ECRONICON OPEN ACCESS

EC PULMONOLOGY AND RESPIRATORY MEDICINE

Short Communication

Does the Current Database Catch Up with Real World Medicine?

Xiaoshun Shi[1,2] and Kaican Cai[1]*

[1]*Department of Thoracic Surgery, Nanfang Hospital, Southern Medical University, Guangzhou, China*

[2]*Harry Perkins Institute of Medical Research, QEII Medical Centre and Centre for Medical Research, University of Western Australia, Nedlands, WA, Australia*

***Corresponding Author:** Kaican Cai, Department of Thoracic Surgery, Nanfang Hospital, Southern Medical University, Guangzhou, China.

Received: March 25, 2019; **Published:** April 12, 2019

Cancer is a common malignancy with high morbidity and mortality [1]. The development of high-throughput sequencing in recent years has helped elucidate the genome, transcriptome, methylome and other genetic/expression data of cancer cells [2]. This unprecedented amount of data has not only improved our understanding of the development and progression of cancer, but also unearthed new therapeutic targets and diagnostic markers. However, the individual clinical and genetic information of cancer patients are often not integrated in the face of continuous big data output, resulting in separate clinical and genetic databases which hinder optimum translation of the available data.

Some widely used clinical databases in oncology studies are SEER (Surveillance, Epidemiology and End Results) [3] and the National Cancer Database (NCDB) [4]. The former was established by the National Cancer Institute in 1973 and is one of the most representative large-scale tumor registration databases in North America. It records the patient registration number, personal data, primary lesion location, tumor size, tumor code, treatment plan, the cause of death etc. The NCDB records the demographic characteristics (age, gender, ethnicity, type of insurance, etc.), comorbidities, tumor characteristics, patient survival data, and first-line treatment including surgery, RT, chemotherapy etc [5]. Studies conducted on the basis of these databases consist of the following steps: 1) patient grouping according to treatment methods, 2) generating Kaplan-Meier survival curves to determine patient prognosis in response to different therapies, and 3) performing univariate and multivariate Cox regression analysis to explore other prognostic factors. In other words, a prognostic model is established using clinical parameters to predict the outcome of treatment. However, there is considerable heterogeneity in cancer clinical data in terms of ethnicity, and a model developed using the above databases may not be extrapolated to the Asian cohort. In addition, the lack of genomic information in these clinical databases limits the validation of the prognosis models.

Databases like The Cancer Genome Atlas (TCGA) and the tumor alterations relevant for genomics-driven therapy (TARGET) include transcriptome and methylome data, along with some aspects of the corresponding clinical data. In general, the steps for establishing a prognostic/survival model using these databases are as follows: 1) the identification of differentially expressed genes (DEGs), 2) univariate analysis for each DEG, and 3) multivariate survival regression analysis with DEGs of prognostic significance [6]. In contrast, the exosomal databases [7] and the COSMIC database [8] lack corresponding clinical and prognostic data, which limits the predictive power of the differentially expressed genes and RNAs. Combining both clinical and genetic data can significantly increase the accuracy of predicting treatment outcomes. At the ASCO Annual Meeting 2017, Foundation Medicine and Flatiron Health jointly released the Clinico-Genomic Database, which integrates the diagnostic, therapeutic and prognostic data of cancer patients in the Flatiron Health platform with the genomic data from 100,000 cancer patient samples sequenced by Foundation Medicine [9]. Preliminary studies indicate better prognostic value of this clinico-genomic model compared to models based on clinical covariates alone [10].

TCGA and the Clinico-Genomic Database are databases with translational potential. We propose that, under the regulation of the local ethics committee and with patient consent, future databases should include the precise demographic, clinical and genomic information of individual patients to optimize the use of patient data. Big data models can not only help cancer patients with early intervention and prevention, but also enable personalized diagnosis and treatment.

Bibliography

1. Siegel RL., *et al*. "Cancer statistics, 2019". *CA: A Cancer Journal for Clinicians* 69.1 (2019): 7-34.
2. Goodwin S., *et al*. "Coming of age: ten years of next-generation sequencing technologies". *Nature Reviews Genetics* 17.6 (2016): 333-351.
3. Cronin KA., *et al*. "The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute". *Cancer* 120.23 (2014): 3755-3757.
4. Fremgen AM., *et al*. "Clinical highlights from the National Cancer Data Base, 1999". *CA: A Cancer Journal for Clinicians* 49.3 (1999): 145-158.
5. Boffa DJ., *et al*. "Using the National Cancer Database for Outcomes Research: A Review". *JAMA Oncology* 3.12 (2017): 1722-1728.
6. Shi X., *et al*. "An expression signature model to predict lung adenocarcinoma-specific survival". *Cancer Management and Research* 10 (2018): 3717-3732.
7. Li S., *et al*. "exoRBase: a database of circRNA, lncRNA and mRNA in human blood exosomes". *Nucleic Acids Research* 46.D1 (2018): D106-D112.
8. Forbes SA., *et al*. "COSMIC: somatic cancer genetics at high-resolution". *Nucleic Acids Research* 45.D1 (2017): D777-D783.
9. Singal G., *et al*. "Development and validation of a real-world clinicogenomic database". *Journal of Clinical Oncology* 35.15 (2017): 2514-2514.
10. Bovelstad HM., *et al*. "Survival prediction from clinico-genomic models--a comparative study". *BMC Bioinformatics* 10 (2009): 413.

Volume 8 Issue 5 May 2019