

bioinformatics

Tumour espionage

On the road to individualised cancer therapy

Dr. Hubert Hackl and Prof. Dr. Zlatko Trajanoski

Section for Bioinformatics, Innsbruck Medical University, Biocenter

Cancer is a multifactorial disease with high prevalence in Western Europe. Although considerable characterisation of possible environmental influences and mechanisms are known [1], the identification of markers for the cancer process, which allow a good prognosis or are reliable for clinical usage, is not yet sufficient.

New findings in the molecular pathogenesis of cancer have fundamentally changed the discovery and development of new tumour therapeutics. The hope is that in the future these therapies can be better tailored to individual patients [2-3].

The aim of Oncotyrol is to translate the findings of cancer research in the areas of cell biology, genetics and inflammation research to clinical daily routine and to develop innovative, individualised and cost-effective approaches for the prevention, diagnosis and therapy of cancer, whereby a focus is set on breast cancer, chronic leukaemias (CML, ALL) and prostate cancer (www.oncotyrol.at).

The compilation of the entire, comprehensive genetic information (and its epigenetic modulation) contributes considerably to the research of complex diseases, in particular cancer and its molecular mechanisms. The newest developments of high-throughput methods, that is to say new sequencing technologies, have made it possible to determine the entire genomic information (genome) of healthy or tumour cells of a person with affordable effort. Where as for the decoding of the first human genome 100 million dollars were spent, currently the sequencing of a human genome at 30-times coverage costs 5000 dollars and is rapidly dropping. The analysis can be completed in a few days and it

is possible to detect genetic variations, such as translocations or point mutations in the genome, which could be the triggers for the development of cancer [4]. In the future it should be possible to sequence the genome of a patient – comparable with the current use of imaging processes such as computer tomography. Although obviously technological and cost-effective improvements are required and the sequencing of hundreds of genomes is necessary for genetic studies of complex diseases, it is the beginning of a new era: the personalised cancer medicine.

In view of the facts that an huge amount of data accumulates (1 sequence run delivers 160 million 100 base long sequences which have a memory requirement of 1TB) is not surprising that Bioinformatics plays an increasing role.

The important roles of bioinformatics in this context are the building up of an appropriate infrastructure, the combined data management from high-throughput data and clinical data, integrative data analysis, network analysis and mathematis-

ical modelling, to be able to predict the biomarkers for diagnosis and therapy.

The key to personalised medicine: management and integration of clinical and genome-wide data.

Today we have the unprecedented opportunities to test cancer-specific molecular processes genome-wide: activity of genes and proteins, protein-protein interactions, genomic and genetic variations, epigenetic modifications (e.g. methylation patterns), metabolic products, cell characterisation, regulatory RNA. Moreover, there is corresponding clinical, patient-related information, the use of which is often very difficult from a technical, ethnic or legal point of view. The volume and diversity of both the experimental and clinical data require the systematic and well organised management and that the scientist has computer access for the analysis and interpretation. The first steps are usually data collection and entry; and this must take place in the laboratory. Here one relies on a laboratory information management system (LIMS). The subsequent data pre-processing especially with sequencing data is a major challenge, not

only in the storage of the data but also in processing (algorithms for combination and mapping of sequences as well as the detection of variations), this is done with the newest methods of high performance computing, such as for example cloud computing, where appropriate external computer clusters can be used. For data integration one works with information technical solutions, such as for example in the design of a warehouse where data from individual data sets (that is in general processed data and not raw data) and data banks are consolidated, in order to be able to perform systematic comprehensive data retrieval. This also allows access to data for the implementation of analytical methods and modelling, which only then makes possible the identification of diagnostic markers, therapeutic aims and molecular mechanisms.

The central question is, how can one complement the available data from a large population with personalised medicine? The answer lies precisely in the well-illustrated design and development of the massive data storage, data mining and effective web-based processes [3], whereby correspondingly clinical information and individual response and data have to be returned to the system (see Fig. 1).

According to our experience, hypothesis formation in science is driven today by a large volume of data, and therefore data management should be an integral part of research activity. Retrospective data management requires not only a lot of effort, but is often not possible. There must be commitment to data management over an extended period as the efforts for infrastructure and personal are not insignificant. Iterative cycles between computer analyses and experiments can not only improve the methods and deliver the corresponding data, but also enable scientific questions to be answered [5].

Biomolecular networks as an example of integrative data analysis

Biomolecular networks facilitate an integration of medical information. Thereby, a greater understanding of the contexts of diseases is possible, based on genome-wide data. For example, it is possible to identify the factors which have a large influence on the cancer development, evi-

dent in the network where they are linked to many other factors. A number of approaches to network modelling have already been exposed as very promising for cancer [6–8].

The first step in the modelling of networks is normally the building of a gene-network, whereby two nodes (genes) are linked, should their activity throughout a row of different tumour samples/patients be (significantly) similar. In addition to gene expression a number of other resources and experimental data can be drawn upon, to be integrated into the network in order to gain new perspectives which would otherwise remain hidden in the complex data set. In particular cases, protein-protein interactions (associations) are suitable as complemen-

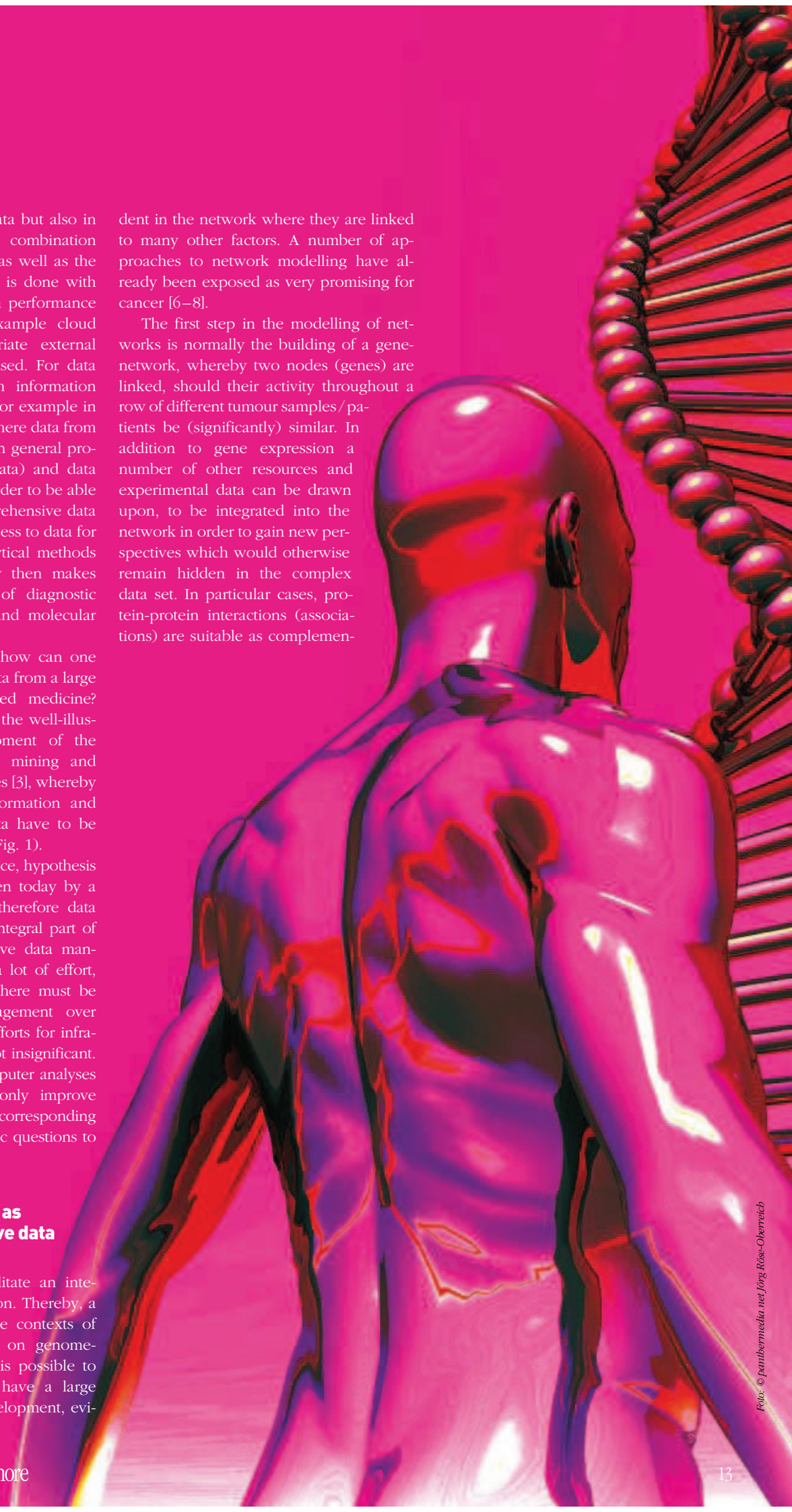


Foto: © panthermedia.net/Jörg Röss-Oberreich

bioinformatics

tary data sources, as is information about point mutations in the genome or regulatory interactions.

Clinical patient data such as prognoses can also be integrated into the network: tumor samples (patients), based on the activity of a factor (e.g. gene) can be divided into 2 groups, in order to determine whether there is a significant difference between the groups within a time period without resurgence of the disease.

Another possibility of integration is the organisation and visualisation of quantitative data (e.g. expression data from genes with a similar profile) together with data from already known interactions which have been confirmed in the laboratory. This could be a combination of enzymatic metabolic reactions (pathways) or signal transduction paths.

Not only is it almost impossible to predict the causal interrelationships, there is a further problem: missing or incorrect data. For this reason it is of utmost importance, to experimentally test at least some of those interrelationships. Nevertheless, pathway and network analyses enjoy great popularity, as proteins are “social” and do not operate individually in the cellular context, in complex diseases whole pathways and not only individual genes are deregulated and in many cases a robust prediction of the interactions is possible, with a high probability of being confirmed in experimental testing.

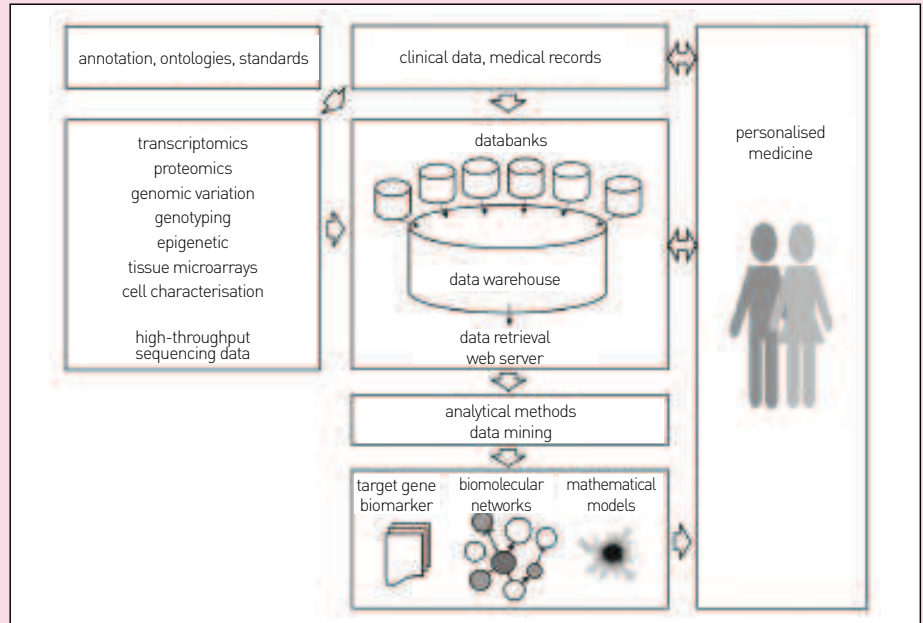


Fig. 1 Information technological solution and course of the data flow for personalised medicine
 Diagram partly adapted from: Hackl H, Stocker G, Charoentong P, Mlecnik B, Bindea G, Galon J, Trajanoski Z. Information technology solutions for integration of biomolecular and clinical data in the identification of new cancer biomarkers and targets for therapy. *Pharmacol Ther.* 2010. (128:488-498) and created using CorelDrawX4

Can immune cells provide information about the tumour and offer a prognosis?

We have recently integrated biomolecular data with clinical data for colorectal cancer, in order to identify new prognostic biomarkers. For this purpose a data bank was implemented, comprised of pre-processed and normalised data (clinical data, gene expression (qPCR), FACS-data for cell characterisation and data from tissue microarray analysis) of a cohort of more than 1700 patients [9]. The high dimensionality and

complexity of the data meant that the data analysis was a considerable challenge. Using suitable statistical (survival analysis) and bioinformatic methods (network analysis) new hypotheses were formulated, e.g. including the influence of the environment of the tumour (tumour microenvironment) on the tumour. And indeed it was shown that the concentration and localisation of immune cells which can be found surrounding the tumour and characterised using specific surface markers, permit a considerably better prognosis for colorectal

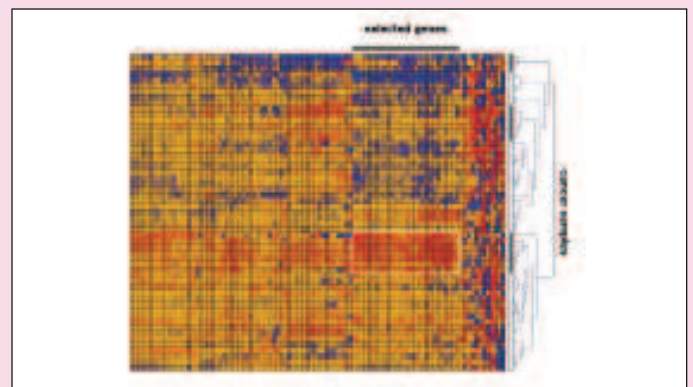
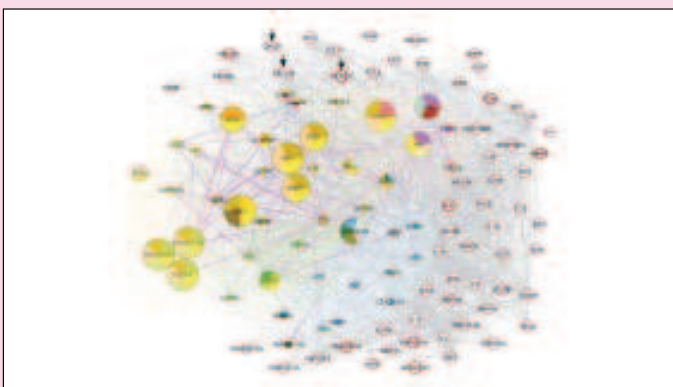


Fig. 2. Example of a biomolecular network and a heatmap for gene expression profiles of cancer samples

Adapted from Mlecnik B, Tosolini M, Charoentong P, Kirilovskiy A, Bindea G, Berger A, Camus M, Gillard M, Bruneval P, Fridman WH, Pages F, Trajanoski Z, Galon J. Biomolecular network reconstruction identifies T cell homing factors associated with survival in colorectal cancer. *Gastroenterology.* 2010. 138:1429-1440

Data from Verbaak RG, Wouters BJ, Erpelinck CA, Abbas S, Beverloo HB, Lugthart S, Löwenberg B, Delwel R, Valk PJ. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica.* 2009. 94:131-4 Gene Expression Omnibus GEO (GSE 6891) clustered (hierarchical clustering) and visualised using Genesis (Sturn A, Quackenbush J, Trajanoski Z. Genesis: cluster analysis of microarray data. *Bioinformatics.* 18: 207-208 (2002))



Zlatko Trajanoski, born 1964 in Skopje, completed academic studies and a doctorate in Biomedical Engineering at the Graz University of Technology. From 1995 he worked as a university assistant at the TU Graz and as a postdoc at the Department of Internal Medicine, Yale University, New Haven, CT/USA (1997–1998). In 1999 he qualified as a professor in Biomedical Engineering and founded a working group for Bioinformatics at the TU Graz. During 2000–2001 he was a visiting scientist at the National Institutes of Health (NIH), Bethesda, MD/USA. In 2003 he became University Professor for Bioinformatics at the TU Graz and Director of the Institute for Genomics and Bioinformatics. In 2010 he was appointed as University Professor for Bioinformatics at the Biocenter of the Innsbruck Medical University and heads the Section for Bioinformatics. He is the Coordinator of the Bioinformatics Integration Network of the Austrian Genome-Program (GEN-AU) and has assumed the management of the area “Bioinformatics and System Biology” at Oncotryol.

Hubert Hackl, born 1969 in Kirchdorf a.d. Krems, studied Electrical and Biomedical Engineering at the Graz University of Technology and received his doctorate there in 2004 in the specialist area of Bioinformatics. He was a visiting scientist at the research facility The Institute for Genomic Research (TIGR) (now the J. Craig Venter Institute), Rockville, MD, USA. Subsequently, he was employed as a scientific assistant and since 2007 as a university assistant at the Institute for Genomics and Bioinformatics at the TU Graz. Since 2010 he is employed as a university assistant in the newly formed Section for Bioinformatics at the Biocenter of the Innsbruck Medical University. His work is concerned with integrative data analyses, transcriptional regulation and computational biology.

cancer than the previously used classification method, based on tumour histopathology [10].

The next challenge will be the improved characterisation of the biomolecular interactions between tumour and immune cells. The ultimate question from the Bioinformatics point of view is however, whether the tumour development (e.g. the size of the tumour) can be predicted using a mathematical model, based on information about immune cells in the area of the tumour. It is a multi-scale problem; this

means that the model must embrace many levels (molecular interactions in the cells and between the cells, quantity, localisation of varying cell types and tissues within a single and between multiple organs). Therefore there are not only spatial but also time components to be considered [11] (see Fig. 3).

→ zlatko.trajanoski@i-med.ac.at
 → www.icbi.at

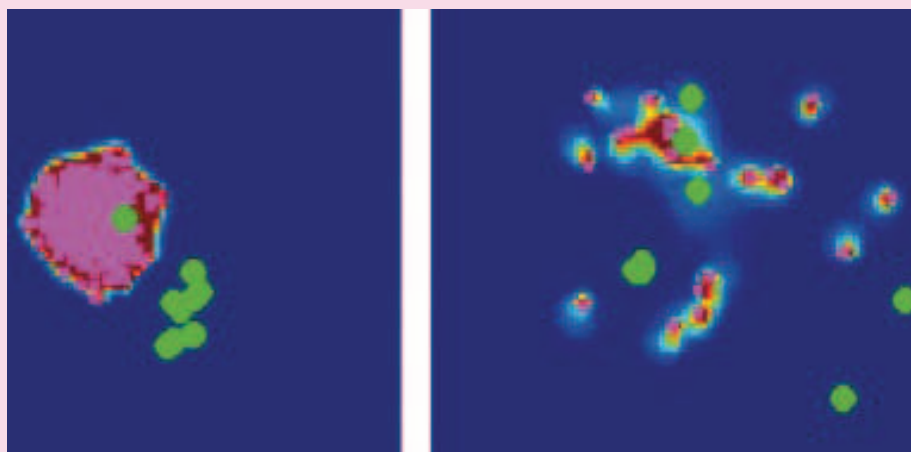


Fig. 3 Simulation of tumour-immune cell-interactions, with the help of a 2-dimensional tumour growth model (tumour is magenta and immune cells (CD4 and CD8+ T-cells) are green)
 according to Anderson AR, Quaranta V: Integrative mathematical oncology. Nat Rev Cancer. 2008. 8:227-234. (not published)

Literature

- [1] Hanahan D, Weinberg RA: The hallmarks of cancer. *Cell* (2000) **100**:57–70.
- [2] Ocana A, Pandiella A: Personalized therapies in the cancer „omics“ era. *Mol Cancer* (2010) **9**:202.
- [3] Deisboeck TS: Personalizing medicine: a systems biology perspective. *Mol Syst Biol* (2009) **5**:249.
- [4] Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, Carnevali P, Nazarenko I, Nisen GB, Yeung G, Dahl F et al: Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* (2010) **327**:78–81.
- [5] Hackl H, Stocker G, Pornpimol C, Mlecnik B, Bindea G, Galon J, Trajanoski Z: Information technology solutions for integration of biomolecular and clinical data in the identification of new cancer biomarkers and targets for therapy. *Pharmacol Ther* (2010) in press.
- [6] Pujana MA, Han JD, Starita LM, Stevens KN, Tewari M, Abn JS, Rennett G, Moreno V, Kirchhoff T, Gold B, Assmann V et al: Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nat Genet* (2007) **39**:1338–1349.
- [7] Kreger PK, Lauffenburger DA: Cancer systems biology: a network modeling perspective. *Carcinogenesis* (2010) **31**:2–8.
- [8] Mlecnik B, Tosolini M, Charoentong P, Kirilovsky A, Bindea G, Berger A, Camus M, Gillard M, Bruneval P, Fridman WH, Pages F et al: Biomolecular network reconstruction identifies T-cell homing factors associated with survival in colorectal cancer. *Gastroenterology* (2010) **138**:1429–1440.
- [9] Mlecnik B, Sanchez-Cabo F, Charoentong P, Bindea G, Pages F, Berger A, Galon J, Trajanoski Z: Data integration and exploration for the identification of molecular mechanisms in tumor-immune cells interaction. *BMC Genomics* (2010) **11 Suppl 1**:S7.
- [10] Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pages C, Tosolini M, Camus M, Berger A, Wind P, Zinzindoboue F et al: Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* (2006) **313**:1960–1964.
- [11] Anderson AR, Quaranta V: Integrative mathematical oncology. *Nat Rev Cancer* (2008) **8**:227–234.