With a number of single cell technologies entering the market, there is now the possibility to measure single cell protein, DNA, and functional profiles over time. Previously, cell type was defined as a series of protein markers identified. With increasing information, including dynamic data and thousands of markers instead of dozens, how should we classify cells in the future? How should we think about cell identity and what is its utility? See interesting opinions from lead researchers already in the field from a special issue of Cell

Cell type is one way to classify a cell and necessary in providing a simplification of rather complicated biological systems. Yet, as highlighted by J.Sanes of Harvard, defining them by one feature is less important than for instance looking at variations among groups. Rather, an integrated approach, as proposed by E. Lundberg, M.Uhlen, and others, considers not only cell types but also cell states, protein composition, protein activities, localization, post-translational modifications, interactions and pathway functions. I agree with the perspective shared by J. Kim from the University of Pennsylvania, who argues that cell properties relevant to their roles at the system level are more important.

In translational research, cells and their types are one piece of information along with clinical datapoints, such as lab results, survival rates, or drug response. For a variety of reasons, including cost, most of the RNA data in tumor studies, is bulk RNA transcripts which are average estimate of single cell RNA expression levels. Single cell deconvolution provides a workaround by estimating cell type ratios within tissue RNA, yet it is only an approximation. Furthermore, the whole human genomic sequencing is about 19,000 genes, this large number of features increase the difficulty of data analysis. Dimension reduction is the next natural processing step in managing this complexity, and numerical mathematical algorithms do it efficiently without losing the information present into the initial data. However, relating the outcomes using this latent space to the original features is not always straightforward. Therefore, I would agree with A. Martinez-Arias from the University of Cambridge, that “we do not have a way to capture this information in a satisfactory manner, and we really need some conceptual breakthrough”.