Quality of Life through Quality of Information
J. Mantas et al. (Eds.)
IOS Press, 2012

doi:10.3233/978-1-61499-101-4-688

A Data-driven Living Review for Pharmacogenomic Decision Support in Cancer Treatment

Gökhan KARAKÜLAH[a,b,1], Aslı SUNER[a,c,1], Klaus-Peter ADLASSNIG[a,d], Matthias SAMWALD[a,e,2]

[a] *Section for Medical Expert and Knowledge-Based Systems, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Austria*
[b] *Department of Biomechanics, Health Sciences Institute, Dokuz Eylül University, İzmir, Turkey*
[c] *Department of Statistics, Faculty of Science, Dokuz Eylül University, İzmir, Turkey*
[d] *Medexter Healthcare GmbH, Austria*
[e] *Institute of Software Technology and Interactive Systems, Technical University of Vienna, Austria*

Abstract. With drastically decreasing costs of genetic sequencing, it has become feasible to use individual genetic markers to optimize treatment selection in cancer therapy. However, it is still difficult for medical practitioners to integrate these new kinds of data into clinical routine, since available information is growing rapidly. We demonstrate how a blend of manual curation and automated data extraction and evidence synthesis can be used to generate a 'living review', a summarization of current evidence on cancer classification, corresponding genetic markers, genetic tests and treatment options that can be used by clinicians to refine treatment choices. In contrast to a classical review, this automated 'living review' offers the opportunity of automatically updating core content when available data changes, making it easier to keep an overview of the best current evidence. We discuss some of the findings we made while creating a prototype of a 'living review' for colorectal cancer pharmacotherapy.

Keywords. Pharmacogenomics, evidence synthesis, web-based technologies, clinical decision support, oncology

1. Introduction

Clinically defined types of cancer are actually a combination of many different molecular subtypes, based on the mutations that turn normal cells into cancerous cells. The response to medication and chemotherapy might vary drastically among these molecular subtypes. With the price of genetic testing becoming lower, the question of how to best interpret personal genetic data to optimize treatment becomes a key issue. The use of information technologies and shared databases of clinical genetic findings is becoming essential to solve this issue. For example, CancerCommons is an initiative that spearheads the creation and sharing of molecular disease models for different types

[1] These authors contributed equally to this work.
[2] Corresponding author, e-mail: samwald@gmx.at

of cancer [1]. CancerCommons introduces the idea of a 'living review' of biomedical information about currently available disease models, pharmacogenomic markers, ongoing clinical and other information relevant to specific disease that can help medical professionals select tailored treatments for patients who do not respond to standard therapy. Contrary to a standard review paper, this 'living' review is continuously updated and extended to reflect the most recent findings about molecular markers for a specific disease.

Here, we describe the creation of an enhanced living review that blends manually curated data with data that was automatically extracted and aggregated through web-based technologies, as well as highlighting the biomedical insight that was gained by creating such a review. As a driving use-case, we focus on colorectal cancer (CRC). CRC is the third most common cancer type in the world, and its incidence rate is increasing [2, 3]. For example, there currently are around 50,000 deaths attributed to CRC each year in the US alone [2].

2. Methods

In order to extract experimentally identified genes which play a crucial role in the pathogenesis of CRC, Online Mendelian Inheritance in Man (OMIM) text data was downloaded [4]. Additionally, all approved human gene symbols and their aliases were retrieved from the Human Gene Nomenclature (HGNC) web site by using the BioMart online tool [5]. A script was developed to extract statements from OMIM text that contained gene symbols. A script was created to automatically generate a living review of the collected information in HTML format, so that researchers and medical practitioners could screen OMIM statements grouped by their relationships with CRC relevant genes.

To collect pharmacogenomic information about chemotherapeutic drugs commonly used in CRC treatment, we automatically extracted drug information such as drug labels, available dosing guidelines and pharmacogenetic tests from several online databases. A list of chemotherapeutic drugs commonly used in the treatment of CRC was obtained from the official web site of the Colorectal Cancer Association of Canada (CCAC) [6]. To create an integrated knowledge source for the drug list, information including indications, mechanisms of action, pharmacodynamics and as well as alternate names of each drug from the DrugBank database were collected [7]. The Google Refine tool was used for this task, which provided sophisticated functionalities for automated data alignment and data fetching from different online sources [8]. Each drug was also linked to its Food and Drug Administration (FDA) label which provided detailed information regarding the drug in question [9].

Where available, dosing guidelines and pharmacogenomics tests of each chemotherapeutic drug were extracted from the Pharmacogenomics Knowledge Base (PharmGKB) [10]. Additionally, gene-drug, drug-disease and drug-drug relationships relevant to CRC therapy were extracted from PharmGKB.

To extend the content of our integrated knowledge base beyond the knowledge found in existing databases, several decision making algorithms for CRC classification and the classification approaches of CRC syndromes were manually collected and curated by screening the scientific literature and web resources. Classifications and decision rules based on clinical and molecular features were extracted from Mayo

Clinic's Molecular Genetic Laboratory online source [11], the CancerCommon's CRC disease model [12], and ARUP Laboratories [13].

3. Results

The majority of datasets and source codes (excluding those subject to license restrictions) we created are available from [14].

We identified 43 different CRC related genes after screening OMIM text data which were included in the living review. We also identified 11 major chemotherapeutic drugs used in the treatment of CRC (capecitabine, fluorouacil, leucovorin, irinotecan, mytomycin, oxaliplatin, tegafur, raltitrexed, bevacizumab, cetuximab and panitumumab).

After gathering the data sets obtained from different online databases, we found that there currently were pharmacogenomic tests available for 5 of these drugs (capecitabine, fluorouracil, irinotecan, cetuximab and panitumumab). We were also able to collect pharmacogenomic dosing recommendations for the drugs capecitabine, fluorouracil and tegafur. Beside this, we determined relationships between variations in genes and resultant changes in drug efficacy/safety for the gene-drug pairs CCND1-cetuximab, GSTP1-fluorouracil/oxaliplatin and ERCC2/KLC3-fluorouracil/leucovorin (efficacy). We included these drugs in the living review.

Our analysis of molecular CRC disease classification methodologies resulted in the identification of two distinct major classification methods. The first method was based on mutation patterns of cancerous tissue, as described in the CancerCommon's CRC disease model. This method classifies CRC into molecular subtypes based on mutations of specific genes (Figure 1).

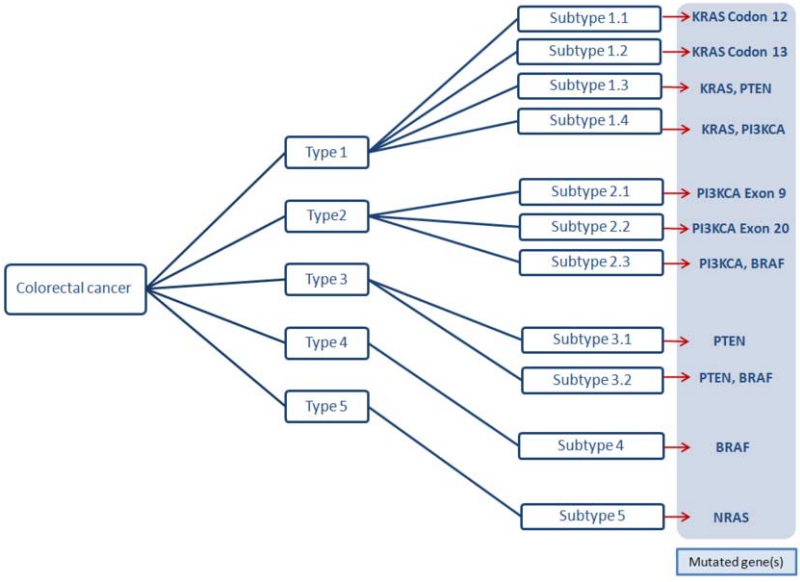

Figure 1. The CancerCommon's approach for molecular classification of CR

The second classification method of CRC was obtained from Mayo Clinic's Molecular Genetic Laboratory online source (Figure 2). We merged both classifications, effectively unifying molecular knowledge of both hereditary and non-hereditary forms of CRC [15]. The decision tree and additional visualizations of the knowledge base can be viewed at the project web site [14].

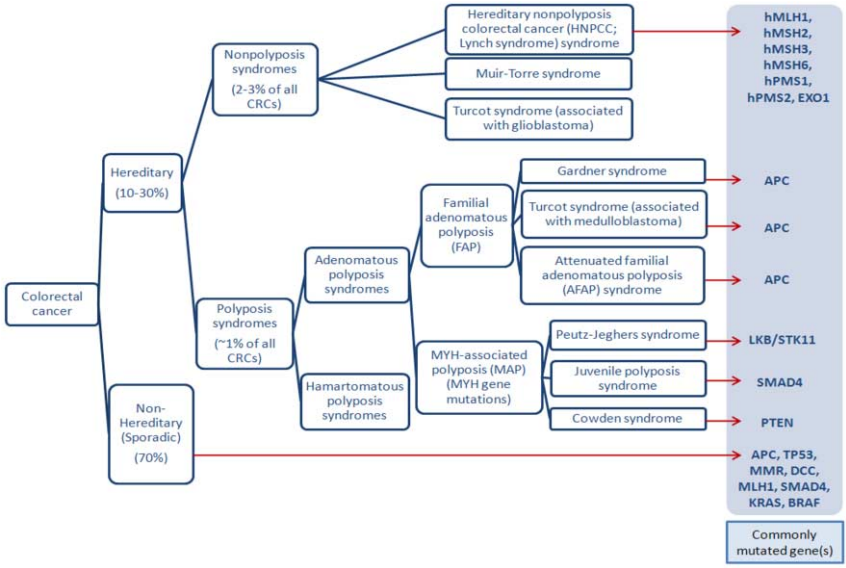

Figure 2. The Mayo Clinic's Molecular Genetic Laboratory approach for classification of CRC syndromes.

Our analysis also identified several genetic markers that might be of crucial importance in terms of CRC drug effectiveness and toxicity, but which are not backed by sufficient clinical evidence yet (examples are variations in CCND1, GSTP1, ERCC2 and KLC3).

Our analysis of cost-effectiveness of genetic testing in CRC therapy showed that most current data focused on screening strategies for carriers of hereditary forms of CRC such as HNPCC [16, 17] and FAP [18]. There was no extensive data on the cost-effectiveness of genetic tests related to CRC classification for pharmacogenomics decision making available, except for some key genes such as UGT1A1 [19] and KRAS [20]. Hence, we are working on extending our knowledge base with data about costs of available genetic tests to assist decision makers in the cost-effectiveness analyses of using less established genetic biomarkers.

4. Discussion

Herein, we described an initial framework to create an extended, automatically updated living review about personalized treatment options in colorectal cancer based on a mixture of automated extraction of publicly available data sets on the web with manual curation of the scientific literature. Our resources go beyond the previous state-of-the-art by providing a more complete view of genetic markers for colorectal cancer treatment, incorporating genetic markers for both hereditary and non-hereditary types of cancer, as well as the ability to automatically update key parts of the living review

without manual work by medical experts. For example, this enabled us to merge decision trees with automatically updated data on genetic markers relevant for specific decisions in the treatment workflow. Still, the resources presented in this paper are at an early stage of development, and more work is needed to efficiently balance manual curation and automated data harvesting and presentation.

We are currently preparing to practically apply the methodologies developed in this work in two ways: First, to collaborate with clinicians at the Vienna General Hospital in order to provide them with an evidence base for refining standard operating procedures for cancer treatment based on the latest information from scientific datasets. Second, we are developing clinical decision support systems that can make use of the structured knowledge and which can be embedded into hospital information systems.

References

[1] Vidwans SJ, Flaherty KT, Fisher DE, Tenenbaum JM, Travers MD, Shrager J. A Melanoma Molecular Disease Model. Plos One.6(3):e18257
[2] Howlader N, Noone AM, Krapcho M, Neyman N, Aminou R, Waldron W, et al. SEER Cancer Statistics Review, 1975-2008. http://seer.cancer.gov/csr/1975_2008/: National Cancer Institute.
[3] Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer. 2010 Dec 15;127(12):2893-917
[4] http://www.omim.org. [last accessed date: 20.11.2011].
[5] Guberman JM, Ai J, Arnaiz O, Baran J, Blake A, Baldock R, et al. BioMart Central Portal: an open database network for the biological community. Database (Oxford). 2011;2011:bar041.
[6] http://www.colorectal-cancer.ca. [last accessed date: 24.01.2012].
[7] Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, et al. DrugBank 3.0: a comprehensive resource for 'Omics' research on drugs. Nucleic Acids Res. 2011 Jan;39:D1035-D41
[8] http://code.google.com/p/google-refine. [last accessed date: 24.01.2012].
[9] http://www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics. [last accessed date: 24.01.2012].
[10] McDonagh EM, Whirl-Carrillo M, Garten Y, Altman RB, Klein TE. From pharmacogenomic knowledge acquisition to clinical applications: the PharmGKB as a clinical pharmacogenomic biomarker resource. Biomark Med. 2011 Dec;5(6):795-806
[11] http://www.mayomedicallaboratories.com/media/articles/communique/mc2831-0904.pdf. [last accessed date: 12.01.2012].
[12] http://crdm.cancercommons.org/cr/index.php/A_Colorectal_Cancer_Molecular_Disease_Model. [last accessed date: 12.01.2012].
[13] http://www.arupconsult.com/Algorithms/HNPCC.pdf. [last accessed date: 12.01.2012].
[14] http://www.genomic-cds.org/resources. [last accessed date: 27.01.2012].
[15] Tejpar S, Bertagnolli M, Bosman F, Lenz HJ, Garraway L, Waldman F, et al. Prognostic and Predictive Biomarkers in Resected Colon Cancer: Current Status and Future Perspectives for Integrating Genomics into Biomarker Discovery. Oncologist. 2010;15(4):390-404
[16] Reyes CM, Allen BA, Terdiman JP, Wilson LS. Comparison of selection strategies for genetic testing of patients with hereditary nonpolyposis colorectal carcinoma - Effectiveness and cost-effectiveness. Cancer. 2002 Nov 1;95(9):1848-1856
[17] Mvundura M, Grosse SD, Hampel H, Palomaki GE. The cost-effectiveness of genetic testing strategies for Lynch syndrome among newly diagnosed patients with colorectal cancer. Genet Med. 2010 Feb;12(2):93-104
[18] Lima AOD, del Castillo LS, Mochon LG, Epstein D, Tamayo CB, Portero RV. An economic assessment of genetic testing, for familial adenomatous polyposis. Rev Esp Enferm Dig. 2008 Aug;100(8):470-475
[19] Pichereau S, Le Louarn A, Lecomte T, Blasco H, Le Guellec C, Bourgoin H. Cost-Effectiveness of UGT1A1*28 Genotyping in Preventing Severe Neutropenia Following FOLFIRI Therapy in Colorectal Cancer. J Pharm Pharm Sci. 2010;13(4):615-625
[20] Shiroiwa T, Motoo Y, Tsutani K. Cost-Effectiveness Analysis of KRAS Testing and Cetuximab as Last-Line Therapy for Colorectal Cancer. Mol Diagn Ther. 2010;14(6):375-384