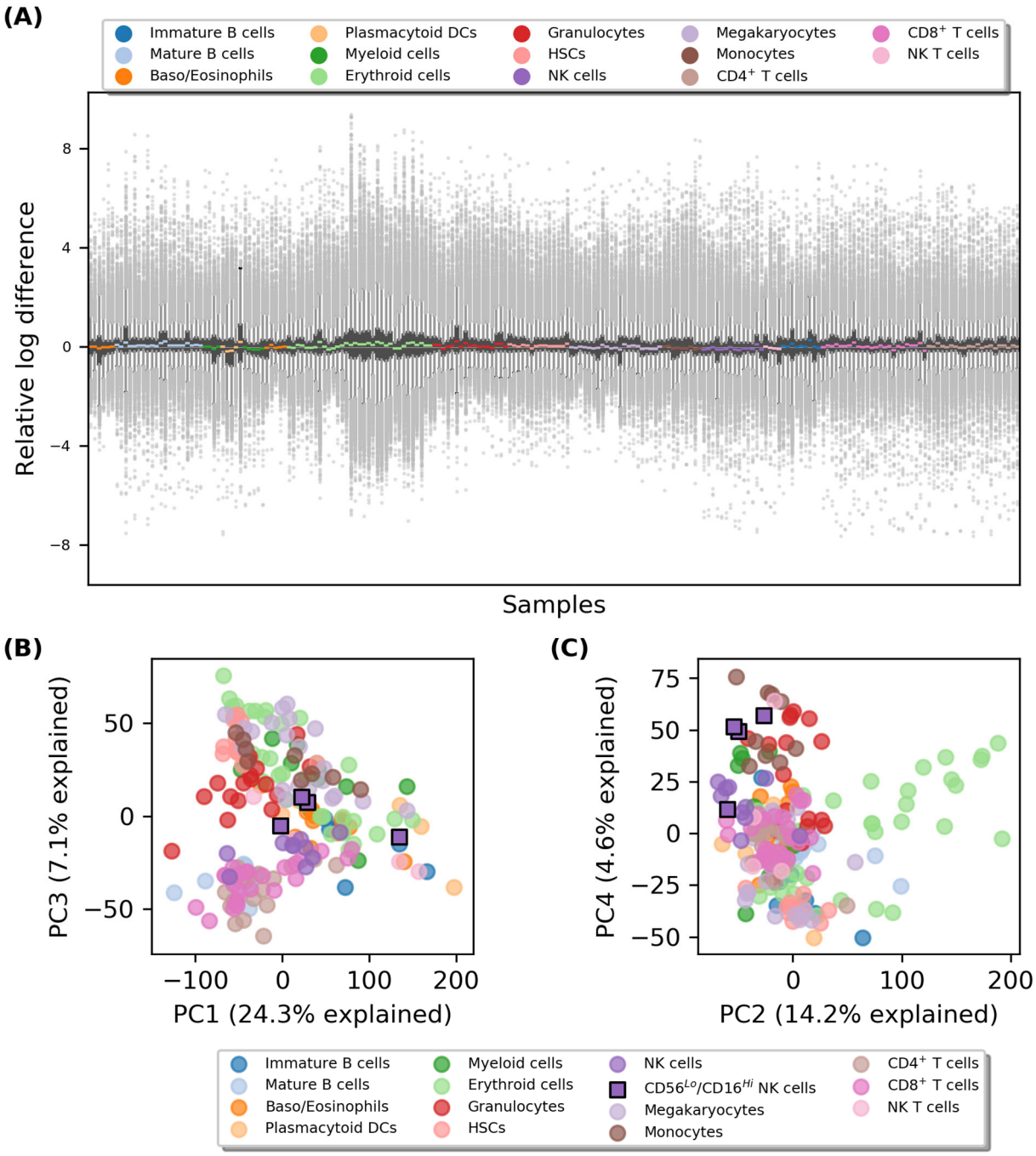
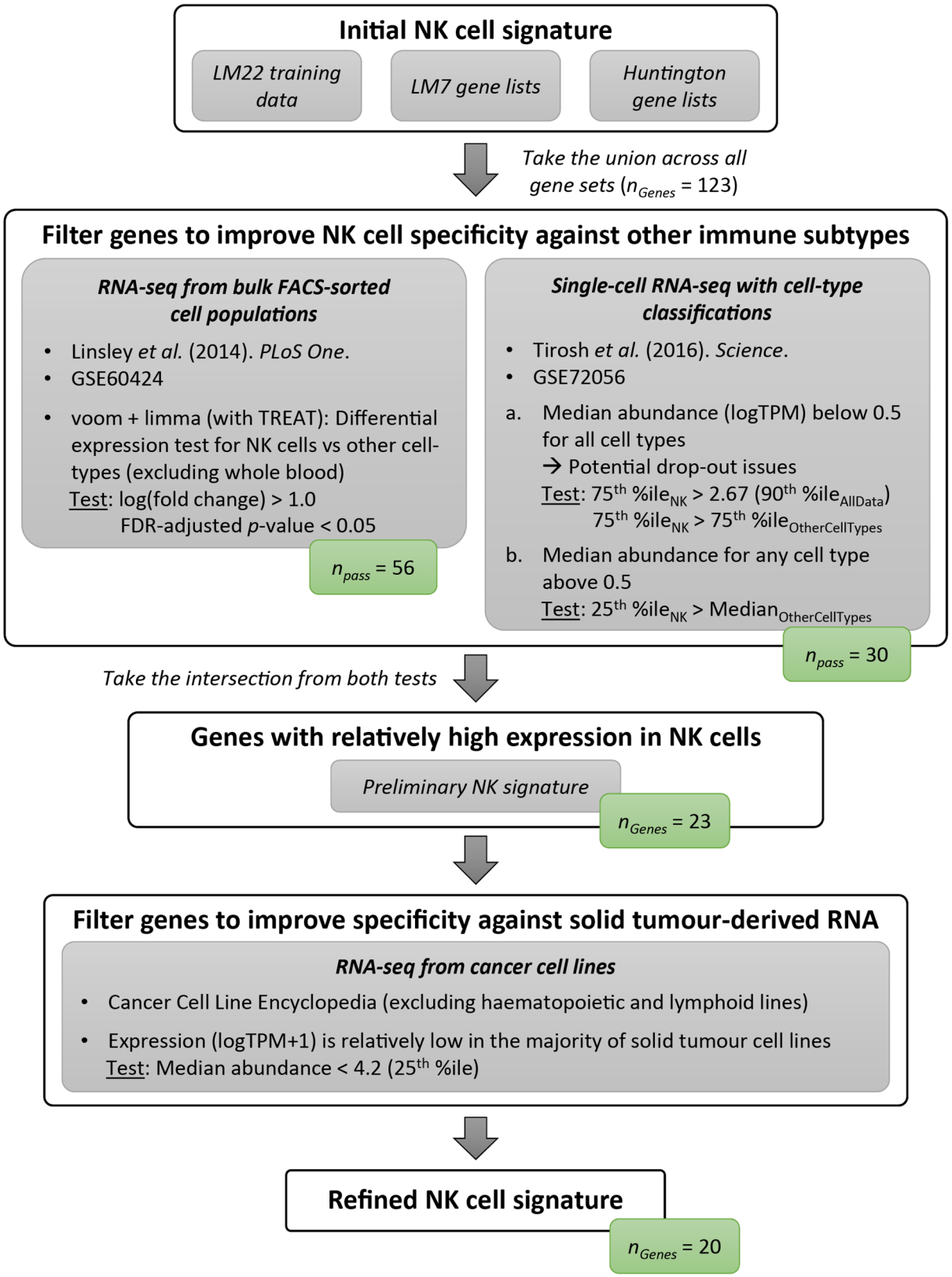
Supplementary Material for Cursons & Souza Fonseca Guimaraes et al.



Supplementary Figure S1: Outlier CD56^{-}/CD16^{+}/CD3^{-} NK cell samples in GSE24759. (**A**) Relative log expression (RLE) plot for all peripheral blood-derived samples included in the microarray data GSE24759, with samples grouped by cell type. (**B**, **C**) principal component analysis of peripheral blood-derived samples included in the microarray data GSE24759, with samples grouped by cell type. As shown (purple square markers with black outline), a subset of CD56^{Lo}/CD16^{Hi}/CD3^{Lo} NK cells appear to separate from the remaining NK cell samples, and thus, these samples were excluded from further analyses.



Supplementary Figure S2: Schematic workflow of natural killer cell gene signature curation. For further details please refer to Materials & Methods/Natural killer cell signature curation. The curated NK signature genes and a subset of test results are given in Supplementary Table S1; further details are given in TableS1.xlsx.

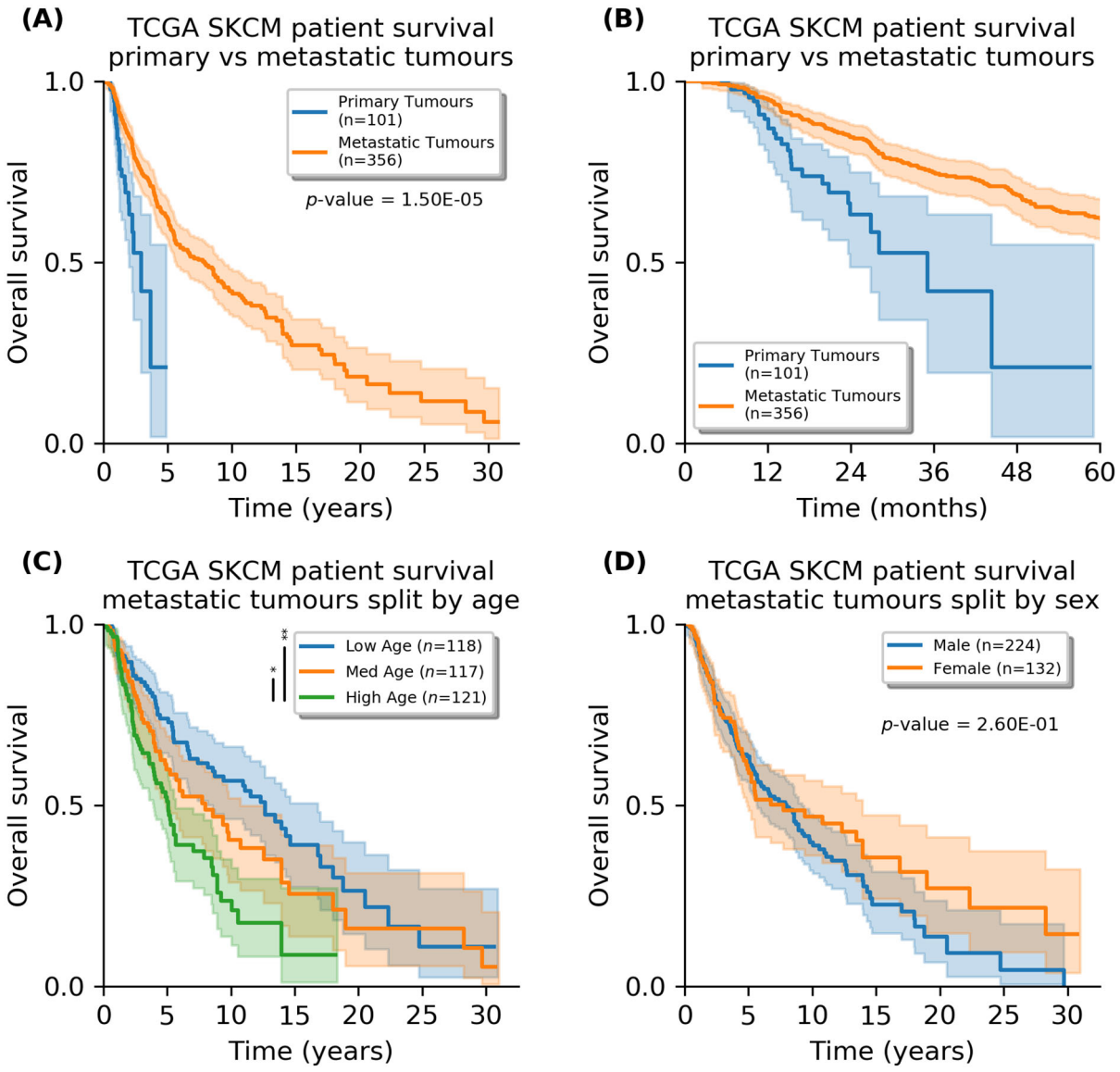
Supplementary Table S2: Cox proportional hazard model coefficients. (**A**) Covariate hazard coefficients for TCGA patients with metastatic melanoma (including site-specific hazard). Male patients were used for the base hazard (*isFemale* reflects the relative hazard associated with patient sex). Given the study of metastatic melanoma, the 'Lymph nodes, NOS' site (corresponding to unspecified lymph nodes) was used for the base hazard group. (**B**) Groupings of metastatic sites from the TCGA SKCM cohort.

(**A**) Cox proportional hazards model coefficients

	Hazard coefficient	exp(Hazard coefficient)	Standard error	z	p-value	lower 0.95	upper 0.95
Age (years)	0.026	1.026	0.005	4.984	6.21E-07	0.016	0.036
isFemale	-0.207	0.813	0.163	-1.266	0.205	-0.527	0.113
Epithelial tissue	0.114	1.120	0.256	0.444	0.657	-0.388	0.616
Internal organ	0.256	1.291	0.263	0.973	0.331	-0.259	0.771
Lung	0.164	1.178	0.288	0.569	0.569	-0.400	0.727
Connective	-0.168	0.845	0.263	-0.641	0.521	-0.683	0.346
Internal, other	0.039	1.040	0.378	0.103	0.918	-0.701	0.779
CNS	0.239	1.270	0.550	0.435	0.664	-0.838	1.316
Lymph nodes of head, face and neck	-0.216	0.806	0.491	-0.440	0.660	-1.179	0.747
Lymph nodes of inguinal region or leg	0.114	1.121	0.430	0.265	0.791	-0.729	0.957
Lymph nodes of axilla or arm	-0.848	0.428	1.022	-0.829	0.407	-2.851	1.156
Skin	0.340	1.404	1.032	0.329	0.742	-1.684	2.363

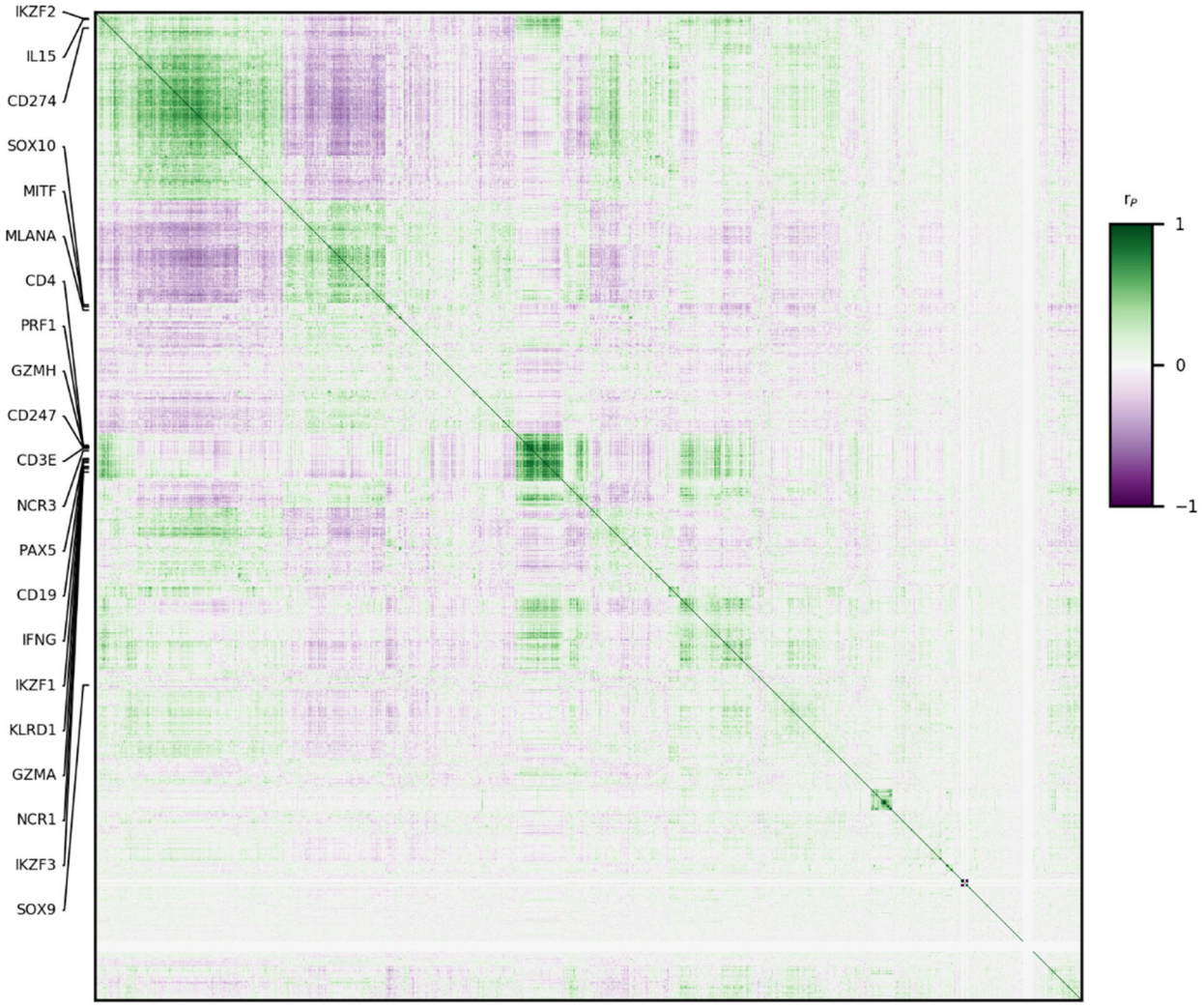
(**B**) Metastatic tumor site groupings

Metastatic site grouping	TCGA SKCM annotated metastatic site
Lymph nodes of axilla or arm (n=90)	"Lymph nodes of axilla or arm" (n=90)
Epithelial tissue (n=11)	"Small intestine, NOS" (n=8), "Vulva, NOS" (n=1), "Vagina, NOS" (n=1), "Colon, NOS" (n=1)
Lymph nodes of head, face and neck (n=19)	"Lymph nodes of head, face and neck" (n=19)
CNS (n=5)	"Spinal cord" (n=1), "Parietal lobe" (n=1), "Brain, NOS" (n=1), "Frontal lobe" (n=2)
Internal organ (n=5)	"Parotid gland" (n=1), "Liver" (n=1), "Adrenal gland, NOS" (n=2), "Spleen" (n=1)
Connective (n=65)	"Connective, subcutaneous and other soft tissues, NOS" (n=13), "Connective, subcutaneous and other soft tissues of lower limb and hip" (n=10), "Connective, subcutaneous and other soft tissues of upper limb and shoulder" (n=9), "Connective, subcutaneous and other soft tissues of thorax" (n=11), "Connective, subcutaneous and other soft tissues of trunk, NOS" (n=11), "Connective, subcutaneous and other soft tissues of head, face, and neck" (n=4), "Connective, subcutaneous and other soft tissues of pelvis" (n=5), "Connective, subcutaneous and other soft tissues of abdomen" (n=2)
Skin (n=51)	"Skin of trunk" (n=21), "Skin of lower limb and hip" (n=4), "Skin of upper limb and shoulder" (n=7), "Skin, NOS" (n=16), "Skin of scalp and neck" (n=3)
Lung (n=8)	"Upper lobe, lung" (n=3), "Lower lobe, lung" (n=4), "Lung, NOS" (n=1)
Internal, other (n=4)	"Thorax, NOS" (n=1), "Abdomen, NOS" (n=1), "Pelvis, NOS" (n=1), "Peritoneum, NOS" (n=1)
Lymph nodes of inguinal region or leg (n=55)	"Lymph nodes of inguinal region or leg" (n=48), "Pelvic lymph nodes" (n=7)
Lymph node, NOS (n=43)	"Lymph node, NOS" (n=42), "Intra-abdominal lymph nodes" (n=1)

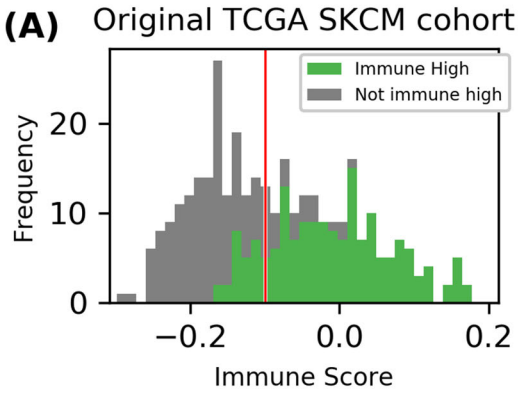


Supplementary Figure S3: Survival functions for TCGA SKCM patients. Kaplan-Meier (KM) survival functions (with 95% confidence intervals) for subsets of the TCGA skin cutaneous melanoma patient cohort; showing (**A**, **B**) survival of patients with either a primary or metastatic lesion (one patient with both tumor types was excluded) over (A) the full duration of the study and (B) the first 5 years after diagnosis; (**C**) survival of patients with a metastatic tumor, partitioned by age (at the 33rd and 66th 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); (**D**) survival of patients with a metastatic tumor, partitioned by sex.

Reduced patient survival for patients with primary tumors over the first 3 years is unexpected. It may be possible that recruitment of patients into the metastatic tumor cohort implies a relatively accessible metastatic tumor site with less deleterious effects, whereas patients within the primary tumor cohort may have carried an undiagnosed metastatic tumor burden within deeper, and potentially more dangerous locations.

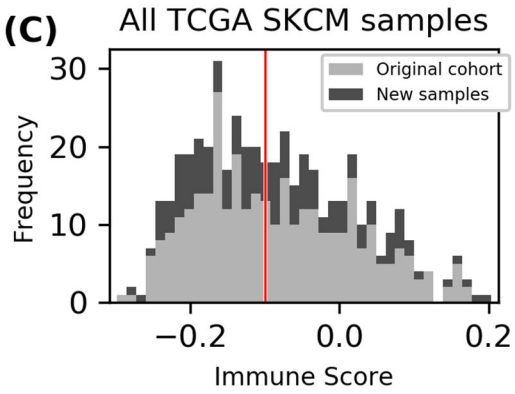


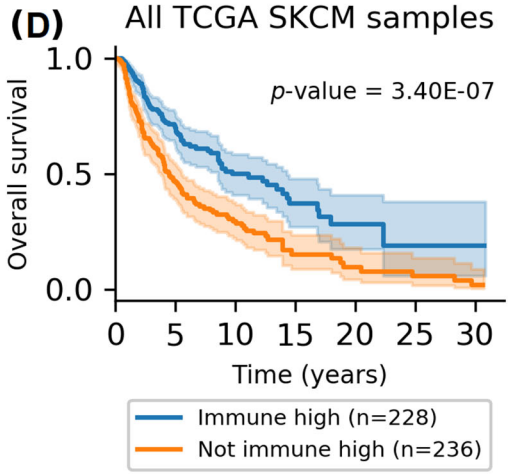
Supplementary Figure S4. Gene-gene cross-correlation across the TCGA SKCM transcriptomic data. A clustered heat map of the gene-gene Pearson's correlation (r_P) across the TCGA skin cutaneous melanoma data. A common criticism of immune deconvolution methods is that the inferred immune cell subsets are highly cross-correlated. While some of this reflects the fact that many immune marker genes are not unique to an individual subset, there are a number of immune gene subsets which are highly cross-correlated due to the nature of immune cell recruitment, where one population secretes the necessary chemokines and cytokines to effectively recruit subsequent sub-populations. Selected genes associated with melanoma biology or immune cell sub-populations are shown.



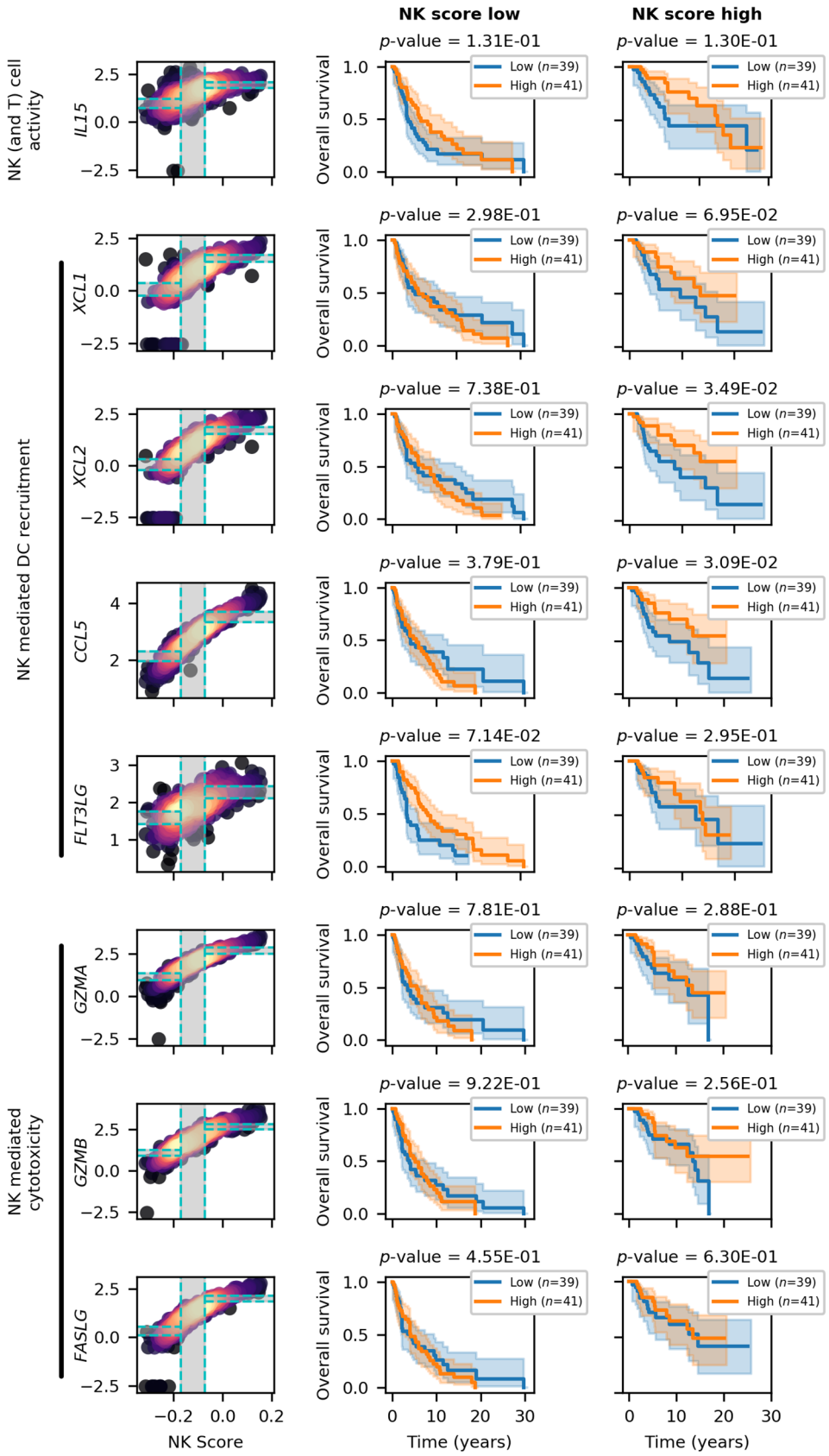
(B)

	Score Low	Score High	Total
TCGA Imm. Low	140	24	164
TCGA Imm. High	27	141	168
Total	167	165	

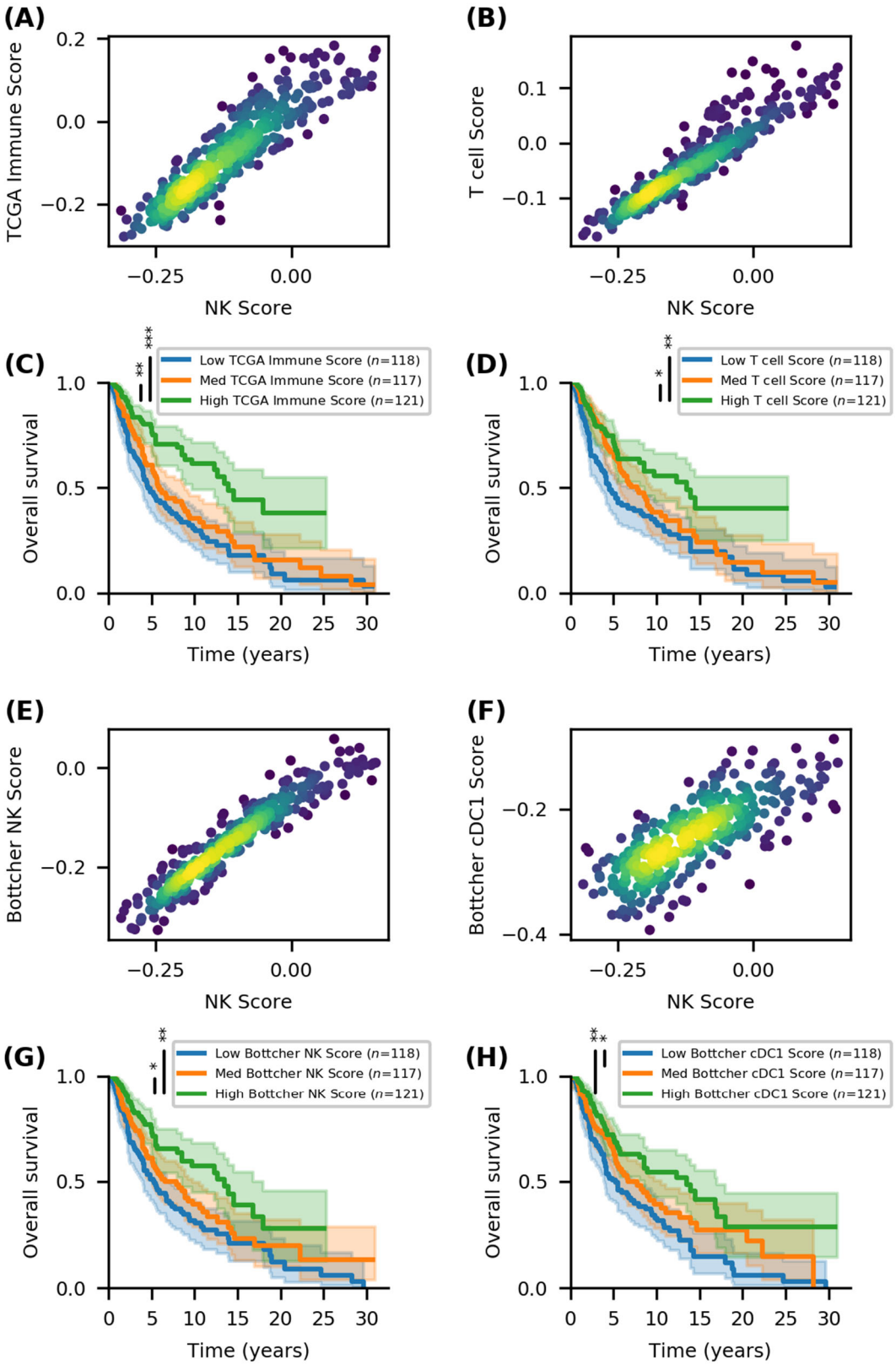




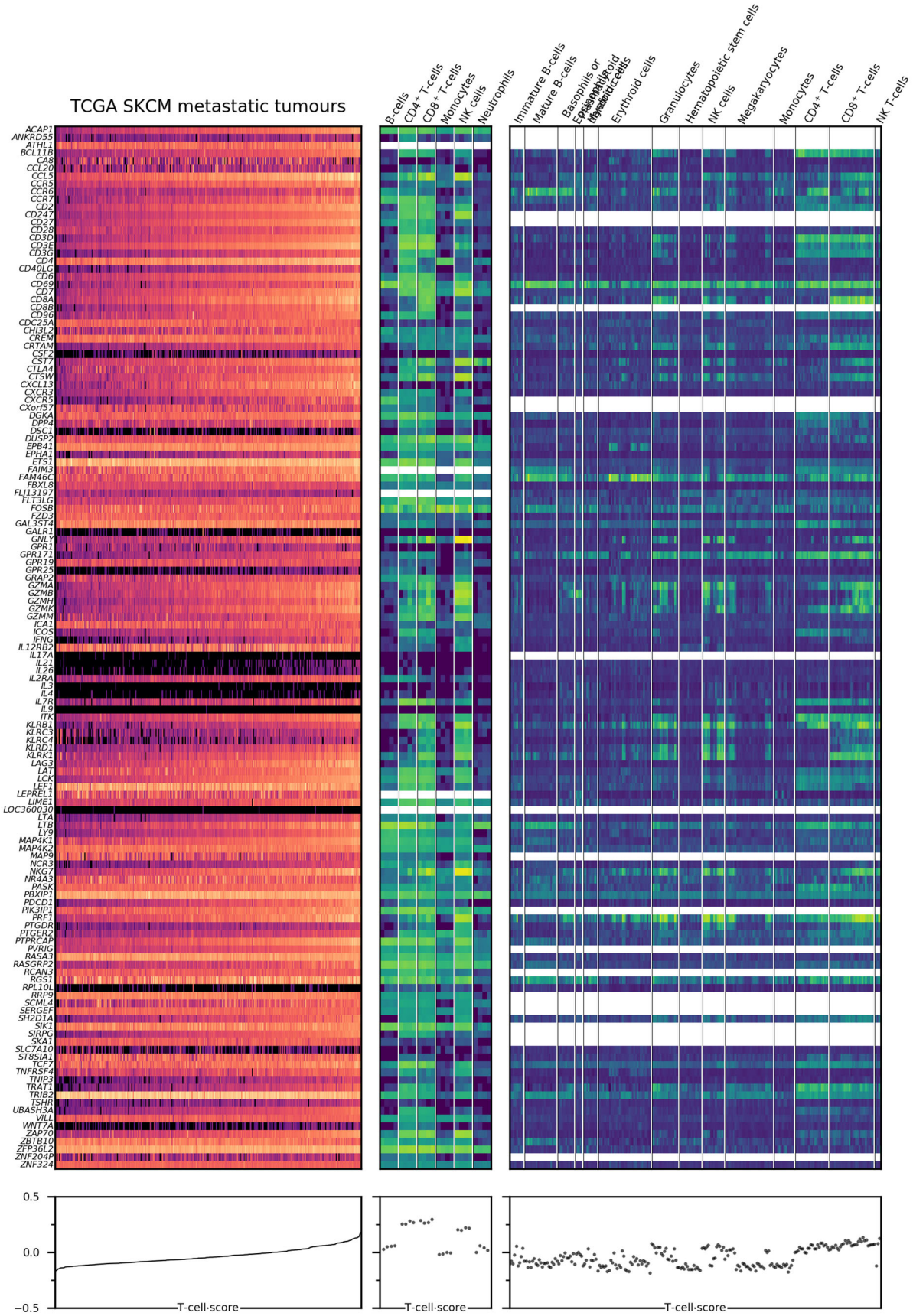
Supplementary Figure S5: A comparison of patient partitions between original TCGA clustering and gene set scoring. Within the original TCGA SCKM manuscript an immune gene cluster was identified, and patients with high expression of this immune cluster were shown to have improved survival outcomes. (**A**) Using the *singscore* method immune genes identified within the TCGA SKCM paper were scored across all samples and those with an 'Immune high' annotation are shown. (**B**) A contingency table listing the number of patients assigned to each class relative to original TCGA immune classifications with the threshold of -0.1 as shown in (A). (**C**) The immune score for all TCGA SKCM data with new samples shown. (**D**) Kaplan-Meier survival functions across all samples separated by Immune score with a threshold of -0.1 (as shown in (A) and (C)).



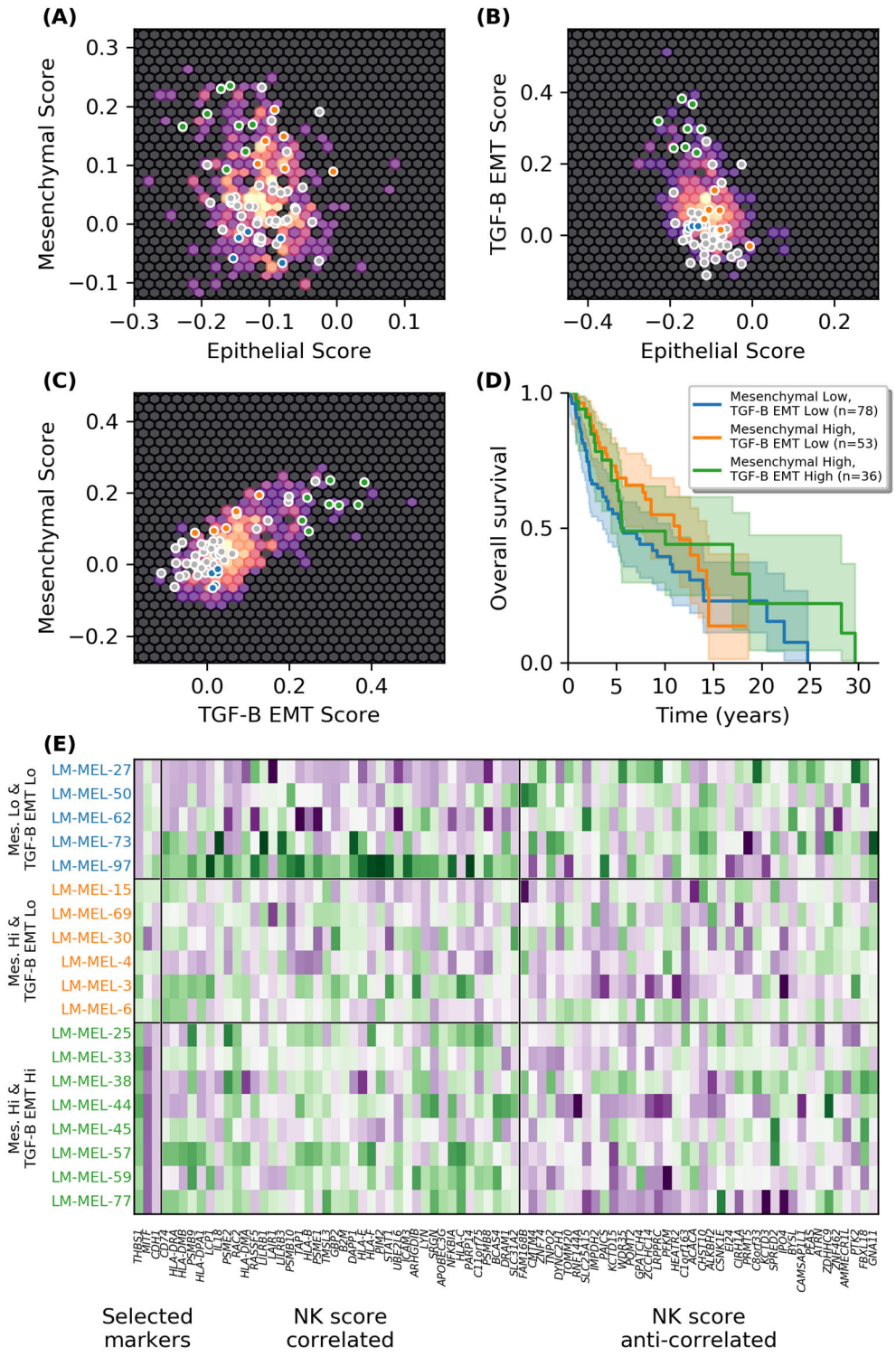
Supplementary Figure S6. Survival effects of genes associated with NK function. TCGA SKCM tumor samples within the bottom 33% or top 33% of samples ranked by NK score (Fig. 4A) were further partitioned by selected genes associated with NK function (top vs bottom 33%; *i.e.* shaded regions were excluded). Kaplan-Meier survival functions for each corresponding patient subset are shown together with an unadjusted *p*-value (calculated using Kaplan-Meier log rank test)



Supplementary Figure S7. Associations between our NK score and other immune-related gene scores and corresponding survival curves. (**A, B, E, F**) Associations between NK score and various immune signature gene scores across TCGA SKCM metastatic tumor samples, and (**C**, **D**, **G**, **H**) Kaplan-Meier (KM) survival functions across the corresponding patient subsets. 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).



Supplementary Figure S8. T cell signature genes. Transcript abundance of T cell marker genes across TCGA SKCM metastatic tumors ranked by T cell score (*at left*); RNA-seq data for sorted immune cell populations (GSE60424; *at center*); microarray data for sorted immune cell populations (GSE24759; *at right*). Note some genes could not be directly mapped due to changes in gene symbol (blank rows).



Supplementary Figure S9. Exploring melanoma-expressed transcripts associated with increased or decreased NK score. (**A, B, C**) Relative distributions of selected phenotype associated gene sets across TCGA SKCM metastatic tumor samples (*hexbin density plots*) overlaid with scores for LM-MEL cell lines (only cell lines shown in (E) are colored), (**D**) together with survival functions of corresponding TCGA patients subset by mesenchymal scores and TGF-β EMT score. (**E**) The relative abundance of genes (purple = low; green = high) selected by their correlation with/against NK score for a subset of LM-MEL cell lines subset by mesenchymal score and TGF-β EMT scores.