1. **Sort-of-basics**: Consider the venerable p-value. It is ubiquitously used by the bioinformatics community to determine the viability of their hypotheses results. However, it is running into a maelstrom of criticism as noted in the following commentaries found in the first two links. The other two in the list below are more informed manuscripts that dwell on this controversy.   
     
   You are asked to consider the criticism in the context of single-cell technologies. How will the large inherent dimensionality and relatively large samples (number of cells) will impact hypotheses testing in general for populations of cells ? Will the use of CI, or confidence interval be of more value ? Please carefully answer this question after perusing the commentaries and critiques. You are also welcome to peruse other material.
   1. <http://debunkingdenialism.com/2015/04/01/new-nature-methods-paper-argues-that-p-values-should-be-discarded/>.
   2. <http://www.nature.com/news/statistics-p-values-are-just-the-tip-of-the-iceberg-1.17412>
   3. <http://www.nature.com/news/scientific-method-statistical-errors-1.14700>
   4. <http://www.nature.com/nmeth/journal/v12/n3/full/nmeth.3288.html>

2. **A-survey-of-sorts**: Many argue that single-cell methods are here to stay as stated in the following publications:

1. <http://www.ncbi.nlm.nih.gov/pubmed/22323135>
2. <http://www.nature.com/nmeth/journal/v9/n1/full/nmeth.1819.html>

There is much work on the actual acquisition of the single cell measurements through appropriate micro-fluidics and chemistry. However, there is a paucity of work and surveys on techniques of analysis. Now the questions -

1. Given your understanding of acquisition technologies, please provide a systematic and mathematical formal description replete with symbols and detailed formulation of available signal and confounding noise.
2. Please list all the sources of noise, outliers, and confounding (and possibly latent) factors.
3. What are the essential and underlying reasons for the sources of noise and how are they characterized ?
4. List various quantitative models of signal that have been reported and what are their deficiencies.
5. What are the main questions that are sought with the quantitative models ? How well have they been answered.
6. You are welcome to consult any manuscript to answer this question including these two:
7. Quantitative assessment of single-cell RNA-sequencing methods, *Nature Methods* **11**, 41–46 (2014) doi:10.1038/nmeth.2694 <http://www.nature.com/nmeth/journal/v11/n1/full/nmeth.2694.html>
8. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells *Nature Biotechnology* **33**, 155–160 (2015), doi:10.1038/nbt.3102 <http://www.nature.com/nbt/journal/v33/n2/full/nbt.3102.html>

3. **Use-of-hypervariability**: Very large variance or hypervariability in expression can be used in differential studies as demonstrated in the following two manuscripts:

1. <http://www.ncbi.nlm.nih.gov/pubmed/23088656>
2. <http://www.ncbi.nlm.nih.gov/pubmed/26078586>

The authors essentially propose the use of anti-profiles. The premise of this method is that tumor phenotypes are best characterized by extreme variability.

Some of your own proposed methods do not take this extreme variability into account. For instance, consider the boolean quantification techniques you wish to borrow and use.

Evaluate the notion of hyper-variability in the context of single-cell measurements. If you believe that this notion has no merit what-so-ever, please provide careful reasoning to support your stand. Otherwise, propose ways to include hypervariability in your methods. Also, how will you extend this idea to co-expression networks. Could you one discuss extreme hypervariability of “functional groups” and “networks”. What are the possible perils and useful outcomes of such an approach ?