Given the complexity of biological systems, machine learning methods are critically needed for building systems models of cell and tissue behavior and for studying their perturbations. Such models require accurate information about the subcellular distributions of proteins, RNAs and other macromolecules in order to be able to capture and simulate their spatiotemporal dynamics. Microscope images provide the best source of this information, and we have developed tools to build generative models of cell organization directly from such images. Generative models are capable of producing new instances of a pattern that are expected to be drawn from the same underlying distribution as those it was trained with. Our open source system, CellOrganizer, contains components that can build probabilistic generative models of cell, nuclear and organelle shape, organelle position, and microtubule distribution. These models capture heterogeneity within cell populations, and can be dependent upon each other and can be combined to create new higher level models. The parameters of these models can be used as a highly interpretable basis for analyzing perturbations (e.g., induced by drug addition), and generative models of cell organization can be used as a basis for cell simulations to identify mechanisms underlying cell behavior. Once a common framework for representing the effects of perturbagens can be created, the next step is to learn a comprehensive model that describes the effect of large numbers of potential perturbagens (e.g., small molecule compounds or inhibitory RNAs). We could try to build such a model by measuring all combinations of perturbagens and targets, but the experimentation required would be enormous. We have therefore developed “active” machine learning approaches that iteratively select experiments to perform in order to improve the best model currently available. Results in test cases show that very accurate models can be built with significantly fewer measurements than exhaustive screening.