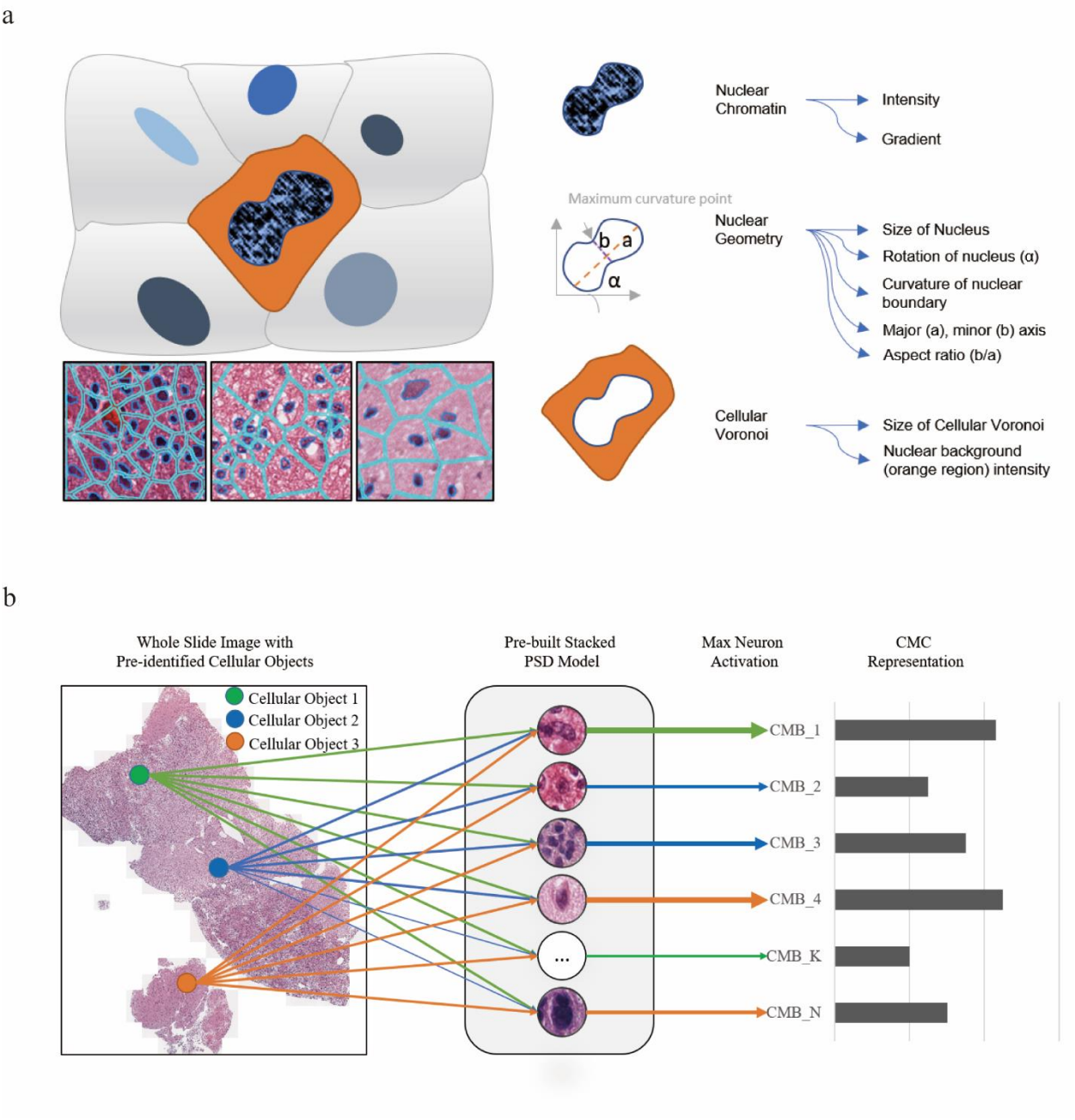
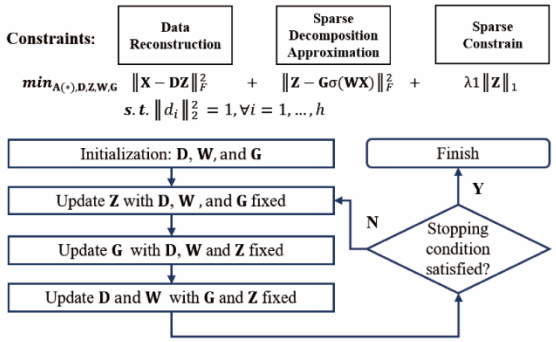


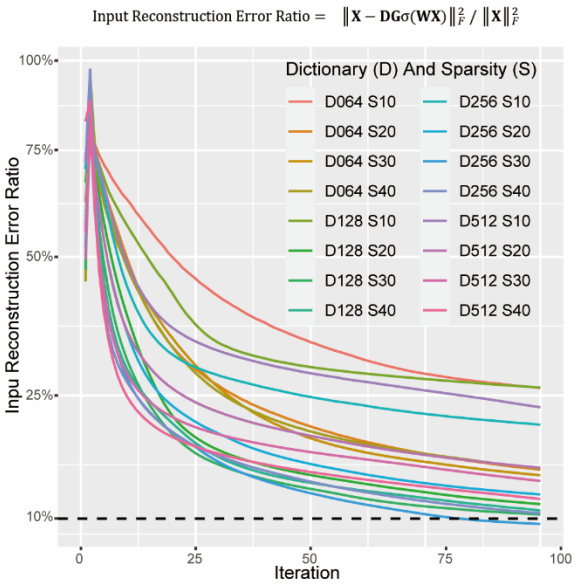
Extended Data Figure 1. **Cellular level phenotype extraction and illustration of workflows for both unsupervised CMB discovery and the construction of patient-level CMC representation**. **a.** Definition of phenotypes related to cellular-object; **b.** Pre-built unsupervised learning model (Stacked PSD) provides efficient inference and construction of patient-level CMC representation.

a

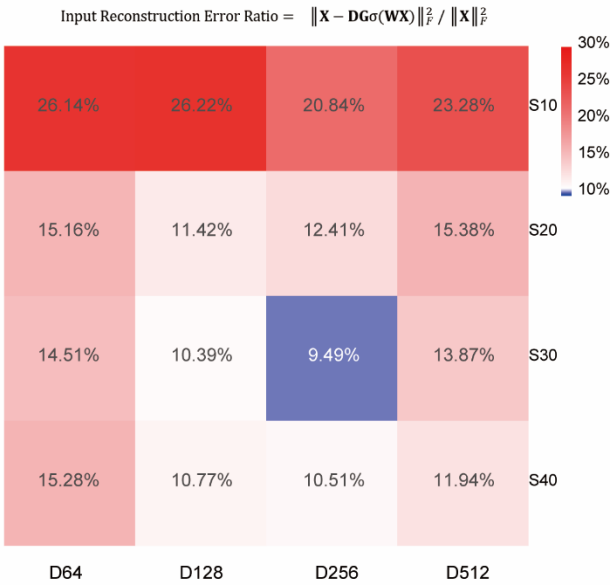


Type	Variable	Definition
Input	$\mathbf{X} = [x_{1,\dots,}x_N] \in \mathbb{R}^{m\times N}$	Cellular Objects with morphometric properties
Input	$\lambda 1$	Sparsity regularization constraint
Input	$\sigma(\cdot)$	Element-wise sigmoid function
Output	$\mathbf{D} = [d_{1,\dots,}d_h] \in \mathbb{R}^{m\times h}$	Cellular morphometric biomarkers
Output	$\mathbf{Z} = [z_{1,\dots,}z_N] \in \mathbb{R}^{h\times N}$	Sparse matrix
Output	$\mathbf{W} \in \mathbb{R}^{h\times m}$	Auto-encoder
Output	$\mathbf{G} = diag(g_{1,\dots,}g_h) \in \mathbb{R}^{h\times h}$	Scaling matrix

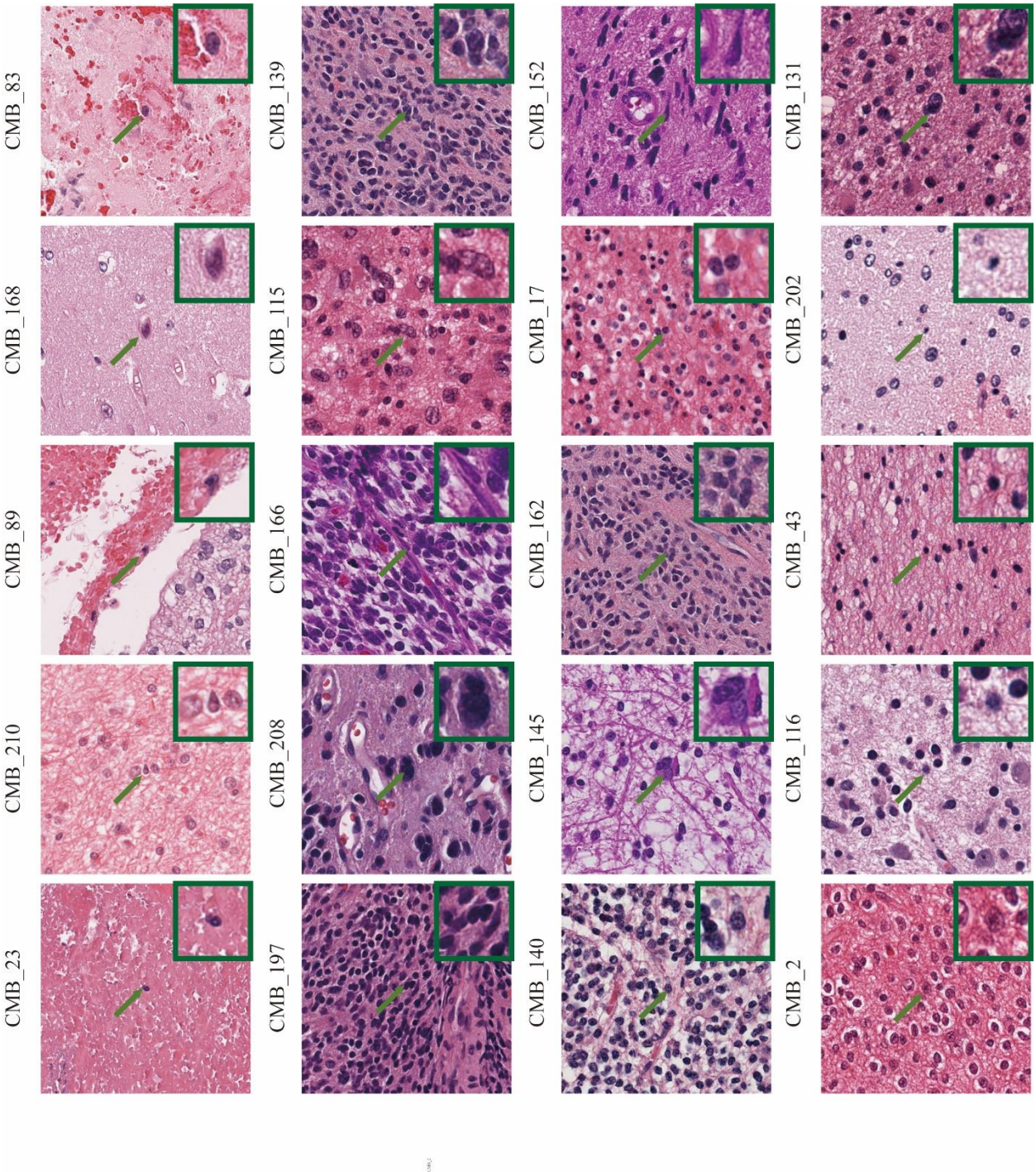
b



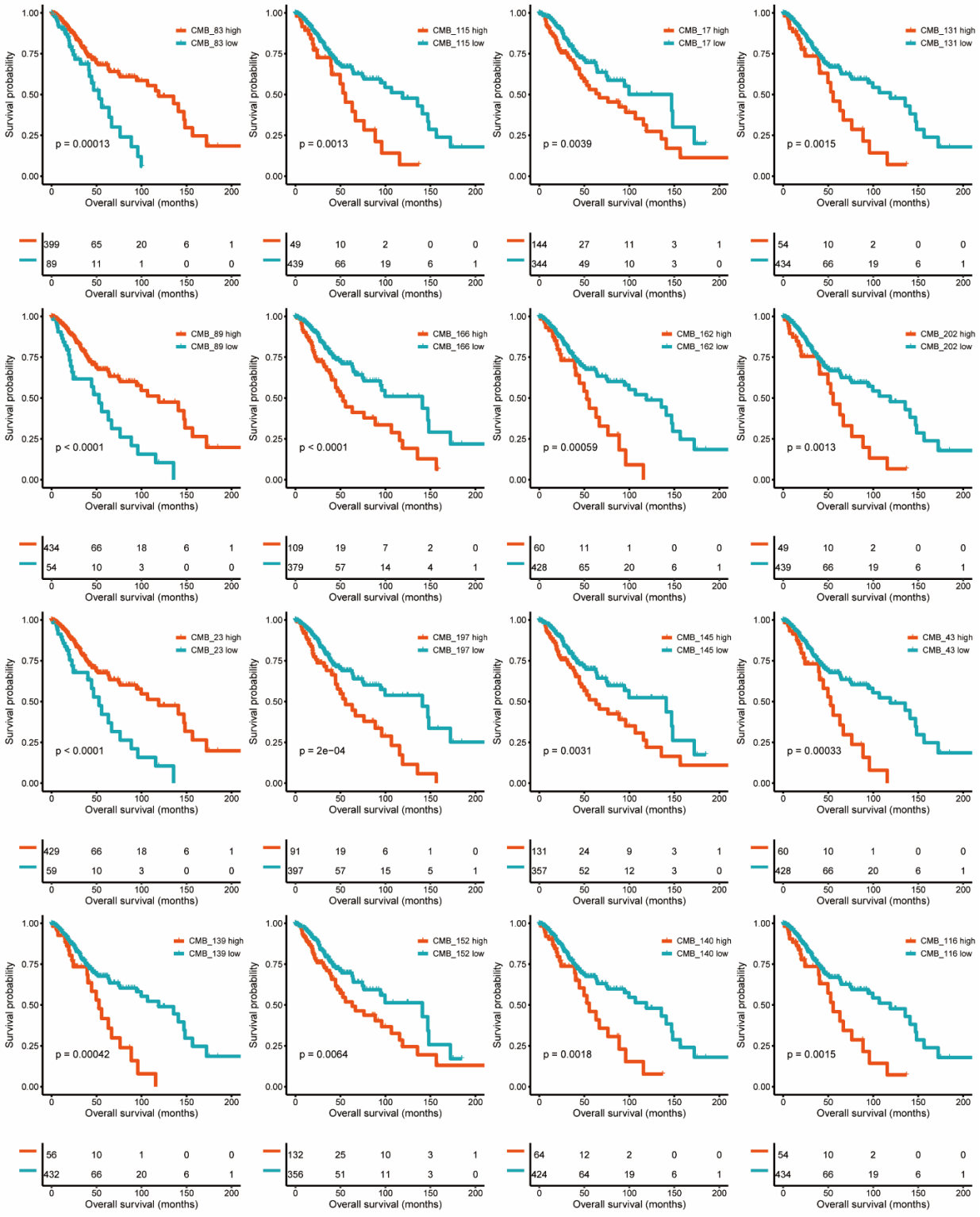
c



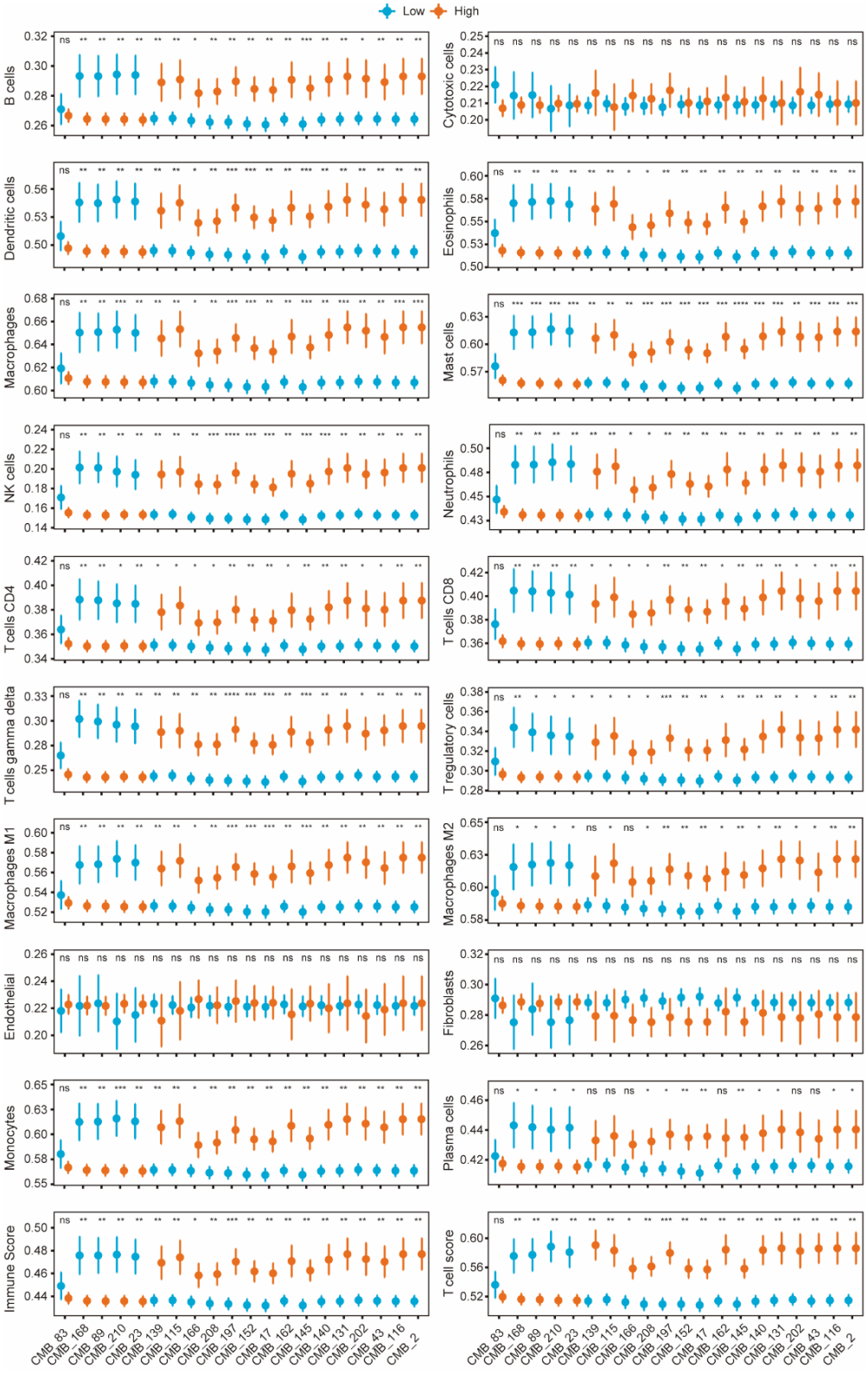
Extended Data Figure 2. **a.** Algorithmic flowchart for our unsupervised representation pipeline; **b.** Evaluation of input reconstruction error ratio during training with different combinations of dictionary size and sparsity; **c.** Final input reconstruction error ratio after training (at 100 iterations) with different combinations of dictionary size and sparsity.



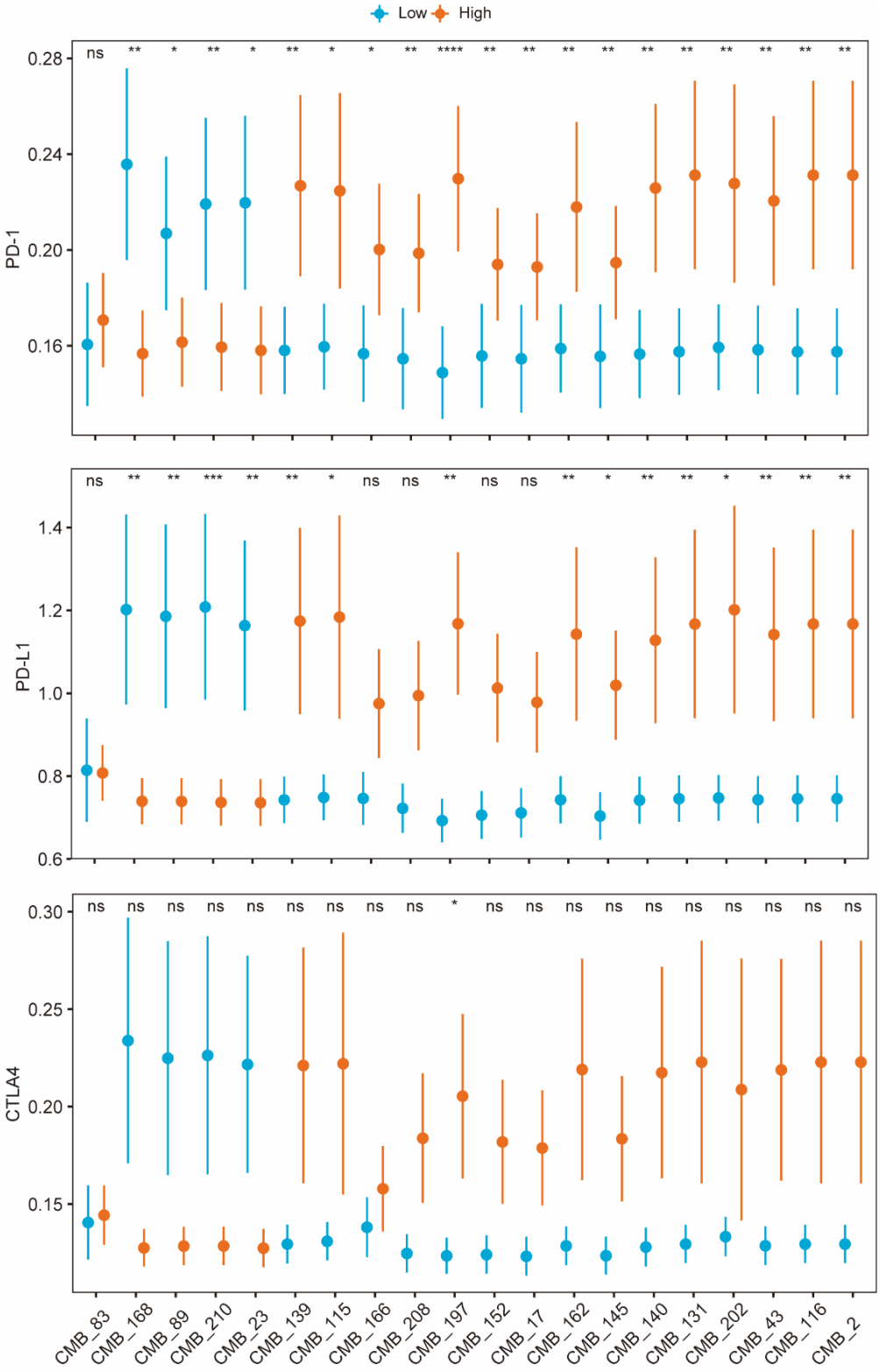
Extended Data Figure 3. **Examples of prognostically significant CMBs.** CMBs (pointed out with green arrows) are displayed within 75.852 μm x 75.852 μm neighborhood, where the top right zoom-in box illustrated each CMB in 25.452 μm x 25.452 μm neighborhood, and the CMBs are sorted ascendingly based on their hazard ratio in column-major order.



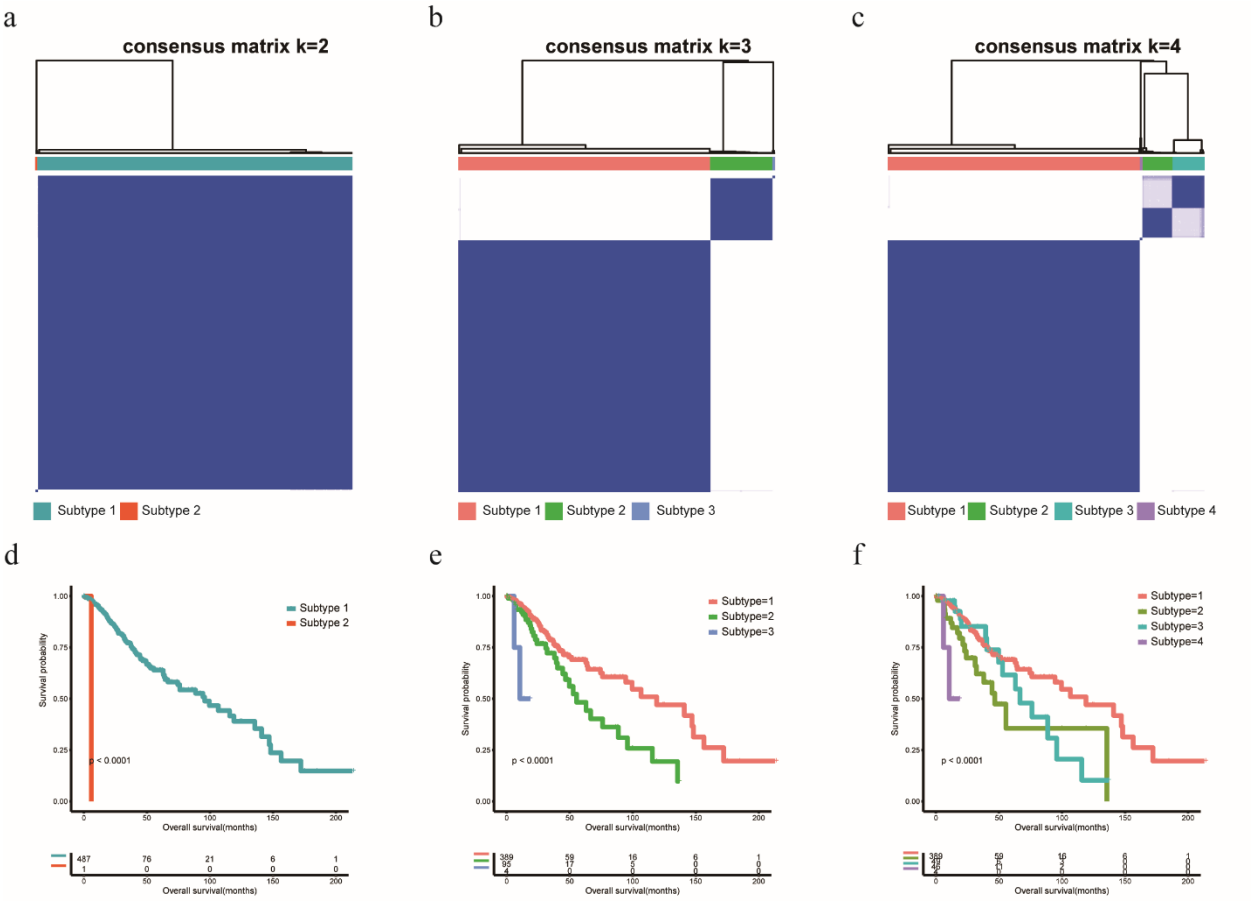
Extended Data Figure 4. The prognostic impact of the level of cellular morphometric biomarker (CMB) on TCGA-LGG cohort. TCGA-LGG patients within CMB-low and CMB-high groups (cutoffs summarized in Supplementary Table 12) show significant difference in survival.



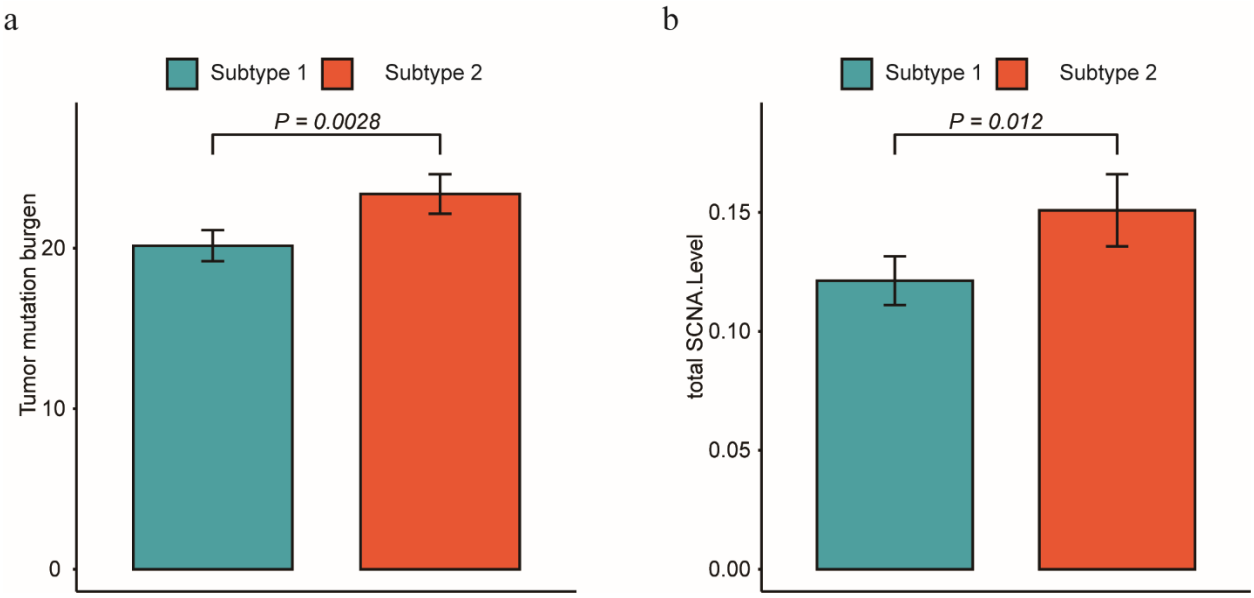
Extended Data Figure 5. TCGA-LGG patients within high and low cellular morphometric biomarker groups (cutoffs summarized in Supplementary Table 12) show significant difference in various tumor microenvironmental factors.



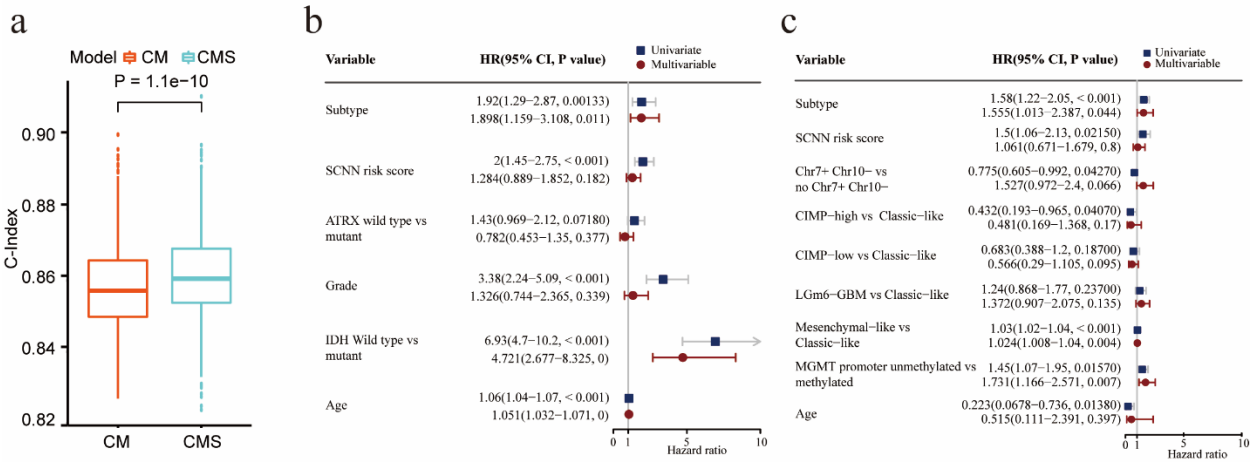
Extended Data Figure 6. TCGA-LGG patients within high and low cellular morphometric biomarker groups (cutoffs summarized in Supplementary Table 12) show significant difference with important biomarkers for immunotherapy.



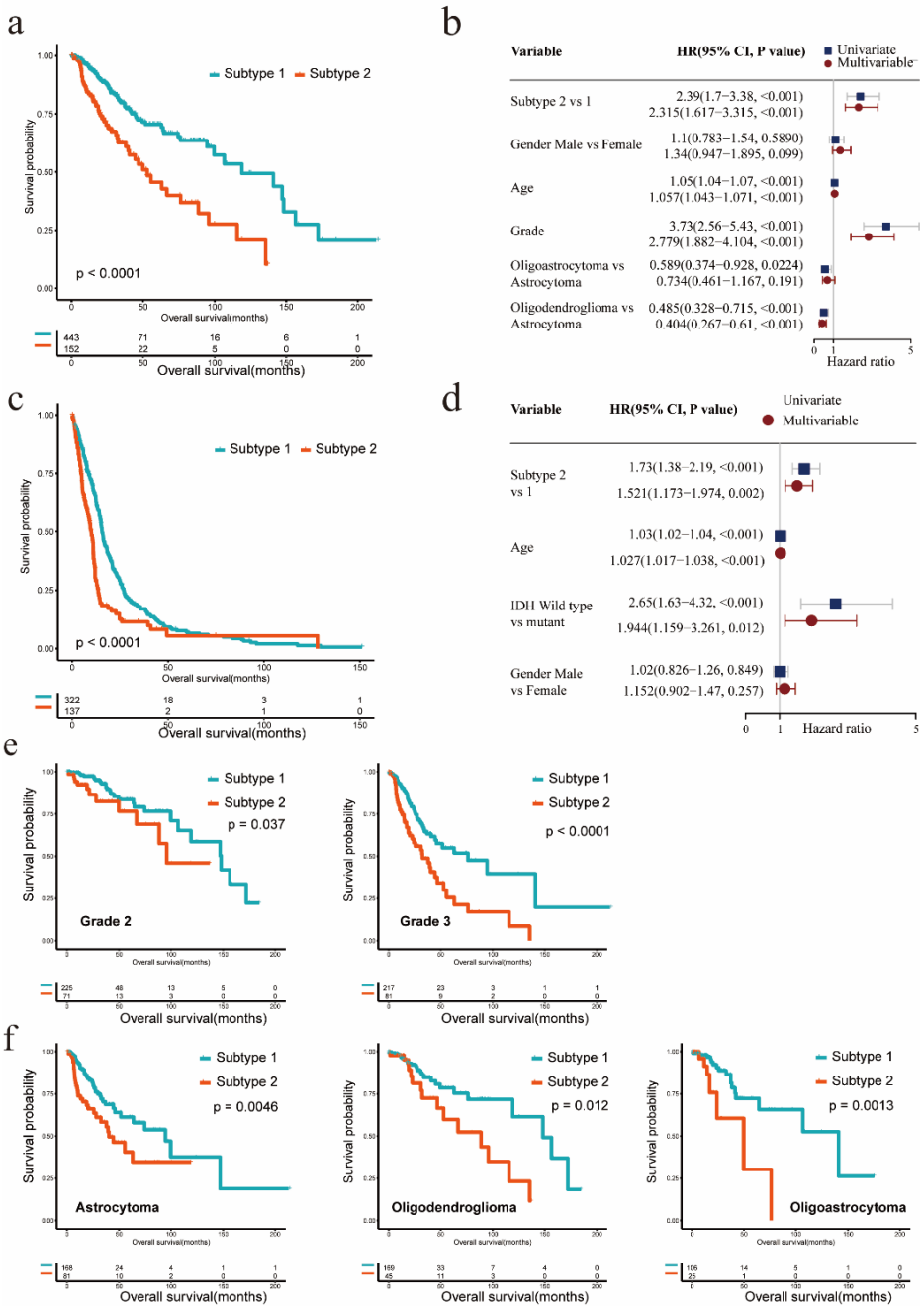
Extended Data Figure 7. **Subtyping model construction with consensus clustering in TCGA-LGG cohort**. **a, d.** consensus matrix and the corresponding Kaplan–Meier curves with two clusters. **b, e.** consensus matrix and the corresponding Kaplan–Meier curves with three clusters. **c, f.** consensus matrix and the corresponding Kaplan–Meier curves with four clusters.



Extended Data Figure 8. Patient subtypes in TCGA-LGG cohort show **a.** significant difference in tumor mutation burden; and **b.** somatic copy number alteration (SCNA).



Extended Data Figure 9. The combination of cellular morphometric subtypes with important clinical and molecular factors (CMS) provides significantly improved predictive power of OS than classical model (CM) with clinical and molecular factors only (**a**). And univariate and multivariate CoxPH analysis in TCGA-LGG (**c**) and TCGA-GBM (**d**) cohorts with SCNN score, cellular morphometric subtype (Subtype) and other important clinical/molecular factors.



Extended Data Figure 10. Analysis on Pooled-LGG cohort, combining TCGA-LGG, ZN-LGG, and SU-LGG patients, confirms the consistency of CMS in terms of the significant difference in OS of all LGG patients (**a**), as well as its significant and independent prognostic value (**b**); analysis on Pooled-GBM cohort, combining TCGA-GBM, and ZN-GBM patients, confirms the consistency of our findings in terms of the significant difference in OS of all GBM patients (**c**), as well as the significant and independent prognostic value of CMS in Pooled-GBM cohort (**d**). The significant difference in OS between CMSs in Pooled-LGG cohort is independent from Grades (**e**), and histology types (**f**). Note, during pooled analysis, only factors available in all cohorts are included.