PNAS
www.pnas.org

Supplementary Information for

Cancer stemness, intratumoral heterogeneity, and immune response across cancers

Alex Miranda, Phineas T Hamilton, Allen W Zhang, Swetansu Pattnaik, Etienne Becht, Artur Mezheyeuski, Jarle Bruun, Patrick Micke, Aurélien De Reynies, Brad H Nelson

Brad H Nelson
Email: bnelson@bccrc.ca

This PDF file includes:

Key Resources Table
Figs. S1 to S9
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Datasets S1 to S2

www.pnas.org/cgi/doi/10.1073/pnas.1818210116

Key Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Processed sequencing and clinical data		Accessed through TCGABiolinks
Cancer Cell Line Encyclopedia	Barretina et al. 2012 (1)	https://portals.broadinstitute.org/ccle
EBI (E-MTAB-6149)	Lambrechts et al.2018 (2)	https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6149/
GEO GSE30652	Nazor et al. 2012 (3) (Used on Malta et al. 2018)	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30652
GEO GSE39582	Marisa et al. 2013 (4)	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39582
GEO GSE81089	Mezheyeuski et al. 2018 (5)	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE81089
GEO GSE15192	Bhat-Nakshatri et al. 2010 (6)	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15192
GEO GSE31257	Jung et al., 2011 (7)	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31257
GEO GSE76009	Ng et al., 2016 (8)	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE76009
Software and Algorithms		
R 3.4.2		https://www.r-project.org/
xCell	Aran (9) et al. 2017	https://github.com/dviraran/xCell
GSVA	Hanzelmann et al. 2013 (10)	https://www.bioconductor.org/packages/release/bioc/html/GSVA.html
biomaRt	Durinck et al. 2009 (11)	https://bioconductor.org/packages/release/bioc/html/biomaRt.html

TCGABiolinks	Colaprico et al. 2015 (12)	https://www.bioconductor.org/packages/release/bioc/html/TCGAbiolinks.html
lme4	Bates et al. 2018 (13)	https://cran.r-project.org/web/packages/lme4/index.html
survival	Therneau 2018 (14)	https://cran.r-project.org/web/packages/survival/survival.pdf
limma	Richie et al. 2015 (15)	https://bioconductor.org/packages/release/bioc/html/limma.html
reactomePA	Yu et al. 2016 (16)	https://www.bioconductor.org/packages/release/bioc/html/ReactomePA.html
genefilter	Gentleman et al. 2018 (17)	https://bioconductor.org/packages/release/bioc/html/genefilter.html
Circos	Krzywinski et al 2009 (18)	http://circos.ca/

Supplementary Figures

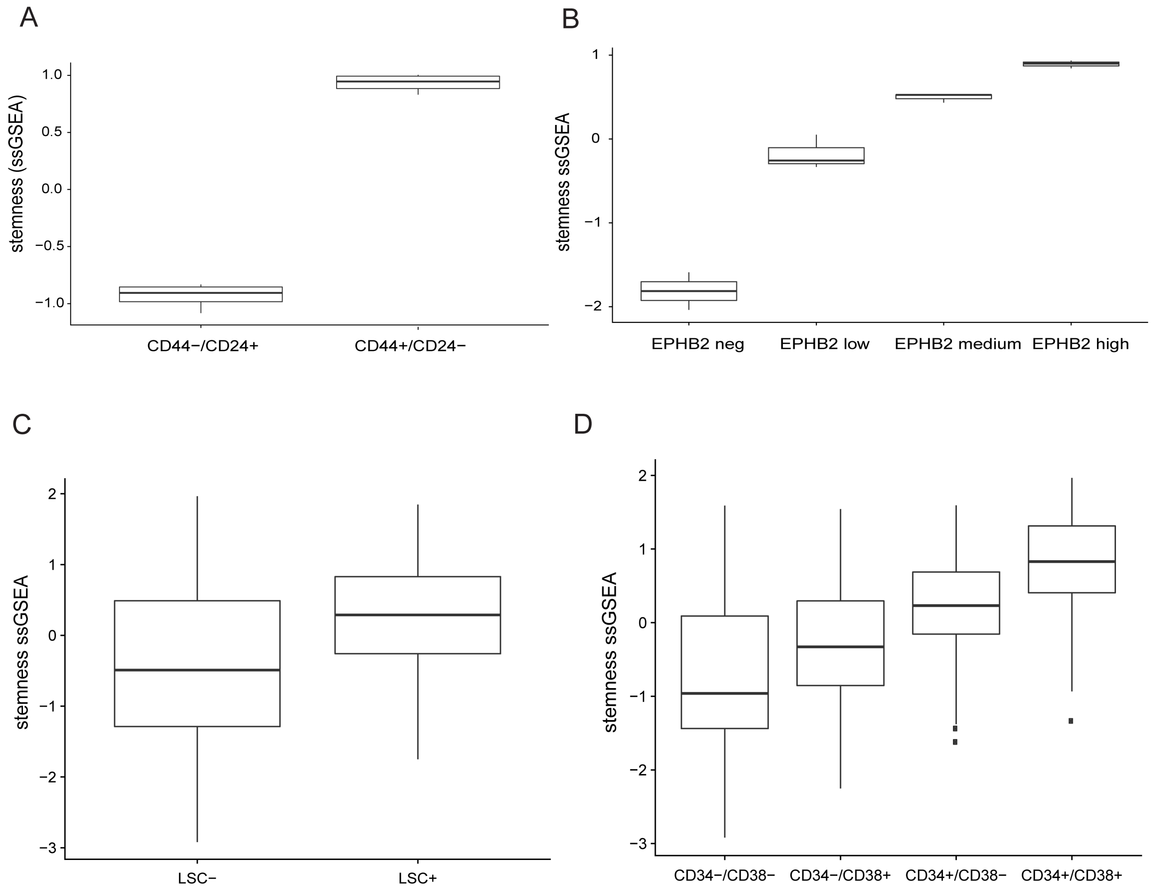

Fig. S1. Stemness scores in validation datasets. **A)** Breast epithelial stem cell, CD44^{-}/CD24^{+} n = 4; Breast epithelial differentiated cell CD44^{+}/CD24^{-}, n = 4. (data from GSE15192) **B)** Intestinal cells with different degree of differentiation. The highest EPHB2 surface levels correspond to epithelial colonic cells with the longest telomeres and elevated expression of intestinal stem cell marker genes. EPHB2-high, n=3; EPHB2-medium, n=3; EPHB2-low, n=3; EPHB2-negative, n=2. (data from GSE31257). **C)** LSC^{+} leukemia stem cells positive fraction, n=138; LSC^{-} negative fraction, n=89. (data from GSE76009) **D)** Leukemia cell fractions with different contents of stem cells. The majority of CD34+ and a minority of CD34− fractions contain leukemia stem cells. CD34^{-}/CD38^{-} , n=48; CD34^{-}/CD38^{+}, n=69; CD34^{+}/CD38^{-} , n=54; CD34^{+}/CD38^{+}. (data from GSE76009).

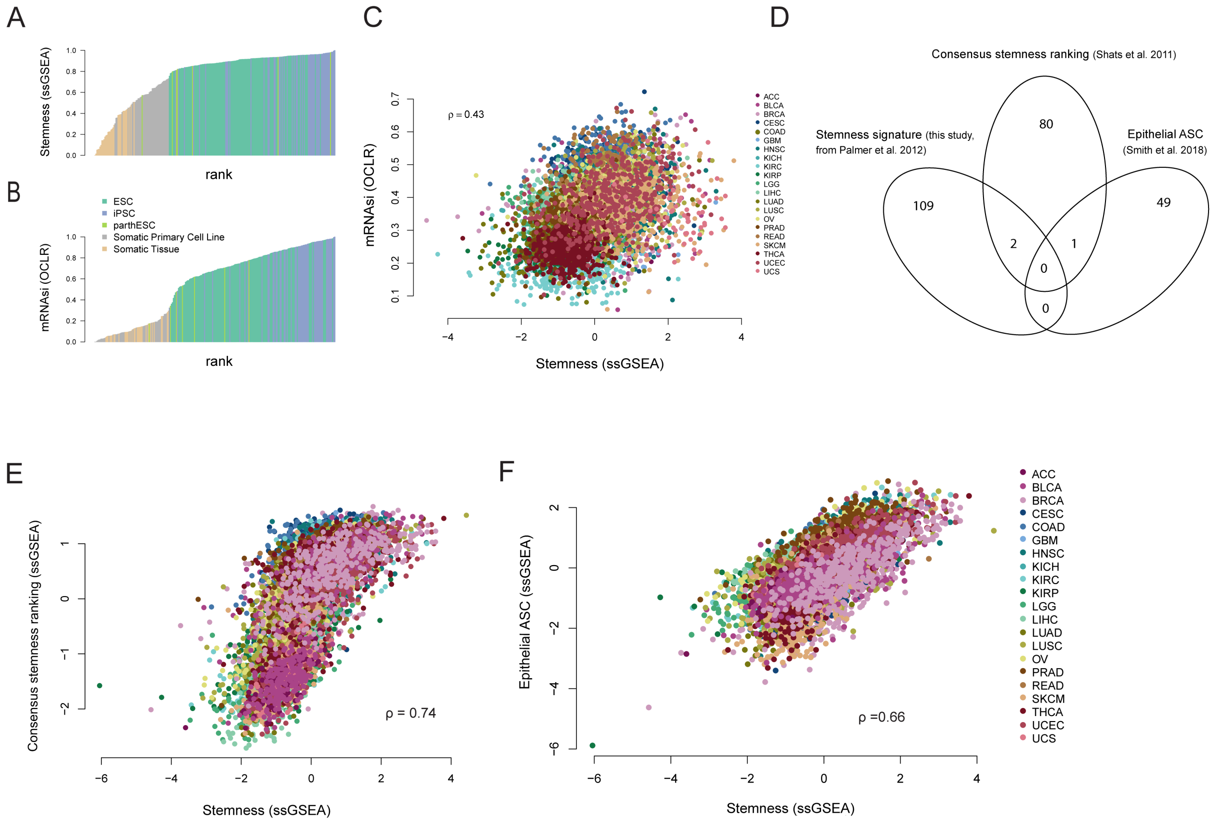

Fig. S2. A) Stemness scores for validation cohort of Malta et al. 2018 (GSE30652) generated via ssGSEA. **B)** mRNAsi stemness score generated by one-class logistic regression in same cohort. **C)** Relationship between the mRNAsi stemness signature (Malta et al., 2018) and stemness computed by ssGSEA in evaluable samples. **D)** The Venn diagram shows the number of genes in common and distinct to each of the gene sets: our curated stemness signature (this study), (Shats et al., 2011) and (Smith et al., 2018). **E)** and **F)** Correlation between ssGSEA computed stemness signature and the Consensus stemness ranking (Shats et al., 2011) and epithelial adult stem cells (ASC) signature (Smith et al., 2018), respectively.

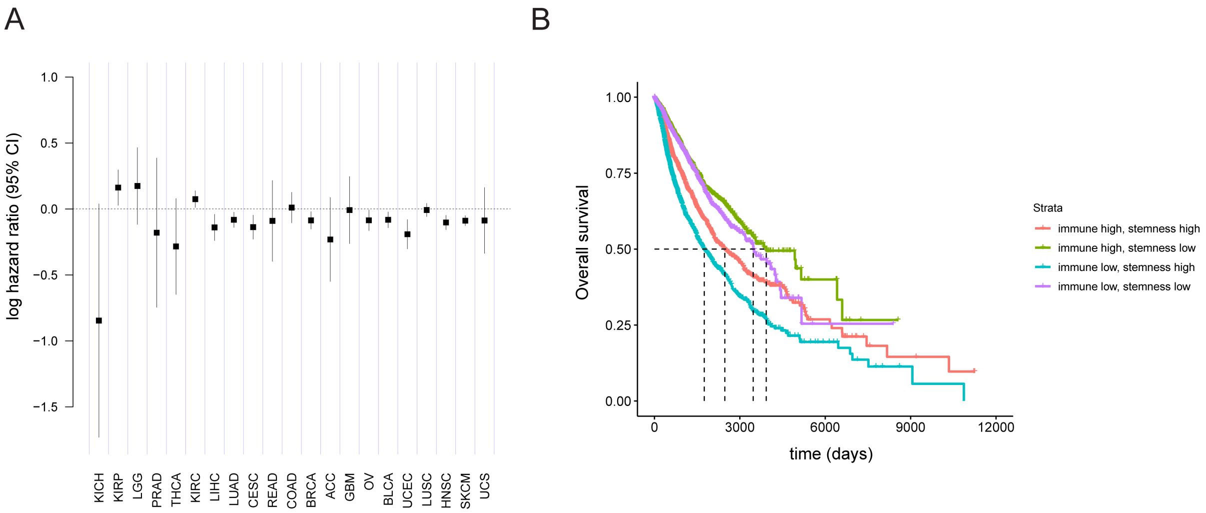

Fig. S3. **A)** Log hazard ratios and 95% confidence interval (CI) for overall survival associated with immune signature for each TCGA cancer (Cox proportional hazards models). **B)** Kaplan-Meier overall survival curves for evaluable TCGA patients, split by median stemness and immune signatures.

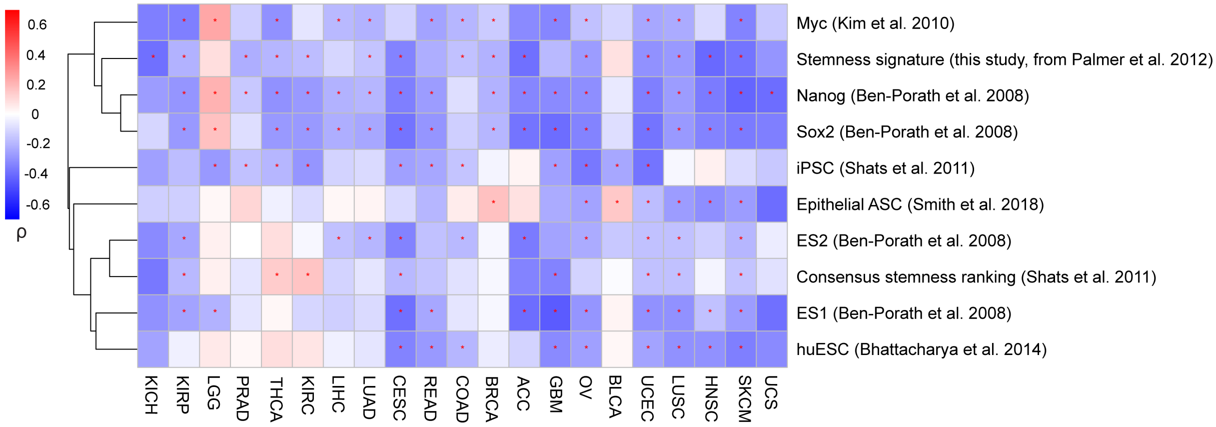

Fig. S4. Heatmap representing color-coded Spearman correlations between alternative stemness signatures and immune signature. Asterisks (*) denote significant associations ($P_{adj} < 0.05$).

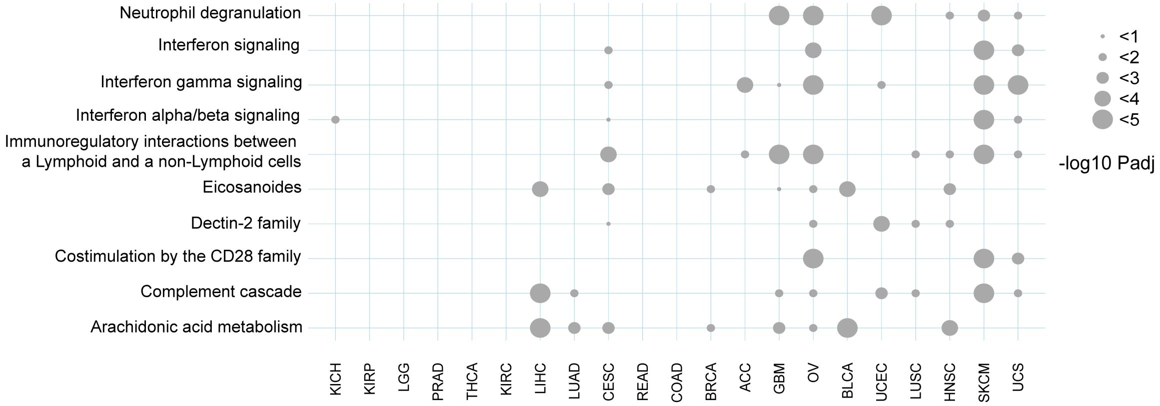

Fig. S5. Curated reactome pathway enrichment analysis of the top 1,000 genes up-regulated in low stemness (<20th percentile) vs. high stemness (>80th percentile) samples. The size of each point reflects -$\log_{10}$ P_{adj}.

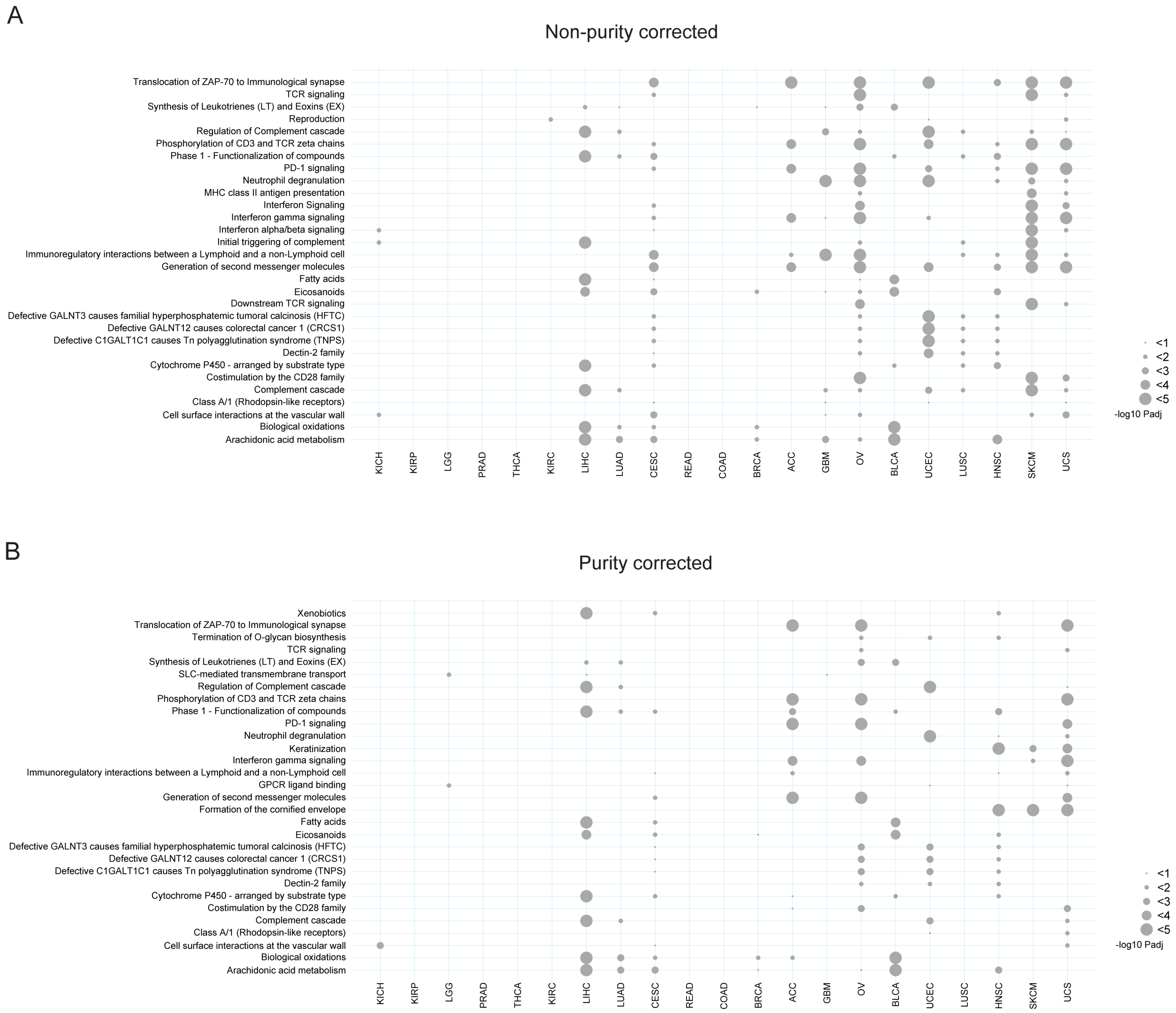

Fig. S6. **A)** Top 30 recurrently enriched Reactome pathways on the top 1000 genes down-regulated in high stemness vs. low stemness samples, for each cancer. **B)** As above, but including log tumor purity in differential expression analysis.

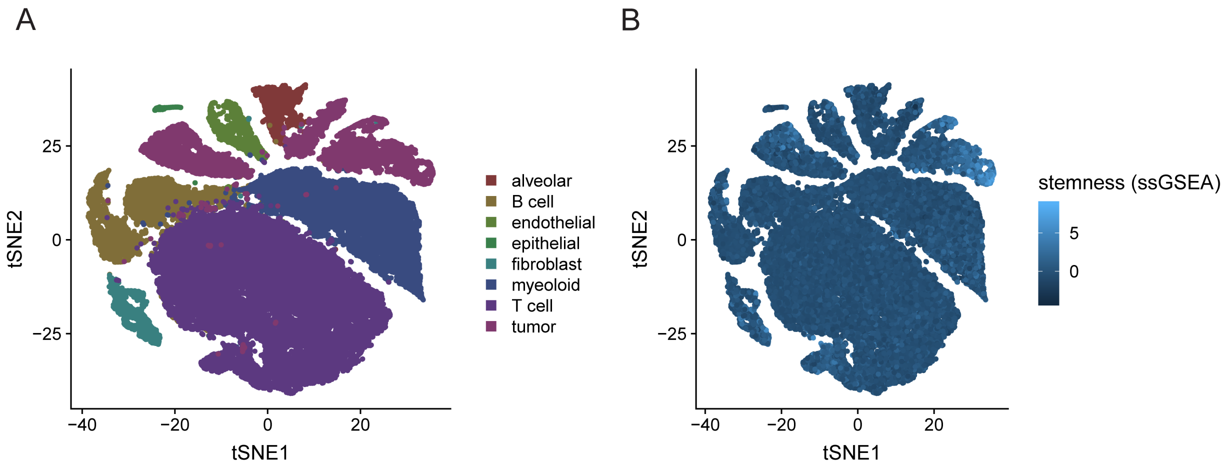

Fig. S7. Stemness score is higher in cancer cells than in other cell types, as evaluated on a lung cancer scRNA-seq dataset of Lambrechts et al. (one-way ANOVA with pairwise Tukey tests; cancer cells have greater average stemness ($P < 10^{-15}$) in all comparisons to other cell types). tSNE plot showing based on published tSNE coordinates of 52,698 cells with each cell color-coded for **A)** the associated cell type and **B)** the stemness score computed on that cell (color scale is defined in the inset).

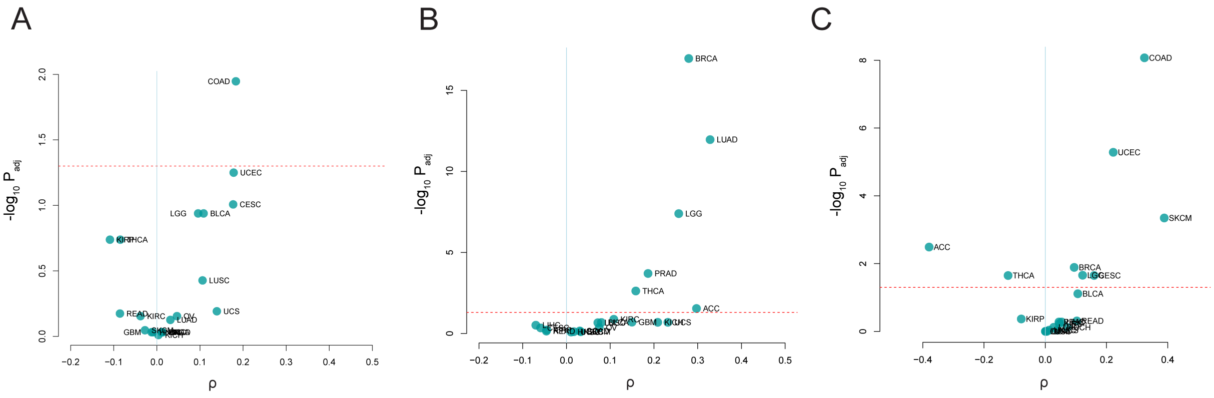

Fig. S8. Volcano plots of Spearman correlation between **A)** immune signature and mutation load (as log non-synonymous mutations), **B)** stemness and neoantigen load and **C)** immune signature and neoantigen load, based on neoantigens predicted by Thorsson et al. (2018). Dashed line of $P_{adj} = 0.05$ is plotted.

Fig. S9. Immune signature and CT antigen expression (ssGSEA of CT antigen gene set) negatively correlate, where significant. Bar plots show the Spearman ρ values for each cancer type, and asterisks denote $P_{adj} < 0.05$.

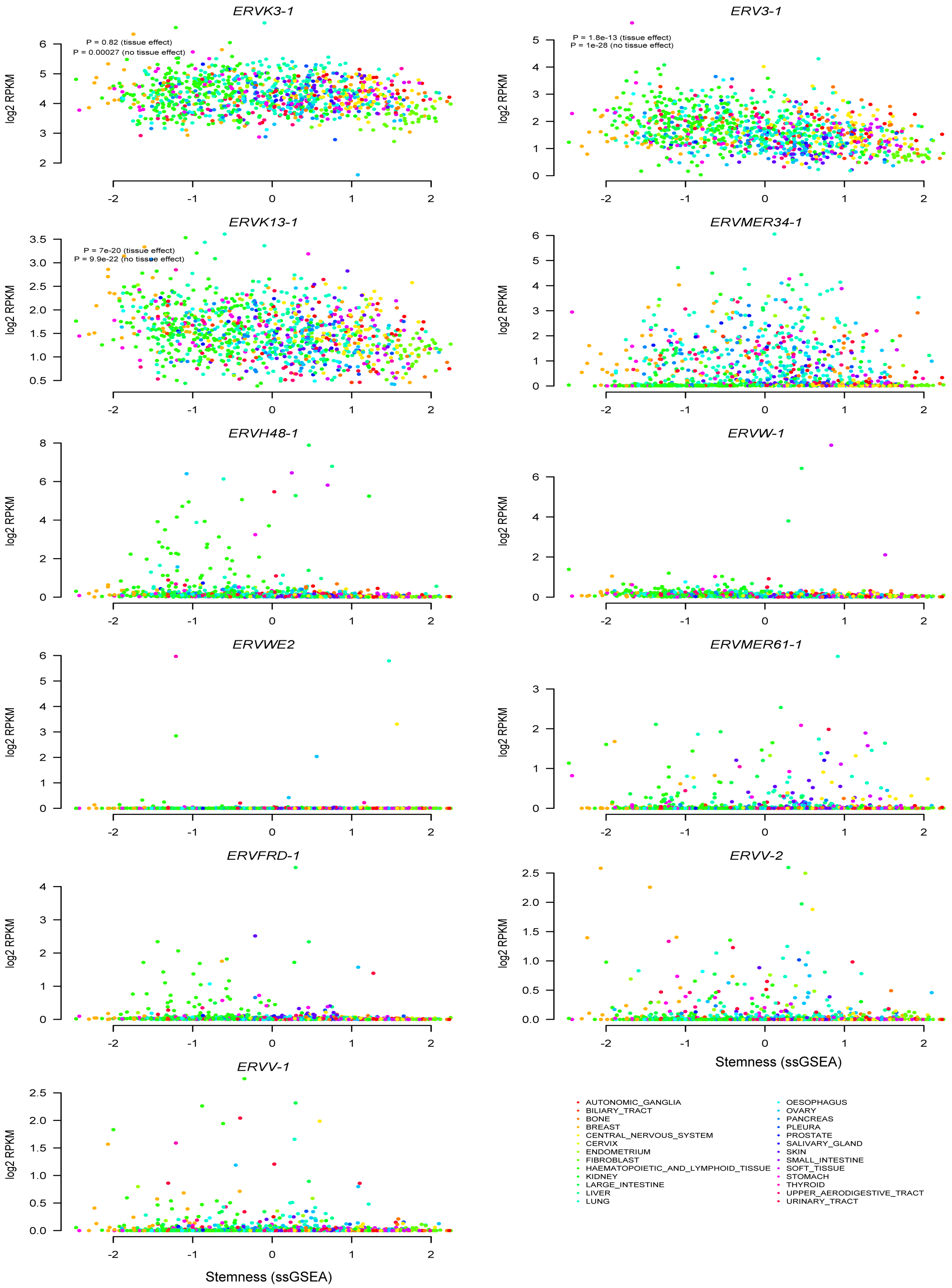

Fig. S10. ERV expression negatively associates with stemness across the Cancer Cell Line Encyclopedia, for evaluable ERVs (ERVK3-1, ERV3-1, ERVK13-5 tested). P-values for main effect of stemness in linear models shown for models either including or excluding tissue of origin as a main effect.

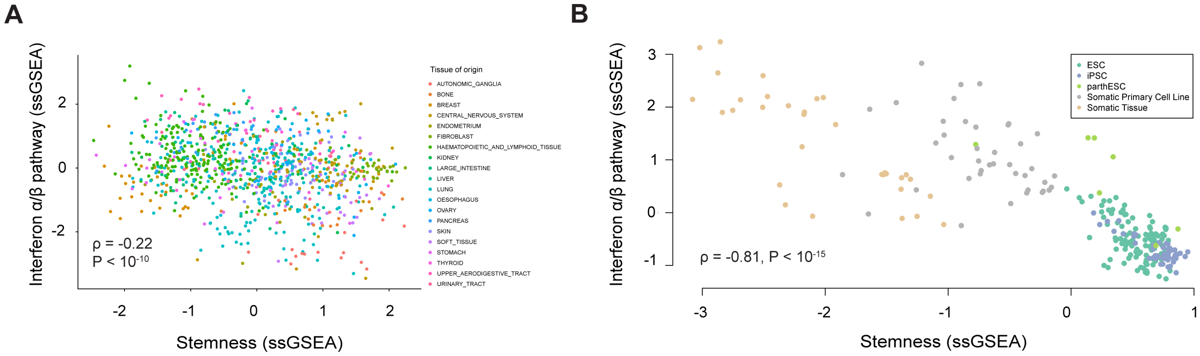

Fig. S11. Stemness negatively associates with type I interferon signaling **A)** Association of stemness score with type I interferon signaling (P < 10-10; Reactome IFN α/β pathway ssGSEA) across cancer cell lines. Colored points represent different cell lines in data from the CCLE. Only tissues represented by more than 10 independent cells lines were included in analysis. B) Association of stemness score with type I interferon signaling across non-malignant stem cells, somatic tissues, and primary cells (P < 10-15). Colored points represent different cell lines and tissues in data from the GSE30652.

Additional Dataset S1 (separate file)
Dataset S1. Stemness gene signatures

Additional Dataset S2 (separate file)
Dataset S2. Cancer testis genes list

References

1. Barretina J*, et al.* (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483(7391):603-607.

2. Lambrechts D*, et al.* (2018) Phenotype molding of stromal cells in the lung tumor microenvironment. *Nature medicine* 24(8):1277-1289.

3. Nazor KL*, et al.* (2012) Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. *Cell stem cell* 10(5):620-634.

4. Marisa L*, et al.* (2013) Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS medicine* 10(5):e1001453.

5. Mezheyeuski A*, et al.* (2018) Multispectral imaging for quantitative and compartment-specific immune infiltrates reveals distinct immune profiles that classify lung cancer patients. *The Journal of pathology* 244(4):421-431.

6. Bhat-Nakshatri P*, et al.* (2010) SLUG/SNAI2 and tumor necrosis factor generate breast cells with CD44+/CD24- phenotype. *BMC cancer* 10:411.

7. Jung P*, et al.* (2011) Isolation and in vitro expansion of human colonic stem cells. *Nature medicine* 17(10):1225-1227.

8. Ng SW*, et al.* (2016) A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* 540(7633):433-437.

9. Aran D, Hu Z, & Butte AJ (2017) xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome biology* 18(1):220.

10. Hanzelmann S, Castelo R, & Guinney J (2013) GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC bioinformatics* 14:7.

11. Durinck S, Spellman PT, Birney E, & Huber W (2009) Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature protocols* 4(8):1184-1191.

12. Colaprico A*, et al.* (2016) TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic acids research* 44(8):e71.

13. Bates D, Mächler M, Bolker BM, & Walker SC (2015) Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1-48).

14. Therneau TM (2018) Survival Analysis. *R package Version 2.42-3*.

15. Ritchie ME*, et al.* (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* 43(7):e47.

16. Yu G & He QY (2016) ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Molecular bioSystems* 12(2):477-479.

17. R. Gentleman, V. Carey, W. Huber, & Hahne F (2018) genefilter: methods for filtering genes from high-throughput experiments. *R package Version 1.62.0*.

18. Krzywinski M*, et al.* (2009) Circos: an information aesthetic for comparative genomics. *Genome research* 19(9):1639-1645.