

Literature assignment 1: Hawkins RD, Hon GC, Ren B. *Next-generation genomics: an integrative approach*. Nat. Rev. Genet. 2010;11:47686.

Due September 27th

As experimental techniques advance and experimental data sizes increase dramatically, it is possible to integrate different types of genomic data to answer biological questions. This article gives examples of what data integration can achieve and describes several methods of data integration. Because functional genomic data sets tend to be extremely large, the authors also describe methods to deal with data complexity. This article provides a foundation for further understanding of bioinformatic data analysis.

Read this article and briefly answer the following questions. You may read additional materials if you wish. If you do, you must cite your sources. **Please turn in your assignment in class or in person to Annette McLeod in MI 646 or electronically to comp-bio@cs.cmu.edu by 5pm on the due date.**

1. Briefly explain the four types of data that are currently available using high-throughput technologies. For each type, describe in one or two sentences what the data is used for and what method or technique can be used to obtain the data?
2. “Integrative analysis” is used to answer some fundamental questions regarding the mechanisms of genome function and disease. Describe the three types of integrative analysis (purpose, method), and give an example of how each analysis technique is used. Your answer should be based on information in the article, not from an online search.

3. When doing integrative analysis, repeating experiments and incorporating other types of data can greatly increase the data complexity.
 - (i) Based on the article, give a brief description of data complexity reduction and an example of how it is used in ChIP-seq.
 - (ii) Unsupervised integration and supervised integration: What are the general steps for unsupervised integration? What is the difference between unsupervised and supervised integration?