

What the Genomics Revolution Means to You and Your Research

Seyeon Weon

Bioinformatics Research Laboratories. Co.

Genomics is often ornamented with the word revolution. The question is what are the things that we need to prepare to join the revolution. The field so-called proteomics is no different at all in the aspects discussed here.

First, large quantity of data is involved. It implies the use of computer, which again implies expertise in computer including database management, at least some script language programming, and system management skills. We can no longer get by with a pencil and notebook, or simple Excel spreadsheet, or even Access database. Second, the data is not isolated. All biological data is intrinsically interconnected to each other. We must always consider the interconnections when doing something about the data. It implies that we need to have an integrated database system that can deal with these interconnections.

Go back to the production of the data. We always use some kinds of automatic machines for the data production. Needless to mention, the expertise in the manipulation of the machines is required. Also, a large number of samples are involved in the data production. Therefore, we need some method to deal with these samples physically. It always implies that some computer system is required to support this activity since the number of samples usually far exceeds what can be traced manually.

Go back to the data itself again. We are, of course, producing the data to obtain some useful knowledge. How this can be done might be the most wanted answer to hear. First, what we must admit is that there is no simple one-to-one correspondence between some useful knowledge and a point in the data. For example, we used to collect useful knowledge (or at least the basic components of the knowledge) by observing a shifting of a band in a lane (or a handful of bands in a few lanes) of an agarose gel. Also, our ingenuity used to be mainly exerted to design the experiment as neatly as possible to obtain a few shifting of bands that support the hypothesis in our hands. Bluntly speaking, things are not working this way any more in the genomics or proteomics world. This is one of the most profound changes by the genomics revolution for the researchers like us.

Then, what and how we should do? Of course, we don't have all the answers right now and we will be searching for the better answers at least for some time now. However, there are a few obvious facts. First, since it is no longer the search for one-to-one correspondence and many data points are involved as well, we need some degree of expertise in statistics at least for the proper design of experiment. Second, since it is no longer possible to analyze, for example, all the spots with our own eyes and also the comparison of the numbers from the spots involves some kind of intricate computer algorithms, we need some degree of understanding for these algorithms. We should not forget the fact that this step is no longer the one that provides us only an extra convenience as before. It has some fundamental influence on the overall results, so we should have the characteristics and limitations of the algorithms in mind from the beginning of the design of experiment. Third, we should of course know how to use the software packages in which the algorithms are implemented. No need to mention that we should have the general knowledge and skills that are commonly involved in the analysis of large amount of data using computer. Since these steps are the major, if not the most important, part of the research itself, it is not always possible to delegate.

Let us conclude this presentation with a short introduction to a more profound issue; what do we have to look for with this powerful, albeit hard to access, new way? The type of

scientific query we ask should not be such one as Does protein A bind to protein B? any more. The query should be more of systems property in which above question is a small part. This may be the most fundamental benefit of genomics or proteomics, since we can finally escape from being the blind men touching an elephant.