Data Graphs for Linking Clinical Phenotype and Molecular Feature Space

Andreas Heinzel, Emergentec Biodevelopment GmbH, Austria
Raul Fechete, Emergentec Biodevelopment GmbH, Austria
Johannes Söllner, Emergentec Biodevelopment GmbH, Austria
Paul Perco, Emergentec Biodevelopment GmbH, Austria
Georg Heinze, Medical University of Vienna, Austria
Rainer Oberbauer, Medical University of Vienna, Austria
Gert Mayer, Medical University of Innsbruck, Austria
Arno Lukas, Emergentec Biodevelopment GmbH, Austria
Bernd Mayer, Emergentec Biodevelopment GmbH, Austria

ABSTRACT

Omics profiling in translational clinical research has provided detailed molecular characterization of disease phenotypes. Integrating this molecular data space with clinical phenotype descriptors has triggered advancements regarding a systems view on disease, resulting in the concept of stratified medicine. The authors present a methodology for patient stratification by analyzing clinical and molecular information on a per-patient level represented as a data graph. This approach rests on linking patient specific clinical data and biomarker profiles with molecular functional units being derived by segmenting a human proteome interaction network. As a result patient strata are built holding sets of affected functional molecular units as common denominator. Annotation of such functional units on the level of associated diseases, biomarkers and drug targets allows reconciliation with respective clinical data for further improving the assignment of patients to specific strata. The authors finally discuss this approach in the light of adaptive clinical trials design and analysis.

Keywords: Adaptive Clinical Trial, Biomarker, Graph, Integration, Network, Omics, Stratified Medicine, Systems Biology

DOI: 10.4018/ijssbbt.2012010102
INTRODUCTION

The Omics revolution has brought forward significant advancements in research methodologies in general molecular biology as well as clinical translational research. Whereas “candidate gene” approaches dominated in the past, naturally being hypothesis driven, explorative analysis became amenable by utilizing high-throughput approaches capable of analyzing significant fractions of genomes, or substantial parts of the protein coding transcripts, proteins, and metabolites. However, similar to some disillusion regarding the immediate insights or even clinical consequences of getting hold of the first human genome sequences also Omics profiling has seen its bottlenecks. Major shortcomings include the pending completion of molecular catalogues for a human cell (with protein coding transcripts in the continuum of genomic stretches eventually transcribed as example), the fact of analyzing snapshots instead of time series rendering the description of cellular processes still significantly underdetermined with respect to quantitative modeling, and finally also conceptual and methodological limitations on the analysis side, the latter mainly hampering an integration of statistical procedures with biological concepts.

Nevertheless, translational clinical research has seen a significant boost in the course of emerging Omics, bringing forward further insights into disease pathophysiology, delineation of novel biomarker candidates and profiles, subsequently supporting tailored medication regimes. On this ground of growing data on specific diseases also the analysis side has experienced conceptual advancements. Specifically integration aspects have been in focus, even bringing forward novel scientific disciplines as Systems Biology (Hood, 2004) and Systems Medicine (Auffray, 2009). Here the assumption is that a broad scale integration of clinical and molecular data space results in detailed models of diseases, in turn providing biomarkers and therapy targets being causally linked to a disease. Such biomarkers, frequently being multimarker profiles, may then be used for supporting patient characterization beyond the classical clinical data space, which in turn is considered as supportive for evidence-based decisions on patient specific therapy in case multiple regimes are available (Friend, 2011).

In the advent of these developments the term “personalized medicine” and “stratified medicine” became en vogue (Trusheim, 2007), in one part based on emerging insights regarding the molecular heterogeneity of diseases which on the clinical level are considered as homogeneous phenotype, and secondly also driven by the increasing therapy options for a number of major diseases. Here the assumption is that integrating clinical and molecular (biomarker) profiles serve for identification of specific patient cohorts (strata) for which therapy can be specifically optimized. Such approach may be in particular of relevance for complex disease etiologies as chronic kidney disease in the context of diabetes and hypertension (Fechete, 2011).

In this work we present a concept on the level of a data graph for integrating the clinical phenotype data space and a biomarker profile-centered molecular data space encoded as molecular functional units, resulting in the delineation of patient strata reconciling both, patient specific clinical as well as molecular characteristics.

Omics in Clinical Translational Research

Technological maturity of central Omics procedures (in particular genome sequencing, transcriptomics, proteomics and metabolomics), as well as development of standard operating procedures for sample retrieval and analysis procedures has in the meantime brought Omics close to the patient. However, complex, multimarker profiles for stratification, as any other IVD (in vitro diagnostic) device, have to meet requirements including i) being tied to a specific phenotype and interventions, ii) showing required test accuracy and iii) providing sufficient diagnostic accuracy (Vitzthum, 2005). As an example, the MammaPrint microarray-based test system (Slodkowska, 2009) (a signature of
Figure Based Biomedical Document Retrieval System using Structural Image Features
[www.igi-global.com/article/figure-based-biomedical-document-retrieval/74694?camid=4v1a](www.igi-global.com/article/figure-based-biomedical-document-retrieval/74694?camid=4v1a)

Predictive Toxicity of Conventional Triazole Pesticides by Simulating Inhibitory Effect on Human Aromatase CYP19 Enzyme