Journal of
Health & Medical Informatics

Kim and Ritchie, J Health Med Informat 2014, 5:4
DOI: 10.4172/2157-7420.1000e122

Editorial Open Access

Data Integration for Cancer Clinical Outcome Prediction

Dokyoon Kim and Marylyn D Ritchie*

Center for Systems Genomics, Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania, USA

Cancer clinical outcome prediction based on the molecular information has received increasing interest for better diagnostics, prognostics, and further therapeutics. Accurate molecular-based predictors of outcome can be used clinically to choose the best of several available therapies for a cancer patient. In the past decade, gene expression profiles have been most widely used to predict clinical outcomes in several cancers [1,2]. There have been also many attempts at cancer clinical outcome prediction using a set of copy number alterations (CNA), miRNA, DNA methylation, and protein expression [3-6].

However, it is still difficult to accurately predict clinical outcome since the cancer genome is neither simple nor independent but rather complicated and dysregulated by multiple levels of the biological system through genome, epigenome, transcriptome, proteome, metabolome, interactome, etc. [7]. For instance, cancer is mainly caused by somatic driver mutations in coding and non-coding sequences or epigenetic changes of methylation, acetylation, and histone. Collectively, these genetic and epigenetic changes can lead to many alternative forms of cause-and-result effect in transcription, translation, and post-translational modification, which are all involved in cancer pathophysiology. Therefore, no single type of genomic data will be sufficient to elucidate the phenotypic end-point of events accumulated through multiple levels of biological systems involved in cancer, and hence, a consideration of incorporating the multi-layered processes in biological systems might provide much more reasonable prediction of cancer clinical outcome.

Recently, emerging multi-omics data and clinical information from cancer patients have been providing unprecedented opportunities to investigate the multi-layered processes involved in cancer development and progression for improving the ability to diagnose, treat, and prevent cancer. The Cancer Genome Atlas (TCGA) is a large-scale collaborative initiative to improve our understanding of multi-layered of molecular basis of cancer and has been generating multi-omics data for 25 cancer types [8]. In addition, the International Cancer Genome Consortium (ICGC) is another comprehensive collaborative initiative to obtain a multidisciplinary description of genomic, epigenomic, and transcriptomic changes in 50 different cancer types [9]. Before exploding multi-omics data in cancer from TCGA or ICGC, there have been many integrative studies for two types of genomic data such as association, regression, or correlation-based methods [10]. However, as multi-scale genomic data have become more available, it is hard to directly use existing integrative methods, which are mainly for two types of genomic data. Thus, the development of multi-scale integrative approaches is more required in order to integrate multiple types of genomic data at hand and investigate an enhanced global view on interplays between different types of genomic data.

In order to solve the current problems for data integration in cancer research, many multi-scale integrative approaches have been recently proposed. Kim et al. proposed a graph-based integration framework for predicting cancer clinical outcomes using CNA, methylation, miRNA, and gene expression data [11]. Sohn et al. [12] proposed an integrative statistical framework based on a sparse regression to model the impact of multi-layered genomic features including CNA, miRNA, and methylation on gene expression traits. Kim et al. [13] also investigated an integrative framework in order to identify interactions between different types of genomic data associated with clinical outcome. In addition, Mankoo et al. [14] predicted time to recurrence and survival in ovarian cancer using CNA, methylation, miRNA, and gene expression data using multivariate Cox Lass model.

While the TCGA and ICGC provide many opportunities to uncover the novel knowledge of the molecular basis of cancer, it is crucial to address the issue of development of an appropriate methodological framework for data integration to better understand different cancer phenotypes, further providing an enhanced global view on the interplays between different genomic features. With an abundance in of multi-omics data and clinical data from cancer patients, relevant integration frameworks will be valuable for explaining the molecular pathogenesis and underlying biology in cancer, eventually leading to more effective screening strategies and therapeutic targets in many types of cancer.

Acknowledgement

This work was funded by NIH grant 5R01 LM010040 and NHLBI grant 2U01 HL065962.

References

1. Fan X, Shi L, Fang H, Cheng Y, Perkins R, et al. (2010) DNA microarrays are predictive of cancer prognosis: a re-evaluation. Clin Cancer Res 16: 629-636.
2. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, et al. (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature 415: 530-536.
3. ten Berge RL, Meijer CJ, Dukers DF, Kummer JA, Bladergroen BA, et al. (2002) Expression levels of apoptosis-related proteins predict clinical outcome in anaplastic large cell lymphoma. Blood 99: 4540-4546.
4. Zhang Y, Martens JW, Yu JX, Jiang J, Sieuwerts AM, et al. (2009) Copy number alterations that predict metastatic capability of human breast cancer. Cancer res 69: 3795-3801.
5. Deneberg S, Grovdal M, Karimi M, Jansson M, Nahi H, et al. (2010) Gene-specific and global methylation patterns predict outcome in patients with acute myeloid leukemia. Leukemia 24: 932-941.
6. Nair VS, Maeda LS, Ioannidis JP (2012) Clinical outcome prediction by microRNAs in human cancer: a systematic review. J Nat Cancer Institute 104: 528-540.
7. Hanash S (2004) Integrated global profiling of cancer. Nature reviews Cancer 4: 638-644.

***Corresponding author:** Marylyn D Ritchie, Center for Systems Genomics, Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania, USA, Tel: +1 814 8634467; E-mail: marylyn.ritchie@psu.edu

Received January 16, 2014; **Accepted** January 18, 2014; **Published** January 24, 2014

Citation: Kim D, Ritchie MD (2014) Data Integration for Cancer Clinical Outcome Prediction. J Health Med Informat. 5: e122. doi:10.4172/2157-7420.1000e122

8. (2008) TCGA Network: Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455: 1061-1068.

9. Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, et al. (2010) International network of cancer genome projects. Nature 464: 993-998.

10. Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, et al. (2002) Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc Natl Acad Sci U S A 99: 12963-12968.

11. Kim D, Shin H, Song YS, Kim JH (2012) Synergistic effect of different levels of genomic data for cancer clinical outcome prediction. J Biomed Inform 45: 1191-1198.

12. Sohn KA, Kim D, Lim J, Kim JH (2013) Relative impact of multi-layered genomic data on gene expression phenotypes in serous ovarian tumors. BMC systems biology 7: S9.

13. Kim D, Li R, Dudek SM, Ritchie MD (2013) ATHENA: Identifying interactions between different levels of genomic data associated with cancer clinical outcomes using grammatical evolution neural network. Bio Data Mining 6: 23.

14. Mankoo PK, Shen R, Schultz N, Levine DA, Sander C (2011) Time to recurrence and survival in serous ovarian tumors predicted from integrated genomic profiles. PLoS One 6: e24709.