REVIEW ARTICLE

The "Virtual Patient" system: modeling cancer using deep sequencing technologies for personalized cancer treatment

Alexander Kühn · Hans Lehrach

Abstract Cancer is a complex disease involving many different genomic and epigenomic changes, mutations, copy number changes, loss of heterozygosity etc. These differences cause tumors with the same pathological classification to respond very differently to the drugs, making therapy decisions difficult. As a result of intensive research in the field of oncology much information is available on molecular interactions and pathways involved in cancer onset and progression. In addition recent increases in the capacity of next-generation sequencing systems will provide huge amounts of genome, epigenome and transcriptome data, making it feasible to apply deep sequencing in the clinic to characterize tumor/patient samples. Both, the complexity of disturbances in interaction networks of biological processes in cancer and the new molecular information generated by this sequence analysis urgently require the development of systems that are able to derive clinically relevant predictions from all available data. The "Virtual Patient" modeling system combines general information available about cancer relevant pathways with the individual (genome, transcriptome) information available on the individual tumor/patient to generate models able to predict the effects and side effects of individual drugs or drug combinations. This opens the way to experiment with the response of the model of the individual patient to different therapy options in the computer, offering new routes to improve oncological practice, reduce health costs but also to accelerate the development and the approval process for new drugs in this area.

Keywords Next-generation sequencing · Systems biology · Cancer · Virtual patient · Personalized medicine

A. Kühn · H. Lehrach (✉)
Max-Planck Institute for Molecular Genetics,
Ihnestraße 63-73, 14195 Berlin, Germany
e-mail: lehrach@molgen.mpg.de

1 Introduction

With ∼12 million new cancer cases per year worldwide cancer is expected to overtake heart disease as the leading cause of death worldwide in 2010 (Boyle and Levin 2008). The basic general mechanisms of cancer are already well known. Research both in clinic and in the lab has clearly shown that cancer is a disease that develops over time as a result of an accumulation of many factors that promote tumor growth and metastasis (Hanahan and Weinberg 2000). Cure rates for most common forms of cancer have however hardly changed over the last decades. Experience in sequencing tumor genomes both by us and by others have shown that tumors typically have thousands of somatic changes and that many more gene mutations drive the development of cancer than previously thought (Greenman et al. 2007). These many alterations make every tumor different,

causing every patient to respond differently to specific therapies. Thus even tumors from the same clinical classification may require very different treatments.

Up to now more than 350 genes have been identified, whose mutations or other genetic alterations are causally implicated in cancer development and hence are known as 'cancer genes' (Greenman et al. 2007). These genes offer starting points for the development of innovative drugs and new therapies. However, the actual molecular blueprint of each individual cancer is poorly known, mostly because until now, only a few known mutations could be monitored in tumors, without any knowledge of the transcriptional landscape, or of the global genome alterations including numerous somatic mutations, aneuploidy status, genome instability and epigenetic changes. For this reason even the most advanced targeted therapies developed so far are typically only effective for a small fraction of patients primarily due to molecular differences between different tumors. It has often been shown that the effectiveness of specific targeted therapies depends on the presence or absence of specific changes. It has for example been shown, that the overexpression or the presence of an activating mutation in the epidermal growth factor receptor (EGFR) is a strong predictor of a response of the tumor to the drug Gefitinib (Mok et al. 2009). Similarly the hormone receptors and HER2 status in breast cancer (Taneja et al. 2010) or the presence or absence of KRAS wild type or BRAF wild type have been used for individualized therapy recommendations (Di Nicolantonio et al. 2008; Karapetis et al. 2008; Van Cutsem et al. 2010; Wilson et al. 2010). However, since hundreds of mutated genes or genetic alterations are implicated in cancer development, in many other indications e.g. ovarian cancer (Hays et al. 2010) or the renal cell carcinoma (Cheng et al. 2010; Shah and Margulis 2010) little progress has been made. Even targeted diagnostics like the KRAS assay are only providing a relatively small improvement, e.g. the KRAS diagnostic improves the response rate of the corresponding EGFR inhibitor/Irinotecan therapy in second-line colon cancer from 10 % for the KRAS mutant group to a 35 % for the KRAS wild type group (Amgen 2009), due to other mutations or other genetic alterations beside KRAS and BRAF, such as a PIK3CA mutation or a loss of PTEN expression (Jhawer et al. 2008), that counteract the drug response. This example underscores the urgent need for a more accurate, reproducible and systematic approach for new diagnostic biomarkers that are more informative

in predicting the response of the tumor to specific therapies.

Since it is expected by 2030 that there could be 27 million incident cases of cancer, 17 million cancer deaths annually and 75 million persons alive with cancer within five years of diagnosis (Boyle and Levin 2008) and the development costs of new cancer drugs have dramatically increased, while the number of new drug approvals in this area keep dropping, both the public health system as well as the pharmaceutical industry are facing big challenges. Therefore a system is needed that on the one hand helps to select the right drug (or drug combination) for each patient, but also identifies the 'right' patient (population) for each drug, reducing the size of clinical studies required to detect a positive response, and decreasing development costs of new cancer drugs.

Two significant developments allow us to take a more integrated approach for selecting appropriate cancer therapies for patients by predicting effects (and side effects) of different drugs/drug combinations on individual patients:

- As a result of decades of cancer research many signaling pathways considered relevant to cancer have been characterized, as illustrated in the "Pathways of Human Cancer" poster showing the molecular interactions of hundreds of proteins (Weinberg 2007). Databases on cellular interaction networks with functional information of cellular systems have been established, such as Reactome (Joshi-Tope et al. 2005) and KEGG (Kanehisa et al. 2004). Moreover the molecular effects of somatic mutations observed in many types of cancer have been characterized and different comprehensive databases have been set up, e.g. the Catalogue Of Somatic Mutations In Cancer (COSMIC), which integrates information on cancer associated mutations (Forbes et al. 2006).

- Since the completion of the first human genome sequence (Lander et al. 2001; IHGSC 2004) new sequencing technologies have been developed. These next-generation sequencing (NGS) technologies provide an increase in sequencing throughput and a decrease in per base pair sequencing costs at the same time (Brenner et al. 2000; Schuster 2008), which now allow the application of genome scale analysis to the tumor and somatic tissue of each cancer patient. Extrapolating current trends in sequencing costs, full genome sequencing would be within the reach of the mass markets within the next few years.

The combination of global information on cancer relevant pathways and detailed characterization of patient's tumor material by deep sequencing now allows, for the first time, the development of predictive models of the onset and the progression of cancer as well as the drug response in individual cancer patients. Systems biology offers computational tools to analyze, integrate and interpret biological data, and provides mathematical and computational concepts for the development, simulation and interpretation of cellular interaction networks (Wierling et al. 2007; Bois 2009; Chen et al. 2009; Klipp et al. 2009). Most of these models (as collected in the BioModels database) address individual pathways (metabolic processes, individual signal transduction pathways, cell cycle regulation). Tumors have however changes in many components of the relevant biological networks. Similarly, most drugs used in oncology interact in a complex pattern with many different proteins, binding to different (main or side) targets with different dissociation constants. Realistic models of therapy effects therefore have to cover more complex situations than usually addressed in systems biology.

Although models on tumor cellular evolution have been developed, such as a stochastic mathematical model of the evolution of tumor metastases in an expanding cancer cell population (Haeno and Michor 2010) or a model that describes the accumulation of driver (and passenger) mutations during tumor development (Bozic et al. 2010), there is no approach in cancer modeling available that can model very large pathway systems with consideration of patient specific environment. But since many of the cancer drugs available are only effective in an often rather small fraction of the patient population, with the majority of patients showing little or no benefits or even suffering often quite severe side effects shows that the progress in the treatment of tumors in individual patients will depend critically on being able to predict the effect of such treatments in the context of the genome/transcriptome involved. We therefore developed a system that is capable of handling the complexity of biological processes like cancer even in the absence of accurate knowledge on kinetic parameters (which are typically unknown) and has the unique opportunity to be fuelled with massive patient specific molecular data arise from NGS analysis and thereby is able to predict effects and side effects of drugs and drug combinations in individual tumor patients: the "Virtual Patient" system.

2 Detailed molecular characterization of tumors using NGS technologies

The molecular characterization of tumor has developed from the systematic sequencing of segments coding for candidate genes or gene families by Sanger sequencing to whole genome sequencing based on NGS techniques (Wood et al. 2007; Parmigiani et al. 2009; Pleasance et al. 2010a, b). These techniques also offer a new way in which cancer patient samples can be analyzed. RNAseq provides comprehensive RNA expression profiles with qualitative and quantitative information relative to transcript expression levels including that of rare transcripts, small RNAs and novel transcribed units, alternative splicing, allele specific expression patterns and RNA editing (Sultan et al. 2008; Richard et al. 2010). Due to improvements of RNAseq protocols the routine analysis of the strand specificity (Parkhomchuk et al. 2009), as well as paired end read information can be generated.

Therefore RNAseq studies on cancer cells and tumors are now increasingly performed (Berger et al. 2010; Stark et al. 2010) showing that both genome and transcriptome information are essential to establish an accurate correlation with the phenotype of the tumor. For example, in the case of a tumor suppressor gene mutated on one chromosome, whereas the second (non-mutated) copy silenced by epigenetic processes, the loss of functioning allele will be only detected by analyzing allele specific expression patterns.

In addition to the molecular analysis of the bulk tumor, the molecular characterization of cancer stem cells (CSC) can be of high value, despite the continuing debate over their definition, how common they are, and whether they even exist. CSC are proposed to originate either from malignant transformation of a normal or a progenitor cell (Reya et al. 2001), to have self-renewal capacities and, since they proliferate throughout life, to be more susceptible to accumulation of oncogenic mutations than differentiated cells with their comparatively short life span (Monzani et al. 2007) and thereby more resistant than the rest of the tumor cells. It is believed that a cancer recurrence occurs because even one single cell that evades the surgeon's blade or a chemotherapy or radiation by its hardiness is enough to recapitulate the whole tumorigenesis. It is therefore likely that CSC harbor quite different molecular characteristics than bulk tumor, and require to be treated differently in order to be killed. We have already applied this deep molecular characterization technology on a

specific patient, a metastatic melanoma patient. For this patient, both the genomes of the bulk tumor and blood were extensively characterized by deep sequencing with a $40\times$ and $30\times$ coverage respectively ($\sim$120 gigabases tumor, $\sim$90 gigabases blood). In addition, a detailed analysis of the transcriptome of the tumor, tumor derived cell lines, and cancer stem cells (selected by growth in ES cell media, followed by magnetic bead selection using antibody against the cell surface antigen CD133, often associated with cancer stem cells) from the same patient, as well as melanocyte cell line as control, were carried out ($\sim$300 million reads in total). This combined RNAseq and genome analysis supplies molecular information about one specific tumor in unprecedented quantity and provides an ultra-deep analysis of a melanoma molecular landscape: More than 1,000 somatic mutations were detected in this patient, of which more than 100 affected expressed coding regions, more than 1,000 genes were induced and over a dozen of autocrine loops were identified that are potential drivers of the tumor (manuscript in preparation).

3 Modeling cancer using the "Virtual Patient" system

The "Virtual Patient" system is based on the modeling and simulating system PyBioS (Wierling et al. 2007) and uses relatively conventional mathematical pathway modeling approaches. It is written in an object oriented programming language (Python) and is designed so that the objects correspond to the components of the pathways relevant for cancer (genes, proteins, DNA and protein modification states, complexes, metabolites etc). Each object is assigned to keep track of its state (e.g. concentration) as well as its functions (e.g. to phosphorylate another protein object at a specific amino acid, to form a complex with a specific dissociation constant). The system is initialized with the starting concentrations of all components, as well as values of the kinetic constants in these differential equations, allowing the system to be solved numerically. The resulting computations allow analysis of both the kinetic changes, as well as the concentrations of the relevant products in the steady state after a perturbation such as a mutation or an autocrine loop. Since many kinetic parameters are unknown, or have at best been determined under laboratory conditions far from the situation in the tumor environment, a Monte-Carlo-based simulation approach is used. In this strategy the kinetic

parameters are sampled from appropriate probability distributions and are used for multiple simulations in parallel. Simulation results from different forms of the model (e.g. a model that resembles a certain mutation pattern) can be compared with an unperturbed control and used for the prediction of the effect of the perturbation (Wierling et al. 2011). The backbone of the "Virtual Patient" model contains several cancer relevant signaling pathways as proposed by Hanahan and Weinberg (Hanahan and Weinberg 2000; Weinberg 2007) plus additional signaling pathways available from pathway databases, like BioCyc (Karp et al. 2005), KEGG (Kanehisa et al. 2004), Reactome (Joshi-Tope et al. 2005; Vastric et al. 2007) and ConsensusPathDB (Kamburov et al. 2009). By rigorous literature screening using different resources, like PubMed, GeneCards, iHOP and Bibliosphere, molecular interaction details were added when required. Overall, in the developed modeling system up to date about 30 different signaling pathways are implemented, amongst others cytokine signaling (e.g. CSF, IFNA, IL8), death receptor signaling (e.g. Fas, TNFa, Trail), DNA repair/cell cycle, ephrin signaling, GCPR/hormone signaling (e.g. insulin, progesterone, testosterone), hedgehog signaling, notch signaling, several RTK signaling (e.g. bNGF, EGF, FGF, IGF, PDGF, VEGF), TGFb signaling (e.g. BMP, TGFb) and Wnt signaling. At the moment the model covers more than 400 genes (reflecting over 2,000 paralogues), corresponding to more than 2,000 components. The potential effects of more than 50 different mutations (loss of function, activating and fusion) in several genes have been implemented (e.g. BCR-ABL, KRAS), and the potential effects of about 50 extra-cellular stimuli, such as growth factors (e.g. IGF, VEGF), hormones (e.g. testosterone) and organic compounds (e.g. phorbol esters) as well as almost 40 different inhibitors (e.g. EGFR inhibitor, MEK inhibitor) have been introduced. In total, these different components are connected to more than 3,000 reactions giving rise to over 3,500 kinetic parameters. Patient specific molecular data and clinical information were then used to generate a comprehensive individual model of the biology of the tumor, as well as key tissues of the patient in the "Virtual Patient" system. Levels of transcripts can be directly set based on the RNAseq results. Concentrations of specific proteins or protein modification states can be either based on proteomic results, or (with unavoidable uncertainty) estimated from protein synthesis rates assumed to be proportional to transcript levels, counterbalanced by a constant protein degradation rate. Secondly, RNAseq

is used to determine whether mutations detected in the patient's genome are expressed and must be modeled or can be ignored. To identify possible effects on surrogate markers of relevant cellular phenotypes (e.g. c-Myc for cell division, Casp-3 for apoptosis) modeling runs comparing the different conditions (every single mutation, autocrine loop, splice variant etc.) with a control state (no mutation, no growth factor) are performed. This allows the identification of key drivers of specific abilities which characterize cancer cells, such as sustained angiogenesis, limitless replicative potential and insensitivity to anti-growth signals, as proposed by Hanahan and Weinberg (2000) and to distinguish between the rare driver mutations and the thousands of passenger mutations typically found in tumors (Greenman et al. 2007; Ley et al. 2008; Mardis et al. 2009). By combining all alterations found, this gives a detailed picture of a patient's tumor. Having a patient-specific tumor model (according to the patient's mutation pattern and expression profile) opens up the possibility to predict a patient's response on various cancer drugs and drug combinations *in silico* and to find an optimal personalized cancer treatment. We therefore created a drug database, based on knowledge from textbooks, primary literature and several databases as part of the "Virtual Patient" system. The drug database up to now contains about 30 different drugs (e.g. Dasatinib, Erlotinib, Imatinib). Drugs are entered into the model by initializing appropriate inhibitors according to the known target proteins of the drug with ideally known values for dissociation constants (Karaman et al. 2008). The effect of a drug is analyzed using the concentration of c-Myc as a surrogate marker for clinical effectiveness, as elevated levels of this protein are indicative of cell division. In an automated procedure first all drugs are analyzed separately (Round1). In a second round all drugs that showed a positive effect on the level of c-Myc in Round1 are analyzed in double combination (Round2). These steps (Round3 uses triple combination etc.) are repeated until a drug combination is found that showed the highest positive effect on the level of c-Myc. Side effects of a drug/drug combination can be predict by identifying the effects of drug/drug combinations on normal cells stimulated by different types of growth factors. For this, modeling runs are carried out for the normal tissue responding to different growth factors in the presence and absence of a drug/drug combination.

The data of the melanoma patient (see Sect. 2) have been used to generate a comprehensive model of the biology of the tumor, as well as the cancer stem cells in the "Virtual Patient" system. Eight of the genes to be found mutated were already part of the model of which five were found to be not expressed by RNAseq analysis. Two of them could be identified as key drivers for proliferation. Beyond we could model the effect of twelve of the identified autocrine loops of which seven seem to be important for driving the tumor growth (see Fig. 1). Based on these mutations and autocrine loops we modeled a patient specific tumor and performed the drug optimization program. All individual drugs showed no or only little effect on the level of c-Myc. Pairwise combinations of effective drugs resulted in some cases in an increase in synergistic effects. The c-Myc level could be further reduced by a triple combination. Increasing doses of the optimal drug combination resulted in a predicted reduction of c-Myc levels to the c-Myc level predicted for non-stimulated normal cells (see Fig. 1; manuscript in preparation).

Since for one of the driving mutations found (we first assumed gain of function), the molecular effect of the missense mutation were not known, we also modeled this mutation as both loss of function and tolerated mutation (wild type function). The model was able to find that this mutation when not modeled as constitutively activating changed the sensitivity of the *in silico* patient's tumor model. In this case a combination of two drugs is already sufficient to stop proliferation as indicated by predicted reduced levels of c-Myc. Subsequent cellular testing in patient derived tumor cells is used to validate the predictions of the model. This is powerful example of how computational modeling coupled with deep sequencing and integrative bioinformatics analysis can identify the key drivers in an individual tumor and can eliminate those mutations that are either passenger, non-expressed or without impact for consideration of appropriate therapies.

4 Discussion and outlook

Here we present the "Virtual Patient" system that effectively provides a place where all current knowledge of cancer pathways and mutations as well as other relevant knowledge can reside. Using a Monte-Carlo-based approach it is able to construct accurate predictive cancer models of much more complex networks than possible up to now and can thereby help to improve the understanding on the consequences of cancer related mutations on a molecular pathway level and their functional effects on the
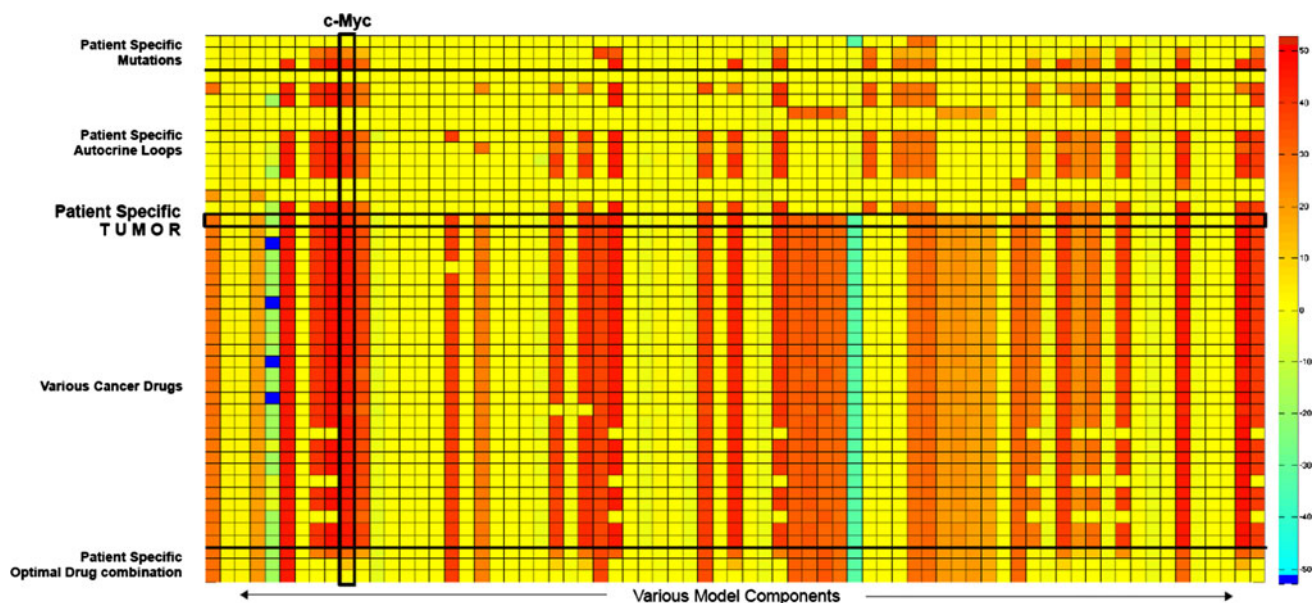
Fig. 1 A 'virtual' patient modeled according to the patient's mutation pattern and expression profile shows predictions of changes in key components of the model under different conditions (stimulation with growth factors, mutations, different drugs and drug combinations at different concentrations). The results show mean ratios of specific treatment experiments vs. a control state (no growth factor, no mutation, no drug). The effects of three mutations of the selected genes as well as the twelve autocrine loops in the patient were simulated individually and in combination to model a tumor of a specific melanoma patient. The model predicts a high c-Myc concentration in the tumor indicating a high rate of cell division. All individual drugs show no or only little effect on the level of c-Myc. Increasing doses of a combination of drugs in contrast can block c-Myc expression in the tumor, expected to stop cell division

cellular and organism level. Together with the complete molecular characterization of the tumor and somatic genomes and transcriptomes by application of next generation sequencing of individual cancer patients, the "Virtual Patient" system combines everything that can be known about a patient's disease with everything that is known about cancer as a whole as reflected in the patient-specific tumor model. Due to the rapid development of ultra high throughput sequencing techniques, promising sequencing cost of a few hundreds of dollars per Gigabase, an enormous effect on the use of predictive modeling for individual tumor fate can be expected. Hence, it is likely that the detailed molecular analysis of the genome of a particular cancer (as well as somatic genome of the patient) becomes a component of routine diagnostics in oncology. The impact on patient care, but also on economic viability, of personalized medicine approaches using predictive models has already been demonstrated. Using an ER response/Herceptin model and a 21-gene assay for breast cancer and through meta-analysis of a series of studies it was shown that the 21-gene assay is not only prognostic for recurrences, but can also predict the impact of chemotherapy and chemotherapy ineffectiveness. Specifically, chemotherapy benefit

was restricted to patients with high-Recurrence Score (RS), while cases with low-RS, despite having other clinical features of higher risk, had virtually no benefit from chemotherapy. Using such an approach it was possible to demonstrate significant overall savings – although the testing cost ca. $7,000/pt. The savings were ascribed to: a substantially lower number (20–35 %) of patients receiving chemotherapy and a higher overall survival rate due to the addition of curative chemotherapy to the 5 % of high-RS patients, otherwise considered in the pre-RS testing practice for treatment with hormones alone (Tsoi et al. 2010). Currently the "Virtual Patient" system is being used in the EU/EFPIA funded Innovative Medicine Initiative, OncoMark, to integrate high throughput 'omics' data and to guide the treatment of 60 cancer patients with non-metastatic colon cancer and in the German BMBF funded, TREAT20 project, where it is providing treatment recommendations for 20 metastatic melanoma patients based on patient's 'omics' data.

Another potential saving is given by the capacity of the "Virtual Patient" system to detect new biomarkers and new drug targets. It is suggested that the KRAS mutational status of metastatic colorectal cancer is predictive for the patient's response to an anti-EGFR therapy and that a routine KRAS analysis

would exclude 30–40 % of patients from anti-EGFR therapy, resulting in a potential financial saving of hundreds of millions of dollars (Patil et al. 2010). Importantly they go on to note, however, that a significant proportion of KRAS wild-type tumors do not, in fact, respond to anti-EGFR therapy and that additional testing of other genes or proteins within the EGFR signaling pathway will probably be required to help guide better patient selection. Since the "Virtual Patient" system integrates the complete patient's 'omics' data, it is expected that new markers beside KRAS can be identified and that an EGFR panel will become standard practice prior to initiating anti-EGFR therapy. Thus, more patients can escape a therapy they do not benefit from (or even make them more sick) and the public health system is financially relieved. A generation of novel markers will be also of significant advantage for patient stratification and/or therapy monitoring during clinical trials. The use of highly characterized markers, derived from a systematic analysis of a tumor type (e.g. breast, lung, colon) could dramatically impact the design of such clinical trials. One might expect a significant reduction in the number of patients required to assess new therapeutic agents – since the biomarkers selected would reflect the specific pathways impacted by the therapeutic agent. Likewise one could envisage dramatically shorter clinical trial times. Both scenarios would significantly reduce the lab bench – to patient bedside time, and markedly improve the efficiency of drug discovery for the pharmaceutical industry. The combined result of these effects would be to cap the spiraling costs of health care by yielding more efficacious drug use in the clinic and a reduced drug development time. In the field of biomarker/drug target discovery the "Virtual Patient" system is already involved in four projects: In the EU/EFPIA funded Innovative Medicine Initiative, OncoMark, is it used to identify biomarkers for colon cancer by analyzing a cohort of 60 colon carcinoma patients; in the project "Systems biology of prostate cancer" funded by the Austrian Nationalstiftung and the Austria Wirtschaftsservice GmbH in the framework of the IMGuS research program, where it supports the detection of prostate specific biomarkers for the prognosis of prostate cancer risk by integrating high-throughput 'omics' data of 50 prostate cancer patients into personalized computer models; and in the PREDICT project funded within the MedSys program of the BMBF, where it is used for the prediction of success of targeted therapies (e.g. cetuximab and erlotinib) on pre-clinical in vitro and in vivo models for lung cancer. In the MUTANOM project the "Virtual Patient" system is used to integrate data from mutational profiling, transcriptome and proteome data from breast/prostate/gastroentestinal cancer cells and to develop predictive models for the clinical use of chemotherapeutics, environmental determinants on diseases as well as for predicting disease progression in general.

The "Virtual Patient" system is a powerful weapon against cancer by finding an optimal drug/drug combination for individual cancer patient treatments and predicting new drug targets and diagnostic biomarkers resulting in improved cure rates in neoplastic disease.

References

Amgen (2009) ECCO15–ESMO34 European Multidisciplinary Congress in Berlin, Germany 2009, Abstract Number 14LBA

Berger MF, Kevin JZ et al (2010) Integrative analysis of the melanoma transcriptome. Genome Res 20:413–427

Bois FY (2009) GNU MCSim: Bayesian statistical inference for SBML-coded systems biology models. Bioinformatics 25(11): 1453–1454

Boyle P, Levin B (2008) World Cancer Report 2008, International Agency for Research on Cancer

Bozic I, Antal T et al (2010) Accumulation of driver and passenger mutations during tumor progression. Proc Natl Acad Sci USA 107(43):18545–18550

Brenner S, Johnson M et al (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol 18(6):630–634

Chen WW, Schoeberl B et al (2009) Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. Mol Syst Biol 5:239

Cheng L, Williamson SR et al (2010) Understanding the molecular genetics of renal cell neoplasia: implications for diagnosis, prognosis and therapy. Expert Rev Anticancer Ther 10:843–864

Di Nicolantonio F, Martini M et al (2008) Wild-type BRAF is required for response to panitumumab or cetuximab in metastatic colorectal cancer. J Clin Oncol 26(35):5705–5715

Forbes S, Clements J et al (2006) COSMIC 2005. Br J Cancer 94:318–322

Greenman C, Stephens P et al (2007) Patterns of somatic mutation in human cancer genomes. Nature 446(7132): 153–158

Haeno H, Michor F (2010) The evolution of tumor mestatses during clonal expansion. J Theor Biol 263(1):30–44

Hanahan D, Weinberg RA (2000) The hallmarks of cancer. Cell 100:57–70

Hays JL, Kim G et al (2010) Proteomics and ovarian cancer: Integrating proteomics information into clinical care. J Proteomics 73(10):1864–1872

International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. Nature 431:931–945

Jhawer M, Goel S et al (2008) PIK3CA mutation/PTEN expression status predicts response of colon cancer cells to epidermal growth factor receptor inhibitor cetuximab. Cancer Res 68:1953–1961

Joshi-Tope G, Gillespie M et al (2005) Reactome: a knowledge-base of biological pathways. Nucleic Acids Res 33(Database issue):D428–D432

Kamburov A, Wierling C et al (2009) ConsensusPathDB—a database for integrating human functional interaction networks. Nucl Acids Res 37:D623–D628

Kanehisa M, Goto S et al (2004) The KEGG resource for deciphering the genome. Nucl Acids Res 32(Database issue):D277–D280

Karaman MW, Herrgard S et al (2008) A quantitative analysis of kinase inhibitor selectivity. Nat Biotechnol 26(1):127–132

Karapetis CS, Khambata-Ford S et al (2008) K-ras Mutations and Benefit from Cetuximab in Advanced Colorectal Cancer. N Engl J Med 359:1757–1765

Karp PD, Ouzounis CA et al (2005) Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. Nucl Acids Res 33:6083–6089

Klipp E, Liebermeister E et al (2009) Systems biology: a textbook. Weinheim, Wiley-VCH

Lander ES et al (International Human Genome Sequencing Consortium) (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Ley TJ, Mardis ER et al (2008) DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature 456:66–72

Mardis ER, Ding L et al (2009) Recurring mutations found by sequencing an acute myeloid leukaemia genome. N Engl J Med 361:1058–1066

Mok TS, Wu YL et al (2009) Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med 361(10): 947–957

Monzani E, Facchetti F et al (2007) Melanoma contains CD133 and ABCG2 positive cells with enhanced tumourigenic potential. Eur J Cancer 43:935–946

Parkhomchuk D, Borodina T et al (2009) Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucl Acids Res 37(18):e123

Parmigiani G, Boca S et al (2009) Design and analysis issues in genome-wide somatic mutation studies of cancer. Genomics 93(1):17–21

Patil DT, Fraser CR, Plesec TP (2010) KRAS testing and its importance in colorectal cancer. Curr Oncol Rep 12:160–167

Pleasance ED, Cheetham RK et al (2010a) A comprehensive catalogue of somatic mutations from a human cancer genome. Nature 463(7278):191–196

Pleasance ED, Stephens PJ et al (2010b) A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature 463(7278):184–190

Reya T, Morrison SJ et al (2001) Stem cells, cancer, and cancer stem cells. Nature 414:105–111

Richard H, Schulz MH et al (2010) Prediction of alternative isoforms from exon expression levels in RNA-Seq experiments. Nucl Acids Res 38(10):e112

Schuster SC (2008) Next-generation sequencing transform today's biology. Nat Methods 5(1):16–18

Shah JB, Margulis V (2010) In search of a better crystal ball: recent advances in prognostic markers for clear-cell renal cell carcinoma. Expert Rev Anticancer Ther 10:837–842

Stark MS, Tyagi S et al (2010) Characterization of the Melanoma miRNAome by Deep Sequencing. PLoS ONE 5(3):e9685

Sultan M, Schulz MH et al (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science 321:956–960

Taneja P, Magkic D et al (2010) Classical and novel prognostic markers for breast cancer and their clinical significance. Clin Med Insights Oncol 20:15–34

Tsoi DT, Inoue M et al (2010) Cost-effectiveness analysis of recurrence score-guided treatment using a 21-gene assay in early breast cancer. Oncologist 15:457–465

Van Cutsem E, Dicato M et al (2010) Molecular markers and biological targeted therapies in metastatic colorectal cancer: expert opinion and recommendations derived from the 11th ESMO/World Congress on Gastrointestinal Cancer, Barcelona, 2009, Ann Onco. (Suppl 6):vi1–vi10

Vastric I, D'Eustachio P et al (2007) Reactome: a knowledge base of biologic pathways and processes. Genome Biol 8:R39

Weinberg RA (2007) The Biology of Cancer, Garland Science, New York

Wierling C, Herwig R, Lehrach H (2007) Resources, standards and tools for systems biology. Brief Funct Genomics Proteomics 6(3):S3–S8

Wierling C, Kühn A et al (2011) Prediction in the face of uncertainty: a Monte-Carlo-based simulation strategy for systems biology of cancer treatment Mutat Res (submitted)

Wilson PM, Labonte MJ, Lenz HJ (2010) Molecular markers in the treatment of metastatic colorectal cancer. Cancer J 16:262–272

Wood LD, Parsons DW et al (2007) The genomic landscapes of human breast and colorectal cancers. Science 318(5853): 1108–1113