https://doi.org/10.1038/s41592-023-01938-4

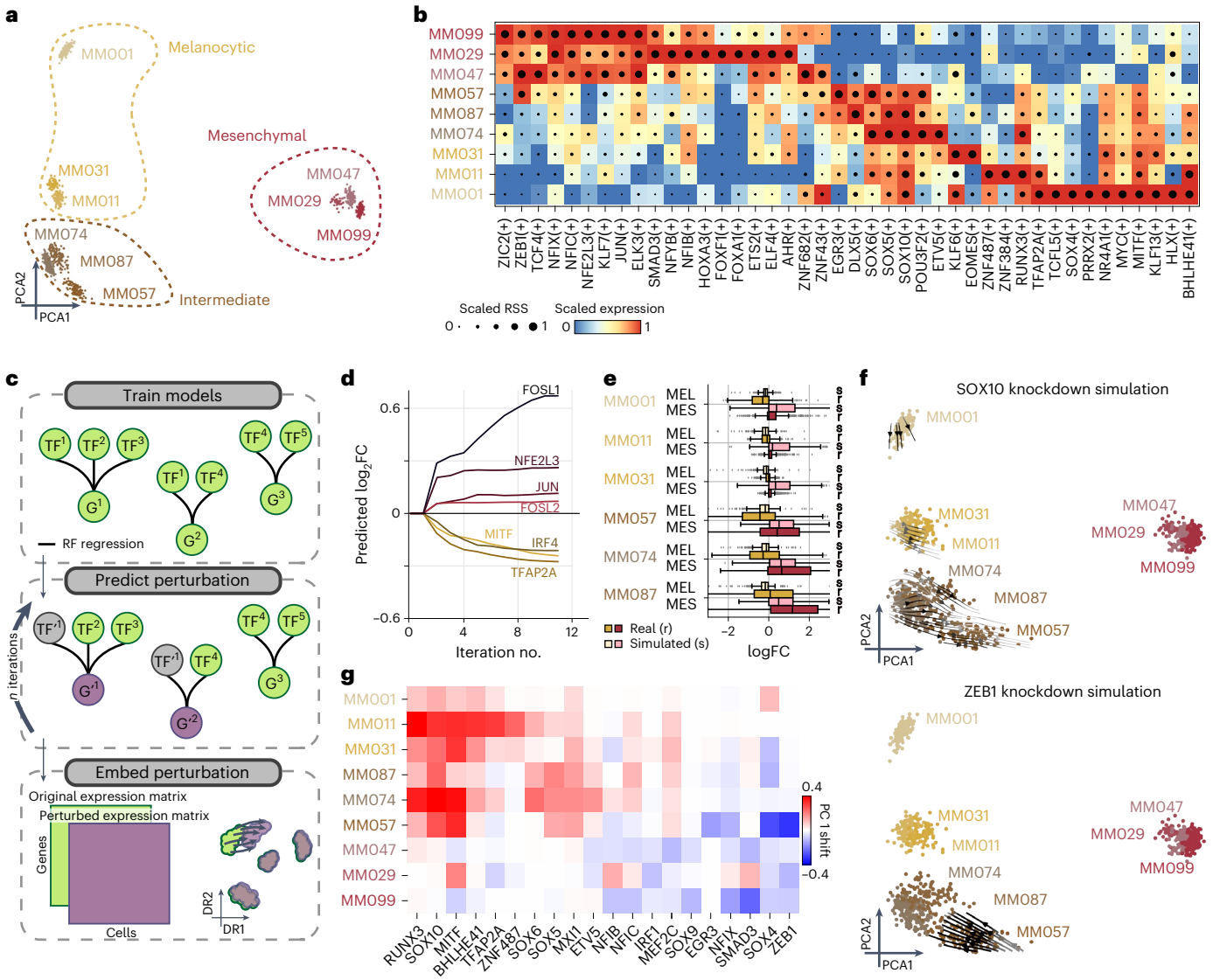

Fig. 4 | SCENIC+ analysis using separate scATAC-seq and scRNA-seq data on a mix of human melanoma lines. a, PCA of 936 pseudo-multiome cells based on target gene and target region enrichment scores. **b**, Heat map/dot-plot showing TF expression of the eRegulon on a color scale and cell-type specificity (RSS) of the eRegulon on a size scale. **c**, Illustration of how predictions from SCENIC+ can be used to simulate TF perturbations. Top: SCENIC+ is used as a feature selection method and RF regression models are fitted for each gene using TF expressions as predictors for gene expression. Middle: the expression of TF(s) is altered in silico and the effect on gene expression is predicted using the regression models, which is repeated for several iterations to simulate indirect effects. Bottom: the original and simulated gene expression matrices are co-embedded in the same dimensionality reduction to visualize the predicted effect of the perturbation on cell states. **d**, Predicted logFC of mesenchymal (red shades) and melanocytic (yellow shades) marker genes over several iterations of SOX10 knockdown simulation. **e**, Simulated (s) and actual (r) distribution of logFCs of melanocytic ($n = 523$) and mesenchymal ($n = 722$) marker genes after SOX10 knockdown across several melanoma lines. Upper/lower hinge represents upper/lower quartile, whiskers extend from the hinge to the largest/smallest value no further than 1.5 × interquartile range from the hinge respectively. The median is used as the center. **f**, Simulated shift after SOX10 and ZEB1 knockdown represented using arrows. Arrows are shaded based on the distance traveled by each cell after knockdown simulation. **g**, Heat map representing the shift along the first principal component of each melanoma line after simulated knockdown of several TFs.

For the mouse cortex, we performed 10x single-cell multiome and for the human cortex we re-used a previously published multiome dataset[60]. We were able to identify matching cell types in both species, including layer-specific excitatory neurons, interneurons derived from the medial and caudal ganglionic eminences (MGEs and CGEs, respectively) and non-neuronal populations (microglia, astrocytes, endothelial cells, oligodendrocytes and oligodendrocyte progenitor cells (OPCs); Fig. 5a,b).

SCENIC+ identified 125 and 142 high-quality eRegulons for mouse and human, respectively, out of which 60 are found in both species (Fig. 5c,d). Notably, we observed a high correlation of the specificity scores of eRegulons for these orthologous TFs in matching cell types (Extended Data Fig. 9a), implying that cell-type identity can be decomposed into these 60 eRegulons. Eight out of 60 conserved TFs have not been described before in the context of the cortex. These include Smad3/SMAD3 in the excitatory neurons of the upper cortical layers, Pparg/PPARG and Bhlhe40/BHLHE40 in L4 excitatory neurons, Etv5/ETV5 and Nfat5/NFAT5 in L5/6 excitatory neurons, Thrb/THRB and Pbx1/PBX1 in L6 excitatory neurons and Meis1/MEIS1 in oligodendrocytes (Fig. 5c,d and Extended Data Fig. 9b–d). Projection of SCENIC+ regulons onto spatial transcriptomics data further validated layer-specific GRNs in the mammalian cortex (Extended Data Fig. 9e–j and Supplementary Note 5).