51. Jahrestagung der
Deutschen Gesellschaft für
Medizinische Informatik,
Biometrie und Epidemiologie

gmds · LEIPZIG 2006

10. – 14. September 2006

# IS-04

## „Clinical Relevance of Genetic Evolutionary Pathways in HIV and Cancer“

Koordination: Dr. J. Rahnenführer, Max-Planck-Institut für Informatik, Saarbrücken

Beschreibung der wissenschaftlichen Problemstellung

The reconstruction of genetic evolutionary pathways is important for understanding disease progression. Recently, the new probabilistic model class of mutagenetic trees mixture models to describe evolutionary processes was introduced [1]. In these models, progression is characterized by the ordered accumulation of permanent genetic changes. The strength of this approach is the implementation of a general concept with explanatory power for diverse applications. The model has been successfully applied for estimating the development of viral drug resistance in HIV patients [2] and for modelling the accumulation of chromosomal aberrations in tumor cells from cancer patients [3].

Successful treatment of HIV infection is complicated by the high genetic variability in combination with high replication rates of the virus. This property enables the virus to escape from selective pressure of the human immune system and of the drug therapy by acquiring beneficial mutations in the course of the disease [Talk 1].

Often, such genetic data are only available in the form of cross-sectional measurements. In this case, every patient is characterized by a subset of genetic events that describe the disease state at one specific point of time. From a statistical viewpoint, the task then is to approximate the multivariate joint distribution of the single events across all patients. From a biological viewpoint, an important goal is to reconstruct the preferred order in which the events occur. A statistical framework that allows the successful handling of these tasks is given by mixture models of directed trees [1]. Based on these models, the genetic barrier can be defined, a quantity that summarizes the difficulty for the virus to escape from drug pressure by acquiring mutations. The genetic barrier has been proven to be a useful tool in designing effective treatment strategies [2]. [Talk 2]

In some hematological and exception solid tumors specific chromosomal alterations are diagnostic. More commonly, cancer cells are aneuploid and to various extents heterogeneous. Common alterations are therefore obscured by a plethora of bystander events and case-specific aberrations. Moreover, most tumor types do not follow a single linear track with successive specific genetic alterations corresponding to distinct stages of clinical progression, but develop along parallel, divergent or convergent pathways. In addition, defined biological (and clinical) properties of a tumor type can be brought about by different genetic or epigenetic alterations that result in different chromosomal alterations or may not even be discernible at the chromosome level. Therefore, as a rule, deriving prognostic information (and in some cases even differential diagnostic information) from chromosomal alterations requires sophisticated statistical models adapted to each tumor type. [Talk 3].

Prediction of time until death or time until relapse after surgery is important for the treatment of cancer patients. Traditional biostatistics research considers clinical and histological measurements like tumor stage, tumor volume, or lymph node status, as prognostic factors. For relapse of prostate cancer patients, the Gleason score reflecting the histological pattern of tumor growth is a common grading system with predictive value. The identification of genetic markers that better reflect tumor biology is eminent. Cytogenetic alterations in the tumor cells derived from CGH (comparative genomic hybridization) can be the starting point for an improved genetic disease characterization. [Talk 4]

The genetic progression score (GPS) has been introduced as a new genetic marker that estimates the progression of a tumor based on trees mixture models [3]. The GPS is derived from the tree models by replacing conditional probabilities with expected waiting times. The GPS directly estimates the age

of tumors up to a scaling. Using Cox regression models it could be demonstrated that the GPS is a medically relevant prognostic factor that can be used to discriminate between patient subgroups with different expected clinical outcome. For prostate cancer patients, restricting to patients with the average Gleason score 7 only, the GPS can be used to further identify subgroups with different prognosis. This shows that the GPS can improve diagnostics even after adjusting for traditional markers. [Talk 5]

References:

[1] Niko Beerenwinkel, Jörg Rahnenführer, Martin Däumer, Daniel Hoffmann, Rolf Kaiser, Joachim Selbig, Thomas Lengauer: Learning multiple evolutionary pathways from cross-sectional data, Journal of Computational Biology, 12(6): 584-598, 2005.
[2] Niko Beerenwinkel, Martin Däumer, Tobias Sing, Jörg Rahnenführer, Thomas Lengauer, Joachim Selbig, Daniel Hoffmann, Rolf Kaiser: Estimating HIV evolutionary pathways and the genetic barrier to drug resistance, The Journal of Infectious Diseases 191(11): 1953-1960, 2005.
[3] Jörg Rahnenführer, Niko Beerenwinkel, Wolfgang A. Schulz, Christian Hartmann, Andreas von Deimling, Bernd Wullich, Thomas Lengauer: Estimating cancer survival and clinical outcome based on genetic tumor progression scores, Bioinformatics 21(10): 2438-2446, 2005.

Anmerkung zur Interdisziplinarität und zur Zielgruppe

The scientific findings of this research project are the result of an interdisciplinary effort involving expertise from virologists, bioinformaticians, computer scientists, molecular biologists, clinicians, and statisticians. Highly effective cooperative alliances between bioinformaticians on the one side and virologists and clinicians on the other side were established.

The session describes methods for quantification of disease progression and optimization of therapy selection applied to the two diseases AIDS and cancer. From a genetic point of view these two disease entities and their corresponding underlying biological processes are entirely different. However, the basis for modeling genetic disease progression is the same in the two considered scenarios. In both cases, typically the underlying data are only available in the form of cross-sectional data. For every patient, only measurements at one specific point in time are available. Based on the assumption of genetic changes in the course of the disease being permanent, the tree mixture models can overcome the problem of estimating the temporal order from measurements at single time points. The strength of the algorithmic approach is the implementation of a general concept with explanatory power for diverse applications.

In this framework, the collaboration between researchers from life science and data analysis is not a one-way street. The multidisciplinarity not only refers to the application of statistical methods to genetic and clinical measurements, but predictions obtained from statistical models serve as an important decision-making aid for doctors. In the HIV context, a server that is running for five years now at the Max-Planck-Institute for Informatics has been used by virologists and medical doctors all over the world for making over 35000 predictions concerning therapy options. A suitable therapy is chosen based on the genetic pattern of the virus of the respective patient. The server can be accessed at the website www.geno2pheno.org. The primary underlying algorithm was created in a collaborative effort with virologists at the University of Cologne and the German National Reference Center for Retroviruses in Erlangen. The corresponding Arevir database was created with the support of the German Research Foundation (DFG). The tree mixture models for estimating disease progression go one step further and allow for a specific patient a glance in the future of his individual disease development. The therapy selection thus now includes the prediction of the next most likely steps in disease progression.

Besides other sources, at the Max-Planck-Institute for Informatics, both projects related to HIV and to cancer are funded in the context of the BioSapiens Network of Excellence (EU contract no. LSHG-CT-2003-503265).

Vorträge

IS-04-1: HIV drug resistance mutations (~15 min)
Dr. Martin Däumer, Institut für Virologie, Universität zu Köln

IS-04-2: Estimating HIV evolutionary pathways and the genetic barrier to drug resistance (~25 min) Dr. Niko Beerenwinkel, Department of Mathematics, UC Berkeley, USA

IS-04-3: Molecular pathways in human cancers (~ 20 min)
Prof. Wolfgang Schulz, Urologische Klinik, Universität Düsseldorf

IS-04-4: Cytogenetic alterations in prostate carcinomas (~ 20 min)
Prof. Bernd Wullich, Klinik für Urologie und Kinderurologie, Universität des Saarlandes

IS-04-5: Clinical relevance of genetic tumor progression scores (~10 min)
Dr. Jörg Rahnenführer, Max-Planck-Institut für Informatik, Saarbrücken