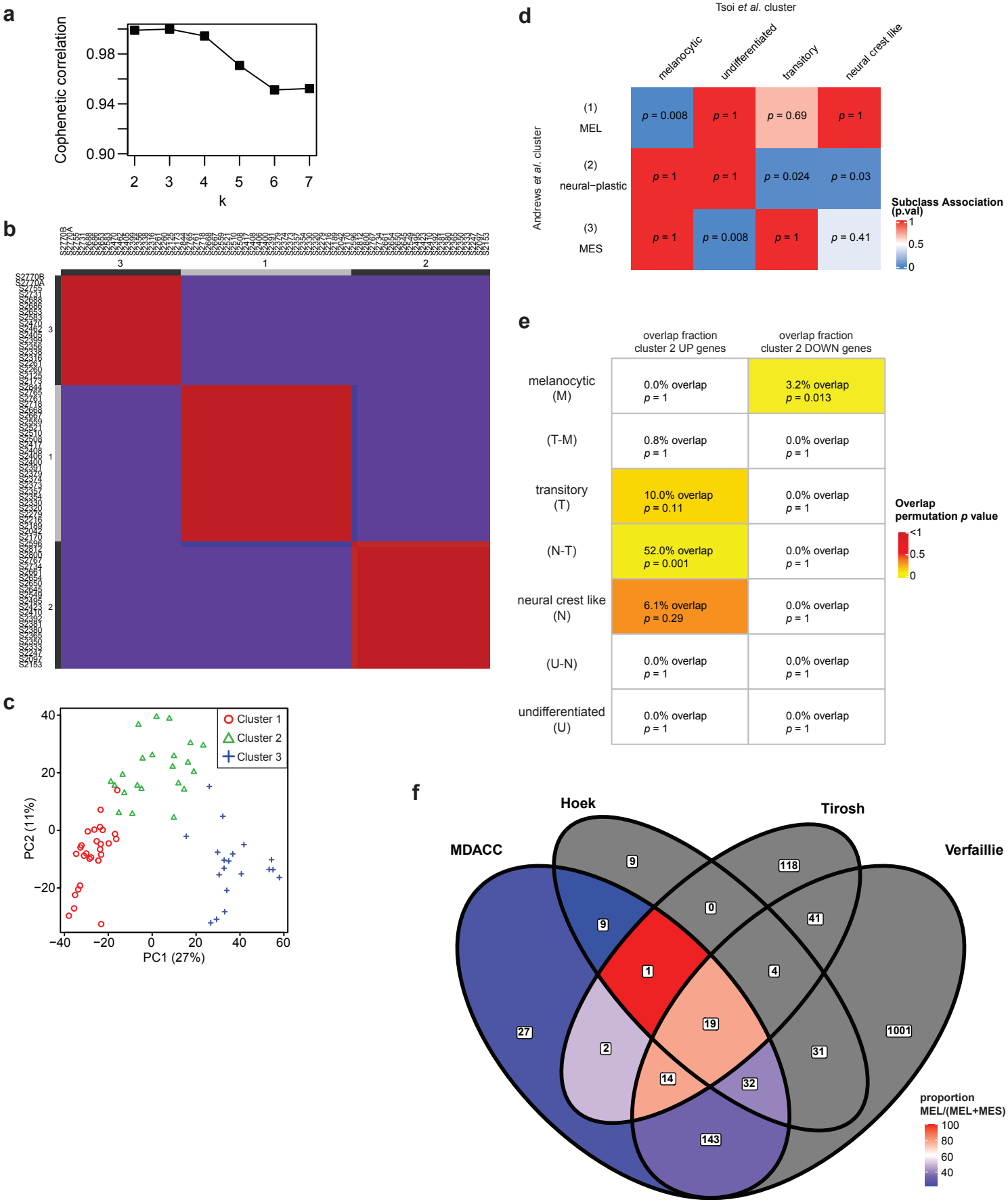
Multi-modal molecular programs regulate melanoma cell state and clinical outcomes.

Authors: Miles C. Andrews, Junna Oba, Chang-Jiun Wu, Haifeng Zhu, Tatiana Karpinets, Caitlin A. Creasy, Marie-Andrée Forget, Xiaoxing Yu, Xingzhi Song, Xizeng Mao, A. Gordon Robertson, Gabriele Romano, Peng Li, Elizabeth M. Burton, Yiling Lu, Robert Szczepaniak Sloane, Khalida M. Wani, Kunal Rai, Alexander J. Lazar, Lauren E. Haydu, Matias A. Bustos, Jianjun Shen, Yueping Chen, Margaret B. Morgan, Jennifer A. Wargo, Lawrence N. Kwong, Cara L. Haymaker, Elizabeth A. Grimm, Patrick Hwu, Dave S.B. Hoon, Jianhua Zhang, Jeffrey E. Gershenwald, Michael A. Davies, P. Andrew Futreal, Chantale Bernatchez, Scott E. Woodman.

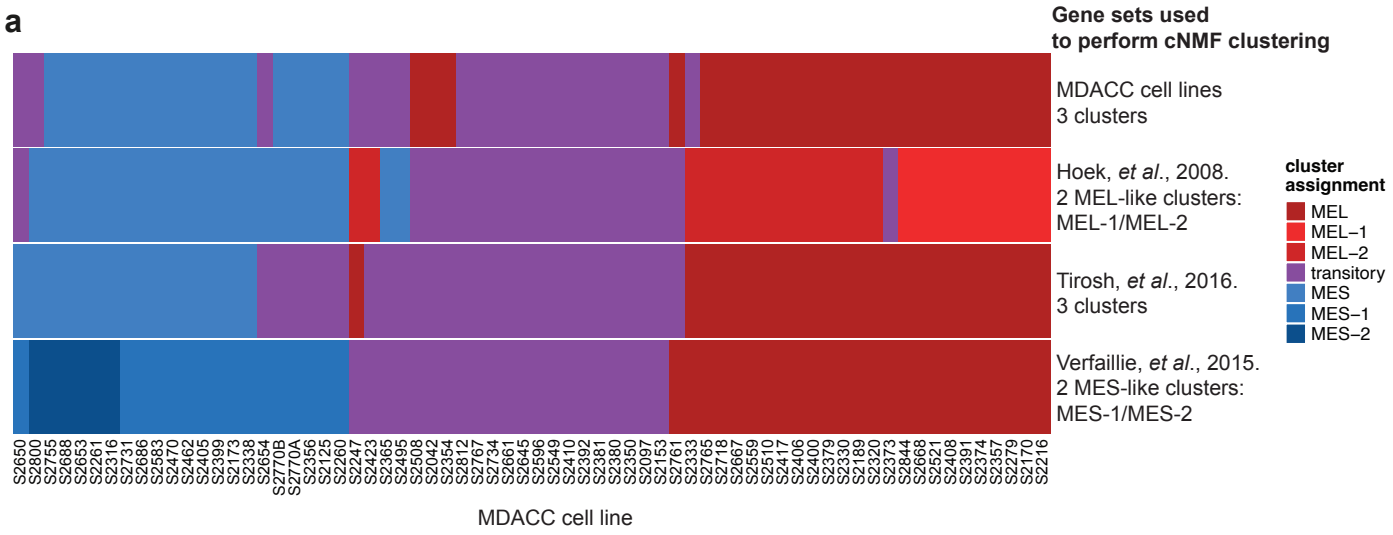
Supplementary Information

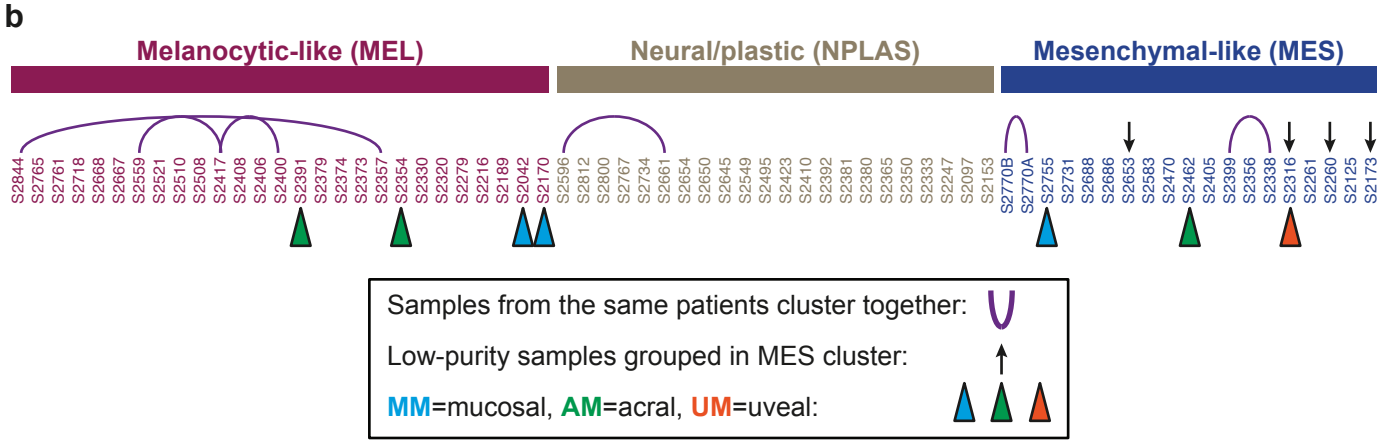
Supplementary Figure 1.

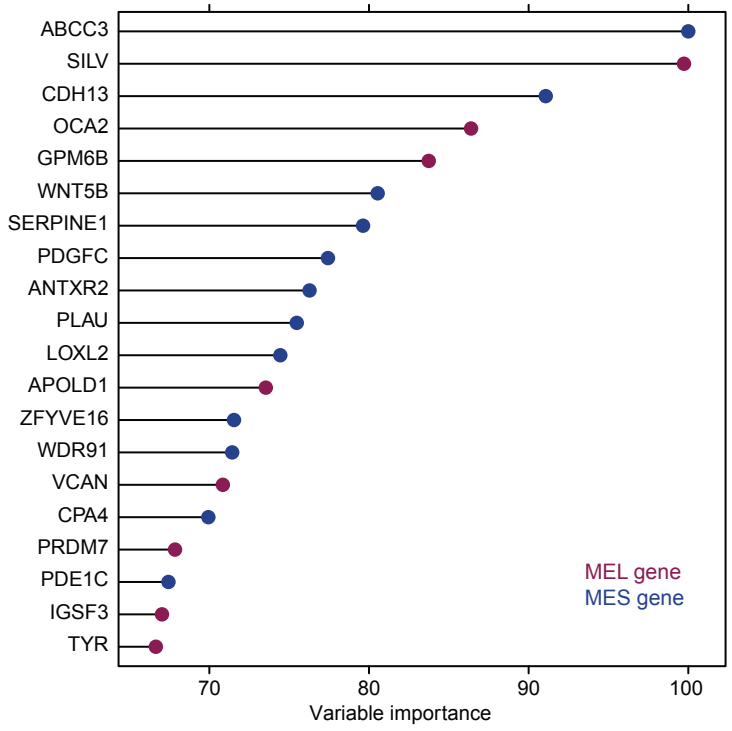


Supplementary Figure 1: Unsupervised cNMF clustering and gene set overlaps of early-passage melanoma MDACC cell lines. Using the 1500 most-variably expressed genes across a cohort of early-passage melanoma cell lines (n=68), cNMF clustering was optimized by cophenetic correlation (**a**) and consensus clustering (**b**) with 1000 iterations revealing 3 distinct mRNA expression-based clusters. (**c**) Principal component analysis demonstrates the clear separation between clusters. (**d**) Heatmap of cross-cohort class similarities measured by SubMap and (**e**) gene set overlaps between the MEL and MES-based clusters and transcriptomic states described by Tsoi *et al*. Shown in (d) are summary subclass association *p* values derived from Bonferroni-corrected one-sided Fisher inverse chi-square statistics estimated on subclass gene set enrichment scores, and in (d) one-sided *p* values of gene set overlap using permutation testing with 1000 iterations. (**f**) Gene set overlaps between MEL and MES genes and the published melanoma transcriptome type gene sets indicated. The relative proportion of MEL to MES genes within each overlap segment is indicated by color, with red representing more MEL and blue representing more MES. Source data are provided as a Source Data file.

Supplementary Figure 2.





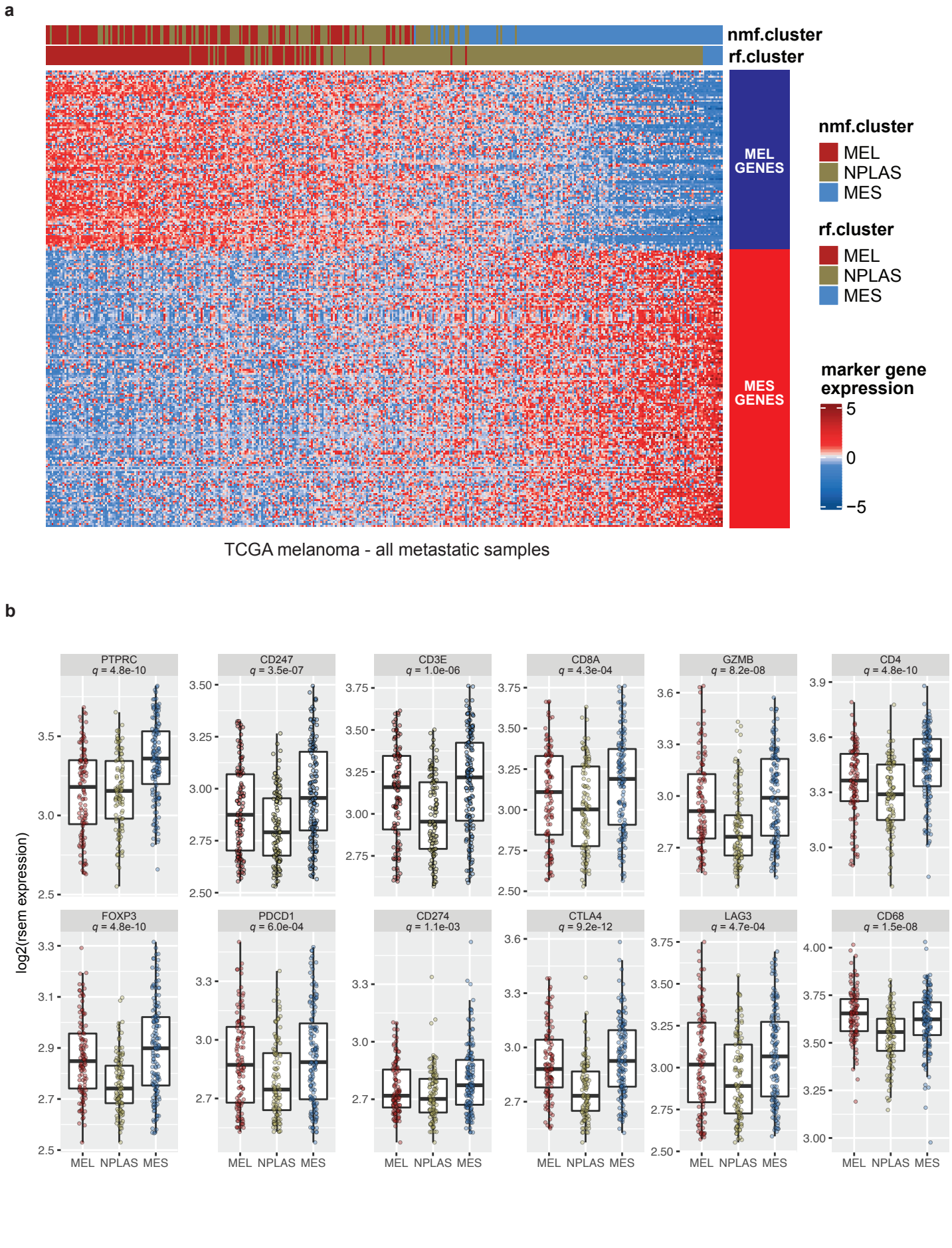


d

	Full cohort (n=68 lines from 62 patients)	Refined cohort (n=53 lines from 53 patients)
	Number (%)	Number (%)
Sex		
Male	37 (60%)	33 (62%)
Female	25 (40%)	20 (38%)
Subtype		
Non-acral	50 (74%)	41 (77%)
Acral lentiginous	3 (4%)	1 (2%)
Mucosal	3 (4%)	-
Uveal	1 (2%)	-
Unknown primary	11 (16%)	11 (21%)

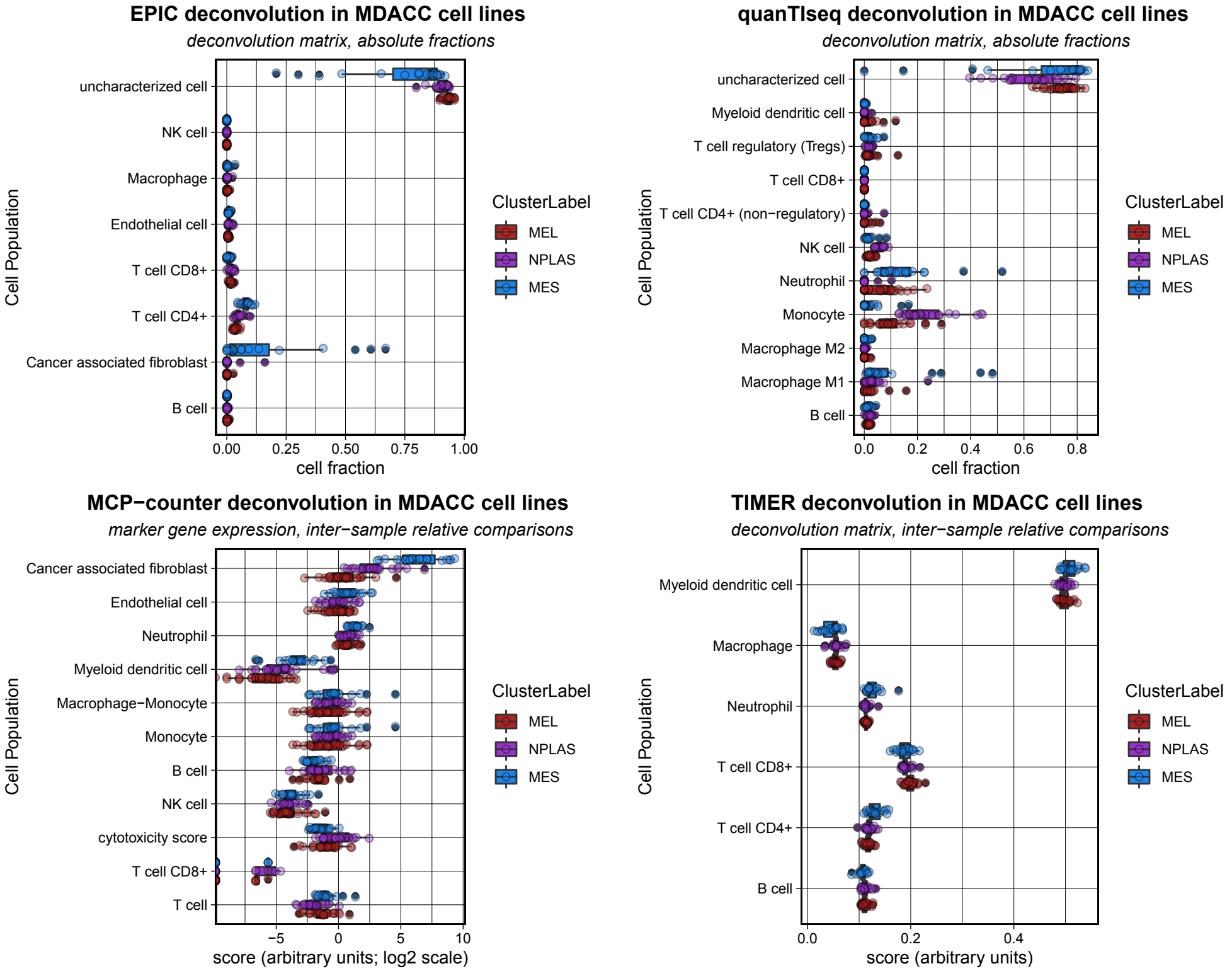
Supplementary Figure 2: Comparative gene set clustering and cohort refinement. (**a**) Comparison of cell line clustering based on MEL and MES genes and three other published melanoma transcriptomic types. Optimal cNMF-based clustering solutions for different cohorts produced 2 MEL-like clusters or 2 MES-like clusters as indicated in the figure. (**b**) Based on clustering derived from the cNMF approach, cell lines derived separately from tumors arising in the same patient remain closely related, and lower-purity cell lines appear within the MES cluster. Cell line names are colored according to the cNMF clustering (red, Cluster 1; khaki, Cluster 2; blue, Cluster 3). Cell lines derived from mucosal and uveal melanomas are also indicated. (**c**) Variable importance plot of the top 20 most informative genes used in the random forest cluster classifier model. (**d**) Full cohort composition used in initial cluster discovery, and refined cohort used for subsequent molecular analyses considering only non-mucosal/uveal subtypes, higher-purity samples, and only one cell lines derived from an individual patient. Source data are provided as a Source Data file.

Supplementary Figure 3.



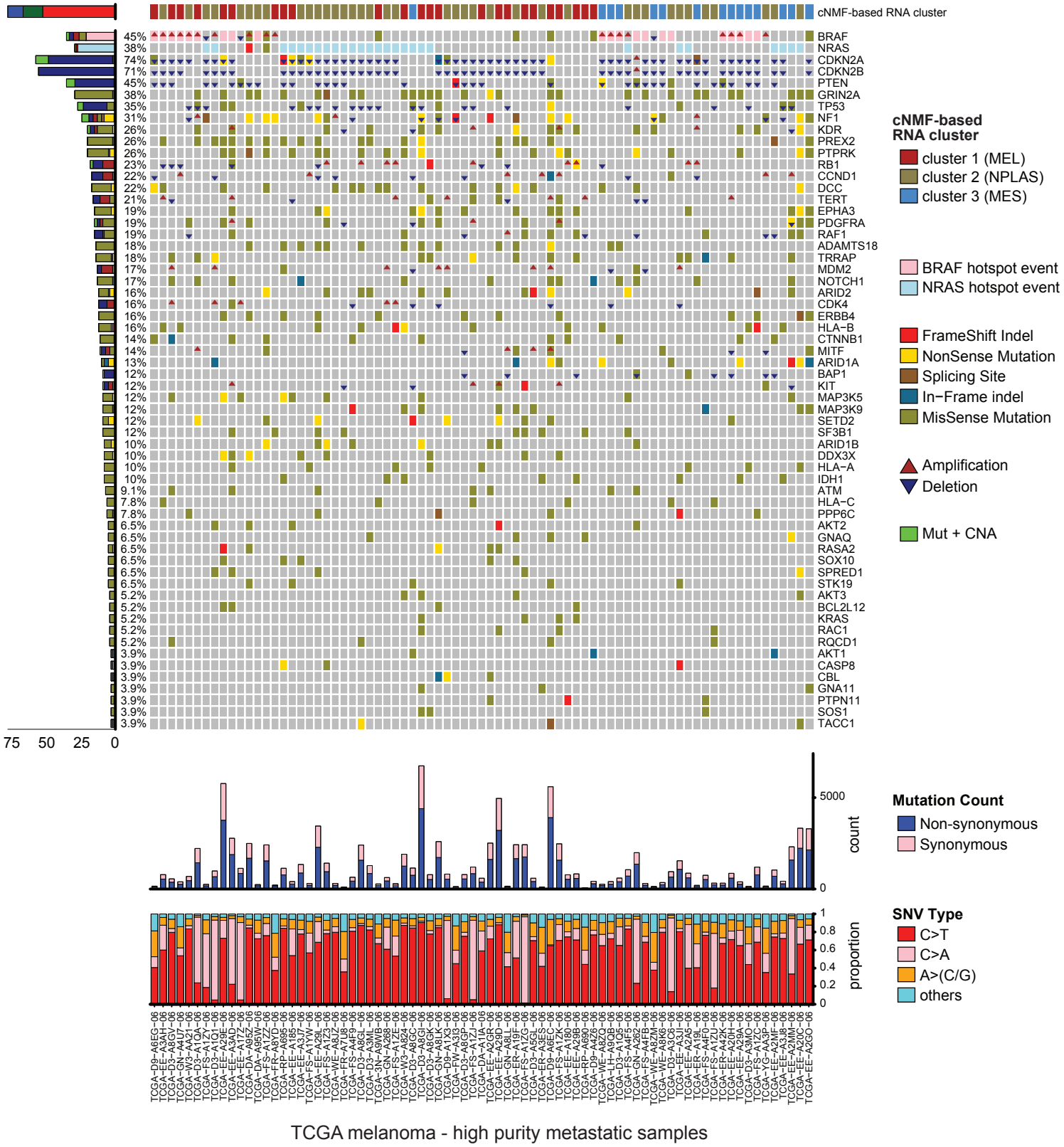
Supplementary Figure 3: MCS clusters and key immune gene expression differences in TCGA melanoma samples. (**a**) Heatmap and unsupervised cNMF clustering analysis was performed on all metastatic TCGA melanoma samples (n=368, purity ≥0.50) based on MEL and MES marker gene expression, showing overall preserved separation into predominantly MEL-like or predominantly MES-like clusters, although with less distinct separation than was seen in only high-purity samples. cNMF-based clusters (nmf.cluster) and the cluster prediction from the cell line-trained random forest model (rf.cluster) are shown above; samples are ordered left-to-right by decreasing MEL-MES score (see Methods). (**b**) Specific marker genes supportive of leukocyte subsets within all metastatic TCGA melanoma samples (n=368) reveal differential abundances between MCS groups. Boxplots indicate the median (thick bar), first and third quartiles (lower and upper bounds of the box, respectively), lowest and highest data value within 1.5 times the interquartile range (lower and upper bounds of the whisker), and all individual data points are shown. The q-values shown are Benjamini-Hochberg corrected two-sided Kruskal-Wallis p-values comparing marker gene expression across all three sample classes. Source data are provided as a Source Data file.

Supplementary Figure 4.



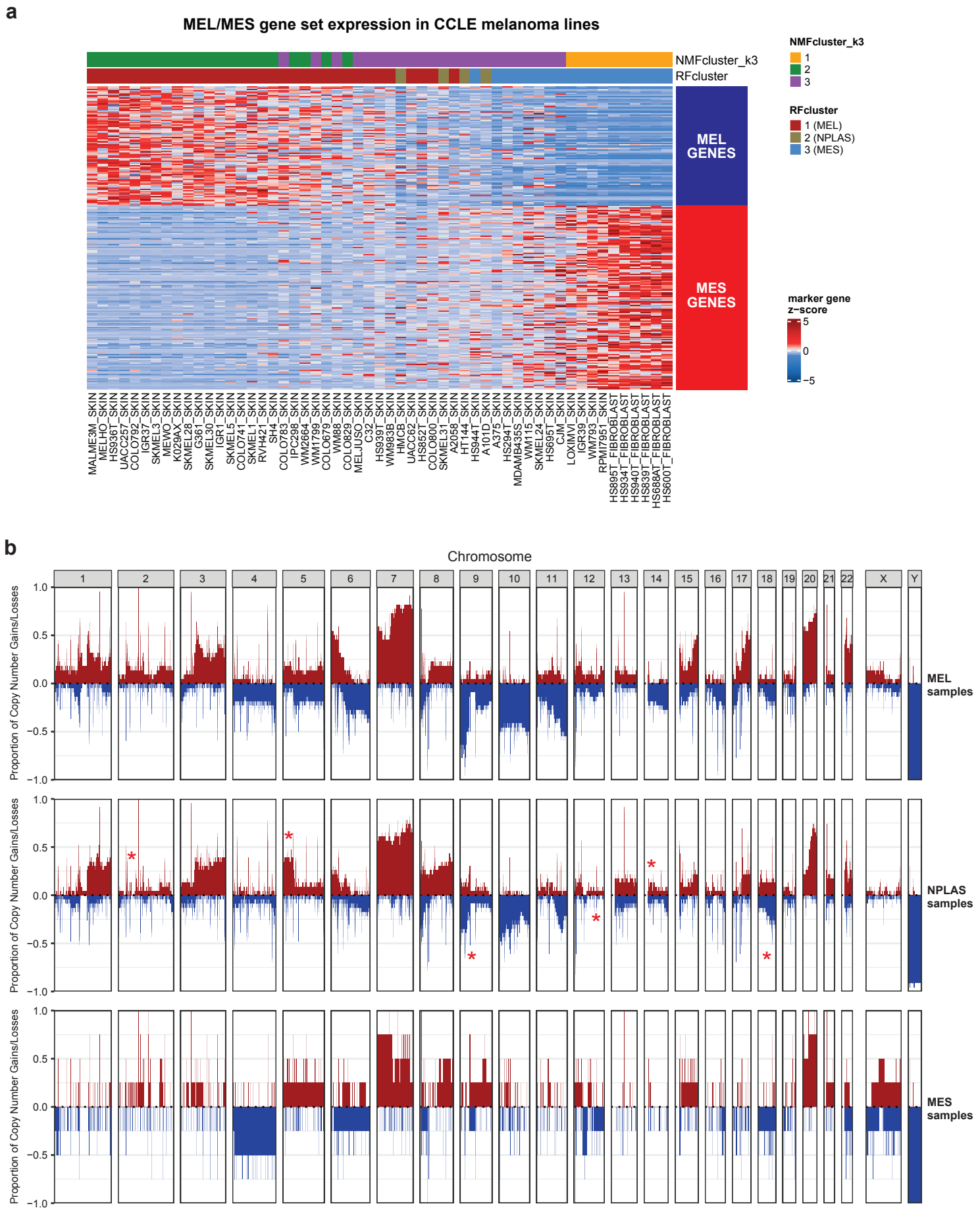
Supplementary Figure 4: Transcriptomic immune deconvolution in MDACC cell lines. Four immune deconvolution methods were applied to gene expression data of the MDACC melanoma cell line panel (n=53), with cell lines grouped by MEL/NPLAS/MES clusters, showing evidence of spurious myeloid and/or stromal signatures particularly in MES samples. The cell populations evaluated by each algorithm are shown along the y-axis with output scores/fractions along the x-axis. Horizontal boxplots indicate the median (thick bar), first and third quartiles (lower and upper bounds of the box, respectively), lowest and highest data value within 1.5 times the interquartile range (lower and upper bounds of the whisker), and all individual data points are shown. The deconvolution strategy and output type are indicated above each graph: EPIC, quanTIseq and TIMER utilize a gene expression deconvolution matrix, whereas MCP-counter employs subset-defining marker gene expression values. EPIC and quanTIseq output absolute cell fractions whereas MCP-counter and TIMER output arbitrary unit values which are suitable for inter-sample relative abundance comparisons. Source data are provided as a Source Data file.

Supplementary Figure 5.



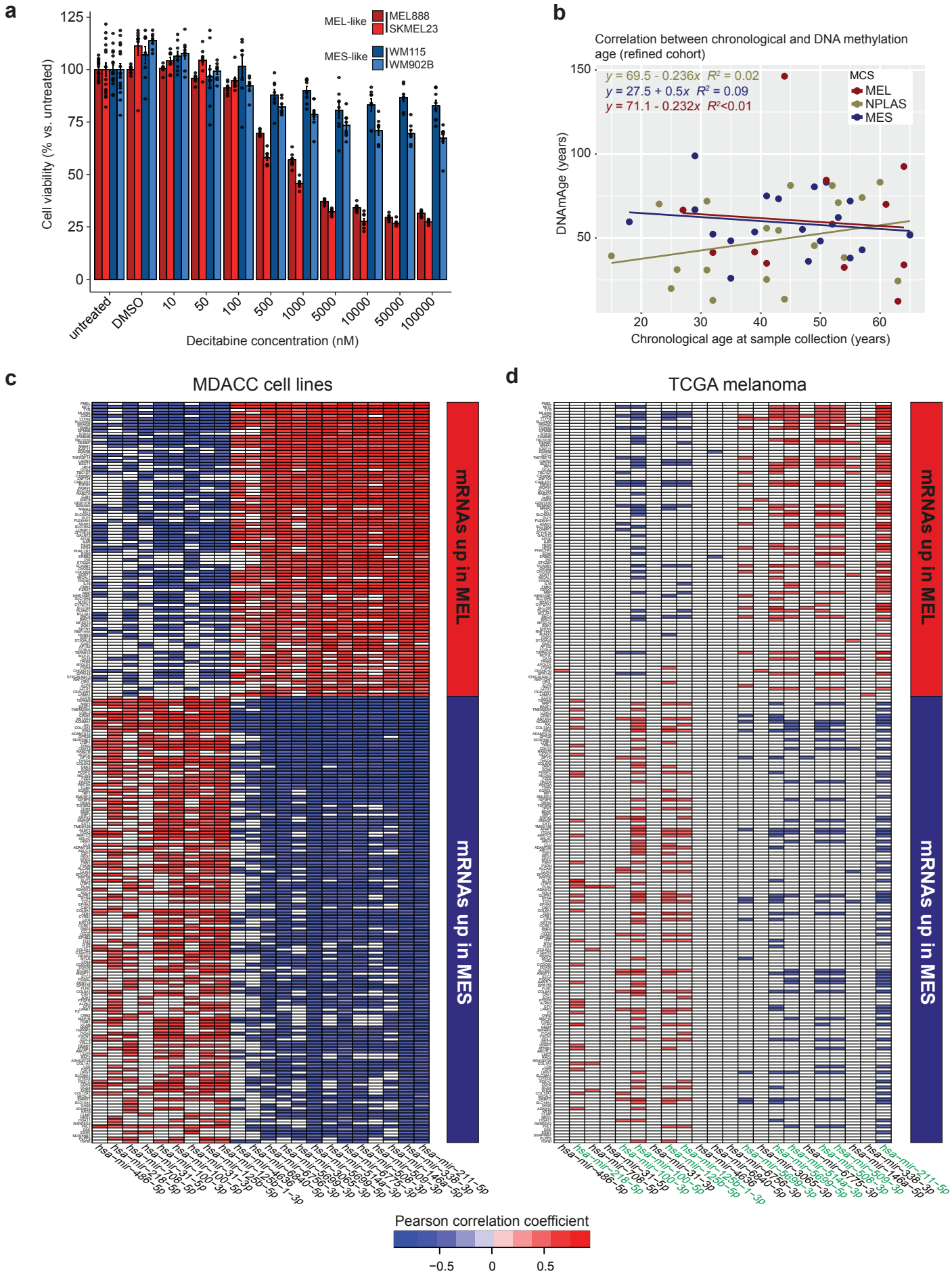
Supplementary Figure 5: Genomic features of high-purity TCGA melanoma metastases. Oncomap displaying somatic variants identified from whole-exome sequencing affecting a panel of genes frequently altered in melanoma. High-purity metastatic samples of the TCGA melanoma cohort (n=77) are shown indicating the cluster assignments assigned by cNMF performed on all TCGA melanoma samples (primaries and all metastatic samples), the presence of *BRAF* and *NRAS* mutations, and other mutations or copy number alterations (CNA). The histogram at left indicates the cumulative frequency of genomic alterations (all types) for each gene. Total synonymous and non-synonymous mutation count (number) and type of single nucleotide variants (proportion) are indicated in the bar graph and stacked bar graph below, respectively. Source data are provided as a Source Data file.

Supplementary Figure 6.



Supplementary Figure 6: MCS and copy number alteration of Cancer Cell Line Encyclopedia (CCLE) melanoma samples. (**a**) Heatmap of MEL and MES gene expression in CCLE melanoma samples with available gene expression data demonstrating clear segregation into MCS groups. cNMF-based clusters (k=3; NMFcluster_k3) and random forest cluster predictions (RFcluster) are shown above; samples are ordered left-to-right by decreasing MEL-MES score (see Methods). (**b**) Genome-wide copy number alteration surveyed as segmented data loss/gain frequency plots contrasting profiles between the three cNMF-derived MCS identified as in (a). Regions of particular note displaying differential CNA profiles between MCS groups are indicated by red asterisks. Source data are provided as a Source Data file.

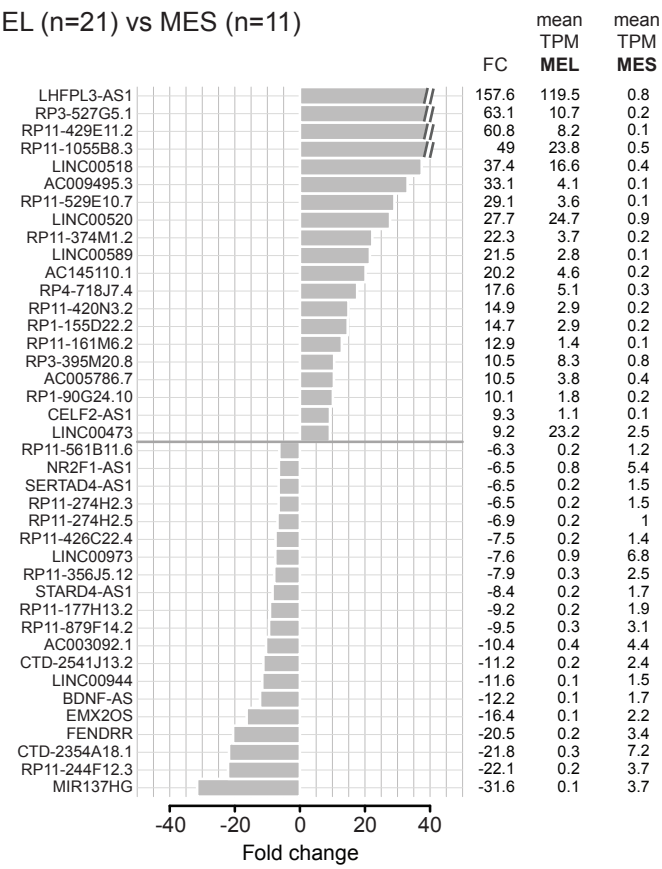
Supplementary Figure 7.



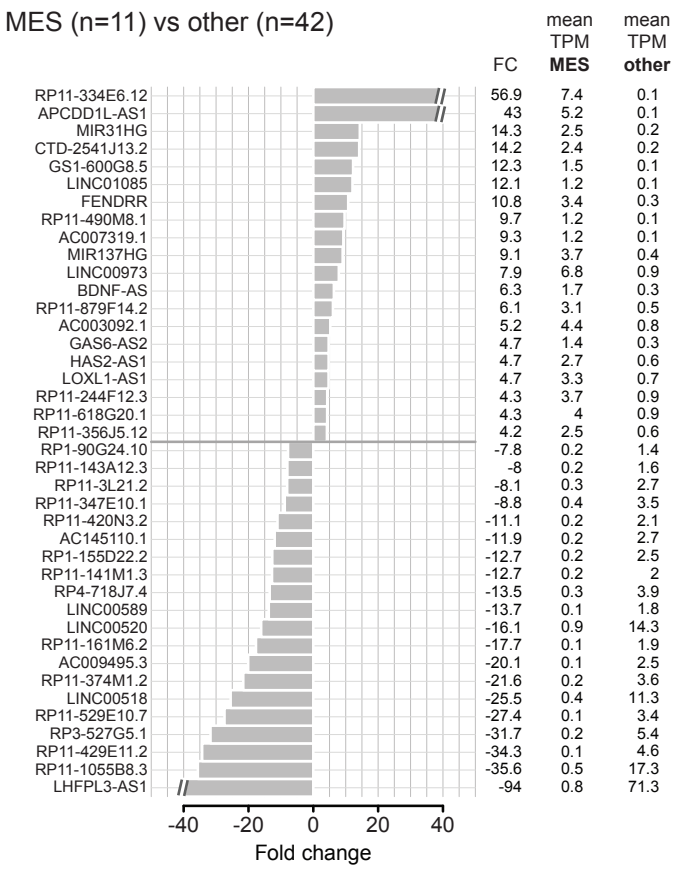
Supplementary Figure 7: Methylation effects and microRNA targeting of MEL and MES genes in MDACC cell lines and TCGA melanoma samples. (**a**) Melanoma cell lines exposed to increasing concentrations of the demethylating agent decitabine reveal greater sensitivity in MEL-like (high ECAD versus NCAD/ZEB1 expression) lines compared with MES-like (low ECAD versus NCAD/ZEB1 expression) lines. Bars show mean ± SEM of biological replicates (n=16 for untreated, n=8 for the other conditions), normalized to untreated cells for each line indicated. (**b**) Correlation between DNA methylation age (DNAmAge) and chronological age of the patient at the time the tumor from which each early-passage melanoma cell line was created was sampled, indicating MCS of each sample, and linear regression lines for each MCS. (**c-d**) Heatmaps of the Pearson correlation between expression of individual microRNAs (columns) and MEL/MES marker gene transcript abundances (rows) in (**c**) the cell line panel and (**d**) TCGA melanoma samples showing preservation of most cell line miR-gene correlations in tissue samples being particularly consistent for half of the marker miRs indicated by green labels in (b). Source data are provided as a Source Data file.

Supplementary Figure 8.

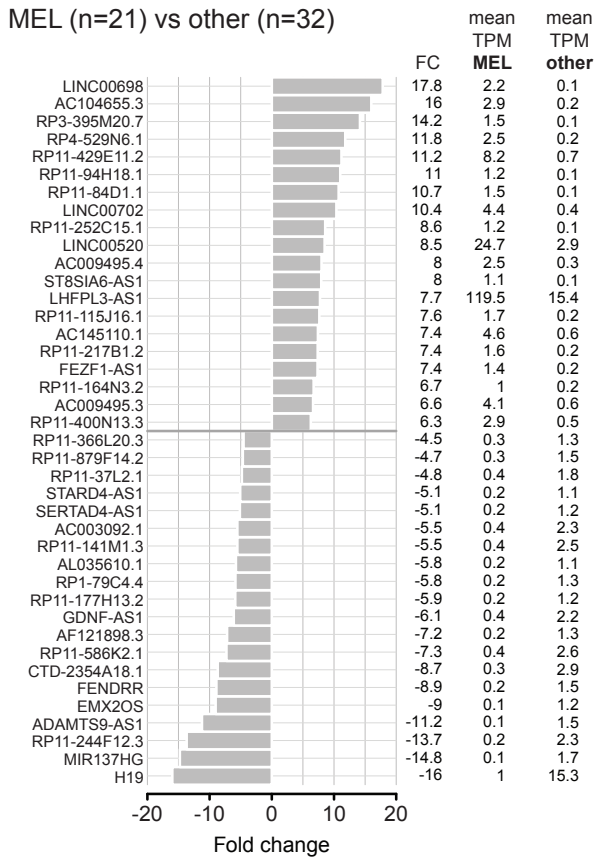
a



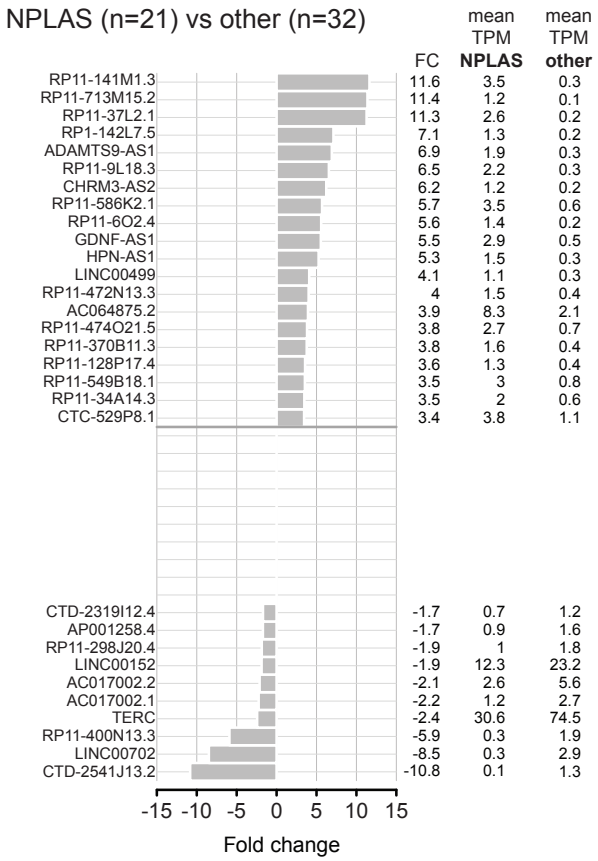
b



c



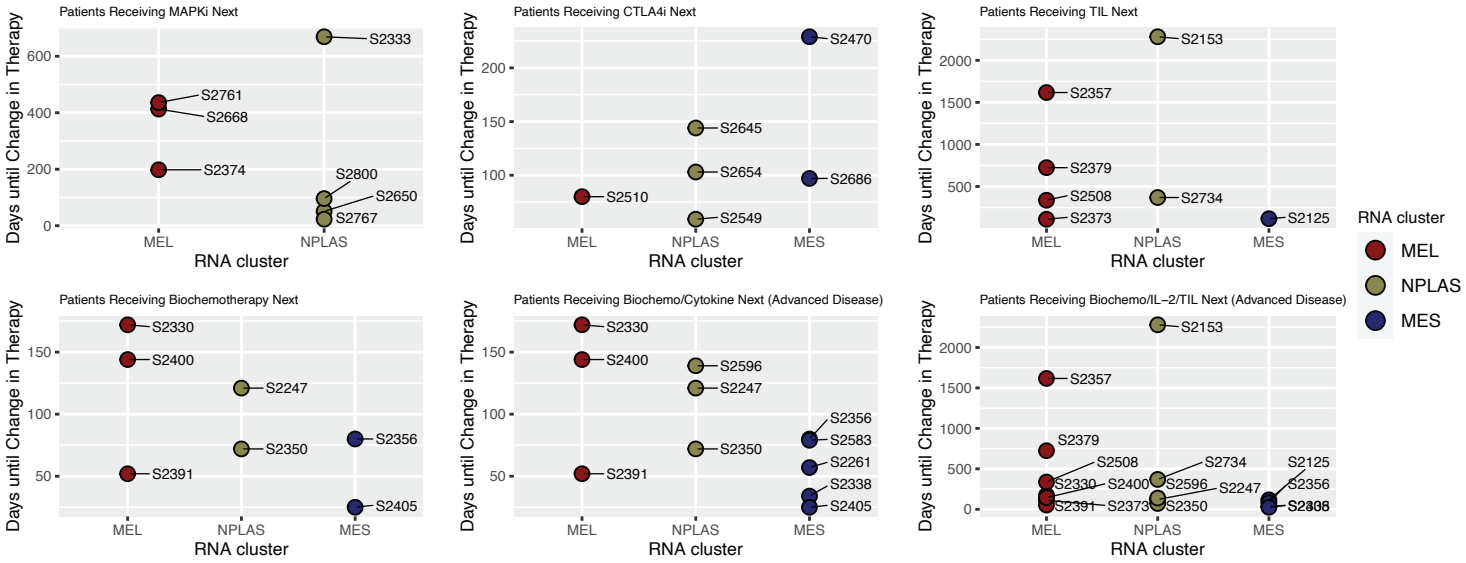
d



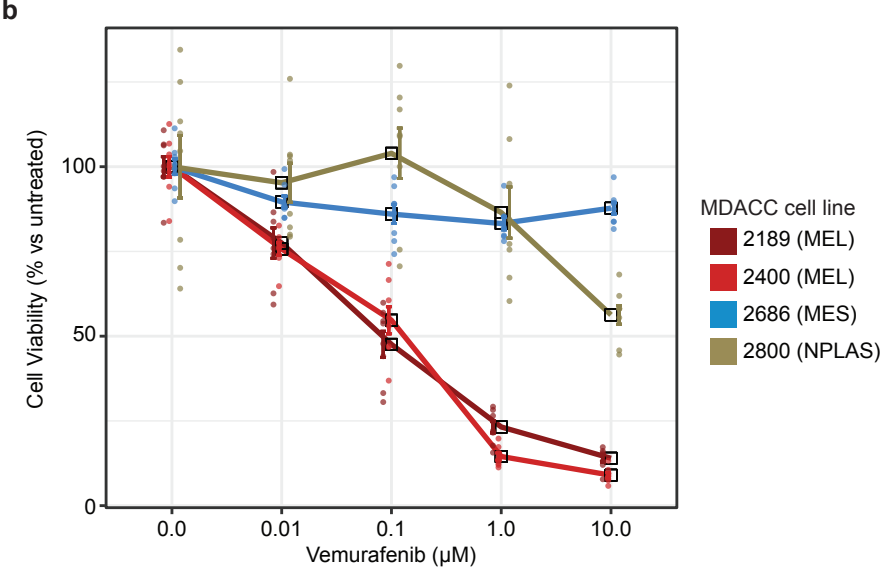
Supplementary Figure 8: Differential lncRNA expression between MCS in MDACC cell lines. Pairwise comparisons of lncRNA expression for (**a**) MEL versus MES, (**b**) MES versus (MEL or NPLAS), (**c**) MEL versus (MES or NPLAS) and (**d**) NPLAS vs (MEL or MES), expressed as fold change showing the top 20 differentially expressed lncRNAs in either direction, with a minimum of 1.5 absolute fold change between the two groups. TPM: transcripts per million. Source data are provided as a Source Data file.

Supplementary Figure 9.

a

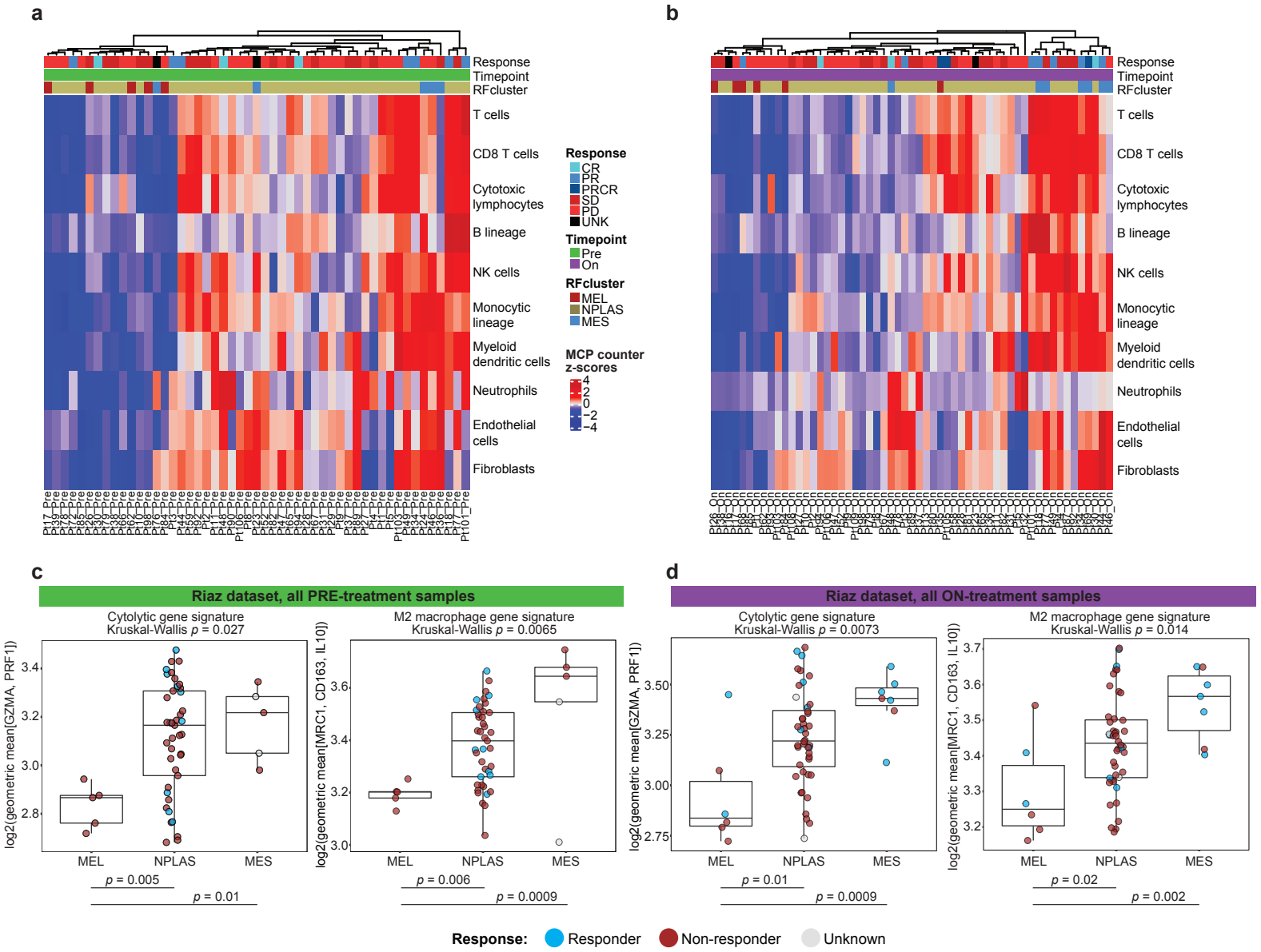


b



Supplementary Figure 9: MCS are associated with differential clinical outcomes. (**a**) Duration from initiation of systemic therapy until requirement for a change in therapy, stratified by MCS for patients receiving the next category of systemic therapy as indicated in each graph following harvest of the tumor from which a melanoma cell line was derived. (**b**) Dose-response viability curves measured by MTT assay for melanoma cell lines exposed to vemurafenib for 24 hours at the doses indicated. Shown are mean ± SD for biological replicates (n=8 each) normalized to untreated controls. Source data are provided as a Source Data file.

Supplementary Figure 10.



Supplementary Figure 10: (a) Pre-treatment (n=51) and (**b**) on-treatment (n=57) samples of the Riaz PD-1 inhibitor treated cohort demonstrate distinct immune/stromal content between samples, with relative lack of immune infiltrates in samples classified as MEL by the cell line-trained random forest classifier model compared with relatively diverse immune and stromal representation in MES samples, similarly to observations in TCGA melanoma samples based on MCP-counter transcriptomic deconvolution. In the same dataset, MEL samples also show lower cytolytic and M2 macrophage-associated markers compared with NPLAS/MES samples at both (**c**) pre-treatment and (**d**) on-treatment time points. Boxplots indicate the median (thick bar), first and third quartiles (lower and upper bounds of the box, respectively), lowest and highest data value within 1.5 times the interquartile range (lower and upper bounds of the whisker), and all individual data points are shown. Two-sided Kruskal-Wallis *p* values comparing across all sample classes are shown with post-hoc Dunn tests indicating pairwise comparisons. Response (binary) is indicated by color (cyan=responder, red=non-responder, gray=unknown) within each cluster group. Source data are provided as a Source Data file.