A gene signature predicting natural killer cell infiltration and improved survival in melanoma patients

Joseph Cursons[1,2,†], Fernando Souza-Fonseca-Guimaraes[2,3,†], Momeneh Foroutan[1,4], Ashley Anderson[1], Frédéric Hollande[4], Soroor Hediyeh-Zadeh[1], Andreas Behren[5,6], Nicholas D. Huntington[2,3,7^,*], Melissa J. Davis[1,2,8,^,*]

[1] Bioinformatics Division, Walter and Eliza Hall Institute of Medical Research. Parkville, Victoria 3052, Australia

[2] Department of Medical Biology, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Parkville, Victoria 3010, Australia

[3] Division of Molecular Immunology, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052, Australia

[4] Department of Clinical Pathology, The University of Melbourne Centre for Cancer Research, Victorian Comprehensive Cancer Centre, Melbourne, Victoria 3000, Australia

[5] School of Cancer Medicine, La Trobe University, Melbourne, Victoria 3086, Australia

[6] Olivia Newton-John Cancer Research Institute, Heidelberg, Victoria 3084, Australia

[7] Biomedicine Discovery Institute and the Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, Australia

[8] Department of Biochemistry, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Parkville, VIC 3010, Australia

†,^ Authors have contributed equally to this work

* Corresponding Authors

Key words: natural killer cell, gene set scoring, TGF-β, transforming growth factor beta, phenotype-switching, melanoma survival

Corresponding author contact details:

A/Prof. Nicholas D Huntington
nicholas.huntington@monash.edu
+61 3 9345 2409

Dr Melissa J Davis
davis.m@wehi.edu.au
+61 3 9345 2597

Total word count: 5,752 (Introduction – Discussion, inclusive)

Total number of Figures: 5

Total number of Tables: 2

Abstract

Natural killer (NK) cells are a promising immunotherapy target. Emerging data show that NK cell activity is essential for initiating anti-tumor responses and may be linked to the success of current immunotherapies. As such, there is interest in exploiting NK cells and other innate immune components to treat cancer, and accordingly, a need for intuitive tools and methods that identify new therapeutic avenues. Here, we extend our gene set scoring method *singscore* to investigate NK cell infiltration within RNA-seq samples from bulk tumors.

Several computational methods have been developed for the deconvolution of immune cell types within solid tumors. We have taken the NK cell gene signatures from several tools, then curated and expanded this list using a comparative analysis of tumors and immune cell types. Using a gene set scoring method to investigate RNA-seq data from The Cancer Genome Atlas (TCGA) we show that patients with metastatic cutaneous melanoma have an improved survival rate if their tumor shows evidence of NK cell infiltration. Furthermore, these survival effects are enhanced in tumors which have a higher expression of NK cell stimuli such as IL-15. Using this signature, we then examine transcriptomic data to identify tumor and stromal components which may influence the penetrance of NK cells into solid tumors.

These data support a role for NK cells in the regulation of human tumors and highlight potential survival effects associated with increased NK cell activity. Our computational analysis identifies putative targets which may help NK cell anti-tumor immunity.

Introduction

Immunotherapies have greatly impacted clinical treatment for several cancer types, including renal cell carcinoma [1], non-small cell lung cancer [2], hematological malignancies [3], and in particular, melanoma [4]. While these treatments show great promise, their efficacy is largely restricted to a certain subset of tumor types and patients with 'immune hot' tumors. Accordingly, there is growing need to understand the role of the innate immune system in mediating a robust anti-tumor response, as this can help to open new therapeutic avenues which will increase the potential of established clinical treatments, increasing their efficacy and expanding the range of targeted malignancies. Natural killer (NK) cells are a subset of innate lymphoid cells and an essential effector of innate immunity. They have an exquisite cytotoxic ability, allowing them to kill tumor cells even at a relatively low ratio (e.g. 1:1) [5], and they are critical in the clearance of cells that carry a viral burden or have undergone oncogenic transformation. A number of *in vivo* studies have demonstrated a role for NK cells in limiting the metastatic dissemination of melanoma [6-9] and there is growing interest in the targeting of NK cells for novel immunotherapeutics [10, 11]. It has recently been demonstrated that NK cells initiate a robust anti-tumor T-cell response by recruiting conventional type-1 dendritic cells (cDC1) through chemokine signaling (*via.* XCL-1 & CCL-5/RANTES) [12], and supporting stimulatory DCs (sDCs) through the expression of Flt3 ligand (*Flt3l* in mice and *FLT3LG* in humans) [13].

Important regulators of NK cell activity include the cytokines IL-15 [14], chemokines such as CCL5 (RANTES) [15], growth factors such as TGF-β [16, 17], and the intracellular JAK-STAT signaling component CIS [18]. Evidence suggest that modulation of NK cell populations is feasible for cancer treatment [19], and treatments based upon systemic administration of IL-15 constructs have shown promise in leukemic and solid tumors [20-22]. Thus, while IL-15 provides a promising treatment to stimulate immune targeting of cancers, we note that cytokine-mediated NK cell activation and expansion can be increased in combination with IL-12 and IL-18 [23] or further amplified through deletion of CIS (encoded by *CISH*), an important negative regulator of cytokine signaling and effector function, such that *Cish*$^{-/-}$ mice are resistant to a range of metastatic cancers [18]. Accordingly, we are yet to fully elucidate the range of molecular regulatory systems that control NK cell activity *in vivo.*

Advances in sequencing technology and associated methods for data analysis over recent decades have allowed the application of transcriptomic profiling to complex tumor samples [24, 25]. Resulting data have enabled the development of mathematical methods, such as CIBERSORT [26] that infer the relative abundance of immune cells which have infiltrated into solid tumor samples. While these tools have had a profound impact on our understanding of immune infiltration [27-30] and led to a new field of cancer research, the complexity of model fitting procedures make it difficult for researchers to modify or investigate these

gene lists. In addition, there are further opportunities to capitalize on public tumor transcriptomic data in order to identify how changes in tumor phenotype are associated with changes in the relative abundance of immune sub-populations.

We have recently developed a single-sample gene set scoring method which uses a rank-based metric to quantify the relative enrichment of a gene set within a sample transcriptome [31]. We have combined NK cell signatures from the LM22/CIBERSORT [26] and LM7 [32] gene sets and curated this list to produce a gene set which reflects the relative abundance of NK cells within a tumor sample. As melanoma tumor cells are highly-immunogenic, we have focused upon the analysis of TCGA skin cutaneous melanoma (SKCM) data [33] demonstrating that the relative expression of NK cell genes within metastatic tumors is associated with a strong survival advantage. Using the SKCM data we show how our scoring approach can be used to explore putative modulators of NK cell activity by examining their association with NK score and survival effects associated with their expression.

Materials and Methods

Data

Data used in this study are available from listed repositories (Table 1). For TCGA SKCM data, RSEM abundance data without normalization were downloaded directly from the genomic data commons. For sorted immune cell populations (GSE60424 [34] & GSE24759 [35]) and melanoma cell line data (E-MTAB-1496), processed transcript abundance data were downloaded and used directly. For GSE24759, only samples derived from peripheral blood were used, data from colony-forming samples was excluded to exclude culturing effects, and CD56-/CD16+/CD3- mature NK cell data were excluded due to apparent batch effects (Fig. S1). In cases with gene multi-mapping (multiple probes/probe sets per gene), median values were used. To better examine the relative expression of marker genes across immune cell subsets, we examined $CD45^+$ ('non-malignant') single cell RNA-seq data from dissociated melanoma samples with annotated cell types (GSE72056) [36]. Further single cell RNA-seq from (checkpoint inhibition naïve) dissociated melanoma samples (GSE120575) [37] were used for subsequent comparison and visualization.

As noted in the results and shown in Fig. S2A & S2B, there are large survival differences between patients with primary and metastatic tumors. To avoid confounding effects from this, unless otherwise stated, we have focused on patients with metastatic tumors only who also had valid age and survival data. One patient with both a metastatic and primary sample was excluded.

Microarray data relative log expression & principal component analysis

Using relative log expression (RLE) plots to explore unwanted variation [38], log-transformed transcript abundance data (downloaded directly from GEO) were median-centered for each gene, and then within each sample the difference between the observed and population median of each gene was calculated. For the principal component analysis (PCA), genes with an expression level above the 10^{th} percentile (5.34) within at least 4 samples (corresponding to the smallest sample group) were retained. Data were normalized using sklearn StandardScaler, before calculating principal components using the sklearn PCA function.

Immune gene sets & previous classifications

Genes identified in Figure 1 were annotated as immune-associated using the Gene Ontology category GO:0002376 - '*immune system process*', together with all descendant or child processes [39]. The original SKCM manuscript [33] was used for the associated TCGA immune gene list and TCGA classifications of 'Immune high' patients.

Single cell RNA-seq analysis and visualization

Pre-processed single-cell RNA data were downloaded from respective sources (Table 1) and were analyzed through the Seurat (v. 2.3.4) pipeline [40].

Single cell RNA-seq data from Tirosh et al. (2016) [36] have annotated cell types and thus $CD45^+$ 'non-malignant' cells were used for signature curation. Cells with abundance data for less than 1,500 genes or more than 10,000 genes were removed, and genes were retained if their abundance exceeded a selected threshold (logTPM > 1) in more than 30 cells. To improve downstream clustering and visualization of immune subtypes used, endothelial cells were removed before performing a PCA across the 1,000 most-variable genes, taking the first 50 principal components for uniform manifold approximation and projection (UMAP) [41] visualization, with a 'minimum distance' of 0.2 and 'number of neighbors' equal to 50.

Pre-processed $CD45^+$ single cell RNA-seq data from treatment-naïve samples were taken from Sade-Feldman et al. (2018) [37]. Cells with abundance data for less than 1,000 genes or more than 5,000 detected genes were removed, as were cells with more than 5% of reads derived from mitochondrial genes. PCA was performed across the 3,550 most variable genes, and the top 20 principal components were used for visualization with UMAP, using a 'minimum distance' of 0.1 and 'number of neighbors' equal to 30.

Natural killer cell signature curation

A schematic overview of the workflow for NK signature curation is given in Fig. S2. With the aim of applying a unidirectional (expected upregulated) gene set with *singscore* [31], a preliminary NK cell gene set was created by combining all 'expected upregulated' genes from the CIBERSORT (LM22) active and resting NK cell gene sets [26], the LM7 NK cell gene

set [32], and human orthologs for established NK cell markers from a number of mouse studies [18, 42, 43] (collectively referred to as the 'Huntington gene list'; Fig. S2).

To improve specificity of this signature against other immune subsets we examined bulk RNA-seq from sorted cell populations and single-cell RNA-seq data from melanoma metastases. SRA files for immune cell bulk RNA-seq data from healthy individuals were downloaded September 2016 and converted to FASTQ format using the SRA toolkit. Reads were mapped to human genome hg19 using Rsubread (v. 1.32.0) [44] and counts were quantified using featureCount. Count-per-million (CPM) and RPKM values were calculated using edgeR (v. 3.24.0). For differential expression analysis, genes were retained if their abundance exceeded a threshold (CPM > 4) within at least 4 samples (i.e. all samples for a sorted cell type). The voom-limma (v. 3.38.2) pipeline [45] was used by applying the TREAT criteria [46] (log|fold change| > 1) to calculate the t-statistics, log|fold change| and adjusted p-values for all genes when comparing NK cells against all other sorted cell populations (i.e. excluding whole blood), and the Homo.sapiens (v. 1.3.1) package was used for annotation. Results for NK signature genes are given in Table S1.

For $CD45^+$ single cell RNA-seq data (Fig. S2), if genes had evidence of extensive dropout (median value below 0.5 for all cell types) they were retained if the 75^{th} percentile value for NK cells was higher than all other cell types and greater than 2.67 (90^{th} percentile of abundance across all genes and cells, for genes with logTPM > 1 in at least 30 cells). All other genes (which passed the above dropout criteria) were retained if the median expression within NK cells was greater than the 75^{th} percentile of all other cell types. Genes were retained if they passed these criteria within both the sorted bulk RNA-seq and single cell RNA-seq data.

Next, we improved the specificity of our NK signature against genes which are expressed across solid tumors. For all genes we examined the expression within adherent cell lines (i.e. hematopoietic and lymphoid cell lines were excluded) using RNA-seq data from the Cancer Cell Line Encyclopedia [47]. Genes were retained if their median expression (logTPM) was below 5.23 (25^{th} percentile of non-zero abundance across all genes and cell lines). Results of these tests are given in Table S1.

Gene set scoring

Gene set scoring was performed using the *singscore* approach [31]. Briefly, genes are ranked by increasing transcript abundance and for a set of target genes the mean rank is calculated and normalized against theoretical minimum and maximum values. If directional gene lists are provided (i.e. a set of genes expected to be upregulated and a set of genes to be downregulated), as with the TGF-β EMT signature [48] then the mean rank of expected up-regulated genes is calculated from genes ranked by increasing abundance, while the mean rank of expected down-regulated genes is calculated from genes ranked by decreasing

abundance, and these values are then normalized and summed. Accordingly, a high gene set score indicates that the pattern of gene expression in a sample is concordant with the pattern captured by the gene expression signature.

Survival analyses

Cox proportional hazard models and Kaplan-Meier survival curves were generated using python lifelines (v. 0.14.6; DOI: 10.5281/zenodo.1303381) with standard parameters. For individual gene Cox hazard models a Bonferroni correction was applied to correct for multiple-hypothesis testing. Unless otherwise specified: separation of patient samples with a single parameter (e.g. age, gene expression) used the 33rd and 66th percentile values to threshold, and; separation of patient samples with two parameters used the 40th and 60th percentile values to threshold.

Code availability

General analyses and visualization were performed using python v3.6 with: pandas [49] for data handling; scipy (v 1.1.0), scikit-learn (v. 0.19.2), and numpy (v. 1.14.5) for numerical calculations, and; matplotlib (v. 2.2.3) for plotting/visualization. For analyses with R, dplyr (v. 0.7.8) and tidyr (v. 0.8.2) were used for formatting data, and; ggplot2 (v 3.1.0), gridExtra (v. 2.3) and RColorBrewer (v. 1.1-2) for visualization.

All computational scripts used in this work will be made freely available from our GitHub repository:

https://github.com/DavisLaboratory/NK_scoring

Results

Cutaneous melanoma is generally associated with a strong immunogenic response

Cutaneous melanoma is an ideal target for immunotherapy as the high mutational burden of this malignancy is associated with the generation of neo-antigens which can induce an immune response [50]. Several reports demonstrate that immune infiltration signatures provide a strong prognostic indicator in melanoma [51], including the TCGA skin and cutaneous melanoma (SKCM) study which demonstrated that this effect was independent of the underlying genomic subtype of the melanoma [33].

Due to significant survival differences between patients with a primary or metastatic tumor (Fig. S3A & B) this report focusses on the 365 patients with only metastatic tumor samples. There are significant survival effects associated with the patient's age at diagnosis (Fig. S3C), and while female patients had better survival rates (Fig. S3D), these effects were not significant (Table 2; Cox proportional hazards model, *p*-value = 0.20). Similarly, while certain metastatic tumor sites (e.g. central nervous system) were associated with a greater hazard coefficient, these effects did not appear to be statistically significant (Table S2A). We

note, however, that certain sites had relatively low sample numbers (Table S2B) which may be contributing to greater standard errors and reducing statistical power.

To analyze survival effects associated with individual genes we built a series of Cox proportional hazard models for each gene where patient age at diagnosis was included as the only covariate (together with transcript abundance for that gene). As shown in Fig. 1A, when focusing upon highly-significant genes with a negative hazard coefficient, many are annotated as modulators of immune function (*genes in bold*) and a number of remaining genes are well-known immune modulators that lack associated GO annotations (e.g. *CLEC2B*, *CD72*, *SRGN* and *MIR155HG*). The associated patient survival curves are shown for patient age (Fig. 1B) and a selection of genes (Fig. 1C-E). Higher expression of the hallmark inflammatory cytokine encoded by *IFNG* corresponds to improved survival outcomes (Fig. 1C), while a number of interferon-induced genes are further associated with a hazard reduction (*e.g. IRF1*, *IFITM1*; Fig. 1A). High tumor transcript abundances of the NK cell marker gene *KLRD1* (Fig. 1D; also known as *CD94*) or the cytokine IL-15 (Fig. 1E) which is an important regulator of NK cell [14, 42, 52] and T-cell activity [53] are also associated with improved long-term survival outcomes. Finally, we note that transcript abundance for the *B2M* gene encoding beta-2 microglobulin has one of the most negative hazard coefficients, likely reflecting its role in MHC class I antigen presentation of neo-antigens to CD8 T cells and consistent with recent reports of the importance of this process for immune control of tumors [54]. Further, a truncation mutant of B2M can confer resistance to PD-1 blockade in melanoma [55], and mutations in B2M have been shown to disrupt immune surveillance in lung cancer [56]. The large negative hazard coefficient associated with *HAPLN3*, encoding a hyaluronan and proteoglycan link protein, suggests that this gene may warrant further investigation in the context of immune recognition and targeting.

Developing a more specific transcriptomic signature for natural killer cells

Several transcriptomic data deconvolution methods have generated gene signatures that are predictive of tumor infiltration by specific immune cell sub-populations. Of note for this work, we examined the training data (transcriptomics from sorted immune cell populations) and NK cell signatures from the LM22 [26] and LM7 [32] gene sets. A common critique of immune deconvolution methods is the high co-linearity/cross-correlation between different signatures [32]. While this can largely be attributed to the similar transcriptional profiles of some lymphocyte subsets (*e.g.* demonstrated by similar positions of sorted NK and T cell populations in PCA plots; Fig. S1B & S1C), to an extent it also represents the cascading series of intercellular interactions that mediate immune activation within complex tissue samples. Accordingly, several immune-associated gene subsets are cross-correlated to a varying extent (Fig. S4). To try and address this, we combined and curated (Fig. S2; details given in Materials and Methods) the NK cell signatures from the LM7 and LM22 gene lists, together

with human orthologues for well-known mouse NK marker genes. Genes were retained if they had higher expression in NK cells relative to other immune populations, as well as relatively low expression across adherent cell lines from the CCLE (Table S1).

The relative expression of these 20 marker genes is shown across immune subsets from melanoma single cell RNA-seq data (Fig. 2A), as well bulk RNA-seq (Fig. 2B) and microarray (Fig. 2C) data from sorted immune cell populations. Many NK marker genes have some expression within $CD4^+$ and $CD8^+$ T-cell populations (Fig. 2B & C) – in particular the cytotoxic mechanisms used by NK cells share many similarities to $CD8^+$ T-cells, including secretion of granzymes (e.g. GZMA, GZMB, GZMK, GZMM) and perforin (PRF1) [57]. As demonstrated by genes without extensive dropout in the single cell RNA-seq data (e.g. *CTSW*, *GZMB*, *IL2RB*, *KLRD1* and *NKG7*), however, the minimum abundance within NK cells tends to be greater than the 75^{th} percentile of abundance within T cell populations. For genes with minor dropout issues (e.g. *NCR1*, *IL18RAP*, *XCL1* and *XCL2;* median abundance for all cell types around 0) the 75^{th} percentile of abundance within NK cells corresponds to relatively high expression levels (logTPM of approximately 5 or above). There does appear to be a small subset of unresolved cells (Fig. 2A; *gray*) which show relatively high expression of these markers, such as *CTSW*, *GZMB* and *NKG7*. Examining the cell classifications by Tirosh *et al.* (2016) many of the unresolved cells appear show expression of various lymphocyte markers from both B cell and T cell populations and it may be possible that these correspond to unresolved cell doublets.

Given our use of annotated cell subsets from Tirosh et al (2016) for NK cell marker curation, we next examined the expression of selected lymphocyte markers across an independent, larger set of single cell RNA-seq data from Sade-Feldman et al (2018). As shown (Fig. 3; at top) annotated NK cells within the Tirosh data have little or no expression of T cell markers such as *CD3D* and *CD4*, and lower expression of *CD3E*, while there is strong expression of remaining NK cell marker genes. Note that *FCGR3A* and *NCAM1* were excluded from our curated signature (Fig. 2) due to overlap with other immune subsets and solid cancer cell lines, respectively, however they are included here as established markers to delineate lymphocyte populations. Examining the Sade-Feldman data there are a subset of cells with lower expression *CD3E* and no expression of *CD4*, yet relatively high expression of other indicated NK marker genes. Of note, *KLRF1* (NKp80) appears to have particularly high expression within this subset that we have annotated as NK cells.

Gene set scoring allows dimensional reduction of RNA-seq data

Gene set enrichment analyses are commonly used after differential expression to assess whether genes with the most significant changes are enriched for classifications of specific pathways or processes. An alternative 'relative approach' [25] is to analyze the gene expression patterns (transcript abundances) of individual samples and calculate the relative

concordance of each one against specific gene set that has been defined to capture a particular molecular phenotype.

We have recently developed a gene set scoring method, *singscore* [31], which uses the normalized mean rank of genes that are associated with a specific molecular phenotype or cellular behavior [48, 58]. With this approach, a difference in score between two samples can be related to the percentile change in mean rank of the gene set, providing a metric which summarizes the concordance between the gene expression profile of an individual sample and the specified gene sets. Using this scoring method with "Immune cluster" genes from the original TCGA SKCM publication [33] we can largely recapitulate the original sample clustering (Fig. S5), and furthermore, we can easily extend this analysis to samples that have more recently been added to the TCGA SCKM cohort.

The NK cell score is associated with improved patient survival

We used our curated 20 gene NK signature to score metastatic tumors from the TCGA SKCM cohort. When samples are sorted by increasing NK score the highly concordant expression pattern of these genes across metastatic melanoma samples becomes apparent, and as expected, all these genes carry a positive correlation with the NK score (Fig. 4A). Using our NK signature to partition patients, there are strong survival effects for patients with either a high or moderate NK score (Fig. 4B). These survival effects are largely recapitulated across a selection of marker genes (Fig. 4B) associated with: T cell infiltration (*CD3D*), cytokine signaling which promotes NK cell and T cell infiltration (*IL15*), a shared component of the IL-2 and IL-15 receptors (*IL2RB*), and interferon-gamma responsive checkpoint markers (*CD274*), NK secreted chemokines/cytokines associated with dendritic cell recruitment and stimulation (*CCL5*, *XCL1*) and cytotoxic effector molecules (*GZMB*, *FASLG*).

Next, we investigated the relative survival effects of key cytokine and cytotoxic components linked to NK cell function (*IL15*, *XCL1*, *XCL2*, *CCL5*, *FLT3LG*, *GZMA*, *GZMB* and *FASLG*) within patient samples from the low NK score and high NK score groups (Fig. S6). For metastatic tumors with a low NK score, *IL15* and *FLT3LG* showed some survival benefit, however, none of the genes examined (including these) showed a significant (p-value < 0.05) association with patient survival. Conversely, when examining metastatic tumors with a high NK score, there appears to be improved patient survival linked to relatively high expression of the chemokines *XCL2* (also shown in Fig. 4C; p-value = 0.035) and *CCL5* (p-value = 0.031), while there are no survival differences in this group for the cytotoxic genes investigated (Fig. S6), exemplified by *GZMB* (Fig. 4C).

We note that it is difficult to disentangle the survival effects of other cytotoxic/effector immune cells, as improved survival effects are also seen for the T cell marker CD3E (Fig. 4B). This is further reflected by the very strong association between our NK score, and scores derived from the TCGA Immune cluster genes (Fig. S7A) or the T cell signature (Fig. S7B &

Fig. S8), which also share similar survival effects (Fig. S7C & D). As discussed above, however, there is increasing evidence that NK cells play a key role initiating the intercellular signaling cascade which is necessary for strong immune recruitment. Using the Böttcher 5-gene NK cell signature [12] there is good concordance with our NK signature score (Fig. S7E), as well as between our NK score and a score calculated using the Böttcher 4 gene DC cell signature (Fig. S7F), again with similar survival effects (Fig. S7G & H).

NK cell targeting of mesenchymal-like melanoma tumors with evidence of low TGF-β activity is associated with favorable patient survival

A key advantage of the *singscore* approach is that it allows gene signatures associated with cell phenotype to be combined for further investigation. Recent work on innate anti-PD-1 resistance in melanoma found similarities with markers of MAPK inhibitor resistance [59], and as noted earlier, melanoma phenotype switching has been linked to general drug/MAPK inhibitor resistance [60]. Thus, to investigate this further, we examined the relative association between our NK score, and several phenotype-switching associated gene set scores, including: a proliferative, epithelial, phenotype; an invasive, mesenchymal phenotype; and a mesenchymal phenotype where EMT has been induced specifically by TGF-β [48]. We found no association between NK score and epithelial score (Fig. 5A), however there was an association with mesenchymal score (Fig. 5B), such that less mesenchymal tumors have lower NK scores, while highly mesenchymal samples had a range of NK scores, suggesting that a subset of these tumors had higher NK cell infiltration. Intriguingly, there was no association with NK score relative to the TGF-β specific EMT gene score (Fig. 5C), despite the fact that there was a relatively strong positive association between TGF-β EMT score and mesenchymal score (Fig. 5D) – as we had previously observed however, while all samples with a high TGF-β EMT score scored high against general mesenchymal gene expression, a subset of highly mesenchymal samples showed no evidence of TGF-β driven EMT. Further, while neither the TGF-β EMT score (Fig. 5E) or NK score (Fig. 5F) had any association with age, when we partitioned patients by NK score and TGF-β EMT score, those with evidence of good NK cell infiltration and low TGF-β activity had favorable survival outcomes (Fig. 5G & H), particularly for patients within the younger cohort.

Although most *in vivo* experiments indicate a primary role for NK cells in limiting metastatic colonization [7], these data suggest that not only are NK cells associated with established metastatic tumors, but the presence of NK cell infiltrate is associated with an improved prognostic outcome.

Natural killer cells offer a promising avenue for targeted immunotherapeutics to control melanoma metastasis

As noted above, a critical role for NK cells in driving a robust immune response has further support from a recent study which has demonstrated an important role for the NK-DC cell axis in modulating responsiveness to immunotherapy [13]. To further investigate potential modulators of NK cell infiltration we next examined transcriptomic data from the LM-MEL panel which contains representative cell lines for both the proliferative and invasive phenotype. Gene sets were filtered to retain only those present in both the TCGA and LM-MEL data, and gene set scoring was repeated for both data sets to facilitate comparison between tumor samples and the corresponding cell line models (Fig. S9A-C). As shown, in the absence of a high NK score, patients with a high mesenchymal score show no survival effects associated with the TGFβ-EMT score (Fig. S9D).

A number of melanoma cell lines from the LM-MEL panel appear to be associated with various subsets of high/low mesenchymal score and TGFβ-EMT score (Fig. S9A-C; colored scatter markers). By contrasting genes correlated or anti-correlated with NK score across the TCGA data against these cell line data we can identify markers which may be derived from the melanoma tumor which exert an immunomodulatory effect (Fig. SE). To demonstrate the association with phenotype-switching the markers CDH1 and MITF are included [61, 62]. Further, TGF-β activity in these cell lines has a demonstrated association with *THBS1* [63]. Consistent with our observation that NK score tends to be higher in more mesenchymal tumor samples, many of the positively correlated genes tend to have higher expression in the MITF-low cell lines. Similarly, many of the anti-correlated genes tend to have lower expression in the MITF-low cell lines.

Several notable genes are present within these lists (Fig. S9E). Again, B2M is identified, together with a number of HLA- genes, suggesting that more mesenchymal-like melanomas may be more immunogenic in part because of increased antigen presentation associated with this phenotype. Given data linking IL-18 to NK cell activity [64], it is interesting to note that in the more mesenchymal cell lines, the expression of IL-18 appears to be slightly higher in the TGF-β EMT low samples. From the genes anti-correlated with NK score we note that CMTM4 was recently identified as a positive regulator of PD-L1 (together with its paralog CMTM6 which is not present) [65] and this appears to have lower expression within more mesenchymal cell lines.

Discussion

Immune “checkpoint” inhibitor antibodies, which function by reactivating tumor-resident cytotoxic lymphocytes, have revolutionized cancer therapy. Although much clinical research is currently directed towards programs that underlie immunotherapy resistance and immune related adverse events, we still lack an in-depth understanding of the fundamental mechanisms dictating response, and we do not yet have robust markers to identify patients who are the most likely to respond in the context of metastatic disease. Checkpoint

inhibitors primarily block inhibitory pathways in tumor-resident T cells, however interest in other effector populations, such as NK cells, is growing [10], with recent studies showing that NK cells have a critical role in immunotherapy success [13].

Clinically, peripheral blood NK cell activity has been shown as inversely correlated with cancer incidence [66]. More recent evidence has shown that NK cell infiltration in human tumors is associated with better prognosis in squamous cell lung carcinoma, as well as gastric and colorectal carcinomas [7]. In melanoma cells, researchers have previously found specific HLA-I allelic losses in up to 50% of patients analyzed, and even when expressed on melanoma cells, specific HLA class I molecules are often at insufficient levels to inhibit NK cell-mediated cytotoxicity [67]. These data hint that metastatic melanoma is an ideal target for NK cell-mediated killing and therapies that enhance NK cell activity should be investigated further. Along these lines, cutaneous, subcutaneous and lymph node melanoma biopsies from a small cohort of stage III and IV unresectable metastatic melanoma patients being treated with anti-PD1 (Pembrolizumab) were recently analyzed for immune infiltrate [13]. Metastatic melanoma biopsies from patients who responded to Pembrolizumab had significantly higher NK cell infiltration compared to non-responders, and this NK infiltration was strongly correlated with DC infiltration. What was surprising however, is that this report also observed an association between regulatory T cell and effector T cell populations in both responders and non-responders [68], as $CD8^+$ T cell proliferation has previously been reported as a marker of Pembrolizumab on-target effect and tumor regression [68].

While the kinetics of immune infiltration into tumors are difficult to study directly, Barry and colleagues proposed a role for NK cells in supporting DC persistence/survival via their production of the key DC growth factor FLT3 ligand, and demonstrated these effects using a *Flt3l* transgenic mouse model [13]. Together with another report [12], it was proposed that NK cell infiltration may proceed the majority of DC infiltration since intratumor NK cells are a major source of *XCL1*, a chemoattractant for $XCR1^+$ DCs. These emerging data on melanoma immune infiltrate have motivated this study, and we have performed a detailed investigation of the transcriptomic and matched clinical data available through the TCGA skin cutaneous melanoma cohort [33].

Metastatic melanoma tumors which have a high NK score are associated with much better patient survival (Fig. 4B), consistent with the results from Böttcher *et al* (2018) and a range of *in vivo* animal survival studies [6, 17]. This effect was largely recapitulated by the important NK cell-secreted chemokines *XCL1* and *CCL5*, as well as NK cell effectors *GZMB* and *FASLG* (Fig. 4B). Interestingly, a comparison of cytokine and cytotoxic gene expression within the patient subset with NK score high metastatic tumors suggests that higher expression of cytokines such as *CCL5* and *XCL2* may have a greater effect on survival than

higher expression of cytotoxic effectors such as *GZMB* and *FASLG* (Fig. 4C; Supp. Fig. S6). This may represent a saturation of the effect of cytotoxic molecules within activated cytotoxic lymphocytes (already expressing relatively high levels of these molecules), but this may also reflect the ability of NK cells to recruit and support DCs, driving amplification of the anti-tumor response through intercellular signaling programs. We note that while the gene set derived by Böttcher and colleagues performs well [12], they calculated NK scores using mean $\log_2$ abundance data which can be susceptible to outliers and places a greater weighting on genes with high transcript abundance. As demonstrated (Fig. S6E & S6F), while *singscore* has been developed for larger gene sets, it still performs relatively well with the 5 gene Böttcher NK signature (*NCR3*, *KLRB1*, *PRF1*, *CD160*, *NCR1*).

While it is hard to directly compare the accuracy of these signatures without validation data for NK cell infiltration, these results demonstrate the application of our computational method in estimating the abundance and heterogeneity of different immune subsets across different tumors and patients. Importantly, our methods allow variations in these relative immune scores to be easily compared against other relevant phenotype-associated gene sets as demonstrated by the striking survival effects we observed when our NK score is examined in the context of melanoma phenotype switching and TGF-β signaling (Fig. 5G & H). Phenotype-switching is an important regulatory program involved in the progression of melanoma [62, 69] which has been linked to vemurafenib resistance [70] and general drug resistance [60]. It allows tumor cells to transition between proliferative (“epithelial-like”) and invasive (“mesenchymal-like”) behaviors, and in melanoma, there is strong evidence that TGF-β is an important driver of EMT/phenotype-switching programs [48], mediated in part by signaling molecules such as thrombospondin 1 [63].

Further, while there are prominent long-term survival effects associated with patient age (Fig. 1B), we note that there is no association between NK score and age (Fig. 5E). Yet, the survival advantages associated with a low TGFβ-EMT score and a high NK score are particularly pronounced within younger patients (Fig. 5G). These results suggest that younger patients may receive a greater benefit from NK targeted immunotherapies, perhaps reflecting a higher capacity of the immune system in young patients to produce a robust anti-tumor response.

A clear link between tumor inflammation and response to immune checkpoint blockade has been well established [68]. Why certain tumor types are largely devoid of immune infiltrate and why dramatic heterogeneity in immune infiltration exists for similar tumor types are outstanding questions that drive a considerable amount of immunotherapy drug R&D efforts in order to increase indications for class-leading drugs such as Pembrolizumab and Nivolumab. A large pan-cancer analysis of such immune checkpoint inhibitors recently found that tumor mutational burden (TMB) was correlated with patient outcomes, with high

TMB patients (top 10% TMB by histology) having significantly better survival than low TMB patients (bottom 80%) [71]. While immune infiltration data was not presented in this study, it would be of interest to examine links between TMB and immune infiltrate, especially since there appears to be a clear cut-off (>15%) for TMB to impact on the Hazard ratio in this large cohort study. A recent overview of the literature in non-small cell lung cancer revealed that smokers have higher TMB [72] and a higher PD-1L level compared to non-smokers [73] which can account for superior PD-1 blockade response rates. Not surprisingly, our analysis of the TCGA data also found PD-1L levels (*CD274*) to correlate strongly with predicted immune infiltrate, NK cell score and melanoma patient survival (Fig. 4B), suggesting TMB may be linked to immune infiltration and highlighting the need for further work to examine this interplay.

The NK cell gene signature and NK cell gene score that we describe here can be readily applied to other cancer datasets which are becoming increasingly available thanks to the efforts of large cancer research consortia. The information from such gene signature analyses will also allow researchers to stratify responders and non-responders to conventional treatments, identify the patients that are likely to profit from NK cell-based immunotherapies, and facilitate the development of prognostic markers for personalized immunotherapeutics.

Disclosure of Potential Conflicts of Interest

N. Huntington has ownership interest (including patents) in oNKo-innate Pty Ltd. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: J. Cursons, F. Souza Fonseca Guimaraes, N.D. Huntington, A. Behren, M.J. Davis

Development of methodology: J. Cursons, M. Foroutan, F. Hollande, M.J. Davis

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): J. Cursons, M. Foroutan, F. Hollande, A. Anderson, S. Hediyeh Zadeh, F. Souza Fonseca Guimaraes, N.D. Huntington, M.J. Davis

Writing, review, and/or revision of the manuscript: J. Cursons, F. Hollande, F. Souza Fonseca Guimaraes, N.D. Huntington, A. Behren, M.J. Davis

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): J. Cursons, A. Anderson, M. Foroutan

Study supervision: J. Cursons, F. Souza Fonseca Guimaraes, N.D. Huntington, M.J. Davis

Acknowledgments

This work is supported in part by project grants from the National Health and Medical Research Council (NHMRC) of Australia (#1147528 to J.C.; #1128609 to M.J.D.; #1124784, #1066770, #1057852, #1124907 to N.D.H; #1164081 to F.H., and; #1140406 to F.S.F.G). F.S.F.G. was supported by NHMRC Early Career Fellowship (1088703), National Breast Cancer Foundation (NBCF) Fellowship (PF-15-008), and grant #1120725 awarded through the Priority-driven Collaborative Cancer Research Scheme and funded by Cure Cancer Australia with the assistance of Cancer Australia. N.D.H is a NHMRC CDF2 Fellow (1124788), a recipient of a Melanoma Research Grant from the Harry J Lloyd Charitable Trust, Melanoma Research Alliance Young Investigator Award, and a CLIP grant from Cancer Research Institute. M.J.D was supported by NBCF Career Development Fellowship ECF-14-043 and is the recipient of the Betty Smyth Centenary Fellowship in Bioinformatics. This study was made possible through Victorian State Government Operational Infrastructure Support and Australian Government NHMRC Independent Research Institute Infrastructure Support scheme.

Results published here are based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/.

References

1. Cho, Y.H., et al., *Novel immunotherapy in metastatic renal cell carcinoma.* Investig Clin Urol, 2017. **58**(4): p. 220-227.
2. Hellmann, M.D., et al., *Nivolumab plus Ipilimumab in Lung Cancer with a High Tumor Mutational Burden.* N Engl J Med, 2018. **378**(22): p. 2093-2104.
3. Nelson, M.H. and C.M. Paulos, *Novel immunotherapies for hematologic malignancies.* Immunol Rev, 2015. **263**(1): p. 90-105.
4. Ribas, A. and J.D. Wolchok, *Cancer immunotherapy using checkpoint blockade.* Science, 2018. **359**(6382): p. 1350-1355.
5. Huntington, N.D., C.A. Vosshenrich, and J.P. Di Santo, *Developmental pathways that generate natural-killer-cell diversity in mice and humans.* Nat Rev Immunol, 2007. **7**(9): p. 703-14.
6. Sathe, P., et al., *Innate immunodeficiency following genetic ablation of Mcl1 in natural killer cells.* Nat Commun, 2014. **5**: p. 4539.
7. Krasnova, Y., et al., *Bench to bedside: NK cells and control of metastasis.* Clin Immunol, 2017. **177**: p. 50-59.
8. Viant, C., et al., *Cell cycle progression dictates the requirement for BCL2 in natural killer cell survival.* J Exp Med, 2017. **214**(2): p. 491-510.
9. Lopez-Soto, A., et al., *Control of Metastasis by NK Cells.* Cancer Cell, 2017. **32**(2): p. 135-154.
10. Guillerey, C., N.D. Huntington, and M.J. Smyth, *Targeting natural killer cells in cancer immunotherapy.* Nat Immunol, 2016. **17**(9): p. 1025-36.
11. Souza-Fonseca-Guimaraes, F., J. Cursons, and N.D. Huntington, *The Emergence of Natural Killer Cells as a Major Target in Cancer Immunotherapy.* Trends Immunol, 2019.

12. Bottcher, J.P., et al., *NK Cells Stimulate Recruitment of cDC1 into the Tumor Microenvironment Promoting Cancer Immune Control.* Cell, 2018. **172**(5): p. 1022-1037 e14.
13. Barry, K.C., et al., *A natural killer-dendritic cell axis defines checkpoint therapy-responsive tumor microenvironments.* Nat Med, 2018.
14. Rautela, J. and N.D. Huntington, *IL-15 signaling in NK cell cancer immunotherapy.* Curr Opin Immunol, 2017. **44**: p. 1-6.
15. Mgrditchian, T., et al., *Targeting autophagy inhibits melanoma growth by enhancing NK cells infiltration in a CCL5-dependent manner.* Proc Natl Acad Sci U S A, 2017. **114**(44): p. E9271-E9279.
16. Viant, C., et al., *Transforming growth factor-beta and Notch ligands act as opposing environmental cues in regulating the plasticity of type 3 innate lymphoid cells.* Sci Signal, 2016. **9**(426): p. ra46.
17. Gao, Y., et al., *Tumor immunoevasion by the conversion of effector NK cells into type 1 innate lymphoid cells.* Nat Immunol, 2017. **18**(9): p. 1004-1015.
18. Delconte, R.B., et al., *CIS is a potent checkpoint in NK cell-mediated tumor immunity.* Nat Immunol, 2016. **17**(7): p. 816-24.
19. Liu, E., et al., *Cord blood NK cells engineered to express IL-15 and a CD19-targeted CAR show long-term persistence and potent antitumor activity.* Leukemia, 2018. **32**(2): p. 520-531.
20. Miller, J.S., et al., *A First-in-Human Phase I Study of Subcutaneous Outpatient Recombinant Human IL15 (rhIL15) in Adults with Advanced Solid Tumors.* Clin Cancer Res, 2018. **24**(7): p. 1525-1535.
21. Liu, B., et al., *Evaluation of the biological activities of the IL-15 superagonist complex, ALT-803, following intravenous versus subcutaneous administration in murine models.* Cytokine, 2018. **107**: p. 105-112.
22. Romee, R., et al., *First-in-human Phase 1 Clinical Study of the IL-15 Superagonist Complex ALT-803 to Treat Relapse after Transplantation.* Blood, 2018.
23. Lusty, E., et al., *IL-18/IL-15/IL-12 synergy induces elevated and prolonged IFN-gamma production by ex vivo expanded NK cells which is not due to enhanced STAT4 activation.* Mol Immunol, 2017. **88**: p. 138-147.
24. Finotello, F. and Z. Trajanoski, *Quantifying tumor-infiltrating immune cells from transcriptomics data.* Cancer Immunol Immunother, 2018.
25. Cieslik, M. and A.M. Chinnaiyan, *Cancer transcriptome profiling at the juncture of clinical translation.* Nat Rev Genet, 2018. **19**(2): p. 93-109.
26. Newman, A.M., et al., *Robust enumeration of cell subsets from tissue expression profiles.* Nat Methods, 2015. **12**(5): p. 453-7.
27. Gentles, A.J., et al., *The prognostic landscape of genes and infiltrating immune cells across human cancers.* Nat Med, 2015. **21**(8): p. 938-945.
28. Clancy, T. and E. Hovig, *Profiling networks of distinct immune-cells in tumors.* BMC Bioinformatics, 2016. **17**(1): p. 263.
29. Iglesia, M.D., et al., *Genomic Analysis of Immune Cell Infiltrates Across 11 Tumor Types.* J Natl Cancer Inst, 2016. **108**(11).
30. Thorsson, V., et al., *The Immune Landscape of Cancer.* Immunity, 2018. **48**(4): p. 812-830 e14.
31. Foroutan, M., et al., *Single sample scoring of molecular phenotypes.* BMC Bioinformatics, 2018. **19**(1): p. 404.

32. Tosolini, M., et al., *Assessment of tumor-infiltrating TCRVgamma9Vdelta2 gammadelta lymphocyte abundance by deconvolution of human cancers microarrays.* Oncoimmunology, 2017. **6**(3): p. e1284723.
33. Cancer Genome Atlas, N., *Genomic Classification of Cutaneous Melanoma.* Cell, 2015. **161**(7): p. 1681-96.
34. Linsley, P.S., et al., *Copy number loss of the interferon gene cluster in melanomas is linked to reduced T cell infiltrate and poor patient prognosis.* PLoS One, 2014. **9**(10): p. e109760.
35. Novershtern, N., et al., *Densely interconnected transcriptional circuits control cell states in human hematopoiesis.* Cell, 2011. **144**(2): p. 296-309.
36. Tirosh, I., et al., *Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq.* Science, 2016. **352**(6282): p. 189-96.
37. Sade-Feldman, M., et al., *Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma.* Cell, 2018. **175**(4): p. 998-1013 e20.
38. Gandolfo, L.C. and T.P. Speed, *RLE plots: Visualizing unwanted variation in high dimensional data.* PLoS One, 2018. **13**(2): p. e0191629.
39. Maetschke, S.R., et al., *Gene Ontology-driven inference of protein-protein interactions using inducers.* Bioinformatics, 2012. **28**(1): p. 69-75.
40. Butler, A., et al., *Integrating single-cell transcriptomic data across different conditions, technologies, and species.* Nat Biotechnol, 2018. **36**(5): p. 411-420.
41. Becht, E., et al., *Dimensionality reduction for visualizing single-cell data using UMAP.* Nat Biotechnol, 2018.
42. Huntington, N.D., et al., *IL-15 trans-presentation promotes human NK cell development and differentiation in vivo.* J Exp Med, 2009. **206**(1): p. 25-34.
43. Huntington, N.D., et al., *Innate lymphoid cells: parallel checkpoints and coordinate interactions with T cells.* Curr Opin Immunol, 2016. **38**: p. 86-93.
44. Liao, Y., G.K. Smyth, and W. Shi, *The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote.* Nucleic Acids Res, 2013. **41**(10): p. e108.
45. Law, C.W., et al., *voom: Precision weights unlock linear model analysis tools for RNA-seq read counts.* Genome Biol, 2014. **15**(2): p. R29.
46. McCarthy, D.J. and G.K. Smyth, *Testing significance relative to a fold-change threshold is a TREAT.* Bioinformatics, 2009. **25**(6): p. 765-71.
47. Barretina, J., et al., *The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity.* Nature, 2012. **483**(7391): p. 603-7.
48. Foroutan, M., et al., *A Transcriptional Program for Detecting TGFbeta-Induced EMT in Cancer.* Mol Cancer Res, 2017. **15**(5): p. 619-631.
49. McKinney, W. *Data structures for statistical computing in python*. in *Proceedings of the 9th Python in Science Conference*. 2010. van der Voort S, Millman J.
50. Luke, J.J., et al., *Targeted agents and immunotherapies: optimizing outcomes in melanoma.* Nat Rev Clin Oncol, 2017. **14**(8): p. 463-482.
51. Davoli, T., et al., *Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy.* Science, 2017. **355**(6322).
52. Huntington, N.D., *The unconventional expression of IL-15 and its role in NK cell homeostasis.* Immunol Cell Biol, 2014. **92**(3): p. 210-3.
53. Huntington, N.D., et al., *IL-15 transpresentation promotes both human T-cell reconstitution and T-cell-dependent antibody responses in vivo.* Proc Natl Acad Sci U S A, 2011. **108**(15): p. 6217-22.

54. Sade-Feldman, M., et al., *Resistance to checkpoint blockade therapy through inactivation of antigen presentation.* Nat Commun, 2017. **8**(1): p. 1136.
55. Zaretsky, J.M., et al., *Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma.* N Engl J Med, 2016. **375**(9): p. 819-29.
56. Pereira, C., et al., *Genomic Profiling of Patient-Derived Xenografts for Lung Cancer Identifies B2M Inactivation Impairing Immunorecognition.* Clin Cancer Res, 2017. **23**(12): p. 3203-3213.
57. Voskoboinik, I., J.C. Whisstock, and J.A. Trapani, *Perforin and granzymes: function, dysfunction and human pathology.* Nat Rev Immunol, 2015. **15**(6): p. 388-400.
58. Cursons, J., et al., *Combinatorial Targeting by MicroRNAs Co-ordinates Post-transcriptional Control of EMT.* Cell Syst, 2018.
59. Hugo, W., et al., *Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma.* Cell, 2016. **165**(1): p. 35-44.
60. Kemper, K., et al., *Phenotype switching: tumor cell plasticity as a resistance mechanism and target for therapy.* Cancer Res, 2014. **74**(21): p. 5937-41.
61. Behren, A., et al., *The Ludwig institute for cancer research Melbourne melanoma cell line panel.* Pigment Cell Melanoma Res, 2013. **26**(4): p. 597-600.
62. Widmer, D.S., et al., *Systematic classification of melanoma cells by phenotype-specific gene expression mapping.* Pigment Cell Melanoma Res, 2012. **25**(3): p. 343-53.
63. Jayachandran, A., et al., *Thrombospondin 1 promotes an aggressive phenotype through epithelial-to-mesenchymal transition in human melanoma.* Oncotarget, 2014. **5**(14): p. 5782-97.
64. Chaix, J., et al., *Cutting edge: Priming of NK cells by IL-18.* J Immunol, 2008. **181**(3): p. 1627-31.
65. Mezzadra, R., et al., *Identification of CMTM6 and CMTM4 as PD-L1 protein regulators.* Nature, 2017. **549**(7670): p. 106-110.
66. Imai, K., et al., *Natural cytotoxic activity of peripheral-blood lymphocytes and cancer incidence: an 11-year follow-up study of a general population.* Lancet, 2000. **356**(9244): p. 1795-9.
67. Sottile, R., et al., *HLA class I downregulation is associated with enhanced NK-cell killing of melanoma cells with acquired drug resistance to BRAF inhibitors.* Eur J Immunol, 2016. **46**(2): p. 409-19.
68. Tumeh, P.C., et al., *PD-1 blockade induces responses by inhibiting adaptive immune resistance.* Nature, 2014. **515**(7528): p. 568-71.
69. Andrews, M.C., et al., *Systems analysis identifies miR-29b regulation of invasiveness in melanoma.* Mol Cancer, 2016. **15**(1): p. 72.
70. Li, F.Z., et al., *Phenotype switching in melanoma: implications for progression and therapy.* Front Oncol, 2015. **5**: p. 31.
71. Samstein, R.M., et al., *Tumor mutational load predicts survival after immunotherapy across multiple cancer types.* Nat Genet, 2019.
72. Davis, A.A., et al., *Association of tumor mutational burden with smoking and mutation status in non-small cell lung cancer (NSCLC)*. 2017, American Society of Clinical Oncology.
73. Norum, J. and C. Nieder, *Tobacco smoking and cessation and PD-L1 inhibitors in non-small cell lung cancer (NSCLC): a review of the literature.* ESMO Open, 2018. **3**(6): p. e000406.

Tables

Table 1. Data used in this report. TCGA: The Cancer Genome Atlas; SKCM: Skin cutaneous melanoma; CCLE: Cancer Cell Line Encyclopedia

Resource	Data source & identifier	Reference
TCGA SKCM	NIH Genomic Data Commons: https://gdc.cancer.gov/	[33]
RNA-seq data for sorted immune cells	NCBI Gene Expression Omnibus: GSE60424 https://www.ncbi.nlm.nih.gov/geo/	[34]
Microarray data for sorted immune cells	NCBI Gene Expression Omnibus: GSE24759 https://www.ncbi.nlm.nih.gov/geo/	[35]
LM-MEL melanoma cell line panel	EBI Array Express: E-MTAB-1496 https://www.ebi.ac.uk/arrayexpress/	[61]
Single cell RNA-seq data from melanoma	NCBI Gene Expression Omnibus: GSE72056 https://www.ncbi.nlm.nih.gov/geo/	[36]
Single-cell RNA-seq data from melanoma	NCBI Gene Expression Omnibus: GSE120575 https://www.ncbi.nlm.nih.gov/geo/	[37]
CCLE RNA-seq data	https://portals.broadinstitute.org/ccle/data (requires free user registration)	[47]

Table 2. Covariate hazard coefficients for TCGA patients with metastatic melanoma. Variables from a Cox proportional hazards model (together with metastatic site) tested against the null hypothesis that the hazard coefficient is equal to 0. Metastatic sites were compared against unspecified lymph nodes for the baseline hazard and none were found to be significant; statistics and site groupings given in Table S2.

Variable	Coefficient mean (95% CI)	*p*-value
Age at diagnosis (*years*)	0.026 (0.016, 0.036)	6.21×10^{-7}
Gender (is female)	-0.207 (-0.527, 0.113)	0.20

Figure Legends

Figure 1: Hazard ratios associated with transcript abundance of individual genes. (**A**) A Cox proportional hazard model was created for each gene with patient age as the only co-variate. The top 50 genes were selected by significance and ranked by hazard coefficient (red dot, 95% confidence intervals shown with black lines). (**B**) Kaplan-Meier (KM) survival curves for patients with metastatic melanoma partitioned by age at diagnosis. (**C** - **F**) KM

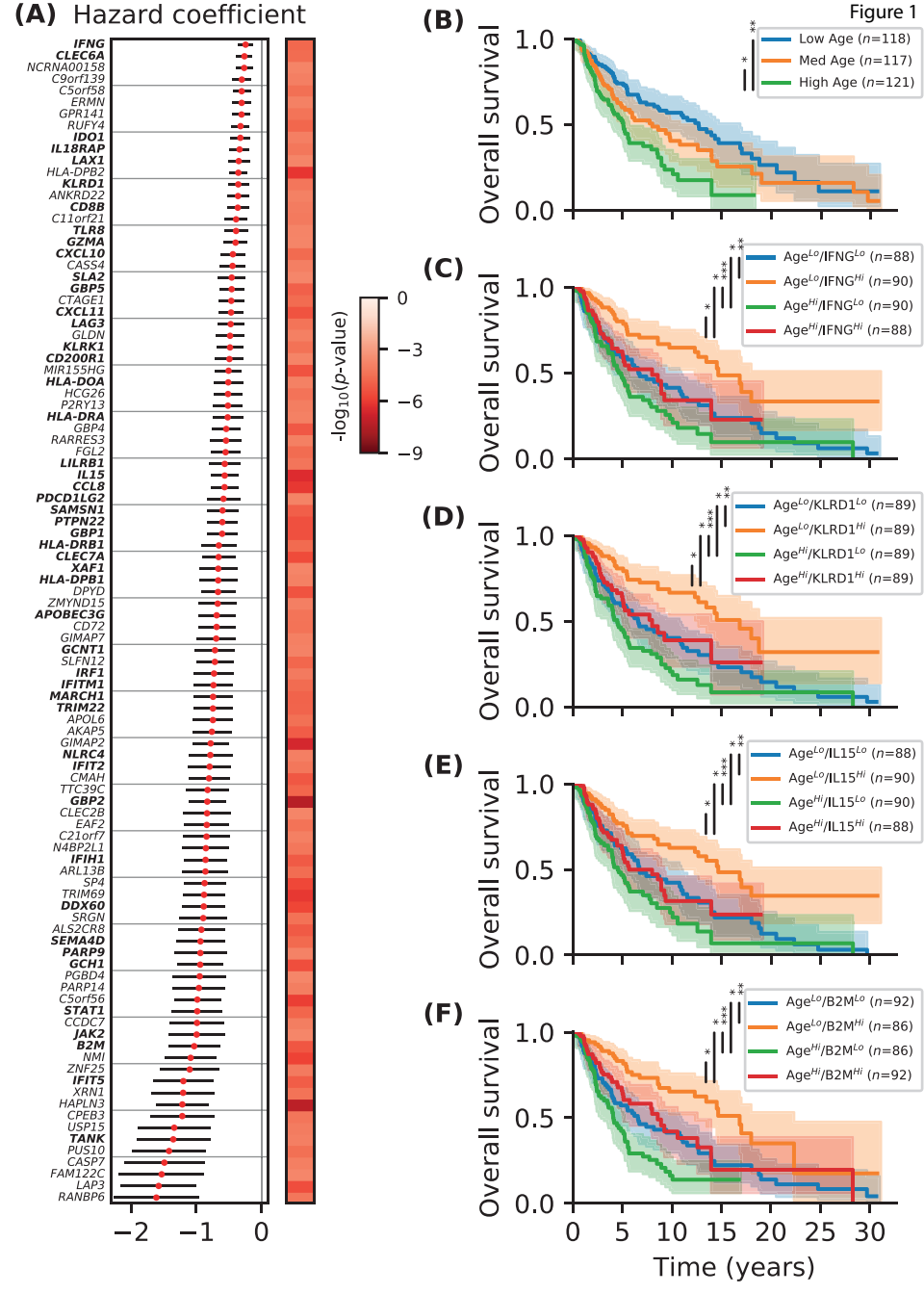
survival curves for patients partitioned by age and (C) *IFNG*, (D) *KLRD1*, (E) *IL15*, or (F) *B2M* transcript abundance. Survival curves differences were tested using a KM log rank test and significant comparisons are indicated (*: p-value < 0.05; **: p-value < $1x10^{-3}$; ***: p-value < $1x10^{-6}$).

Figure 2: A refined natural killer cell gene signature. Transcript abundance of NK cell marker genes across selected data with identified cell populations. (**A**) Metastatic melanoma single cell RNA-seq data (GSE72056), including NK cells (NK), macrophages (Macro.), B cells (B), unresolved/unidentified cells (Unres.); cancer associated fibroblasts (CAF), T cells (T) and endothelial cells (Endo.). (**B**) RNA-seq data from sorted immune populations (GSE60424). (**C**) Microarray data from sorted immune cell populations (GSE24759). NK: natural killer; DCs: dendritic cells; HSCs: hematopoietic stem cells.

Figure 3: Expression of NK cell markers across cellular sub-populations. Abundance of selected markers in $CD45^+$ cells from dissociated melanoma samples. Maximum expression for each gene is given in parentheses. For details on UMAP clustering refer to Materials and Methods.

Figure 4: Survival outcomes for TCGA SKCM patients. (**A**) Expression of NK marker genes within TCGA SKCM metastatic tumor samples sorted by NK score, together with the correlation of each gene against NK score. (**B**) Kaplan-Meier (KM) survival curves for patients separated by NK score and indicated genes (separated at 33^{rd} and 66^{th} percentiles). Survival curves differences were tested using a KM log rank test and significant comparisons are indicated (*: p-value < 0.05; **: p-value < $1x10^{-3}$; ***: p-value < $1x10^{-6}$). (**C**) KM survival curves for patients with high NK score tumors further separated by key cytokine (*XCL2*) and cytotoxic effector (*GZMB*) genes.

Figure 5: Melanoma tumors with a high NK score and evidence of a mesenchymal-like phenotype but low TGF-β activity show favorable patient outcomes. Associations across the TCGA SKCM metastatic tumor samples (r_P: Pearson's correlation; r_S: Spearman's correlation), between (**A**-**C**) NK score and scores associated with EMT/phenotype-switching, (**D**) mesenchymal score and a score of specific TGF-β induced EMT; and (**E**, **F**) NK score or TGF-β EMT score and age. (**G**, **H**) Kaplan-Meier (KM) survival curves for patients partitioned by TGF-β EMT score and NK score, and split by age. Survival curves differences were tested using a KM log rank test and significant comparisons are indicated (*: p-value < 0.05; **: p-value < $1x10^{-3}$; ***: p-value < $1x10^{-6}$).


(A) Hazard coefficient
IFNG
CLEC6A
NCRNA00158
C9orf139
C5orf58
ERMN
GPR141
RUFY4
IDO1
IL18RAP
LAX1
HLA-DPB2
KLRD1
ANKRD22
CD8B
C11orf21
TLR8
GZMA
CXCL10
CASS4
SLA2
GBP5
CTAGE1
CXCL11
LAG3
GLDN
KLRK1
CD200R1
MIR155HG
HLA-DOA
HCG26
P2RY13
HLA-DRA
GBP4
RARRES3
FGL2
LILRB1
IL15
CCL8
PDCD1LG2
SAMSN1
PTPN22
GBP1
HLA-DRB1
CLEC7A
XAF1
HLA-DPB1
DPYD
ZMYND15
APOBEC3G
CD72
GIMAP7
GCNT1
SLFN12
IRF1
IFITM1
MARCH1
TRIM22
APOL6
AKAP5
GIMAP2
NLRC4
IFIT2
CMAH
TTC39C
GBP2
CLEC2B
EAF2
C21orf7
N4BP2L1
IFIH1
ARL13B
SP4
TRIM69
DDX60
SRGN
ALS2CR8
SEMA4D
PARP9
GCH1
PGBD4
PARP14
C5orf56
STAT1
CCDC7
JAK2
B2M
NMI
ZNF25
IFIT5
XRN1
HAPLN3
CPEB3
USP15
TANK
PUS10
CASP7
FAM122C
LAP3
RANBP6
−2
−1
0
-log10(p-value)
0
−3
−6
−9
(B)
Low Age (n=118)
Med Age (n=117)
High Age (n=121)
Overall survival
1.0
0.5
0.0
(C)
Age^Lo/IFNG^Lo (n=88)
Age^Lo/IFNG^Hi (n=90)
Age^Hi/IFNG^Lo (n=90)
Age^Hi/IFNG^Hi (n=88)
(D)
Age^Lo/KLRD1^Lo (n=89)
Age^Lo/KLRD1^Hi (n=89)
Age^Hi/KLRD1^Lo (n=89)
Age^Hi/KLRD1^Hi (n=89)
(E)
Age^Lo/IL15^Lo (n=88)
Age^Lo/IL15^Hi (n=90)
Age^Hi/IL15^Lo (n=90)
Age^Hi/IL15^Hi (n=88)
(F)
Age^Lo/B2M^Lo (n=92)
Age^Lo/B2M^Hi (n=86)
Age^Hi/B2M^Lo (n=86)
Age^Hi/B2M^Hi (n=92)
0
5
10
15
20
25
30
Time (years)

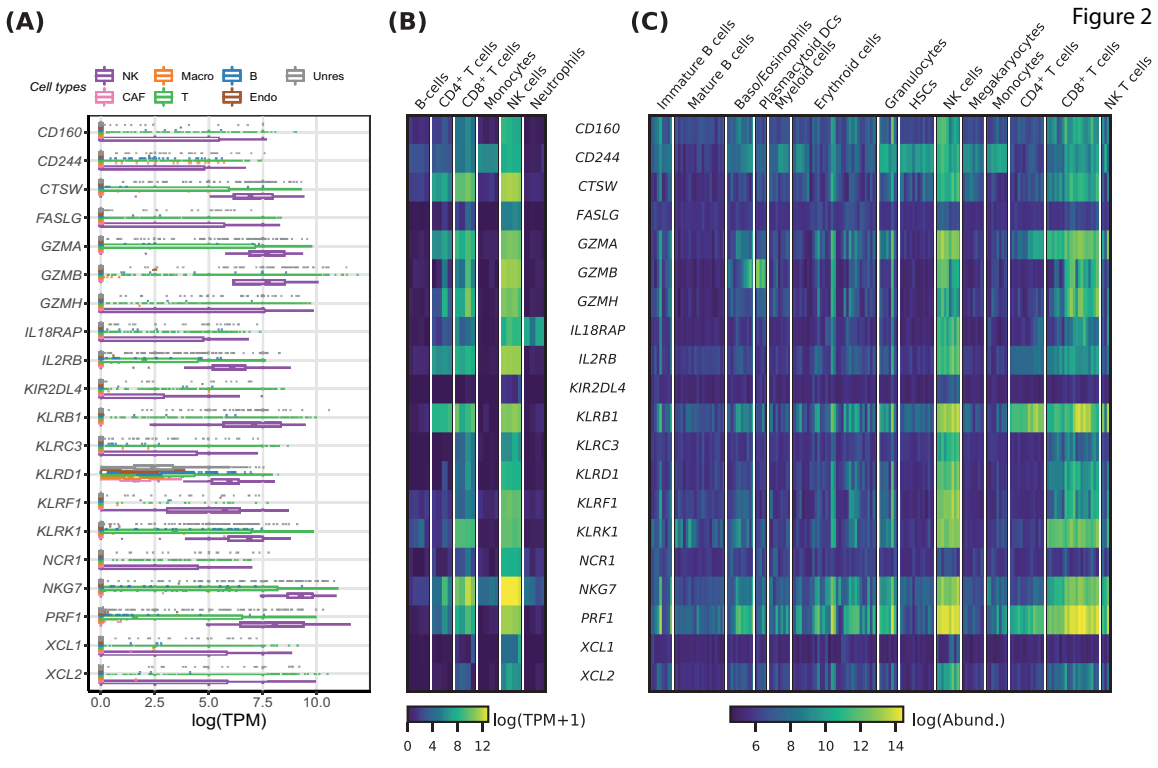
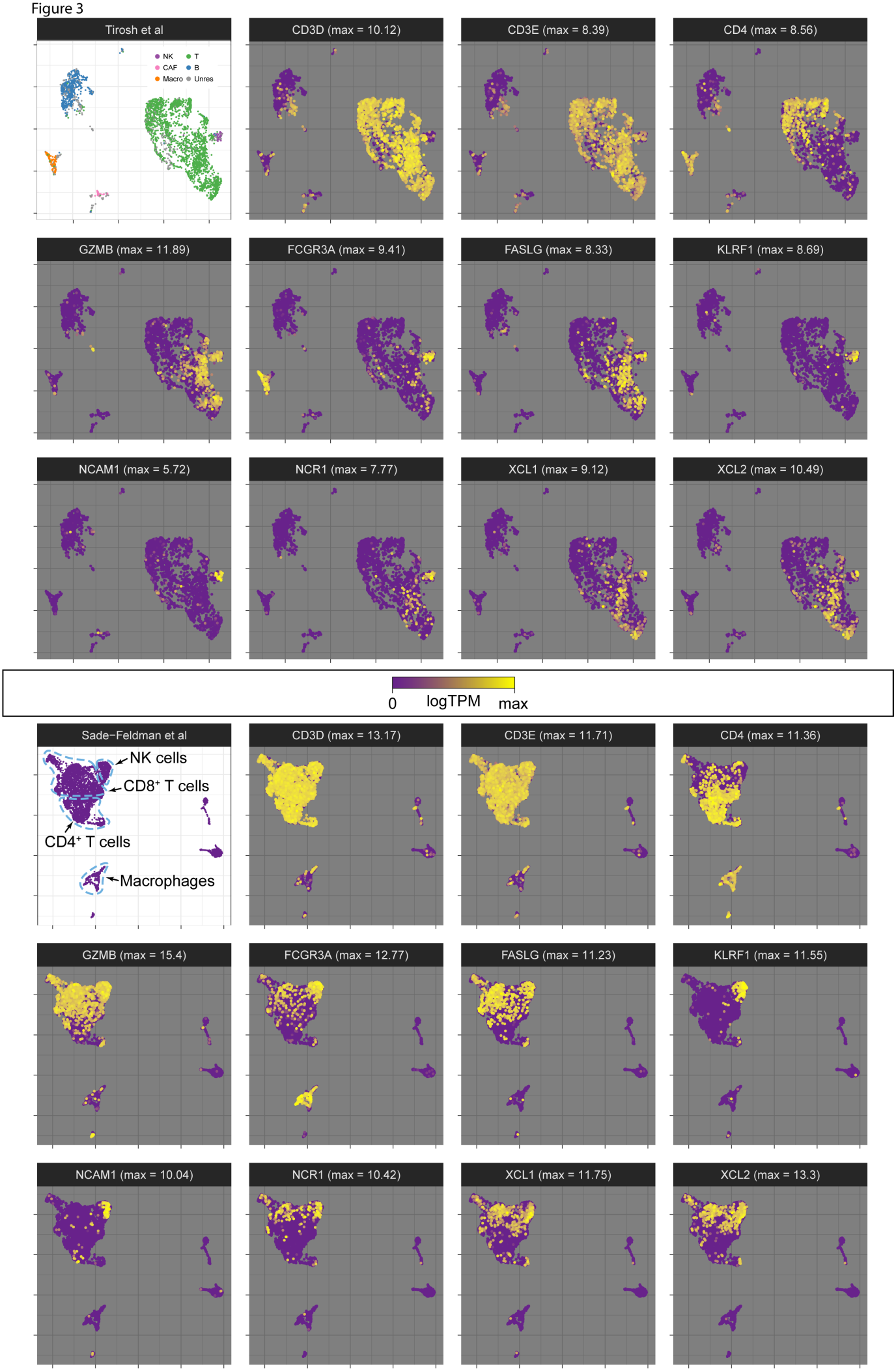

(A)
Cell types
NK
CAF
Macro
T
B
Endo
Unres
CD160
CD244
CTSW
FASLG
GZMA
GZMB
GZMH
IL18RAP
IL2RB
KIR2DL4
KLRB1
KLRC3
KLRD1
KLRF1
KLRK1
NCR1
NKG7
PRF1
XCL1
XCL2
0.0
2.5
5.0
7.5
10.0
log(TPM)
(B)
B-cells
CD4+ T cells
CD8+ T cells
Monocytes
NK cells
Neutrophils
log(TPM+1)
0
4
8
12
(C)
Immature B cells
Mature B cells
Baso/Eosinophils
Plasmacytoid DCs
Myeloid cells
Erythroid cells
Granulocytes
HSCs
NK cells
Megakaryocytes
Monocytes
CD4+ T cells
CD8+ T cells
NK T cells
log(Abund.)
6
8
10
12
14

Figure 3


Tirosh et al
NK
T
CAF
B
Macro
Unres
CD3D (max = 10.12)
CD3E (max = 8.39)
CD4 (max = 8.56)
GZMB (max = 11.89)
FCGR3A (max = 9.41)
FASLG (max = 8.33)
KLRF1 (max = 8.69)
NCAM1 (max = 5.72)
NCR1 (max = 7.77)
XCL1 (max = 9.12)
XCL2 (max = 10.49)
0
logTPM
max
Sade−Feldman et al
NK cells
CD8⁺ T cells
CD4⁺ T cells
Macrophages
CD3D (max = 13.17)
CD3E (max = 11.71)
CD4 (max = 11.36)
GZMB (max = 15.4)
FCGR3A (max = 12.77)
FASLG (max = 11.23)
KLRF1 (max = 11.55)
NCAM1 (max = 10.04)
NCR1 (max = 10.42)
XCL1 (max = 11.75)
XCL2 (max = 13.3)

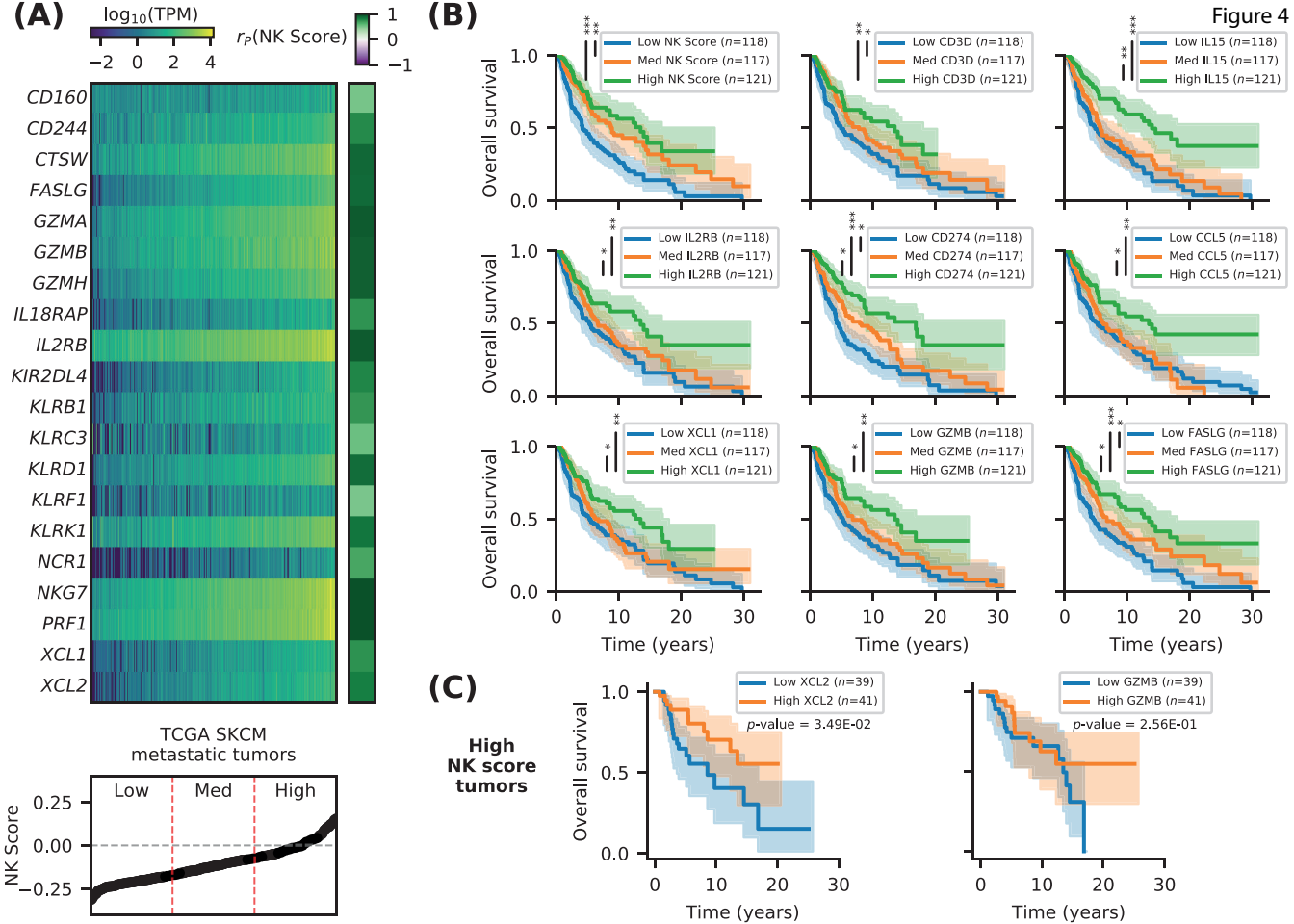
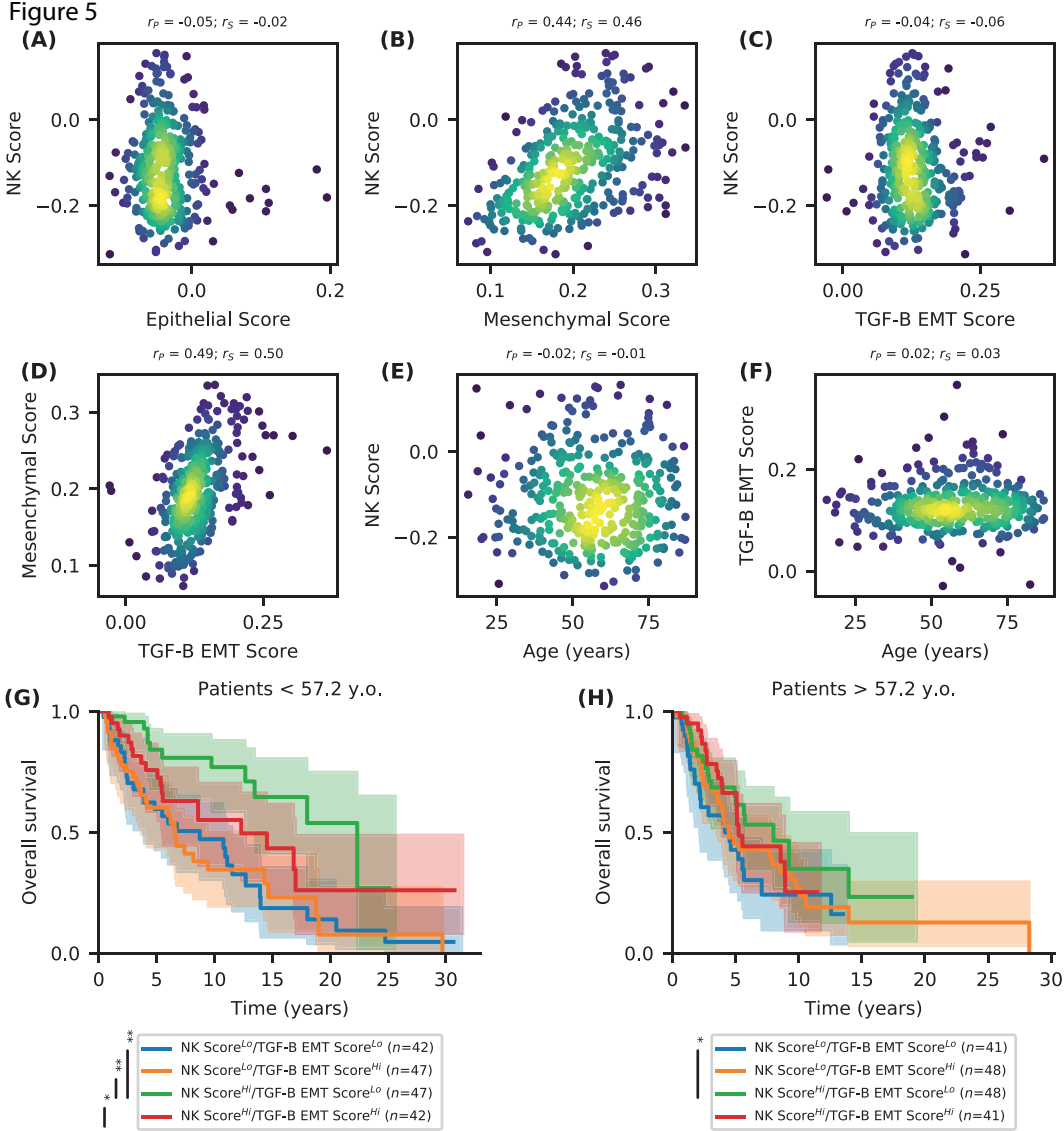

(A)
log10(TPM)
−2 0 2 4
rP(NK Score)
1 0 −1
CD160
CD244
CTSW
FASLG
GZMA
GZMB
GZMH
IL18RAP
IL2RB
KIR2DL4
KLRB1
KLRC3
KLRD1
KLRF1
KLRK1
NCR1
NKG7
PRF1
XCL1
XCL2
TCGA SKCM
metastatic tumors
Low
Med
High
NK Score
0.25
0.00
−0.25
(B)
Low NK Score (n=118)
Med NK Score (n=117)
High NK Score (n=121)
Low CD3D (n=118)
Med CD3D (n=117)
High CD3D (n=121)
Low IL15 (n=118)
Med IL15 (n=117)
High IL15 (n=121)
Low IL2RB (n=118)
Med IL2RB (n=117)
High IL2RB (n=121)
Low CD274 (n=118)
Med CD274 (n=117)
High CD274 (n=121)
Low CCL5 (n=118)
Med CCL5 (n=117)
High CCL5 (n=121)
Low XCL1 (n=118)
Med XCL1 (n=117)
High XCL1 (n=121)
Low GZMB (n=118)
Med GZMB (n=117)
High GZMB (n=121)
Low FASLG (n=118)
Med FASLG (n=117)
High FASLG (n=121)
Overall survival
Time (years)
(C)
High NK score tumors
Low XCL2 (n=39)
High XCL2 (n=41)
p-value = 3.49E-02
Low GZMB (n=39)
High GZMB (n=41)
p-value = 2.56E-01
Overall survival
Time (years)

Figure 5


(A)
r_P = -0.05; r_S = -0.02
NK Score
Epithelial Score
(B)
r_P = 0.44; r_S = 0.46
NK Score
Mesenchymal Score
(C)
r_P = -0.04; r_S = -0.06
NK Score
TGF-B EMT Score
(D)
r_P = 0.49; r_S = 0.50
Mesenchymal Score
TGF-B EMT Score
(E)
r_P = -0.02; r_S = -0.01
NK Score
Age (years)
(F)
r_P = 0.02; r_S = 0.03
TGF-B EMT Score
Age (years)
(G)
Patients < 57.2 y.o.
Overall survival
Time (years)
NK Score^{Lo}/TGF-B EMT Score^{Lo} (n=42)
NK Score^{Lo}/TGF-B EMT Score^{Hi} (n=47)
NK Score^{Hi}/TGF-B EMT Score^{Lo} (n=47)
NK Score^{Hi}/TGF-B EMT Score^{Hi} (n=42)
(H)
Patients > 57.2 y.o.
Overall survival
Time (years)
NK Score^{Lo}/TGF-B EMT Score^{Lo} (n=41)
NK Score^{Lo}/TGF-B EMT Score^{Hi} (n=48)
NK Score^{Hi}/TGF-B EMT Score^{Lo} (n=48)
NK Score^{Hi}/TGF-B EMT Score^{Hi} (n=41)