The Role of the Electronic Health Record in Support of Genomic Research

Vincent Normandeau-Babin[a], Hanad Nwilati[b], Sherif Abou-Elela[b], Andrew Grant[a]

[a] *Collaborative Research for Effective Diagnostics (CRED), Université de Sherbrooke, Québec, Canada*
[b] *Laboratoire de Génomique Fonctionnelle de l'Université de Sherbrooke (LGFUS), Université de Sherbrooke, Québec, Canada*

Abstract

Electronic health records (EHR) are a source of information which could be used to better understand the chronological evolution of disease for optimising and innovating treatment. In parallel, translational research focuses on the analysis of gene expression signatures (GES) with the aim to diagnose subtypes of diseases in order to predict clinical outcome. The combination of genomic and clinical information is opening the opportunity to explore the concept of clinical phenotype. Here we present the first step of an overall informatics architecture for exploiting phenotyping information from an EHR for genomic research purposes. This work, developed in studies of gene expression for ovarian cancer, describes the use of EHR data to document specimen quality and clinical context with significant impact on the selection of specimens for genomic studies.

Keywords:

Electronic health record (EHR), Data warehouse, Translational research, Clinical phenotype, Gene expression signature (GES), Ovarian cancer, Specimen quality.

Introduction

The electronic health record (EHR) gathers extensive data of different disease populations that make it a valuable source of information to better understand the chronological evolution of disease and to help the development of better diagnostic tests and treatment in the future [1].

The medical field is complex and medical data occurs in many different forms. This makes it difficult to manually explore and extract new knowledge from these data. Predictive data mining is an emerging process of selecting, exploring and modeling large amounts of data from patient information in order to link them to an outcome of interest.

Predictive data mining can also be used for translational research which focuses on the analysis of gene expression signatures (GES) with the aim to diagnose subtypes of diseases in order to predict clinical outcome [2].

In oncologic translational research, GES are obtained from representative specimens of the patient tumour. One important role of the EHR to identify the quality of specimen collection and of the specimen itself has to date not been fully emphasized. Specimens of poor quality increase the risk of spurious results.

Methods

The Master Specimen File (MSF) is a web-based tool that enables a 3 step specimen assessment and quality evaluation based on data from an EHR. The first step consists to confirm the diagnosis of the tumor and to describe the *pathological representivity* of each specimen which we define as the set of clinical and pathological data describing best the specimen to enhance selection for genomic testing. For exemple the MSF contains data relevant to the time of the extraction of the specimen such as the surgical record, a macroscopic and microscopic pathological description (histology, grade, stage) of the specimen, tumour markers and recent therapy like chemotherapy. Overall, the MSF includes about 100 data fields which have been elaborated taking into account actual best practice in biobanking [3]. The second step consists of evaluating the cellular characteristics of the chosen specimens assessed by the percentage of cancerous cell versus normal cell and the ratio of epithelial versus stromal tissue in the specimens. The last step consists in the assessment of molecular integrity of the specimen which is done by a score called RIN for RNA integrity number [4].

Conclusion

By getting acces to the complete set of clinical data strored in the MSF for query, research pathologist can readily find specimens that correspond to a specific research purpose and can then assess specimen quality.

References

[1] Grant A, Plante I, Leblanc F. The TEAM methodology for the evaluation of information systems in biomedicine. Comput.Biol.Med. 2002 5;32(3):195-207

[2] Bellazzi R, Zupan B. Predictive data mining in clinical medicine: Current issues and guidelines. Int J Med Informat 2008 2;77(2):81-97.

[3] Elisa E, Gabrielle B, Jennifer B, Noreen C, Stuart O. Case Studies of Existing Human Tissue Repositories: Best Practices for a Biospecimen Resource for the Genomic and Proteomic Era. ; 2004.

[4] Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Molecular Biology 2006;7(1):3.