplastic brain tissue, and peripheral blood from patients. Using information theory, we partitioned the diversity of these TCR repertoires into that from the distribution of VJ cassette combinations and diversity due to VJ-independent factors, such as selection due to antigen binding. Tumor-infiltrating lymphocytes (TILs) possessed higher VJ-independent diversity than nonneoplastic tissue, stratifying patients according to tumor grade. We found that the VJ-independent components of tumor-associated repertoires diverge more from their corresponding peripheral repertoires than T-cell populations in nonneoplastic brain tissue, particularly for low-grade gliomas. Finally, we identified a "signature" set of TCRs whose use in peripheral blood is associated with patients exhibiting low TIL divergence and is depleted in patients with highly divergent TIL repertoires. This signature is detectable in peripheral blood, and therefore accessible noninvasively. We anticipate that these immunophenotypes will be foundational to monitoring and predicting response to antiglioma vaccines and immunotherapy.

T-cell receptor | immunoprofiling | glioma | glioblastoma | immunooncology

The potential for immunotherapy to alleviate progression and recurrence in glioma has inspired intense study of the immunological phenotypes underlying the regulation and selection of tissue-infiltrating lymphocytes (TILs). Motivated by the prospect of targeted therapy, previous studies of glioma have undertaken large-scale genomic and gene expression analysis. These studies revealed distinct phenotypic states or subtypes that stratify gliomas and resemble different glial lineages (1–3). Although immunological gene expression classifications have been associated with clinical outcomes and prognosis (4, 5), precision immunotherapy will ultimately rely on manipulation of the T-cell population that infiltrates gliomas and its underlying repertoire of T-cell receptors (TCRs). However, the structure and intertumoral heterogeneity of the TIL population in gliomas has not been described. Here, we use whole repertoire sequencing of TCRs from glioma tissue and matched peripheral blood to discover novel immunological phenotypes with implications for noninvasive patient monitoring.

Local antitumor potential is, in part, a function of the TCR repertoire expressed by T cells that surveil the CNS beyond the blood–brain barrier. Through somatic V(D)J recombination including random nucleotide insertion in each T-cell, the α- and β-chain loci each encode a complementarity-determining 3 (CDR3) domain that interacts directly with target epitopes in the heterodimeric receptor. The ability to mount an adaptive immune response relies on the diversity of binding specificities conferred by these receptors, which determine functional activation, clonal expansion, and selection for individual T cells. Methods for repertoire-wide amplification and deep sequencing now allow massively parallel repertoire profiling (6) of the TCRs in whole populations of T cells (TCRseq) (7–9). Besides providing unprecedented insight into the determinants of TCR diversity in healthy individuals, previous applications of this strategy have provided new understanding of xenoreactivity in transplants, infection in human patients and animal models, and therapeutic response, residual disease, and relapse in cancer (for a review, see ref. 10).

We applied TCRseq to peripheral blood and tumor tissue samples from both low- and high-grade glioma patients. By combining V and J cassette use and amino acid sequence information, we describe the diversity and divergence of these populations. We found previously undescribed phenotypes characterized by the diversity of clonotypes in the TIL repertoire and its divergence from the peripheral repertoire. Importantly, we also discovered a signature set of TCRs in the peripheral repertoire that reflect its divergence from the TIL repertoire. These immunological phenotypes defined by the TCR repertoire offer new insights into intertumoral heterogeneity among low- and high-grade glioma patients and may provide useful insights for noninvasive monitoring of glioma progression and response to immunotherapy.

Significance

High-throughput sequencing of T-cell receptor (TCR) repertoires provides a high-dimensional biomarker for monitoring the immune system. We applied this approach, measuring the extent to which the TCR repertoires of T-cell populations infiltrating malignant brain tumors diverge from their peripheral blood. Our analytical strategy separates the statistical properties of the repertoire derived from VJ cassette combination usage from the VJ-independent contribution that reflects the antigen-binding component of the receptor. We discovered a TCR signature strongly inversely correlated with the VJ-independent divergence between the peripheral and tissue-infiltrating repertoires of these patients. Importantly, this signature is detectable in peripheral blood and could serve as a means of noninvasively monitoring immune response in patients.

Author contributions: J.S.S., P.C., Y.S., P.A.S., and J.N.B. designed research; J.S.S., B.G., Y.F., T.H.U., J.A.N., J.L.S., P.C., Y.S., and P.A.S. performed research; J.S.S., B.G., Y.F., P.C., Y.S., and P.A.S. analyzed data; and J.S.S., B.G., P.C., Y.S., P.A.S., and J.N.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequencing data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE79338).

[1]To whom correspondence may be addressed. Email: ys2411@columbia.edu, pas2182@columbia.edu, or jnb2@columbia.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1601012113/-/DCSupplemental.

Results

TCRseq of Tumor-Infiltrating and Peripheral Repertoires in Glioma Patients. We prepared α- and β-chain TCRseq libraries from biopsies of glioma and nonneoplastic brain tissue along with peripheral blood mononuclear cells (PBMCs) procured during surgery to compare TILs with the peripheral population from which they were derived (Fig. 1*A*). Subjects included three low-grade glioma (LGG), eight primary glioblastoma (GBM), and three nonglioma patients from which we obtained nonneoplastic brain tissue (Table S1). Briefly, from the total RNA of each 2.5 million PBMC sample or the mRNA of each fresh-frozen tissue biopsy, we used reverse transcription amplicon-rescue multiplexed PCR (arm-PCR) to produce dsDNA sequencing libraries for TCRα and TCRβ (Fig. 1*A*) (7, 11, 12), which were subjected to 2 × 220- or 2 × 250-base paired-end sequencing. After mapping the merged reads to V and J cassette genes in the reference genome and in silico translation, we described each TCRα and TCRβ repertoire of the PBMC and TIL of each patient as (*i*) the repertoire of VJ cassette combinations, (*ii*) the repertoire of antigen-binding amino acid motifs encoded by the CDR3 region (CDR3 amino acid sequences), and

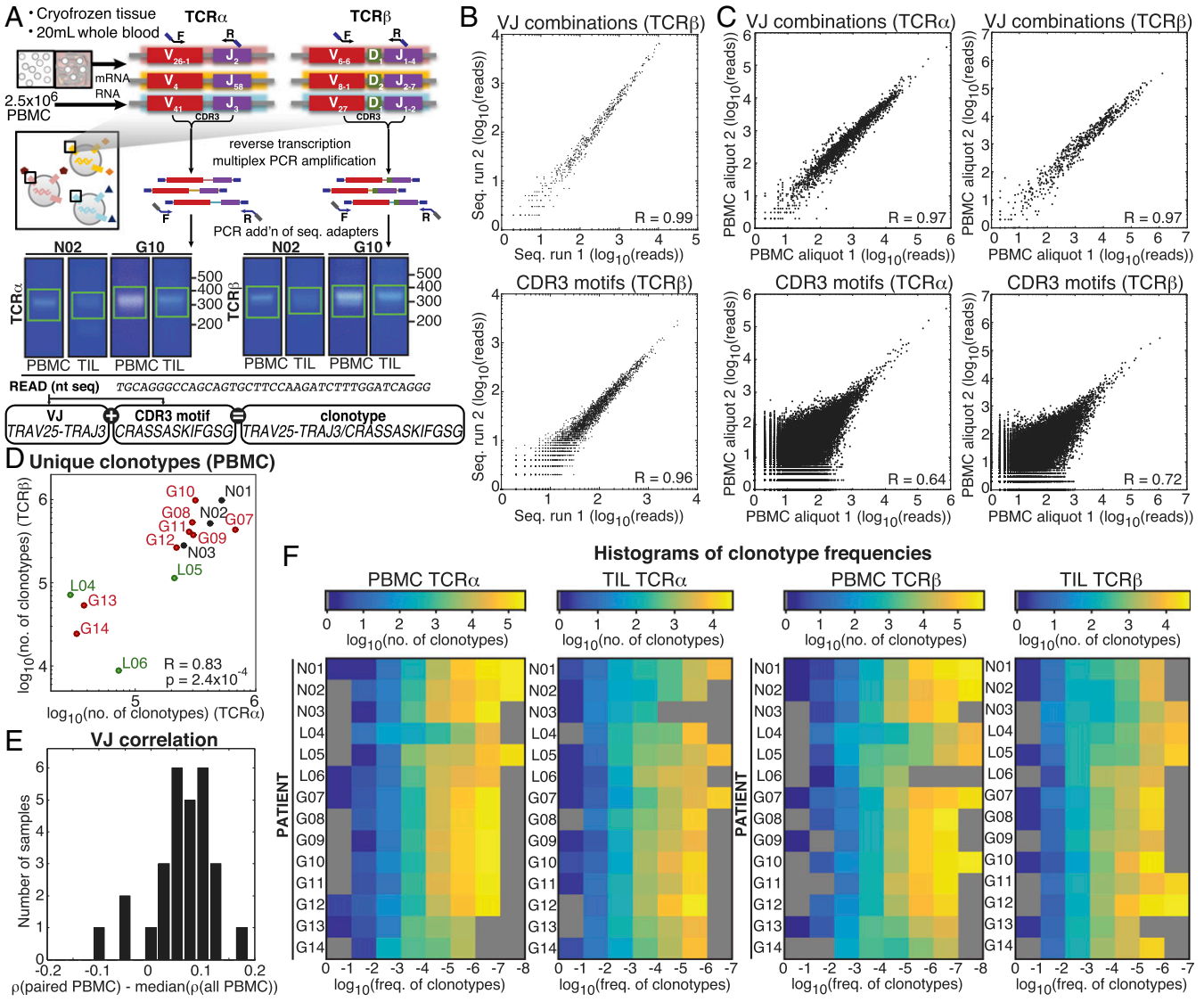

Fig. 1. TCRα and TCRβ repertoires from brain tissue and matched peripheral blood. (*A*) Gene-specific reverse transcription and PCR of the CDR3 region using pan-repertoire primers (V-J/C) was performed on mRNA of tissue biopsies and total RNA of ∼1 × 10^6 T cells, followed by bead purification and second amplification incorporating sequencing adapters. Libraries were purified by gel electrophoresis (example extracted regions in green) and sequenced (PE250). Following error analysis and read filtration (*SI Materials and Methods*), merged reads were translated in silico, and productive reads were tabulated by V and J cassette identity, amino acid CDR3 motif, and the combination of VJ cassette combinations and amino acid CDR3 used to identify unique clonotypes. (*B*) Correlation between the abundance detected in two sequencing runs of a single TCRseq library (L06 TIL TCRβ) of VJ combinations and CDR3 motifs. (*C*) Correlation between the abundance in two aliquots of PBMC from the same draw (H15) of VJ cassette combinations and CDR3 motifs. Pearson correlation coefficients between VJ cassette combination abundance ($R = 0.97$ for TCRα and 0.97 for TCRβ) and CDR3 motif abundance ($R = 0.64$ for TCRα and 0.72 for TCRβ). (*D*) The number of clonotypes (the unique combination of a V and J cassette pair with a CDR3 amino acid motif) observed in the PBMC of patients. Pearson correlation between $\log_{10}$(clonotypes) of TCRα and TCRβ is shown ($R = 0.83$, $P = 2.4 \times 10^{-4}$). Label colors indicate clinical status (GBM, red; LGG, green; and nonneoplastic, black). (*E*) The Spearman correlation (ρ) of VJ combination frequencies between each TIL sample and its paired PBMC was calculated, and the median (ρ) of its correlation with all PBMCs was subtracted. Resulting values for each TIL library are displayed as a histogram. (*F*) Heat maps of the number of clonotypes (colorbars) occurring at frequencies (*x* axis) for TCRα (*Left*) and TCRβ (*Right*) chains of the PBMC and TIL of each patient (*y* axes). Unpopulated frequency bins are displayed in gray.

(*iii*) the repertoire of clonotypes—which includes all observed unique combinations of CDR3 amino acid sequences and VJ cassette combinations. To avoid overcounting as a result of, for example, sequencing error, reads were filtered on the basis of their nucleotide similarity to other reads (*SI Materials and Methods*).

The quantitative reproducibility of the TCR repertoires detected by this platform was assessed first between sequencing runs of the same prepared library (Fig. 1*B*), yielding strong correlations for both VJ cassette combinations and CDR3 amino acid sequences, as expected, and in different aliquots of PBMC from the same blood draw, which were similarly well-correlated at the level of VJ cassette combinations. Because different PBMC aliquots will sample different TCRs, particularly for low-abundance clones, correlations between aliquots at the CDR3 amino acid sequence level were strong (Fig. 1*C*) but significantly lower than those of the technical replicates in Fig. 1*B*.

The total number of clonotypes observed was strongly correlated between TCRα and TCRβ across patients in PBMC libraries (Fig. 1*D*, Pearson $R = 0.83$, $P = 2.4 \times 10^{-4}$). Although most PBMC repertoires possessed at least one clonotype at a very high frequency (>5%, Dataset S1), our results agreed with previous observations in healthy individuals (13, 14) that the vast majority of clones appear at low frequencies (Fig. 1*F*)—only 19 or fewer clonotypes appeared at a frequency >1% in the PBMC of any

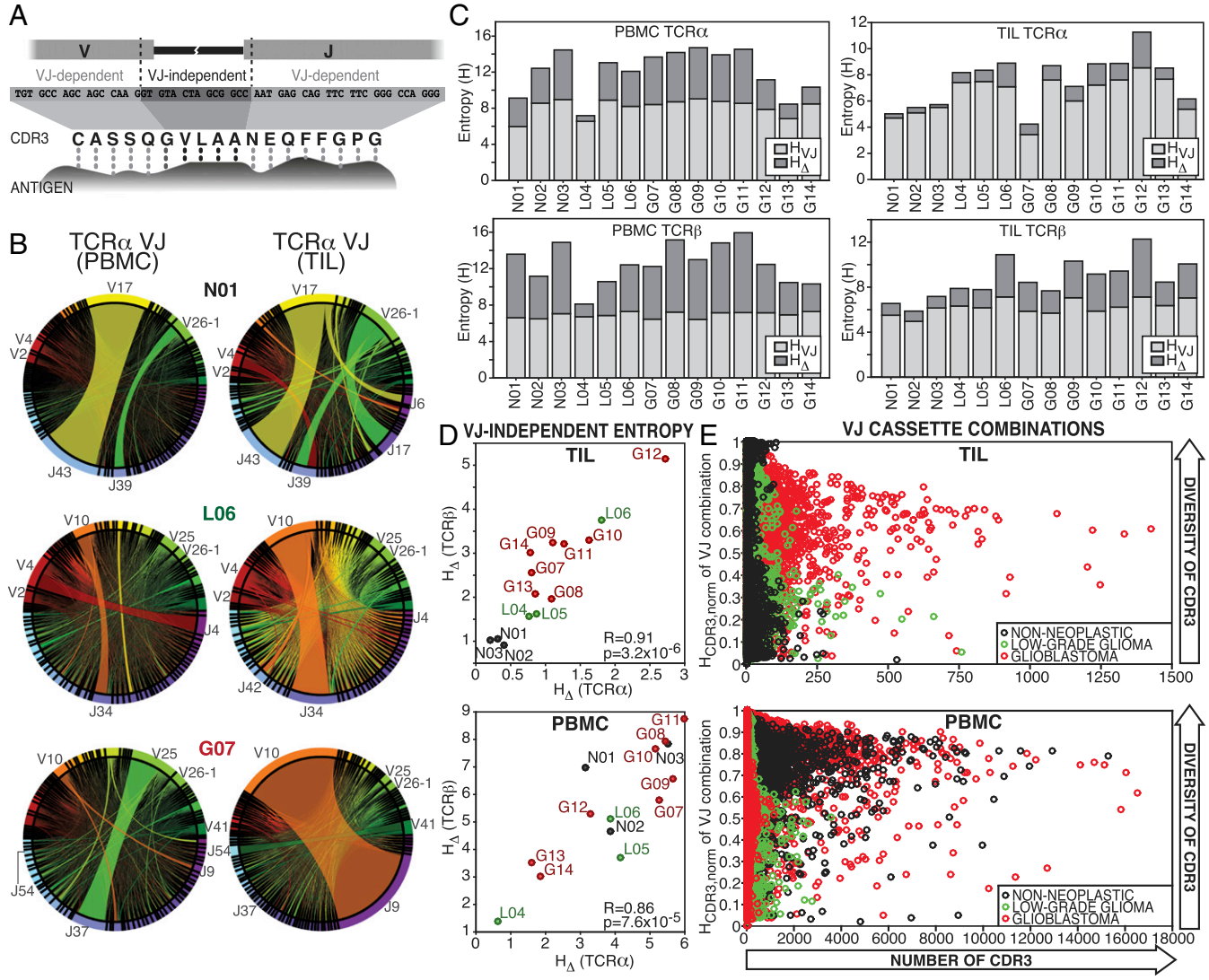

Fig. 2. Entropy-based dissection of TCR diversity. (*A*) The components of the Shannon entropy of a single CDR3 sequence are displayed, reflecting the VJ-dependent (light gray) and VJ-independent (dark gray) origins of the information given by each residue, including the addition of random nucleotides upon the joining of cassettes and the replacement of some cassette-encoded nucleotides with random bases (dashed lines). Although all amino acids in the CDR3 motif participate in antigen binding (dotted connectors), some are determined entirely by V and J cassette sequence, whereas others are entirely independent. (*B*) The frequencies of VJ combinations in select PBMC (*Left*) and TIL (*Right*) TCRα repertoires are displayed in circular plots, with frequency of each V or J cassette represented by its arc length and that of the VJ cassette combination by the width of the joining ribbon. (*C*) The entropy of the clonotype repertoire ($H_{clonotype}$) for the PBMC (*Left*) and TIL (*Right*) of each patient (*x* axis) is displayed as the sum of the entropy of the VJ repertoire (H_{VJ}, light gray) and the VJ-independent entropy (H_{Δ}, dark gray). (*D*) For each patient, the VJ-independent entropy (H_{Δ}) of the TIL (*Top*) and PBMC (*Bottom*) repertoires were calculated, as well as the Pearson correlations between TCRα (*x* axis) and TCRβ (*y* axis) across patients (PBMC $R = 0.86$, $P = 7.6 \times 10^{-5}$; TIL $R = 0.92$, $P = 3.2 \times 10^{-6}$). Label colors indicate clinical status (GBM, red; LGG, green; and nonneoplastic, black). (*E*) To illustrate relationship between the VJ-independent entropy of the clonotype repertoire and the CDR3 amino acid diversity of each VJ cassette combination, these combinations were plotted for each patient according to the number of CDR3 motifs encoded (*x* axis) and the normalized Shannon entropy of those CDR3 (*y* axis), with colors of each VJ cassette combination indicating clinical status of the patient (GBM, red; LGG, green; and nonneoplastic, black; TCRα and TCRβ are displayed together).

patient—but the vast majority of reads comprised high-frequency clonotypes (Fig. S1 *C* and *D*). In the TIL repertoires, representing T cells infiltrating various regions of the brain parenchyma (Fig. S2*A*), the number of clonotypes detected was 2- to ∼100-fold lower than their respective PBMC libraries (Dataset S1 and Fig. S1*B*), with no significant correlation to sequencing coverage (Fig. S1*A*). Consistent with shared patient origins, VJ cassette combination use in TIL repertoires was more strongly correlated with that of their matching PBMC repertoires than with those of the overall cohort (Fig. 1*E*).

Diversity of the Glioma-Infiltrating TCR Repertoire. The vast diversity of the TCR repertoire and its responsiveness to stimuli through amplification of individual T-cell clones serve both as the basis for cell-mediated adaptive immunity and as a readout of those responses in an individual (15, 16). We used information theory to describe the TCR repertoires of the TIL and PBMC at a population-wide level. As a measure of diversity, we calculated the Shannon entropy, *H*—a function of the number of unique elements (e.g., VJ cassette combinations) in the population and their frequencies—for VJ cassette combinations, CDR3 amino acid sequences, and clonotypes (Dataset S2). As described previously, the entropy of each individual TCR can be described as the sum of the contributions from different parts of its sequence (17). Because the entropy of the repertoire is a separable function of those TCRs, we sought to partition the diversity of the population into components that reflect the factors affecting the distribution of these distinct components of the receptor (18), simplified as (*i*) the genome-encoded V and J cassette sequences and (*ii*) the VJ-independent sequence. These components are affected differentially by many factors (for a review see ref. 19). For example, epigenetically biased cassette use (20), age (21), and postdevelopmental enrichment for functional lineages (e.g., CD8+) (22) strongly influence the frequency of VJ cassette combinations in T-cell populations. However, as described in Fig. 2*A*, the binding affinity between CDR3 and an antigen depends on amino acids that are either not encoded by or not specific to the V or J cassette used by the receptor (the VJ-independent component). Thus, we describe the entropy of the clonal population, $H_{clonotypes}$ as the sum of these contributions:

$$H_{clonotypes} = H_{VJ} + H_{\Delta}, \quad [1]$$

where H_{VJ} is the entropy of the distribution of VJ cassette combinations and H_{Δ} is the entropy from the VJ-independent component. H_{VJ} varied among patients and between the PBMC and TIL of individuals (Fig. 2*B* and Fig. S6), with smaller contributions from VJ-independent diversity among TIL than in PBMC (Fig. 2*C* and Dataset S2). H_{Δ} was highly correlated between TCRα and TCRβ across patients for both the PBMC and TIL repertoires (Fig. 2*D*) and saturated at the coverage of our TCRseq libraries (Fig. S3 *A* and *B*). TILs from nonneoplastic tissue were distinguished by a very low proportion of $H_{clonotypes}$ contributed by H_{Δ} (average 5.7% for TCRα and 15.2% for TCRβ) compared with those from glioma tissue, with an average of 15.8% for TCRα and 31.8% for TCRβ (Fig. 2*D*, *Top* and Fig. S3*C*), although this trend was not observed in the PBMC (Fig. 2*D*, *Bottom* and Fig. S3*D*).

Although H_{VJ}, H_{Δ}, and $H_{clonotypes}$ represent the diversity of the whole TCR repertoire, we can use the same formalism to assess the diversity within each VJ cassette combination. A given combination of a V and a J cassette will often encode multiple CDR3s in the repertoire, which gives that cassette combination a certain amount of entropy. For each VJ combination observed in each PBMC and TIL repertoire, we calculated the entropy imparted by the diversity of CDR3 amino acid sequences encoded (Fig. 2*E*). This analysis is analogous to spectratyping the products of certain cassette-specific primer pairs to assess clonal diversity (23). We observed both higher numbers (*x* axis) and broader distributions (normalized H_{CDR3}, *y* axis) of CDR3 amino acid sequences encoded by each VJ combination in glioma tissue than in nonneoplastic tissue (Fig. 2*E*, *Top*), consistent with the comparatively higher H_{Δ} of glioma TIL repertoires (Fig. S3*E*). Neither the H_{Δ} (Fig. 2*D*, *Bottom*) nor the distribution of CDR3 motifs among VJ cassette combinations in the PBMC repertoires stratified the patients by clinical status (Fig. 2*E*, *Bottom*). The repertoire-wide property H_{Δ}, as a measure of the component of CDR3 diversity most closely tied to antigen binding function, facilitates quantitative comparison between TCR repertoires. Here, we observed increasing VJ-independent diversity in the TIL repertoires—but not the matched PBMC—of patients with increasing glioma grade, suggestive of the diversified antigen targeting associated with active T-cell responses in both cancer and infectious diseases (24–26).

VJ-Independent Divergence of TIL and PBMC Repertoires Distinguishes Glioma Patients. We sought to determine how the composition of the TIL repertoire differed from the peripheral blood and whether these differences might distinguish T-cells infiltrating tumor compared with nonneoplastic tissue. Comparison of VJ cassette combination and CDR3 amino acid sequence use between the PBMC and TIL of each patients revealed that more than half of each patient's composite VJ cassette combination repertoire was present in both tissues (average 59.5% for TCRα and 71.9% for TCRβ, Fig. S4*A*), whereas very few CDR3 amino acid sequences were shared (average 1.9% for TCRα and 1.6% for TCRβ, Fig. S4*B*), as expected given the larger potential diversity of the CDR3 amino acid sequences (17) and limited sampling in each experiment.

To resolve the differences between the PBMC and TIL, we calculated Jensen–Shannon divergence (*JS*) and its related metric (*JSM*), derived from the joint entropy of two distributions, which characterize the difference in the frequencies of their members (Dataset S2). *JS* and *JSM* are bounded by zero (when two distributions are identical) and one (where two distributions are completely nonoverlapping) and have been widely applied to the comparison of TCR repertoires (13, 14, 27). Inferring that the distinctiveness of the TIL population might therefore be resolved by separating the VJ-dependent and VJ-independent components, we described the components of the divergence of the TIL clonotypes repertoire from the PBMC, $JS_{clonotypes}$ as follows:

$$JS_{clonotypes}(PBMC,TIL) = JS_{VJ}(PBMC,TIL) + JS_{\Delta}(PBMC,TIL), \quad [2]$$

where JS_{VJ} is the Jensen–Shannon divergence of their VJ combination repertoires and JS_{Δ} represents the VJ-independent divergence between PBMC and TIL (detailed in *SI Materials and Methods*). These divergences were computed for each PBMC–TIL pair for TCRα and TCRβ (Fig. 3*A*, Fig. S4 *C* and *D*, and Dataset S2).

Some of the divergence between PBMC and TIL repertoires can be explained simply by the differences in repertoire size. To correct for this and extract differences due to composition, we calculated $JS_{clonotypes}$, JS_{VJ}, and thus JS_{Δ} between the PBMC repertoire and a down-sampled repertoire (PBMC′) containing the same number of clonotypes as the TIL repertoire, randomly sampled from the PBMC distribution (Fig. 3*B*). The comparison between the VJ-independent divergence of PBMC–TIL and PBMC–PBMC′ could be made using the *JSM*, which is a distance metric defined as

$$JSM = \sqrt{JS}. \quad [3]$$

Thus, we write the VJ-independent Jensen–Shannon divergence metric between the PBMC and TIL, corrected for population size, as

$$JSM_{\Delta,corr}(PBMC,TIL) = JSM_{\Delta}(PBMC,TIL) - JSM_{\Delta}(PBMC,PBMC'), \quad [4]$$

where $JSM_{\Delta}(PBMC,TIL) = \sqrt{JS_{\Delta}(PBMC,TIL)}$ and $JSM_{\Delta}(PBMC,PBMC') = \sqrt{JS_{\Delta}(PBMC,PBMC')}$ for the down-sampled control population PBMC′ (per Eqs. **2** and **3**). Because $JSM_{\Delta,corr}(PBMC,TIL) = 0$ between the PBMC repertoire and a random sample of its TCR distribution, positive values indicate more divergence associated with the VJ-independent component than by random selection.

To evaluate the VJ-independent divergence between peripheral blood and TIL populations, differentially exposed to local and tumor antigens, we calculated $JSM_{\Delta,corr}(PBMC,TIL)$ for both TCRα and TCRβ of all patients (Fig. 3*C* and Dataset S2). VJ-independent divergence was below expectation (negative) for the three nonneoplastic patients, whereas the divergence of all glioma TIL was above expectation (positive), with high Pearson correlation between the α- and β-chains ($R = 0.86$, $P = 8.5 \times 10^{-5}$). The GBM patients displayed heterogeneity, with VJ-independent divergence ranging from slightly to strongly above expectation, and all three LGG patients exhibited high divergence (Fig. 3*D*). $JSM_{\Delta,corr}(PBMC,TIL)$ had no significant correlation with overall white blood cell count in peripheral blood or presurgical steroid therapy (Fig. S2 *B* and *C*), suggesting that this variability did not reflect lymphopenia or steroid-induced immunosuppression. Furthermore, no significant relationship was found between VJ-independent divergence and gene expression markers of vascularization (Fig. S2*D*). Although higher divergence of glioma TIL overall (Fig. S4*E*) supported the inferred link between the binding-associated TCR repertoire and novel tumor antigens, $JSM_{\Delta,corr}(PBMC,TIL)$ also allowed quantitative comparison of VJ-independent divergence between patients, revealing that the TILs of LGG patients were among the most divergent, despite the lower clinical grade of tumor.

To confirm our interpretation that this divergence metric, $JSM_{\Delta,corr}$, represents the degree to which the VJ-independent diversity of the TIL repertoire exceeds that predicted by its PBMC, we compared $JSM_{\Delta,corr}$(PBMC,TIL) to the difference between the VJ-independent diversity of the TIL and of the PBMC [H_{Δ}(TIL) − H_{Δ}(PBMC)] for each patient (Fig. 3*E*). These values correlated strongly (Pearson $R = 0.84$, $P = 1.7 \times 10^{-4}$), indicating that high $JSM_{\Delta,corr}$ reflects diversity in the local T-cell population that (*i*) distinguishes it from the peripheral blood and (*ii*) is associated with its VJ-independent component. Thus, the proportions of diversity and divergence associated with the amino acid binding function of the TIL TCR repertoire, rather than its underlying VJ cassette combination distribution, may provide insight into differential selective pressure in the local tumor microenvironment and may indicate its capacity to respond to antigens endemic to the tissue.

A Common CDR3 Signature in Peripheral Blood Is Associated with Divergence of TIL from PBMC Repertoire. Perturbations to the peripheral T-cell population have been linked to intratumoral immune status in glioma progression and therapy (28–31), implying the potential for noninvasive clinical monitoring through TCRseq. Therefore, we investigated the properties of the peripheral blood repertoires in the context of clinical status, in particular, whether TCRseq-defined immunophenotypes of the TIL corresponded to detectable signatures in the PBMC. We identified the 1,000 highest-abundance amino acid CDR3 motifs (C…FGXG)

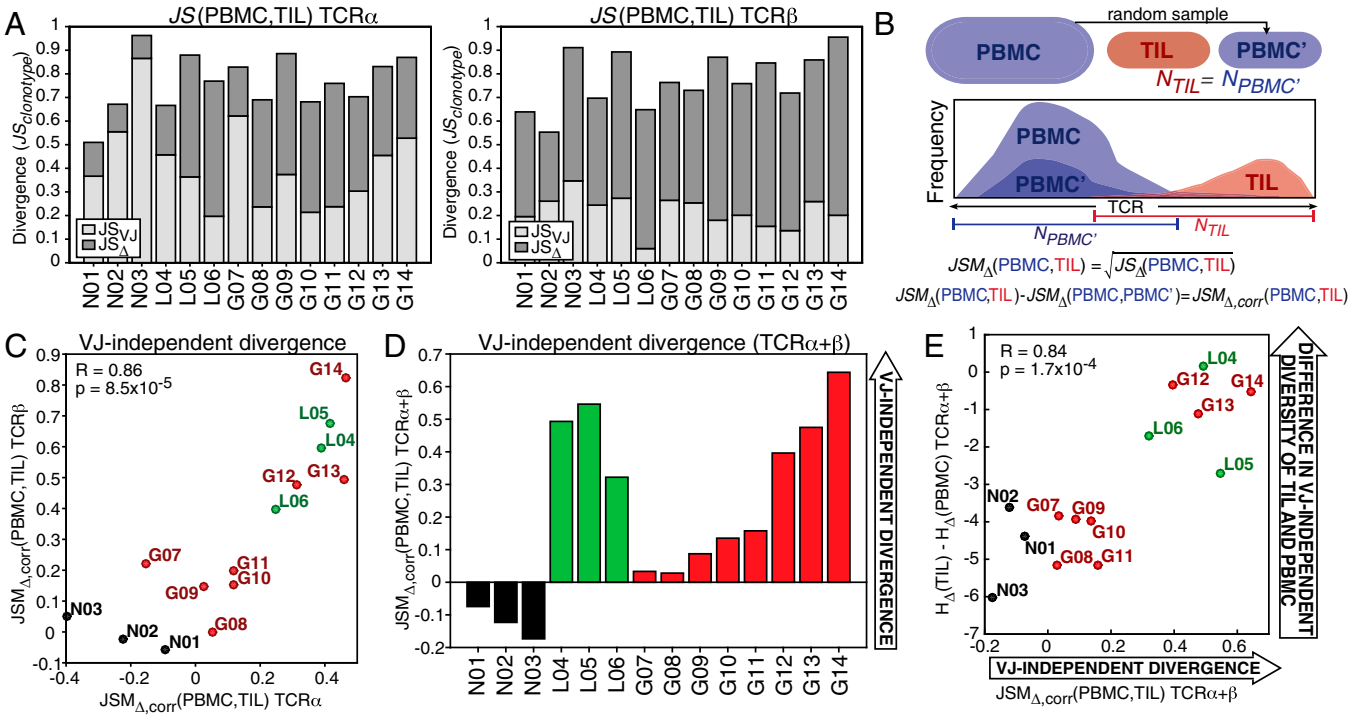

Fig. 3. VJ-independent divergence distinguishes TIL of glioma from nonneoplastic tissue. (*A*) The VJ-dependent (light gray) and -independent (dark gray) Jensen–Shannon divergence, *JS*(PBMC,TIL), was calculated for each patient (*x* axis). (*B*) Statistical divergence between PBMC and TIL due to population size was simulated using a control population (PBMC′), randomly sampled from the PBMC repertoire, containing the same number of clonotypes as the TIL (N_{TIL}), and the JSM_{Δ} between PBMC and PBMC′ was subtracted from JSM_{Δ} between the PBMC and TIL, yielding $JSM_{\Delta,corr}$(PBMC,TIL). (*C*) Correlation between $JSM_{\Delta,corr}$(PBMC,TIL) of TCRα and TCRβ across patients (Pearson $R = 0.86$, $P = 8.5 \times 10^{-5}$). (*D*) The average $JSM_{\Delta,corr}$ of the α and β chains is plotted with colors indicating clinical status (GBM, red; LGG, green; and nonneoplastic, black). (*E*) The $JSM_{\Delta,corr}$(PBMC,TIL) (average of TCRα and TCRβ) of each patient was compared with the difference between the VJ-independent diversity H_{Δ} of the TIL and of the PBMC (Pearson $R = 0.84$, $P = 1.7 \times 10^{-4}$).

from the peripheral blood of each patient, compiling an inclusive top use list for TCRα (11,638) and TCRβ (13,561) with frequencies $\geq 10^{-5}$. For both chains, hierarchical clustering of a binary matrix indicating the presence or absence of these CDR3 amino acid sequences across PBMC samples segregated the patients into two major groups: patients distinguished by common use of a subset of CDR3 amino acid sequences (1242 TCRα, 84 TCRβ, Dataset S3) and a cluster in which these same sequences were less commonly observed (Fig. 4 *A* and *B*, dendrograms). The patient groupings were consistent between the two TCR chains (Fig. 4 *A* and *B*, zoomed panels). Similar patient clustering resulted from multidimensional scaling (MDS) of the frequencies of the top 1,000 CDR3 amino acid sequences (Fig. S4*F*). These "signature" CDR3s were expressed in a distinctively high number of patients compared with the total top use sets (Fig. 4 *A* and *B*, histograms).

The occurrence of these signature amino acid CDR3s correlated with the total number of clonotypes observed in the PBMC repertoires (Fig. S4*G*), which is a function of both the number of cells included in the library and the intrinsic diversity of the repertoire from which they are drawn. Because there was little variation in the number of cells used to produce each PBMC library (Dataset S1), we hypothesized that the high fractional use of these CDR3s reflected repertoire diversity, particularly VJ-independent entropy. Indeed, the fraction of signature CDR3s observed in each PBMC repertoire was significantly correlated with its H_Δ (Fig. 4*C*, Pearson correlation $R = 0.93$, $P = 1.7 \times 10^{-6}$), tying the presence

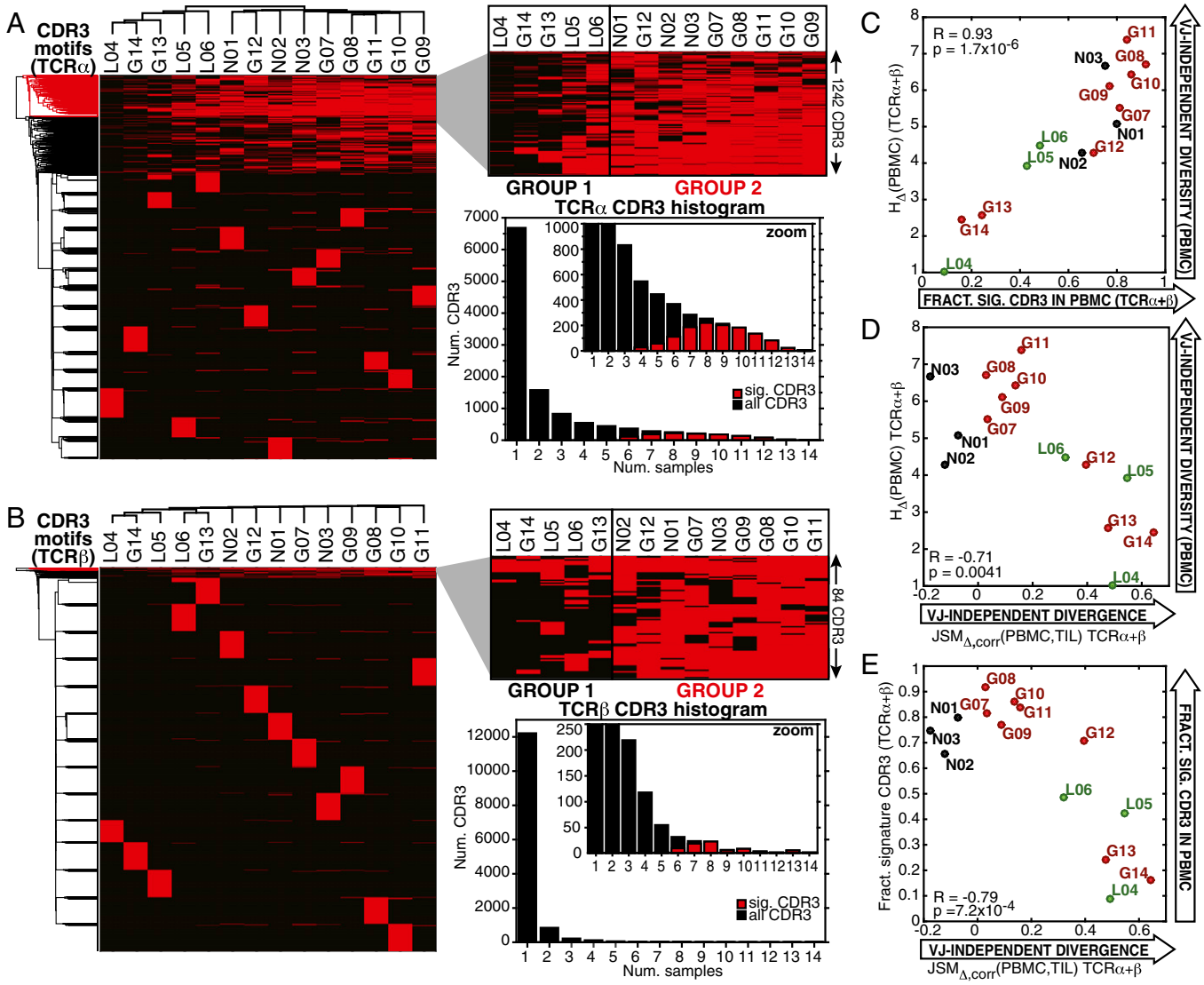

Fig. 4. Use of highly shared CDR3 motifs in peripheral TCR repertoire predict VJ-independent divergence of PBMC and TIL. (*A* and *B*) Hierarchical clustering of top use TCRα and TCRβ *CDR3* amino acid motifs. The top 1,000 CDR3 amino acid motifs from each of the 14 PBMC repertoires were compiled into a high-use list of 11,668 TCRα and 13,561 TCRβ with frequencies $\geq 10^{-5}$. Patients were hierarchically clustered by the presence or absence of these CDR3 motifs (*y* axes) across samples (*x* axes), revealing a subset of subjects (group 2) in which a cluster of 1,242 TCRα and 84 TCRβ motifs were commonly used (red dendrogram clusters) and a cluster of subjects in which these motifs were less frequently observed (group 1). Histograms of the number of samples in which each CDR3 was observed, with the 1,242 signature TCRα (median = 9) and 84 TCRβ (median = 8) highlighted in red, and the total top use set shown in black (median = 1, both chains). (*C*) The correlation between the VJ-independent entropy (H_Δ) of PBMC repertoires and the fraction of the signature CDR3s they contain (average of TCRα and TCRβ for each patient) is plotted (Pearson $R = 0.93$, $P = 1.7 \times 10^{-6}$). (*D*) The correlation between VJ-independent divergence, $JSM_{\Delta,corr}$(PBMC,TIL), and the VJ-independent entropy (H_Δ) of the PBMC (average of TCRα and TCRβ) is plotted (Pearson $R = -0.71$, $P = 0.0041$). (*E*) The VJ-independent divergence, $JSM_{\Delta,corr}$(PBMC,TIL), of each patient is plotted against the fraction of signature CDR3 motifs observed in the PBMC repertoire (average of TCRα and TCRβ), revealing significant anticorrelation (Pearson $R = -0.79$, $P = 7.2 \times 10^{-4}$).

of these motifs to VJ-independent diversity. VJ-independent divergence, $JSM_{\Delta,corr}$ was strongly anticorrelated with H_{Δ}(PBMC) (Fig. 4*D*, Pearson correlation $R = -0.71$, $P = 0.0041$) and was significantly lower among patient group 2, with high use of the signature CDR3s (Fig. S4*H*). The relationship between the use of this signature subset of CDR3 motifs and H_{Δ}(PBMC) allows this diversity signature to function as a peripherally accessible correlate of the divergence between the TIL and PBMC repertoires (Fig. 4*E*, Pearson correlation $R = -0.79$, $P = 7.2 \times 10^{-4}$).

In our dataset, the "signature" CDR3 motifs were strongly represented in healthy individuals (Fig. S5 *D* and *E*), which cluster in group 2. Previous studies have characterized the TCR repertoires in the peripheral blood of healthy individuals (13), and so we tested whether these datasets would also cluster with group 2. Fig. S5 *F–I* shows that 6/6 TCRβ profiles and 5/6 TCRα profiles from healthy individuals obtained in a previously reported study cluster with group 2 through expression of CDR3s which strongly overlap with our signature CDR3 motifs. Thus, the relationship between signature CDR3 use and a highly diverse peripheral repertoire (high H_{Δ}), which coincided with low VJ-independent divergence in our tissue-paired cohort, was robust among the PBMC of healthy subjects. Not only is this signature of repertoire diversity and low PBMC-TIL divergence amenable to noninvasive monitoring, it requires only detection and identification of the CDR3s, rather than precise quantification, at a sampling depth readily accessible from peripheral blood.

A recent study from Britanova et al. (21) identified CDR3 motifs that are shared among many individuals and associated with diverse repertoires. Given that the signature CDR3 motifs in our study are also associated with the diverse repertoires of noncancer patients and certain GBM patients, we reasoned that there might be significant overlap between these two sets of CDR3 motifs. Because the previous study reported only TCRβ motifs, we compared them to our TCRβ signature CDR3s (Table S2, Dataset S3, and *SI Materials and Methods*). Indeed, 62/74 signature CDR3 motifs from our study appear in the Britanova et al. (21) list (~83%), whereas only 632/13,693 top clones that are not among the signature CDR3s appear in the list (<5%). Although we found that the Britanova et al. (21) CDR3 motifs did not cluster our patients as effectively as the signature CDR3s reported here (Fig. S5), the remarkable overlap between these two sets of motifs provides valuable insight into the potential generality of these clones for assessing immune response to disease.

One potential source of shared or "public" CDR3 motifs among individuals arises from exposure to common pathogens. There have been several reports of CDR3 motifs associated with pathogens such as *Clostridium tetani* (toxoid), *Candida albicans*, *Mycobacterium tuberculosis*, HSV, CMV, EBV, and influenza A (32–40). We investigated the overlap of the signature CDR3 motifs with these pathogen-associated motifs for TCRβ (Dataset S3). We found enrichment of our signature CDR3 motifs in those associated with tetanus toxoid (8.1% among "signature," 0.5% among nonsignature), *M. tuberculosis* (2.7% vs. 0.08%), and *C. albicans* (27% vs. 1%) (Table S2). Similar to the "public" motifs from Britanova et al. (21), these pathogen-associated motifs were strongly represented in the PBMC repertoires (Dataset S4), but were not able to cluster our patients as effectively as the signature CDR3 motifs identified here (Fig. S5), but exposure to common pathogens could nonetheless contribute to the prevalence of these CDR3 motifs in the population.

Discussion

Monitoring the strength and specificity of antitumor T-cell reactivity remains a crucial but elusive component of precision immunotherapy. Whereas the immune response itself provides selective pressure and has been associated with both effective tumor clearance and immunoediting of the cancer cell population (41), the combination of the identities and distribution of tumor-responsive TCRs represent a "footprint" of the conditions they face, such as neoantigenic load and dysfunctional immune signaling. Thus, identifying and tracking such T-cell clones has risen to priority as a potential high-dimensional biomarker for tumor development and personalized predictor of the efficacy of immunotherapeutic interventions in cancer (42, 43).

High-throughput sequencing (TCRseq) provides population-wide profiles at a resolution that connects clonal identification to the selectable, functional binding properties of the receptor. Here, we have separated the Shannon entropy of the TCR repertoire into contributions from the diversity of the VJ cassette combination distribution and VJ-independent diversity. The VJ-dependent component is strongly influenced by developmental and lineage restriction, whereas the VJ-independent component includes selection for the antigen-binding affinity of the CDR3 sequence (18, 19). In glioma, these statistical attributes indicated that TILs, especially those of GBM, possess greater VJ-independent diversity than those of nonneoplastic tissue and diverge from their respective PBMC repertoires on the basis of this VJ-independent diversity.

These TCR repertoire phenotypes were consistent with known characteristics of brain TILs. In healthy CNS tissue, surveilling lymphocytes are limited to a relatively small, functionally restricted subpopulation of T cells, which may be expanded following injury or infection, with varying expansion of antigen specificity depending on microenvironmental immune signaling (see review in ref. 44). Our observation of high divergence between PBMC and TIL repertoires in LGG and some GBM provides evidence of tumor-associated T-cell responses in glioma. This overall approach may offer new insight into the coevolution of antitumor reactivity with tumor progression (41) independent and synergistically with gene expression-based immune profiling (45).

Although noninvasive, longitudinal monitoring would have tremendous utility in glioma, efforts to describe a TCR signature predictive of intratumoral immune status have focused on TCR sequences with tumor specificity (46–48). Although our initial calculation of VJ-independent divergence, $JSM_{\Delta,corr}$, required both peripheral blood and biopsied tumor tissue, we discovered a signature set of CDR3 amino acid sequences whose high use in peripheral blood distinguished those patients with low $JSM_{\Delta,corr}$ from those with high $JSM_{\Delta,corr}$, providing a predictor of PBMC–TIL divergence that could be assayed noninvasively. This signature set of CDR3s overlaps not only with previously reported CDR3 motifs associated with TCRβ diversity (21), but also with CDR3s found to be associated with certain common pathogens. Thus, instead of detecting the presence of tumor-reactive TCRs, our signature set reflects peripheral diversity that is depleted in glioma patients with high PBMC–TIL divergence ($JSM_{\Delta,corr}$). These motifs distinguish them from the healthy and tumor-free individuals as well as low-divergence GBM patients, reminiscent of repertoire narrowing observed following strong T-cell responses to certain diseases (10, 16, 19). Our ability to infer high VJ-independent TIL divergence based on a repertoire-wide shift away from the properties of healthy PBMC, not on a shift toward antitumor TCRs, circumvents confounding variability in the antigenic specificities of these TCRs in GBM and other heterogeneous tumors. Furthermore, this signature requires only detection and identification of certain CDR3s, not precise quantification of their frequency—and thus may be translatable across TCRseq platforms.

Direct analysis of repertoire-wide diversity and divergence based on the functionally distinct components of the receptor provides an orthogonal immunophenotype to both the neoantigen load of the tumor and immune signaling in the microenvironment. Diagnosing the capacity of adequately stimulated T cells to recognize tumor antigens and respond locally could provide key guidance for rational combination of immunotherapies, and TCRseq of the peripheral repertoire could enable

longitudinal monitoring of systemic and intratumoral responses to therapy, empowering precision treatment for glioma.

Materials and Methods

Clinical Sample Procurement. Tissue samples of nonneoplastic, low-grade and high-grade glioma were cryofrozen and stored in liquid nitrogen immediately following surgical resections, along with concurrent peripheral blood cells, performed at Columbia University Medical Center through The Neurological Institute, as were tissue samples from nontumor control subjects. Glioma patients included subjects with WHO grade II oligodendrogliomas and astrocytomas, WHO grade IV glioblastoma upon initial (not recurrent) diagnosis. Surgical specimens were collected under Columbia University Institutional Review Board protocol #AAAA4666 (not considered human subjects research) were coded and released as deidentified tissue samples to researchers according to Columbia University Office of Human Research Protection guidelines. Additional healthy peripheral blood was obtained through separate volunteer donation (Table S1).

RNA Isolation and TCR Repertoire Amplification. Total RNA was isolated from cryofrozen human tissue samples using the TissueLyzer system with Qiazol and steel beads (Qiagen). For TCR library preparation, mRNA was isolated using magnetic oligo-dT Dynabeads (Life Technologies), according to manufacturer protocol, and final concentration determined by Qubit (Life Technologies). Tumor tissue samples in this study were ≤300 mg, or $\sim 1 \times 10^6$ total cells, predicted to contain $\leq 1 \times 10^5$ lymphocytes [4–10% of cellularity in glioma tissue (49), and less in nonneoplastic brain tissue]. Total RNA was obtained from 2.5×10^6 PBMCs for each library (the only notable exception was G14, $\sim 7 \times 10^5$ PBMC due to the availability of material) using the RNeasy system (Qiagen) according to manufacturer instructions, with a goal of including up to $\sim 1 \times 10^6$ T cells (Dataset S1, "Input").

We used the commercially available iRepertoire platform (7) for nested amplicon arm-PCR (11) of the CDR3 of the human TCRα and TCRβ chains and addition of adaptors for Illumina platform sequencing. Reverse transcription of 100–400 ng of mRNA (tissue libraries) or total RNA (PBMC libraries) was conducted with a one-step reverse transcription and amplification kit (Qiagen) according to the manufacturer's protocol. The PCR product was purified using AmpureX-100 magnetic beads (Agencourt), and secondary amplification of 40% of the resulting product was performed (Multiplex PCR Kit, Qiagen), allowing addition of Illumina adapter sequences (manufacturer's protocol). Libraries were purified by agarose gel electrophoresis, cutting between 200–400 bp (Fig. 1*A*, predicted amplicon size 210–310 bp), extracted (Qiagen), and sequenced as described below.

Analysis of High-Throughput Sequencing Data. Using an Illumina MiSeq, we obtained 220–250 nt paired-end reads (95 ± 2% reads passing filter, averaged over runs). We targeted 4.5–10× sequence coverage of $\sim 1 \times 10^6$ peripheral T cells (*SI Materials and Methods*).

Raw paired-end fastq files were first demultiplexed based on the internal 6-nt barcode sequences added during library construction. The paired reads were then merged using FLASH 1.2.11 (flash –M 250 -O) (50) and aligned to the human genome (GRCh37) using the Burrows–Wheeler Aligner (bwa mem) (51). Reads mapping to the T-cell receptor loci (TRA and TRB) were then extracted, associated with V- and J-cassettes, and translated in silico in all three reading frames using the genetic code. Reading frames containing a C. . .FGXG amino acid motif that was uninterrupted by a stop codon were identified as productive CDR3 amino acid sequences. For each demultiplexed sample, all V- and J-cassettes were then reference-corrected and the number of reads identified with each unique combination of V and J cassettes encoding a CDR3 amino acid sequence were counted. Reads were filtered by Hamming distance as thresholded by the sequence error rate in the mapped regions of the sample (detailed in *SI Materials and Methods*).

Shannon Entropy-Derived Statistics. We selected Shannon entropy (*H*) as our fundamental measure of the diversity (52), which has been widely used in previous TCR repertoire studies for its inclusion of both richness (number of members) and diversity (evenness of distribution) and its applicability to more complex models of repertoire size (17, 18). We calculated entropy for the CDR3 amino acid motif, VJ cassette combination, and clonotype repertoire of each library (Dataset S2) and described the entropy of the VJ-dependent (genome-encoded cassette sequence) and VJ-independent (all other sources of sequence diversity and bias) as separable components of each clonotype repertoire. Saturation of the entropy properties of the repertoires was demonstrated by in silico downsampling of the reads of each PBMC and TIL library, yielding H_Δ (Fig. S3*A*) for numbers of reads and clonotypes (Fig. S3*B*) spanning approximately two orders of magnitude for TCRα and TCRβ. Similarly, we calculated the Jensen–Shannon divergence (*JS*) between each TIL repertoire from its paired PBMC, which has been used to compare the peripheral blood repertoires of individuals (13). The square root of *JS*, the Jensen–Shannon divergence metric (*JSM*), is then corrected for divergence due to the size of the TIL population, allowing us to report the size-corrected, VJ-independent divergence metric, $JSM_{\Delta,corr}$. These calculations are detailed in *SI Materials and Methods*. All entropy characterization of the repertoires was performed using code in Python2.7.5, tabulated and displayed using MATLAB (R2014b).

Clustering by CDR3 Amino Acid Motifs in Peripheral Blood. We compiled the 1,000 highest-abundance TCRα or TCRβ CDR3 amino acid motifs (including [C. . .FGXG], the amino acids relevant to CDR3-region binding, both VJ-encoded and VJ-independent) with a frequency of $\geq 10^{-5}$ from each peripheral blood library into a top use list for each chain across all patients. Two binary matrices, reflecting the presence or absence of each TCRα or TCRβ CDR3 motif in each sample, were hierarchically clustered (Euclidean distance) in MATLAB (R2014b), producing the clustergrams in Fig. 4 *A* and *B*. The distinct upper clusters of each CDR3 motif dendrogram (in red in Fig. 4, 1242 TCRα and 84 TCRβ) were designated signature CDR3 motifs. The fraction of these signature CDR3 motifs present in each PBMC repertoire was tabulated, and Pearson correlations of this fraction with other properties of the patient repertoires [e.g., $JSM_{\Delta,corr}$(PBMC,TIL)] in Fig. 4 were calculated in MATLAB (R2014b). MDS was performed using the frequencies of the same top use TCRα and TCRβ *CDR3* motifs (combined list) in MATLAB using the "cmdscale" function. This associated the same "group 2" patients as the hierarchical CDR3 motif clustering (Fig. S4*F*), onto which the fraction of signature CDR3 motifs in each PBMC repertoire was projected in color to illustrate the relationship between their fractional use and patient clustering by MDS.

ACKNOWLEDGMENTS. We thank Erin Bush and Drs. Xiaojun Feng and Xiaoyun Sun (JP Sulzberger Columbia Genome Center) for their technical assistance and Dr. Tao Su (Herbert Irving Comprehensive Cancer Center Molecular Pathology Shared Resource) and Dr. Anthony Sireci and Samantha Cano (Department of Pathology) for sharing instrumentation. We also thank the surgeons and clinical staff of the Columbia Neurological Institute, particularly Dr. Michael B. Sisti and Dr. Guy M. McKhann II, for procuring surgical specimens. J.N.B. and P.C. were supported by an anonymous donor. This work was supported by NIH/National Institute of Biomedical Imaging and Bioengineering Grant K01EB016071 (to P.A.S.), funds from the Columbia Department of Systems Biology and the JP Sulzberger Columbia Genome Center (Y.S.), and NIH/National Institute of Neurological Disorders and Stroke Grant R01NS066955 (to P.C.).

1. Phillips HS, et al. (2006) Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* 9(3):157–173.
2. Verhaak RG, et al.; Cancer Genome Atlas Research Network (2010) Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 17(1):98–110.
3. Ivliev AE, 't Hoen PA, Sergeeva MG (2010) Coexpression network analysis identifies transcriptional modules related to proastrocytic differentiation and sprouty signaling in glioma. *Cancer Res* 70(24):10060–10070.
4. Doucette T, et al. (2013) Immune heterogeneity of glioblastoma subtypes: Extrapolation from the cancer genome atlas. *Cancer Immunol Res* 1(2):112–122.
5. Murat A, et al. (2009) Modulation of angiogenic and inflammatory response in glioblastoma by hypoxia. *PLoS One* 4(6):e5947.
6. Weinstein JA, Jiang N, White RA, 3rd, Fisher DS, Quake SR (2009) High-throughput sequencing of the zebrafish antibody repertoire. *Science* 324(5928):807–810.
7. Wang C, et al. (2010) High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets. *Proc Natl Acad Sci USA* 107(4):1518–1523.
8. Boyd SD, et al. (2009) Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. *Sci Transl Med* 1(12):12ra23.
9. Freeman JD, Warren RL, Webb JR, Nelson BH, Holt RA (2009) Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res* 19(10):1817–1824.
10. Woodsworth DJ, Castellarin M, Holt RA (2013) Sequence analysis of T-cell repertoires in health and disease. *Genome Med* 5(10):98.
11. Han J, et al. (2006) Simultaneous amplification and identification of 25 human papillomavirus types with Templex technology. *J Clin Microbiol* 44(11):4157–4162.
12. Yang Y, et al. (2015) Distinct mechanisms define murine B cell lineage immunoglobulin heavy chain (IgH) repertoires. *eLife* 4:e09083.
13. Zvyagin IV, et al. (2014) Distinctive properties of identical twins' TCR repertoires revealed by high-throughput sequencing. *Proc Natl Acad Sci USA* 111(16):5980–5985.
14. Putintseva EV, et al. (2013) Mother and child T cell receptor repertoires: Deep profiling study. *Front Immunol* 4:463.
15. Zarnitsyna VI, Evavold BD, Schoettle LN, Blattman JN, Antia R (2013) Estimating the diversity, completeness, and cross-reactivity of the T cell repertoire. *Front Immunol* 4:485.

16. Merkenschlager J, Kassiotis G (2015) Narrowing the gap: Preserving repertoire diversity despite clonal selection during the CD4 T cell response. *Front Immunol* 6:413.
17. Murugan A, Mora T, Walczak AM, Callan CG, Jr (2012) Statistical inference of the generation probability of T-cell receptors from sequence repertoires. *Proc Natl Acad Sci USA* 109(40):16161–16166.
18. Elhanati Y, Murugan A, Callan CG, Jr, Mora T, Walczak AM (2014) Quantifying selection in immune receptor repertoires. *Proc Natl Acad Sci USA* 111(27):9875–9880.
19. Turner SJ, Doherty PC, McCluskey J, Rossjohn J (2006) Structural determinants of T-cell receptor bias in immunity. *Nat Rev Immunol* 6(12):883–894.
20. Ndifon W, et al. (2012) Chromatin conformation governs T-cell receptor Jβ gene segment usage. *Proc Natl Acad Sci USA* 109(39):15865–15870.
21. Britanova OV, et al. (2014) Age-related decrease in TCR repertoire diversity measured with deep and normalized sequence profiling. *J Immunol* 192(6):2689–2698.
22. Emerson R, et al. (2013) Estimating the ratio of CD4+ to CD8+ T cells using high-throughput sequence data. *J Immunol Methods* 391(1-2):14–21.
23. Gregersen PK, Hingorani R, Monteiro J (1995) Oligoclonality in the CD8+ T-cell population. Analysis using a multiplex PCR assay for CDR3 length. *Ann N Y Acad Sci* 756:19–27.
24. Robert L, et al. (2014) CTLA4 blockade broadens the peripheral T-cell receptor repertoire. *Clin Cancer Res* 20(9):2424–2432.
25. Tumeh PC, et al. (2014) PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* 515(7528):568–571.
26. Radebe M, et al. (2015) Broad and persistent Gag-specific CD8+ T-cell responses are associated with viral control but rarely drive viral escape during primary HIV-1 infection. *AIDS* 29(1):23–33.
27. Morris H, et al. (2015) Tracking donor-reactive T cells: Evidence for clonal deletion in tolerant kidney transplant patients. *Sci Transl Med* 7(272):272ra10.
28. Kmiecik J, et al. (2013) Elevated CD3+ and CD8+ tumor-infiltrating immune cells correlate with prolonged survival in glioblastoma patients despite integrated immunosuppressive mechanisms in the tumor microenvironment and at the systemic level. *J Neuroimmunol* 264(1-2):71–83.
29. Okada H, Khoury SJ (2012) Type17 T-cells in central nervous system autoimmunity and tumors. *J Clin Immunol* 32(4):802–808.
30. Waziri A, et al. (2008) Preferential in situ CD4+CD56+ T cell activation and expansion within human glioblastoma. *J Immunol* 180(11):7673–7680.
31. Fong B, et al. (2012) Monitoring of regulatory T cell frequencies and expression of CTLA-4 on T cells, before and after DC vaccination, can predict survival in GBM patients. *PLoS One* 7(4):e32614.
32. Becattini S, et al. (2015) T cell immunity. Functional heterogeneity of human memory $CD4^+$ T cell clones primed by pathogens or vaccines. *Science* 347(6220):400–406.
33. Venturi V, et al. (2008) TCR beta-chain sharing in human CD8+ T cell responses to cytomegalovirus and EBV. *J Immunol* 181(11):7853–7862.
34. Lim A, et al. (2000) Frequent contribution of T cell clonotypes with public TCR features to the chronic response against a dominant EBV-derived epitope: Application to direct detection of their molecular imprint on the human peripheral T cell repertoire. *J Immunol* 165(4):2001–2011.
35. McCluskey J, Kanaan C, Diviney M (2003) Nomenclature and serology of HLA class I and class II alleles. *Curr Protoc Immunol* 52(1S):A.1S.1–A.1S.8.
36. Wucherpfennig KW, Strominger JL (1995) Molecular mimicry in T cell-mediated autoimmunity: Viral peptides activate human T cell clones specific for myelin basic protein. *Cell* 80(5):695–705.
37. Tosato G, Cohen JI (2007) Generation of Epstein-Barr virus (EBV)-immortalized B cell lines. *Curr Protoc Immunol* Chap 7:Unit 7.22.
38. Bhaduri-McIntosh S, Rotenberg MJ, Gardner B, Robert M, Miller G (2008) Repertoire and frequency of immune cells reactive to Epstein-Barr virus-derived autologous lymphoblastoid cell lines. *Blood* 111(3):1334–1343.
39. Lossius A, et al. (2014) High-throughput sequencing of TCR repertoires in multiple sclerosis reveals intrathecal enrichment of EBV-reactive CD8+ T cells. *Eur J Immunol* 44(11):3439–3452.
40. Suessmuth Y, et al. (2015) CMV reactivation drives posttransplant T-cell reconstitution and results in defects in the underlying TCRβ repertoire. *Blood* 125(25):3835–3850.
41. Angelova M, et al. (2015) Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. *Genome Biol* 16:64.
42. Drake CG, Lipson EJ, Brahmer JR (2014) Breathing new life into immunotherapy: review of melanoma, lung and kidney cancer. *Nat Rev Clin Oncol* 11(1):24–37.
43. Sims JS, Ung TH, Neira JA, Canoll P, Bruce JN (2015) Biomarkers for glioma immunotherapy: The next generation. *J Neurooncol* 123(3):359–372.
44. Carson MJ, Doose JM, Melchior B, Schmid CD, Ploix CC (2006) CNS immune privilege: Hiding in plain sight. *Immunol Rev* 213:48–65.
45. Guan X, et al. (2014) Molecular subtypes of glioblastoma are relevant to lower grade glioma. *PLoS One* 9(3):e91216.
46. Sensi M, Parmiani G (1995) Analysis of TCR usage in human tumors: A new tool for assessing tumor-specific immune responses. *Immunol Today* 16(12):588–595.
47. Gros A, et al. (2014) PD-1 identifies the patient-specific $CD8^+$ tumor-reactive repertoire infiltrating human tumors. *J Clin Invest* 124(5):2246–2259.
48. Linnemann C, et al. (2013) High-throughput identification of antigen-specific TCRs by TCR gene capture. *Nat Med* 19(11):1534–1541.
49. Hitchcock ER, Morris CS (1988) Mononuclear cell infiltration in central portions of human astrocytomas. *J Neurosurg* 68(3):432–437.
50. Magoč T, Salzberg SL (2011) FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27(21):2957–2963.
51. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5):589–595.
52. Shannon CE (1948) A mathematical theory of communication. *At&T Tech J* 27(3): 379–423.
53. Warren RL, et al. (2011) Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes. *Genome Res* 21(5):790–797.
54. Lefranc MP, et al. (2015) IMGT®, the international ImMunoGeneTics information system® 25 years on. *Nucleic Acids Res* 43(Database issue):D413–D422.
55. Krzywinski M, et al. (2009) Circos: An information aesthetic for comparative genomics. *Genome Res* 19(9):1639–1645.
56. Endres DM, Schindelin JE (2003) A new metric for probability distributions. *Ieee T Inform Theory* 49(7):1858–1860.
57. Klarenbeek PL, et al. (2012) Deep sequencing of antiviral T-cell responses to HCMV and EBV in humans reveals a stable repertoire that is maintained for many years. *PLoS Pathog* 8(9):e1002889.
58. Emerson R, et al. (September 10, 2015) Immunosequencing reveals diagnostic signatures of chronic viral infection in T cell memory. *BioRxiv*, 10.1101/026567.
59. Dziubianau M, et al. (2013) TCR repertoire analysis by next generation sequencing allows complex differential diagnosis of T cell-related pathology. *Am J Transplant* 13(11):2842–2854.
60. Hong J, et al. (1999) A common TCR V-D-J sequence in V beta 13.1 T cells recognizing an immunodominant peptide of myelin basic protein in multiple sclerosis. *J Immunol* 163(6):3530–3538.
61. Junker A, et al. (2007) Multiple sclerosis: T-cell receptor expression in distinct brain regions. *Brain* 130(Pt 11):2789–2799.
62. Bolotin DA, et al. (2012) Next generation sequencing for TCR repertoire profiling: Platform-specific features and correction algorithms. *Eur J Immunol* 42(11): 3073–3083.