14

Bayesian Graphical Models for Integrating Multiplatform Genomics Data

WENTING WANG, VEERABHADRAN BALADANDAYUTHAPANI, CHRIS C. HOLMES, AND KIM-ANH DO

14.1 Introduction

The major known genomic alterations related to cancer include nucleotide substitution mutations, small insertions/deletions, copy number gains and losses, chromosomal rearrangements, and nucleic acids of foreign origin. Early genomics studies focused on examining only one type of genomic alteration at a time and achieved some success. For example, copy number variations have enabled the discovery of many oncogenes in ovarian cancer (Nanjundan et al., 2007), melanoma (Scott et al., 2009), and lung carcinoma (Bass et al., 2009). Similarly, directed sequencing technologies have found many genes related to specific types of cancer (Pao et al., 2004; Stephens et al., 2004; Mosse et al., 2008).

However, because different types of genomic alteration illuminate different aspects of the cancer genome, we can integrate several types of alteration derived from the same set of tumors to determine important genes involved in cancer initiation, development, and progression. There are two main advantages of such integration studies. The first is that integration can increase the precision, accuracy, and statistical power of identifying cancer-related genes compared with analyzing any single type of alteration. The reason for this is that cancer is thought to be primarily caused by random genetic alterations via different mechanisms. Although each type of alteration may be rare, the cumulative number of different alterations can indicate that a gene is important in a certain cancer. For example, The Cancer Genome Atlas (TCGA) glioblastoma project integrated targeted sequencing, copy number, and expression profiling of more than 400 tumor samples to define core pathways of deregulation in glioblastoma (The Cancer Genome Atlas Network, 2008) and discovered four molecular subtypes (Noushmehr et al., 2010; Verhaak et al., 2010).

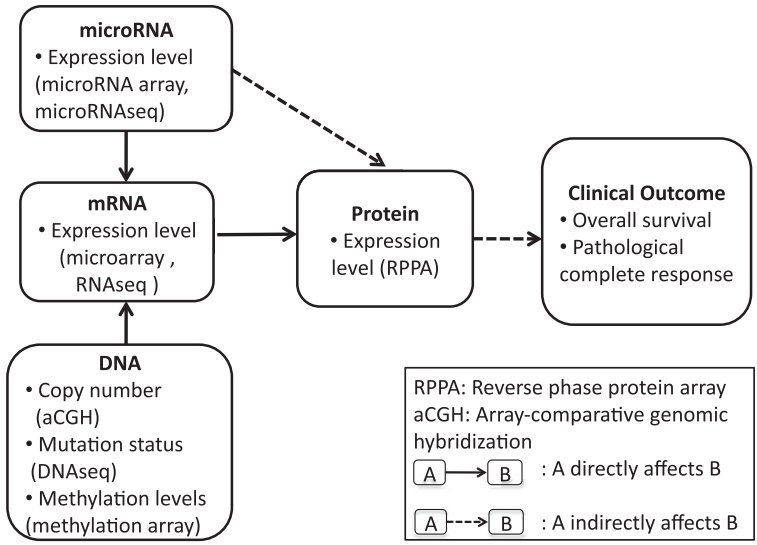

Figure 14.1 The relationships among different molecular features and clinical outcome. Common technologies to obtain each molecular feature are listed within parentheses.

The second advantage of integration studies is to improve the understanding of complex networks of biological processes underlying cancer and how these networks affect patient clinical outcome (pathological complete response, progression-free survival time, etc.). For example, Zhu et al. (2004) reconstructed gene networks by integrating gene expression and genetic data. Talluri et al. (2011) constructed protein signaling pathways using reverse-phase protein array (RPPA) data. Stingo et al. (2010) integrated microRNAs with the expression of their target genes information to infer microRNA-regulated networks.

Figure 14.1 shows the relationships among different molecular features and their associations with clinical outcome. The overall underlying biological mechanisms are (1) that molecular features measured at the transcriptional level (e.g., mRNA expression) affect clinical outcome (e.g., patient overall survival time and pathological complete response) more directly than molecular features measured at the DNA level (e.g., copy number, methylation, and mutation status); and (2) that the molecular features related to post-translational modification (e.g., protein expression) affect clinical outcomes more directly than mRNA expression. For example, molecular features measured at the DNA level affect clinical outcome by regulating mRNA expression (Glinsky, 2006; de Tayrac et al., 2009; Fabiani et al., 2010). Similarly, microRNAs are