Circulating Clonally Expanded T Cells Reflect Functions of Tumor Infiltrating T Cells

Liliana Lucca[1], Pierre-Paul Axisa[1], Benjamin Lu[1-2], Brian Harnett[1], Shlomit Jessel[2], Le Zhang[1], Khadir Raddassi[1], Lin Zhang[2], Kelly Olino[3], James Clune[3], Meromit Singer[4,5], Harriet M. Kluger[2]*, David A. Hafler[1,5]*†

Affiliations:

[1]Departments of Neurology and Immunobiology, Yale School of Medicine, New Haven, CT, USA.

[2]Department of Medicine, Yale School of Medicine, New Haven, CT, USA

[3]Department of Surgery, Yale School of Medicine, New Haven, CT, USA

[4]Dana Farber Cancer Institute, Harvard Medical School, Boston, MA 02115, USA

[5]Broad Institute of MIT and Harvard University, Cambridge, MA USA 02142, USA

*These authors contributed equally to the work

† Corresponding author: david.hafler@yale.edu; Departments of Neurology and Immunobiology, Yale School of Medicine, 15 York Street, New Haven, CT 06520 orcid.org/0000-0003-4664-535X

1 **Running title**

2 Single-cell RNAseq of circulating TILs in Melanoma

3

4 **Summary**

5 Combining transcriptomic and TCRαβ repertoire analysis of circulating and tumor-infiltrating

6 CD8 T cells from patients with metastatic melanoma, we identify a blood-based population with

7 effector properties that reflect those of clonally related tumor-resident T cells.

8

9 **Abstract**:

10   Understanding the relationship between tumor and peripheral immune environments could

11 allow longitudinal immune monitoring in cancer. Here, we examined whether T cells that share

12 the same TCRαβ and are found in both tumor and blood can be interrogated to gain insight into

13 the ongoing tumor T cell response. Paired transcriptome and TCRαβ repertoire of circulating and

14 tumor-infiltrating T cells were analyzed from matched tumor and blood from patients with

15 metastatic melanoma at the single cell level. We found that in circulating T cells matching clonally

16 expanded tumor-infiltrating T cells (circulating TILs), gene signatures of effector functions, but

17 not terminal exhaustion, reflect those observed in the tumor. In contrast, features of exhaustion are

18 displayed predominantly by T cells present only in tumor. Finally, genes associated with a high

19 degree of blood-tumor TCR sharing were overexpressed in tumor tissue after immunotherapy.

20 These data demonstrate that circulating TILs, identified by TCRs shared with T cells in tumors,

21 have unique transcriptional expression patterns that may have utility for the interrogation of T cell

22 function in cancer immunotherapy.

Introduction

Tumor infiltration by T cells, particularly CD8$^+$ cells, is a significant prognostic factor in cancer (reviewed in Barnes and Amir, 2017). Cytotoxic T cells, characterized by lytic granule components and NK-associated receptors, play a pivotal role in tumor rejection by directly destroying tumor cells and associated structures in an antigen-specific fashion (Breart et al., 2008; Schietinger et al., 2013). In contrast, exhausted T cells are defined by sequential loss of effector functions, limited proliferative potential (Blank et al., 2019; Moskophidis et al., 1993) and increased, stable expression of co-inhibitory receptors (reviewed in Blank et al., 2019; van der Leun et al., 2020). Inhibition of co-inhibitory signals has proven highly successful in the treatment of solid malignancies (Ribas and Wolchok, 2018). Understanding the relationship between tumor and peripheral immune environments could allow longitudinal immune monitoring in situations in which tumor biopsies are unfeasible (Cohen and Buchbinder, 2019). Moreover, serial sampling for predicting and monitoring response to therapy is often necessary when weighing risk/benefit ratios for continuation of treatment.

Here, we analyzed paired transcriptome and TCRαβ repertoire of circulating and tumor-infiltrating T cells from matched tumor and blood samples from patients with metastatic melanoma at the single cell level. We tested the hypothesis that circulating sister clones of clonally expanded TILs, i.e. clones with matching TCR sequences, can be interrogated to gain insight into the ongoing T cell response in the tumor. We found that the degree of cytotoxicity of the tumor infiltrate, but not exhaustion, can be inferred from circulating TILs. The most pronounced features of exhaustion are consistently displayed by tumor-exclusive clonal T cells. Thus, the degree of TCR repertoire sharing across tissues predicts the T cell phenotype, and genes associated with a high degree of

1 blood-tumor TCR sharing are overexpressed by tissue-infiltrating T cells after successful

2 immunotherapy.

3

4 **Results and Discussion**

5 *Detection of clonally-expanded TILs in the circulation of melanoma patients*

6 We analyzed paired transcriptomic and TCRαβ libraries from a total of 123,868 T cells

7 derived from matched metastatic tumors and blood from 11 stage IV melanoma patients selecting

8 subjects with diverse treatment history (Table 1). A single TCRαβ clonotype (see Methods) could

9 be detected in 66.1% (48.5% − 76.6%) of the cells, and only these cells were retained for

10 subsequent analyses. A reduction in the diversity of immune repertoires signals an ongoing

11 immune response. We computed four estimates of repertoire diversity and compared tumor and

12 blood repertoires across patients and observed a significant reduction in the richness of clonotype

13 species in the tumor as calculated by the Chao (Chao et al., 1992) and ACE indexes (Laydon et

14 al., 2014) (Fig 1A). To identify patterns of similarity, we performed hierarchical clustering based

15 on clonal size distribution and observed an overall separation of blood and tumor samples with

16 two tumor samples, Mel UT-2 and Mel T-2, clustering closer to blood samples (Fig. 1B). These

17 two samples consistently ranked lowest when measuring the proportion of the repertoire occupied

18 by the top 10 most frequent clonotypes (Fig 1C). Of note, Mel UT-2 was a mucosal melanoma and

19 therefore sun-shielded (Table 1), consistent with the correlation between UV radiation exposure

20 and immune infiltration (Wang et al., 2017). These analyses confirmed previous results showing

21 restricted TCR repertoires in melanoma, indicative of ongoing antigen-specific responses (Kalaora

22 et al., 2018). We then classified individual TIL clonotypes into clonally expanded or not, based on

23 the proportion of the repertoire occupied by each clonotype. Because of significant variability in

24 the number of analyzed TILs across patients (1,127 - 5,404), for this classification we established

1   individual percentage-based thresholds (Fig. S1). We next examined whether sister clones of

2   expanded TILs could be detected in the blood based on exact clonotype matching. Comparing the

3   proportion of tumor and blood repertoires occupied by the 20 most expanded TIL clonotypes

4   highlights the existence of both shared and tumor exclusive large clones (Fig 1D). Based on the

5   largely mutually exclusive expression of the *CD8A* and *CD4* genes, we found that both tumor-

6   resident and circulating expanded TILs were disproportionally CD8 (Fig 1E-G). Finally, we

7   observed a positive correlation between the frequency of circulating TILs in blood, (average 7.5%

8   ± 10.4%, Fig 1H) and the duration of metastatic disease (Table 1 and Fig 1I).

9   *Cytotoxic profile of circulating TILs*

10   We then examined whether clonally expanded T cells in the tumor displayed specific

11   functional traits by analyzing transcriptional profiles of individual CD8 T cells from all patients

12   by grouping similar cells into clusters. We identified seventeen clusters (Fig 2A), and clonally

13   expanded TILs largely map onto clusters overrepresented in the tumor (Fig 2B). We further

14   annotated each cluster by identifying cluster-specific gene markers (Table S1) and comparing

15   these markers with gene sets derived from previous studies of the immune infiltrate in melanoma

16   at the single cell level (Li et al., 2019; Sade-Feldman et al., 2018; Tirosh et al., 2016). A large

17   subgroup of blood-dominated clusters (6, 13, 16, 8) corresponded to naïve and central memory

18   CD8 T cells, defined by the high expression of *CCR7* and *IL7R*, overlapping with published

19   signatures of central memory (Fig 2C-D). Clusters 2 and 12 display both signatures of central and

20   effector memory, while clusters 3, 9 and 14 express high levels of cytotoxic molecules and cluster

21   together with previously reported cytotoxic populations (Fig 2C-D). Tumor-dominated clusters 0,

22   5, 10, 15 were enriched in signatures of exhaustion and terminal differentiation and are indeed

23   marked by expression of *PDCD1* (Fig 2C-D). Clusters 1, and 11 and presented an intermediate

4

1   profile, with both signatures of exhaustion and cytotoxicity, and higher expression of *GZMK* than

2   *GZMB* (Zhang et al., 2019) (Fig 2C-D), possibly representing pre-dysfunctional T cells (Li et al.,

3   2018). We also identified two clusters of actively proliferating cells (4 and 7), expressing high

4   levels of *MKI67* and predominantly containing tumor-derived, clonally expanded T cells.

5   In the blood, circulating TILs map to both cytotoxic and pre-dysfunctional clusters. To

6   further define transcriptional traits associated with this population, we compared transcription

7   profiles of circulating TILs and other circulating T cells. We detected higher expression of multiple

8   cytotoxic molecules such as granzymes, perforin and NK-associated receptors, and also markers

9   of tissue residence (*LGALS1*), migration (*CD99*) and tissue homing (*ITGB7*). Finally, *LAG3*, but

10   not other coinhibitory receptors, was also significantly upregulated by circulating TILs (Fig 2E).

11   *Tumor signature of circulating TILs*

12

13   We then examined whether circulating TILs are enriched for features of tumor resident,

14   clonally expanded TILs. We focused on seven patients where a minimum of 20 circulating TILs

15   could be detected (Fig 1H). For each patient, we defined expanded TIL gene sets consisting of

16   genes differentially expressed between clonally expanded TILs and tumor-resident T cells with a

17   unique TCR (singletons, Fig 3A). Similarly, we compared clonally expanded, blood-derived T

18   cells (based on thresholds shown in Fig. S2) and blood-derived singletons (Fig 3A). To obtain a

19   list of genes specifically associated with clonal expansion in the tumor, we incorporated a term for

20   the interaction between clonal expansion and tissue into the linear model used for differential

21   expression detection (Fig 3B). This analysis yielded two types of gene sets: those containing all

22   genes relevant to clonal expansion in the tumor, including hallmarks of general activation

23   (expanded TIL gene sets), and those containing only genes found exclusively when T cells expand

24   in the tumor (tumor-specific gene sets, Table S3). We then calculated the overlap between

1   expanded TIL gene sets from individual patients. The expanded TIL gene sets of Met-T-1 and

2   Mel-T-2, the two patients with BRAF/NRAS/c-kit wild-type tumors, presented a high degree of

3   overlap and shared multiple markers of cytotoxicity. A high degree of similarity was also observed

4   for subjects Mel-T-5 and Mel-T-12, enriched for signatures of exhaustion (Fig 3C, Table S3).

5   We then used expanded TIL gene sets to score individual CD8 T cells from the blood, and

6   compared circulating TILs to other circulating T cells. Circulating TILs indeed had higher levels

7   of aggregated expression of genes contained in expanded TIL gene sets in six out of seven patients

8   (Fig 3D). Because expanded TIL gene sets contain multiple markers of activation, and, as shown

9   in Fig 2A-B, circulating TILs are an effector population, it is possible that this result may just

10 reflect a difference in activation. Therefore, we also measured the expression of tumor-specific

11 gene sets in circulating TILs. For four out of seven patients, circulating TILs were also enriched

12 for tumor-specific gene sets (Fig 3E). To visualize trends of expression of individual genes

13 upregulated by clonally expanded TILs, we selected genes contained in the expanded TIL modules

14 of at least three patients and compared differential expression between tumor-resident expanded

15 TILs, circulating TILs and other circulating T cells. Of 566 genes tested, 283 were more highly

16 expressed in circulating TILs compared to other circulating T cells, including multiple markers of

17 effector functions. In contrast, 200 genes had similarly lower expression in blood compared to

18 tumor regardless of TCR sharing. These genes included inhibitory receptors and other markers of

19 terminal differentiation (Fig 3F, top 50 genes). Likewise, of 105 genes induced in the tumor-

20 specific sets of at least two patients, 29 were significantly upregulated in circulating TILs, while

21 35 were not differentially expressed between blood populations. The tumor-specific genes

22 reflected in circulating TILs included effector molecules (*IL32, IFNG*), enzymes involved in

23 carbohydrate metabolism (*GAPDH, GPI, GYG1*) and genes involved in cytoskeleton dynamics

1 (*ACTG, MYL6, PFN1, CFL1*), possibly mediating adaptations required for tissue residency

2 (Kumar et al., 2017). Tumor-specific genes not reflected in circulating TILs also included some

3 regulators of cell structure (*TUBA1B, TUBB, COTL1*), and additional inhibitory molecules

4 (*HAVCR2, CTLA4*) and transcription factors associated with exhaustion (*TOX*) and tissue

5 residency (*ID2*, Fig 3G). Thus, these data indicate that circulating TILs reflect certain features of

6 the ongoing T cell response in the tumor including adaptations to tissue residency, but also reveal

7 that key features of exhaustion in the tumor cannot be inferred from the periphery.

8 *Co-expression of KLRD1 and CD74 enriches for circulating TILs*

9 We then examined whether circulating TILs can be identified in the blood based on

10 expression of gene markers. We classified circulating CD8 T cells populations based on sharing

11 of TCRs expanded in the tumor. The transcriptomes of individual cells with this annotation were

12 used as the input of COMET (Combinatorial Marker Detection from Single-Cell Transcriptomic

13 data, (Delaney et al., 2019), a tool that uses the minimal hypergeometric test to rank single and

14 paired gene marker panels from a curated list of surface-expressed genes (Chihara et al., 2018) for

15 their ability to identify predefined populations (Fig S3A). Of fourteen candidate gene marker pairs

16 identified in five out of seven patients, the *KLRD1-CD74* module had the highest average rank and

17 a highly significant q value in most patients (Fig S3B). Moreover, *KLRD1* and *CD74* were also

18 highly ranked as candidate gene markers of circulating TILs in the single gene marker analysis for

19 most patients (Fig S3C). We determined whether cells expressing high levels of the *KLRD1-CD74*

20 module were enriched for circulating TILs. Indeed, selecting for cells that co-express *KLRD1* and

21 *CD74* genes (Fig S3D) enriches for circulating TILs (Fig S3E). Interestingly, a significantly higher

22 frequency of circulating TILs could not be found by classifying CD8 T cells by *PDCD1* gene

23 expression alone (Fig S3F). Analysis of PD-1 expression at the protein level consistently revealed

that while tumor-resident expanded TILs display higher levels of both *PDCD1* transcript and PD-1 protein than unexpanded cells, circulating TILs express similarly low levels of PD-1 as other circulating T cells (Fig S3G). Overall, these data suggest that it may be possible to identify a significant proportion of circulating T cells that are clonally related to tumor-resident T cells by measuring transcripts encoding for cell surface proteins in the absence of TCR sequence information.

Tumor-blood sharing as a predictor of cytotoxic phenotype

Because many markers of exhaustion could not be detected in the circulation, we hypothesized that these are predominantly expressed by tumor-exclusive clones. To this end, we grouped CD8 TILs and circulating CD8 T cells into clonal groups, i.e. groups of cells sharing the same TCR, and calculated expansion and blood-tumor sharing scores for each clonal group. Mapping the tumor expansion and sharing scores onto the TIL tSNE plot shows clearly that most highly expanded clonal groups are not highly shared with the blood (Fig 4A-B).

We then determined whether gene expression in the tumor is associated with these metrics by calculating average Spearman's rho coefficients for each gene. The expansion score appears to be correlated with an activation signature, with markers of naïve cells, (*IL7R, CCR7, LEF1*) displaying negative coefficients. Low values of the sharing score, instead, were associated with markers of exhaustion and terminal differentiation, (*HAVCR2, PDCD1, CXCL13, FABP5, IFNG*, Fig 4A-B). This approach did not allow us to separate the relative contribution of expansion and degree of sharing to these phenotypes. In fact, cumulative measures calculated for all the expanded clonal groups of each patient confirm that these two metrics are not independent, but rather inversely correlated, specifically in the tumor (Fig 4C).

1   To separate the effects of expansion and degree of sharing, we selected clonal groups for

2   which either of these parameters was constant. Specifically, we identified four clonal group pairs

3   with similar expansion scores within pairs, and opposite degrees of blood-tumor sharing. Similarly,

4   we selected six clonal group pairs that were equally tumor exclusive and presented different

5   degrees of clonal expansion (Fig 4D). We noted that pairs that differ by degree of sharing occupy

6   distinct areas of the cell space, whereas pairs with variable expansion scores all map to the same

7   area (Fig 4E-F). This difference could be quantified by calculating the average Euclidean distances

8   for each clonal group pair. Indeed, the distances between exclusive versus shared clonal groups

9   were significantly larger than the ones between highly versus lowly expanded clonal groups (Fig

10   4G). These data indicate that in the tumor infiltrate, the degree of sharing with blood explains more

11   variation than the degree of clonal expansion. Moreover, we identified differentially expressed

12   genes among cells belonging to these clonal groups and were able to confirm that TILs with a TCR

13   shared with blood T cells were highly enriched for signatures of central memory and cytotoxicity,

14   and presented less features of terminal exhaustion (Fig 4H).

15   As we demonstrated that circulating TILs express cytotoxic genes but low levels of

16   exhaustion markers, we examined whether the cytotoxic score of circulating TILs was predictive

17   of the cytotoxic score of the tumor infiltrate of the same patient. We scored individual tumor-

18   derived CD8 TILs and circulating TILs for expression of a set of cytotoxicity markers previously

19   reported in tumor-infiltrate T cells in single-cell RNAseq-based studies of multiple human cancers

20   (see Methods and van der Leun et al., 2020). We observed a significant direct correlation between

21   cytotoxicity scores in the tumor versus the circulating TIL population (Fig 4I). Moreover, this was

22   not the case when scoring tumor-derived and circulating TILs for an exhaustion signature (Fig 4I),

1   indicating that the degree of cytotoxicity of the tumor infiltrate, but not exhaustion, can be inferred

2   from the periphery by analyzing circulating TILs.

3   *Signature of shared clones in the tissue of responders*

4   It has been recently reported that clinical response to immunotherapy is accompanied by

5   an influx of new T cell clones from the periphery into the tumor microenvironment, suggesting

6   that TILs with a high degree of TCR sharing with blood can be involved in tumor rejection (Yost

7   et al., 2019). To address this, we classified expanded TILs from the immunotherapy naïve patient

8   Mel-T-1 into shared with blood or tumor exclusive, and scored individual cells for gene sets

9   associated with response or disease progression after checkpoint immunotherapy (Sade-Feldman

10   et al., 2018). While tumor exclusive TILs resembled TILs of progressors, genes predicting

11   response were more represented in TILs shared with blood (Fig 5A-B). Furthermore, we examined

12   whether defining sets of genes based on their correlation with the degree of blood-tumor sharing

13   (Table S4) was sufficient to identify genes relevant to tumor rejection after immunotherapy. We

14   analyzed residual tissue from two patients whose tumors responded clinically to checkpoint

15   inhibitor immunotherapy. These samples did not contain any discernible tumor tissue, but rather

16   focal aggregation of inflammatory cells including T cells and activated macrophages. We refer to

17   these as tissue-infiltrating T cells. Similar to what we observed for tumor recurrent patients, in the

18   two responder samples, less than 25% of tissue-infiltrating T cells carried a clonally expanded

19   TCR (Fig 5C). We then analyzed the expression of genes that we had previously established to be

20   associated with blood-tumor sharing or tumor-exclusivity in the tissue (Fig 4A, Table S5). Genes

21   positively correlated with blood-tumor sharing were highly expressed in the clusters dominated by

22   clonally expanded tissue-infiltrating T cells (Fig 5D-E). These included cytotoxic effector

23   molecules and receptors (*GNLY*, *GZMH*, *GZMM*, *KLRD1*). Contrary to patients with recurrent

1 disease, clonally expanded tissue-infiltrating T cells did not express high levels of co-inhibitory

2 receptors (*TIGIT, HAVCR2, PDCD1, CTLA4*) and other markers of exhaustion (*TOX, TOX2,*

3 *CXCL13*) that we demonstrated to be negatively correlated with the blood-tumor sharing score

4 (Fig 5D-E). Overall, these results suggest that genes positively correlated with the degree of blood-

5 tumor sharing mark T cells that are involved in tumor rejection in response to immunotherapy.

6   Recent development in single-cell RNAseq technology has significantly improved our

7 understanding of the relationship between TCR identity and functional properties. A number of

8 studies have analyzed the overlap between TCR repertoires of discrete TIL clusters to examine

9 developmental continuity across functional states (Tirosh et al., 2016; Li et al., 2019; Azizi et al.,

10 2018). Here, we show that the degree of TCR overlap between the distinct tumor and blood tissue

11 compartments can be indicative of the functional state of TILs. Specifically, we show that tumor-

12 derived T cells belonging to TCR clonal groups highly shared across tumor and blood display the

13 highest cytotoxic signatures in the tumor. Of note, circulating sister clones of these TILs also

14 display cytotoxic functions proportionally to what is observed in the tumor, suggesting that

15 analysis of circulating TILs could be used to predict effector functions in the tumor. Previous work

16 using similar approaches in various cancers indicate that cytotoxicity represents one of two main

17 developmental states of CD8 TILs, the other being T cell exhaustion (Zheng et al., 2017; Guo et

18 al., 2018; Zhang et al., 2018; Azizi et al., 2018). These two phenotypes are conceptualized as the

19 extremes in a bifurcation developmental model, where T cells belonging to the same clonal group

20 tend to follow one or the other differentiation pathway. We consistently observe that cytotoxic and

21 exhaustion phenotypes were also dichotomously associated with the pattern of TCR sharing across

22 blood and tumor; while cytotoxic sister clones can be found in both tumor and blood, the most

23 exhausted clonal groups appear to be tumor exclusive.

1   That the pattern of TCR sharing recapitulates CD8 T cell differentiation has a number of

2   implications. In regard to the natural history of the antigen-specific response in the tumor, our data

3   suggest that an exhausted phenotype is more likely to arise in T cells that have been primed locally

4   in the tumor compared to central memory T cells seeded from the periphery. Indeed, tumor

5   microenvironments contain antigen presenting cells that are unfit for priming T cell responses,

6   partly as a consequence of interactions with regulatory T cells (Wolf et al., 2015; Magnuson et al.,

7   2018).

8   The most exhausted TILs belong to tumor-exclusive clonal groups, implying that features

9   of exhaustion cannot be inferred by analyzing circulating TILs. Nevertheless, it is known that

10   tumor-resident exhausted T cells have only modest potential for functional reinvigoration, partly

11   due to hard-wired epigenetic changes (Jadhav et al., 2019; Ghoneim et al., 2017; Pauken et al.,

12   2016). Recent work has established that in metastatic melanoma patients, clinical response to

13   checkpoint inhibition immunotherapy is associated with peripheral expansion of large clonal

14   groups expressing high levels of cytotoxic and effector markers, with strong overlap with genes

15   that we found to be associated with a high degree of blood-tumor sharing (Fairfax et al., 2020).

16   This observation would suggest that circulating TILs contribute to response to immunotherapy,

17   although the specificity of this population for tumor antigens remains to be demonstrated.

18   Moreover, a study of basal cell carcinoma lesions sampled before and after anti-PD1 therapy found

19   that pre-existing exhausted TILs were largely replaced by novel expanded clonotypes, reinforcing

20   the idea that exhausted clonal groups might be short-lived and unlikely to participate in long-term

21   immune surveillance (Yost et al., 2019). We report that gene associated with high degree of blood-

22   tumor sharing are enriched for effector signatures. Consistently, a recent meta-analysis of four

23   clinical trials of an anti-PD-L1 antibody (atezolizumab) in a variety of solid tumors found that in

1   the tumor, gene signatures associated with clonotypes that were expanded in both tumor and

2   normal adjacent tissue predicted a better response to anti-PD-L1 (Wu et al., 2020).

3   As we wished to examine the circulating TILs in a variety of clinical settings, we did not

4   select patients by prior exposure to immunotherapy or variability in clinical presentation. We thus

5   were able to focus on findings that were reproducible in the patients analyzed, with the goal that

6   these findings could be further corroborated in larger, uniformly treated patient cohorts. While we

7   have further shown that it could be possible to enrich for circulating TILs by analyzing co-

8   expression of genes encoding for surface markers, critical validation steps are required for

9   implementation into a scalable immune monitoring assay. Nevertheless, we note that, in a

10   companion paper by Pauken et al., circulating TILs could be identified by the expression of a set

11   of gene markers partially overlapping with the ones that we found. The authors describe a

12   combination of four negation markers (*CCR7*, *FLT3LG*, *LTB*, and *GYPC*) that allows for

13   enrichment of circulating TILs in treatment naïve metastatic melanoma patients. Indeed, three of

14   these markers were also highly significantly associated with circulating TILs in our dataset in the

15   two patients with the highest frequency of circulating TILs. Therefore, it appears that the approach

16   of using the TCR as a molecular barcode for circulating TILs followed by an unbiased search for

17   candidate markers can return the same genes across distinct patient cohorts and independent

18   laboratories. Finally, we could not assess the directionality of circulating TILs with respect to their

19   tumor-resident sister clones to determine whether circulating TILs include cells that have left the

20   tumor bed and re-entered the circulation. This possibility has been described for tissue-resident

21   memory T cells in the skin: a subset of these cells can enter the circulation, where they remain

22   transcriptionally representative of the skin-resident population (Klicznik et al., 2019). Additional

23   work is required to assess if a similar phenomenon also occurs in cancer, and whether related

1 clonal T cells can sequentially patrol multiple metastases, including tissues to which immune cell

2 trafficking is tightly regulated, such as the central nervous system.

3   In conclusion, we propose that in human cancer the pattern of TCR repertoire sharing

4 across tissues recapitulates T cell differentiation states. Moreover, we report that it is possible to

5 assess the ongoing immune response in the tumor by analyzing a circulating T cell population.

6 Further work is necessary to better define circulating TILs as blood-based biomarkers to increase

7 the precision of existing 'liquid biopsy' assays.

8

14

1 **Author contributions**

2 L.E.L. planned and performed experiments, analyzed data and wrote the manuscript with inputs

3 from P-P. A., B.L., B.H., J.S., M.S., H.K. and D.A.H; P-P. A. and B.L. analyzed data; B.H.

4 performed experiments, L.Z. and K.R. provided technical help with single-cell genomics; L.Z.

5 provided technical help with sample acquisition; K.O. and J.C. harvested patient tissue samples,

6 S.J. collected and reviewed patient clinical information; M.S. oversaw data analysis; L.E.L., M.S.,

7 H.K. and D.A.H. conceptualized the study; H.K. and D.A.H. planned and supervised the study and

8 acquired funding.

20 **Conflicts of interest**

21 Though we believe that none of these relationships are conflicts of interest, Dr. Hafler has received

22 research funding from Bristol-Myers Squibb, Sanofi, and Genentech for work unrelated to this

23 project. He has been a consultant over the past ten years for Bristol Myers Squibb, Compass

1   Therapeutics, EMD Serono, Genentech, Juno therapeutics, Novartis Pharmaceuticals, Proclara

2   Biosciences, Sage Therapeutics, and Sanofi Genzyme. Dr. Kluger has received research funding

3   from Bristol-Myers Squibb, Apexigen and Merck. She has been a consultant for Corvus, Biodesix,

4   Roche-Genentech, Pfizer, Iovance, Immunocore, Celldex, Array Biopharma, Merck, Elevate Bio

5   and Instil Bio.

6

7

1 **Materials and Methods**

2 *Human subjects*

3 This study was approved by the Yale University Institutional Review Boards. All participants

4 provided written informed consent for the study. All patients had histologically confirmed Stage

5 IV metastatic melanoma. Surgical resection was clinically indicated in all cases. Patients donated

6 a portion of surgically resected tissue (for TIL isolation) and a time-matched blood sample (for

7 PBMC isolation).

8 *Tumor dissociation and PBMC isolation*

9 Freshly resected tumor samples were mechanically dissociated and digested in HBSS medium with

10 Collagenase IV (2.5 mg/mL) and DNAse I (0.2 mg/mL) (Worthington Biochemical Corporation)

11 at 37°C for 30 minutes. TIL-enriched single cell suspensions were then isolated using Lymphoprep

12 gradient centrifugation. PBMCs were isolated from whole blood using Lymphoprep gradient

13 centrifugation. For sample Mel-T-9, TILs and PBMCs were cryopreserved in GemCell human AB

14 serum with 10% DMSO in liquid nitrogen. For sample Mel-T-11, cryopreserved TILs and PBMCs

15 were kindly provided by the Advanced Cell Therapy Laboratory in the Department of Laboratory

16 Medicine (Yale School of Medicine).

17 *Cell sorting*

18 All samples but Mel-T-9 and Mel-T-11 were sorted directly after PBMC and TIL isolation. For

19 sample Mel-T-9 and Mel-T-11, cryopreserved TIL and PBMC aliquots were thawed following

20 the protocol CG00039 from 10X Genomics (https://support.10xgenomics.com/). Samples were

21 sorted on a BD FACS Aria II using a live cell (Live/Dead Cell Viability Assay, Life

22 Technologies), CD45+ (eBioscience, PerCP-Cy5.5 conjugated), TCRαβ+(eBioscience, PE-Cy7

17

1 conjugated) gating strategy. At least 5,000 live T cells were sorted per sample for inclusion in the

2 10X library preparation.

3 *Preparation of scRNAseq and scTCRseq libraries*

4 Sorted T cell samples were centrifuged at 300 rcf for 10 minutes at 4°C, resuspended in PBS with

5 0.04% BSA and counted on a hemocytometer. Cell concentrations were adjusted to be between

6 700 and 1,000 cells/µL. Paired single-cell gene expression and TCR libraries were prepared and

7 sequenced at the Yale Center for Genome Analysis following standard protocols from 10X

8 genomics   (https://medicine.yale.edu/keck/ycga/sequencing/10x/singcellsequencing/).   Briefly,

9 single cells were isolated in 1 nanoliter GEMs (Gel Bead in Emulsion), using the GemCode

10 technology. Depending on the initial cellularity of the sorted samples, between 5,000 and 10,000

11 cells were targeted. Cell barcoding, lysis, and reverse transcription of mRNA occurred within each

12 GEM. cDNA libraries were then generated using next-generation sequencing PCR amplification

13 with the 5-prime chemistry. Both scRNAseq libraries and scTCRseq libraries were generated for

14 each sample.

15 *CITE-seq staining*

16 Sorted live/CD45+/TCRαβ+ cells derived from blood and tumor samples were incubated with 0.5

17 µg/sample of the following antibody mix: αCD4 clone RPA-T4, αCD8a clone RPA-T8, αPD-1

18 clone EH12.2H7 (all from BioLegend, conjugated with TotalSeq™ oligos) in PBS:BSA 0.1% for

19 30 minutes on ice. Stained cells were centrifuged at 300 rcf for 10 minutes at 4°C, resuspended in

20 PBS with 0.04% BSA, counted on a hemocytometer and used for preparation of paired scRNA-

21 seq, scTCR-seq and scCITE-seq libraries, as described above.

22

18

1   *Sequencing*

2   scRNAseq libraries were sequenced on an Illumina NovaSeq S4 instrument at a read length of

3   26x8x91 base pairs and at a depth of 300 million reads per sample. CITE-seq libraries were also

4   sequenced on at a read length of 26x8x91 base pairs and at a depth of 30 million reads per sample

5   scTCRseq libraries were sequenced on an Illumina NovaSeq S4 at a read length of 150x150 base

6   pairs and at a depth of 20 million reads per sample.

7   *Data processing of scRNAseq and scTCRseq libraries*

8   ScRNAseq reads were aligned to the reference genome GRCh38 using the cellranger count

9   pipeline from the CellRanger software v3.0.2 (10X Genomics). The mean number of reads per cell

10   was 49,499 (28,471 − 103,776) for blood T cell samples, and 92,357 (43,243 − 283,598) for tumor

11   T cell samples. The median number of unique molecular identifiers (UMI) per cell was 4,383

12   (2,624 − 5,820) for blood and 3,960 (2,440 − 4,836) for tumor samples. The median number of

13   unique genes per cell was 1,444 (1,105 − 1,802) for blood and 1,465 (1,060 − 1791) for tumor

14   samples. Filtered gene-barcode matrices containing barcodes that had passed the threshold for cell

15   detection were used for downstream analyses.

16   ScTCRseq reads were processed using the cellranger vdj pipeline from CellRanger v3.0.2 (10X

17   Genomics, reference genome GRCh38). The mean number of read pairs per cell was 4,792 (1,358

18   − 10,040) for blood-derived TCR libraries and 6,442 (1,765 − 27,411) for tumor-derived TCR-

19   libraries. The VDJ enrichment, defined as the percentage of reads mapping to any VDJ gene, was

20   89% ± 2.4 for blood and 90% ± 2.58 for tumor. TCR contigs called by cellranger vdj with high

21   confidence, listed in the filtered_contig_annotations.csv output file, were retained for downstream

22   analyses. Nucleotide and amino acid CDR3α and β sequences and count matrices of gene

1   expression were matched for each cell based on barcode identities using costume R scripts and

2   imported into Seurat v3.0 (www.satijalab.org/seurat/). When more than one CDR3α and β

3   sequence was found for a given cell, additional sequences were appended, while NA values were

4   assigned for missing sequences. For downstream analyses, all cells for which multiple CDR3α

5   and β sequences existed or NA values were present were removed from the dataset.

6   *Cluster analysis*

7   Cell-level quality control was performed for individual patients and for blood and tumor gene

8   expression matrices separately. Thresholds for removing low-quality cells were selected based on

9   the distribution of the number of unique genes per cell and the percentage of genes that map to the

10   mitochondrial genome. On average, cells that expressed between 258 and 3,896 unique genes and

11   with a percentage of mitochondrial reads below 13.75% were retained for downstream analyses.

12   For the cluster analysis, we filtered out T cell receptor genes and genes encoded by the Y

13   chromosome and selected for cells with *CD8A* count >= 1 and *CD3E + CD3D + CD3G* >= 1. We

14   then joined count matrices from individual patients to generate one merged Seurat object for blood-

15   derived cells and one for tumor-derived cells. Counts were log-normalized and 2,000 variable

16   genes were identified for blood and tumor separately using the Seurat function

17   FindVariableFeatures with the "vst" method (variance stabilizing transformation). Tumor and

18   blood datasets were subsequently integrated into a single dataset using the Seurat functions

19   FindIntegrationAnchors and IntegrateData using 20 dimensions. For the resulting integrated

20   object, scaled z-scores for each gene were computed using the ScaleData function and regressed

21   against sequencing batch and fresh or frozen processing. Principal component analysis was then

22   run to identify 30 principal components, of which the first 8 were selected for t-SNE

23   dimensionality reduction using the elbow method on the distribution of the standard deviation of

each PC. For clusters identification, we initially used the FindNeighbors function to build a shared nearest neighbor graph based on the k-nearest neighbor of each cell, with k = 5. Clusters are then determined with the FindClusters function, which uses the Louvain algorithm for modularity optimization (ref, resolution = 0.4). Markers of individual clusters were defined as the differentially expressed genes between cells belonging to each cluster versus every other cell using the Wilcoxon Rank Sum test as implemented in the function FindMarkers. Genes with an adjusted p value < 0.05 were retained.

TCRαβ repertoire analysis

The TCRαβ repertoires of blood and tumor samples were analyzed using scRepertoire (Borcherding et al., 2020). Filtered contig annotation files from the output of cellranger vdj were used as input. Indexes of diversity (Shannon entropy, Inverted Simpson) and species richness (Chao, ACE) were calculated using the clonalDiversity function (cloneCall = "aa") and the significance of group differences was tested with the Wilcoxon rank-sum test. Hierarchical clustering of samples based on clone size distribution was performed with the function clonesizeDistribution (cloneCall = "aa") derived from the package PowerTCR (Koch et al., 2018). The proportion of occupied repertoire space was plotted using the function clonalProportion (cloneCall = "aa") and alluvial plots were generated with compareClonotypes after selecting the top 20 most expanded clonotypes in the tumor.

To obtain a binary definition of clonal expansion, thresholds were selected by inspection of histograms displaying the percentage of the tumor TCRαβ repertoire occupied by a given clonotype on the y axis, and individual TIL clonotypes ordered by increasing percentage of the tumor TIL repertoire on the x axis to identify a knee of the curve.

1   *TIL expanded gene set analysis*

2   Gene sets of clonally expanded TILs were calculated for each individual patient. To determine

3   which genes are specifically associated with T cell expansion in the tumor we modeled raw gene

4   counts in a generalized linear model as implemented in DESeq2 (Love et al., 2014) including an

5   interaction term between the expansion and the tissue variable:

6   ~Tissue+Expansion+Tissue:Expansion. We ran our model on T cells from each patient separately,

7   using a local fit. To generate expanded TIL gene sets, we filtered genes with

8   log2FoldChange_tumor > 0 at an adjusted p value < 0.1. To generate tumor-specific gene sets, we

9   filtered genes with log2FoldChange_tumor > 0 at an adjusted p value < 0.1 and

10   log2FoldChange_interaction > 0 at an adjusted p value < 0.1. We then used the function

11   AddModuleScore with default parameters to calculate the average expression of each gene set on

12   single cells from the blood, for each patient. Blood cells were annotated based on TCR identity as

13   circulating TILs (those with a TCR found to be above the threshold for clonal expansion in the

14   tumor) and or not. Statistical significance of the difference between the means of each group of

15   cells was established for each patient using the Wilcoxon rank-sum test.

16   *COMET analysis*

17   The input of the COMET analysis consisted of log-normalized counts of CD8 T cells from the

18   blood of individual patients. Each cell was annotated as having or not a TCR clonally expanded in

19   the tumor. Expression matrices and cell identity annotations were uploaded to

20   http://www.cometsc.com/comet and comet was run with default parameters. For each patient, we

21   obtained an output file containing lists of gene pairs ranked by their power to identify circulating

22   TILs. We joined these lists by the gene pair identity and filtered out all gene pairs that were not

23   detected in at least 5 out of 7 patients. For the remaining gene pairs, we calculated an average rank

1   and ordered them by decreasing values of the average rank, the highest of which corresponded to

2   the gene pair *KLRD1-CD74*.

3   *Expansion and blood-tumor sharing score*

4   At the clonotype level, the expansion score (Fig 4B) was defined as follows. For each clonotype,

5   we calculated the percentage of the total tumor T cell infiltrate occupied by cells sharing that same

6   clonotype (clonal groups). The patient-level expansion score, as displayed in Fig 4C, was

7   calculated by adding up the expansion scores of individual clonal groups that were identified as

8   clonally expanded as exemplified in Fig 1A. Separate patient-level expansion scores were

9   calculated for blood and tumor. The sharing score of individual tumor-expanded TCRs (Fig 4A)

10   was defined as the percentage of cells in each clonal group that were also found in blood. For

11   blood-expanded TCRs, the sharing score was a measure of the frequency of cells with that TCRs

12   that were also found in the tumor infiltrate. At the patient-level, the degree of sharing was

13   calculated by averaging the sharing score of individual clonal groups that were identified as

14   clonally expanded as exemplified in Fig 1A. Separate patient-level sharing scores were calculated

15   for blood and tumor.

16   *Exhaustion and cytotoxicity gene signatures*

17   Consensus markers of exhaustion and cytotoxicity were obtained from a recent review of CD8

18   states in human cancer (van der Leun et al., 2020). Specifically, the exhaustion gene set included

19   *HAVCR2, IFNG, ITGAE, PDCD1, CXCL13, LAYN, LAG3, TIGIT, CTLA4, ENTPD1*, while the

20   cytotoxicity gene set consisted of *CX3CR1, PRF1, GZMA, GZMH, GNLY, FGFBP2, KLRG1,*

21   *GZMM, LYAR, TXNIP, FCRL6, NKG7, KLRD1*. We then used the function AddModuleScore with

22   default parameters to calculate the average expression of each gene set on single cells. For each

1   patient, we averaged the exhaustion and cytotoxicity scores of all tumor-infiltrating CD8 T cells

2   and circulating TILs.

3

4

5

6   **Supplementary Materials**

7   Fig. S1 Individual thresholds of clonal expansion in the tumor.

8   Fig. S2 Individual thresholds of clonal expansion in the blood.

9   Fig. S3 Co-expression of KLRD1 and CD74 enriches for circulating TILs.

10   Table S1 Top 10 marker genes for each cluster, with percentage of positive cells in the cluster of

11   interest, delta percentage of positive cells between that cluster and the rest of the dataset,

12   and adjusted p-value.

13   Table S2 List of gene marker pairs of circulating TILs found in at least five out of seven patients,

14   with rank and q value for each patient.

15   Table S3 Expanded TIL gene sets and tumor-specific gene sets for each patient.

16   Table S4 List of correlation coefficients between genes and expansion score, filtered for -LogP >=

17   100. Same for sharing score.

18   Table S5 Amino acid and nucleotide sequences of the TCR$\alpha\beta$ CDR3 regions of clonotypes

19   displayed in Figure 4.

20

References

Azizi, E., A.J. Carr, G. Plitas, A.E. Cornish, C. Konopacki, S. Prabhakaran, J. Nainys, K. Wu, V. Kiseliovas, M. Setty, K. Choi, R.M. Fromme, P. Dao, P.T. McKenney, R.C. Wasti, K. Kadaveru, L. Mazutis, A.Y. Rudensky, and D. Pe'er. 2018. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell*. 174:1293-1308.e36. doi:10.1016/j.cell.2018.05.060.

Barnes, T.A., and E. Amir. 2017. HYPE or HOPE: the prognostic value of infiltrating immune cells in cancer. *Br. J. Cancer*. 117:451–460. doi:10.1038/bjc.2017.220.

Blank, C.U., W.N. Haining, W. Held, P.G. Hogan, A. Kallies, E. Lugli, R.C. Lynn, M. Philip, A. Rao, N.P. Restifo, A. Schietinger, T.N. Schumacher, P.L. Schwartzberg, A.H. Sharpe, D.E. Speiser, E.J. Wherry, B.A. Youngblood, and D. Zehn. 2019. Defining "T cell exhaustion." *Nat. Rev. Immunol.* 19:665–674. doi:10.1038/s41577-019-0221-9.

Borcherding, N., N.L. Bormann, and G. Kraus. 2020. scRepertoire: An R-based toolkit for single-cell immune receptor analysis. *F1000Research*. 9:47. doi:10.12688/f1000research.22139.2.

Breart, B., F. Lemaître, S. Celli, and P. Bousso. 2008. Two-photon imaging of intratumoral CD8+ T cell cytotoxic activity during adoptive T cell therapy in mice. *J. Clin. Invest.* 118:1390–1397. doi:10.1172/JCI34388.

Chao, A., S.M. Lee, and S.L. Jeng. 1992. Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal. *Biometrics*. 48:201–216.

Chihara, N., A. Madi, T. Kondo, H. Zhang, N. Acharya, M. Singer, J. Nyman, N.D. Marjanovic, M.S. Kowalczyk, C. Wang, S. Kurtulus, T. Law, Y. Etminan, J. Nevin, C.D. Buckley, P.R. Burkett, J.D. Buenrostro, O. Rozenblatt-Rosen, A.C. Anderson, A. Regev, and V.K. Kuchroo. 2018. Induction and transcriptional regulation of the co-inhibitory gene module in T cells. *Nature*. 558:454–459. doi:10.1038/s41586-018-0206-z.

Cohen, J.V., and E.I. Buchbinder. 2019. The Evolution of Adjuvant Therapy for Melanoma. *Curr Oncol Rep*. 21:106. doi:10.1007/s11912-019-0858-3.

Delaney, C., A. Schnell, L.V. Cammarata, A. Yao-Smith, A. Regev, V.K. Kuchroo, and M. Singer. 2019. Combinatorial prediction of gene-marker panels from single-cell transcriptomic data. *bioRxiv*. 655753. doi:10.1101/655753.

Dysfunctional CD8 T Cells Form a Proliferative, Dynamically Regulated Compartment within Human Melanoma. *Cell*. 176:775-789.e18. doi:10.1016/j.cell.2018.11.043.

Eisenhauer, E.A., P. Therasse, J. Bogaerts, L.H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe, and J. Verweij. 2009. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer*. 45:228–247. doi:10.1016/j.ejca.2008.10.026.

Fairfax, B.P., C.A. Taylor, R.A. Watson, I. Nassiri, S. Danielli, H. Fang, E.A. Mahé, R. Cooper, V. Woodcock, Z. Traill, M.H. Al-Mossawi, J.C. Knight, P. Klenerman, M. Payne, and M.R. Middleton. 2020. Peripheral CD8+ T cell characteristics associated with durable responses to immune checkpoint blockade in patients with metastatic melanoma. *Nat. Med.* 26:193–199. doi:10.1038/s41591-019-0734-6.

Gershenwald, J.E., R.A. Scolyer, K.R. Hess, V.K. Sondak, G.V. Long, M.I. Ross, A.J. Lazar, M.B. Faries, J.M. Kirkwood, G.A. McArthur, L.E. Haydu, A.M.M. Eggermont, K.T. Flaherty, C.M. Balch, J.F. Thompson, and for members of the American Joint Committee on Cancer Melanoma Expert Panel and the International Melanoma Database and Discovery Platform. 2017. Melanoma staging: Evidence-based changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin*. 67:472–492. doi:10.3322/caac.21409.

Ghoneim, H.E., Y. Fan, A. Moustaki, H.A. Abdelsamed, P. Dash, P. Dogra, R. Carter, W. Awad, G. Neale, P.G. Thomas, and B. Youngblood. 2017. De Novo Epigenetic Programs Inhibit PD-1 Blockade-Mediated T Cell Rejuvenation. *Cell*. 170:142-157.e19. doi:10.1016/j.cell.2017.06.007. Guo, X., Y. Zhang, L. Zheng, C. Zheng, J. Song, Q. Zhang, B. Kang, Z. Liu, L. Jin, R. Xing, R. Gao, L. Zhang, M. Dong, X. Hu, X. Ren, D. Kirchhoff, H.G. Roider, T. Yan, and Z. Zhang. 2018. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat. Med.* 24:978–985. doi:10.1038/s41591-018-0045-3.

Jadhav, R.R., S.J. Im, B. Hu, M. Hashimoto, P. Li, J.-X. Lin, W.J. Leonard, W.J. Greenleaf, R. Ahmed, and J.J. Goronzy. 2019. Epigenetic signature of PD-1+ TCF1+ CD8 T cells that act as resource cells during chronic viral infection and respond to PD-1 blockade. *Proc. Natl. Acad. Sci. U.S.A.* 116:14113–14118. doi:10.1073/pnas.1903520116.

Kalaora, S., Y. Wolf, T. Feferman, E. Barnea, E. Greenstein, D. Reshef, I. Tirosh, A. Reuben, S. Patkar, R. Levy, J. Quinkhardt, T. Omokoko, N. Qutob, O. Golani, J. Zhang, X. Mao, X. Song, C. Bernatchez, C. Haymaker, M.-A. Forget, C. Creasy, P. Greenberg, B.W. Carter, Z.A. Cooper, S.A. Rosenberg, M. Lotem, U. Sahin, G. Shakhar, E. Ruppin, J.A. Wargo, N. Friedman, A. Admon, and Y. Samuels. 2018. Combined Analysis of Antigen Presentation and T-cell Recognition Reveals Restricted Immune Responses in Melanoma. *Cancer Discov*. 8:1366–1375. doi:10.1158/2159-8290.CD-17-1418.

Klicznik, M.M., P.A. Morawski, B. Höllbacher, S.R. Varkhande, S.J. Motley, L. Kuri-Cervantes, E. Goodwin, M.D. Rosenblum, S.A. Long, G. Brachtl, T. Duhen, M.R. Betts, D.J. Campbell, and I.K. Gratz. 2019. Human CD4+CD103+ cutaneous resident memory T cells are found in the circulation of healthy individuals. *Sci Immunol*. 4. doi:10.1126/sciimmunol.aav8995.

Koch, H., D. Starenki, S.J. Cooper, R.M. Myers, and Q. Li. 2018. powerTCR: A model-based approach to comparative analysis of the clone size distribution of the T cell receptor repertoire. *PLOS Computational Biology*. 14:e1006571. doi:10.1371/journal.pcbi.1006571.

Kumar, B.V., W. Ma, M. Miron, T. Granot, R.S. Guyer, D.J. Carpenter, T. Senda, X. Sun, S.-H. Ho, H. Lerner, A.L. Friedman, Y. Shen, and D.L. Farber. 2017. Human Tissue-Resident Memory T Cells Are Defined by Core Transcriptional and Functional Signatures in Lymphoid and Mucosal Sites. *Cell Rep*. 20:2921–2934. doi:10.1016/j.celrep.2017.08.078.

Laydon, D.J., A. Melamed, A. Sim, N.A. Gillet, K. Sim, S. Darko, J.S. Kroll, D.C. Douek, D.A. Price, C.R.M. Bangham, and B. Asquith. 2014. Quantification of HTLV-1 clonality and TCR diversity. *PLoS Comput. Biol.* 10:e1003646. doi:10.1371/journal.pcbi.1003646.

van der Leun, A.M., D.S. Thommen, and T.N. Schumacher. 2020. CD8+ T cell states in human cancer: insights from single-cell analysis. *Nat. Rev. Cancer*. 20:218–232. doi:10.1038/s41568-019-0235-4.

Li, H., A.M. van der Leun, I. Yofe, Y. Lubling, D. Gelbard-Solodkin, A.C.J. van Akkooi, M. van den Braber, E.A. Rozeman, J.B.A.G. Haanen, C.U. Blank, H.M. Horlings, E. David, Y. Baran, A. Bercovich, A. Lifshitz, T.N. Schumacher, A. Tanay, and I. Amit. 2019. Dysfunctional CD8 T Cells Form a Proliferative, Dynamically Regulated Compartment within Human Melanoma. *Cell*. 176:775-789.e18. doi:10.1016/j.cell.2018.11.043.

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

Magnuson, A.M., E. Kiner, A. Ergun, J.S. Park, N. Asinovski, A. Ortiz-Lopez, A. Kilcoyne, E. Paoluzzi-Tomada, R. Weissleder, D. Mathis, and C. Benoist. 2018. Identification and validation of a tumor-infiltrating Treg transcriptional signature conserved across species and tumor types. *Proc Natl Acad Sci U S A*. 115:E10672–E10681. doi:10.1073/pnas.1810580115.

Moskophidis, D., F. Lechner, H. Pircher, and R.M. Zinkernagel. 1993. Virus persistence in acutely infected immunocompetent mice by exhaustion of antiviral cytotoxic effector T cells. *Nature*. 362:758–761. doi:10.1038/362758a0.

Pauken, K.E., M.A. Sammons, P.M. Odorizzi, S. Manne, J. Godec, O. Khan, A.M. Drake, Z. Chen, D.R. Sen, M. Kurachi, R.A. Barnitz, C. Bartman, B. Bengsch, A.C. Huang, J.M. Schenkel, G. Vahedi, W.N. Haining, S.L. Berger, and E.J. Wherry. 2016. Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. *Science*. 354:1160–1165. doi:10.1126/science.aaf2807.

Ribas, A., and J.D. Wolchok. 2018. Cancer immunotherapy using checkpoint blockade. *Science*. 359:1350–1355. doi:10.1126/science.aar4060.

Sade-Feldman, M., K. Yizhak, S.L. Bjorgaard, J.P. Ray, C.G. de Boer, R.W. Jenkins, D.J. Lieb, J.H. Chen, D.T. Frederick, M. Barzily-Rokni, S.S. Freeman, A. Reuben, P.J. Hoover, A.-C. Villani, E. Ivanova, A. Portell, P.H. Lizotte, A.R. Aref, J.-P. Eliane, M.R. Hammond, H. Vitzthum, S.M. Blackmon, B. Li, V. Gopalakrishnan, S.M. Reddy, Z.A. Cooper, C.P. Paweletz, D.A. Barbie, A. Stemmer-Rachamimov, K.T. Flaherty, J.A. Wargo, G.M. Boland, R.J. Sullivan, G. Getz, and N. Hacohen. 2018. Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma. *Cell*. 175:998-1013.e20. doi:10.1016/j.cell.2018.10.038.

Schietinger, A., A. Arina, R.B. Liu, S. Wells, J. Huang, B. Engels, V. Bindokas, T. Bartkowiak, D. Lee, A. Herrmann, D.W. Piston, M.J. Pittet, P.C. Lin, T. Zal, and H. Schreiber. 2013. Longitudinal confocal microscopy imaging of solid tumor destruction following adoptive T cell transfer. *Oncoimmunology*. 2:e26677. doi:10.4161/onci.26677.

Stuart, T., A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W.M. Mauck, Y. Hao, M. Stoeckius, P. Smibert, and R. Satija. 2019. Comprehensive Integration of Single-Cell Data. *Cell*. 177:1888-1902.e21. doi:10.1016/j.cell.2019.05.031.

Tirosh, I., B. Izar, S.M. Prakadan, M.H. Wadsworth, D. Treacy, J.J. Trombetta, A. Rotem, C. Rodman, C. Lian, G. Murphy, M. Fallahi-Sichani, K. Dutton-Regester, J.-R. Lin, O. Cohen, P. Shah, D. Lu, A.S. Genshaft, T.K. Hughes, C.G.K. Ziegler, S.W. Kazer, A. Gaillard, K.E. Kolb, A.-C. Villani, C.M. Johannessen, A.Y. Andreev, E.M. Van Allen, M. Bertagnolli, P.K. Sorger, R.J. Sullivan, K.T. Flaherty, D.T. Frederick, J. Jané-Valbuena, C.H. Yoon, O. Rozenblatt-Rosen, A.K. Shalek, A. Regev, and L.A. Garraway. 2016. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*. 352:189–196. doi:10.1126/science.aad0501.

Wang, J., C.J. Perry, K. Meeth, D. Thakral, W. Damsky, G. Micevic, S. Kaech, K. Blenman, and M. Bosenberg. 2017. UV-induced somatic mutations elicit a functional T cell response in the YUMMER1.7 mouse melanoma model. *Pigment Cell Melanoma Res*. 30:428–435. doi:10.1111/pcmr.12591.

Wolf, D., S. Sopper, A. Pircher, G. Gastl, and A.M. Wolf. 2015. Treg(s) in Cancer: Friends or Foe? *J. Cell. Physiol.* 230:2598–2605. doi:10.1002/jcp.25016.

Wu, T.D., S. Madireddi, P.E. de Almeida, R. Banchereau, Y.-J.J. Chen, A.S. Chitre, E.Y. Chiang, H. Iftikhar, W.E. O'Gorman, A. Au-Yeung, C. Takahashi, L.D. Goldstein, C. Poon, S. Keerthivasan, D.E. de Almeida Nagata, X. Du, H.-M. Lee, K.L. Banta, S. Mariathasan, M. Das Thakur, M.A. Huseni, M. Ballinger, I. Estay, P. Caplazi, Z. Modrusan, L. Delamarre, I. Mellman, R. Bourgon, and J.L. Grogan. 2020. Peripheral T cell expansion predicts tumour infiltration and clinical response. *Nature*. 579:274–278. doi:10.1038/s41586-020-2056-8.

Yost, K.E., A.T. Satpathy, D.K. Wells, Y. Qi, C. Wang, R. Kageyama, K.L. McNamara, J.M. Granja, K.Y. Sarin, R.A. Brown, R.K. Gupta, C. Curtis, S.L. Bucktrout, M.M. Davis, A.L.S. Chang, and H.Y. Chang. 2019. Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nat. Med.* 25:1251–1259. doi:10.1038/s41591-019-0522-3.

Zhang, F., K. Wei, K. Slowikowski, C.Y. Fonseka, D.A. Rao, S. Kelly, S.M. Goodman, D. Tabechian, L.B. Hughes, K. Salomon-Escoto, G.F.M. Watts, A.H. Jonsson, J. Rangel-Moreno, N. Meednu, C. Rozo, W. Apruzzese, T.M. Eisenhaure, D.J. Lieb, D.L. Boyle, A.M. Mandelin, Accelerating Medicines Partnership Rheumatoid Arthritis and Systemic Lupus Erythematosus (AMP RA/SLE) Consortium, B.F. Boyce, E. DiCarlo, E.M. Gravallese, P.K. Gregersen, L. Moreland, G.S. Firestein, N. Hacohen, C. Nusbaum, J.A. Lederer, H. Perlman, C. Pitzalis, A. Filer, V.M. Holers, V.P. Bykerk, L.T. Donlin, J.H. Anolik, M.B. Brenner, and S. Raychaudhuri. 2019. Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat. Immunol.* 20:928–942. doi:10.1038/s41590-019-0378-1.

Zhang, L., X. Yu, L. Zheng, Y. Zhang, Y. Li, Q. Fang, R. Gao, B. Kang, Q. Zhang, J.Y. Huang, H. Konno, X. Guo, Y. Ye, S. Gao, S. Wang, X. Hu, X. Ren, Z. Shen, W. Ouyang, and Z. Zhang. 2018. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature*. 564:268–272. doi:10.1038/s41586-018-0694-x.

Zheng, C., L. Zheng, J.-K. Yoo, H. Guo, Y. Zhang, X. Guo, B. Kang, R. Hu, J.Y. Huang, Q. Zhang, Z. Liu, M. Dong, X. Hu, W. Ouyang, J. Peng, and Z. Zhang. 2017. Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. *Cell*. 169:1342-1356.e16. doi:10.1016/j.cell.2017.05.035.

Figures and Tables

Table 1 Demographics and clinical features of patients included in the study. All patients are stage IV melanoma (Gershenwald et al., 2017)**.** *Tumor burden* is defined based on a lesion volume threshold of 2 cm^3. *Disease duration* corresponds to the time between the diagnosis of metastatic disease and sample collection. *Response* is defined according to the RECIST guidelines (v1.1, (Eisenhauer et al., 2009). N.A. = not applicable, P.D. = progressive disease, P.R. = partial response. <u>*Tumor burden* was calculated using the formula (4/3)π r3, where r is the maximum radius of each metastasis, except for three lesions for which the formula for the volume of an ellipsoid was deemed more appropriate (V=4/3π*L/2*H/2*W/2).</u>

Figure Legends

Fig. 1 Detection of clonally-expanded TILs in the circulation of melanoma patients. A, dot plots displaying indexes of diversity and species richness for tumor and blood TCRαβ repertoires of individual patients. Horizontal bars indicate medians. Statistical significance was determined using Wilcoxon rank-sum test. **B**, hierarchical clustering of tumor and blood TCRαβ repertoires based on clonal size distribution. Red = blood, Grey = tumor. **C**, bar graph displaying the proportion of TCRαβ repertoire space occupied by the top n clonotypes (bins in legend) for the tumor samples of individual patients. D, alluvial plots displaying the frequency of the top 20 most expanded TIL clonotypes in blood and tumor for individual patients. E-F, tSNE plot of TILs (**B**) and circulating T cells (**C**) divided by

expression of CD8A (normalized counts > 1) and CD4 (normalized counts > 1) for one representative patient. T cells expressing TCRs expanded in the tumor are highlighted in blue. **D**, table displaying the absolute number and frequency of tumor-resident and blood-derived CD8 T cells expressing a TCR expanded in the tumor for each patient. Bold indicates patients in which the absolute number of blood-derived T cells with a TCR expanded in the tumor (circulating TILs) was deemed large enough for subsequent analyses. **E**, percentage of CD4+ (white) and CD8+ (grey) T cells among clonally expanded T cells in tumor and circulating TILs. Each data point represents the average for one patient. Data represent mean +/- standard deviation. Statistical significance was determined using t test with Bonferroni-Dunn correction for multiple comparisons. **F** scatter plots showing the relationship of the frequency of circulating TILs and disease duration. Adjusted R-squared and F test p value are shown.

Fig. 2 Cytotoxic profile of circulating TILs. A, tSNE plot of CD8 T cells from 11 patients, split by tissue of origin and coloured by cluster. Clustering was obtained with a SNN modularity optimization-based algorithm (k parameter = 5, resolution = 0.5). **B**, tSNE plot of CD8 T cells from 11 patients, split by tissue of origin and coloured by expression of a TCR expanded in the tumor (blue, threshold for expansion as in Supplementary Fig 1), or not (gold). **C**, tSNE plots coloured by scaled expression of indicated genes. **D**, clustered heatmap diplaying the degree of overlap between lists of cluster markers (rows) and gene sets derived from published studies of melanoma TIL transcriptome (1 = Li et al. 2019, 2 = Sade-Feldman et al. 2018, 3 = Tirosh et al. 2016). **E,** heatmap of differentially expressed genes between circulating TILs and blood T cells with a TCR not expanded in the tumor with a delta % of expressing cells between the two groups of at least +/-10%. Rows

5

represent genes ordered by delta % of expressing cells. Columns represent individual cells grouped by identity class as indicated on the horizontal bars. Values are scaled and mean-centered log2-transformed gene counts. For visualization purposes, the dataset was randomly down sampled to include 1000 cells. Values are scaled and mean-centered log2-transformed gene counts.

Fig. 3 Circulating TILs reflect features of clonally expanded tumor-infiltrating T cells. A, tSNE plots highlighting tumor-infiltrating cells with a TCR clonally expanded in the tumor (dark blue) or singletons (light blue) and blood-derived cells with a TCR clonally expanded in blood (dark red) or singletons (tomato red). **B**, diagram illustrating the definition of expanded TIL gene sets and TIL-specific gene sets. **C**, clustered heatmap of the similarity matrix between expanded TIL gene sets of individual patients. **D**, quantification of the expression of expanded TIL gene sets in circulating TILs (blue) versus other circulating T cells (red) for each patient. Statistical significance was determined using Wilcoxon rank-sum test. **E**, quantification of the expression of tumor specific gene sets in circulating TILs (blue) versus other circulating T cells (red) for each patient. Statistical significance was determined using the Wilcoxon rank-sum test. **F-G** for selected genes belonging the expanded TIL gene sets of at least three patients, expression is compared between tumor-resident clonally expanded TILs, circulating TILs and the rest of blood CD8 T cells. **F**, heatmap displaying expression of expanded TIL genes differentially expressed (left panel) or not (right panel) between circulating TILs and other circulating T cells. **G** heatmap displaying expression of tumor-specific genes differentially expressed (left panel) or not (right panel) between circulating TILs and other circulating T cells. For heatmaps, values are scaled and mean-centered log2-transformed gene counts. Statistical significance was

determined with Wilcoxon rank-sum tests Bonferroni-corrected by the number of genes in the dataset (17.045) multiplied by the number of comparisons (3).

Fig. 4 Tumor-blood sharing as a predictor of cytotoxic phenotype. A, tSNE plot of tumor CD8 T cells coloured by expression of a bood-tumor sharing score (left) calculated for each clonal group in the tumor (= cells sharing the same TCR) and volcano plot representing the Spearman's rho coefficients and the adjusted p values of the correlations between sharing score and gene expression for CD8 TILs from 12 pooled patients (right). **B**, tSNE plot of tumor CD8 T cells coloured by expression of an expansion score (left) and volcano plot representing the Spearman's rho coefficients and the adjusted p values of the correlations between expansion score and gene expression for CD8 TILs from 12 pooled patients (right). **C**, scatter plots showing the relationship between expansion score in the blood (left) or in the tumor (right) and degree of sharing across organ. Each dot represents the cumulative scores for each patient. Adjusted R-squared and F test p value are shown. **D**, table summarizing the number of cells expressing selected tumor exclusive (T. e.) and shared (Sh.) clonotypes in tumor and blood for the indicated patient (top). Table summarizing the number of cells expressing selected highly expanded (Hi) and lowly expanded (Lo) tumor-exclusive clonotypes in tumor and blood for the indicated patient (bottom). The indicated cells are mapped on a tSNE plot split by organ. **E**, **F** the indicated cells are mapped on a tSNE plot split by organ. **G**, Euclidean distances between the average of the PCA embeddings of each cell in the tumor exclusive versus shared clonal groups and highly expanded versus lowly expanded clonal groups. Distances were calculated using the PCs that were determined to be statistically significant in an elbow plot (8 PCs). Statistical significance between groups was determined using the Mann-Whitney test. **H**,

heatmap displaying expression of genes differentially expressed between tumor exclusive versus shared clonal groups (rows, as in **E**). Cells are grouped and colour-coded according to the clonal group to which they belong. Differential expression was established with the Wilcoxon rank sum test, retaining genes with an adjusted p value < 0.05. Values are scaled and mean-centered log2-transformed gene counts. **I**, scatter plots showing the relationship of exhaustion signature (left) or cytotoxicity signature (right) between tumor infiltrating CD8 T cells and blood-derived circulating TILs. Adjusted R-squared and F test p value are shown.

Figure 5. Features associated with blood-tumor sharing are highly expressed in the tissue of two patients who responded to immunotherapy. A, tSNE plots of blood and tumor T cells from patient Mel-UT-1 highlighting tumor-infiltrating cells with a TCR clonally expanded in the tumor and tumor-exclusive (dark blue) or shared with blood (aqua). **B**, quantification of the expression of gene sets associated with response to immunotherapy or progression as described by (Sade-Feldman et al. Cell 2019) in tumor-resident cells indicated in **A**. Statistical significance was determined using Wilcoxon rank-sum test. **C**, top, tSNE plot of CD8 T cells from 2 patients, split by tissue of origin and coloured by cluster. Clustering was obtained with a SNN modularity optimization-based algorithm (k parameter = 5, resolution = 0.5). Bottom, tSNE plot of CD8 T cells from 2 patients, split by tissue of origin and coloured by expression of a TCR expanded in tissue (blue), or not (gold). **D**, heatmap of genes previously established to be positively and negatively correlated with TCR sharing score at an adjusted Log(p value) <= -100 (as in Fig. 5A). Rows represent genes grouped by correlation with TCR sharing. Columns represent individual cells grouped by cluster identity (horizontal bars), with clusters ordered by

decreasing percentage of blood cells. Values are scaled and mean-centered log2-transformed gene counts. Pie charts stacked on top of the grouping colour bar indicate tumor vs blood composition and expanded TIL composition for each cluster. E, tSNE plot of CD8 T cells from 2 patients, split by tissue of origin and coloured by scaled expression of the indicated genes.

Fig. S1 Individual thresholds of clonal expansion in the tumor. Bar graphs displaying the frequency of TILs expressing a given TCRαβ clonotype on the y axis, and individual TIL clonotypes ordered by increasing frequency on the x axis, for individual patients.

Fig. S2 Individual thresholds of clonal expansion in blood. Bar graphs displaying the frequency of TILs expressing a given TCRαβ clonotype on the y axis, and individual TIL clonotypes ordered by increasing frequency on the x axis, for individual patients.

Fig. S3 Co-expression of KLRD1 and CD74 enriches for circulating TILs. A, workflow for identification of gene markers of circulating TILs. Singleton and paired gene markers of circulating TILs were identified from the blood of each patient using COMET. Gene marker pairs detected in most patients were chosen and blood T cells were classified based on co-expression of the selected gene markers. The percentage of circulating TILs in cells positive and negative for the selected gene markers was compared for each patient. **B**, table summarizing the rank of the gene marker pair *KLRD1-CD74*, the q value and the total number of gene marker pairs identified by the COMET analysis. N.D. indicates that the gene marker pair *KLRD1-CD74* was not detected in that patient. N.A. = not applicable. **C**, heatmap of the reverse rank (rank * -1, rows) of singleton gene markers of tumor directed T cells that were shared by at least 5 out of 7 patients (columns), ordered by the average across patients. For pairs not detected in a specific patient, a reverse rank of -200 was

assigned. Arrows indicate *KLRD1* and *CD74*, detected by both the singleton and paired gene marker analysis. **D**, scatter plots displaying the relationship between the integrated (Stuart et al., 2019) expression of *KLRD1* and *CD74* transcripts in each patient. Dashed boxes indicate double positive cells, with expression of *KLRD1* >= 1.5 and *CD74* >= 3. **E**, percentage of cells expressing a TCR expanded in the tumor among cells that are positive or negative for the *KLRD1-CD74* combination in each patent. **F**, percentage of cells expressing a TCR expanded in the tumor among cells that are positive or negative for *PDCD1* expression in each patent. Statistical significance was determined using paired t tests. **G**, Pie charts displaying the percentage composition in terms of TCRs expanded in the tumor, shared but not expanded in the tumor, and not shared with the tumor for blood T cells classified as positive or negative for the *KLRD1-CD74* marker combination. Percentages are calculated on the sum of all blood CD8 T cells from the 7 patients.
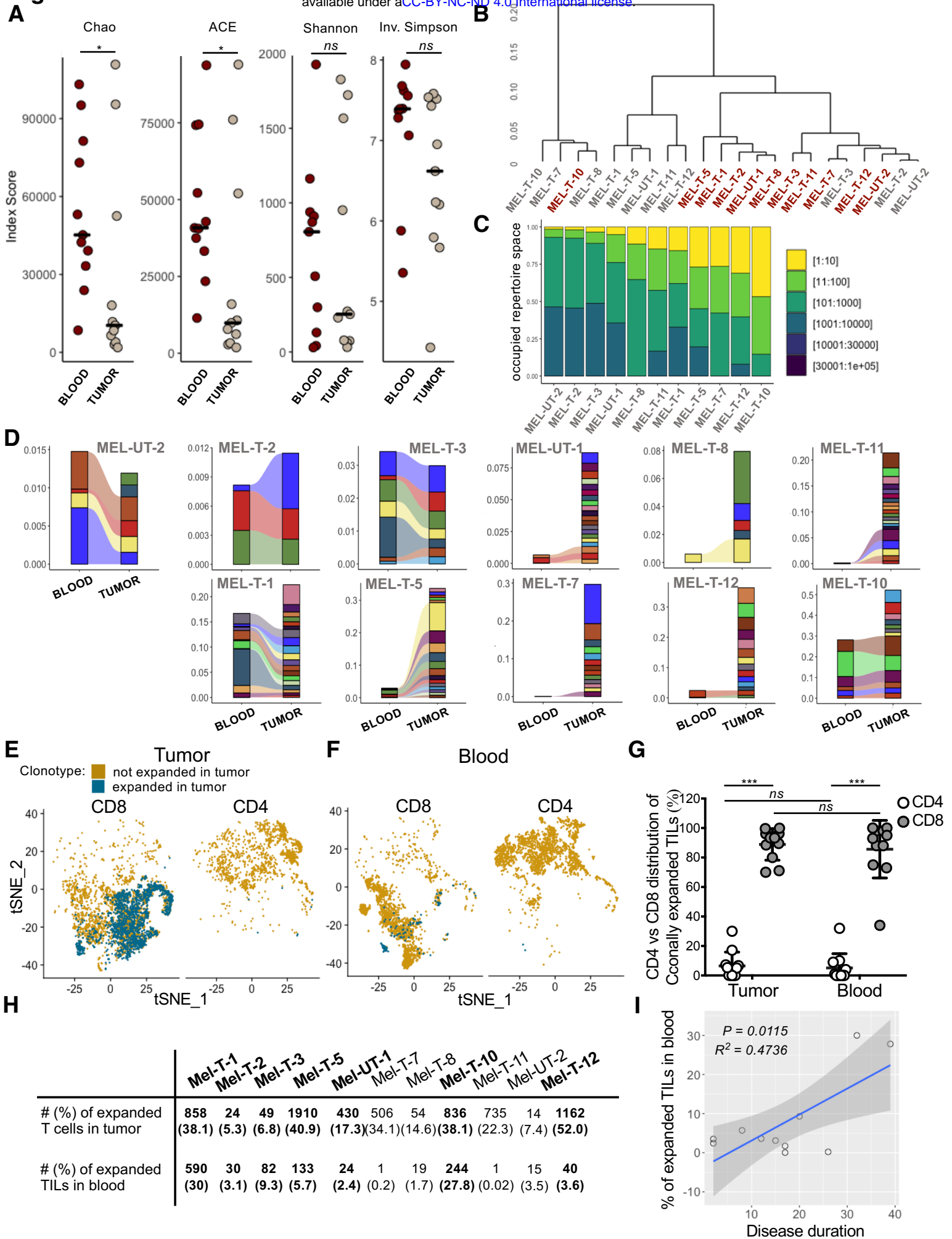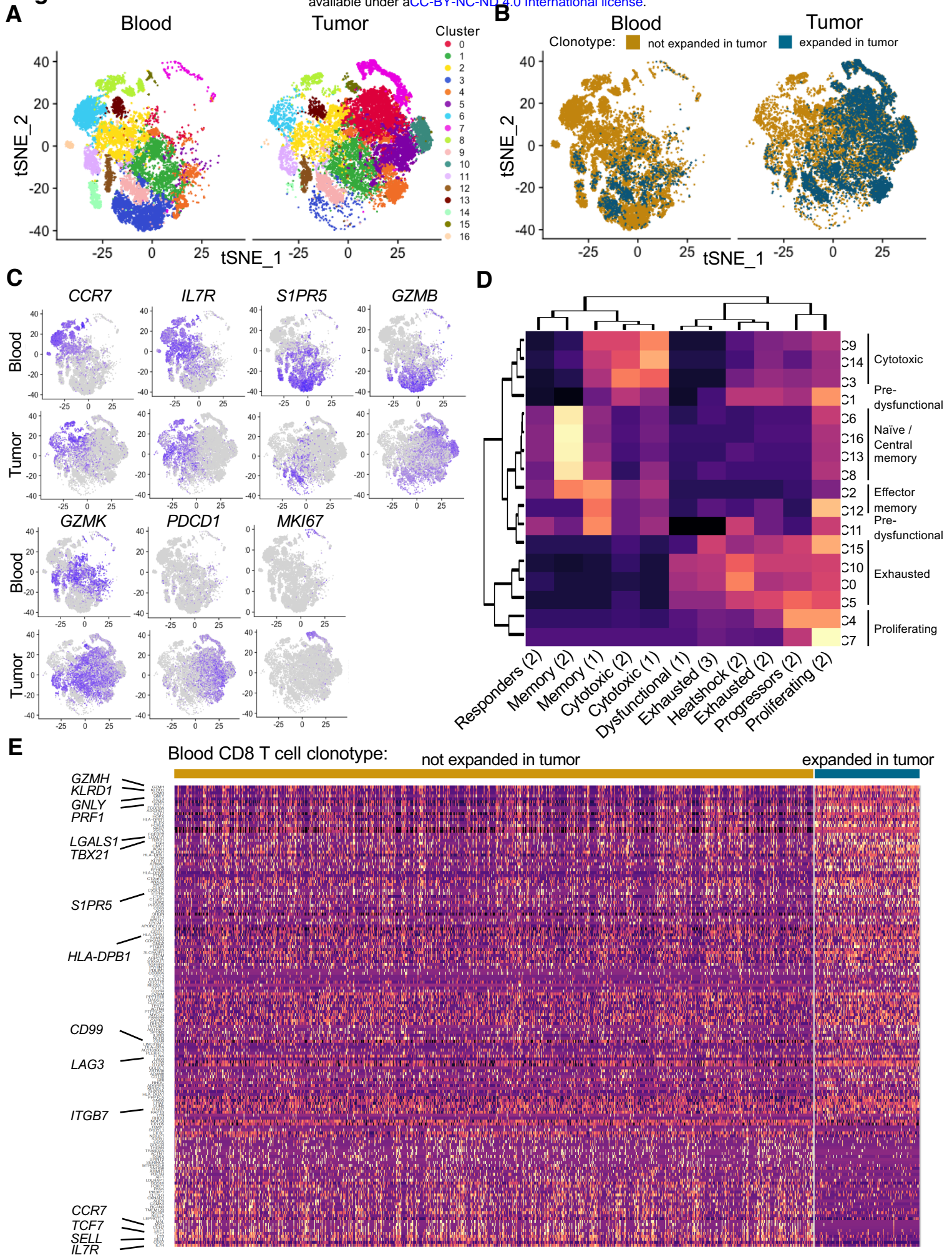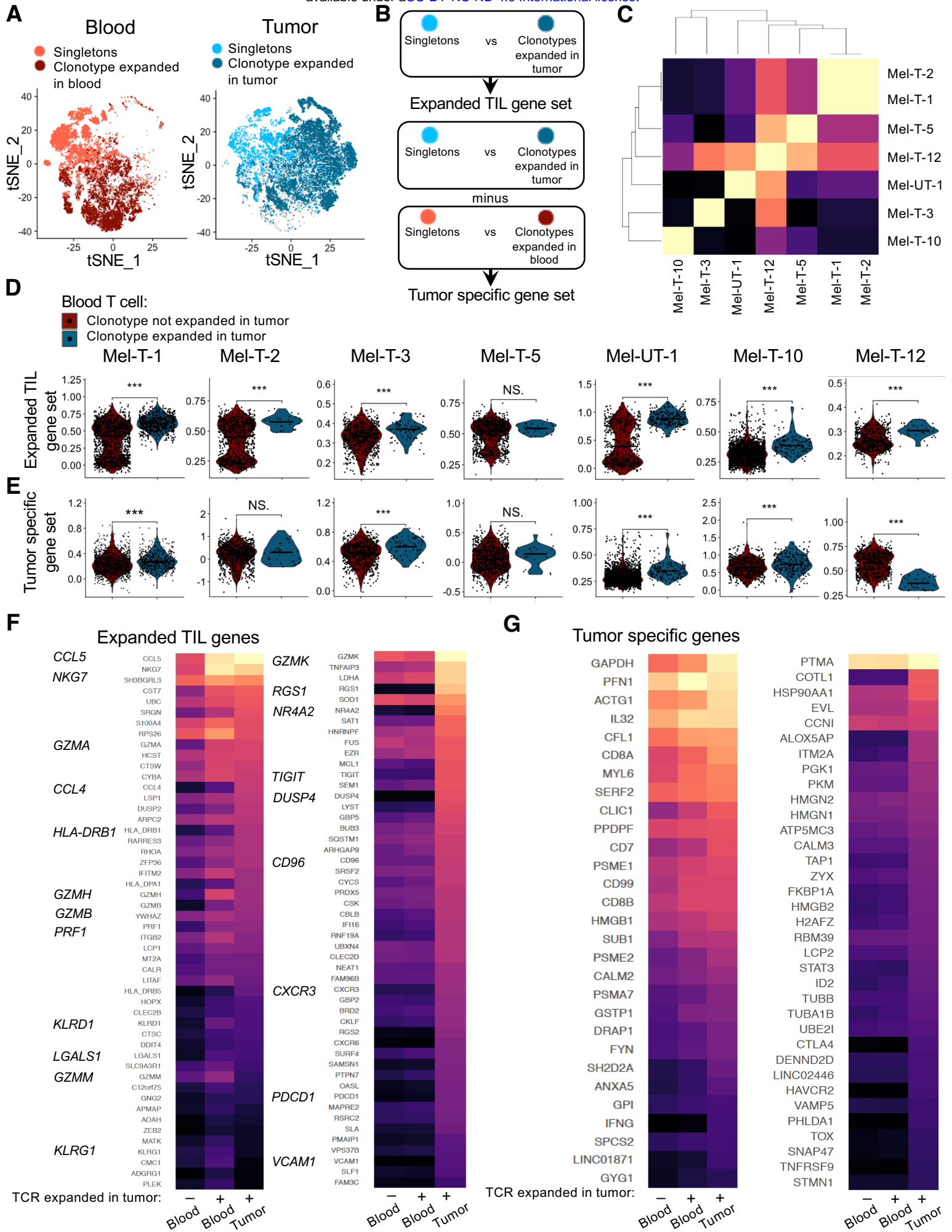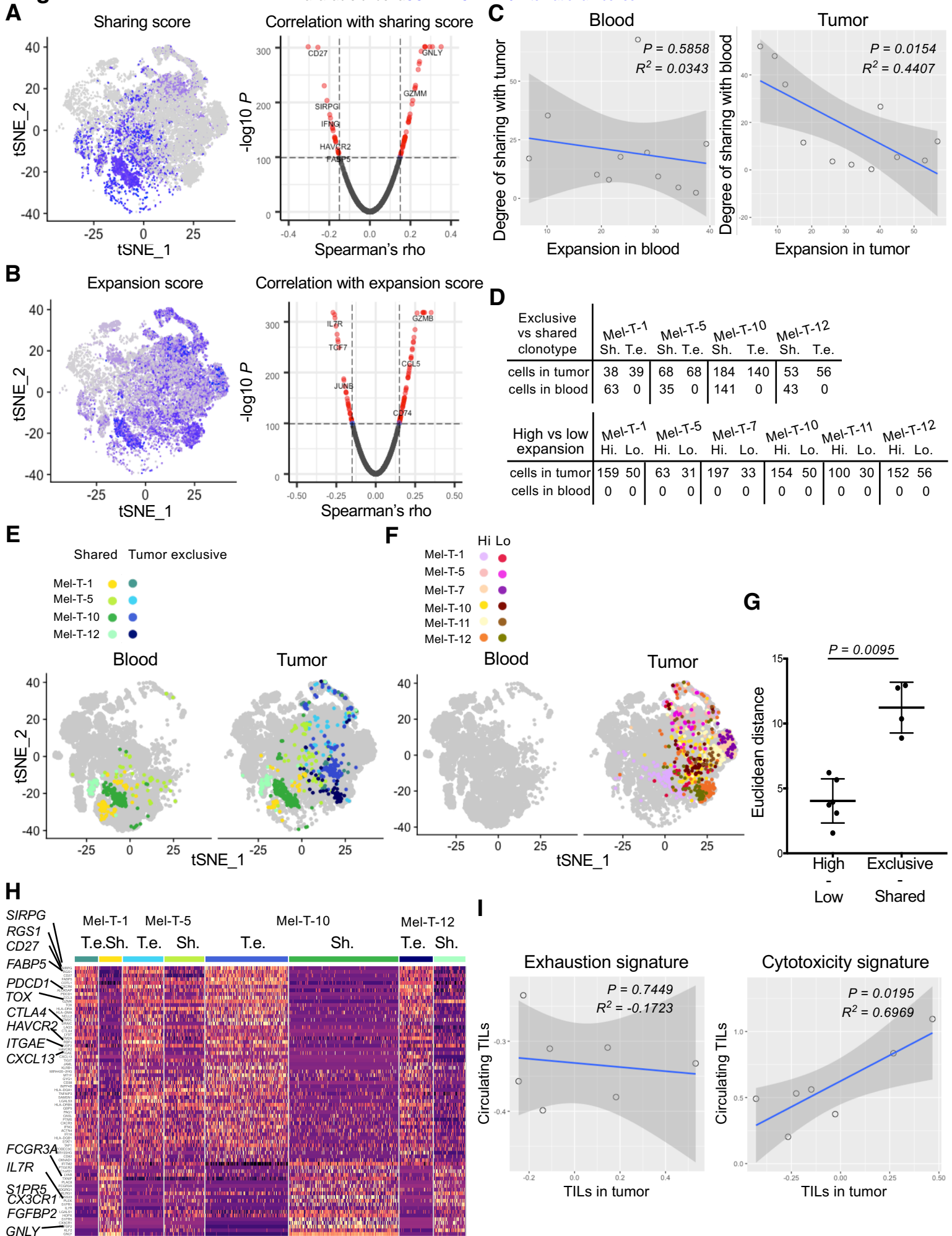
Figure 1

Figure 2
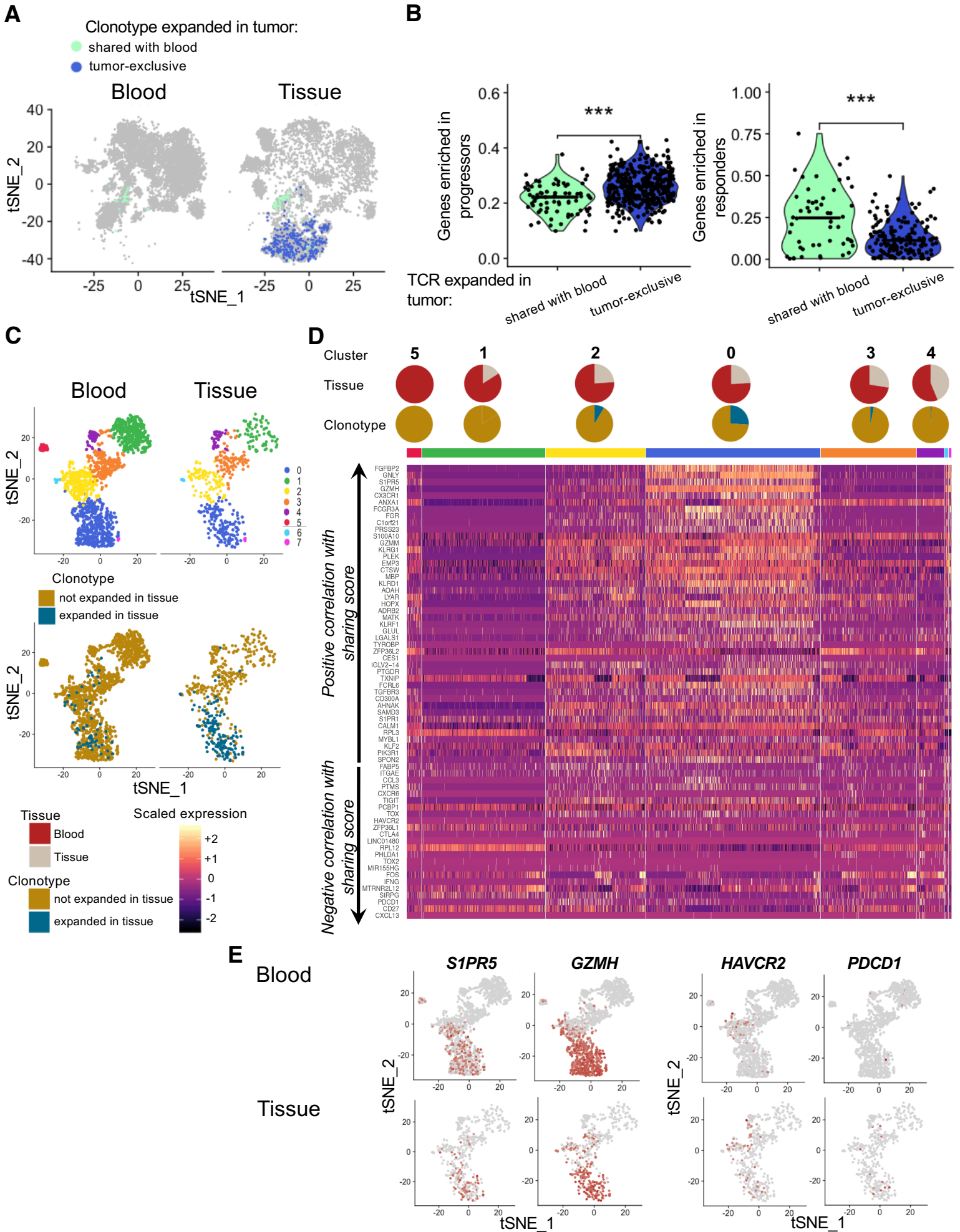
Figure 3

Figure 4

Figure 5

Patient I.D.	Sex	Age	Lesion location	BRAF/NRAS/c-Kit mutation	Sun exposure	Tumor burden	Disease duration (mo)		ICB history	Most recent ICB infusion	Response
Mel-T-1	F	62	Lymph node	WT	Exposed	Large	32		Treated	Pembrolizumab	P.D.
Mel-T-2	M	40	Lymph node	WT	Exposed	Large	15 (IIIC)		Treated	APX005M + Cabiralizumab +	P.D.
Mel-T-3	M	71	Lymph node	NRAS	Exposed	Small	20 (IIIB)		Treated	Nivolumab	P.D.
Mel-T-5	F	44	Lymph node	NRAS	Exposed	Large	8		Treated	Pembrolizumab	P.D.
Mel-T-6	M	57	Lung	NRAS	Exposed	Small	29		Treated	Nivolumab + Ipilimumab	P.R.
Mel-UT-1	F	79	Lymph node	BRAF	Exposed	Small	2 (IIIB)		Naïve	N.A.	N.A.
Mel-T-7	F	40	Gall bladder	BRAF	Exposed	Large	26		Treated	APX005M + Cabiralizumab +	P.D.
Mel-T-8	M	68	Small bowel	BRAF	Exposed	Large	17		Treated	Pembrolizumab	P.D.
Mel-T-9	M	43	Lymph node	BRAF	Exposed	Small	22		Treated	Nivolumab + Ipilimumab	P.R.
Mel-T-10	M	61	Soft tissue	BRAF	Exposed	Large	39		Treated	APX005M + Cabiralizumab +	P.D.
Mel-T-11	M	36	Lymph node	BRAF	Exposed	Small	17		Treated	Nivolumab + Ipilimumab	P.D.
Mel-UT-2	M	71	Lymph node	WT	Shielded (mucosal)	Large	2 (IIID)		Naïve	N.A.	N.A.
Mel-T-12	M	34	Skin metastasis	BRAF	Exposed	Large	12		Treated	Nivolumab + Ipilimumab	P.R.