SPONSORED BY THE

Federal Ministry
of Education
and Research

NGFN
National Genome
Research Network

SCIENCE INSIDE 2

THE NATIONAL GENOME RESEARCH NETWORK

SCIENCE INSIDE 2

THE NATIONAL GENOME RESEARCH NETWORK

FOREWORD

The decoding of the human genome at the beginning of this century was a milestone in the history of science. At the same time it was a starting point for the next but not lesser challenge: to elucidate the functions of the individual genes and to understand the full scope of their intricate interplay in the human organism. The results of genome research help us to understand the causes and development of human diseases and thus to develop new means of diagnosis, therapy and even prevention of diseases which have hitherto been difficult to treat or incurable.

Research on a high, internationally competitive level – and this is especially true for genome research – cannot be carried out in an isolated way. The deciphering of complex disease mechanisms, which are triggered by the interplay between different genes and between genes and the environment, requires the collaboration of many scientists with quite varied experience and expertise. In 2001 Germany's Federal Ministry of Education and Research (BMBF) launched a unique research program, the National Genome Research Network (NGFN). It combines systematic genome research with applied clinical research and bundles existing research competence and resource capacities.

In more than 300 laboratories at university and non-university research institutions in Germany, scientists from different disciplines are working intensively at the highest level to achieve the same objective: to elucidate the genetic causes and mechanisms of disease by means of genome research and thus open up new possibilities for improving health. The network structure of the NGFN has created the precondition for outstandingly productive forms of collaboration that are unequaled worldwide. The impulses emanating from this unique science network are remarkable; already at present they often form the nucleus for networks on the national or international level. Due to networks like the NGFN, Germany plays a major role internationally in the highly competitive field of genome research.

Meanwhile, the second funding phase of the NGFN is coming to a close. It is a success story: the combination of state-of-the-art technologies with systematic and clinical research approaches has produced many outstanding research results and generated an enormous increase in knowledge. Just to mention two examples: NGFN scientists have developed technologies to improve diagnostics and therapy planning in cancer diseases, and German research groups have played a major role in the identification of risk genes for alcohol addiction, asthma, intestinal diseases, heart attack and Parkinson's disease.

Scientific results obtained from medical genome research help to identify causative genes which were not anticipated, thus calling into question hypotheses gained from animal models. In the future, such new approaches to therapy shall be put on a faster track to industrial application, thus promoting the development of new diagnostics and medications. For that reason, it is absolutely essential to intensify and strengthen cooperation between universities, non-university research institutions and research companies. Only a combined effort will lead to market-relevant and transferable innovations – for the health, well-being and quality of life of each and every patient. The ultimate aim of many scientists is targeted prevention, i.e. to take action to preserve the health of genetically disposed individuals before the disease process begins.

With this brochure we warmly invite you to take a look into our laboratories and gain insight into the research conducted in the various research areas of the NGFN, based on selected projects as examples. Share the experiences of NGFN researchers and learn about the interlinked, innovative and future-oriented research being done in the network!

We wish you interesting and enjoyable reading!

Professor Stefan Schreiber

- Spokesperson of the Project Committee of the NGFN
- Department Head at the Clinic for General and Internal Medicine and Director of the Institute for Clinical Molecular Biology at the University of Kiel

Professor Martin Hrabé de Angelis

- Spokesperson of the Project Committee of the NGFN
- Director of the Institute of Experimental Genetics, GSF – National Research Center for Environment and Health, and Chair of Experimental Genetics at Technical University Munich

CONTENTS

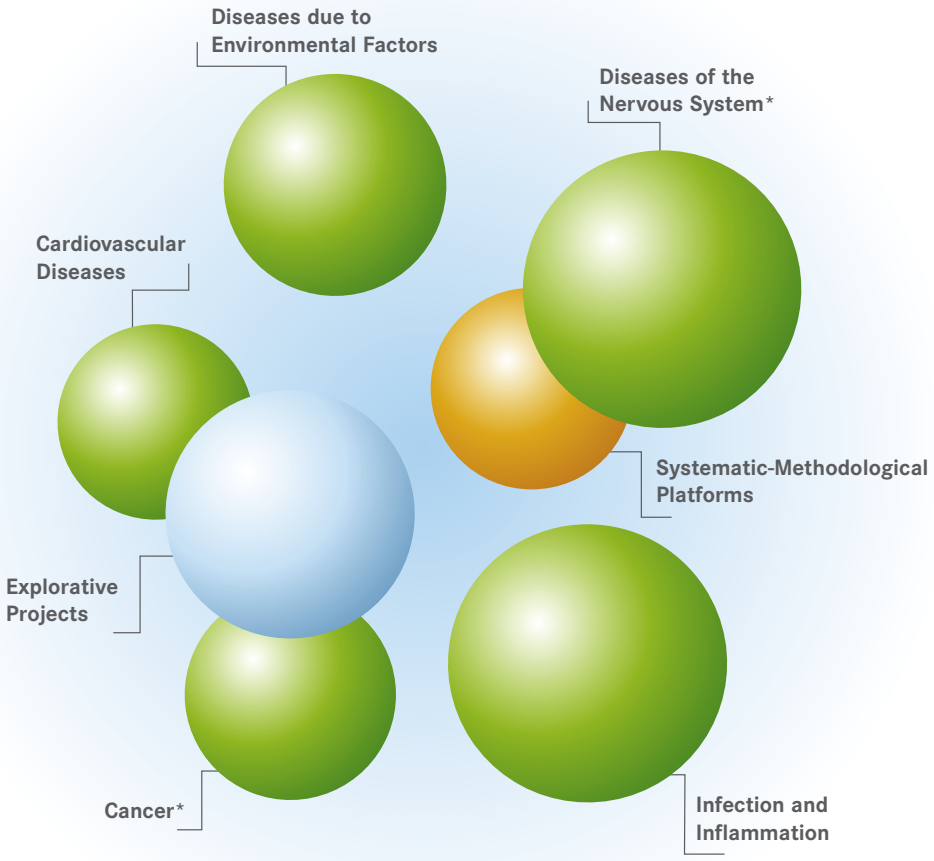
CONTENTS

CONTENTS

SYSTEMATIC-METHODOLOGICAL PLATFORMS (SMP)

CONTENTS

THE NATIONAL GENOME RESEARCH NETWORK



*three genome networks

Since 2001 the German Federal Ministry for Education and Research (BMBF) has been funding the National Genome Research Network (NGFN) to promote the elucidation of genes and how they function. The research work of the NGFN focuses on the investigation of the genetic causes of common diseases. To this end, a unique network structure has been created in which scientists specializing in systematic genome research and clinical research work closely together. The shared aim of all researchers in the network is to understand the intricate mechanisms that regulate how our body functions at the DNA, RNA and protein levels and thus to find new approaches for the treatment of complex diseases.

The NGFN builds on the German Human Genome Project (DHGP), which from 1995 to 2004 carried out fundamental genetic analyses to decipher the human genome, followed by functional analyses. Already during the initial funding phase of the National Genome Research Network from 2001 to 2004, NGFN scientists achieved impressive results. In 2003 the work of the NFGN was evaluated by an international panel of experts. The results of the evaluation gave rise to the second, current phase of funding.

DISEASE-RELATED AND SYSTEMATIC GENOME RESEARCH

NGFN scientists carry out their research within nine disease-related genome networks, which investigate disease processes in cancer, cardiovascular diseases, and diseases of the nervous system. They also investigate diseases resulting from infection, inflammation, or environmental factors. In the NGFN, patient-oriented research goes hand in hand with systematic genomic, transcriptomic and proteomic analysis. In order to reduce the costs and time required for such large-scale investigations to reasonable levels, research activities were organized in Systematic-Methodological Platforms (SMP). In 12 SMPs highly specialized experts utilize the semi- and fully automated technologies of modern high-throughput

screening and develop these further in a continual process. In addition, all NGFN scientists have access to state-of-the-art data processing. In 2004 the Explorative Projects (EP) were created to provide an instrument for tapping new technologies and application areas for human genome research. In a total of 19 EPs, scientists can test and realize their innovative ideas.

ESTABLISHING STANDARDS AND UTILIZING RESULTS

Comprehensive quality management ensures that all NGFN scientists work according to the same quality standards, which are oriented on international guidelines. Consistent high quality of the materials and results that are produced in the NGFN and optimal exchange of data within the network are ensured. In addition, a Technology Transfer Coordination Agency (KTT) was established within the NGFN. Its task is to identify commercially interesting findings made in the NGFN and to facilitate their utilization in industrial applications as quickly as possible. Patients will thus have greater chance of benefiting – in a timely manner – from new therapeutic and diagnostic methods developed in the National Genome Research Network.

ADVISING, COORDINATING AND ORGANIZING

The internal self-management of the NGFN takes place through the Project Committee. It consists of representatives of the Systematic-Methodological Platforms, the nine disease-related genome networks, and one representative for all of the Explorative Projects together. The members of the Project Committee monitor the status of the scientific projects and coordinate the research and the public relations work. The Project Management supports the Project Committee in its coordination and management tasks. It regularly informs about the situation and development of the overall project and organizes scientific conferences and symposia. With public relations measures the Project Management fosters the acceptance of medical genome research in the general public.

FAST TRACK FROM LAB TO CLINIC

In February 2007 BMBF launched a new call for proposals for a third funding phase of the NGFN. This includes two grant initiatives. The first initiative “Integrated Associations of Medical Genome Research (**NGFN*plus***)” shall expand the understanding of molecular biological and pathophysiological processes in common diseases under a clinical aspect and create approaches for the development of innovative methods and products for diagnosis and therapy. Molecular-biological know-how, relevant clinical-medical expertise and systematic and systems-biological approaches shall be combined with each other to achieve this objective.

In the second grant initiative “Innovation Alliances of Medical Genome Research (**NGFN*transfer***)” the transfer of scientific findings from medical genome research to their applications shall be supported. The goal is to more rapidly transfer results from academic research to industrial applications and thus to promote the development of new diagnostics and medications. Jointly and in close cooperation, research companies, universities and non-university research institutions shall generate market-relevant and transferable innovations.

The assessment process for both grant initiatives is currently ongoing.

LEUKEMIC STEM CELLS AND COLLABORATING GENETIC EVENTS – INSIGHTS INTO ACUTE MYELOID LEUKEMIA

MICHAELA FEURING-BUSKE, CHRISTIAN BUSKE

DEPARTMENT OF MEDICINE III, UNIVERSITY HOSPITAL GROSSHADERN & CLINICAL COOPERATIVE GROUP "LEUKEMIA"
GSF – NATIONAL RESEARCH CENTER FOR ENVIRONMENT AND HEALTH, MUNICH

INTRODUCTION

Experimental data have shown that acute leukemia originates from a malignant hematopoietic progenitor cell. The leukemic transformation is caused by critical genetic alterations such as fusion genes or mutations, which often affect proteins known to be key regulators of normal hematopoiesis. However, observations in patients with acute myeloid leukemia (AML) and experimental models have demonstrated that single genetic alterations are often not able to induce leukemia but need to collaborate with additional genetic hits to transform the hematopoietic cell. The biology of acute leukemias is determined not only by the nature of the (leukemia-associated) genetic alteration, but also by the characteristics of the cell initially affected and transformed by the genetic event. This leukemic stem cell (LSC) stands at the top of the hierarchy of the leukemic clone and is able to initiate and propagate the leukemia at the single cell level. Experimental data have shown that hematopoetic stem cells – and also more committed myeloid progenitors – can function as leukemic stem cells in AML. In the NGFN projects we characterized the collaborative potential of two of the most frequent genetic alterations in AML, the AML1-ETO fusion gene and the FLT3 length mutation (FLT3-LM) in a murine leukemia model. Furthermore, we succeeded in characterizing the leukemic stem cell population in a murine model of AML, characterized by the expression of the CALM-AF10 Fusion gene.

PROJECT STATUS/RESULTS

Collaboration of AML1-ETO and FLT3-LM in AML

We first screened 135 patients with *AML1-ETO*-positive AML for additional genetic alterations focusing on activating mutations in the receptor tyrosine kinases *FLT3* and *KIT* as well as in *NRAS* (*KITD816, NRAS* codon 12/13/61). Nearly one in three patients harbored such an activating mutation (28.1%) including *FLT3* (11 patients) or *KIT* (25 patients in total) or in *NRAS* (13 patients). These data demonstrated that FLT3 LM are recurrently found in patients with *AML1-ETO*-positive AML.

To test the functional significance of the association of FLT3-LM with the AML1-ETO fusion gene, we established a murine AML1-ETO positive AML model, using the murine bone marrow (BM) transplantation model and retroviral gene transfer to express AML1-ETO, FLT3-LM or both genetic alterations together in murine hematopoietic progenitor cells. We carried out long-term BM transplantation and over an observation period extending to 20.6 months, no disease developed in recipients of BM singly transduced with *AML1-ETO* (n=9) or *FLT3-LM* (n=9). In clear contrast, all recipients of doubly transduced BM (n=11 from 5 independent experiments) expressing both the fusion gene and the FLT3-LM died of an aggressive acute leukemia after a median latency time of 233 days post transplantation. Using short-term repopulating stem cell assays, we could demonstrate that the transforming activity of the FLT3-LM in collaboration with AML1-ETO was blocked by the kinase inhibitor PKC412.

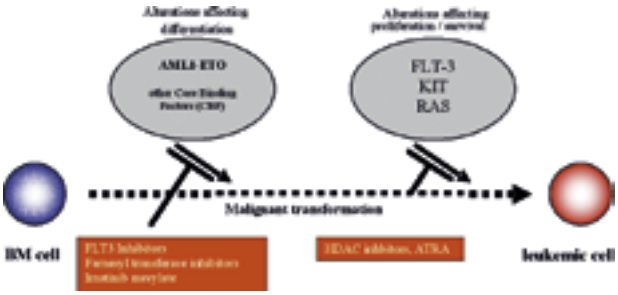


Fig. 1: Collaboration of genetic events in AML[2]

These data provided the first functional proof that AML1-ETO collaborates with a mutation recurrently found in patients with this subtype of AML (Fig. 1). Interestingly, mice developed different phenotypes of acute leukemia such as AML, but also acute lymphoblastic leukemia. This is not observed in patients carrying the AML1-ETO fusion gene since all of these patients develop AML. However, it is well known that the vast majority of these AML cases express at least one lymphoid antigen, and expression of the PAX5 transcription factor, known to commit cells to the B-cell lineage, has been described in a substantial proportion of patients with AML1-ETO positive leukemia. Although the underlying mechanism is not well understood, one possible explanation might be that an early progenitor cell with lymphoid characteristics belongs to the pool of leukemic stem cells in this AML subtype.

COORDINATOR
Professor Walter Birchmeier
Max Delbrück Center for Molecular Medicine, Berlin
wbirch@mdc-berlin.de

Characterization of the leukemic stem cell AML

The proof that a lymphoid cell can be the leukemic stem cell in AML could be demonstrated in a murine model of CALM-AF10-positive leukemia, which we established by using the retroviral gene transfer and murine BM transplantation. In this model, mice transplanted with bone marrow retrovirally engineered to express the leukemia-specific *CALM/AF10* fusion gene consistently developed an acute leukemia with short latency. The leukemia showed characteristic myeloid features such as the presence of myeloid marker positive cells, infiltrating multiple hematopoietic and non-hematopoietic organs, the positivity of blasts for myeloid specific markers and the depletion of the lymphoid compartment in lymphoid organs. Apart from the major population of cells expressing myeloid but not lymphoid markers, a smaller population of cells expressing myeloid markers as well as the lymphoid marker B220 and a minor population expressing only the B220 marker could be detected in all mice. We sought to identify which of these populations showed cancer stem cell characteristics in this mouse model of leukemia and found that the minor fraction of cells with B220 marker expression, but lacking myeloid markers was the population with the highest leukemogenicity. This was demonstrated in limiting dilution transplantation assays which demonstrated that one in 36 cells is able to cause AML in secondary recipients in the B220-positive, myeloid marker negative cell compartment.

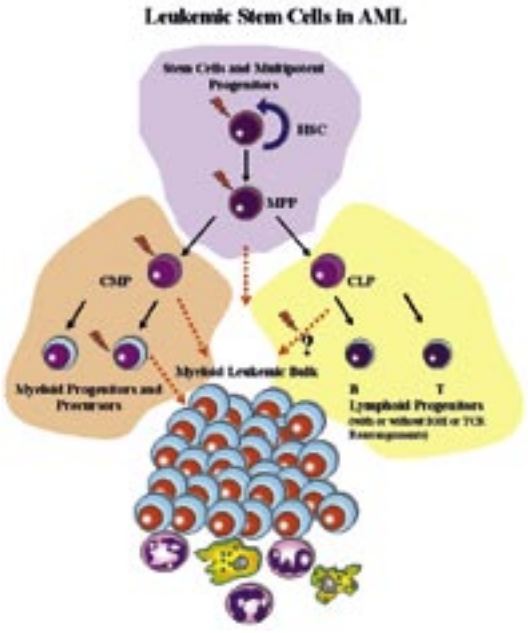


Fig. 2: Leukemic stem cell candidates in AML[4]

Moreover, at a single cell level, the leukemic B220$^+$/Mac1$^-$ cells, which showed IGH DJ rearrangements could give rise to the B220$^+$/Mac1$^+$ as well as the Mac1$^+$/B220$^-$ populations with the identical clonal DJ rearrangements *in vitro*, induced identical leukemias *in vivo*, demonstrating their capacity to give rise to the myeloid leukemic bulk at the single cell level. Antibody-based depletion of the B220$^+$ cells from the leukemia before secondary transplantation effectively prevented leukemia development in mice.

This murine model closely recapitulated C/A$^+$ human AML: BM cells of a majority of patients (7 of 9) displayed clonal IGH DJ rearrangements and in 3 of 3 patients tested, a significant proportion of C/A$^+$ IGH rearranged CD45RA$^+$ (B220) cells could be detected.

Taken together, these results indicate that progenitor cells with lymphoid characteristics can have leukemic stem cell characteristics in myeloid leukemias. This suggests that LSCs can have surface markers different from both normal stem cells and the tumor bulk, which might be useful for specific targeting of such LSCs while sparing the normal stem cell pool.

OUTLOOK

Ongoing projects will extend the analysis of collaborating events in acute myeloid leukemia, focusing on the AML1-ETO positive AML and the AML with normal karyotype and NPM mutation. They also focus on the characterization of leukemic stem cells in these two named subtypes of AML.

References

[1] *Schessl C et al. The AML1-ETO fusion gene and the FLT3 length mutation collaborate in inducing acute leukemia in mice. J Clin Invest. 2005; 115:2159-2168.*

[2] *Kuchenbauer F et al. AML1-ETO needs a partner: new insights into the pathogenesis of t(8;21) leukemia. Cell Cycle. 2005; 4:1716-1718.*

[3] *Deshpande AJ et al. Acute myeloid leukemia is propagated by a leukemic stem cell with lymphoid characteristics in a mouse model of CALM/AF10-positive leukemia. Cancer Cell. 2006; 10:363-374.*

[4] *Deshpande AJ et al. Lymphoid Progenitors as Candidate Cancer Stem Cells in AML: New Perspectives. Cell Cycle, 2007; 6:543-5.*

"NO GO!" FOR CANCER CELLS – A NOVEL THERAPEUTIC CONCEPT?

ROLAND H. STAUBER[1], THORSTEN BERG[2]

[1] ENT DEPARTMENT, MOLECULAR AND CELLULAR ONCOLOGY, UNIVERSITY HOSPITAL, MAINZ
[2] DEPARTMENT OF MOLECULAR BIOLOGY, MAX PLANCK INSTITUTE OF BIOCHEMISTRY, MARTINSRIED

INTRODUCTION

Translation of "genomic knowledge" into innovative strategies for diagnosis and therapy of cancer represents one of the biggest challenges for biomedical research, the healthcare economy and the biomedical industry. Major progress in this complex field requires dedicated interdisciplinary cooperation between basic, translational and clinical scientists.

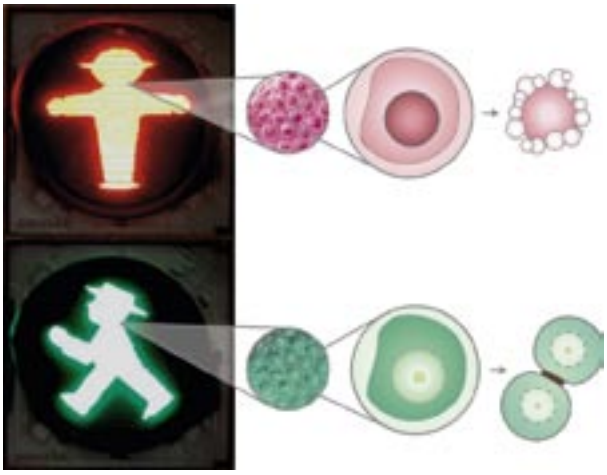


Fig. 1: "No Go!" for cancer cells. Targeted interference with the nucleocytoplasmic transport of proteins may force cancer cells to die.

The transfer of findings from laboratory-based human genome exploitation into the clinical arena requires: (1) a detailed understanding of the underlying molecular mechanism, and (2) the development and implementation of innovative knowledge-driven therapeutic strategies. Consequently, scientists at the University Hospital in Mainz and the Max Planck Institute in Martinsried combined the application of systematic genomic analysis together with a detailed molecular and patient-based characterization to demonstrate the implication of nucleocytoplasmic transport for the tumor promoting functions of the "inhibitor of apoptosis protein" (IAP) survivin.

The activation of many cancer-relevant signal transduction pathways depends on regulated nucleocytoplasmic transport, which is executed via specific protein-protein interactions. Hence, protein-protein interaction inhibitors (PPIs) are being intensively explored as novel therapeutics by academia and industry.[6] Transport takes place through the nuclear pore and is regulated by specific signals and transport factors. Nuclear import is mediated by nuclear import signals (NLS) interacting with import receptors in the cytoplasm. Nuclear export signals (NESs) are leucine-rich, interact with the export receptor Crm1 in the nucleus and depend on the small GTPase Ran, which controls the Crm1/substrate interaction.

Survivin is highly expressed in most cancers and proposed to function as a mitotic regulator and an apoptosis inhibitor. Importantly, survivin's expression has been correlated with resistance against cancer therapy-induced apoptosis and abbreviated patient survival, and hence is regarded a *bona fide* therapeutic target.

PROJECT STATUS/RESULTS

Survivin is a highly mobile protein overexpressed in tumor cells

Genome-wide expression profiling of tumor material from patients suffering from breast, colon, and head and neck cancer revealed that survivin is significantly overexpressed in the tumor tissue compared to non-malignant cells.[4] By applying innovative microscopic techniques and computer-guided microinjection in living cells, we could show that survivin is able to shuttle between the nucleus and the cytoplasm. The function of Crm1 as an export receptor creates a cytoplasmic survivin concentration gradient. A pronounced cytoplasmic survivin localization seems to promote survivin's cytoprotective activity against radiation- and chemotherapy-induced cell death by facilitating its interplay with the apoptotic machinery[5] (Fig. 2). During mitosis, the Crm1/survivin interaction is critically involved in tethering the chromosomal passenger complex (CPC) to the centromeres and thus may facilitate tumor progression by providing a mechanism to tolerate increasing chromosomal abnormalities[3] (Fig. 2). In addition, our results provide novel insights into how interphase transport receptors take over novel functions during mitosis.

Nuclear survivin – better patient survival

The clinical significance of our findings is underlined by analyzing the intracellular localization of survivin in a large number of tumor biopsies by immunohistochemistry. Uni- as well as multivariate analysis revealed that a preferential nuclear localization of survivin, indicative of disturbed nuclear export, turned out to be a statistically highly significant favorable prognostic marker in colon and head and neck cancer patients.[1]

COORDINATOR
Professor Walter Birchmeier
Max Delbrück Center for Molecular Medicine, Berlin
wbirch@mdc-berlin.de

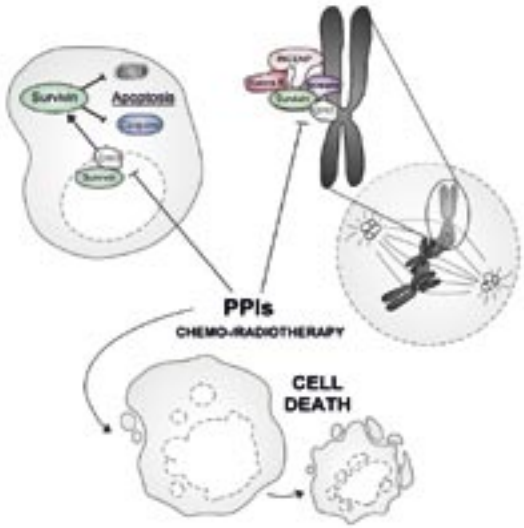


Fig. 2: In interphase cells, nuclear export promotes a high cytoplasmic concentration of survivin to counteract pro-apoptotic stimuli. During mitosis, the Crm1/survivin interaction is critically involved in tethering the CPC to the centromere. Pharmacological targeting of the survivin/Crm1 interaction by protein–protein interaction inhibitors (PPIs) in combination with current chemoradiation treatment protocols may result in increased tumor cell death and promote patient survival.

OUTLOOK

"No Go" for cancer cells as a novel therapeutic concept

Since the survivin network is exploited in virtually every human cancer, our data combined with our cell-based screening assays[2] emphasize the exploitation of the pharmacogenetic interference with survivin's export as a novel therapeutic strategy.[7] "The challenge for future NGFN programs will be whether a survivin-tailored "*NO GO! - for Cancer Cells*" (Fig. 1) can be achieved by applying chemical biology approaches," Professor Stauber predicts. "In addition, we will have to see whether it will be effective in a clinical setting."

References

[1] *Engels K et al. Dynamic intracellular survivin in oral squamous cell carcinoma: underlying molecular mechanism and potential as an early prognostic marker. J Pathol 211, 532 (2007).*

[2] *Knauer SK et al. Translocation biosensors to study signal-specific nucleo-cytoplasmic transport, protease activity and protein-protein interactions. Traffic 6, 594 (2005).*

[3] *Knauer SK et al. The Survivin-Crm1 interaction is essential for chromosomal passenger complex localization and function. EMBO Rep 7, 1259 (2006).*

[4] *Knauer SK et al. The Survivin Isoform Survivin-3B is Cytoprotective and can Function as a Chromosomal Passenger Complex Protein. Cell Cycle 6, 553 (2007a).*

[5] *Knauer SK et al. Nuclear export is essential for the tumor-promoting activity of survivin. Faseb J 21, 207 (2007b).*

[6] *Schust J et al. Stattic: a small-molecule inhibitor of STAT3 activation and dimerization. Chem Biol 13, 1235 (2006).*

[7] *Stauber RH et al. Nuclear and Cytoplasmic Survivin: Molecular Mechanism, Prognostic, and Therapeutic Potential. Cancer Research 67, 5579 (2007).*

TARGETING CLASS I HISTONE DEACETYLASES IN NEUROBLASTOMA

OLAF WITT, HEDWIG DEUBZER, INA OEHME, DENNIS WEGENER

CLINICAL COOPERATION UNIT PEDIATRIC ONCOLOGY, GERMAN CANCER RESEARCH CENTER, HEIDELBERG

INTRODUCTION

Childhood neuroblastoma is a unique cancer with clinical courses ranging from spontaneous differentiation and regression to fatal aggressive growth and metastasis. The tumor originates from precursor cells of the sympathetic nervous system and develops early in childhood. One of the prominent genetic alterations occurring in neuroblastoma is amplification of the oncogene MYCN. Amplification of MYCN is associated with poor prognosis and is a current molecular risk stratification parameter in neuroblastoma trials. Thus, neuroblastoma covers the whole spectrum of cancer biology and provides an interesting model system to study fundamental processes in tumor biology.

Histone deacetylases (HDACs) are a family of enzymes that remove acetyl groups from lysine residues of histone as well as non-histone proteins, such as transcription factors. HDACs play an important function in the regulation of chromatin structure by controlling the acetylation level of histone proteins, thereby influencing histone-DNA interactions and transcriptional activity. More recently, the activity of several other proteins, for example of p53, has also been described to be controlled by HDACs through regulation of their acetylation levels. The discovery of HDACs has been facilitated by inhibitors of their enzymatic activity and subsequent affinity purification. Most interestingly, these inhibitors have been found to change fundamental biological features of cancer cells such as proliferation, cell cycle, apoptosis, differentiation and angiogenesis. Several HDAC inhibitors are currently in clinical trials. As most if not all of these currently available compounds are so called pan-inhibitors with unselective targeting of several HDACs, little is known about the contribution of individual HDAC family members in controlling malignant genetic programs in cancer cells. However, identification of the responsible HDAC(s) would enable the design of specific compounds and targeted therapy and contribute to the molecular understanding of tumorigenesis.

Within the Neuroblastoma Network, the goal of our research team is to (1) develop novel HDAC inhibitors, (2) identify functionally relevant HDAC family members and (3) characterize key molecular players mediating differentiation programs through HDAC inhibition in neuroblastoma.

PROJECT STATUS/RESULTS

Class I HDACs as novel potential drug targets for neuroblastoma therapy

We investigated the expression of all HDAC family members in a panel of 100 primary neuroblastoma (NB) samples from the German Neuroblastoma Trial using a neuroblastoma-customized microarray developed by groups of the Neuroblastoma Network. We found all 11 HDAC family members to be expressed in all samples and several neuroblastoma cell lines. Interestingly, only the expression of distinct class I HDACs were associated with poor prognostic markers such as advanced stage metastatic disease and MYCN oncogene amplification. In collaboration with Cenix Biosciences, Dresden, a functional screening of 100 neuroblastoma-related candidate genes including HDACs was performed using RNA interference technology. In this screen, class I HDACs showed functional relevance in a cultured neuroblastoma model. This prompted us to look in more detail into the functional role of these genes for neuroblastoma tumorigenesis. We discovered that certain class I HDACs are involved in regulation of cell proliferation, cell cycle progression and clonogenic growth and that apoptotic and differentiation programs are controlled by distinct HDACs. Thus, only a limited number of HDACs appear to be cancer relevant.

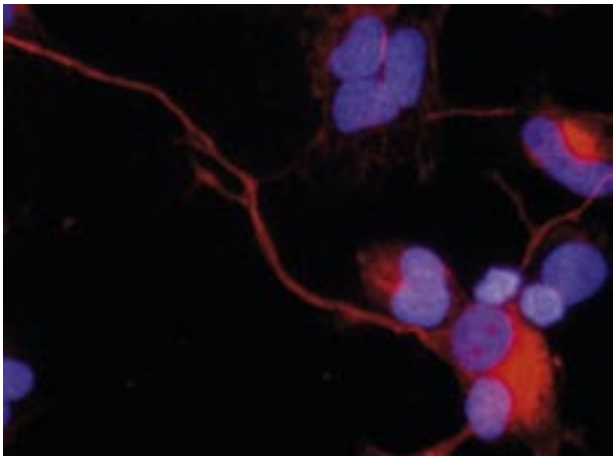


Fig. 1: Knockdown of HDACs induces differentiation of neuroblastoma cells

COORDINATOR
Professor Manfred Schwab
German Cancer Research Center, Heidelberg
m.schwab@dkfz.de

Histone deacetylase inhibitor Helminthosporium carbonum (HC) toxin suppresses the malignant phenotype of neuroblastoma cells

The survival rate of children with high-risk neuroblastoma (NB) is unsatisfactory despite intensive multimodal therapy. Limited efficacy and serious side effects of currently used therapeutic regimens necessitate new, less toxic treatments.

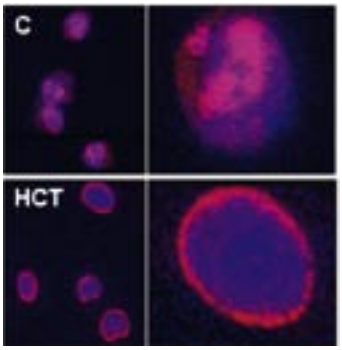


Fig. 2: HC toxin changes histone acetylation patterns in the nuclei of neuroblastoma cells

The histone deacetylase inhibitor (HDACI) *Helminthosporium carbonum* (HC) toxin, a fungal metabolite of tetrapeptide structure, was found to suppress the malignant phenotype of NB cells with and without amplified MYCN more efficiently than retinoids and all other HDACIs tested at dosages below 20 nM. The shift to a benign and differentiated phenotype is associated with an activation of the retinoblastoma (RB) tumor suppressor network. This includes the transcriptional repression of cell cycle regulators found at high levels in NBs associated with poor prognosis, like E2F-1, N-myc, Skp2, Mad2 and survivin proteins.

Whole genome expression analysis of HC toxin-treated NB cells followed by the functional characterization of 100 candidate genes with medium-throughput RNA-interference identified novel early regulated candidate genes for NB cell differentiation and potential novel therapeutic intervention strategies. In the future, we aim to further analyze the potential of HC toxin for NB therapy in both *in vitro* and *in vivo* preclinical models and to identify the underlying molecular mechanisms.

A novel HDAC inhibitor identified by high throughput screening, has potent anti-neuro-blastoma activity

The promising effects of HDAC inhibition on neuroblastoma cells prompted us to further evaluate novel compounds obtained by high throughput screening of a natural compound library. In collaboration with the Institute of Microbiology and Genetics, University of Göttingen, we evaluated a compound with a novel lead structure, which was identified by a screening campaign. The compound inhibits several HDAC preparations in the sub-micromolar range and showed potent activity against several neuroblastoma cell lines, i.e. inhibition of proliferation, cell viability, clonogenic growth and induction of apoptosis.

OUTLOOK

Our data now provide a rational basis for the development of selective HDAC inhibitors. This will require establishment of appropriate assay systems which are currently being set up with the aim of discovering the next generation of HDAC inhibitors for neuroblastoma treatment.

Because of their fundamental impact on cancer biology, the molecular pathways controlled by individual HDAC family members need to be studied in detail.

The promising preclinical activity of HC toxin against neuroblastoma now warrants safety and efficacy studies in mouse models in order to further develop the compound towards clinical application. Finally, candidate master regulator genes and molecular pathways involved in differentiation of neuroblastoma cells by HC toxin are currently being explored.

References

[1] *Deubzer HE et al. Means and methods for treating advanced stage neuroblastoma. 23.12.2005: EP 05 028 301.9.*

[2] *Deubzer HE et al. Histone deacetylase inhibitor Helminthosporium carbonum toxin for suppressing malignant qualities of neuroblastoma cells. 20.10.2006: US 60/853,022.*

[3] *Deubzer HE et al. Means and methods for treating advanced stage neuroblastoma. 22.12.2006: PCT/EP2006/069776.*

BEYOND GENOMICS: ROLE OF microRNAS IN NEUROBLASTOMA

JOHANNES H. SCHULTE, ALEXANDER SCHRAMM, ANGELIKA EGGERT

PEDIATRIC ONCOLOGY RESEARCH, DEPARTMENT OF PEDIATRIC HEMATOLOGY AND ONCOLOGY, UNIVERSITY HOSPITAL ESSEN

INTRODUCTION

MicroRNAs (miRNAs) are a class of small, noncoding RNAs (22nt) with a specific hairpin structure that act as negative regulators of protein expression. They promote mRNA degradation and repress mRNA translation by sequence specific interaction with mRNAs. Several hundreds of human miRNAs have been identified within the last years. They are involved in the regulation of multiple physiological processes including apoptosis, proliferation and differentiation. There is growing evidence that miRNA dysregulation is involved in the pathogenesis of various diseases, in particular cancer. High-throughput miRNA expression profiling has identified specific miRNAs induced or repressed in different tumor entities. Functional evidence for miRNA involvement in cancerogenesis exists: dysregulated miRNAs may inhibit apoptosis or downregulate angiogenic inhibitors.

Fig. 1: Studying the characteristics of neuroblastoma cells

We asked if miRNA dysregulation also plays a role in the biology of neuroblastoma, a common solid tumor of childhood, which is derived from primitive sympathetic neuroblasts. Amplification of the *MYCN* oncogene is a strong predictor of poor prognosis in neuroblastoma, and is used worldwide as a parameter for treatment stratification in clinical neuroblastoma trials. The MYCN protein is a transcription factor expressed during neural crest development. MYCN has been shown to promote cell cycle progression in neuroblastoma cells *in vitro*, and transgenic overexpression of MYCN is sufficient to induce the development of neuroblastoma in mice. Using high-throughput gene expression profiling, distinct sets of MYCN-regulated target genes have been identified in neuroblastoma by members of the NGFN Neuroblastoma Network and others.[1-3] In addition to numerous genes induced by MYCN, all studies have reported significant numbers of genes to be downregulated. MYCN induces gene transcription via specific E-box sequences in target gene promoters. The mechanisms by which MYCN silences genes are currently less well understood.

Given that MYCN is a major regulator of neuroblastoma tumor biology, we investigated (1) if specific miRNAs are regulated by MYCN in neuroblastoma cells, and (2) if these miRNAs are also differentially expressed between MYCN non-amplified and MYCN amplified neuroblastomas *in vivo*.

PROJECT STATUS/RESULTS

Establishing a cell culture model for conditional MYCN expression

To analyze the effect of MYCN on the miRNA expression profile in neuroblastoma cells *in vitro*, a novel model system for ectopic MYCN overexpression was established and validated by cooperation partner S. Horn, Marburg: the MYCN non-amplified human neuroblastoma cell line, SH-EP, was stably transfected with an expression vector encoding for MYCN under the control of an estrogen responsive domain (MYCN-ER). Activation of MYCN-ER in SH-EP cells induced morphological changes resulting in undifferentiated cells resembling highly aggressive neuroblastomas. We also observed an increase of cells in the S and G2/M phase.

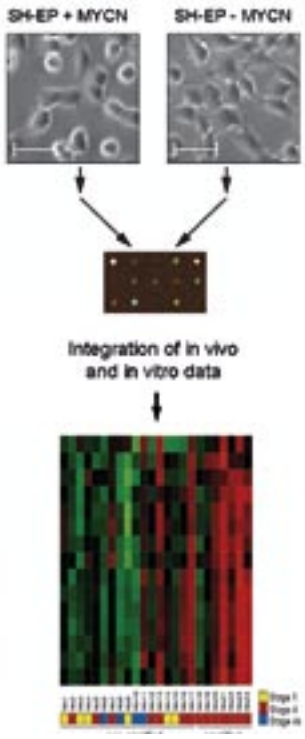


Fig. 2: Workflow – analysis of miRNA expression in neuroblastoma

Identification of MYCN target miRNAs *in vitro*

We analyzed the global regulation of miRNA expression by MYCN in this inducible SH-EP MYCN-ER cell culture model (Fig. 2). A custom-made miRNA microarray ("miRNA-chip") based on a library containing probes directed against 384 miRNAs was developed in a joint effort of the University Children's Hospital Essen (AG Eggert) and the University of Marburg (AG Eilers). Upon MYCN induction, expression of 15 unique miRNAs was induced at least

COORDINATOR
Professor Manfred Schwab
German Cancer Research Center, Heidelberg
m.schwab@dkfz.de

two-fold. Surprisingly, no miRNA was downregulated. A set of 160 primer pairs was then employed to analyze and validate miRNA expression using qPCR with stem-loop reverse transcription. The qPCR was implemented in a high-throughput manner using pipetting robots. In this way, we validated expression levels of 11 out of the 15 differentially regulated miRNAs.

E-box sequences were detected in the promoter regions of these 11 regulated miRNAs. To further validate miRNAs induced by MYCN expression in our inducible model system, we analyzed expression of selected miRNAs in MYCN amplified versus non-amplified neuroblastoma cell lines. Indeed, expression of the previously identified regulated miRNAs was low in MYCN non-amplified SH-EP and SK-N-AS cells and significantly higher in MYCN amplified Kelly, LAN1 and BE2C cells.

Analysis of miRNA expression in primary neuroblastomas

We then used the miRNA chip to analyze miRNA expression in pre-treatment primary neuroblastomas of different tumor stages with and without MYCN amplification. Differential expression of 14 miRNAs was detected between MYCN-amplified and non-amplified neuroblastomas according to the significance analysis of microarrays (SAM) algorithm. All of these miRNAs were upregulated in the MYCN-amplified tumors.

Integration of *in vitro* and i*n vivo* data

Comparison of miRNAs upregulated in both the in vitro culture system overexpressing MYCN and tumors with amplified MYCN revealed that 7 miRNAs were commonly upregulated. QPCR confirmed MYCN-mediated regulation of 6 of the 7 miRNAs. Interestingly, 4 of these 7 miRNAs are encoded by three paralogous miRNA clusters, the miR-17 cluster on chromosome 13, the miR-106b cluster on chromosome 7 and the miR-106a cluster on the X chromosome.

OUTLOOK

Several lines of evidence point to the oncogenic properties of the miRNAs of the miR-17 and miR-106a cluster. In carcinomas, overexpression of miR-17-5p and miR-106a is a general feature and in lymphomas the miR-17 cluster is both amplified and overexpressed. Overexpression of the miR-17 cluster in haematopoietic stem cells led to lymphoma formation in mice, providing functional evidence for the involvement of those miRNAs in lymphomagenesis. In addition, the miR-17 miRNA cluster induced by c-Myc was demonstrated to enhance angiogenesis by downregulating angiogenic inhibitors in a murine colon tumor model.

Additionally, we found and validated one specific miRNA to be induced by MYCN *in vitro* and to be differentially regulated between MYCN-amplified and non-amplified neuroblastomas. This miRNA has not been linked to the Myc family of transcription factors before, suggesting that it is either an miRNA targeted specifically by MYCN or by MYCN in the context of neuroblastoma.

In addition to the potential role of MYCN-regulated miRNAs in neuroblastoma pathogenesis, miRNA induction by MYCN might be a general mechanism of MYCN-mediated protein regulation. Translational silencing mediated by MYCN-induced miRNA upregulation is an independent additional mechanism by which MYCN could downregulate gene expression.

The functional implications of miRNA deregulation in neuroblastoma, as well as evaluation of their potential application as novel targets for therapy using “antagomirs” deserves attention in future studies. A prerequisite to establish miRNA directed therapies is the generation of appropriate preclinical mouse models. This includes the generation of conditional transgenic mice allowing deletion or overexpression of oncogenic miRNAs. We intend to use such models to analyze (1) the physiological role of oncogenic miRNAs in embryonal development and in the adult animal and (2) the potential role of oncogenic of NB tumorigenesis *in vivo*.

References

[1] *Schramm A et al. Prediction of clinical outcome and biological characterization of neuroblastoma by expression profiling. Oncogene, 24: 7902-7912, 2005.*

[2] *Oberthuer A et al. Customized oligonucleotide microarray gene expression-based classification of neuroblastoma patients outperforms current clinical risk stratification. J Clin Oncol, 24: 5070-5078, 2006.*

[3] *Berwanger B et al. Loss of a FYN-regulated differentiation and growth arrest pathway in advanced stage neuroblastoma. Cancer Cell, 2: 377-386, 2002.*

[4] *Schulte JH et al. MYCN regulates oncogenic microRNAs in neuroblastoma. Int. J. of Cancer, in press.*

IDENTIFICATION OF PATHOPHYSIOLOGICAL RELEVANT CANDIDATE GENES IN GLIOMA

MEINHARD HAHN[1], BJÖRN TEWS[1], GRISCHA TÖDT[1], JÖRG FELSBERG[2], MARIETTA WOLTER[2], PETER ROERIG[2], DETLEF TROST[3], ANNEGRET KUNITZ[4], CHRISTIAN HARTMANN[5], RUTHILD G. WEBER[3], ANDREAS VON DEIMLING[5], OTMAR D. WIESTLER[6], PETER LICHTER[1], GUIDO REIFENBERGER[2]

[1] DIVISION OF MOLECULAR GENETICS (B060), GERMAN CANCER RESEARCH CENTER, HEIDELBERG
[2] DEPARTMENT OF NEUROPATHOLOGY, UNIVERSITY OF DÜSSELDORF
[3] DEPARTMENT OF HUMAN GENETICS, UNIVERSITY OF BONN
[4] DEPARTMENT OF NEUROPATHOLOGY, CHARITÉ – UNIVERSITY MEDICINE BERLIN
[5] DEPARTMENT OF NEUROPATHOLOGY, UNIVERSITY OF HEIDELBERG
[6] GERMAN CANCER RESEARCH CENTER, HEIDELBERG

INTRODUCTION

Gliomas, tumors of the primary central nervous system, are the most common primary brain tumors, originating from glial cells. They represent a heterogeneous group of neoplasms with respect to morphological appearance, biological behavior and genetic alterations, as well as response to therapy and clinical outcome. Oligodendroglioma represents a subgroup of low-grade glioma (WHO grade II and III), which derives from oligodendrocytes and frequently shows combined loss of the chromosomal arms 1p and 19q (loss of heterozygosity, LOH, in >75% of cases). It has become evident that LOH 1p/19q is an independent predictive marker of better response to radio- and chemotherapy as well as longer survival. Presumably, these deleted chromosomal regions carry yet unknown genes with a pathogenetic role and relevance for therapy response.

Within our brain tumor research network (BTN) we were interested in identifying pathophysiologically relevant candidate genes in low-grade glioma using genome-wide approaches for the detection and analysis of both genomic imbalances and aberrations at the level of the tumor transcriptome. Such genes might be relevant molecular markers for prognosis and therapy response.

PROJECT STATUS/RESULTS

Genomic aberrations associated with shorter overall survival in patients with oligodendro-glioma

To assess in a comprehensive manner the prognostic significance of genomic aberrations in oligodendroglioma (beside the frequent 1p/19q LOH), a retrospective study of 70 patients was carried out using the genome-wide screening method of comparative genomic hybridization (CGH). CGH data were correlated with long-term overall patient survival (OS). The most frequent aberrations were losses of chromosome 19q (64%), 1p (59%), 9p (26%) and gains on 7q (24%), 19p (19%) and 7p (17%). Combined 1p/19q and 19q losses were associated with longer OS, gains on 7, 8q, 19q, 20, and losses on 9p, 10 18q, Xp with shorter OS. The most significant prognostic factors for OS of patients with any oligodendroglioma are, besides the WHO grade, 7p gains and 9p losses. In summary, several independent genomic markers of shorter survival in oligodendroglioma patients could be identified. Therefore, molecular diagnostic testing of this tumor entity should be refined by additionally appraising the prognostically unfavorable genomic aberrations identified in this study[1].

Identification of novel oligodendroglioma-associated candidate tumor suppressors in 1p36 and 19q13

To identify downregulated candidate tumor suppressor genes in the minimal deleted chromosome regions 1p36 and 19q13, we performed cDNA microarray-based RNA expression profiling of 35 low-grade gliomas, using a specific microarray type with complete coverage of the genes located in 1p36.13-p36.31 and 19q13.2-q13.33. We identified 8 genes from these regions which are significantly downregulated in gliomas with 1p/19q LOH: *MGC4399, SRM, ICMT, RPL18, FTL, ZIN, FLJ10781* and *DBP*. Data were confirmed by quantitative real-time RT-PCR. This set of interesting novel candidate genes may play important roles in the pathogenesis of oligodendroglioma[2].

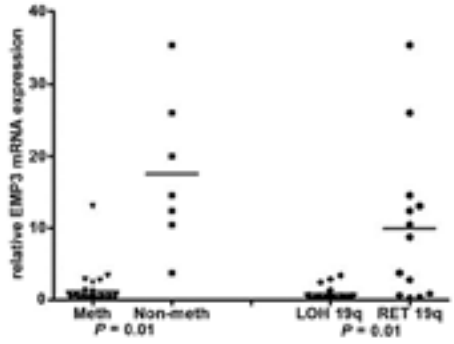


Fig.1: Relative EMP3 mRNA expression levels in oligodendroglial tumors with (n=34) and without (n=7) hypermethylation of EMP3. The left part of the dot plot diagram presents the relative EMP3 expression against the hypermethylation status of each case (Meth: EMP3 hypermethylated; Non-meth: EMP3 not hypermethylated), the right part against the 19q LOH status (LOH: loss of heterozygosity; RET: retention of heterozygosity). Horizontal bars: mean expression levels. Expression levels are significantly reduced in tumors with EMP3 hypermethylation and tumors with 19q LOH.

COORDINATOR
Professor Otmar D. Wiestler
German Cancer Research Center, Heidelberg
o.wiestler@dkfz.de

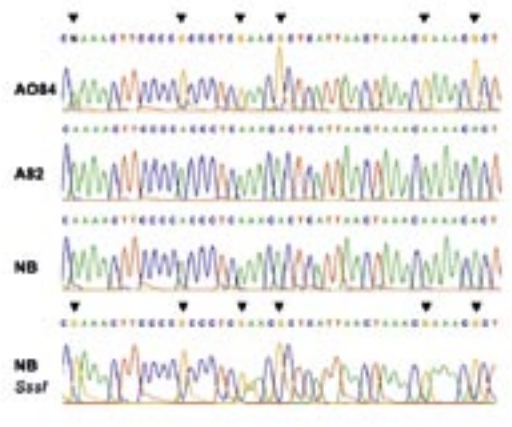


Fig. 2: Sequencing of sodium bisulfite treated DNA identifies methylated CpG-sites within the CITED4 promoter-associated CpG-island. The genomic DNA of glioma AO84 (LOH 1p/19q positive) shows methylated CpGs (indicated by arrow heads), in contrast to glioma A82 (LOH 1p/19q negative) and normal brain tissue (NB). The lower panel shows the positive control: normal brain DNA which was methylated in vitro using CpG methylase Ssp I.

DNA hypermethylation and aberrant mRNA expression of the gene *EMP3* at 19q13.3 in glioma

The gene *EMP3* at 19q13.3 was among those genes showing the most pronounced expression difference in cDNA microarray expression profiles of low-grade gliomas. The molecular analysis of this candidate TSG by mutation analysis of 162 human gliomas did not reveal *EMP3* alterations. In oligodendroglial tumors, we found that aberrant methylation in the 5'-region of *EMP3* was significantly associated with reduced mRNA expression and LOH 19q (see Fig. 1). Hypermethylation was detected in >80% of secondary glioblastomas while primary glioblastomas mostly lacked *EMP3* hypermethylation and frequently overexpressed this gene. These findings suggest that these two glioblastoma groups significantly differ in their epigenetic aberrations[3].

Hypermethylation and transcriptional down-regulation of *CITED4* at 1p34.2 in glioma with LOH at 1p/19q

By using microarray-based expression profiling, we found that oligodendroglial tumors with 1p/19q LOH showed significantly lower expression of the gene *CITED4* at 1p34.2 as compared to tumors without these losses. Mutational analysis showed no *CITED4* mutations in glioma. However, 1p/19q losses as well as low expression of *CITED4* transcripts were significantly associated with hypermethylation of the *CITED4*-associated CpG island (see Fig. 2). *CITED4* hypermethylation was significantly associated with longer recurrence-free and overall survival of patients with oligodendroglial tumors (see Fig. 3). Taken together, our results indicate that *CITED4* is epigenetically silenced in the vast majority of oligodendroglial tumors with 1p/19q LOH and suggest *CITED4* hypermethylation as a novel prognostic marker in oligodendroglioma patients[4].

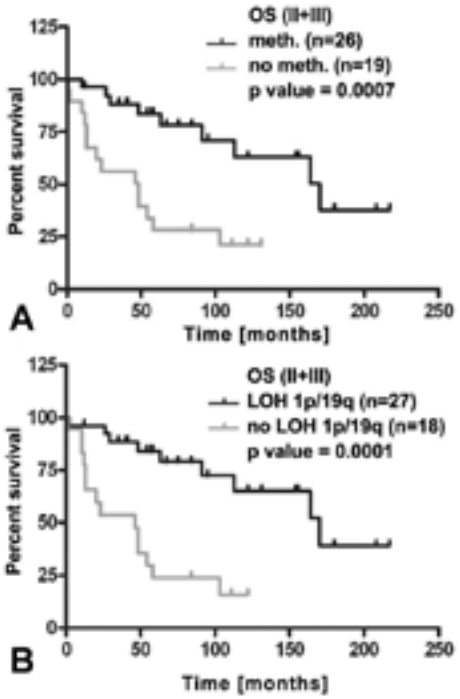


Fig. 3: Kaplan-Meier survival curves obtained for 45 glioma patients with tumors of WHO grade II (n=27) and grade III (n=19) showing the associations **(A)** between hypermethylation of *CITED4* and overall survival (OS) and **(B)** LOH of chromosomal regions 1p / 19q and OS.

OUTLOOK

The identified marker genes will be functionally characterized and their potential as robust biomarkers for prognosis and/or prediction of therapy response of glioma patients will be validated.

References

[1] *Trost D et al. (2007) Identification of genomic aberrations associated with shorter overall survival in patients with oligodendroglial tumors. Int J Cancer. 120(11):2368-2376.*

[2] *Tews B et al. (2006) Identification of novel oligodendroglioma-associated candidate tumor suppressor genes in 1p36 and 19q13 using microarray-based expression profiling. Int J Cancer. 119(4):792-800.*

[3] *Kunitz A et al. (2007) DNA hypermethylation and aberrant expression of the EMP3 gene at 19q13.3 in human gliomas. Brain Pathol. Epub 2007 Jul 4.*

[4] *Tews B et al. (2007) Hypermethylation and transcriptional downregulation of the CITED4 gene at 1p34.2 in oligodendroglial tumours with allelic losses on 1p and 19q. Oncogene. 2007; 26(34):5010-5016. Epub 2007 Feb 19.*

IDENTIFICATION OF MOLECULAR MARKERS IN PEDIATRIC MEDULLOBLASTOMA

STEFAN PFISTER[1,2], FRANK MENDRZYK[1], ANDREY KORSHUNOV[3], CHRISTOF SCHLAEGER[1], STEFAN JOOS[1], WOLFRAM SCHEURLEN[4], GUIDO REIFENBERGER[5], PETER LICHTER[1], BERNHARD RADLWIMMER[1]

[1] DIVISION MOLECULAR GENETICS, GERMAN CANCER RESEARCH CENTER (DKFZ), HEIDELBERG
[2] DEPARTMENT OF PEDIATRIC ONCOLOGY, HEMATOLOGY & IMMUNOLOGY, UNIVERSITY OF HEIDELBERG
[3] DEPARTMENT OF NEUROPATHOLOGY, BURDENKO NEUROSURGICAL INSTITUTE, RUSSIA
[4] CNOPF'SCHE KINDERKLINIK, NÜRNBERG CHILDREN'S HOSPITAL, NÜRNBERG
[5] DEPARTMENT OF NEUROPATHOLOGY, UNIVERSITY OF DÜSSELDORF

INTRODUCTION

Medulloblastoma is the most common malignant brain tumor and a significant cause of cancer mortality in children. Despite considerable therapeutic advances, prognosis is still poor, with a five-year-survival rate of about 60%. Radiochemotherapy, generally applied to all patients during standard therapy, causes severe long-term cognitive, neurological and endocrinological side effects in many of the survivors, emphasizing the urgent need for molecular markers that could be used for molecular stratification of disease risk.

Applying genome-wide approaches for the detection of genomic imbalances and DNA methylation, we sought to identify molecular markers that could potentially be used to improve the stratification of patients or as new therapeutic targets. Having one of the largest collections of paraffin blocks from medulloblastomas at hand, we were additionally able to investigate potential markers on the protein level by immunohistochemistry, a routine method in clinical practice.

PROJECT STATUS/RESULTS

Array- (or Matrix-) Comparative Genomic Hybridization (CGH) is a microarray-based method that allows for the genome-wide detection of chromosomal imbalances in tumors. In this study, array-CGH was applied to a total of 105 cerebellar medulloblastomas and 10 supratentorial primitive neuroectodermal tumors (stPNETS), a related tumor localized in the cerebrum. Of the medulloblastoma patients, 80 received comparably effective treatments and could therefore be included in statistical analyses.

CDK 6 protein expression is an independent prognostic marker in medulloblastoma

A very interesting finding in the array-CGH study was the presence of high-level genomic amplifications of the *CDK6* locus in three medulloblastomas. This prompted us to investigate CDK6 protein expression by immunohistochemistry on paraffin sections of 189 medulloblastomas. Strong CDK6 expression was detected in 30% of cases and was significantly associated with poor overall survival (Fig. 1; P<0.01)[1].

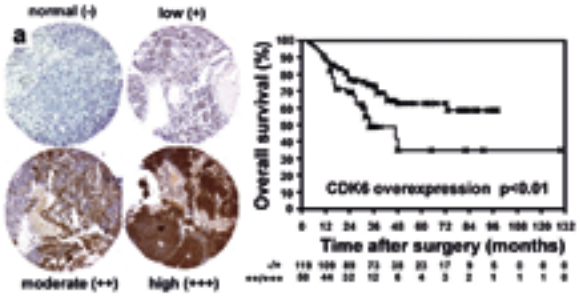


Fig. 1: CDK6 protein expression was analyzed in 189 medulloblastomas by immunohistochemistry. Tumors with normal or low expression (upper panel) and tumors with moderate or high expression (lower panel) were grouped. Overall survival in patients with tumors with moderate or high expression was significantly worse (right).

Role of chromosome 17 aberrations in the pathogenesis of medulloblastoma and stPNETs

The most frequent chromosomal rearrangement detected in medulloblastoma, is a so-called isochromosome 17 i(17q)[1,2]. This term describes a condition in which large parts of one allele of the short arm of chromosome 17 are deleted, whereas the long arm as well as the very centromeric part of the short arm are over-represented (three or more copies) in the tumor genome (Fig. 2).

Detailed genomic analysis of the breakpoints within chromosome band 17p11.2 identified two major breakpoint clusters at megabases 19.0 and 21.3. Since the breakpoints do not directly disrupt a gene and are flanked by DNA sequence repeats, we concluded that effects of i(17q) in medulloblastoma are mediated by gene dosage effects of genes on 17p or 17q rather than by the disruption or deregulation of a specific "breakpoint gene"[2]. Knowing that the important tumor suppressor gene *TP53* is located on 17p, it has long been assumed that the deletion of 17p is the biologically relevant event in i(17q) formation. However, the frequency of mutations of the remaining *TP53* allele was found to be below 10%, suggesting additional genes of interest on chromosome 17. Interestingly, we identified gains of 17q (either in conjunction with 17p deletions or isolated) as an independent prognostic marker for overall survival in medulloblastoma. This indicates the presence of one or several oncogenes on 17q that might be targeted by a gene-dosage effect (Fig. 2 D). Furthermore,

COORDINATOR
Professor Otmar D. Wiestler
German Cancer Research Center, Heidelberg
o.wiestler@dkfz.de

we found that i(17q) only rarely occurs in stPNETs, pointing towards a specific role of this aberration in the pathogenesis of medulloblastoma. In contrast, *CDKN2A* deletions and gains of 19p are frequently encountered in stPNETs and virtually not in medulloblastomas[3]. These data further support the hypothesis of two different tumor entities of embryonal neuroepithelial tumors with characteristic genetic aberrations.

ZIC2 is downregulated by CpG island methylation in medulloblastoma

CpG islands are stretches of genomic DNA enriched in CG-dinucleotides frequently encountered in the promoter region of genes. Associated genes can be silenced by cytosine-methylation within CpGs. To investigate the role of CpG island methylation in medulloblastoma, we developed a new protocol for the genome-wide assessment of DNA methylation, designated aPRIMES (array-based profiling of reference-independent methylation status)[4]. Array-PRIMES is based on the differential restriction and competitive hybridization of methylated and unmethylated DNA by methylation-specific and methylation-sensitive restriction enzymes, respectively (Fig. 3). Applying this protocol to a series of 18 sporadic medulloblastomas, we identified *ZIC2* as a frequent target of aberrant methylation and consecutive downregulation in medulloblastoma[4].

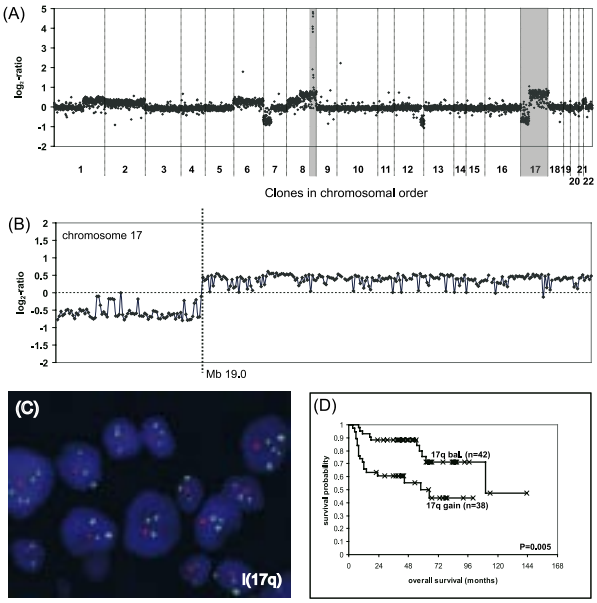


Fig. 2: (A) Typical array-CGH profile of a medulloblastoma including a high-level amplification of c-MYC and a so-called isochromosome 17 i(17q) indicated by gray-shaded areas. (B) Breakpoint clustering on chromosome 17 identified two major breakpoint clusters, the more frequent mapping at megabase 19. (C) FISH analysis confirming the presence of one copy of 17p and 3 or more copies of 17q in a case with i(17q). (D) Gains of 17q were identified as an independent prognostic marker for overall survival in a cohort of 80 patients.

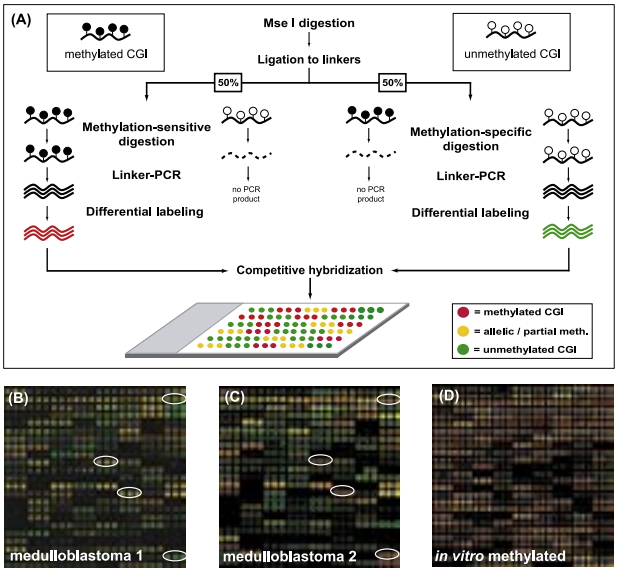


Fig. 3: (A) Flowchart and overview of expected results for methylated and unmethylated CGIs. (B)-(D) Performance of aPRIMES on CGI microarrays. All clones are spotted in triplicate. A representative block is shown for medulloblastoma with low (B) and high (C) methylation, and (D) in vitro-methylated DNA (positive control). Red spots indicate methylated clones, green spots indicate unmethylated clones, yellow spots indicate mixed or allelic methylation.

OUTLOOK

The potential of identified markers to be used for the stratification of patients in prospective clinical trials is currently under investigation.

References

[1] *Mendrzyk F et al. Genomic and protein expression profiling identifies CDK6 as novel independent prognostic marker in medulloblastoma. J Clin Oncol. 2005; 23(34):8853-8862.*

[2] *Mendrzyk F et al. Isochromosome breakpoints on 17p in medulloblastoma are flanked by different classes of DNA sequence repeats. Genes Chromosomes Cancer. 2006; 45(4):401-410.*

[3] *Pfister S et al. Supratentorial primitive neuroectodermal tumors of the central nervous system frequently harbor deletions of the CDKN2A locus and other genomic aberrations distinct from medulloblastomas. Genes Chromosomes Cancer. 2007; 46(9):839-851.*

[4] *Pfister S et al. Array-based profiling of reference-independent methylation status (aPRIMES) identifies frequent promoter methylation and consecutive downregulation of ZIC2 in pediatric medulloblastoma. Nucleic Acids Res. 2007; 35(7):e51.*

THE GENETICS OF MYOCARDIAL INFARCTION: WITH GENOME-WIDE ASSOCIATION STUDIES ON THE FAST TRACK

JEANETTE ERDMANN[1], CHRISTIAN HENGSTENBERG[2], HERIBERT SCHUNKERT[1]

[1] MEDIZINISCHE KLINIK II, UNIVERSITÄT ZU LÜBECK
[2] KLINIK UND POLIKLINIK FÜR INNERE MEDIZIN II, UNIVERSITÄT ZU REGENSBURG

INTRODUCTION

Myocardial infarction (MI) is the most prevalent and often the most fatal manifestation of coronary artery disease (CAD). MI occurs when thrombosis, induced by either a ruptured or eroded atherosclerotic plaque, occludes a coronary artery, leading to necrosis of myocardium.

The prevalence of the disease is very high; ~750 000 Europeans die of the disease each year.

MI is a so-called complex disease, like schizophrenia, hypertension, diabetes or multiple sclerosis, with a strong genetic component. The complexity of this disease is obvious by the many cell types that are involved in the formation of atherosclerotic plaques and by the multiple processes that can affect risk of MI, such as inflammation, plaque calcification, extracellular matrix turnover, apoptosis, and thrombosis.

During the last decade enormous financial and technical efforts have been undertaken to unravel the genetics of MI and CAD. Nevertheless, the results from linkage and association studies were often disappointing because only very few genetic variants could be replicated in further studies. Therefore, until very recently, it was not obvious which genes/gene regions harbor genetic variants responsible for the genetic component of MI.

Epidemiology of CAD/MI

The incidence of MI in the European Union varies from 500 cases/ 100 000/ year in eastern European countries to 125 cases/ 100 000/ year in southern Europe. Overall, the disease is the leading cause of death with 744 000 fatal cases every year in Europe (www.heartstats.org). MI is frequently the first, and potentially fatal, manifestation of CAD. In addition, CAD and MI substantially contribute to morbidity, direct health costs and indirect economic burdens and have important socioeconomic implications like premature retirement and disability.

PROJECT STATUS/RESULTS

The German MI Family Study

Starting in 1997, research teams from the University of Lübeck (Professor Heribert Schunkert and Dr Jeanette Erdmann) and from the University of Regensburg (Professor Christian Hengstenberg) systematically screened more than 200 000 patient records from 17 rehabilitation clinics all over Germany. From this a total of 1400 families with at least two affected siblings were recruited for the German MI Family Study. Additionally, unaffected siblings and spouses who married into the family were included, leading to more than 7500 test persons in the MI registry and making it one of the largest family-based MI registries worldwide.

Studies to unravel the genetics of CAD/MI

Enormous efforts, both technical and financial, have been undertaken during the past decade to explore the genetics of complex diseases like CAD/MI. In many cases this research was funded by the NGFN. For these explorations two different methodological approaches – linkage studies and association studies – are available.

Linkage analysis

By analyzing genetic markers (short tandem repeats) located at short intervals throughout the entire human genome in families or affected siblings, regions can be identified in which a gene causing disease is localized with a high probability. Using this approach, our research team was able to identify the first CAD/MI locus on chromosome 14.[4] To date, more than 25 gene loci for CAD/MI have been identified using linkage information, but the underlying gene is known for only four of these regions.

Association studies

Candidate gene approach

Since the first positive association study for MI published in *Nature* in 1992 by Cambien and co-workers, more than 6200 (PubMed search 01/2007) articles have reported the association of genes with CAD/MI. The most relevant candidate pathways are: lipid metabolism, inflammation, coagulation and fibrinolysis, circulation and vascular growth.

Overall, the expectations raised by early reports of positive associations have been tempered by inconsistent results for almost all the genes studied. All of these contradictory results appear to be due to differences in ethnicity, age (and possibly gender), diagnostic criteria, and environmental factors (e.g., diet and smoking) among the study populations.

COORDINATOR
Professor Hugo Katus
University of Heidelberg
hugo.katus@med.uni-heidelberg.de

The fact that the majority of these studies were underpowered helps to explain why most of these studies failed to be replicated in other populations. These more or less very disappointing association reports for CAD and MI can be circumvented by applying new state-of-the-art methods like genome-wide association studies.

Genome-wide association studies – moving on the fast track

The development of platforms allowing simultaneous typing of hundreds of thousands of polymorphisms now enables a systematic search for the inherited components of complex diseases like CAD/MI.

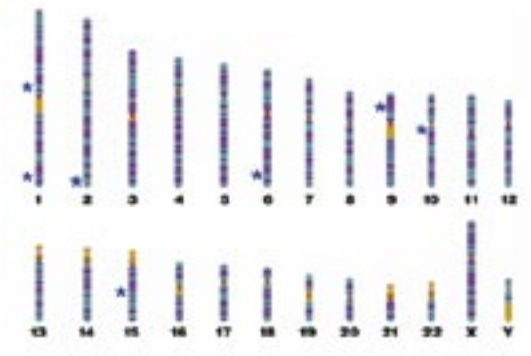


Fig. 1: Schematic overview of the seven gene loci identified in our GWA study[1]

The research teams in Lübeck and Regensburg, together with colleagues in Leicester (Professor Nilesh Samani) and Munich (Professor Thomas Meitinger and Professor H.-Erich Wichmann) were able to conduct the worldwide largest genome-wide association study on CAD/MI. A total of 3000 CAD/MI patients and 4500 controls were studied and several chromosomal regions strongly associated with CAD/MI could be identified.[1] Actually, three of these regions (chromosomes 2q36.3, 6q25.1, and 9p21.3) could be replicated in the German MI Family Study. At present, the main locus on chromosome 9 is the most strongly associated gene region for CAD/MI ever published, because two further GWA studies conducted in Iceland[5] and Canada[6] identified the same loci. This chromosomal region is of special interest because the disease-associated allele is frequent and the increase in risk is almost doubled for homozygote carriers of the risk allele. Unequivocally, these findings demonstrate a major genetic risk variant at this site. The region contains the coding sequences of genes for two cyclin-dependent kinase inhibitors, *CDKN2A* and *CDKN2B*, which play an important role in the regulation of the cell cycle and may be implicated in the pathogenesis of atherosclerosis through their role in TGF-β-induced growth inhibition. Additionally, the combined analysis of both studies identified four loci with a high probability (> 80%) of true association on chromosomes 1p13.3, 1q41, 10q11.21, and 15q22.33[1] (Figure 1). For all of these gene regions further studies are ongoing – first, to fine map these regions and second, to identify the disease-causing variants.

OUTLOOK

At present, the study from Lübeck and Regensburg, conducted with the financial support of the NGFN and the EU, is the most extensive study to unravel the genetic architecture of CAD/MI. Over the middle term these results will help to develop patient-oriented tests for early risk prediction. In the long run we hope to gain more insights into the pathophysiology of the disease and the chance to develop new therapeutics.

References

[1] *Samani NJ et al. WTCCC and the Cardiogenics Consortium. Genomewide association analysis of coronary artery disease. N Engl J Med. 2007 Aug 2; 357(5):443-53.*

[2] *Mayer B et al. Association between arterial pressure and coronary artery calcification. J Hypertens. 2007 Aug; 25(8):1731-8.*

[3] *Fischer M et al. Distinct heritable patterns of angiographic coronary artery disease in families with myocardial infarction. Circulation. 2005 Feb 22; 111(7):855-62.*

[4] *Broeckel U et al. A comprehensive linkage analysis for myocardial infarction and its related risk factors. Nat Genet. 2002 Feb; 30(2):210-4.*

[5] *Helgadottir et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science. 2007 Jun 8; 316(5830):1491-3.*

[6] *McPherson et al. A common allel on chromosome 9 associated with coronary heart disease. Science. 2007 Jun 8; 316(5830):1488-91.*

IDENTIFICATION OF VKORC1 LEADS TO NEW TREATMENT PARADIGMS

JOHANNES OLDENBURG

INSTITUTE OF EXPERIMENTAL HEMATOLOGY AND TRANSFUSION MEDICINE, UNIVERSITY OF BONN

INTRODUCTION

Since 1950, coumarins such as warfarin and phenprocoumon have been widely used as orally administered anticoagulants for therapy and prophylaxis of thromboembolic conditions, such as thrombosis, pulmonary embolism, stroke and myocardial infarction. Coumarins rank among the most prescribed drugs worldwide, and millions of patients begin coumarin therapy each year. Although oral application is convenient for the patient, clinical use of coumarins is complicated by their narrow therapeutic window and a high rate of bleeding complications amounting to 7.5% per year of treatment. It is estimated that in Germany 900 000 patients are treated with coumarins and about 2000 of them suffer from fatal bleeds.

In 2004 our group discovered the VKORC1 protein that is responsible for the recycling of vitamin K and also represents the molecular target of coumarins. The story of VKORC1 began in 1997 with the diagnosis of a single family with deficiency of all vitamin K-dependent coagulation factors (VKCFD2). Homozygosity linkage in this family followed by positional cloning led to the successful identification of the VKORC1 gene.[1]

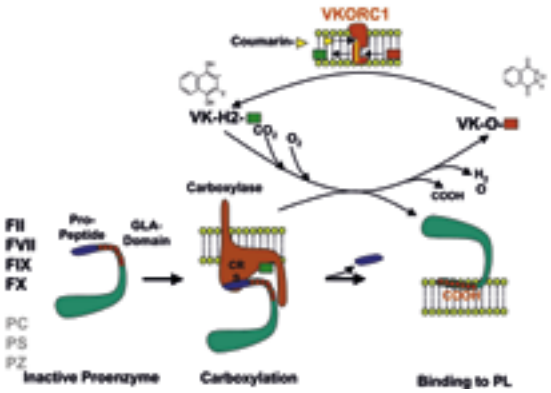


Fig. 1: Vitamin K cycle and principle of oral anticoagulation[4]

After joining the Cardiovascular Diseases Network in the National Genome Research Network, we started to establish an SNP map of the VKORC1 gene. We and other groups soon realized that different VKORC1 haplotypes greatly influence coumarin dosing. Other aspects driven by the identification of VKORC1 were VKORC1 haplotypes as risk factor for cardiovascular disease, more efficient production of recombinant clotting factors VII and IX by coexpression of VKORC1 and the concomitant administration of coumarins and low-dose vitamin K as a new paradigm of oral anticoagulant therapy[4]. The work on VKORC1 began with a clinical case and now, three years after the identification of VKORC1, the translation of the findings back into clinical practice has started.

PROJECT STATUS/RESULTS

Pharmacogenetics of VKORC1

In Caucasians three main haplotypes were identified (distribution: VKORC1*2: 42%, VKORC1*3: 38%, and VKORC1*4: 20%). Haplotype VKORC1*2 includes an SNP in the promoter region (c.1-1639G>A, dbSNP: rs9923231) that was identified as a marker for low-dose warfarin requirement. In human liver tissue expressing the AA (= homozygous VKORC1*2) genotype, VKORC1 mRNA levels as well as enzyme activity were reduced to only 30% of wild type. As a consequence, patients who are homozygous VKORC1*2 only need half of the coumarin dose compared to the other VKORC1 haplotypes. Moreover, in 93% of patients with increased coumarin sensitivity but in none of the patients with partial coumarin resistance, we found homozygosity for the VKORC1*2 haploytpe. Vice versa, non-VKORC1*2 haplotypes were found homozygous in 86% of patients with partial coumarin resistance but in none of the patients with increased coumarin sensitivity.[2]

VKORC1 haplotypes also explain the interethnic differences in coumarin requirement. The VKORC1*2 haplotype frequency distribution of 95% in Chinese, 14% in Africans and 42% in Caucasians is in perfect agreement with the reported low warfarin dose requirement in Chinese and a higher warfarin dose requirement in Africans as compared to Caucasians.[2]

VKORC1 haplotypes as risk factor for cardiovascular disease

Some vitamin K-dependent proteins, especially matrix Gla protein, osteocalcin and Protein S have regulating effects on calcium homoeostasis and cell signaling and therefore have been suggested as playing a role in the pathogenesis of atherosclerosis, myocardial infarction and stroke. Recently, Wang et al. (2006) identified non-VKORC1*2 haplotypes as strong genetic risk factors for coronary heart disease (CHD). These haplotypes nearly doubled the risk for CHD in Chinese. However, in a large cohort of CHD patients from northern Germany (derived from the POPGen biobank, analysis performed at the NGFN technical platform at Kiel) we could not confirm this association.[3] A possible explanation for these

COORDINATOR
Professor Hugo Katus
University of Heidelberg
hugo.katus@med.uni-heidelberg.de

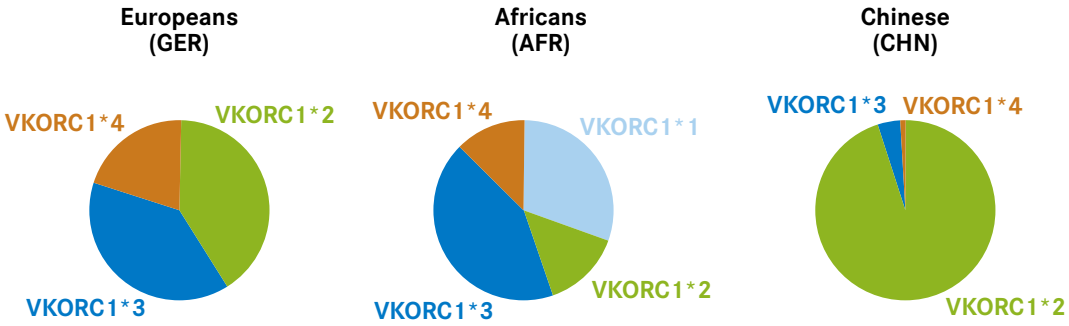


Fig. 2: VKORC1 haplotype distribution in various ethnic groups[2]

discrepant findings might be the ethnic differences in genetic and perhaps environmental predisposition, modifying the polygenic CHD phenotype by interacting with VKORC1 variants and thus conferring disease susceptibility in some populations, but not in others.

Concomitant administration of coumarins and vitamin K

The concomitant application of agonist and antagonist might be a paradox at first glance, but becomes reasonable when viewed in more detail. The difficulty in controlling coumarin therapy has multiple causes: (i) variable vitamin K intake by food, (ii) lack of a vitamin K reservoir, (iii) autocatalytic regulation of gamma carboxylase by vitamin K, and (iv) variable VKOR activities such as the rate-limiting step in vitamin K recycling. With low-dose vitamin K substitution within the range of the WHO-recommended daily intake (80 µg), the above mentioned effects are buffered, thus narrowing down the VKORC1 function almost exclusively to the effect of coumarin.[4] Recent data show that low-dose vitamin K application up to 150 µg of vitamin K per day does not change INR significantly but has the potential to significantly increase the time spent within the therapeutic range (measured by the INR). In the face of an incidence of 0.25% of fatal bleeding complications per treatment year, should it not be worthwhile to test this hypothesis in a prospective randomized study?

OUTLOOK

The identification of VKORC1 has greatly advanced understanding of the vitamin K cycle. The VKORC1 protein has been identified as the molecular target of coumarin anticoagulants and represents the rate-limiting enzyme of the vitamin K cycle and possibly even the sole component of VKOR activity. Mutations and SNPs within the translated and non-translated regions of the *VKORC1* gene cause coumarin resistance and sensitivity. Testing for these variants might be helpful for more rapidly adjusting patient coumarin dosage to an effective and safe level. Additionally, the concomitant oral administration of low-dose vitamin K might change the paradigm of oral anticoagulation treatment, and likely lead to less inter-individual variation in coumarin anticoagulant dosage requirement. These findings – just three years after the discovery of *VKORC1* – have illuminated new perspectives for a safer and more individualized oral anticoagulation therapy standard in the near future.

References

[1] *Rost S et al. Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2. Nature 2004; 427: 537–41.*

[2] *Geisen C et al. VKORC1 haplotypes and their impact on the inter-individual and inter-ethnical variability of oral anticoagulation. Thromb Haemost 2005; 94: 773–9.*

[3] *Watzka M et al. Functional promoter polymorphism in the VKORC1 gene is no major genetic determinant for coronary heart disease in Northern Germans. Thromb Haemost 2007; 97: 998–1002.*

[4] *Oldenburg J et al. VKORC1: molecular target of coumarins. J Thromb Haemost. 2007;5 Suppl 1:1-6.*

GENETIC DISSECTION OF PARKINSON'S DISEASE: GENES AND PATHWAYS

THOMAS GASSER

HERTIE INSTITUTE OF CLINICAL BRAIN RESEARCH, UNIVERSITY OF TÜBINGEN

INTRODUCTION

Parkinson's disease (PD) is a syndrome characterized by variable combinations of akinesia, rigidity, tremor and postural instability. These symptoms are caused by a degeneration of dopaminergic neurons of the substantia nigra, leading to a deficiency of dopamine in their striatal projection areas.

Genetic research over the past years, in particular the mapping and cloning of a number of genes which cause – when mutated – monogenically inherited forms of the disorder, has shown that PD is actually not a disease entity, but rather a heterogeneous group of diseases associated with a spectrum of clinical and pathological changes.

PROJECT STATUS/RESULTS

LRRK2

In two large families, linkage to a novel locus for autosomal dominantly inherited Parkinson's disease (PARK8) on chromosome 12 was confirmed by two research teams funded in NGFN-2, in cooperation with researchers at the Mayo Clinic in Jacksonville, FL, USA.[1] Extensive candidate gene sequencing revealed disease-causing mutations in a large gene called leucine-rich repeat kinase 2 (LRRK2) in these and three additional families.[2] This finding turned out to be of far-reaching importance, as LRRK2-associated PD is remarkable for several reasons. First and foremost, mutations in the LRRK2 gene are clearly the most common cause of inherited parkinsonism discovered so far. In a number of studies across several different populations, between 5 to 15% of dominant families carry mutations in LRRK2. The single most common mutation alone, G2019S, is responsible for familial PD in up to 7% in different Caucasian populations, but also in 1 to 2% of sporadic cases. Surprisingly high prevalence rates of up to 40% were found in genetically more isolated populations, such as the Ashkenazi Jewish or the North African Berber Arab populations, both in sporadic and familial cases.

Another common variant, G2385R, was found in up to 9% of Chinese patients with PD.

Clinical signs and symptoms of LRRK2-related disease closely resemble typical sporadic PD. This is also true for age of onset, which is on average in the late fifties. However, age at onset as well as severity of the disease may be highly variable, even within families.[3]

Pathological changes in patients with LRRK2 mutations are consistent with typical Lewy body Parkinson's disease in most cases reported so far, but also include diffuse Lewy body disease, nigral degeneration without distinctive histopathology and, rarely, even progressive supranuclear palsy-like tau aggregation. LRRK2 mutations may therefore be an upstream event in the cascade leading to neurodegeneration with different pathologies. By sequence homology, LRRK2 can be assigned to the group of recently identified ROCO proteins and contains a protein kinase domain of the MAPKKK class, suggesting a role in intracellular signaling pathways. Although the natural substrate of LRRK2 is unknown, cell culture studies suggest that pathogenic mutations seem to be associated with an increase, rather than a loss, of kinase activity.

The discovery of relatively common mutations in a gene coding for a potentially "drugable" target has considerable implications for future research. Screening for drugs modulating LRRK2 kinase activity is already being undertaken by numerous drug companies. Efficacy assays and biomarkers will be developed that will allow to monitor drug effects, and populations of patients and presymptomatic carriers of LRRK2 mutations are being identified as potential candidates for the first truly causative or even preventive treatment of this devastating disease.

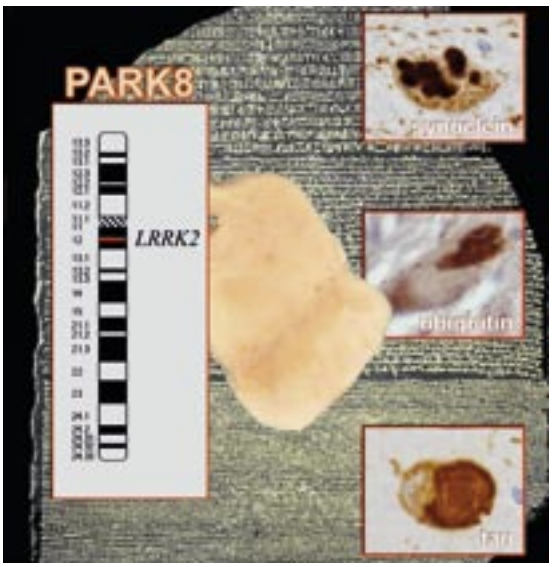


Fig. 1: The disease caused by mutations in the LRRK2 gene (PARK8) is associated with diverse pathologies. It has therefore been called the "Rosetta Stone" of Parkinson's disease.

COORDINATOR
Professor Peter Propping
University of Bonn
propping@uni-bonn.de

α-synuclein and its role in sporadic Parkinson's disease

The first PD gene (PARK1) was mapped to chromosome 4 in a large family with dominantly inherited PD and Lewy body pathology. The mutated gene was later identified to be α-synuclein (SNCA). Over the years only two additional point mutations have been recognized, each in a large, dominant family, reflecting the low prevalence, but high penetrance of these mutations.

However, the identification of SNCA mutations led to the discovery that the encoded protein α-synuclein is the major fibrillar component of the Lewy body, the characteristic intraneuronal inclusion, found in familial as well as sporadic cases of PD.

The currently favored hypothesis states that the amino acid changes in α-SYN lead to an increased tendency of the protein to form oligomers and later on fibrillar aggregates, although the precise relationship between aggregation, cellular dysfunction and cell death is still unknown. Many studies favor the hypothesis that the mature aggregates (Lewy bodies), are actually not the toxic moiety, but rather an attempt of the cell to form small toxic oligomers.

A further direct link between α-synuclein and PD is supported by the recent discovery that not only point mutations, but also multiplications of the wild-type sequence of the α-synuclein gene (duplications and triplications) cause autosomal dominant parkinsonism with or without dementia with αSYN inclusions, indicating that a mere increase in αSYN levels can be directly toxic to neurons. Analysis of one family by a research team working within NGFN-2 identified patients with both duplications and triplications in the SNCA-gene. Patients with duplications have late onset disease, resembling the typical sporadic disorder, while those with a triplication have early onset of symptoms, resembling the more severe phenotype of diffuse Lewy body disease. This finding proves a direct effect of α-synuclein dosage on the disease manifestation.[4]

If an increase of the cellular α-synuclein protein load by 50 to 100% (as found in gene duplications and triplications, respectively) causes familial PD with high penetrance, it can be suspected that alterations in regulatory regions of the gene, which might result in just slightly increased gene expression, may be associated with a higher risk to develop the disease. In fact, there is increasing evidence from multiple studies that this is indeed the case. Several case-control studies found a complex polymorphic dinucleotide repeat polymorphism (NACP-Rep1) located close to the promoter region to be associated with sporadic PD. In a more comprehensive approach, Müller et al. of NGFN-2 first analyzed the entire SNCA gene with more than 50 single-nucleotide repeats (SNPs) and found a strong association of the disease with a haplotype block comprising exons 5 and 6 and the 5'-UTR.[5] This finding has been confirmed in a recent Japanese study. It is of course likely that more than one mechanism regulates gene expression. Preliminary evidence suggests that these variants influence SNCA-expression by transcriptional regulation and/or by altering mRNA stability. If confirmed, pharmacologic manipulation of SNCA expression may be a possible therapeutic or preventive strategy in susceptible individuals.

OUTLOOK

Work within the NGFN has defined promising targets for pharmacological modification of the disease course of PD, and maybe even for its causative treatment or prevention. Further studies are planned to validate these findings in a number of animal and cellular models and to identify lead compounds and biomarkers that will allow the translation of these findings into tangible progress for patients with PD.

References

[1] *Zimprich A et al. The PARK8 Locus in Autosomal Dominant Parkinsonism: Confirmation of Linkage and Further Delineation of the Disease-Containing Interval. Am J Hum Genet 2004;74(1):11-19.*

[2] *Zimprich A et al. Mutations in LRRK2 Cause Autosomal-Dominant Parkinsonism with Pleomorphic Pathology. Neuron 2004;44(4):601-607.*

[3] *Berg D et al. Type and frequency of mutations in the LRRK2 gene in familial and sporadic Parkinson's disease. Brain 2005;128:3000-3011.*

[4] *Fuchs J et al. Phenotypic variation in a large Swedish pedigree due to SNCA duplication and triplication. Neurology 2007;68(12):916-922.*

[5] *Mueller JC et al. Multiple regions of alpha-synuclein are associated with Parkinson's disease. Ann Neurol 2005;57(4):535-541.*

GENOME-WIDE ASSOCIATION STUDY SUGGESTS MANY GENES UNDERLIE BIPOLAR AFFECTIVE DISORDER

PETER PROPPING, JOHANNES SCHUMACHER
INSTITUTE OF HUMAN GENETICS, UNIVERSITY OF BONN

INTRODUCTION

A recent genome-wide association (GWA) study supports the notion that many genetic variants each contribute a little to the risk of developing bipolar affective disorder (BPAD).

BPAD is a common psychiatric disorder posing a severe burden on the individual and the society. The core feature of BPAD is a pathological disturbance in mood ranging from extreme elation or mania to severe depression. Typically, BPAD is an episodic illness, usually with full recovery between episodes.

Family, twin, and adoption studies provide strong evidence of the importance of genes in the predisposition to BPAD. It has been consistently shown that the BPAD risk is increased approximately 7-fold in first-degree relatives of affected test persons. The estimated heritability is hereby in the range of approximately 85%. The pattern of inheritance is complex, reflecting the actions and interactions of multiple genetic and environmental factors, which has led to difficulties in mapping individual risk genes so far.

Genome-wide association (GWA) studies are a new approach to the genetics of complex diseases. By using dense maps of single nucleotide polymorphisms (SNPs) and high-throughput genotyping methods, it is now possible to test essentially every gene and most intergenic regions for evidence of association with a disease. Therefore, this approach offers a powerful alternative to genetic linkage studies, which are often underpowered to detect genes contributing to complex phenotypes, and to candidate gene association studies, which are biased by the choice of genes included and may not reveal novel pathways to disease. In view of the advantages of GWA methods, major efforts are now underway to detect genes contributing to complex diseases. However, one disadvantage of GWA studies includes the costs, which are substantial in sufficiently sized case-control or family-based samples. DNA pooling is one solution to this dilemma. By combining individual DNA samples in pools, it is possible to perform GWA studies at a fraction of the cost of individual genotyping. Although haplotype-based analysis is not feasible in DNA pooling studies, it has been shown that this approach – along with robust genotyping platforms and novel, validated statistical methods – allow for remarkably accurate estimates of allele frequency differences in cases and controls.

PROJECT DESIGN/RESULTS

In a recent collaborative study that appeared online in the journal Molecular Psychiatry on May 8, research teams from the United States (National Institute of Mental Health, Bethesda) and Germany (Institute of Human Genetics and Department of Genomics, University of Bonn, and the Central Institute of Mental Health, University of Heidelberg) searched for BPAD risk genes by performing a GWA study using over 550 000 SNPs (Illumina HumanHap550 chip) and a DNA pooling approach. The US site recruited 461 unrelated individuals with BPAD who had affected siblings, and 563 healthy controls; all had solely European ancestry. As a check against the false positives that can compromise GWA studies, the researchers replicated their findings in a sample recruited by the German site. This sample consists of 772 individuals with BPAD and 876 controls with no history of BPAD.

Within the GWA study, which was followed by genotyping the most interesting findings individually, the research groups detected 88 SNPs, representing 80 different genes, meeting the prior criteria for replication in both samples. Effect sizes were modest: no single SNP of large effect was detected. The highest odds ratio (OR) based on individual genotyping was 1.67 (95% CI 1.32-2.13) and seen for an SNP in the *SORCS2* gene. *SORCS2* maps to a region on chromosome 4p that has been widely linked to BPAD and encodes a VPS10 domain-containing receptor prominently expressed in the developing brain. Of 37 further SNPs selected for individual genotyping, the strongest association signal was detected at a marker within the first intron of a gene called *DGKH* ($P=1.5 \times 10^{-8}$), experiment-wide $P<0.01$, OR=1.59). This gene encodes a key protein in the lithium-sensitive phosphatidyl inositol pathway. Lithium represents the medication of first choice for the long-term prevention of BPAD. Although the research groups had no lithium-response data available in the studied samples, it may be of particular interest for future studies to test this finding in samples where lithium response data are available.

Two other strong replicated signals were found in the *NXN* and *VGCNL1* genes. *NXN* encodes the protein nucleoredoxin, which inhibits Wnt-catenin signaling by binding to the PDZ domain of the dishevelled protein. In addition to the phosphatidyl inositol pathway, the Wnt pathway has also been hypothesized to play an important role in the mechanism of

COORDINATOR
Professor Peter Propping
University of Bonn
propping@uni-bonn.de

action of mood-stabilizing medications. *VGCNL1* encodes a voltage-gated ion channel highly expressed in the brain. Its *Drosophila* homolog, narrow abdomen (na), regulates circadian rhythms, suggesting that this gene may play a role in the circadian disruption that is a hallmark of BPAD.

CONCLUSION AND OUTLOOK

In conclusion, this first GWA study of BPAD has detected reproducible association with markers in several genes. None of the replicated genes confers a large risk of disease and no one gene appears to be necessary or sufficient for disease. These findings imply that major gene effects in BPAD are very uncommon, if they exist at all. Future studies may benefit from multi-locus approaches that embrace the genetic heterogeneity of BPAD. In addition, the DNA pooling approach used in the NIMH and German study demonstrates how GWA studies can make them more manageable.

Reference

[1] *Baum AE et al. A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. Molecular Psychiatry. May 8, 2007. Advance online publication.*

IDENTIFICATION OF CANDIDATE GENES FOR HUMAN OBESITY

A. SCHÜRMANN, H. AL-HASANI, H.G. JOOST

GERMAN INSTITUTE OF HUMAN NUTRITION (DIFE) POTSDAM-REHBRÜCKE, DEPARTMENT PHARMACOLOGY, NUTHETAL

INTRODUCTION

Obesity is a chronic disease characterized by excessive accumulation of body fat mass. It is based on a polygenic predisposition (dependent on multiple gene variants) in combination with an adipogenic lifestyle (hypercaloric diet, low physical activity), both of which result in a positive energy balance. Obesity induces several secondary complications (hypertension, cardiovascular diseases, insulin resistance and type-2 diabetes), summarized as the *metabolic syndrome*. The aim of our project is to identify and characterize obesity genes and their interaction with dietary factors to elucidate the molecular basis and the pathophysiology of obesity. So far, genome-wide association studies in humans have led to the identification of only a limited number of candidate genes for the metabolic syndrome. Because of the complexity and limitations of human studies, our strategy mainly involves mouse genetics to identify candidate genes for human obesity. Our mouse model, the New Zealand obese (NZO) mouse (Fig. 1), closely resembles the human polygenic metabolic syndrome, and is useful for analyzing gene-gene and gene-environment interactions.

Fig. 1: Photograph of C57BL/6 (left) and NZO (right) mice.

PROJECT STATUS/RESULTS

To identify relevant chromosomal regions within the genome, we generated several outcross populations of mice that either are susceptible (NZO) or resistant (SJL, NZB and C57BL/6) to the disease. In other words, we generated populations with mixed genomes and correlated the genotype and phenotype of the animals. Genome-wide linkage analysis identified several QTL (quantitative trait loci) that are associated with the metabolic syndrome (obesity, insulin resistance, dyslipidemia, glucose intolerance). These QTL usually span 10-20 cM (centi Morgan), which is equivalent to 20-40 Mbp (million base pairs) or 200-500 genes. To identify obesity genes, we followed different strategies in parallel. First, we identified polymorphic and non-polymorphic segments (haplotypes) by analyzing genetic markers (SNPs) of the parental mouse strains. Because most of the genetic variation in these inbred mice is ancestral, regions identical by descent (IBD) are unlikely to contain disease candidate genes and thus can be excluded from the analysis. In fact, our data indicate that only 20-30% of the NZB and NZO genomes consist of polymorphic haplotypes (size 2-10 Mbp), allowing exclusion of 70-80% of the NZO genome. Second, for characterizing relevant regions in the genome, we generate recombinant congenic strains (RCS) in which selected chromosomal segments are transferred to a control strain via successive backcrossing. The analysis of these RCS may allow a refinement of the critical regions to the level of single candidate genes. Third, cross-species comparisons of genomic segments that are associated with the disease are used for identification of candidate genes.

By utilizing these strategies, we identified the cholesterol transporter ABCG1 as an obesity gene. Genetic manipulation of the genes in the fruit fly *Drosophila melanogaster* increased the triglyceride level. In the NZO mouse genome, the *Abcg1* gene is located within a QTL associated with elevated body fat mass. An NZO-specific insertion within the promoter of the *Abcg1* gene is responsible for a higher expression in adipose tissues of NZO mice. Backcross mice carrying the NZO-specific variation exhibited a 2.9 g higher body weight than the controls in week 12 ($p=0.018$). Targeted disruption of *Abcg1* in mice resulted in reduced body weight and fat mass and protected against diet-induced obesity. This effect was due to increased body temperature, locomotor activity and energy expenditure.[3]

The NmuR2$^{V190M/I202M}$ variant of the neuromedin U receptor 2 *(nmur2)* participates in the hyperphagia of the NZO mouse. Neuromedin U (NmU) is a neuropeptide which has recently been shown to inhibit food intake and increases energy expenditure when administered centrally. We demonstrated that the appetite-inhibiting (anorexigenic) effect of NmU in lean C57BL/6J mice was mainly due to a reduction of the number of meals and that this effect was markedly reduced in NZO mice (15% vs. 60% reduction in C57BL/6J). In addition, *in vitro* studies indicated that binding of NmU to the

COORDINATOR
Professor Johannes Hebebrand
University of Duisburg-Essen
johannes.hebebrand@uni-duisburg-essen.de

NmuR2$^{V190M/I202M}$ variant and signal transduction were markedly reduced. Mice of a F2[NZOxC57BL/6] outcross population exhibited significantly ($p<0.03$) higher body weight (59.4±8.1 g) carrying the NZO variant of *nmur2* than mice with the corresponding C57BL/6 allele (56.2± 6.3 g).[5]

Finally, an obesity repressor originating from the lean SJL strain which carries a loss-of-function mutation of an insulin signaling protein was discovered. One major QTL for high-fat diet (HFD)-induced obesity (*Nob1*; LOD score 7.9) was localized on chromosome 5. Additional crossbreeding experiments indicated that *Nob1* represents an obesity suppressor from the SJL strain. In a combined approach of gene expression analysis and sequencing, we identified a SJL-specific 7 bp deletion in the *Nob1* gene which results in a truncated protein lacking a functional relevant domain. Introgression of the *Nob1* segment of the SJL chromosome 5 into a mixed NZO/C57BL/6J background markedly reduces body weight and blood glucose levels. *Nob1* has recently been linked to human obesity. It is related to an insulin-signaling protein, and is highly expressed in skeletal muscle, heart, pancreas, kidney and hypothalamus. Our data strongly suggest that deletion of Nob1 suppresses high-fat diet-induced obesity caused by the polygenic NZO background. Thus, *Nob1* might be involved in a novel pathway that regulates energy homeostasis in both mice and humans. Since this gene is a promising target for the search of anti-obesity agents, a patent application was filed[1] within the NGFN-2 funding period.

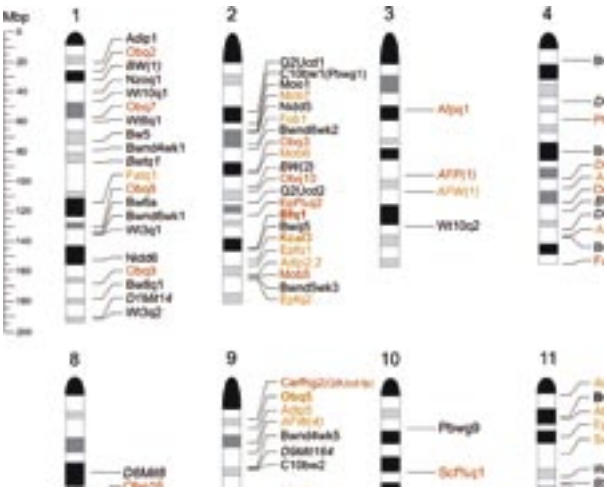


Fig. 2: Physical map of mouse QTL for obesity-related traits. The peak positions for 279 QTL from 34 published genome-wide scans for body weight and weight and body fat were localized to the mouse genome.[6] An interactive physical map of the QTL is available online at http://www.obesitygenes.org.

In addition, we identified several distinct chromosomal segments that are significantly associated with body weight and adiposity in mice (LOD scores 14.8-21.8) in a meta-analysis of 279 non-redundant QTL for obesity-related traits (Fig. 2).[6]

OUTLOOK

The aim of our ongoing study is to develop assays for high-throughput drug screens and to characterize gene-gene and gene-diet interactions. Genetically modified mouse models which have been generated in NGFN-2 and RCS will be used to investigate gene-gene and gene-environment interactions. Preliminary results obtained with our RCS demonstrate elevated body weight and fat mass in mice carrying multiple-obesity QTL, indicating gene-dose dependent effects on body composition. Furthermore, we will focus on analyzing the effect of different nutrient qualities (dietary lipids, carbohydrates) on body weight, body composition, insulin sensitivity and glucose homeostasis in these mice as described earlier.[4]

References

[1] *Al-Hasani, European Patent Application (07007072.7; 04.04.2007).*

[2] *Al-Hasani H et al. Best Pract Res Clin Endocrinol Metab.19, 589-603 (2005).*

[3] *Buchmann J et al. Endocrinology 148, 1561-1573 (2007).*

[4] *Jürgens HS et al. Diabetologia 50, 1481-1489 (2007).*

[5] *Schmolz K et al. Obesity and Metabolism, 3, 28-37 (2007).*

[6] *Wuschke S et al. Int J Obes (Lond). 31, 829-841 (2007).*

OBESITY AND RELATED DISORDERS – FIRST POLYGENES FOR OBESITY

JOHANNES HEBEBRAND, TIMO D. MÜLLER, SUSANN FRIEDEL, GÜNTER BRÖNNER, ANKE HINNEY

DEPARTMENT OF CHILD AND ADOLESCENT PSYCHIATRY AND PSYCHOTHERAPY, UNIVERSITY OF DUISBURG-ESSEN

INTRODUCTION

The past 30 years have witnessed a dramatic increase in prevalence rates for obesity throughout most parts of the world; the obesity epidemic shows no signs of abating.[11] The WHO has declared overweight as one of the top ten risk conditions in the world and one of the top five in developed nations; worldwide more than 1.6 billion adults are overweight and over 400 million are obese (World Health Report, 2005).

Aim of the Project

The major scientific goal of the NGFN-2 consortium 'Obesity and Related Disorders' is the identification of genes/alleles predisposing to obesity and their subsequent evaluation in epidemiological, developmental, clinical and therapeutic terms.

PROJECT STATUS/RESULTS

The central melanocortigenic system controls body weight regulation and energy homeostasis mainly through the hypothalamic melanocortin 4 receptor (MC4R). More than 90 functionally relevant mutations in the *MC4R* gene have so far been described; approximately 1-6% of extremely obese children, adolescents and adults harbor such mutations[12, 13]; many of these mutations result in either a reduced or a total loss of function.[18, 19] We have helped to clarify the role of the *MC4R* by showing that male and female mutation carriers are on average 4.5 and 9.5 kg/m² heavier than their wild-type family members[2], thus providing the first effect-size estimation for such mutations. We also demonstrated that within the general population, carriers of functionally relevant mutations are not necessarily obese.[12] Additionally, we detected the first truly validated polygenic variant (Val103Ile) in the *MC4R*.[4] A meta-analysis, including 7713 individuals from 12 independent studies, provided solid evidence for a negative association of the Ile103 allele with obesity (odds ratio 0.69; 95% confidence interval 0.50-0.96; p=0.03.[4] The Ile103 variant entails a mean BMI reduction of 0.5 kg/m². An effect of the variant was detected in both adolescents and adults[4, 8]; no independent effect was detected on waist circumference.[6] Recently, a second meta-analysis including 29 563 individuals (in which our data were also incorporated), confirmed the negative association of the Ile103 allele with obesity; carriers of the Ile103 allele had an 18% decreased risk of developing obesity.[20] A functional study revealed that the Ile103 variant increases MC4R receptor signaling, thus providing functional evidence for the protective role of this variant in obesity.[19] We further analyzed the Ile103 allele and several cardiovascular parameters (e.g. serum lipids, blood pressure) in fasting patients with coronary heart disease (CHD), revealing reduced serum triglyceride levels in carriers of the Ile103 allele among these patients.[1]

Within an international cooperation framework, we participated in the detection of a second polygenic variant (single nucleotide polymorphism; SNP). The SNP is located approximately 10kb upstream of the insulin induced gene 2 (INSIG2), which plays a key role in fat and cholesterol metabolism. The SNP was identified by a genome-wide association study (GWA) that was based on a 100K Affymetrix scan on 694 participants of the Framingham Heart Study.[10] The SNP near the *INSIG2* was associated with an increased risk of obesity (OR: 1.33, 95% confidence interval: 1.2–1.48; mean BMI increment 0.8 kg/m²). The association of the C allele with obesity was independently confirmed in different ethnicities including 368 of our obesity trios (extremely obese child and both biological parents, p=0.0017) and other samples from our NGFN-2 network (KORA S4; p=0.008).[10] It is noteworthy that the association of this variant with obesity was most significant in our obesity trios, thus substantiating the power of our approach to detect true polygenic variants. Later on, the association of the respective variant with obesity has been confirmed in independent samples including additional samples of members from our NGFN-2 network.[15] However, a positive association of this SNP with obesity has not been detected in all studies.[14, 3, 16] As of today, analyses based on the *INSIG2* SNP (case/control and family-based designs) have been performed in 47 000 individuals; a meta-analysis is currently ongoing within our NGFN-2 network.[7]

Early in 2007, the first GWA (case/control design) for type 2 diabetes mellitus (T2DM) in French patients was published (392 935 SNP Illumina assay).[17] The previously detected[5] and subsequently solidly replicated association of T2DM to SNPs in the transcription factor 7-like 2 gene (TCF7L2) was again confirmed. A minor role of the *TCF7L2* in body weight regulation has recently been detected[9]; we contributed to the respective study by supporting that this minute effect is detectable in males only. Interestingly, the haplotype which predisposes to overweight was distinct from the T2DM haplotype.

COORDINATOR
Professor Johannes Hebebrand
University of Duisburg-Essen
johannes.hebebrand@uni-duisburg-essen.de

OUTLOOK

Due to the rapid progress achieved via GWA, we expect to identify novel polygenes involved in obesity within the next two years. Using GWA data based on more than 3100 subjects - including extremely obese and healthy lean individuals – we will identify novel polygenes underlying body weight regulation and energy homeostasis. We will validate the respective findings and assess the relevance of detected alleles in relationship to different developmental stages in epidemiological samples. Even though success will depend on the mean effect sizes of the gene variants, our samples have shown impressive power in the detection and replication of true polygenes. We will also determine the significance of such polygenes for selected disorders associated with obesity; such studies have proven very valuable for co-operations with other networks/research groups. Furthermore, we will participate in functional studies which will focus on the implications of the detected genetic variation using *in vitro* and *in vivo* models.

References

[1] *Brönner et al. J Clin Endocrinol Metab 2006;91:535-8.*

[2] *Dempfle et al. J Med Genet 2004;41:795-800.*

[3] *Dina et al. Science 2007;315:187.*

[4] *Geller et al. Am J Hum Genet 2004;74:572-81.*

[5] *Grant et al. Nat Genetics 2006;38:320-3.*

[6] *Heid et al. Obesity, in press.*

[7] *Heid et al., in preparation.*

[8] *Heid/Vollmert et al. J Med Genetics 2005;42:e21.*

[9] *Helgason et al. Nat Genetics 2007;39:218-25.*

[10] *Herbert A et al. Science 2006;312:279-83.*

[11] *Hill et al. Science 2003;299:853-5.*

[12] *Hinney et al. J Clin Endocrinol Metab 2006;91:1761-9.*

[13] *Hinney et al. J Clin Endocrinol Metab 2003;88:4258-67; 2006.*

[14] *Loos et al. Science 2007;315:187.*

[15] *Lyon et al. PLoS Genet 2007;3:e61.*

[16] *Rosskopf et al. Science 2007;315:187.*

[17] *Sladek et al. Nature 2007;445:881-5.*

[18] *Tao et al. J Clin Endocrinol Metab 2005;90:5632-8.*

[19] *Xiang et al. Biochemistry 2006;45:7277-88.*

[20] *Young et al. Int J Obes 2007; Epub ahead of print.*

GENE REGULATION IN STROKE

MARKUS SCHWANINGER[1], ARMIN SCHNEIDER[2], OLIVER HERRMANN[1], IOANA INTA[1], MARION N. SCHÖLZKE[1]

[1] DEPARTMENT OF NEUROLOGY AND PHARMACOLOGY, UNIVERSITY OF HEIDELBERG
[2] SYGNIS PHARMA AG, HEIDELBERG

INTRODUCTION

Stroke triggers a multifaceted response in the brain that includes both immediate changes (e.g. ion fluxes) and delayed processes (e.g. inflammation, apoptosis, and neurodegeneration). To alleviate the neurological consequences of stroke and to enhance recovery, we are investigating the molecular pathophysiology of stroke. For practical reasons, delayed events are appealing because they fall into the time window that is accessible to therapeutic interventions. Delayed events often rely on gene expression. We therefore set out to analyze gene expression and its regulation in a mouse model of stroke.

PROJECT STATUS/RESULTS

To analyze gene expression in stroke we have employed various strategies. First, we used open approaches to profile gene expression. At several time points after stroke we isolated RNA from the ischemic brain and investigated the gene expression pattern using different approaches (MPSS, RMDD, microarrays). This led to the identification of multiple genes that are induced in cerebral ischemia. Next, we explored the functional significance of regulated genes (see "TWEAK signaling in stroke"). This indicated the involvement of signaling cascades that are triggered by TWEAK. Indeed, we found evidence that the kinase IKK downstream of TWEAK contributes to ischemic neurodegeneration (see "The kinase IKK promotes neurodegeneration in stroke"). Although this strategy based on gene expression in whole brain led to the discovery of novel mechanisms of acute neurodegeneration in stroke, it is not useful for the investigation of rare cell types in the ischemic brain that might still be of utmost importance. To avoid this pitfall, endeavors are ongoing to analyze gene expression in rare cell populations (see "Outlook").

TWEAK signaling in stroke

While profiling gene expression in stroked brains using MPSS technology, we identified TWEAK as an upregulated gene in cerebral ischemia. TWEAK is a member of the TNF family of cytokines. It binds to the membrane receptor Fn14, which belongs to the TNF receptor superfamily. In addition, the expression of Fn14 was enhanced after cerebral ischemia. A neutralizing antibody against TWEAK reduced the infarct size, suggesting that TWEAK is involved in the molecular mechanism of acute neurodegeneration. Indeed, TWEAK induced apoptosis in primary cortical neurons. Interestingly, the receptor Fn14 lacks typical death domains that have been shown to trigger apoptosis in the case of other members of the TNF receptor superfamily such as FAS. However, Fn14 contains binding sites for TRAFs that function as adapter proteins and link receptors to several intracellular signaling pathways. We could show that TWEAK stimulates the transcription factor NF-κB in primary cortical neurons. NF-κB has been shown to exert neuroprotective properties in many experimental paradigms. Therefore, it was a surprise to find that NF-κB promotes apoptosis in the case of TWEAK-induced neuronal cell death. Other research groups have confirmed our finding that TWEAK is detrimental in cerebral ischemia. As an extracellular mediator, TWEAK is a potentially "drugable" target in stroke. The neutralizing anti-TNF antibody infliximab or the soluble TNF receptor etanercept may be cited as examples of how to inhibit cytokine signaling. These compounds are routinely used in the clinic for severe forms of rheumatoid arthritis or inflammatory bowel disease. For stroke a single application of a biological compound is likely to be sufficient, thus reducing safety concerns.

The kinase IKK promotes neurodegeneration in stroke

TWEAK activates the transcription factor NF-κB, which plays a pivotal role in the regulation of inflammation and apoptosis in many cell types. Activation of the preformed NF-κB complex is mediated by the protein kinase IKK. Therefore, we investigated the function of IKK in cerebral ischemia. We obtained evidence that IKK is activated in cerebral ischemia in neurons. To inhibit IKK in neurons, we joined forces with a research group in Ulm headed by Professor Brunhilde Wirth and another research group in Cologne headed by Professor Manolis Pasparakis. The colleagues in Ulm had generated mice that express either an inhibitor of IKK or an active form in neurons. In these animals we found that inhibition reduced the infarct size, whereas activation of IKK enlarged the infarct volume. The findings were confirmed by studying conditional knockout mice generated in Cologne by the Pasparakis group. Neuronal deficiency of IKK significantly reduced ischemic brain damage. Interestingly, the role of IKK in cerebral ischemia can be exploited by pharmacological tools. Several pharmaceutical companies have developed small molecule inhibitors of IKK. So far we have tested one compound, BMS345541, and found a pronounced neuroprotective effect. BMS345541 was active up to 4.5 h after onset of stroke.

In summary, starting with gene expression profiling, we have identified a signaling cascade that promotes neuronal cell death in cerebral ischemia. This cascade offers numerous steps accessible to pharmacological intervention.

COORDINATOR
Professor Peter Seeburg
Max Planck Institute for Medical Research, Heidelberg
seeburg@mpimf-heidelberg.mpg.de

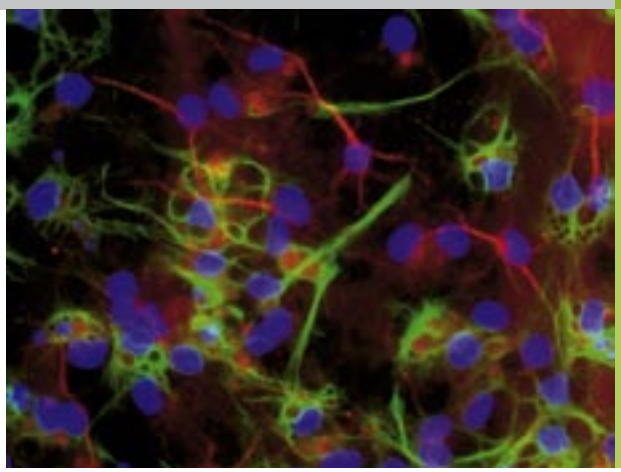


Fig. 1: Differentiated neurosphere cultures. Cells were isolated from the mouse subventricular zone and were treated for 7 days with serum in vitro. Red, neuronal marker MAP2. Green, astrocyte marker GFAP. Blue, nuclear stain DAPI.

OUTLOOK

Our previous approaches have ignored gene expression in rare cell populations such as neural progenitor cells. In recent years it has been recognized that neurogenesis persists into adulthood and is clearly stimulated by stroke. Furthermore, growth factors are able to stimulate neurogenesis *in vivo* and *in vitro*. To understand the ischemia-specific regulation of neurogenesis including mechanisms for migration of progenitors to the ischemic area, it will be important to characterize gene profiles in neural progenitor cells. This goal can be achieved *in vitro* (Fig. 1) or *in vivo*. We are currently establishing a protocol to isolate neural progenitor cells from the ischemic brain for further analysis.

References

[1] *Schneider A et al. (2004) Restriction-mediated differential display (RMDD) indentifies pip92 as a pro-apoptotic gene product induced during focal cerebral ischemia. J Cereb Blood Flow Metab 24: 224-236.*

[2] *Potrovita I et al. (2004) Tumor Necrosis Factor-like Weak Inducer of Apoptosis-induced neurodegeneration. J Neurosci 24: 8237-8244.*

[3] *Herrmann O et al. (2005) IKK mediates ischemia-induced neuronal death. Nat Med 11: 1322-1329.*

[4] *Schölzke MN et al. (2007) Transcriptional regulation of neurogenesis: potential mechanisms in cerebral ischemia. J Mol Med 85: 577-588.*

NEURAL STEM CELLS IN LOW OXYGEN – UN AMOUR POSSIBLE?

MARTIN H. MAURER, WOLFGANG KUSCHINSKY
DEPARTMENT OF PHYSIOLOGY AND PATHOPHYSIOLOGY, UNIVERSITY OF HEIDELBERG

INTRODUCTION

Neural stem cells exist in the adult human mammalian brain throughout the entire lifetime. These cells are capable of self-renewal, proliferation, and differentiation into brain cells, most predominantly into neurons and astrocytes. Neural stem cells in the adult mammalian brain are restricted to defined areas of ongoing neurogenesis, including the dentate gyrus of the hippocampus and the subventricular zone. Neural stem and progenitor cells can be isolated from these brain regions and expanded in cell culture, known as the neurosphere assay.

Recently, stem cells in the brain have become interesting targets for stroke and neurodegenerative diseases. In stroke, the blood supply in the brain is diminished, mainly by vessel occlusion (ischemia) or bleeding (hemorrhagia). Hence, low oxygen tension (hypoxia) and low glucose concentrations cause neuronal cell death in the brain tissue. These conditions can be mimicked in cell culture (oxygen and glucose deprivation, OGD) and are a standard model for *in vitro* ischemia.

One of our main goals in the National Genome Research Network was to compare the reaction of neural stem cells, which are considered immature cells, to mature neurons with regard to cell survival. We were especially interested in the proliferation and differentiation behavior of neural stem cells under hypoxic and low glucose conditions.

New therapies based on the replacement of damaged tissue after stroke with the aim to re-establish lost function can either be achieved by activating endogenous stem cells or by transplanting stem cells from an exogenous source. For either kind of therapy, neural stem cells will need an alignment of their genetic program to survive, proliferate and differentiate within the brain. For these functions, growth factors play an essential role in the brain. Growth factors identified in neural stem cell proliferation and differentiation include fibroblast growth factor, endothelial growth factor, vascular endothelial growth factor, granulocyte-colony stimulating growth factor, and erythropoietin.

Erythropoietin is a 34 kDa glycoprotein primarily produced in the adult kidney upon hypoxic stimuli. It is the principal growth factor regulating red blood cell production. Several studies have shown that erythropoietin is an important survival factor for neurons under hypoxic conditions *in vivo* and *in vitro*. During brain development, erythropoietin acts as a neurotrophin and has anti-apoptotic effects. In the adult, erythropoietin may also serve as a neuroprotective factor in stroke. With regard to neural stem cells, erythropoietin directly promotes the generation of neuronal stem cells from progenitors and induces stem cell migration.

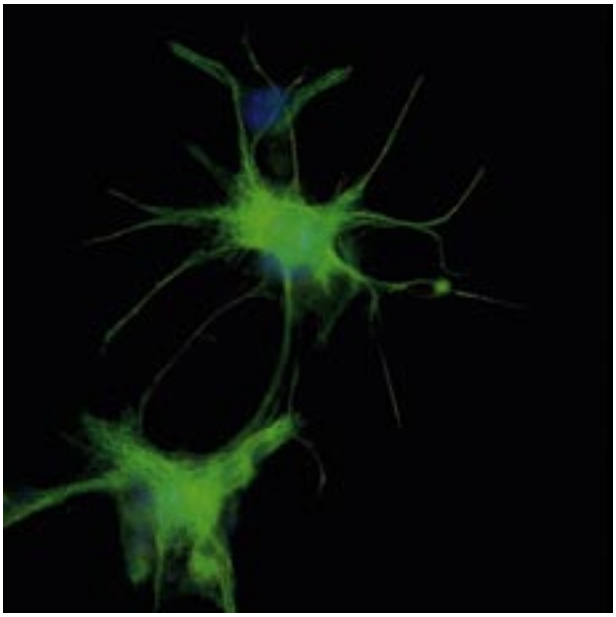


Fig. 1: Neural stem cells from the rat brain express the intermediate filament protein nestin (green). Cell nuclei are marked in blue. (Photo: D. Schelshorn)

PROJECT STATUS/RESULTS

The proteome of a cell is the set of all proteins, and can be regarded as protein complement to the genome, which is the set of all genes. In the current project, we used a technology called “proteome analysis” which allowed us to measure the expression of several hundreds or even thousands of proteins simultaneously. We used an approach based on two-dimensional gel electrophoresis for protein separation and mass spectrometry for protein identification.

In the current project, we applied the proteome analysis technique to cultured neural stem cells exposed to hypoxia, low glucose, and erythropoietin. Surprisingly, the neural stem cells did not react like mature neurons: whereas the neurons were dramatically harmed, the neural stem cells proliferated in the hypoxic environment – we counted a more than 1.5-

COORDINATOR
Professor Peter Seeburg
Max Planck Institute for Medical Research, Heidelberg
seeburg@mpimf-heidelberg.mpg.de

fold increase in cell number when no oxygen was present. It was only when we combined low oxygen tension with low glucose levels that the cell number of neural stem cells was reduced.

Interestingly, erythropoietin did not affect the differentiation of neural stem cells, but enhanced their survival. We searched for the molecular pathways involved in this phenomenon and found that programmed cell death (apoptosis) can be reduced by erythropoietin in the neural stem cells.

Glycogen Synthase Kinase 3-beta (GSK3β) is a key regulator of neural stem cell differentiation

Moreover, we compared undifferentiated and *in vitro*-differentiated neural stem cells by the proteome analysis method. We identified the multifunctional enzyme Glycogen Synthase Kinase 3-beta (GSK3β) as one of the key regulators of stem cell differentiation. GSK3β is involved in several signaling pathways related to cell death and differentiation, for example the developmental Wnt signaling pathway.

Disturbances in both neural stem cells and GSK3β have been associated with brain disease, such as bipolar disorders, depression, or schizophrenia. In this context, it is of special interest that we also identified small molecules which inhibited GSK3β and, hence, increased the number of neurons derived from differentiated neural stem cells.

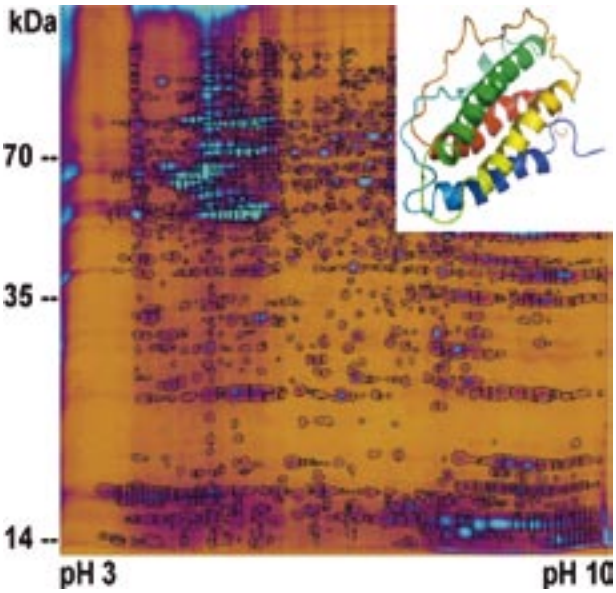


Fig. 2: Two-dimensional gel image of the neural stem cell proteome. Insert: molecular structure of human erythropoietin (Image: M. Maurer)

OUTLOOK

Our recent findings provide novel aspects of stem cell neurobiology which may provide additional pathophysiological insights into their differentiation and proliferation. For example, growth factors such as erythropoietin, or granulocyte-colony stimulating factor, may become an additional option in the treatment of stroke, with increased drug safety and an enlarged time window.

Additionally, neural stem cells may become important targets of drugs used in neurology and psychiatry. But one of the prerequisites is to understand their cellular behavior, i.e. proliferation, migration, survival, integration, and differentiation. Our projects have opened up the prospect of achieving this aim, both with regard to the methodology and the results.

References

[1] *Maurer MH et al. Comparison of statistical approaches for the analysis of proteome expression data of differentiating neural stem cells. J Proteome Res. 2005; 4(1):96-100.*

[2] *Maurer MH et al. Adult neural stem cells express glucose transporters GLUT1 and GLUT3 and regulate GLUT3 expression. FEBS Lett. 2006; 580(18):4430-4.*

[3] *Maurer MH et al. Glycogen Synthase Kinase 3beta (GSK3-beta) regulates differentiation and proliferation in neural stem cells from the rat subventricular zone. J Proteome Res. 2007; 6(3):1198-208.*

[4] *Maurer MH et al. Screening the brain: molecular fingerprints of neural stem cells. Curr Stem Cell Res Ther. 2006; 1(1):65-77.*

SIGNATURES, PATHWAYS AND GENES: NEW SOLUTIONS FOR EXPRESSION PROFILING

THOMAS HÄUPL[1], ANDREAS GRÜTZKAU[2], JOACHIM GRÜN[2], ADRIANE MENSSEN[1], ANDREAS RADBRUCH[2], GERD-RÜDIGER BURMESTER[1]

[1] DEPARTMENT OF RHEUMATOLOGY AND CLINICAL IMMUNOLOGY, CHARITÉ-UNIVERSITY MEDICINE BERLIN
[2] GERMAN RHEUMATISM RESEARCH CENTER, BERLIN

INTRODUCTION

Researchers are becoming more and more interested in microarray analysis of transcriptomes. However, correct interpretation of data is still a challenge. Although public repositories of chip data are constantly expanding, they do not yet provide convenient and instantaneous access for retrieving array information. Furthermore, results of different experiments cannot be compared based on the complete set of array data, but only on the level of published interpretation and is usually performed with different strategies. This constricts the potential development of expression profiling as a tool for systems biology. Due to the current uncertainty of data interpretation in this field, the expression results in each transcript must be individually confirmed using independent technologies. This task cannot be achieved in a reasonable time frame without the use of array technology. It is therefore imperative that new strategies are developed to better analyze the wealth of array data currently available.

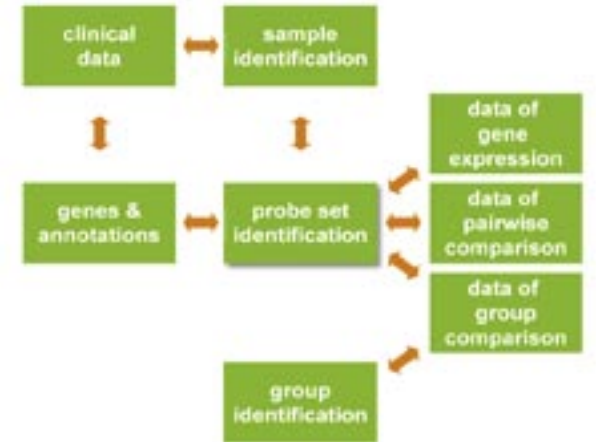


Fig. 1: Structure of the SiPaGene database: The relational database is centered to the probe set and gene. The link of transcriptome to clinical data is not implemented for online retrieval. Annotations are currently adopted from Affymetrix. Own annotation networks are under construction.

PROJECT STATUS/RESULTS

Chronic inflammatory diseases exemplify the complexity of transcriptome data. Our effort was focused on systematically developing the transcriptome signatures of the contributing cell types and extending them to the different profiles induced by various major stimuli. To explore these many data sets, an Oracle database (SiPaGene) was built that provides standardized analysis and tools to compare profiles online (www.bioretis.de). Access was restricted by ownership and individual sharing options to respect concerns about loss of intellectual property and chances of each individual scientist to exploit own data adequately. In a first version, the database was set up to manage information generated with the currently leading profiling technology GeneChip produced by Affymetrix.

Users initiate their analysis by selecting the arrays of interest and starting an automated standardized group comparison. After calculation of 19 different statistical parameters per group comparison, queries can be performed by defining thresholds for these parameters and combining them with the Boolean operators. Besides single group comparisons, multiple group comparisons can be retrieved for lists of overlapping, extended or exclusive candidate genes. Overviews and detailed views on candidate genes allow the user screen to the data. Download options provide statistical parameters of the group comparison as well as signal information from all individual arrays for further analysis by the user (e.g. clustering).

Based on a long-standing experience with the different query parameters, we developed an optimized strategy to set thresholds and combinations of the parameters to identify the relevant candidate genes. These selection conditions are integrated as default queries and were tested with the Affymetrix Latin Square spike-in data set. Access to these data is open to the guest user. Compared to other established tools like significance analysis of microarrays (SAM) or the DNA-chip analyzer (dChip), this new instrument shows improved capabilities to include true positive and exclude true negative transcripts.

The database is currently used as an Internet platform by about 100 registered users in several networks funded by the BMBF (NGFN) or the European Union (Autocure, Genostem, Autorome). It stores about 500 arrays, 6000 pairwise comparisons and 300 group comparisons, and the number of projects, users and arrays is constantly increasing.

Another important aim was to improve data interpretation with new algorithms. The systematic collection of expression

COORDINATOR
Professor Trinad Chakraborty
University of Giessen
trinad.chakraborty@mikrobio.med.uni-giessen.de

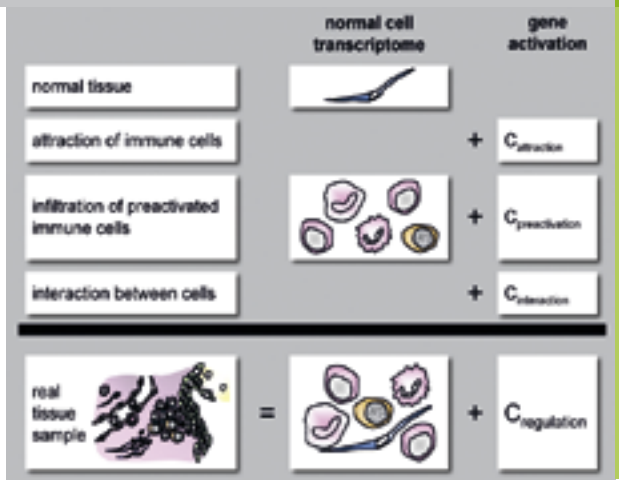


Fig. 2: Functional profile component analysis: Profiles are composed of different functional components. These include components of normal cell profiles and components that consist of gene activation as part of the molecular processes induced by the disease.

profiles from different cell types and major stimuli in combination with the query tools of SiPaGene already improves analysis. Nevertheless, restrictions are obvious when complex samples like whole blood or inflamed tissue are analyzed. Both consist of various cell types with variable composition. Differences between transcriptomes of disparate cell types are extreme – even under normal conditions. Therefore, a shift in cellular composition is accompanied by a broad range of differential gene expression. However, activation of genes in defined cell types as part of the disease process is much more relevant to understand the molecular mechanisms and to decide on the therapeutic strategy. Unfortunately, both aspects are intermixed, and systematic analysis of inflamed joint tissue as a relevant example demonstrated that differential cellular composition contributes 3-4 times more genes to the changes in gene expression than activation of pathophysiologically relevant genes.

Therefore, we developed a new strategy to quantify and normalize differential cellular composition. This new instrument, functional profile component analysis (FPCA), relies on reference profiles from isolated cell types. Based on marker genes, the fraction of each cell type is quantified and used to calculate a virtual profile that would be expected if all cells of the mixture would present a normal expression profile. Comparison between the real profile of the blood or tissue and the virtual profile reveals the functional component related to gene activation induced by the disease. This algorithm is currently integrated into the SiPaGene database for automated application with various types of samples (blood, inflamed or cancer tissue, etc.).

Furthermore, reference signatures of the various cell types were used to develop a concept for functional annotation. A score for the association of a given gene with a particular cell type was developed which allows estimating the influence of differential cellular composition independently of FPCA. It can also be applied to other array platforms.

OUTLOOK

New concepts such as presented here are needed to open new perspectives for systematic and comprehensive analysis of transcriptome data. Although, scientists are concerned about giving access to their array data early after generation, we experienced that the new options of SiPaGene to exploit own data faster and to share only with individually selected colleagues fuels networking. The concept of FPCA will be of particular importance for the analysis of whole blood samples, an indispensable requirement for clinical studies. Thus, new concepts and tools such as presented here will stimulate research and improve our perception and understanding of transcriptome data in particular as tools for recognizing key processes that shape biological systems.

References

[1] *Häupl T et al. SiPaGene: An integrative database for complex analysis and sharing of Affymetrix gene expression data. Ann Rheum Dis 2007; 66(Suppl II):130.*

[2] *Häupl T et al. Expression analysis of rheumatic diseases, prospects and problems. In: The Hereditary Basis of Rheumatic Diseases, Editor Holmdahl: Progress in Inflammation Research, 2006.*

[3] *Häupl T et al. Perspectives and limitations of gene expression profiling in rheumatology: new molecular strategies. Arthritis Res Ther. 6(4):140-6 (2004).*

[4] *Häupl T et al. Methods for recognizing signatures in complex gene expression profiles. WO002005095644A3.*

PROSPECTIVE EVALUATION OF CLINICAL SEPSIS PHENOTYPES BY FOCUSED GENE EXPRESSION PROFILING AND ASSOCIATION WITH GENETIC RESISTANCE FACTORS

TRINAD CHAKRABORTY
INSTITUTE OF MEDICAL MICROBIOLOGY, UNIVERSITY OF GIESSEN

INTRODUCTION

Mortality due to sepsis, septic shock and septic organ failure has not improved substantially over the past 20 years and represents a leading cause of mortality in critically ill adults despite best life-supportive interventions available. According to a recent study in US hospitals, more patients die from sepsis than from acute heart attack, making sepsis one of the most underestimated killers in developed countries. Current techniques of disease monitoring are insufficient to assess the individual risk for sepsis development in critically ill patients, and they fail to reliably predict the individual clinical course in septic patients regarding rapid deterioration, recovery or persistent septic organ failure. The role of the patient's genetic background and predisposition and the extent of the activation of the inflammation/coagulation appear to be crucial. Rationally guided therapeutic intervention strategies in both adults and preterm neonates are, however, critically dependent on early, sensitive and potent surrogate markers for the individual patient's clinical course and probable response to therapies targeting inflammatory/coagulation cascades.

PROJECT STATUS/RESULTS

Polytrauma

The prospective study of polytrauma presented here is currently the largest prospective study of its kind on severe injury in the world. The study was performed with an extremely well-defined cohort suffering from severe multiple trauma. The aim was to uncover the biological reasons why patients with multiple trauma can have dramatically different outcomes after suffering similar traumatic insults. Peripheral whole blood obtained at the time of admission to the ICU, i.e. 12-24 hours after trauma, was used for transcriptome analysis, and the sequence of events from the admission to the ICU to the very onset of sepsis and finally the full development of multiple organ dysfunction syndrome (MODS) was clinically monitored.

In a time-dependent analysis of transcriptional changes in patients prior to, during the onset and undergoing recovery from sepsis, clear differences in the inflammatory responses were found: Many of the genes encoding pro- and anti-inflammatory mediators are induced immediately within the first 24h post injury but are turned off dramatically at time points when the patients are experiencing sepsis. On the other hand, genes involved in innate cellular defenses such as defensins remain highly induced. At time points where patients recover from sepsis, there is restoration of the expression of the genes involved in the inflammatory response.

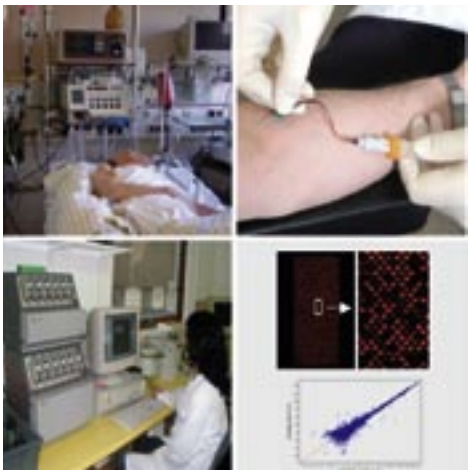


Fig. 1: Microarray analyses of peripheral blood may soon lead to better sepsis diagnosis.

The study rendered several important results: First, peripheral whole blood is a suitable, easily accessible surrogate marker for the host response. Secondly, gene expression profiles can allow stratification of patients into risk classes 12–24 h after trauma, thus indicating patients who will develop sepsis and those who will not. Prediction analyses revealed that a profile of 90 genes can be used as a predictive expression pattern for sepsis. Moreover, the study provided comprehensive insights into the mechanisms that lead to sepsis.

In addition to the explorative profiling, single nucleotide polymorphisms (SNPs) in the tumor necrosis factor (TNF) and lymphotoxin alpha (LTA) genes were focused upon. A candidate-based approach for detecting susceptibility in the first 159 patients in this cohort revealed that harboring the TNF-308 A allele is associated with high serum TNF concentrations on the first day after trauma and during follow-up, with development of sepsis syndrome and with fatal outcome. Gene

COORDINATOR
Professor Trinad Chakraborty
University of Giessen
trinad.chakraborty@mikrobio.med.uni-giessen.de

expression studies on these patients also indicated that the carriage of the TNF rs1800629 A allele was associated with differential expression of genes representing stronger pro-inflammatory and apoptotic responses.

Congenital sepsis

In newborns, preterm birth is the major risk factor for Early Onset Sepsis (EOS). Preterm sepsis constitutes a major clinical problem with 14 000 cases annually. It is associated with a high mortality rate (up to 40%) and the development of severe and long-term complications. Clinical signs of infection are diffuse, and markers for early diagnosis of infection are lacking. In NGFN-2, a study of preterm infants was therefore initiated in which samples from umbilical cord blood obtained from infants at birth were subjected to microarray analysis. Clinical courses of these infants were monitored comprehensively. It turned out that expression profiles obtained at birth allowed clear differentiation of infants who developed clinical and laboratory signs of EOS to those remaining uneventful within 72 hours. Several genes upregulated in infants with EOS were genes involved in inflammatory and apoptotic events and in neutrophil activation, while downregulated genes were mainly involved in activation and proliferation of B, T and NK cells. The results clearly suggested increased neutrophil activation and lymphocyte suppression. This could be explained by the occurrence of an intrauterine inflammatory response known as the fetal inflammatory response syndrome (FIRS). Based on this vulnerable state of the developing organism at birth, further postnatal stress, such as delivery, mechanical ventilation, and invasive procedures might lead to severe perturbations of the immune system that contribute to the development of EOS. Transcriptional profiling emerged as a powerful tool, far superior to current clinical and laboratory diagnostic procedures in identifying preterms at risk for developing sepsis.

Various surfactant protein genotypes and genetic variations in surfactant pathways have been associated with susceptibility to neonatal respiratory distress syndrome (RDS). Epidemiological studies described associations between the SNPs of the human gene coding surfactant protein-D (SFTPD) and infectious pulmonary diseases. To determine the genetic influence of the sequence variations within the SFTPD gene on the constitutional serum levels of SP-D, three different regions of the SFTPD gene of 32 randomly selected blood donors were sequenced in an additional study. Serum SP-D levels were analyzed and the association of SFTPD haplotype estimates with the quantitative phenotype serum SP-D level was determined. One single SFTPD haplotype revealed a negative association with serum SP-D levels, which was confirmed in a second prospectively collected group of blood donors. The discovery of a frequent negative variant of the SFTPD gene provides a basis for genetic analysis of the function of SP-D in the resistance against pulmonary infections and inflammatory disorders in humans.

OUTLOOK

The results of the gene expression profiling studies of polytrauma patients and preterm infants with and without EOS will be transferred into the development of diagnostics at the RNA, metabolite and protein level. The query of the relevant genes and gene products at different biological levels (transcriptome, metabolome, proteome) will allow an early diagnosis and a comprehensive biological view of the patients' status during the course of the disease. The tests will be based on selected technical platforms that allow fast and point–of-care diagnostics (POCD) in order not only to minimize the sampling time but also to minimize the time from patients' admission to diagnosis of septic onset. Collaborations have already been initiated with several industrial partners.

The work described was supported by: NGFN-1: IE-S08T03, IE-S08T04; NGFN-2: NIE-S14T04, NIE-S14T05

References

[1] *Heidinger K et al. Polymorphisms in the human surfactant protein-D (SFTPD) gene: strong evidence that serum levels of surfactant protein-D (SP-D) are genetically influenced. Immunogenetics. 2005 Apr; 57(1-2):1-7.*

[2] *Hossain H et al. Host gene expression profiling in pathogen-host interactions.Curr Opin Immunol. 2006 Aug; 18(4):422-9.*

[3] *Hilgendorff A et al. Identification of protective and pathological pathways in early onset infection of preterm infants by genome-wide expression profiling. Submitted.*

[4] *Menges T et al. Sepsis syndrome and death in trauma patients is associated with variation in the gene encoding TNF. Crit Care Med. Accepted.*

[5] *Hilgendorff A et al. Association of polymorphisms in the mannose-binding lectin gene and pulmonary morbidity in preterm infants. Genes Immun. 2007 Sep 27; (Epup ahead of print).*

Patents

A patent application "Invention regarding expression profiles for the prediction of septic states" registered under the number DE 10 2005 050 933.9 was filed on Oct. 21, 2005.

CROHN'S DISEASE – A MICROBIAL DEFENSE PROBLEM OF THE INTESTINAL BARRIER?

STEFAN SCHREIBER

INSTITUTE OF CLINICAL MOLECULAR BIOLOGY, UNIVERSITY OF SCHLESWIG-HOLSTEIN, CAMPUS KIEL

INTRODUCTION

Crohn's disease is a chronic, relapsing inflammatory condition that mainly affects the small and large bowel. It is the paradigm of a disease of civilization with significant rises in incidence after World War II but with a clear polygenic susceptibility component (i.e. multiple disease genes interacting). It affects up to 0.5% of the population in Western industrialized nations. The main manifestation age is early adulthood and the dramatic decrease in quality of life through pain, diarrhea and severe complications is often debilitating. In 2001 the first disease gene, NOD2, was discovered in parallel by two international groups and in the NGFN.[1-3] Surprisingly, this gene was found to code for an intracellular, antimicrobial defense protein and not – as expected – for a molecule that would be tightly involved in the regulation of an immune cell-driven inflammatory process. The large cohorts of Crohn's disease patients that are available in the NGFN and through the BMBF competence network Inflammatory Bowel Disease allowed annotation of a well-defined sub-phenotype that is related to the genetic variants.[4] The polygenic nature of the disease was further explained through additional disease genes discovered in Canada (OCTN1/2 – epithelial cell-based kation transporters) and in the NGFN (DLG5 – a gene responsible for mechanical barrier function in the colon epithelium).[5,6] These three disease genes clearly point to a key role of the epithelial lining of the intestine in the disease process.

PROJECT STATUS/RESULTS

New strategies to explore genetic susceptibility in polygenic disorders

Early discoveries of disease genes were made using an algorithm adapted from the exploration of monogenic diseases, i.e. the definition of susceptibility regions through linkage analysis of affected sib pairs and the subsequent association studies for fine-mapping in large sets of cases and controls using single nucleotide polymorphisms (SNPs). A new era has started with the advent of the SNP chip/array technology allowing genome-wide, systematic association studies that either use a dense coverage based on the HapMap (i.e. using 500 000 – 1 million SNPs per individual) or a comprehensive set of coding SNPs.

The success of systematic association studies

The first systematic association studies used a low density set of SNPs (i.e. between 100 000 and 300 000 SNPs). They already documented that the new technologies deliver a fast track to discovery by annotating a prostaglandin receptor (PTGER4), its upstream regulator NELL1 through NGFN projects and the IL-23 receptor gene as susceptibility factors for Crohn's disease.[7-9] The prostaglandin system is thought to be responsible for much of the unspecific damage in the inflamed mucosa and IL-23 is both a regulator of T cell activation and an important driver for microbial defense by epithelial cells.

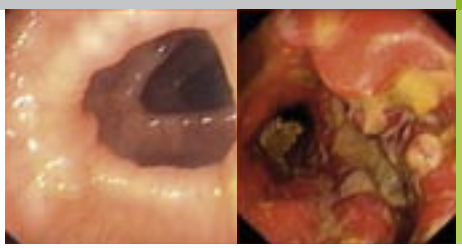
Further successes came through the use of a genome-wide coding SNP set that identified ATG16L1 (and confirmed the IL23R and the NOD2 signals obtained earlier).[10] ATG16L1 is involved in the autophagy pathway that is used to dispose of microbial remnants (in particular tuberculosis) and protein debris. Together with NOD2, ATG16L1 is among the most consistently annotated disease genes in Crohn's disease. A massive experiment using 500 000 SNPs in a very large set of patient and control samples confirmed many of the previous discoveries and newly identifies IRGM, a further gene in the autophagy pathway and PTPN2, a protease that most interestingly is a disease gene for both Crohn's disease and diabetes.[11] Finally, a three-stage experiment in Crohn's disease patients from a genetically restricted population (i.e. the Quebec French-Canadian population) followed by confirmation and fine mapping in German patients came to a comprehensive identification of a whole series of further disease genes including additional components of the IL-23R signaling cascade, further bacterial defense genes and some genes coding for proteins with yet unknown functions.[12]

The next steps to complete understanding of disease mechanisms

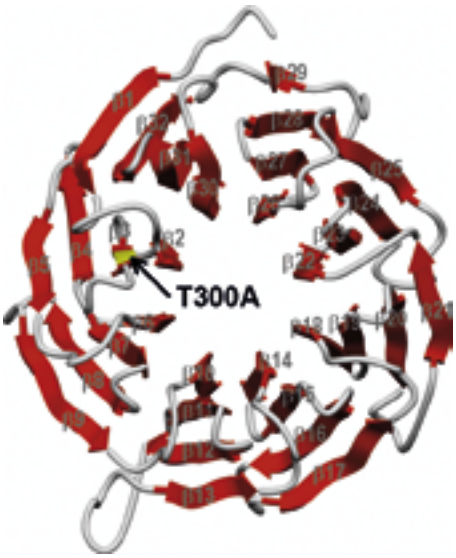
Obviously, the systematic and comprehensive discovery of disease genes in a complex disease has been greatly advanced through the availability of genome-wide association studies. However, the identification of a whole series of diseases genes alone will not satisfy the need for knowledge, since the responsible disease mechanisms still remain to be identified. For this task, systematic resequencing of the disease genes and regions is needed in large numbers (i.e. thousands) of individuals. NGFN has already made investments into new technologies providing the necessary throughput for this task in Berlin, Heidelberg and Kiel.

However, the functional annotation of disease genes will remain a bottleneck. While genetic discovery allows identifying genes in an unprecedented manner, many of the cell biol-

COORDINATOR
Professor Stefan Schreiber
University Hospital Schleswig-Holstein, Campus Kiel
s.schreiber@mucosa.de



ogy techniques are still not automated. A final component will be systematic modeling of the complex disease based on the multiple genetic factors and molecular pathways, requiring a medical systems biology.



Potential benefit for patients

An important issue is the immediate benefit these discoveries may have for patients and their families. Genetic discovery will lead to a fundamental change in the therapeutic goals in Crohn's disease. While the pharmaceutical industry has focused on new therapies to suppress the adaptive immunity process in order to alleviate the inflammatory symptoms[e.g. 13], it is now clear that future therapies will be directed towards augmenting epithelial defense.

OUTLOOK

Inflammatory diseases of the different barrier organs (e.g. intestinal mucosa, lung, mouth, skin) are related. This is documented by increased coincidences but also a remarkable overlap of the diseases genes found thus far (e.g. NOD2 is a disease gene for Crohn's disease but also for asthma, IL23R for Crohn's disease but also for psoriasis). The simultaneous and comprehensive study of inflammatory barrier diseases in the NGFN will therefore lead to a new medical perspective on how to define and treat inflammatory barrier diseases. The medical grouping into indications that presently are determined by anatomic definitions will be replaced by functional definitions.[14] The National Genome Research Network offers a unique platform with large patient cohorts for genome-wide studies of Crohn's disease, ulcerative colitis, atopic eczema, psoriasis, asthma, sarcoidosis and periodontitis.

In the future, we hope to gain full understanding of the genetic mechanisms and the trigger factors that are necessary in our environment to precipitate genetic susceptibility for disease. We expect that the final consequence of understanding the genetic etiology of inflammatory barrier disease will not only be new therapies but also a targeted prevention directed at maintaining health rather than treating disease.

References

[1] *Lancet 2001, 357: 1925-1928 (NGFN publication).*

[2] *Nature 2001, 411, 603-6.*

[3] *Nature 2001, 411, 599-603.*

[4] *Lancet 2002, 359, 1661-1665 (NGFN publication).*

[5] *Nat Genet 36, 2004, 471-5.*

[6] *Nat Genetics 2004, 36: 476-480 (NGFN publication).*

[7] *PLoS Genet 2007, 3: e58.*

[8] *Franke A et al. A genome-wide association scan identifies NELL1 as a novel IBD disease gene and annotates 5p13.1 as a gender-related susceptibility region for Crohn's disease. PLOS ONE 2007, in press (NGFN publication).*

[9] *Science 2006, 314: 1461-1463.*

[10] *Nat Genetics 2007, 39: 207-11 (NGFN publication).*

[11] *Nat Genetics 2007, doi:10.1038/ng2061.*

[12] *Raelson J et al. Genome-wide Association Study for Crohn Disease in the Quebec Founder Population Identifies Multiple Validated Disease Loci. Proc Natl Acad Sci USA 2007, in press (NGFN publication).*

[13] *New Engl J Med 2007, 357: 228-238 and 230-250.*

[14] *Nat Rev Genet 2005, 6: 376-88 (NGFN publication).*

MAJOR ASTHMA GENE IDENTIFIED

MICHAEL KABESCH
LUDWIG-MAXIMILIANS-UNIVERSITÄT, MUNICH

INTRODUCTION

Asthma is a complex chronic inflammatory disease of the airways, leading to airway obstruction and difficulties in breathing (Fig. 1). It affects up to 30% of children worldwide and has been identified as one of the major health burdens of the 21st century by the World Health Organization (WHO).

PROJECT STATUS/RESULTS

The combinations of genetic and environmental factors which cause asthma have been poorly understood so far. Thus, asthma was in the focus of genetic research in the National Genome Research Network (NGFN) from the beginning. By combining clinical and genetic expertise with networking and the application of the newest technologies, major breakthroughs have now been achieved.

Fig. 1: Asthma causes shortness of breath in millions of children worldwide

Patient populations were recruited in a multicenter approach in specialized pediatric pulmonology departments in Germany, Austria and the UK. Phenotypes of asthma were characterized in great detail and at the highest clinical level in almost 1000 patients with the purpose of identifying the genetic and environmental risk factors leading to asthma. Patients donated blood for DNA analyses and gave extensive information on life style factors. The clinical recruitment of childhood asthma cases for genetic studies in Germany and Austria, the Multicenter Asthma Genetics in Childhood (MAGIC) study, was established with the help of NGFN-1 and involved numerous centers in both countries.

Together with colleagues from the UK, France, and the US, a genome-wide association (GWA) study was then performed, analyzing more than 317 000 genetic markers, so-called single nucleotide polymorphisms, in each study participant (Fig. 2). Altogether, almost 2300 children in the UK and Germany were genotyped, including 994 cases and 1243 controls. At the same time, the influence of 100 000 of these genetic markers on global gene expression was studied in lymphocytes of 400 children.

A striking effect of a cluster of polymorphisms on chromosome 17 on the risk of asthma in children was identified (Fig. 3). These effects were found in the whole genome-wide association study population as well as in German and British children when analyzed separately. In a subsample of the participants family data was available. and transmission disequilibrium tests between generations confirmed the results obtained by genome-wide associations.

Next, global gene expression was studied in 400 samples. The single nucleotide polymorphisms common to both populations which were associated with asthma also altered the expression level of a new gene called ORMDL3, located in the vicinity of the polymorphism cluster. ORMDL3 was present at a higher level in the blood cells of children with asthma than in those without. Association results could also be replicated in over 2000 children from Germany and over 3000 UK subjects in independent experiments.

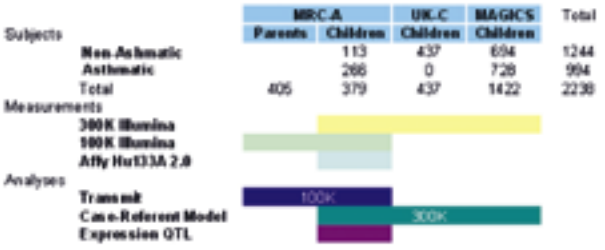


Fig. 2: Study design of asthma genome-wide association studies

How these SNPs regulate ORMDL3 expression and if other genes in the vicinity of the polymorphic locus on chromosome 17 are also affected by this polymorphic regulatory element on chromosome 17 is the focus of further investigations. How ORMDL3 is interfering with the development of asthma remains to be elucidated and is currently being studied using animal models and human cell lines. Further suggestive associations between asthma and other polymorphisms from the genome-wide association study are also being investigated.

COORDINATOR
Professor Stefan Schreiber
University Hospital Schleswig-Holstein, Campus Kiel
s.schreiber@mucosa.de

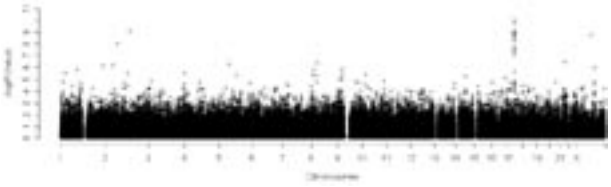


Fig. 3: Genome-wide association of 317 447 SNPs and asthma in 994 asthmatic children and 1243 non-asthmatic children. Numerous markers on chromosome 17q21 show association to asthma in the region of maximum association.

The data, which have now been published in *Nature*, represent the strongest genetic effect on asthma so far discovered. A new understanding of disease mechanisms in asthma can be derived from these findings. It gives rise to the hope that, based on these results, new strategies for the prevention, diagnosis and therapy of asthma can be developed in the near future.

OUTLOOK

Genome-wide association studies similar to those performed in asthma are ongoing in the NGFN to decipher the genetic susceptibility for atopic dermatitis, hay fever and the regulation of immunoglobulin E production. These studies, which are expected to be published during 2007, may also change our views on allergic diseases.

Reference

[1] *Moffatt MF et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. Nature. 2007. 404:470-474.*

THE NATIONAL GENOTYPING PLATFORM

PETER NÜRNBERG

COLOGNE CENTER FOR GENOMICS (CCG), INSTITUTE OF GENETICS AT THE UNIVERSITY OF COLOGNE

INTRODUCTION

Positional cloning is widely used for the identification of gene defects that underlie inherited diseases. A necessary first step for positional cloning is the mapping of the gene locus that cosegregates within families with a particular disease or trait, which allows allocating a specific chromosomal position to the responsible gene. Although mapping was initially developed for monogenetic traits, it has become a widely used strategy to localize genetic factors involved in the etiology of multifactorial diseases. One of the key missions of the National Genome Research Network (NGFN) is to identify the genetic factors involved in complex diseases by establishing the link between phenotypic data and the genome of the patients. This approach requires substantial resources in high-throughput genotyping of “short tandem repeat” (STR) and “single nucleotide polymorphism” (SNP) markers. Therefore, a consortium was founded comprising the most potent genotyping centers of Germany to form a National Genotyping Platform. This platform acts in a coordinated fashion to meet all genotyping requirements of the projects supported by the NGFN. The capacity of the platform is sufficient to process several thousands of samples within a couple of weeks.

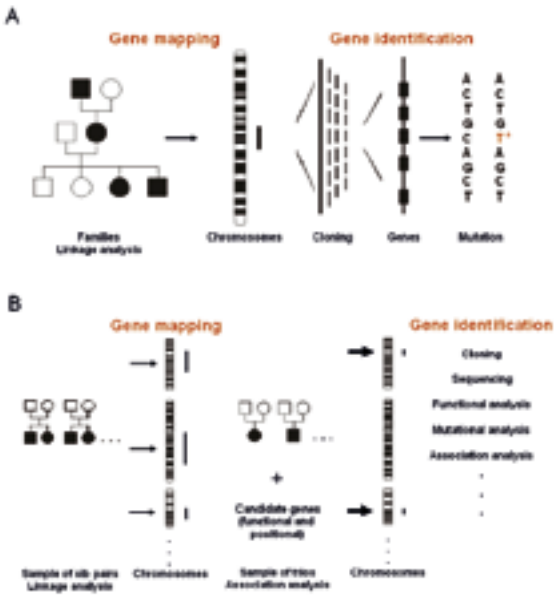


Fig. 1: Positional cloning approach for Mendelian disorders (A) and complex diseases (B)

PROJECT STATUS/RESULTS

Microsatellite or STR marker typing at high-throughput level is technically demanding and needs special expertise. The National Genotyping Platform (NGP) provides this service at the Cologne Center for Genomics and the Gene Mapping Center in Berlin. Although STRs are being replaced by SNPs step by step, there is still a need for STR marker-based genome scans in model organisms as long as sufficient SNP resources are not available. Furthermore, due to their highly polymorphic nature, STR markers are very useful for supplementing SNP scans, especially if reliable haplotype information is necessary. Marker sets for human, mouse, rat, and rhesus macaques are available within the NGP.

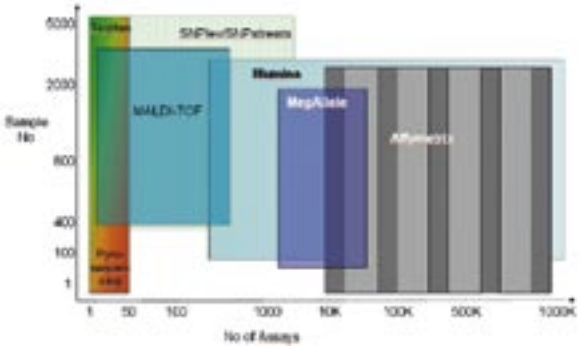


Fig. 2: Choice of SNP typing technology is highly dependent on the numbers of samples and SNPs to be genotyped.

For SNP genotyping, a plethora of different methods is available – each with specific advantages in particular set-ups. There are two important criteria for cost-efficient genotyping, the number of DNA samples and the number of SNPs under investigation. The NGP provides efficient technologies for all conceivable scenarios in SNP typing. Array-based procedures, such as the methods supported by Illumina and Affymetrix, are particularly useful when large numbers of SNP markers are to be typed. However, a common problem of all highly parallel microarray-based technologies is their inability to efficiently type subsets of SNPs for stepwise approaches, replication studies, or candidate gene approaches. This opens a window for SNP typing methods working in single or low-plex mode such as MALDI-TOF mass spectrometry and pyrosequencing. The SNPlex assay, which is based on a multiplexed version of the oligoligation assay (OLA) with an automated read-out on capillary DNA sequencers, is a powerful alternative technology to all mentioned procedures in cases in which medium numbers of SNPs (fifty to several hundred) and several thousand DNAs have to be genotyped. The same type of projects may be run using the SNPstream

COORDINATOR
Professor Hans Lehrach
Max Planck Institute for Molecular Genetics, Berlin
lehrach@molgen.de

TABLE 1: GENOTYPING TECHNOLOGIES PROVIDED BY THE CENTERS

Method	Berlin	Bonn	Kiel	Cologne	Munich
STR marker	•			•	
Pyrosequencing	•			•	•
TaqMan	•	•	•	•	•
MALDI-TOF	•	•			•
SNPlex	•		•		
SNPstream				•	
Illumina	•	•		•	•
MegAllele	•				
Affymetrix	•			•	•

assay from Beckman Coulter which is based on single base extension.

The participating genotyping centers have developed superior expertise with the different technologies and closely cooperate during project management (Table 1).

The activities of the centers also included technical improvements of genotyping methods[1] as well as the development of software tools to facilitate and streamline the workflow from sample input to data analysis[2]. Finally, all these efforts were made to better fulfill the mission of the NGP, namely, assisting clinical partners in their positional cloning projects to identify disease genes. Especially by using array-based SNP technologies, mapping projects could be accelerated enormously (Fig. 3).

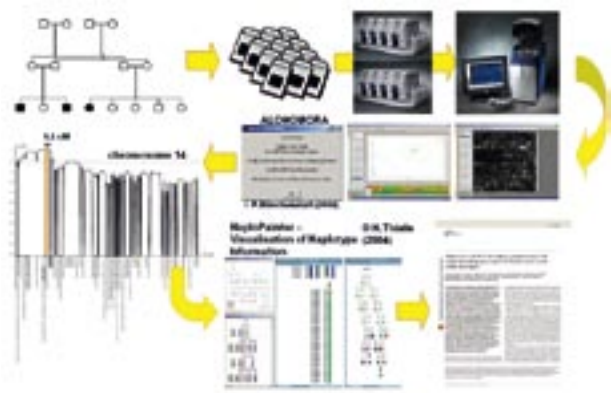


Fig. 3: High-speed mapping pipeline of the NGP

References

[1] *Sauer S et al. (2006) Single-nucleotide polymorphisms: analysis by mass spectrometry. Nature Protoc. 1:1761-71.*

[2] *Franke A et al. (2006) GENOMIZER: an integrated analysis system for genome-wide association data. Hum Mutat. 27:583-8.*

[3] *Hinkes B et al. (2006) Positional cloning uncovers mutations in PLCE1 responsible for a nephrotic syndrome variant that may be reversible. Nat Genet. 38: 1397-405.*

[4] *Buch S et al. (2007) A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease. Nat Genet. 2007 Jul 15; [Epub ahead of print].*

[5] *Winkelmann J (2007) et al. Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. Nat Genet. 2007 Jul 18; [Epub ahead of print].*

THE NGFN PROMOTER RESOURCE

WILFRIED NIETFELD, ANNA GUERASIMOVA, STEFAN HAAS, BERNHARD G. HERRMANN, MICHAL JANITZ, SZYMON M. KIELBASA, THOMAS MANKE, LORENZ NEIDHARDT, ILARIA PICCINI, ROBERT QUERFURTH, MARTIN VINGRON, HANS-JÖRG WARNATZ, MARIE-LAURE YASPO AND HANS LEHRACH

MAX PLANCK INSTITUTE FOR MOLECULAR GENETICS, BERLIN

INTRODUCTION

Genetic networks regulate all biological processes, and transcriptional regulation is a key component. Transcription factors (TFs) interact with DNA elements of promoter regions and regulate the expression of the corresponding gene(s). To dissect such regulatory networks, we have established an analysis pipeline allowing the systematic analysis of promoter functions. Since the mid-1990s many research groups have focused on the analysis of individual promoters[1] to determine the function of even individual bases. The availability of the human reference sequence and the advent of genome-wide expression profiling data allow us to correlate the transcript expression levels with the promoter elements of the respective genes.[2] The NGFN promoter pipeline provides a resource of human and mouse promoter regions of about 2.5 Kb upstream of the transcription start site, which are cloned as reporter constructs for functional analysis (Fig. 1). Promoter regions are annotated and PCR primers designed in the SMP DNA project "Promoter Informatics". In the SMP DNA project "Promoter Resource" the identified promoter regions are amplified and cloned into reporter constructs which then are functionally analyzed in the SMP DNA projects "Functional Promoter Analysis" and "Functional Promoter Analysis in Mice" as well as other SMPs and Disease-oriented Genome Networks.

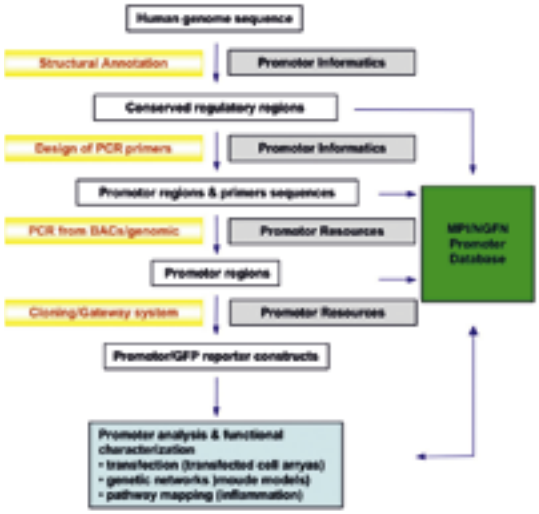


Fig. 1: Overview of the promoter pipeline and the inter-action with other SMP DNA projects.

PROJECT STATUS/RESULTS

The SMP DNA project "Promoter Informatics" has developed a central data resource, an analysis pipeline and a user-friendly interface for promoter related data and annotations (http://promotion.molgen.mpg.de), in order to accommodate the information flow from experimental promoter studies and computational predictions (Fig. 2).

Fig. 2: A graphical outline of the annotation of the IFNAR1 promoter region on HSA21 (from http://promotion.molgen.mpg.de).

This platform enabled us to contribute to the experimental efforts of our NGFN-partners in several ways: (I) *Data integration*: Different data sets were combined (CAGE-tags and ESTs) to more accurately predict the transcription start sites (TSS). For 10% of all genes this amounts to a re-definition of the best TSS compared to ENSEMBL-based annotations. (II) *Primer design*: we have designed primers for all genes of the human chromosome 21 (HSA21) and a number of genes which were selected by our NGFN partners and submitted to the analysis pipeline. (III) *Evolutionary conservation*: We have calculated sequence alignments over a wide range of evolutionary distances including species like human, mouse, chicken and fugu, allowing us to identify regions of functional importance. (IV) *Promoter annotation*: We have developed and tested a biophysical framework to predict binding affinities of TFs to promoter regions. This framework can account for the observed affinities of ChIP-chip data more accurately than traditional methods[3]; (V) *Computational improvements*: Through various optimizations we have achieved a significant speed-up of our algorithms so that genome-wide annotations are now feasible within a short time.

In the SMP DNA project "Promoter Resource" we have established protocols for high-throughput amplification and cloning of human and mouse promoter regions from BAC clones and human genomic DNA. The analysis of the obtained PCR products revealed an approximate 80% success rate. BLAST results of the sequences of the cloned promoter regions

COORDINATOR
Professor Hans Lehrach
Max Planck Institute for Molecular Genetics, Berlin
lehrach@molgen.de

showed the successful cloning of most PCR products. The results indicate that we have a well-established promoter pipeline, including promoter annotation, BAC clone selection, primer design, and PCR amplification and cloning procedures. For some genes we could not identify a BAC clone. We have therefore further optimized the amplification from human genomic DNA by initially using a set of about 40 genes involved in inflammation processes, with a success rate of about 80%. After the initial optimization phase, we started with a larger set of about 400 promoter regions from man and mouse and are in the process of cloning a second batch of about 700 promoters from the NGFN target list. The inflammation set has been extended to about 120 promoters. Currently, a set of about 1200 promoter regions have been selected from man and mouse to be cloned as reporter constructs. About 500 reporter constructs (HSA21, inflammation, etc.) are already available as promoter-reporter constructs and are currently being studied in different NGFN projects.

Within the SMP DNA project "Functional Promoter Analysis" a human promoter set for HSA21 genes has been functionally evaluated in different cell lines and different activation stages within the same cell type using transfected cell arrays. Of 180 HSA21 promoters, about 60 were active in Hek 293T cells, examples of active promoters are shown in Fig. 3. Thus, obtained cell type-specific promoter activity patterns were compared with the gene expression data measured by quantitative RT-PCR in order to establish an integrated picture of HSA21 gene activities on the DNA and RNA level. Currently, the set of 50 truncated HSA21 promoter fragments (600 bp) is being evaluated for more precise determination of the crucial regulatory elements determining promoter activity in particular cell types. Initial experiments revealed that about 50% showed no change in activity. However, we observed considerably higher activity of the longer fragments for 5 genes and for 7 genes of the shorter fragments.

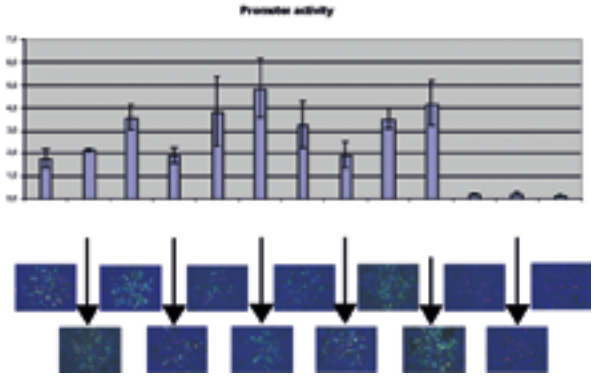


Fig. 3: Examples of active HSA21 promoters in HEK 293T cells using transfected cell arrays.

Within the SMP DNA project "Functional Promoter Analysis in Mice" an efficient system to analyze promoter activity *in vivo* has been successfully established. The integration of reporter constructs in mouse ES cells is based on the cre/lox-driven recombination mediated cassette exchange, followed by tetraploid aggregation to generate fully ES cell-derived transgenic animals. The reporter construct allows fluorescent (YFP) and conventional (lacZ) detection of promoter activity. Its features have been confirmed by analyzing the promoter of 5 mouse genes.[4] Currently, the promoters of 3 TFs on human HSA21 and of 6 mouse genes are being functionally characterized. For several genes, more than 1 construct has been designed to cover distinct regions of interest (alternative TSS, intronic regulatory elements, mutated elements). In total, 21 constructs are under investigation.

OUTLOOK

A more detailed analysis will be performed for the promoter HSA21 candidates showing differential activity. As the next step, we intend to compare the influence of different conditions, i.e. stress induction or transcription activators on the promoter induction and compare it with the correspondent gene expression alterations in the same cell line. The proposed strategy will lead to comprehensive functional evaluation of sequence polymorphisms in promoters regions in respect to their transcription inducing activities. Moreover, application of RNA interference to particular TFs, in combination with our assay, should help in elucidation of the role of individual DNA binding proteins on a genome-wide level. Also the findings for selected promoter features will be confirmed and extended by analyzing them in vivo in the mouse system.

References

[1] *Ptashne M and Gann. Transcriptional activation by recruitment. Nature. 1997 Apr 10; 386(6625):569-77.*

[2] *Pugh BF et al. Genome-wide analysis of protein-DNA interactions in living cells. Genome Biol. 2001; 2(4): Reviews1013. Epub 2001 Apr 4.*

[3] *Roider H et al. (2007). Predicting transcription factor affinities to DNA from a biophysical model. Bioinformatics 12(10): 1523-32.*

[4] *Wittler L et al. (2007). Expression of Msgn1 in the presomitic mesoderm is controlled by synergism of WNT signalling and Tbx6. EMBO Reports AOP, doi:10.1038/sj.embor.7401030.*

UNRAVELING THE EPIGENETIC CODE

MICHAELA SCHANNE[1], MARKUS BEIER[2], JÖRG HOHEISEL[1]

[1] GERMAN CANCER RESEARCH CENTER, HEIDELBERG
[2] FEBIT BIOTECH GMBH, HEIDELBERG

INTRODUCTION

In the human genome, cytosine residues which are immediately followed by a nucleotide with the base guanine, forming a CG dimer sequence usually referred to as CpG, can be modified enzymatically by methylation at their carbon-5 position. DNA methylation plays a key role in genomic imprinting and the maintenance of genome integrity by transcriptional silencing of repetitive DNA sequences and endogenous transposons, thus preserving chromosome conformation and preventing aberrant homologous recombination events. Also, DNA methylation represents an epigenetic mark that regulates the expression of a large number of genes. Epigenetic programming of the genome is therefore a major factor in the interpretation of the basic genetic information. Different cell types, for example, can be distinguished by their cell type-specific epigenetic programs.

Changes in genomic DNA methylation patterns are also an early and consistent hallmark of disease. However, progress in utilization of this information has been hampered by the unavailability of methods for genomic epigenetic profiling. Improving the ability of analyzing a large number of epigenetic changes across the genome was the overall objective of the *Systematic Methodological Platform Epigenetics* (Fig. 1), which also comprises two company partners.

Fig. 1: Locations of the groups participating in the Systematic-Methodological Platform Epigenetics.

The methylation status of a defined nucleotide can be translated into a single nucleotide polymorphism by means of a bisulfite treatment. Unmethylated cytosine is converted into uracil, which turns into thymine during PCR amplification. In contrast, methylated cytosine remains unaffected by the bisulfite treatment. We utilize microarray technology for the analysis of the conversion of individual nucleotides.

PROJECT STATUS/RESULTS

The *Geniom One* system of partner *febit biotech* (Fig. 2) allows an entirely flexible design of each individual microarray at low cost per oligomer – due to light-controlled but mask-free *in situ* synthesis – and sufficient capacity with currently up to 64 000 oligomers per array, split into eight independent channels. At the same time, a synthesis chemistry was established that conforms to biological synthesis direction: the resulting oligomers are attached via their 5'-ends, while the 3'-termini are freely accessible.

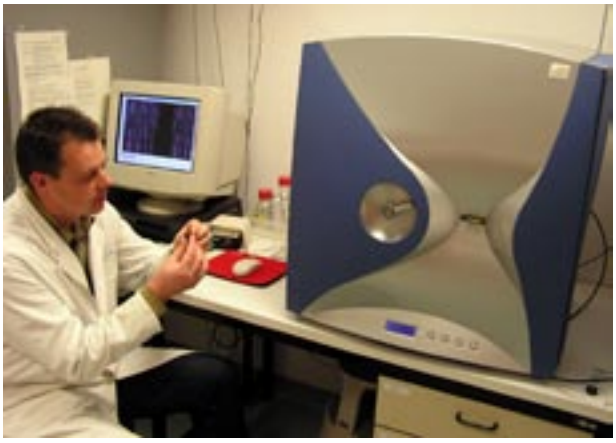


Fig. 2: The *Geniom One* system of the company partner *febit biotech* is used for the analyses.

Initially, we performed hybridization experiments to analyze the sequence variations caused by the bisulfite treatment (e.g., Fig. 3 A). However, this method limited throughput significantly, since a relatively high degree of experimental redundancy was required for accurate measurements. Also, the results were more difficult to interpret.

Within the SMP, we developed a technique for the detection of epigenetic variations by a particular kind of primer extension. The strategy relies on the use of a polymerase for discrimination, since the enzymatic specificity of base-calling is by a factor of 10- to 100-fold better than the detection of mere differences in duplex stability. The oligonucleotides are designed to fit the sequence of the polymorphic site. If a labeled dideoxynucleotide is added, extension occurs only at the oligomer of fully complementary sequence. The detection of cytosine methylation is of reduced complexity, since in either strand only an extension of two oligonucleotides is required, which represent the cytosine or the thymine sequence variant, respectively. In addition to improved dis-

COORDINATOR
Dr. Jörg Hoheisel
German Cancer Research Center, Heidelberg
j.hoheisel@dkfz.de

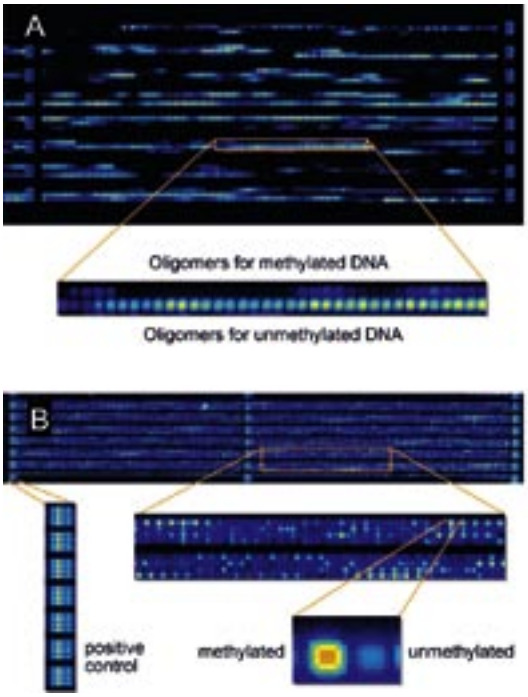


Fig. 3: (A) A typical result of a hybridization-based analysis is shown in the top panel. Upon hybridization of an unmethylated DNA-fragment, binding occurred at the oligonucleotides that represented the unmethylated sequence, while only slight background signals were visible in the row of probes which represented the methylated DNA.(B) The lower panel shows a result of an on-chip primer extension reaction. A reaction took place at the oligonucleotides which represented the methylation status of the respective CpG dinucleotide. Far fewer probes per CpG site are required for an accurate base calling.

crimination and reduced complexity, which in turn increase the number of CpG sites that can be analyzed in a single experiment, the selection of the oligonucleotide sequences is simplified, since the respective duplex stability is a much less critical factor. Finally, detection is possible with the addition of only a single labeled dideoxynucleotide, saving costs and simplifying the read-out.

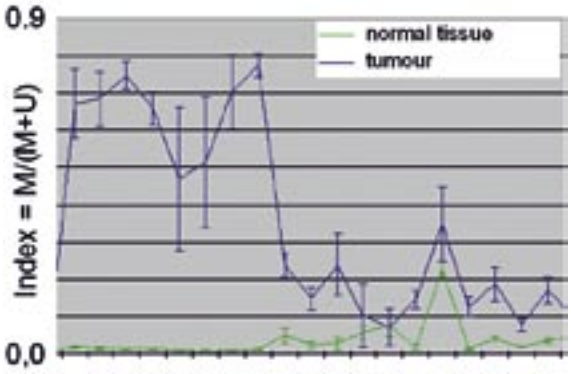


Fig. 4: Methylation status of a CpG-island within a promoter region. While there is a very low degree of methylation in cells of normal tissue, about half the CpG dinucleotides exhibit significantly more methylation if extracted from related tumor tissue.

Using this technology, a large number of individual CpG sites within promoters of interesting genes can be analyzed. We study variations in various cancer tissues, including pancreatic and breast cancer. Besides protein-encoding genes, the promoters of other genes can also be analyzed. We found in many instances that there is a large variation in the methylation status of a promoter, even within individual CpG-islands (e.g., Fig. 4), which are genomic areas with a particularly large number of CpG sites. For diagnosis, it is critical to identify the changes which are highly correlated with the occurrence and stage of disease.

The data obtained from the analyses is evaluated in combination with clinical information and results from transcriptional profiling as well as transcription factor binding studies. Since particularly methylated CpG dimers are known to affect extensively the three-dimensional structure of DNA, a combined analysis of the interaction of such DNA with relevant regulative active proteins is prerequisite for an understanding of regulative effectors. Such analysis permits fundamental insights into the role of DNA methylation during disease development and provides the foundation for an epigenetic classification of disease.

References

[1] *Hoheisel JD (2006). Microarray technology: beyond transcript profiling and genotype analysis. Nature Rev. Genet. 7, 200-210.*

[2] *Beier V et al. (2007). Monitoring methylation changes in cancer. Adv. Biochem. Eng. Biotechnol. 104, Volume: Analytics of Protein-DNA Interactions (Seitz, H., Hrsg.), Springer Verlag, 1-11.*

[3] *Brueckner B et al. (2007). The human let-7a-3 locus contains an epigenetically regulated microRNA gene with oncogenic function. Cancer Res. 67, 1419-1423.*

[4] *Weber B et al. (2007). Methylation of human microRNA genes in normal and neoplastic cells. Cell Cycle 6, 1001-1005.*

[5] *Pullat J et al. (2007). Optimisation of candidate-gene genotyping by flexible oligonucleotide microarrays: analysing variations in immune regulator genes of hay-fever samples. BMC Genomics 8, 282.*

FROM GENE SIGNATURES TO MOLECULAR MECHANISMS OF CANCER

HOLGER SÜLTMANN, RUPRECHT KUNER, TIM BEISSBARTH

DIVISION OF MOLECULAR GENOME ANALYSIS, GERMAN CANCER RESEARCH CENTER, HEIDELBERG

INTRODUCTION

Cellular models are important biological tools to understand distinct steps towards complex disease phenotypes. While many genes have been identified that enhance tumor progression or inhibit tumor growth, the downstream processes associated with the activity of such genes are often unknown. Microarray analysis of global transcription in cells allows to both select disease-relevant genes and to gain insight into their downstream effects. The SMP RNA (http://www.dkfz.de/smp-rna/) combines two technologies, gene intervention in cellular systems by RNA interference (RNAi; collaboration with SMP Cell) and global gene expression profiling, with the development of novel computational biology tools (collaboration with SMP Bioinformatics) in order to construct interaction networks between disease genes.

The goal is to find transcriptomic signatures which are representative for the activity status of biological pathways and disease-relevant processes. Furthermore, we combine several experimental series to unravel novel gene-gene interaction networks that can subsequently be translated to the *in vivo* disease scenario. This strategy could be applied to clinically relevant questions like therapy response or resistance (Fig. 1).

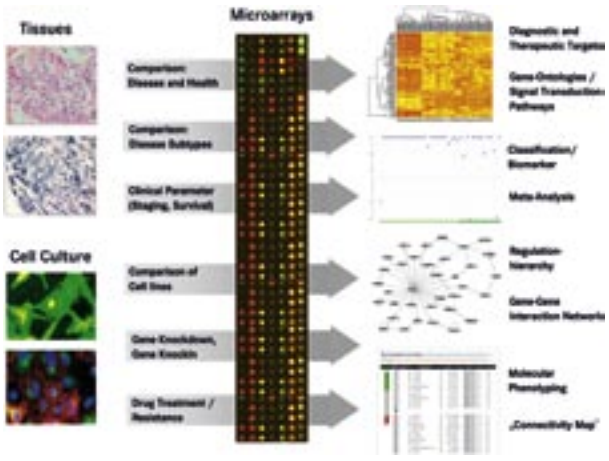


Fig. 1: Use of microarrays to characterize disease genes in clinical studies and *in vitro* systems. To understand the downstream mechanism of target genes, gene intervention experiments are performed in cell lines. The subsequent analysis of global transcriptomic changes helps to unravel the involvement of biological processes and to identify novel gene-gene interaction networks.

Several projects have successfully applied this novel technology described in the following.

PROJECT STATUS/RESULTS

RNAi in MCF7 breast cancer cells

Estrogen receptor expression is of crucial importance for the decision on the therapy of breast cancer patients. Thus, we set out to unravel novel gene-gene interactions associated with ER expression. We performed knock-down experiments via RNAi of several genes we had previously found to be associated with ER status of breast cancer patients. To this end, gene perturbation experiments of 15 genes were performed in ER-positive MCF-7 breast cancer cells. Comparing the transcriptomic changes of individual gene knockdowns upon microarray analysis, we were able to find novel links between the affected genes and downstream cancer-related signal transduction pathways using comprehensive data mining tools.[1]

The observed perturbation effects of each experiment can also be used to reversely engineer interdependencies between gene products on a non-transcriptional level. The observation of the nested structure of significant up or downregulations of affected genes allows us to reconstruct connections of the upstream signaling pathway. In our work we developed algorithms which allow inferring these nested effect structures on a large scale (up to 100 genes) in a robust fashion.[2] Moreover, our methods are able to integrate prior knowledge in a suitable way. In our example, we included known interactions within the ER signaling pathway. Using realistic simulations, we showed that our algorithms are able to infer network structures with an accuracy of over 80%. We applied our approach to reversely engineer the signaling pathway interactions between 10 genes in the ER network of MCF-7 breast cancer cells and found our reconstruction to be insensitive against variations in the data. Several links could be confirmed by prior knowledge. Novel network structures indicate a high relevance of the target genes that have not been described to be related to ER signaling in the past.

RNAi in HPV transformed cervical cancer cells

Some cancer diseases are very well characterized with respect to risk factors and initial causal events. Specific types of human papillomaviruses (HPVs) cause cervical cancer. Both cellular transformation and the maintenance of the oncogenic phenotype of HPV-positive tumor cells are linked to the expression of the viral oncogenes E6 and E7. The aim of this project was to systematically identify primary cellular target genes by performing global gene expression profiling upon siRNA-based E6 and E7 inhibition.[3] We compared the E6/E7-driven gene

COORDINATOR
Professor Annemarie Poustka
German Cancer Research Center, Heidelberg
a.poustka@dkfz.de

signature with independent microarray studies of HPV-positive cervical cancer biopsies and found a substantial overlap between the proposed early events of our HPV *in vitro* model and the accumulated changes during cervical cancer progression. In addition to established changes in cancer-relevant biological processes like apoptosis control, cell cycle regulation or spindle formation, our results indicate the early involvement of a MYC-associated gene-gene interaction network relevant for RNA processing and RNA splicing (Fig. 2).

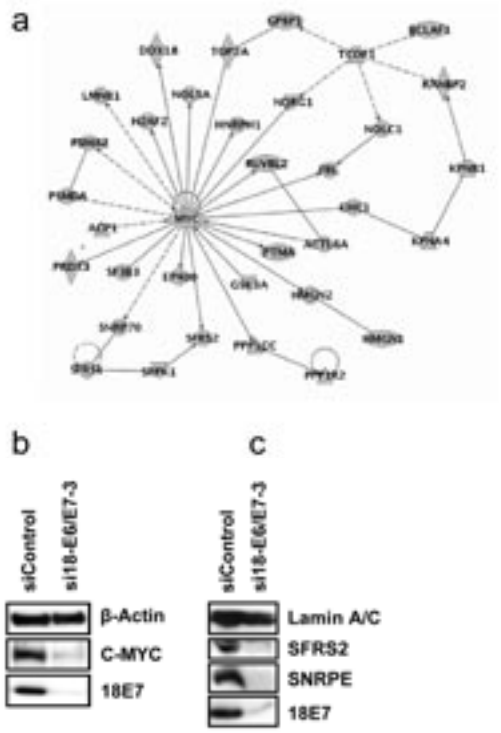


Fig. 2: (a) Mining of global transcriptomic changes upon siRNA-based E6 and E7 inhibition in HPV transformed HeLa cells resulted in good concordance to the *in vivo* condition and highlights a MYC-related gene interaction network associated with RNA processing and splicing; (b) and (c) Western Blot analysis of several proteins confirmed the downregulation upon inhibition of E6 and E7 oncogenes.

microRNA expression and lung cancer

MicroRNAs represent the native cellular complements of artificial siRNA molecules for transcriptional repression. It has been suggested that altered expression of members of the let-7 miRNA gene family might contribute to human tumorigenesis. The cellular target genes of specific miRNAs are widely unknown or only hypothetical. While overexpression of let-7a-3 miRNA in the lung cancer cell line A549 resulted in an increased anchorage-independent growth rate, the aim was to identify cellular target genes that help to unravel the molecular mechanism behind this oncogenic phenotype.

Using the microarray technology, we identified about 200 differentially expressed genes after increased let-7a-3 expression.[4] Data mining of this signature revealed over-representation of genes involved in cell proliferation, adhesion, and differentiation. In line with the suggested role of let-7 in the regulation of RAS, we found 19 RAS-responsive genes to be deregulated in L7-A549 cells, but also potential downstream targets that are already described to promote lung cancer progression.

OUTLOOK

The measurement of genome-wide transcriptomic changes in *in vitro* and *in vivo* models helps to enhance the knowledge about the molecular mechanism behind the pathogenesis of complex diseases. In particular, the translation of *in vitro* generated data into *in vivo* conditions holds great promise for understanding the relevance of single events in advanced-stage phenotypes. The combination of disease-relevant target identification, specific cellular models for target intervention and subsequent genome-wide microarray screening is highly useful to assess the disease-relevance of potential target genes and the associated biological processes. The integration of data from several gene intervention experiments focusing on a common disease-relevant pathway allows generating novel hypotheses about gene-gene interaction networks and regulation hierarchy. These findings will be useful to understand the progression, prognosis and potential therapy response in complex diseases.

References

[1] *Frohlich H et al. GOSim -- an R-package for computation of information theoretic GO similarities between terms and gene products. BMC Bioinformatics (2007) 22; 8:166.*

[2] *Fröhlich H et al. Estimating Large Scale Signaling Networks through Nested Effects Models from Intervention Effects in Microarray Data, Proc. German Conf. On Bioinformatics (2007) in press.*

[3] *Kuner R et al. Identification of cellular targets for the human papillomavirus E6 and E7 oncogenes by RNA interference and transcriptome analyses. J Mol Med. (2007).*

[4] *Brueckner B et al. The human let-7a-3 locus contains an epigenetically regulated microRNA gene with oncogenic function. Cancer Res. (2007) 15;67(4):1419-23.*

GENE EXPRESSION SIGNATURES FOR DISEASE DIAGNOSIS AND PROGNOSIS

HOLGER SÜLTMANN, RUPRECHT KUNER

DIVISION OF MOLECULAR GENOME ANALYSIS, GERMAN CANCER RESEARCH CENTER, HEIDELBERG

INTRODUCTION

Within less than 10 years, gene expression profiling with microarrays has become a key technology in biomedical research. Gene expression analyses of large numbers of patient samples have provided significant advances for the classification of diseases and the identification of target genes relevant to human disease. Microarray studies have also made important contributions to understanding the regulation of biological networks and gene functions. This knowledge is fundamental to the elucidation of the entire network of molecular changes in diseases, and an absolute requirement for the knowledge-driven medicine of the future.

The identification of diagnostic and prognostic gene expression signatures in diseases is a major focus of the SMP RNA; http://www.dkfz.de/smp-rna/. Gene expression profiles were detected in numerous preclinical studies, including breast and prostate cancer as well as heart diseases.

PROJECT STATUS/RESULTS

Diagnostic and prognostic signatures in breast cancer

Breast cancer is the most frequent type of malignancy in women with drastically increased incidence over the past 20 years. Yet, accurate diagnostic and prognostic markers are lacking, and new diagnostic and therapeutic markers are urgently needed.

The division of Molecular Genome Analysis at the DKFZ has analyzed 180 microdissected breast cancer samples using breast cancer-specific microarrays containing 8200 clones. These analyses revealed gene expression signatures associated with tumor type and tumor grading, estrogen receptor (ESR1) expression and morphological parameters. In particular, the association of 56 tumors with the respective ESR1 status yielded a gene expression signature which comprises only 10 genes (Fig. 1).[1] With the exception of five samples, this signature was largely consistent with ESR1 diagnosis in the histopathology laboratory. After revision of the immunohistochemical data, all five samples turned out to be ESR1-positive. Thus, the molecular signature suggested false classified samples which had initially escaped ESR1 expression detection, and the patients might have been susceptible to ESR1 inhibitory treatment if the signature had been available at the time of diagnosis.

In the next step, the 10-gene ESR1 signature was applied to data from previously published microarray studies on ESR1 expression in breast cancer. This “meta-analysis” of 257 samples from three different array platforms turned out to be better than the classifications in the original sample sets. Thus, the usage of the molecular ESR1 signature is superior to currently used ESR1 histopathology.

Identification of genomic amplifications from gene expression array data

In cancer, chromosomal imbalances (amplifications, deletions) influence the transcriptional activity. These alterations are often not limited to single genes but affect several genes of a genomic region and may be relevant for the disease status. For example, the ERBB2 amplicon (17q21) in breast cancer is associated with poor patient prognosis.

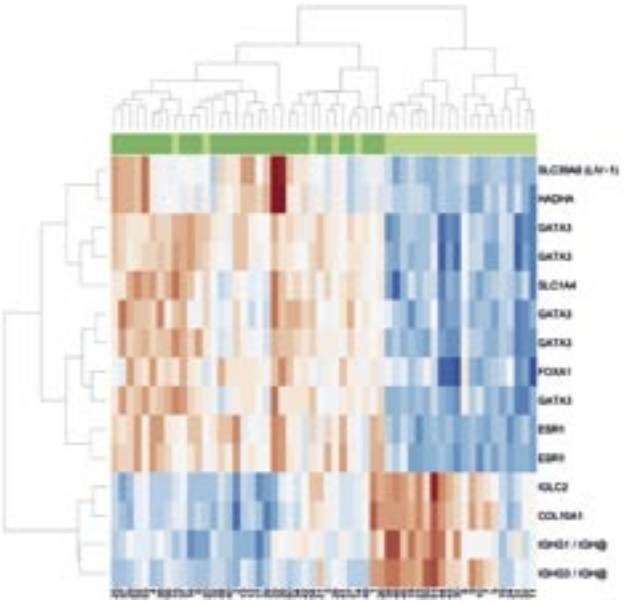


Fig. 1: Diagnostic gene expression profile for estrogen receptor (ESR1) status in breast cancer. The different ESR1 types (light and dark green) can be distinguished by only 10 genes (right).

In order to maximally exploit gene expression microarray data, we developed an unsupervised method to systematically detect chromosomal amplification regions based on their distinct transcriptional activities.[2] We applied our approach to 12 independent human breast cancer microarray studies comprising 1422 tumor samples. We prioritized chromosomal regions and genes predominantly found across all studies. The result highlighted regions which are well known to

COORDINATOR
Professor Annemarie Poustka
German Cancer Research Center, Heidelberg
a.poustka@dkfz.de

be amplified like 17q21 and 11q13, but also others like 8q24 (distal to MYC) and 17q24-q25, which may harbor novel putative oncogenes. Since our approach can be applied to any microarray study, it may become a valuable tool for the exploration of transcriptional changes in diverse disease types.

Diagnostic signatures in prostate cancer

Prostate cancer is the most frequent tumor type in males and the second most frequent cause of male death due to malignancy. Currently available methods do not allow for a sufficiently precise prediction of the natural development of individual prostate. The high sensitivity of the currently used serum marker prostate specific antigen PSA (90%) is accompanied by a low specificity (35%), causing a large number of unnecessary biopsies. Thus, the need for functionally validated novel targets as well as a better understanding of critical markers in PCa formation and progression is large.

Using genome-wide cDNA microarrays, we aimed at identifying gene expression differences between samples of microdissected prostate tumor and normal tissue[3]. We found that a large number of genes were significantly differentially expressed between both sample classes. Among these were several genes known to be associated with prostate cancer as well as novel genes identified. A functional analysis of the latter using RNA interference in PC3 prostate cancer cell lines suggested that several of the selected genes are involved in tumor-related biological processes.

A diagnostic signature for heart failure

Not only tumors are examined in the projects of the SMP RNA. Another project was designed to identify a common gene expression signature in dilated cardiomyopathy (DCM) across different microarray studies. DCM is a major cause of heart failure in Western countries. Several gene expression studies were performed in this disease. However, differences in platform technologies, tissue heterogeneity and small sample sizes obscured the underlying pathophysiological events as yet. As a consequence, the interpretation of previous microarray studies in heart failure was inconclusive.

We accounted for tissue heterogeneity and technical aspects by performing two genome-wide expression studies based on cDNA and short-oligonucleotide microarray platforms[4], which comprised independent septal and left ventricular tissue samples from non-failing (NF, n=20) and DCM hearts (n=20).

Concordant results emerged for the predominant functional classes of deregulated genes between cDNA and oligonucleotide microarrays. Notably, immune response processes displayed the most pronounced downregulation on both microarray platforms, linking this process to the pathogenesis of end-stage DCM. Furthermore, a robust set of 27 genes was identified, which classified DCM and NF samples with >90% accuracy in a total of 108 myocardial samples from our cDNA and oligonucleotide microarray studies as well as two publicly available datasets.

In summary, for the first time, independent microarray datasets pointed to significant involvement of immune response processes in end-stage DCM. Moreover, based on four independent microarray datasets, we identified a robust gene expression signature of DCM, encouraging future prospective studies for the implementation of disease biomarkers in the management of patients with heart failure.

OUTLOOK

The SMP RNA in NGFN-2 has generated a comprehensive data on differentially expressed genes in human diseases. The integration of strong bioinformatics know-how effective in immediate collaboration with the experimental projects has proven to be highly successful. In collaboration with other NGFN partners (e.g. SMP Cell), the gene expression data give rise to cellular functional analyses and proteomics assays in order to understand disease pathomechanisms. The challenge for the future is to employ these data for the translation into the clinical setting.

References

[1] *Schneider J et al. A minimal gene expression signature for the accurate diagnosis of estrogen receptor status in breast cancer. Int J Canc 119(12), 2974-2979, 2006.*

[2] *Buneß A et al. Identification of aberrant chromosomal regions from gene expression microarray studies applied to human breast cancer. Bioinformatics, Jun 28, 2007 [Epub ahead of print].*

[3] *Schlomm T et al. Extraction and processing of high quality RNA from impalpable and macroscopically invisible prostate cancer for microarray gene expression analysis. Int J Oncol 27(3), 713-720, 2005.*

[4] *Barth AS et al. Dilated Cardiomyopathy: Identification of shared functional gene classes and a robust predictive gene expression set across independent microarray platforms in human myocardium. J Am Coll Cardiol 48(8), 1610-1617, 2006.*

THE "esi WAY" FOR TARGET SPECIFICITY IN RNAi

MIRKO THEIS, FRANK BUCHHOLZ

MAX PLANCK INSTITUTE OF MOLECULAR CELL BIOLOGY AND GENETICS, DRESDEN

INTRODUCTION

RNA interference (RNAi) is a powerful tool for loss of gene function studies through the knockdown of specific mRNA transcripts.[1] RNAi is induced by long double-stranded (ds)RNA that is homologous to the target gene. Introduced into cells, this dsRNA is chopped into shorter fragments by an endoribonuclease of the RNase III family (Dicer).

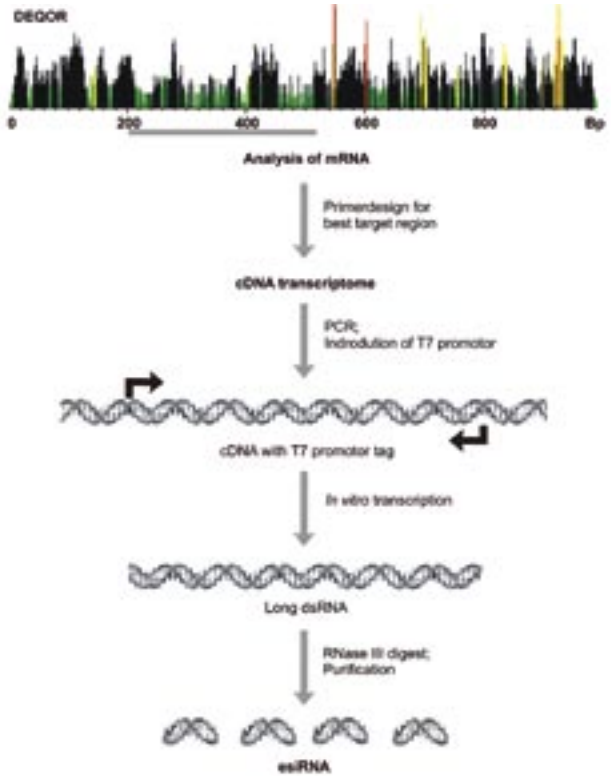


Fig. 1: Flow chart of esiRNA production

These fragments of approximately 21 base pairs length are bound by the RNA-induced silencing complex (RISC) and separated into single strands, where one strand remains bound by RISC. This strand serves as the template for the recognition of the corresponding mRNA. Once a target mRNA is recognized, the protein Argonaut, which is a component of the RISC, now cleaves the mRNA and initiates its further degradation. The path from gene to protein is abrogated at the mRNA level.

Shortly after its discovery, RNAi became a widely used technology for loss-of-function studies in invertebrates like the nematode *Caenorhabditis elegans* or the fruit fly *Drosophila melanogaster*. In most mammalian cells, however, long dsRNA induces a potent and often detrimental interferon response. Hence, the approach of delivering long dsRNA is not useful for studying the specific function of individual genes in most mammalian cells.

This problem was solved in an elegant way in the year 2001[2]: Short double-stranded RNAs – similar to the Dicer products – do not induce an interferon response but mediate gene-silencing in mammalian cells. Importantly, such short interfering RNAs (siRNAs) are easily accessible through chemical solid phase synthesis, making RNAi experiments in mammalian cells straightforward.

PROJECT STATUS/RESULTS

Design of siRNAs

A key step for good siRNA design is the determination of the best target region in the mRNA sequence. Since the target regions differ in their susceptibility for RNAi, the corresponding siRNAs differ largely in their knock-down efficacies. A high-quality siRNA should therefore efficiently destroy the intended target mRNA.

An equally important criterion for a high-quality siRNA is its specificity. It is now well established that siRNAs also silence other transcripts alongside their target genes. These so-called off-target effects arise mostly from partial homologies of the siRNA sequence to other mRNAs. Consequently, observed phenotypes can be due to the knockdown of the intended target, or due to other, unintended silenced transcripts. This explains why off-target effects are a central challenge in the RNAi field, especially in high-throughput screening.

Some improvements to increase the specificity of siRNAs have recently been made, e.g. by avoiding the 3'-untranslated regions, but a design algorithm that can faithfully predict and prevent off-target effects is presently not available.

An alternative concept to the chemical synthesis of siRNAs is the enzymatic digestion of long double-stranded RNAs *in vitro*[3]. In this case a cDNA template is amplified by PCR and tagged with two T7 promoter sequences. T7-RNA polymerase is then used to generate long double-stranded RNA that is homologous to the target gene cDNA. This RNA is subsequently digested with RNase III from *Escherichia coli* to generated short overlapping fragments of siRNAs with a length between 18-25 base pairs. This complex mixture of short double-stranded RNAs is similar to the mixture generated by Dicer cleavage *in vivo* and is therefore called endoribonuclease-prepared siRNA or short esiRNA.

COORDINATOR
Professor Wolfgang Wurst
GSF – National Research Center for Environment and Health, Neuherberg
wurst@gsf.de

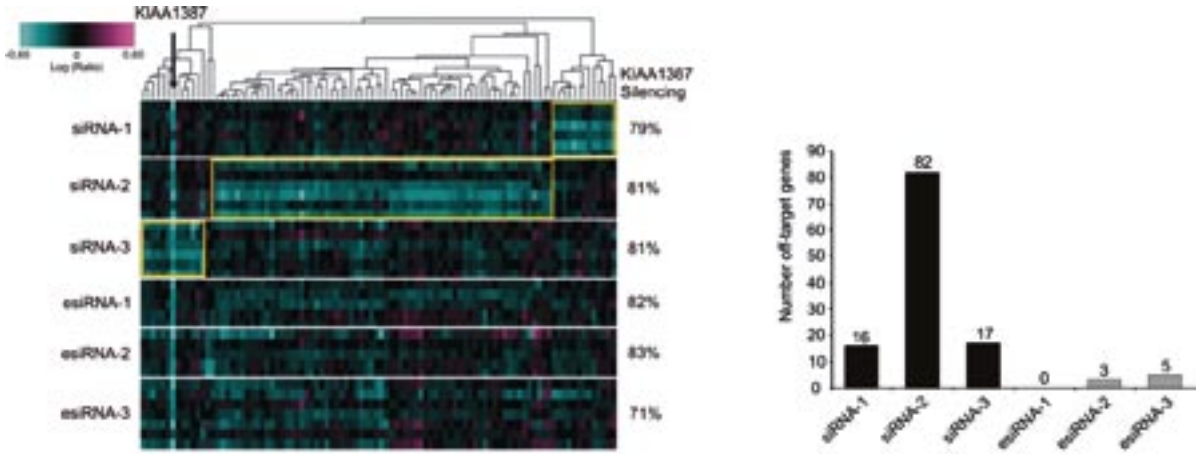


Fig. 2: Microarray analysis of the specificity of esiRNAs compared to siRNA

Off-target effects

For the production of a high-quality esiRNA a 300-600 base-pair-long region in the target mRNA is selected that has a high susceptibility for RNAi without homology to other genes of the same species. For the automatic localization of the best target region we have developed the algorithm DEQOR (Design and Quality Control of RNAi). This algorithm (http://cluster-1.mpi-cbg.de/Deqor/deqor.html) initially determines every possible 21mer in a given mRNA. Based on the quality constraints for siRNA design, each of these 21mer sequences is assigned an efficacy and specificity score. On the basis of this data the region with the highest percentage of good quality silencers is determined. An esiRNA synthesized from a cDNA representing the DEQOR-predicted region will therefore contain a high percentage of good quality siRNAs[4] (Fig. 1). In order to evaluate the performance of esiRNAs, we compared the silencing efficacy and specificity of DEQOR-optimized esiRNAs with chemically synthesized siRNAs using microarray expression analysis and quantitative real-time PCR. These studies revealed a similar silencing efficacy of siRNAs and esiRNAs, making them equally suitable for efficient transcript silencing. In contrast, esiRNAs showed a more than 10-fold higher target specificity in comparison to individual siRNAs[4] (Fig. 2), making esiRNAs the silencing trigger of choice for transcript silencing with minimal off-target effects.

A possible explanation for the increased specificity lies in the complexity of the esiRNA pool. The high numbers of different siRNAs in the mixture share the same on-target signature but differ in their sequence-dependent off-target signatures. As a consequence, many individual siRNAs add to the overall silencing of the target, while off-target effects are diluted out.

High-throughput production of esiRNAs

The generation of esiRNAs can be adapted to high-throughput protocols. Through funding by NGFN-2, a production pipeline for esiRNAs has been established at the Max Planck Institute for Molecular Cell Biology and Genetics in Dresden. This pipeline allows the rapid generation of genome-wide libraries of esiRNAs at relatively low costs.

A human genome-scale library with more than 16000 esiRNAs has been prepared[4] within the NGFN-2 funding period and a number of NGFN partner groups are already using this valuable resource. The target region of each esiRNA and all primer sequences are publicly available via the RiDDLE database (http://cluster-12.mpi-cbg.de/cgi-bin/riddle/search), and detailed protocols for how to prepare esiRNAs have been published[5], providing the information to other groups who want to generate their own esiRNA libraries.

OUTLOOK

In the near future we are planning to complete the human esiRNA library and build additional libraries for other model organisms like *Mus musculus* or *Rattus norvegicus*. These libraries will likely present the most specific mammalian RNAi resource available and hence will aid many research groups in carrying out high-quality large-scale loss-of-function studies.

References

[1] *Fire A et al. (1998) Nature 391, 806-811.*

[2] *Elbashir SM et al. (2001) Nature 411, 494-498.*

[3] *Yang D et al. (2002) Proc. Natl. Acad. Sci.USA 99, 9942-9947.*

[4] *Kittler R et al. (2007) Nat. Methods 4, 337-344.*

[5] *Kittler R et al. (2005) Nat. Methods 2, 779-784.*

RNA INTERFERENCE TECHNOLOGY AS NEW TOOL TO GENERATE MOUSE DISEASE MODELS

RALF KÜHN, WOLFGANG WURST

INSTITUTE OF DEVELOPMENTAL GENETICS, GSF – NATIONAL RESEARCH CENTER FOR ENVIRONMENT AND HEALTH, MUNICH-NEUHERBERG

INTRODUCTION

Silencing of gene expression by RNA interference (RNAi) has become a powerful tool for functional genomics in mammalian cells. RNAi is a sequence-specific gene-silencing process that occurs at the messenger RNA (mRNA) level. In invertebrate cells, long double-stranded RNAs (dsRNA), which are processed into short interfering RNAs (siRNA) by the ribonuclease Dicer, induce efficient and specific gene silencing. In this sequence-guided process, the siRNA antisense strand serves as a template for the RNA-induced silencing complex (RISC). RISC recognizes and cleaves the complementary mRNA, which is then rapidly degraded (Fig. 1).[1] In mammalian cells, long dsRNAs (> 30 bp) elicit an interferon response resulting in the global inhibition of protein synthesis and non-specific mRNA degradation. However, it has been shown that short synthetic dsRNAs trigger the specific knockdown of mRNAs in mammalian cells without interferon activation, if their length is below 30 bp. Such synthetic siRNAs can be easily introduced into cultured cells and induce a transient knockdown that enables the study of mammalian gene function within a short time frame. Due to advances made in the delivery and design of siRNAs, gene silencing has now developed into a routine method for *in vitro* use. Shortly after the establishment of siRNA-mediated transient gene silencing, DNA-based expression vectors were developed that allow the endogenous production of small dsRNAs in mammalian cells. The vector-derived transcripts are designed to contain a sense and an antisense region that is complementary to a selected mRNA segment. These transcripts can fold back into a stem-loop structure and form short hairpin RNAs (shRNAs) that are processed by Dicer in a similar way as the siRNAs. Since shRNA expression vectors can be stably integrated into the genome, they allow permanent, long-lasting gene silencing in cell lines and organisms.

Soon after these technologies were introduced for use in cultured cells, it became an obvious task to explore RNAi-mediated gene silencing in mice as well.[2] Within the framework of SMP RNAi we could successfully develop a shRNA-based technology that enables us to study the function of disease genes in adult mice.

PROJECT STATUS/RESULTS

Within the SMP RNAi *in vivo* section we developed RNAi into a standardized, scalable and time-saving procedure for gene silencing in mice. This approach enables us to produce phenotypic knockdown mutants on a larger scale and within less time as presently achieved with gene knockout methodology. To this end, we developed an advanced strategy for complete or conditional gene knockdown in mice, where the Cre/loxP system is used to activate RNAi in a time- and tissue-dependent manner.

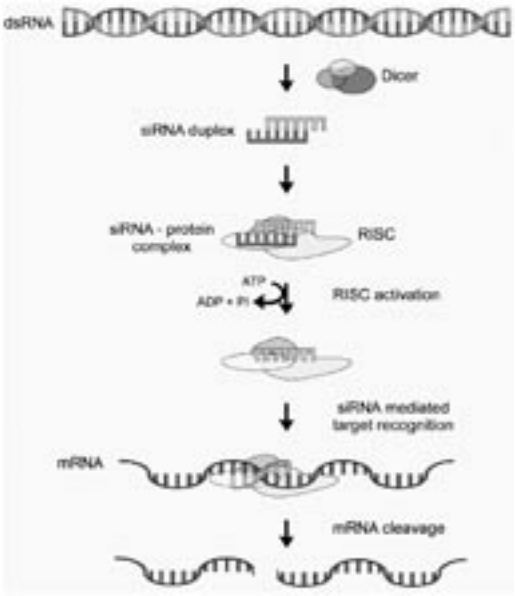


Fig.1: The RNAi Pathway

The site-specific insertion of single copy shRNA vectors allows the expedition and reproducible production of knockdown mice and provides an easy and fast approach for assessing gene function *in vivo*. The basic principle shown in Fig. 2 is based on the insertion of shRNA vectors into the *Rosa26* locus of ES cells by recombinase-mediated cassette exchange and the subsequent generation of chimeric mice. Using the same set of tools, either constitutive or conditional Cre/loxP regulated knockdown mice can be generated.

For our ES cell-based approach, shRNA expression cassettes are inserted into the Rosa26 locus as defined genomic location through recombinase-mediated cassette exchange (RMCE). We derive ES cell lines that contain recognition sites for the site-specific recombinase C31 integrase. One Rosa26 allele harbors

COORDINATOR
Professor Wolfgang Wurst
GSF – National Research Center for Environment and Health, Neuherberg
wurst@gsf.de

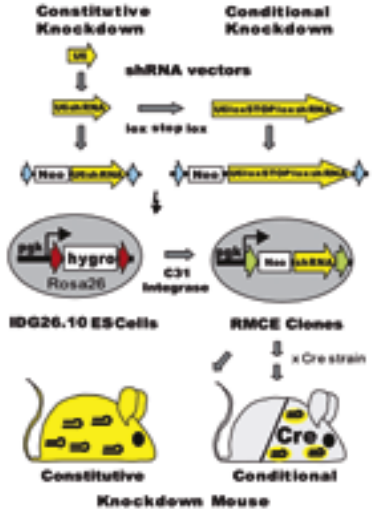


Fig.2: Generation of shRNA transgenic Mice by recombinase-mediated cassette exchange (RMCE) in ES Cells.

Constitutive Knockdown Mice
CRHR1
LRRK2
P2X7
Conditional Knockdown Mice
Braf
Mek1
Mek2
GSK3β
GSK3α
PIK3ca
PIK3cb
BMPRIIa

Table 1: Target genes for which constitutive or conditional knockdown mouse strain have been established.

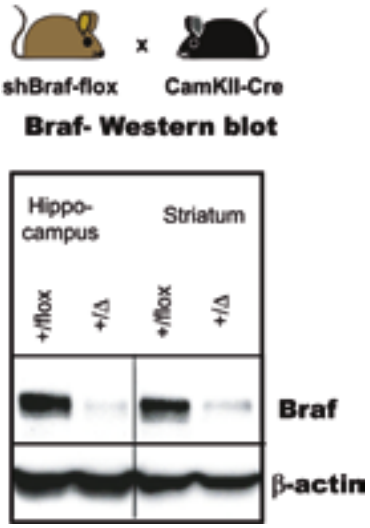


Fig.3: Conditional knockdown of Braf in forebrain regions of shBraf/CamKII-Cre double transgenic mice.

the coding region of the hygromycin resistance gene flanked by a pair of attP recognition sites. These ES cells are transfected with an exogenous donor vector that contains a promoterless neo resistance linked to a shRNA expression vector together with a recombinase expression plasmid. Upon selection for neo resistance about half of the surviving clones contain a RMCE event (Fig. 2). Thus, only few ES cell colonies are required to obtain several clones that contain a shRNA construct of chosen specificity. Subsequently, the modified ES cells are introduced into diploid blastocysts to generate germ line chimeras or aggregated with tetraploid embryos to produce completely ES cell derived mice. The shRNA transgenic offspring can be either directly used for phenotypic analysis or are further crossed to a Cre transgenic strain to activate conditional shRNA vectors.

We successfully applied this approach to more than 10 different target genes and generated either constitutive or conditional knockdown mice (Table 1). These genes code for signal transduction molecules whose function has not yet been characterized in the adult brain (e.g. Braf, GSK3) or resemble Parkinson disease and depression candidate genes identified in human patient cohorts (LRRK2, P2X7). A representative example of the knockdown of a target protein in the mouse brain is shown in Fig. 3. In this case, a Braf specific shRNA vector was activated in neurons of the forebrain by crossing conditional Braf shRNA mice to a second transgenic strain expressing Cre recombinase under the control of the neuron specific CamKII promoter. In double transgenic mice (Fig.3, +/Δ) the shRNA vector becomes activated in forebrain regions like the hippocampal formation and striatum, leading to the knockdown of Braf protein. As shown by Western blot analysis using a Braf specific antibody (Fig. 3), we found a strong reduction of the levels of the 96 kD Braf protein in brain extracts from Braf knockdown mice (+/Δ) as compared to wild-type controls (+/flox). As loading control, the same Western blot was developed with a beta-actin specific antibody.

The Braf knockdown mice as well as most of the other strains are currently being phenotypically analyzed for behavioral alterations. In particular, we focus on anxiety and depression-like behavior that can be well assessed in mice using specific test procedures, like the elevated plus maze (anxiety) and the forced swim test (depression). As first preliminary result of the ongoing studies we found that male knockdown mice for Mek1/2 exhibit altered anxiety behavior.

OUTLOOK

Within this project we demonstrated that transgenic RNAi can be used as a powerful tool to downregulate specific proteins in the brain of adult mice. We developed a streamlined production procedure that enables us to generate knockdown mice in shorter time and with less effort as compared to knockout mice.

Since RNAi targets the mRNA in a reversible fashion, the upcoming generation of RNAi technology will focus on the establishment of inducible knockdown systems in adult mice that allow to switch genes on and off in a reversible fashion.

References

[1] *Dykxhoorn DM et al. (2003). Killing the messenger: short RNAs that silence gene expression. Nat Rev Mol Cell Biol 4(6): 457-67.*

[2] *Kühn R et al. (2007) RNA interference in mice. Handbook of Experimental Pharmacology, Volume 178: Conditional Mutagenesis: An Approach to Disease Models. p. 147-174, Springer Verlag.*

[3] *Hitz C et al. (2007). Conditional brain-specific knockdown of MAPK using Cre/loxP regulated RNA interference. Nucleic Acids Research PMID: 1758681.*

COMBINATORIAL RNAi FOR QUANTITATIVE PROTEIN NETWORK ANALYSIS

STEFAN WIEMANN, DORIT ARLT, ULRIKE KORF, ANNEMARIE POUSTKA
DIVISION OF MOLECULAR GENOME ANALYSIS, GERMAN CANCER RESEARCH CENTER, HEIDELBERG

INTRODUCTION

The fate of cells in an organism is substantially regulated via a number of signaling pathways. Until now these pathways and their impact on disease processes have mostly been studied in isolation. However, they comprise complex networks that interfere and influence one another. This has direct implications on the efficacy of drugs that target individual proteins in this circuitry. To reach a comprehensive understanding of the routes and cross-talk between individual pathways, novel technological and experimental approaches are required. SMP Cell follows a functional profiling strategy[1], (PI: Stefan Wiemann, Division of Molecular Genome Analysis at DKFZ) where high-throughput cellular assays for hit identification and validation are carried out in order to improve our understanding of how proteins impact disease-relevant processes.

This strategy has recently been extended to the "combinatorial RNAi-strategy" for the systematic knockdown of several proteins in parallel[2] (PI: Dorit Arlt, Division of Molecular Genome Analysis at DKFZ), allowing to assess the impact these proteins and the respective pathways have within signaling networks. Single, double, and triple knockdowns were carried out (Figure 1 A, B), and the cells were tested for phenotypic changes in an *in vitro* invasion assay (Figure 1 C). Results were integrated with quantitative data on residual mRNA and protein levels of the targeted proteins to achieve a systems biological understanding in signaling research. To this end, we applied the quantitative protein array technology to monitor the endogenous levels of targeted proteins.

Our results demonstrate that the parallel knockdown of at least three different proteins is feasible, while the silencing of untargeted genes and proteins as well as cytotoxicity can be kept at low levels. We have validated this new approach by investigating the cross-talk of the tyrosine kinase receptor ErbB2 with its downstream targets Akt-1 and MEK1 and the impact on cell invasion.

PROJECT RESULTS

Efficiency of siRNAs in RNAi

Applying siRNA pools (Dharmacon siGenome siRNAs) as well as several individual siRNAs that targeted the respective genes, we first validated the RNAi resources. Residual endogenous levels of mRNA and protein were quantified by Taqman and reverse phase protein arrays (RPPA)[2], respectively. Of twelve individual siRNAs tested, eleven knocked down both

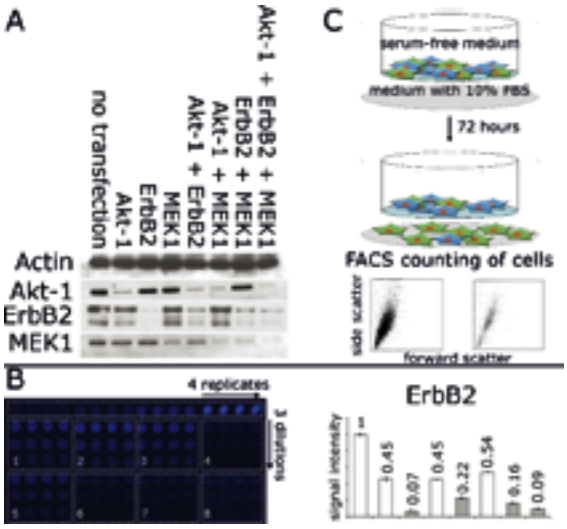


Fig. 1: Akt-1, ErbB2, and MEK1 were knocked down individually or in combinations. A) Western blot analysis of knockdown. B) Quantitative analysis of knockdown via protein array technology, with ErbB2 as an example. Scheme of targets is the same as in the Western blot. C) Unperturbed and perturbed cells were tested in an invasion assay to monitor effects of knockdown on the invasion phenotype. Cells having passed the matrigel-coated membrane were counted by flow cytometry (FACS); reduced numbers of counted cells indicate a phenotypic effect of the respective perturbation.

the targeted mRNA and protein efficiently (>70%). Exploring the minimally required concentration of individual siRNAs, we found a concentration of 20 nmol/L to sufficiently degrade the target mRNAs, and to result also in a substantial reduction at the protein level (Figure 1 A, B). Pools appeared to be superior to silencing with individual siRNAs because off-target and untargeted effects were reduced, as was determined by measuring silencing of actin mRNA and induction of P38 signaling, respectively.

Combinatorial knockdown & RPPA

ErbB2, Akt-1 and MEK1 were targeted via RNAi either individually or in different combinations to test whether the parallel knockdown of several proteins is feasible. Figure 1 A shows the Western blot analysis of typical experiments, demonstrating qualitatively that this is indeed possible. However, Western blots are reported to have standard deviations of up to 30%, impeding a quantitative correlation with phenotypic changes.[3] In contrast, standard deviations in RPPA are commonly <5%, permitting to quantitatively assess the efficiency of knockdown and, most important, the amount of residual targeted proteins. Reverse phase protein arrays have been

COORDINATOR
Dr. Stefan Wiemann
German Cancer Research Center, Heidelberg
s.wiemann@dkfz.de

developed in the Explorative Project "Near infrared nanocrystal-based quantitative protein arrays" (PI: Ulrike Korf, Division of Molecular Genome Analysis at DKFZ) to provide a highly reliable and reproducible technology for the quantitative analysis of proteins and protein modifications.[3] Cell lysates are spotted on glass slides and target proteins are specifically inspected with appropriate antibodies. Detection is in the near infrared and exploits the low level of autofluorescence biomaterials have at these wavelengths. Applying RPPA, we demonstrated that the parallel knockdown of the three proteins did not influence the efficiency of protein depletion (Fig. 1 B), and therefore validated the combinatorial RNAi strategy as a suitable method for functional analysis of phenotypes.

Cell invasion analysis

The highly invasive breast cancer cell line HCC1954 was selected as cell system to test the combinatorial knockdown of the signaling proteins and their potential to reduce the invasive phenotype of that cell line. Applying a standard matrigel assay (Fig. 1 C), we counted the cells that were able to pass the matrigel-coated membrane upon perturbation with the different siRNAs. While Akt-1 had the strongest effect when knocked down either alone or in combination with either of the other proteins, this effect was attenuated when all three proteins were targeted in parallel.

Discussion

We systematically knocked down ErbB2, Akt-1 and MEK1 separately and in combinations to test the impact of the respective signaling pathways on the invasion potential of HCC1954 cells. There, we found attenuating effects that are likely due to cross-talk of the PI3K and ERK1/2 pathways with other and thus far unidentified signaling events. In this study, we validated the combinatorial RNAi strategy to have a huge potential to unravel cross-talk between individual such pathways. The correlation with the quantitative amounts and concentrations of target proteins via RPPA proved indispensable for the analysis and interpretation of the phenotypic effects (i.e. cell invasion). Even subtle differences that had been induced by the various combinations of knocked-down target proteins could be measured and correlated. Hence, the quantitative analysis of cell invasion, as well as the conclusions drawn in Figure 2 became feasible. Our results clearly demonstrate that the connection of PI3K and ERK1/2 signaling is more complex than previously thought, and that more efforts need to be made to unravel the nature and dynamics of cellular signaling networks.

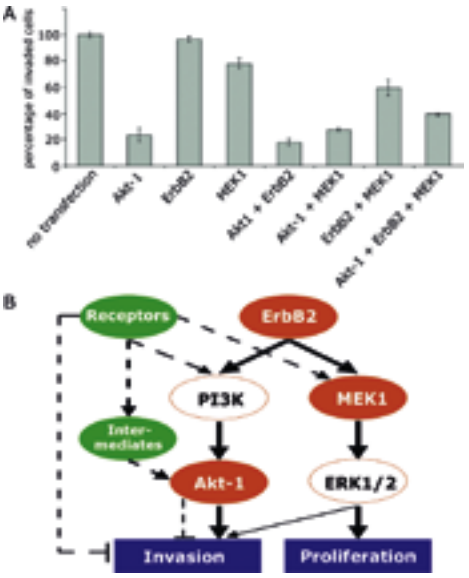


Fig. 2: Impact of signaling molecules on the invasion potential of HCC1954 cells.[2] A) ErbB2, AKT-1 and MEK1 proteins were knocked down individually or in combinations, and the impact on cell invasion was measured. While Akt-1 alone and in combination with ErbB2 was capable of efficiently reducing the invasive capacity of HCC1954 cells, the triple knockdown also with MEK1 had a lesser effect. B) This hints towards an attenuating effect of a thus far not identified pathway (e.g. integrin or Wnt) on the inhibiting function of Akt-1 knockdown (indicated with dashed lines).

OUTLOOK

We have devised and validated the combinatorial RNAi strategy as one novel experimental approach that is suited to drive the analysis of signaling and protein networks towards a systems biological level. Along with the application of reverse phase protein arrays for the time-resolved and quantitative analysis of cell signaling, we aim to extend our focus within NGFN*plus* to achieve a comprehensive understanding of how signaling pathways act and interact in complex networks. This will be the basis to understand drug resistance mechanisms, for example, in breast cancer and to identify novel strategies and target molecules for therapeutic intervention.

References

[1] *Wiemann S et al. From ORFeome to biology: a functional genomics pipeline. Genome Res. (2004) 14:2136-44.*

[2] *Sahin O et al. Combinatorial RNAi strategy: The next generation of quantitative protein network analysis. Proc Natl Acad Sci USA. (2007) 104:6579-84.*

[3] *Loebke C et al. Infrared-based protein detection arrays for quantitative proteomics. Proteomics. (2007) 7:558-64.*

MIACA – THE MINIMUM INFORMATION ABOUT A CELLULAR ASSAY

STEFAN WIEMANN, ALEXANDER MEHRLE, HEIKO ROSENFELDER, ANNEMARIE POUSTKA
DIVISION OF MOLECULAR GENOME ANALYSIS, GERMAN CANCER RESEARCH CENTER, HEIDELBERG

INTRODUCTION

High-throughput cell biological analyses can be triggered via RNAi, or with expression constructs[1], small chemical compounds, ligands for receptors, through environmental stresses (such as temperature shift, serum starvation, oxygen deprivation, etc.) or combinations thereof. The cellular responses to such perturbations are analyzed in order to identify molecular events in the biological processes addressed and to understand biological principles (Fig. 1).[2] Given the availability of such data, there is a growing need to compare and integrate information that originated from different approaches, and to elucidate higher order principles. Minimum information about the rationale, materials, the conditions prior to, during, and after the perturbation must be documented in order to fully describe cellular assay projects and to create a general knowledge base. Only then will an efficient integration of data become possible, allowing researchers to assess the quality and relevance of the gathered data and the deduced information. To this end, the Minimum Information About a Cellular Assay (MIACA) standards initiative[3] (PI: Stefan Wiemann, Division of Molecular Genome Analysis at DKFZ) develops an information guideline and a modular Cellular Assay Object Model (CA-OM) (PI: Heiko Rosenfelder, Division of Molecular Genome Analysis at DKFZ) that is capable of covering the range of cellular assays possible and which is the basis for efficient data exchange. This initiative was pioneered in NGFN-2 SMP Cell and has since been further developed in collaboration with SMPs RNA, RNAi, DNA, Protein, and Bioinformatics as well as with scientists and companies outside of NGFN. The MIACA homepage is http://miaca.sf.net and at http://www.smp-cell.org.

PROJECT STATUS

A number of closely related and complementary communities representing suppliers of genomic and functional genomic resources, experimentalists who perform (large-scale) cell-based assays and/or bioinformatics data analysis, as well as developers of controlled vocabularies and data exchange formats have joined forces[3, 4] to define minimal reporting requirements for experimental details and data from cell-based assays. In addition, they are developing an accompanying data exchange format that shall provide the basis for efficient data dissemination and integration.

This guideline project has been named MIACA, the Minimum Information About a Cellular Assay. A modular Cellular Assay Object Model (CA-OM) is developed in parallel, and is required to flexibly cover a whole range of different cellular assays.

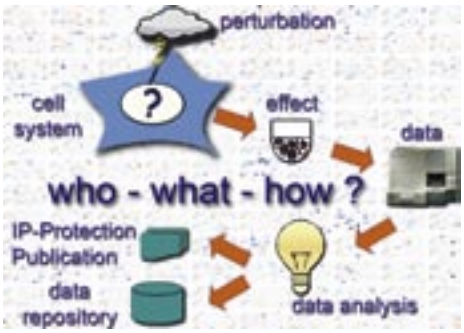


Fig. 1: Information about the rationale, the materials used, as well as details on the experimental and analysis processes of a cellular assay need to be documented and should accompany the final data, in order to permit scientists to review the relevance and quality of the experiment and of the data.

While in principle conditions of any cell-based assays should be covered in sufficient detail, the guideline needs to be flexible enough to allow for customization towards a great variety of applications without compromising the “minimum” information principle. This has been achieved by dividing the experimental work-flow into individual modules of the object model. Common data exchange formats are necessary for standardized transfer of information. A Cellular Assay Markup Language (CA-ML) based on the CA-OM object model is currently under development. MIACA, CA-OM and CA-ML are anticipated to stimulate and support providers of widely accepted data repositories to develop databases that take data from cellular assays.

MIACA structure

The design of the MIACA minimum information guideline comprised the development of three related building blocks. 1.) The MIACA document outlines the individual terms and features that comprise the minimum information required to describe the individual components and processes of a cell-based assay in the necessary depth. 2.) These features have been placed into the modular structure of the MIACA Object Model (CA-OM). The implementation of modules in CA-OM was prerequisite to achieve the level of flexibility that is required to cover the full range of very different cellular assays. 3.) Controlled ontologies are adopted, where possible, from established initiatives (e.g. OBI – http://obi.sf.net), and MIACA contributes to their further development.

COORDINATOR
Dr. Stefan Wiemann
German Cancer Research Center, Heidelberg
s.wiemann@dkfz.de

MIACA minimum information requirements

Terms that define the amount of information that should be sufficient to allow for a reasonably precise description and an assessment of data quality and relevance is outlined in the MIACA text structure. As the level of detail needed for different uses varies, e.g. in LIMS *vs.* data repositories *vs.* publications, these terms are flagged for their appropriate application(s). Controlled vocabularies are used with these terms making MIACA compatible with other community efforts and being a prerequisite for efficient data exchange.

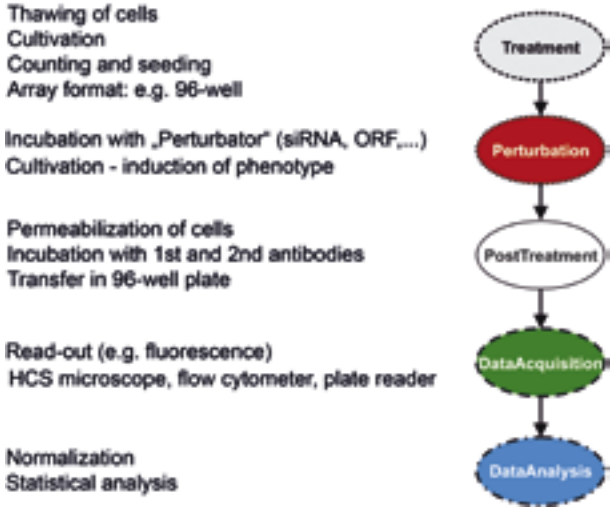


Fig. 2: The structure of the CA-OM separates experimental processes into individual modules.[3] This modular structure has been prerequisite to achieve the flexibility CA-OM has and requires in order to cover the full range of assays possible.

The Cellular Assay Object Model (CA-OM)

The individual reporting terms are structured in different modules (Figure 2). General information about the rationale and concept of the study, details on the laboratory, as well as on materials are given in the "Header" module. The experimental procedures are split into "Treatment", where any manipulation of cells prior to the perturbation is described. Then, the "Perturbation" gives details on how cells are treated in order to induce phenotypic changes. Since very different stresses can be applied to cells ranging from temperature shifts to infection with shRNAs, this module needed to be kept flexible. In many cases, cells are fixed and treated with antibodies prior to data acquisition. Such processes are described in "Posttreatment". The "Data acquisition" module describes the parameters and instrument-settings that are specific for the particular assay. To create hit lists with significant effectors, raw data needs to be processed which is described in the "Data analysis" module. All of these modules can be linked and grouped as needed. Individual modules may be repeated, for example, the "Perturbation" module can be duplicated to reflect a situation where two different perturbations are administered to cells in subsequent steps.

The Cellular Assay Markup Language (CA-ML)

Building on the FuGE environment and experiment models (http://fuge.sf.net), an XML-based markup language has been established to permit efficient data exchange and database integration.

OUTLOOK

While the MIACA initiative already comprises scientists from several communities and continents, the acceptance of this guideline as a common reporting standard requires additional community involvement. To this end, documents (MIACA, CA-OM, CA-ML) are posted and updated at the development page http://sf.net/projects/miaca where they can be downloaded and reviewed. In preparation of common data repositories that should take and disseminate data from cellular assays, initial contacts have been established with EBI (Cambridge, UK) and NCBI (Bethesda, USA). There, the MIACA model is currently under investigation for its potential to serve as basis for such database. In NGFN^{plus} we will enhance our efforts to further improve and to have MIACA become an accepted reporting standard.

In the coming months and years it will be interesting to follow the progress made in this rapidly evolving field of research, where the importance of QM & Standards is increasingly appreciated.

References

[1] *Sauermann M et al. High throughput FACS based assay to identify apoptosis inducing proteins. J Biomol Screening. (2007) 12:510-20.*

[2] *Arlt D et al. Functional Profiling: From Microarrays via Cell-Based Assays to Novel Tumor Relevant Modulators of the Cell Cycle. Cancer Res. (2005) 65:7733-42.*

[3] *Wiemann S et al. MIACA - The Minimum Information About a Cellular Assay, and the Cellular Assay Object Model. Nat Biotechnol. (2007): In public review.*

[4] *Orchard S et al. Proteomics and Beyond A report on the 3(rd) Annual Spring Workshop of the HUPO-PSI 21-23 April 2006, San Francisco, CA, USA. Proteomics. (2006) 6:4439-43.*

PROTEIN COMPLEX COMPOSITION AND FUNCTION IN HEALTH AND DISEASE

MARIE-LAURE FOGERON, HANNAH MÜLLER, VERENA LEHMANN, KARIN HABERMANN, TANJA KURTZ, ANNE-KATHRIN SCHOLZ, STEPHANIE SIEBERT, KARL-HEINZ RAK, JOHAN GOBOM, BODO LANGE

MAX PLANCK INSTITUTE FOR MOLECULAR GENETICS, BERLIN

INTRODUCTION

Most processes in the human body are mediated by protein complexes. Their function is manifold: maintaining cell structure, controlling cell metabolism and regulating cell proliferation and cell differentiation throughout development. In diseases such as cancer, mechanisms that control cell proliferation and apoptosis are frequently defective, leading to excessive cell division and the acquisition of properties that induce cell migration and invasion of other tissues. These changes of cellular properties come about through mutations, through posttranslational modification of proteins or are the consequence of elevated or reduced protein levels when compared to healthy cells. However, to understand the role of such changes in cancer cells, we first have to define the protein-complex interactions and function in normal cells.

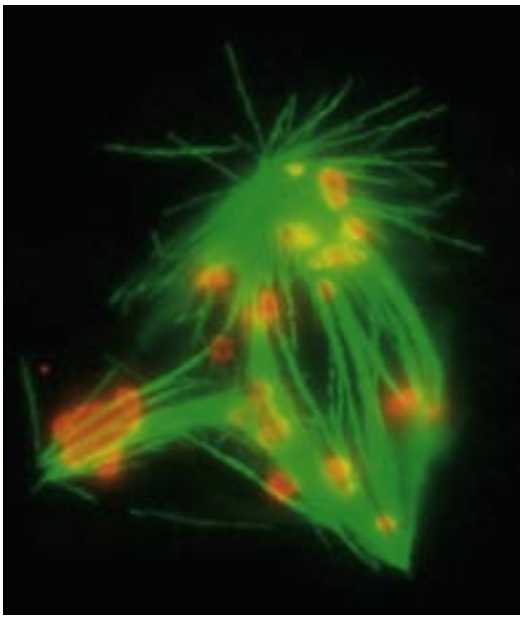


Fig. 1: Abnormal chromosome (red) segregation through the microtubule filaments (green) is caused by depletion of a cell cycle regulatory protein.

We have developed a systematic platform for the characterization of protein complexes in normal and in diseased cells. This approach involves the characterization of protein complexes by expression of bait-proteins in tissue culture cells. These baits are tools to fish out cellular proteins and their natural interactions partners as protein complex. They are isolated through a TAP tag (tandem-affinity purification = TAP tag) that allows affinity-purification and subsequent identification of interaction partners through biophysical and biochemical methods (for example mass spectrometry and Western blotting). After the isolation and identification of protein complexes, we carry out both a functional analysis (Fig. 1) using RNA interference (RNAi) as well as electron microscopy analysis (Fig. 2). This provides us with information on the cellular function of the protein complexes and shows if the isolated proteins are forming an actual complex.

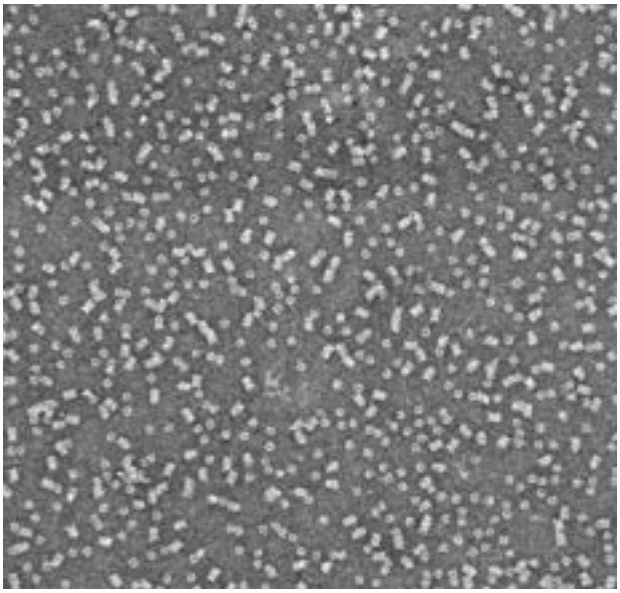


Fig. 2: Electron microscopic image of protein complexes isolated through tandem-affinity purification

PROJECT STATUS/RESULTS

Our research focuses on protein complexes that are involved in the process of cell division (mitosis) and cell proliferation. This process is important because the fidelity of chromosome segregation plays a decisive role for correctly passing on genetic information. It has been suggested that aberrant chromosome segregation (genomic instability) contributes to tumorigenesis and metastasis.

Mitosis is a complicated process with thousands of proteins involved; however, the molecular functioning of most of these proteins is far from clear. We have recently investigated the functional role of a subset of these proteins which

COORDINATOR
Professor Erich Wanker
Max Delbrück Center for Molecular Medicine, Berlin
ewanker@mdc-berlin.de

Fig. 3: Anne-Kathrin Scholz, Karoline Pollok, Bodo Lange, Karin Habermann, Marie-Laure Fogeron, Tanja Kurtz, Verena Lehmann, Stephanie Siebert

constitute the microtubule-organizing center (also called the centrosome), a large protein complex and cell organelle that is involved in the organization of the mitotic spindle in most higher eukaryotic cells. Using an immunoaffinity-based approach combined with mass spectrometry (MS) techniques (in collaboration with the MS group Gobom at our institute) we identified over 250 proteins to be part of this complex. To elucidate their function, we depleted each protein component on the cellular level using RNAi. Interestingly, we identified both centrosomal components that play a "classical" role in microtubule nucleation and additional components that are part of signaling pathways involved in growth control and cell cycle regulation. In summary, this work now provides a molecular and functional map of the centrosome that will make it possible to investigate possible links of this cell organelle to diseases.

Centrosomal proteins interact with spindle checkpoint components

Subsequently, we characterized the γ-tubulin ring proteins (γ-TuRC) that are directly involved in the initiation of the microtubule filaments at the centrosome and hence are critical for spindle organization.

We investigated the interactions of the γ-TuRC proteins to cell cycle regulatory proteins because we wanted to understand how components are controlled that play such an important role in mitotic division. We focused initially on certain cell cycle regulatory proteins (checkpoint proteins), which monitor whether chromosome segregation is carried out correctly. Interestingly, the proteins that fulfill this function have been shown before to be localized to the chromosomes, in particular to the attachment sites of the chromosomes to the microtubules (kinetochores). Our work was able to show that a subset of checkpoint proteins are interacting and functioning together with γ-TuRC proteins as well. This interaction is so vital that it is highly conserved between the fruit fly *(Drosophila melanogaster)* and human cells. The next step is now to define the precise molecular task that γ-TuRC proteins and their interacting checkpoint proteins have in the process of cell division. Also it will be investigated whether mutations of γ-TuRC proteins have a functional role in diseases such as cancer.

OUTLOOK

Prevention, diagnosis and therapy of diseases require a detailed understanding of the molecular processes in health and disease. Application of techniques from the area of functional genomics on the individual patient, combined with the development of bioinformatics systems that model the disease process are now required. Hence, the next logical step in the characterization of protein complexes is the detailed characterization of the effect of mutations on protein-protein interactions, protein complex formation and function. This is planned in a future project ("MUTANOM Project") that will systematically screen mutations either identified in databases by large-scale sequencing projects or through direct sequencing of DNA obtained from disease tissues. A direct side-by-side comparison of protein-protein interactions (normal versus mutated) on a proteomics level as well as on a cellular and in an animal model will elucidate the molecular consequence of frequently identified mutations. It will be the aim of this project to:

- translate our modeling approach and experimental results to the clinical sector for improved prevention, diagnosis treatment and new drug target identification.
- identify new molecular targets for the development of site-specific chemotherapies.
- establish a bedside mutational profiling tool based on a high-throughput sequencing technique and transfer to routine diagnostics.

References

[1] *Müller H et al. A Centrosome-Independent Role for γ-TuRC Proteins in the Spindle Assembly Checkpoint. Science, 314: 654-656, 2006.*

[2] *Lehmann V et al. Immunoisolation of Centrosomes from Drosophila melanogaster. Current Protocols in Cell Biology, Supplement 29, 3.17.1-3.17.13,Dec 2005.*

[3] *Lange BMH et al. Structure and Microtubule-Nucleation Activity of Isolated Drosophila Embryo Centrosomes. Histochemistry and Cell Biology, 124(3-4):325-34, 2005.*

[4] *Sauer S et al. Miniaturization in functional genomics and proteomics. Nature Genetics Review, 6, 2005 6(6):465-476.*

Links

http://www.molgen.mpg.de/~ag_lange
http://www.smp-protein.de/

CHEMICAL GENETICS – BIOCHEMICAL PERTURBATION OF DISEASE MODELS

ANTONIUS DIKMANS, ULRIKE BEUTLING, ARNDT WIEMANN, RONALD FRANK

HELMHOLTZ CENTRE FOR INFECTION RESEARCH, DEPARTMENT OF CHEMICAL BIOLOGY, BRAUNSCHWEIG

INTRODUCTION

The biomedical sciences progress by elucidating molecular mechanisms underlying biological processes. Identification of molecular partners interacting in cellular and organismal processes requires experimental perturbation of such processes, genetically as mutations, and biochemically by using chemical agonist/antagonist ligands of protein targets. Obviously, the latter provides chemical compounds that are potential drug candidates, and their discovery directly creates a link to pharmaceutical drug development and translation into the clinic.

Chemical genomics uses small organic molecules as tools to interfere selectively with protein function (chemical knockdown). Highly parallel and combinatorial chemical synthesis today provide libraries of compounds with diverse chemical structures for biological activity screening comparable in number with molecular biology-based nucleic acid or protein libraries. Screening of these chemical libraries is carried out with either phenotype-directed cell assays (*forward* chemical genetics) or *in vitro* protein binding assays followed by cell-based phenotype validation (*reverse* chemical genetics).

NGFN recognized the value of the chemical genetics approach to systematic studies of disease mechanisms and supported the implementation of the respective infrastructure at one of its core technology sites, the HZI (formerly GBF). This included the building of the required infrastructure and methodologies to provide access to competent compound collections, logistics and robotics for high-throughput screening (HTS), HTS-compatible assays for the processes under investigation and the bio-/cheminformatic tools for data evaluation.

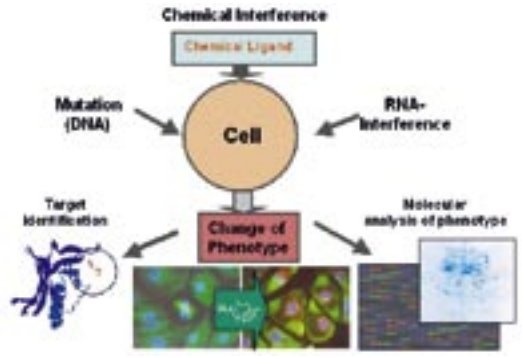


Fig. 1: Three complementary approaches to functional genome analysis.

PROJECT STATUS/RESULTS

The compound repository: Currently, more than 80 000 compounds in different formats are archived and applied in screening campaigns. The HZI compound archive contains

- over 100 natural products from the GBF myxobacterial metabolite collection
- 7000 small organics purchased from EMC Microcollections (Tübingen)
- 15 000 small organics from the ChemBioNet collection
- 40 000 small natural product hybrids (home-made "recombinations" of natural product fragments)
- more than 25 000 cellulose conjugated small molecules (diketopiperazines, triazines, etc.)

The compound storage and application logistics is compliant with most GLP standards by implementing a fully barcoded LIMS system and a cheminformatic compatible database (ChemFinder). All screening results from these compounds are collected and entered into the database, including biological activity data from 20 years research on the myxobacterial metabolites. Currently, also microbial extracts are being formatted for screening and entered into the database.

Screening technologies: Next to a representative collection of commercially available compounds and the unique HZI natural products, our main source for small molecules comes from solid phase synthesis on a special type of cellulose membrane following our combinatorial synthesis technologies like the SPOT synthesis[1] or the Cut&Combine synthesis[2]. These will produce compounds anchored to individual segments of the cellulose support and may be tested *in situ* for protein binding (macroarray). The cellulose segments can be segregated into distinct wells of a microtiter plate and dissolved by treatment with e.g. a trifluoroacetic acid cocktail to yield solutions of cellulose-compound conjugates. Further handling allows for the precipitation and cleansing of the cellulose-conjugates which are finally dissolved in DMSO as stock solutions. From a dilution of these DMSO stocks, small aliquots are transferred to the surface of special glass carriers such as microscope slides either in a mini-format or a micro-format[3]. We can print large numbers of copies from the initial library preparations for systematic protein binding studies.

Proteins are the primary targets for "chemical interference" in disease by drug therapy. Therefore, the identification of chemical compounds that bind with high affinity and selec-

COORDINATOR
Professor Erich Wanker
Max Delbrück Center for Molecular Medicine, Berlin
ewanker@mdc-berlin.de

tivity to disease proteins is paramount and most challenging. Our microarray-based systematic screening particularly matched the major focus of the NGFN-2 platform SMP Protein, which is on the systematic analysis of human and mouse proteins involved in disease processes, signaling cascades and gene regulation.

In addition to array-based target-directed screening, we could implement with the support of the NGFN the required robotics for high-throughput cell-based assays. A cell is a complex biological system containing a huge number of target molecules that are all functionally connected: a high content bio-assay system. The cell reacts to the action of the compound by developing a distinct phenotype. Thus, cell-based phenotypic screening is particularly effective in the search for bioactive compounds, producing hits from relatively small compound libraries although the target is not immediately apparent and needs to be identified subsequently.

OUTLOOK

Since the beginning of NGFN funding, we have successfully built the required infrastructure/methodologies and discovered very promising small molecules. Among these, a new type of proteasome inhibitor with potent *in vivo* anti-tumor activity has the potential for clinical development.[4]

In 2004, the NGFN Chemical Genomics Initiative joined a group of German chemists, biologists and bioinformaticians from academia who realized the need for an interdisciplinary platform to enhance research projects exploiting the systematic usage of small molecules for the study of biological systems and founded the ChemBioNet (www.chembionet.de). ChemBioNet now will focus on establishing and maintaining an appropriate infrastructure for the support of academic screening projects. Three partner institutes (HZI, MDC and FMP) co-financed a shared central compound collection, located at the FMP, to be offered for screening. This repository will be continuously complemented by compound collections donated by chemists of the network. Screening of compounds will be carried out on the individual conditions defined by the donors and on the basis of academic collaboration agreements. ChemBioNet will build up a database for all generated screening results with a regulated access assuring both IP rights and maximum free academic use. This ChemBioNet will be a competent partner for the future NGFN research program.

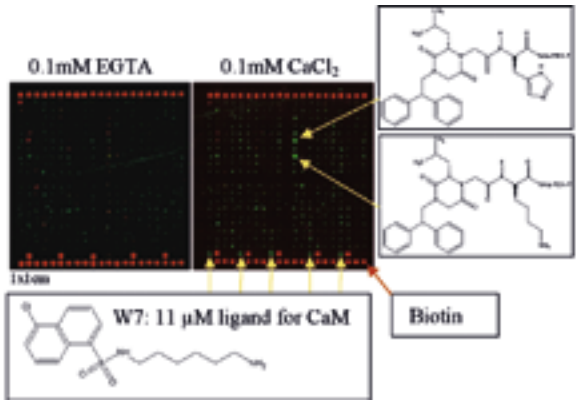


Fig. 2: Discovery of novel calmodulin ligands by probing a diketopiperazine-library array.

ChemBioNet

References

[1] *Frank R. Spot-Synthesis: An easy technique for the positionally addressable, parallel chemical synthesis on a membrane support. Tetrahedron 1992, 48, 9217-9232.*

[2] *Dittrich F et al. "Cut and Combine": An easy membrane-supported combinatorial synthesis technique. Bioorg. Med. Chem. Lett., 1998, 8, 2351-2356.*

[3] *Dikmans A et al. SC2: A novel process for manufacturing multipurpose high-density chemical microarrays. QSAR Comb. Sci. 2006, 25, 1069-1080. DE 10 2004 023 906.1.*

[4] *Malek N et al. Use of inhibitors of the degradation of p27 for the treatment of cancer. 2007, EP 07 004 185.0.*

BLOOD-BASED NEUROCHEMICAL DIAGNOSIS OF EARLY AND INCIPIENT ALZHEIMER'S DEMENTIA

JENS WILTFANG[1], JOHANNES SCHUCHHARDT[2]

[1] DEPARTMENT OF PSYCHIATRY, UNIVERSITY OF ERLANGEN-NUREMBERG
[2] MICRODISCOVERY GMBH, BERLIN

INTRODUCTION

The human brain proteomics project (HBPP) pursues an integrative approach towards understanding molecular mechanisms of neurological pathologies. Our consortium combines novel biomolecular detection and quantification techniques with high-quality clinical phenotyping of samples and patient groups. Here we report on the systematic investigation of a set of blood peptide biomarkers and clinical phenotypic parameters for their ability to predict Alzheimer's disease.

Alzheimer's disease

With increased life expectancy in the Western world, neurodegenerative disorders are becoming a major health concern. Alzheimer's disease (AD), the most prominent example, affects about six to seven percent of people over the age of 65 and more than 30 percent over the age of 85. Although explicit cure still is not available, preventive measures are promising to substantially delay the onset of manifest disease. The long-term goal of this study is establishing a blood-based neurochemical diagnosis of early or incipient AD, replacing the current invasive CSF-based diagnosis.

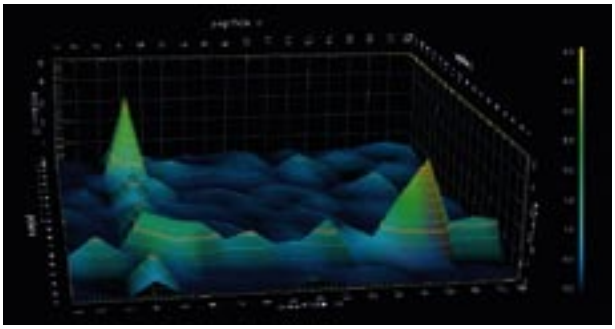


Fig. 1: Systematic combinatory investigation of a peptide marker set. Red lines indicate the significance levels 0.05 and multiple test corrected 0.05. While individual markers are non-significant after multiple test correction, the ratio of peptides 4 and 15 is significant even after multiple test correction.

AD diagnostics

The growing arsenal of molecular imaging methods is making it possible to gain closer insight into processes of the working brain. Nevertheless, the hallmark of disease classification for AD is still detection of molecular plaques and neurofibrillary tangles, accessible to observation only by autopsy. Early diagnostics of neural dementias is difficult, and the best molecular diagnostic assays still rely on the measurement of dementia biomarkers in cerebrospinal fluid (CSF). This approach is more invasive than blood sampling, and it does not allow repetitive measurements during follow-up in larger cohorts of high-risk patients. Blood-based neurochemical dementia diagnostics (Blood-NDD) would allow to conduct large-scale epidemiological studies and also multiple follow-up measurements for therapy monitoring.

PROJECT STATUS/RESULTS

Recent studies indicate that for the first time, blood-based classification may reach comparable levels of accuracy as CSF-based classification. In this study we investigate the discriminative power of a selected spectrum of peptide signatures in blood samples. Our investigation results in the identification of a set of peptide biomarkers that show super additive performance. These markers in combination with the phenotypic parameter age already offer a prediction value of at least 80 percent for early AD and most importantly also for incipient AD.

Data set and methods

The study comprised 64 patients with early AD and non-demented disease controls. Eighteen blood-based peptide markers and 12 clinical parameters were observed. A prediction task was formulated using a classification based on clinical dementia diagnostics, which was further cross-validated by a typical AD dementia biomarker pattern (beta-amyloid peptide 1-42⇓, total-Tau⇑, phospho-Tau⇑) in CSF. Different methods were employed for classification of the data. Here we demonstrate the results achieved using a random forest classifier and support vector machines.

Bootstrapping methods for predictor validation

A classical problem of classification machines is overfitting. Overfitting happens if a predictor system has too many internal degrees of freedom in relation to the size of the data set employed for trimming it. A very robust approach to assess the true value of a set of experimental parameters for prediction is the random forest method. Here bootstrapping techniques are employed to assess the predictive value of each parameter. The result of such an analysis for non-CSF based parameters in our dataset is given in Fig. 2. While imaging-based parameters still have the strongest predictive value, the best molecular parameter ranks second. Our most recent results (data not shown) on blood-NDD indicate that a set of four to six biomarkers promises to rank first within the near future.

COORDINATOR
Professor Helmut E. Meyer
Ruhr-University, Bochum
helmut.e.meyer@rub.de

Pathway modeling

While the above approaches already promise considerable advances in the classification and prediction of molecular patterns related to AD, we cannot claim that the general mechanisms of disease are amenable to this direct approach. It is well known that a number of complex regulatory interactions are involved in the molecular key pathophysiology of AD. Moreover, it is not entirely clear which are the truly relevant or causal mechanisms for disease etiology and progression in sporadic AD. However, it becomes more and more clear that several molecular key events are involved at the same time, linking the disease to protein misfolding and dyscatabolism, deficient immune responses, or a dysbalance of energy homeostasis. Explicit mathematical modeling techniques in combination with highly specific molecular signatures are promising to unravel the underlying pathophysiology.

OUTLOOK

In this project we try to establish a target-driven disease classification based on a selected set of markers with known functional relations. It is evident that only strict adherence to validation – grounded on analytical neurochemistry and molecular biology – can guarantee a set of variables with sufficient reliability. While prediction quality solely employing blood-based parameters still does not match the diagnostic accuracy of CSF-based methods, our approach to multiparametric neurochemical dementia diagnostics seems very promising.

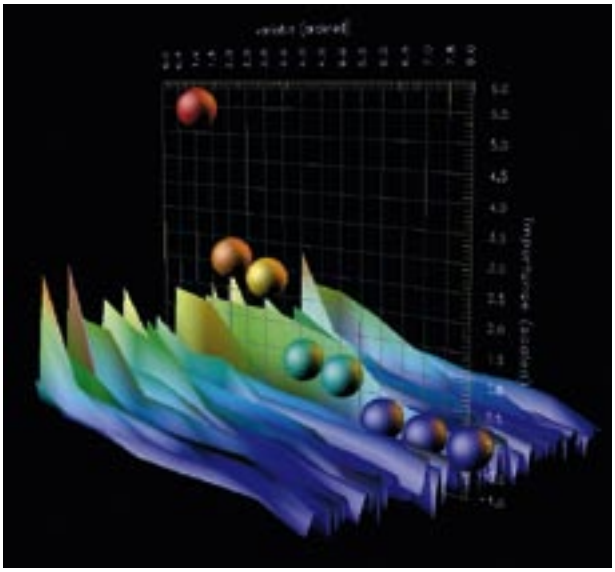


Fig. 2: Estimation of predictive quality of markers using a random forest classifier. The 'Mean Decease Accuracy' of image score (red), peptide 4/15 (orange) and age (yellow) are clearly apart from random (green-blue surface), indicating significant prediction quality between AD and non-AD patients.

A blood-based predictive diagnosis will allow to offer (secondary) preventive treatment to high risk cohorts.

Currently, the study cohort is being extended to validate our results, and various proteomic research strategies are being applied in order to identify additional AD blood biomarkers. Over the longer term we expect that a deepening insight into molecular disease mechanisms and their crosstalk will identify not only novel diagnostic but also therapeutic targets.

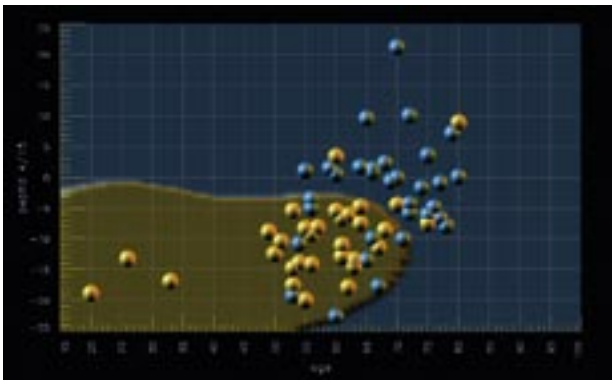


Fig. 3: Classification of diseased vs. non diseased cases employing a support vector machine. The yellow points represent non-AD patients, the blue ones AD positive patients. Yellow and blue areas indicate the distribution of the disease classes predicted by the support vector machine. Peptide ratio 4/15 in combination with patient age allow for correct disease classification in about 80% of the cases.

Reference

[1] *Klafki H-W et al. Therapeutic approaches to Alzheimer's disease, Brain 2006, 129, 2840-2855.*

NEUROMELANIN GRANULES – A HUMAN BRAIN SUBPROTEOME RELATED TO PARKINSON'S DISEASE

FLORIAN TRIBL, HELMUT E. MEYER, KATRIN MARCUS

BRAIN PROTEOMICS GROUP, MEDICAL PROTEOM-CENTER, RUHR-UNIVERSITY BOCHUM

INTRODUCTION

Parkinson's disease (PD) is one of the most frequent neurological disorders of elderly or aged people. About 1 percent of people are affected in the sixth decade of life, rising to 3 percent of people in the eighth decade. Approximately 3 million people in Europe are estimated to suffer from PD, and today there is still no causative treatment for this disease.

PD is a chronically progressing neurodegenerative disorder and is characterized by severely disabling motor symptoms, such as akinesia, rigidity, resting tremor, the 'freezing' phenomenon, a flexed posture and a loss of postural reflexes. Apart from olfactory and autonomic dysfunctions, these cardinal motor symptoms are the phenotypical correlates of the dopamine depletion in the striatum as a result of neurodegeneration occurring in the *substantia nigra pars compacta* (SN).

The underlying causes of PD remain unknown and it is especially unclear why neurons containing the black pigment neuromelanin (NM) are selectively affected in PD. In PD a severe accumulation of iron occurs in the SN and is regarded to account for elevated oxidative stress. Since free iron is cytotoxic, the prevention of cellular damage is generally accomplished by complexation of iron within a suitable storage system. In the SN neurons, such a system was discovered within so-called neuromelanin granules, which represent an organelle in the human brain that has received little attention so far.

The insoluble pigment NM formed from catecholamines is located in specific organelles termed NM granules and causes the characteristic pigmentation of the human SN. Interestingly, these pigmented organelles are absent in laboratory model organisms, such as mice and rats, but occur in most catecholaminergic neurons in primates and reach their maximal appearance in humans. A contribution of NM granules to the pathomechanisms of PD has been based on iron-related oxidative stress and the selective aggregation of alpha-synuclein in PD. The lack of accessible model systems, however, has so far hampered a thorough investigation of NM granules by molecular approaches. Nevertheless, the investigation of NM granules is of high importance as they are increasingly considered to be involved in neuropathological processes in PD.

PROJECT STATUS AND RESULTS

We recently succeeded in the isolation of NM granules from human SN tissue and in the identification of proteins which constitute these organelles. So far, we have published 72 proteins and the first fundamental characterization of NM granules as a lysosome-related organelle.

NM granules were isolated via a top-down approach, stepwise reducing the complexity of the sample. First, human SN tissue was mechanically dissociated through a sieving mesh to provide a cell body suspension. This step was especially necessary to enable an enrichment of the pigmented neuronal bodies by centrifugation through a discontinuous sucrose gradient. While the glial cells and membranous matter accumulated at the density interfaces, the dark brown neuronal cell bodies could be collected as a pellet, isolated and homogenized. The resulting homogenate was then centrifuged through a Percoll cushion that separated the very dense NM granules from the remaining cellular components.

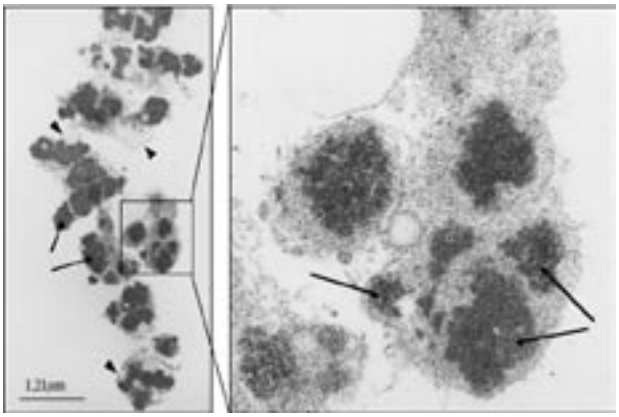


Fig. 1: Neuromelanin granules isolated from the human brain. Dr. Florian Tribl, Brain Proteomics Group, Medical Proteom-Center, Bochum

Quality control was performed by assessing both the morphology and the molecular level: first, an inspection of the isolates by transmission EM revealed morphologically intact organelles (Fig. 1). NM granules are irregularly shaped organelles that are limited by a single membrane and exhibit lipidic bulbs attached to it, and the pigment NM is embedded within a protein matrix. Secondly, the evaluation by Western blotting covering known organellar marker proteins of mitochondria, the nucleus, the Golgi complex, the plasma mem-

COORDINATOR
Professor Helmut E. Meyer
Ruhr-University, Bochum
helmut.e.meyer@rub.de

brane, early endosomes, lysosomes, and the ER, demonstrated a sufficiently high purity of the isolated NM granules as required for a proteomic analysis, but also revealed the presence of lysosomal markers, such as lysosome-associated membrane glycoprotein 1 and cathepsin B.

The mass spectrometric analyses of NM granules revealed 72 proteins, in particular lysosomal membrane proteins, such as lysosome-associated membrane glycoprotein-1 and -3, lysosome-integrated membrane protein II and several subunits of the vacuolar ATPase that mediates the acidification of lysosomes and lysosome-related organelles. Furthermore, NM granules contain enzymes engaged in the turnover of biomolecules including proteins and (glycol)lipids, e.g., cathepsin B and D, di- and tripeptidyl-peptidase, acid ceramidase, angiotensinase C, or sialic acid 9-O-acetylesterase. A few proteins involved in cargo traffic, such as clathrin, tubulins, microtubule-associated protein 2, and SNAP-alpha were identified, in addition to proteins initially identified in mitochondria, e.g. heat shock proteins and components of the ATP-synthase F1 subunit that interestingly are reproducibly identified in organelles deriving from the endosomal system. Similarly, proteins predominantly located in the ER, e.g., calreticulin, ribophorin I, or calnexin, are present in NM granules, while classical ER- marker proteins, such as the Ca^{2+}-binding protein grp78 (BiP/grp78), are absent.

These data suggest that NM granules evolve from the endosomal system (Fig. 2). The comparison of yet existing proteome profiles of lysosomes and lysosome-related organelles shows that NM granules, rather than being conventional lysosomes, more likely belong to the group of lysosome-related organelles (LROs). In contrast to conventional lysosomes that are mainly responsible for biomolecule turnover, LROs, e.g. platelet-dense granules, osteoclast granules, MHC class II compartments, or melanosomes, are found in very specific cell types and are largely engaged in very specific functions, such as blood clotting, bone remodeling, the regulation of the immune system, or pigmentation of the skin, respectively. The function of NM granules, however, remains to be fully elucidated.

OUTLOOK

This first proteomic profile of NM granules forms the valuable basis for future experiments to clarify the role of still uncharacterized proteins. The function of several proteins identified in NM granules is completely unknown and their possible roles in the pathogenesis of PD remain to be elucidated.

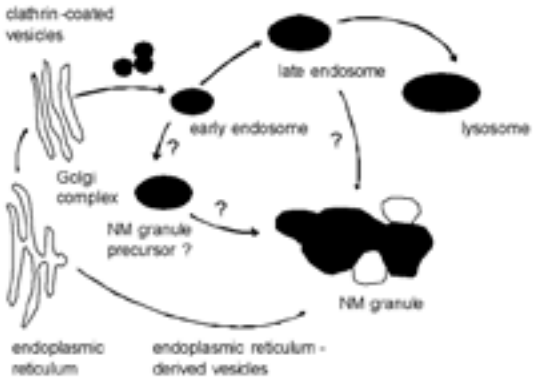


Fig. 2: Suggested biogenesis of neuromelanin granules. Dr. Florian Tribl, Brian Proteomics, Medical Proteom-Center, Bochum

Again, there is currently no powerful experimental approach available to systematically explore NM-associated proteins: since the SN of rodents is devoid of pigmentation, an appropriate model system is lacking. Nevertheless, we recently could overcome this predicament by applying a proteomic approach on NM granules. Now, selected proteins are identified to be further investigated by sophisticated molecular approaches and to open a door eventually towards biomarkers for PD or new therapeutic targets.

References

[1] *Fedorow H et al. Double KL, Prog Neurobiol 75(2): 109-24, 2005.*

[2] *Tribl F et al. Mol Cell Proteomics 4(7): 945-957, 2005.*

[3] *Tribl F et al. J Neural Transm 113, 741-749, 2006.*

DEVELOPMENT OF A HUMAN RECOMBINANT ANTIBODY GENERATION PIPELINE

ZOLTÁN KONTHUR, JOSMAR LANGNER, VOLKER SIEVERT, CAROLA STOSCHEK, JEANNINE WILDE, HANS LEHRACH

MAX PLANCK INSTITUTE FOR MOLECULAR GENETICS, DEPARTMENT OF VERTEBRATE GENOMICS, BERLIN

INTRODUCTION

The major challenge since the sequencing of the human genome has been completed is the characterization of all gene products and their interaction with each other in health and disease. The investigation of individual proteins and protein complexes in functional genomic and proteomic studies demands large sets of detection molecules. Inevitably, antibodies are the most commonly used biological detection reagents in day-to-day laboratory experiments. However, the availability of antibodies recognizing different human gene products is rather limited, as the function of numerous human proteins is yet unknown and many of these proteins have not been studied or are simply not in general focus.

To overcome the bottleneck of antibody availability for research purposes, the NGFN consortium “Antibody Factory” has set out to develop cheap and robust methodologies based on phage display technology to obtain oligo- and monoclonal human antibody fragments opening up a wide range of downstream applications. Within the consortium the aim of the “Panning Automation” subproject is to develop a pipeline for the generation of human recombinant antibodies by phage display as research reagents. The pipeline consists of several modules placed in a virtual “conveyor belt”-like manufacturing flow dealing with different tasks within the production, as depicted in the figure 1.

PROJECT STATUS/RESULTS

In this subproject we are continuously investigating the individual processes of the selection pipeline to achieve the overall goal of selecting human recombinant antibody fragments against 300 target constructs per year. Within the last year we streamlined many aspects of the different modules and have collected data from more than 500 individual selections against ~100 targets. This work resulted in more than 2000 monoclonal antibody expression clones, of which many have been and will be further characterized.

Below, some of the achievements of our work on the different modules, such as expression of target molecules and phage display selection recombinant antibody fragments are highlighted.

Target protein expression

Our focus is on generating antibodies against human proteins, and since availability of homologous proteins of human origin is rather limited, we decided to express our target proteins by recombination. For target expression, the first choice of host is *Escherichia coli*, since the yield of obtained recombinant protein is predominantly high and culturing costs are low. However, the optimal target molecule for successful selection of recombinant antibodies by phage display should be soluble, show homogenous folding and preserve the conformation upon attachment to the selection matrix. Looking closely at these prerequisites, we encountered two bottlenecks within our pipeline: the fair percentage of soluble recombinant protein in sufficient quantity for *in vitro* selection, and the lack of a simple method for uniform presentation of the target protein during the selection process. To overcome these problems, we have designed a set of expression vectors which fulfill both criteria and allow recombinant protein expression as N-terminal fusion protein using a 6His- and a biotinylation-tag (AVI-tag) in *E. coli* and in mammalian cell culture. Firstly, proteins insoluble in *E. coli* can be expressed in human cell lines, resulting in low yield but well-folded and posttranslationally

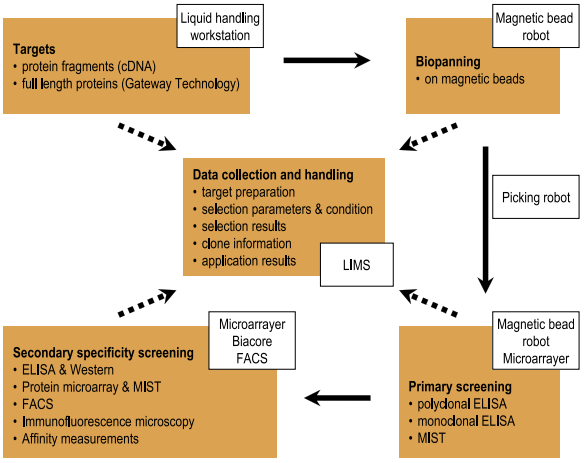


Fig. 1: The schematic representation of the antibody generation pipeline. The applied unit-automation strategy allows continuous modification or exchange of individual modules within the pipeline. Green boxes represent tasks, white boxes equipment in use.

modified proteins. Secondly, orientated immobilization of the recombinant proteins is achieved by applying a biotinylation tag. The AVI-tag is a short peptide sequence which is monobiotinylated *in vivo* by the endogenous biotin ligase in *E. coli*. For efficient biotinylation of the recombinant protein at the AVI-tag also in a eukaryotic expression host, we co-expressed a codon-optimized version of the bacterial biotin ligase encoding *birA* gene in mammalian cells. This biotin ligase is

COORDINATOR
Professor Stefan Dübel
Technical University Braunschweig
s.duebel@tu-bs.de

fully functional and biotinylates only the recombinant target proteins. No unintended biotinylated homologous proteins have yet been encountered.

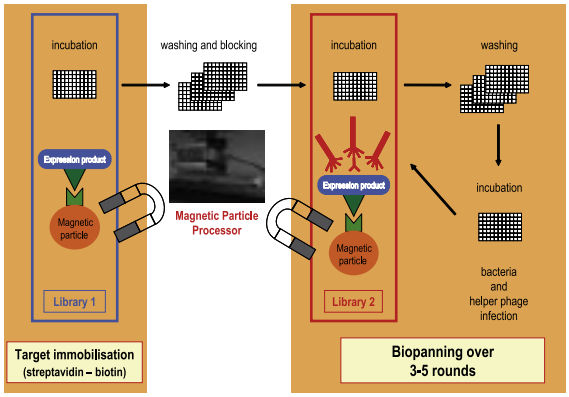
Based on this vector series, standardized large and small scale purification protocols have been developed which allow one- or two-step purification schemes to obtain maximum purity of the recombinant proteins for downstream applications, such antibody selection and evaluation antibody verification experiments.

Additionally, our new set of vectors is Gateway Technology-compatible and – in combination with the available ORFeome clone collection (~12 000 clones) as a resource – enables us to express full-ORF clones of more than 10 000 genes in the future for antibody selection.

In summary, they allow us to do both, simple purification as well as site-specific and directed immobilization of soluble recombinant proteins via the high affinity streptavidin-biotin interaction to magnetic beads, our preferred selection matrix for the in vitro panning of phage display antibody libraries.

Semi-automated selection of antibodies

One of the challenges in setting up parallel selections in a (semi-)automated fashion is the adaptation of the "biological" steps required during the selection process with robotic technology and *vice versa*. Utilizing a simple magnetic bead-handling robot, we managed to introduce valuable features into the selection protocol with minimal intervention to the required biological steps of the biopanning procedure. It allows all steps of the method to be standardized with respect to incubation time, as well as physical forces applied during bead transfer from well to well and resuspension without expensive and complicated automation technology, thus reducing human intervention to a minimum.



Additionally, the system is compliant with the classical 96-well microtitre plate format and therefore allows the use of standard laboratory equipment during all preparation steps. Finally, it enables multiple selections to be performed at the same time. These can be selections against 96 targets in parallel, but also selection against fewer targets with different phage antibody libraries or different buffer conditions in parallel gearing the selections toward different population of antibody clones and specificities for the same selection target.

OUTLOOK

With the current setup of the selection pipeline, it is possible to conduct over 500 selections per year, including repetitions, different selection protocols and the use of different antibody libraries. Full implementation of the new vectors for target expression in combination with full-ORF collections will further increase the number of selection targets expressed and selections performed. We are confident that by the end of the funding period our overall goal of developing a selection pipeline capable of selecting and evaluating antibody molecules against up to 500 target constructs per year for future applications will be achieved.

References

[1] *Konthur Z. Automation of Selection and Engineering. (2007) In Handbook of Therapeutic Antibodies. Wiley-VCH, Weinheim (Dübel, S. Ed.), 413-431.*

[2] *Taussig MJ et al. Proteome Binders: planning a European resource of affinity reagents for analysis of the human proteome. Nat Methods. 4 (2007), 13-17.*

[3] *Hultschig C et al. Recent advances in protein microarrays. Curr Opin Chem Biol 10 (2006), 4-10.*

[4] *Konthur Z et al. Perspectives for systematic in vitro antibody generation. Gene 364 (2005), 19-29.*

Fig. 2: Semi-automated selection procedure using a magnetic bead-handling robot. During the procedure magnetic beads are transferred from plate to plate, which are prefilled with e.g. protein solution for target immobilization, antibody phage library for selection, diverse washing solutions or even bacterial cultures for phage amplification.

IN VITRO ANTIBODY GENERATION

STEFAN DÜBEL, THOMAS SCHIRRMANN, MICHAEL HUST

INSTITUTE OF BIOCHEMISTRY AND BIOTECHNOLOGY, TECHNICAL UNIVERSITY OF BRAUNSCHWEIG

INTRODUCTION

A major 'post-genome' objective is to fully characterize the human proteome and to use this information to understand diseases. Essential to this ambition is a comprehensive collection of specific antibodies directed against all human proteins and their variants. Although a significant number of such reagents are available commercially, their quality is often questionable, and only a small fraction of the proteome has been covered so far. This leaves a huge potential for the discovery and validation of disease-relevant targets (see Editorial Nature Methods 4 (2007) p.1-2). The results of the NGFN-2 "SMP Antibody Factory" consortium matured the separate technologies employed for the high-throughput *in vitro* generation of research antibodies to a point where integration into a robust and upscalable system is the next consequent step. In brief, starting from soluble proteins derived from cDNA or from synthetic peptides, antibodies will be generated by *in vitro* selection (phage display) in a streamlined pipeline optimized for significantly reduced cost per binder.

A rapidly growing list of potential therapeutic and diagnostic target candidates is currently identified by a growing number of clinical projects – mainly from microarray studies or proteome analysis. Despite the rapid increase of knowledge, approximately half of the open reading frames (ORF) identified by bioinformatics tools encode potential protein products with unknown biological functions – not to mention the very large fraction of proteins whose function is assigned solely due to homology without any experimental proof. The elucidation of gene function on a large, genome-wide scale will require systematic analyses of gene products but in turn provides a link between the sequence and the phenotype. To explore the full complexity and function of all these unknown open reading frames (ORFs), it is essential to establish a comprehensive, characterized and standardized collection of specific binding molecules ('binders') directed against these human proteins, including variant forms and modifications. Furthermore, primed with the knowledge of the human genome, a comprehensive repository of affinity reagents would be a resource to understand and exploit the proteome and validate the vast amounts data generated by 'omics' assays (Fig. 1). This task is too big for national initiatives, and consequently, the Antibody Factory is an active part of the EU initiatives Proteome Binder and Quality Binders.[1]

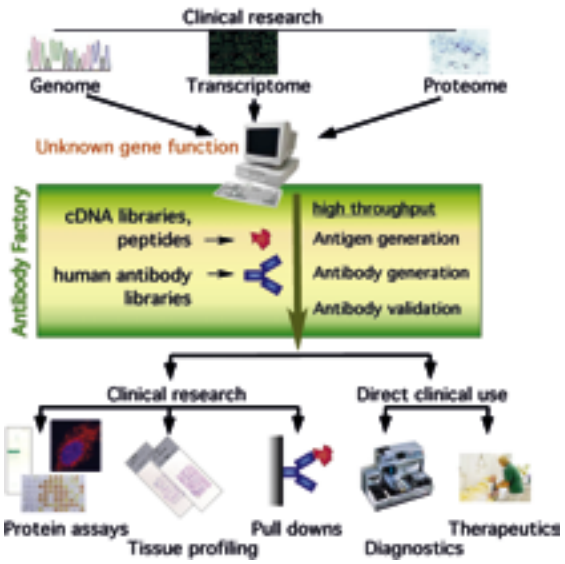


Fig. 1: Why do we need so many antibodies?

PROJECT STATUS/RESULTS

Digging for gold in the world's hugest recombinant human gene repertoires

The size of this problem requires a strategy optimized for throughput and constant quality of the reagents, and compatibility to a huge variety of antigens. The Antibody Factory tackles this challenge by using an antibody selection method which can be done completely *in vitro*, and by consequently optimizing every single step in the selection/production pipeline to the particular needs of research applications (Table 1). The first step is the generation of a huge antibody gene repertoire. Here, the Antibody Factory has successfully built, tested and employed a phage library containing >10^9 different antibody genes – a vast resource to select high affinity binders to almost every possible antigen, as it is even bigger than the estimated size of the antibody repertoire in our own body. So far, with pilot projects involving different external research labs, more than 1000 antibodies to more than 120 different antigens have been made and tested. The key to this success was the use of antibody phage display.[2]

Phage display allows to select binders to almost any target, including 'non-immunogenic', highly homologous, toxic or allosteric variants. The selection is based solely on binding, and the biochemical milieu can be completely defined by the experimenter during the selection process. Hence, we should also get antibodies to a number of interesting targets that normally would evade the immune response of experimental animals, in particular for mouse model studies. This technology has recently been favorably compared to "classical" mouse monoclonal antibodies by generating

COORDINATOR
Professor Stefan Dübel
Technical University Braunschweig
s.duebel@tu-bs.de

TABLE 1: DIFFERENT REQUIREMENTS FOR CONVENTIONAL AND PROTEOMIC SCALE RESEARCH ANTIBODY PRODUCTION

	Conventional antibody generation	Antibody Factory
Antigens	Single protein Known and available Somehow characterized Usually provided by researcher	Large number of proteins Not known, not available Properties unknown Production has to be part of the pipeline
Selection	Number of steps not an issue Optimized for minimal failure rate	Optimized for minimal effort Optimized for throughput
Antibodies	Various polyclonals & monoclonals Large amounts of a few antibodies	Standardized molecule and preparation Small amounts of many different antibodies
Validation	Done individually by end user	Initial assays integrated into process
Process cost	Not a central issue	Optimized for minimal cost

sets of binders to the same set of cDNA-derived proteins.[3] This study shows that the affinity of respective reagents is comparable to or better than that of mouse monoclonals, if a high quality library is used, and, furthermore, is even further improved after selection. Within the antibody factory, we further developed this technology, for example by designing a tailored molecular format, the single chain Fab fragment scFab (Fig. 2).[4] Here, the simplicity of a single polypeptide to be expressed was combined with the advantageous features of the Fab molecule – with its better stability and compatibility to established detection systems for antibodies.

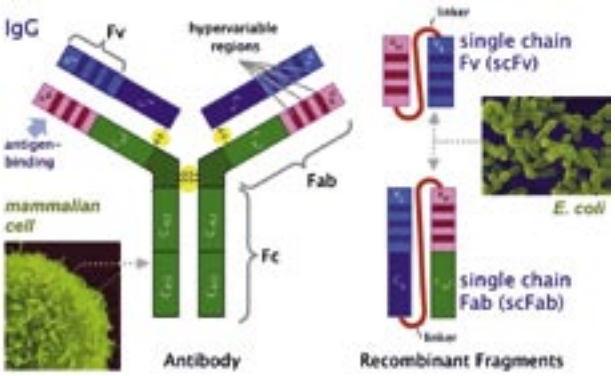


Fig. 2: Recombinant antibody formats

Recombinant antibodies enable novel applications for gene function analysis

Particularly helpful to assess novel research strategies is the recombinant nature of the antibodies. “Cassette” subcloning of individual antibody genes into a variety of different vector systems allows mass production of interesting candidates in cheap microbial production systems like *Pichia* or *Bacillus*. We even can use these antibody genes for experiments not accessible to animal-based antibodies, in particular, by knocking down a protein’s function through intracellular expression of the antibody. This technology has been validated by several labs, including ours, and respective vectors are available. Since the antibody gene is available as a cDNA right from the start in addition to the antibody protein, the antibody factory therefore is the enabling structure to apply this novel knock-down method on a broad scale for gene function research. Finally, by generating human recombinant antibodies from the start, a direct development of interesting candidate products into an *in vivo* diagnostic or therapeutic substance is a straightforward option for the use of the resulting antibodies.

References

[1] *Taussig et al. (2007) Nature Meth. 4, 13-17.*

[2] *Hust et al. (2007) In: Handbook of Therapeutic Antibodies, ed.: S. Dübel, Wiley-VCH.*

[3] *Ohara et al. (2006) Proteomics 6,2638-2646.*

[4] *Hust et al. (2007) BMC-Biotechnology 7:14.*

THE GERMAN GENE TRAP CONSORTIUM: A PUBLIC RESOURCE FOR MUTANT ES CELLS

THOMAS FLOSS[1], HARALD VON MELCHNER[2], PATRICIA RUIZ[3], JENS HANSEN[1], WOLFGANG WURST[1]

[1] INSTITUTE OF DEVELOPMENTAL GENETICS, GSF – NATIONAL RESEARCH CENTER FOR ENVIRONMENT AND HEALTH, MUNICH-NEUHERBERG
[2] DEPARTMENT OF MOLECULAR HEMATOLOGY, UNIVERSITY OF FRANKFURT MEDICAL SCHOOL, FRANKFURT AM MAIN
[3] CENTER FOR CARDIOVASCULAR RESEARCH, CHARITÉ – UNIVERSITY MEDICINE BERLIN AND DEPARTMENT OF VERTEBRATE GENOMICS, MAX PLANCK INSTITUTE FOR MOLECULAR GENETICS, BERLIN

INTRODUCTION

One of the major challenges in the future is the functional annotation of the genome and the establishment of animal models mimicking human disease. Gene trapping is used to introduce insertional mutations into individual genes expressed in mouse embryonic stem cells (ESCs); which we adapted for high throughput to mutagenize the entire genome. Gene trap vectors simultaneously mutate and report the expression of the endogenous gene at the site of insertion and provide a DNA tag for rapid identification of the disrupted gene. Gene trap approaches have been widely used by academic and commercial organizations to create libraries of ESC lines harboring mutations in single genes, to generate mutant mice.[4] Since gene identification is typically achieved by 5' Rapid Amplification of cDNA Ends (RACE) and sequencing, the major determinant for identifying a trapped gene is the level of its expression. During a period in which genomes were poorly annotated and genes were largely defined by cDNAs, RACE tags provided confirmation of a successfully trapped gene, and enabled identification of novel genes via a fusion transcript. However, RACE tags have two major disadvantages. First, they cannot reveal the exact position of insertion sites, which are usually at a considerable distance from the RACE tags and therefore cannot be used directly for mouse genotyping. Second, because 5'RACE is entirely dependent on gene expression, the major limitation for identifying a trapped ESC gene is its expression level. As most high-throughput trapping screens employ highly sensitive G418 selection, more than 50% of the G418-resistant ESC line genes are usually lost because the levels of trapped gene expression are below the gene identification threshold imposed by 5'RACE. We showed that the adaptation of a splinkerette-adaptor mediated PCR (SPLK) for amplification of genomic sequences flanking the gene trap integration sites outperforms the 5'RACE by up to six fold in gene trap screens and more effectively identifies gene insertions that are poorly represented in existing gene trap libraries.[1]

Until recently, the major drawback of gene trapping has been that the null mutations generated by the integration of a gene disruption cassette often result in embryonic lethal phenotypes, which precludes the analysis of gene function in the adult animal. Therefore, we generated conditional gene trap vectors[2], which allow to generate somatic mutations induced in specific cells or tissues at any given time in development and adulthood. Taking advantage of the Cre recombinase and FLP recombinase systems[2], clones carrying conditional gene traps can be postinsertionally modified by recombinase-mediated cassette exchange to introduce any gene of interest into tagged loci, e.g. to integrate Cre recombinase genes to expand the Cre-zoo or point mutated minigenes to mimic human genetic diseases. Further options are the induction of gain-of-function mutations or the ablation of specific cell lineages by inserting gain-of-function cassettes or toxin genes, respectively. These gene traps are now in use by the German Gene Trap Consortium (GGTC; www.genetrap.de) and the European Conditional Mouse Mutagenesis Program (EUCOMM; www.eucomm.org). All of these publicly available gene trap lines have recently been centralized on the International Gene Trap Consortium's (IGTC) web page (www.igtc.ac.org).

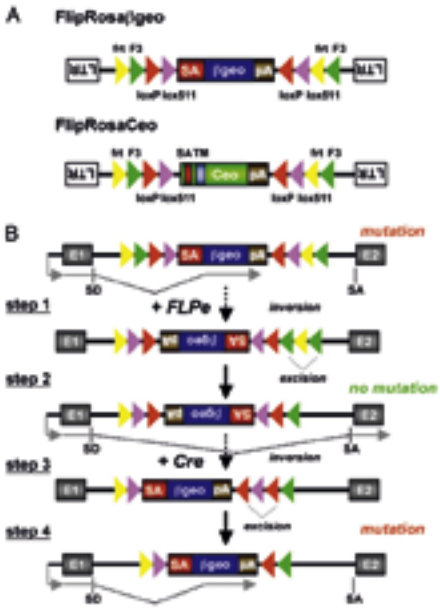


Fig. 1: Conditional gene trap vectors[2]

PROJECT STATUS/RESULTS

Since its formation the German Gene Trap Consortium (GGTC) has generated and sequenced more than 58 801 mutant mouse ES cells. In that way, 42 450 useful GTSTs were detected which represent more than 7044 different genes; 486 of these genes are disease-related. In the framework

COORDINATOR
Professor Martin Hrabé de Angelis
GSF – National Research Center for Environment and Health, Neuherberg
hrabe@gsf.de

of NGFN-2, the GGTC produced and sequenced more than 40 903 of mutant ES cell clones and obtained about 31 348 useful GTSTs. Approximately 800 clones have been distributed to the scientific community worldwide and were used to establish animal models.

OUTLOOK

Taken together, our results are highly relevant to the large-scale ES cell mutagenesis programs recently started in Europe (EuCOMM), the United States (KOMP) and Canada (NorCOMM). Most of these programs employ a combination of gene trapping and gene targeting in the effort to knock out every single gene in the mouse genome, and an optimal balance between the two technologies is sought in order to identify the most highly efficient mutagenesis strategy. Because trapping is cheaper and generally involves less work, targeted mutagenesis is normally reserved for genes that are the least accessible via trapping. Accordingly, genes well represented in the gene trap libraries are generally excluded from gene targeting. Based on our results, however, we predict that the implementation of the SPLK technology in future gene trap screens will significantly increase the pool of genes accessible to trapping and will thus reduce the number of genes requiring gene targeting for inactivation.

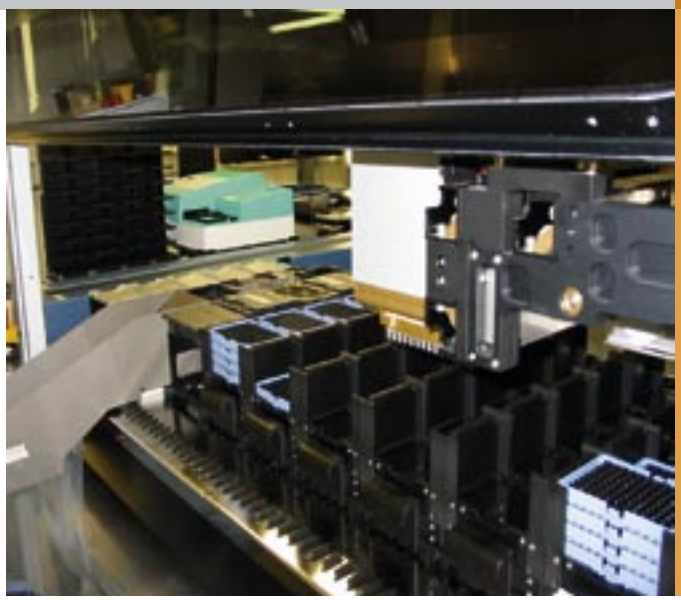


Fig. 2: Automated ES cell facility at the Institute of Developmental Genetics, GSF Munich

References

[1] *Horn C et al. (2007). High-throughput recovery of genomic gene trap sequence tags improves gene trapping efficiency in mouse embryonic stem cells. Nature Genetics, in press.*

[2] *Schnütgen, F et al. (2005). Genomewide production of multipurpose alleles for the functional analysis of the mouse genome. Proc Natl Acad Sci U S A 102, 7221-6.*

[3] *Hansen J et al. (2003). A large-scale, gene-driven mutagenesis approach for the functional analysis of the mouse genome. Proc Natl Acad Sci U S A 100, 9918-22.*

[4] *Wiles MV et al. (2000). Establishment of a gene-trap sequence tag library to generate mutant mice from embryonic stem cells. Nat Genet. 24(1):13-4.*

NEW MOLECULAR MECHANISM DISCOVERED FOR OSTEOGENESIS IMPERFECTA

THOMAS LISSE[1], FRANK THIELE[1], HELMUT FUCHS[1], WOLFGANG HANS[1], KOICHIRO ABE[1], LETICIA QUINTANILLA-FEND[2], BIRGIT RATHKOLB[3], VALERIE GAILUS-DURNER[1], MIEP HELFRICH[4], ECKHARD WOLF[3], STUART RALSTON[5], MARTIN HRABÉ DE ANGELIS[1]

[1] INSTITUTE OF EXPERIMENTAL GENETICS, GSF – NATIONAL RESEARCH CENTER FOR ENVIRONMENT AND HEALTH, MUNICH-NEUHERBERG
[2] INSTITUTE OF PATHOLOGY, GSF – NATIONAL RESEARCH CENTER FOR ENVIRONMENT AND HEALTH, MUNICH-NEUHERBERG
[3] LUDWIG-MAXIMILIANS-UNIVERSITÄT, MUNICH
[4] UNIVERSITY OF ABERDEEN
[5] WESTERN GENERAL HOSPITAL, EDINBURGH

INTRODUCTION

Osteogenesis imperfecta (OI) is an inherited disorder characterized by increased bone fragility, fractures and osteoporosis, and in most cases caused by mutations affecting the type I collagen genes. The estimated prevalence of OI is 1:15 000. OI is classified into seven human subtypes, and to date more than 500 human COL1A1 mutations have been reported representing a clinical heterogeneity dictated by the complex array of mutations. Type II OI represents the most severe form of the disease (consisting of subtypes OI-II A, B and C), whereby most of the affected patients die within the perinatal period, and account for ~20% of reported cases. The clinical phenotype of human OI-II is characterized – to varying degrees – as having abdomen protuberance, respiratory stress, bowed and truncated limbs, woven bone, de-ossified calvaria, generalized and localized osteopenia, intracranial bleeding and extensive fractures.

The possibilities to treat osteogenesis imperfecta patients are limited, as the molecular mechanisms of the disease are still poorly understood. Animal models are a promising tool to obtain more insights into the involved molecules and their interactions. We report on the isolation, cloning and characterization of the skeletal phenotype in a novel mutant mouse line as a model for type II human OI affecting the terminal C-propeptide region of Col1a1. In addition, we found a new mechanism causative for the disease which might be important for new therapeutic strategies.

PROJECT STATUS/RESULTS

The new mouse model for type II osteogenesis imperfecta termed AGA002 (abnormal gait 2) was isolated from the Munich N-ethyl-N-nitrosourea mutagenesis program[2] and exhibited bending of the front and hind limbs. Inheritance testing confirmed an autosomal semidominant mode of inheritance. Using an automated SNP genotyping platform[3], the causal gene was mapped to chromosome 11 by linkage analysis, and a C-terminal frameshift mutation was identified in the Col1a1 (procollagen type I, alpha 1) gene as the cause of the disorder.

First studies in the GMC (NGFN-1) in the Dysmorphology, Bone and Cartilage Screen analyzed the mice according to the human phenotype. AGA002 mice were smaller in body size. Severely affected animals developed scoliosis, provisional rib and long bone calluses, severe body size deficit, ocular edema, greasy skin and eczema. Hearing impairments – a clinical feature of Type I OI – were not observed in AGA002 animals. Reduced bone mass was detected in AGA002 mice by the application of dual-energy X-ray absorptiometry (DXA), peripheral quantitative computed tomography (pQCT) and micro-CT analysis (see Fig. 1). Distal femora, trabecular and cortical vBMD within the metaphysis were substantially decreased compared to controls. In addition, multiple fractures and early lethality were observed. In surviving AGA002 animals, biochemical and hormonal analysis revealed increased osteocalcin, TRACP 5b, parathyroid hormone and calcitonin levels in serum. A subtle transient increase in the total calcium level was detected only in male mutants, but no changes in inorganic phosphate levels were detected.

AGA002 heterozygous animals had markedly increased bone turnover. Using transmission electron microscopy (TEM) and scanning electron microscopy (SEM), we revealed a disrupted native collagen network leading to microfractures,

Fig. 1: Micro-computed tomograph imaging of seven-week-old AGA002 mice (top) and age matched controls (bottom)

COORDINATOR
Professor Martin Hrabé de Angelis
GSF – National Research Center for Environment and Health, Neuherberg
hrabe@gsf.de

and an increased prevalence of osteoblasts and osteoclasts. In *ex vivo* experiments using primary calvarial osteoblasts, we uncovered defects in nodular bone formation and cellular metabolism.

Further studies showed that abnormal pro-α1(I) chains accumulated within the endoplasmic reticulum (ER) in AGA002 dermal fibroblasts and were poorly secreted in the extracellular space. This was associated with the induction of an ER stress-specific unfolded protein response involving BiP/GRP78, with caspases 12 and 3 activation and apoptosis of osteoblasts both *in vitro* and *in vivo*. These studies resulted in the identification of a new model for osteogenesis imperfecta. For the first time we present a role for ER-stress associated osteoblast apoptosis in the pathogenesis of disease.[4]

Patients with OI exhibit increased bone fragility and high bone turnover, but the molecular mechanisms responsible are poorly understood.

As a next step, we rescued the phenotype of the AGA002 mutant mice. The most common therapy for human patients with reduced bone mineral density is a treatment with bisphosphonates. We selected alendronate for the AGA002 mice, and observed a clear increase in BMD values in the treated mutants compared to the nontreated animals. In contrast to the successful fracture prevention – the number of fractures could be reduced to one half (Fig. 2) – the survival rate until the age of weaning did not significantly increase. These findings indicate a systemic effect of the AGA002 mutation on additional organs that remain unaffected by the bisphosphonate treatment.[5] These results stress the importance of a systemic approach for mouse phenotyping to analyze causes of comorbidity in complex diseases.

OUTLOOK

Therefore, as the next step, AGA002 mice will be comprehensively and standardized phenotyped in the systemic primary screen in the German Mouse Clinic (GMC).[1] We expect to find additional phenotypes as a pleiotropic response of the novel Col1a1 gene defect in different modules of the GMC.

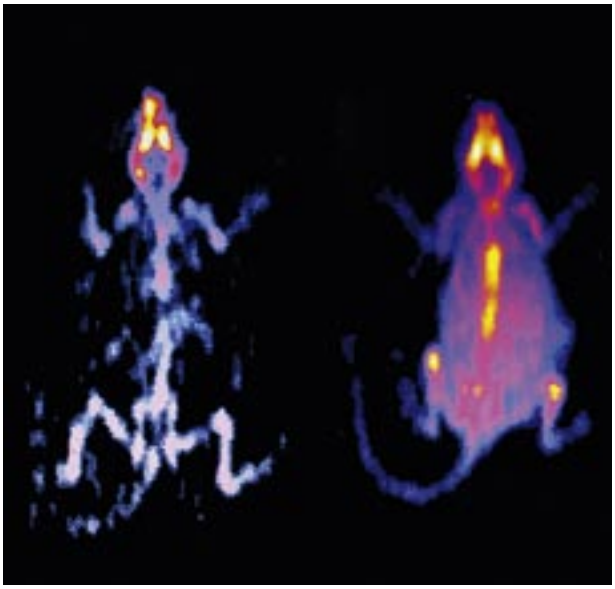


Fig. 2: Bisphosphonate (alendronate) therapy of heterozygous AGA002 mutant mice: Dual-energy X-ray absortiometry (DXA) images of untreated (left) and treated mutants (right). (Yellow-colored areas indicate high bone density, blue and black indicate low bone density)

References

[1] *Gailus-Durner V et al. (2005): Introducing the German Mouse Clinic: open access platform for standardized phenotyping. Nat Methods. 2(6):403-4.*

[2] *Hrabé de Angelis M et al. (2000): Genome-wide, large-scale production of mutant mice by ENU mutagenesis. Nat Genet. 25(4):444-7.*

[3] *Klaften M and Hrabé de Angelis (2005): ARTS: a web-based tool for the set-up of high-throughput genome-wide mapping panels for the SNP genotyping of mouse mutants. Nucleic Acids Res. 33: W496-W500*

[4] *Lisse T et al. (2007): ER stress-mediated apoptosis in a new mouse model of type II osteogenesis imperfecta. PLOS Genetics: in press.*

[5] *Lisse T et al. Pleiotropic effects in type II osteogenesis imperfecta mice. (Manuscript in preparation)*

THE NGFN PORTAL – AGGREGATING NGFN KNOWLEDGE

VOLKER STÜMPFLEN, HANS-WERNER MEWES

MIPS / INSTITUTE FOR BIOINFORMATICS, GSF - NATIONAL RESEARCH CENTER FOR ENVIRONMENT AND HEALTH, MUNICH-NEUHERBERG

INTRODUCTION

NGFN projects focus on information buried in the genome. Since the deciphering of the human and other mammalian genomes, basic as well as clinical research has become strongly data driven. The classical hypothesis-driven experiment is now complemented by high-throughput methods generating huge amounts of data that must be interpreted in the context of biological and medical knowledge. Cohorts of patient-related data are collected internationally and the wealth of information attached to the human genome sequence has grown far beyond intuitive comprehensiveness.

Data and information within NGFN is necessarily heterogeneous due to the different approaches, and it is widely dispersed over a large number of decentralized nodes. The design and implementation asks for an advanced but uncomplicated concept to meet these technical challenges.

Within the recent research period we have developed a web-based solution. The NGFN information portal is capable of aggregating and sharing NGFN-relevant information on the side of the data providers and allows for complex intuitive queries on the side of the user.

PROJECT STATUS/RESULTS

General concepts

The portal is based on a matrix concept, linking every gene of the human genome to the various types of information covered. Thus, it provides central access to distributed heterogeneous information in a context-dependent way (see Fig. 1). The use of standard industrial technologies has enabled us to overcome the contradictory requirements of central access and the need to retrieve data from distributed resources.

We therefore adopted basic concepts of the Web 2.0 technology approach, which resulted in providing the design of different views on genes of interest depending on the requested context. Although the Web 2.0 definition is rather vague, we have successfully combined different web technologies and their associated web applications. The portal employs (1) tagging of information, (2) versatile access to distributed information, and (3) content aggregation.

Tagging of information

Due to the resulting heterogeneity and complexity of the data (reflecting the various research activities within NGFN), a comprehensive physical integration of every piece of information within a centralized resource is not feasible. Here we

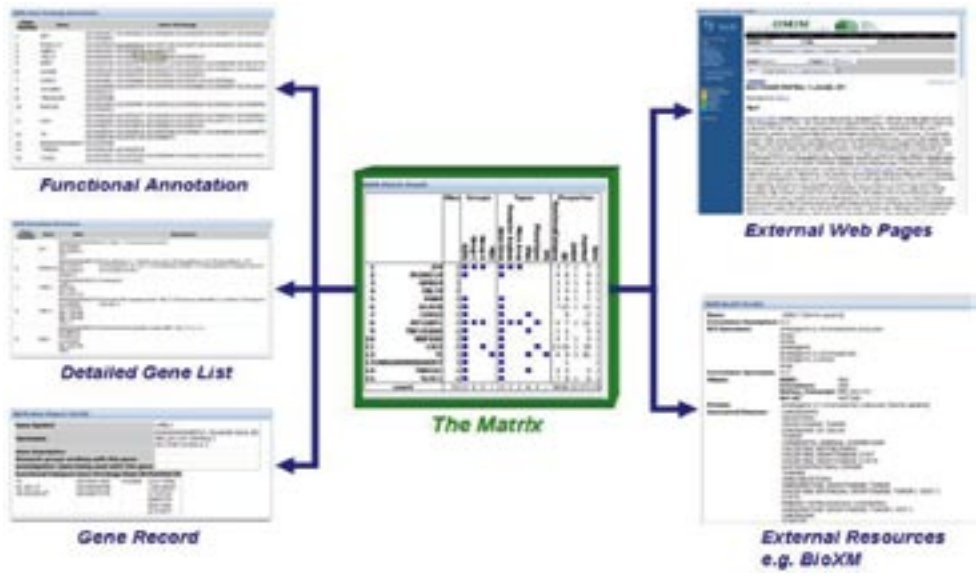


Fig. 1: Matrix view and context dependent views within the portal

COORDINATOR
Professor Roland Eils
German Cancer Research Center, Heidelberg
r.eils@dkfz-heidelberg.de

followed the Web 2.0 approach which involves building "folksonomies" to classify any kind of web-based content. Folksonomies are user-generated taxonomies capable of categorizing and retrieving associated content. To extend the simple but semantically weak tagging approach, we extended this concept to a so-called typed tagging approach, capable of defining the meaning of a tag. For example, a protein interaction investigation is unambiguously described with the type "interaction detection method" and the corresponding tag "Y2H".

Using this approach, one can easily describe various types of features of a gene and/or investigation without the need to develop complex database schemata covering every potential investigation. In addition, semantically defined tags and their relations allow to dynamically link independent data resources.

Accessing distributed information

While the type/tagging approach provides a light-weight layer of important information, it is not the intention of the portal to aggregate data. Instead, the portal uses state-of-the-art technology called Web Services to access external content.

The portal uses Web Services (Fig. 2) based on standard web technologies like HTTP and XML and offers a means of accessing distributed resources within any program. A server delivers a response in the form of an XML document containing any arbitrary complex information. While XML can be easily transformed into HTML for the display in browser applications, the portal provides a simple way to seamlessly combine internal (types/tags) with external distributed information. Public databases such as NCBI's PubMed or ENSEMBL can be accessed through Web Services, and the information retrieved is easily processed to the needs of the query.

Fig. 2: Accessing distributed information with Web Services

Although not yet every potential information resource provides a Web Services within the NGFN interface, the amount of available services within the biomedical domain is growing at an exponential rate.

Content aggregation with portlets

Having a technical solution for the aggregation of information, we extended this principle up to the level of the web interface. The NGFN portal uses portlets. Portlets are independent components within a web page. Each of them provides a different function regarding the rendering of content. The advantage of this solution is the capability of composing context-dependent views based on individual information representation components. This is used to aggregate web pages with different information about a specific gene. One portlet can access and display e.g. information stored in our local database, while another one retrieves and renders the corresponding entries from NCBI Web Service within one page.

User interaction and available data

Queries within the portal are intuitive. Beyond the search for gene names users may add a specific index to the word of interest, using wildcards and combining terms with logical operators. The query "PPI [ITN] AND kinase* [GOD]", for example, searches all genes within PPI investigations annotated with kinase activity (as described with Gene Ontology terms).

At present the NGFN portal contains close to 800 individual investigations within 20 main categories. These investigations reference approximately 13 000 genes and include experiments related to only few genes as well as large scale high-throughput experiments.

References

The portal is accessible at http://portal.ngfn.de

[1] *Stümpflen V et al. "Neue Informationstechnologien zur Nutzung biologischer Daten". Biospektrum, 03, 2007, p. 281-283.*

[2] *Oesterheld M et al. "Analysis of integrated biomolecular networks", Journal of Integrative Bioinformatics, in press.*

BIOINFORMATICS METHODS TO PREDICT PROGNOSIS IN CANCEROUS DISEASES

BENEDIKT BRORS, RAINER KÖNIG, ROLAND EILS

DIVISION OF THEORETICAL BIOINFORMATICS, GERMAN CANCER RESEARCH CENTER, HEIDELBERG

INTRODUCTION

Current methods to predict prognosis, in particular in cancer research, are in many cases unsatisfactory and do not reliably predict the full range of clinically observed courses of the disease. However, in many cases it would be helpful to have an individualized prediction of this course to choose a risk-adapted therapy. Modern methods in molecular genetics, in particular methods based on microarrays or so-called "gene chips", promise to reveal a "molecular portrait" of a tumor. It has been shown that such "portraits" may serve as a valuable source of information to characterize the corresponding tumors[1], and are also capable of predicting the future progress of a cancer disease.[2]

Prognosis prediction requires close collaborations between clinical researchers, molecular biologists and bioinformaticians. Given the large amount of data produced by "gene-chip" analysis, computer-aided analysis is indispensable. A single measurement from a single tumor may result in as many as 40 000 different values which refer to gene expression levels of many if not all transcripts present in a tumor cell. Bioinformatics tries to learn certain "patterns" which may indicate favorable or unfavorable prognosis. The patterns are learned from tumors obtained from a tumor bank, where the course of disease over several years is known for each patient. Thus, tumors can be classified into "good prognosis" and "poor prognosis" groups. Computer programs, so-called classifiers, are then trained with data from these groups and "learn" the characteristic patterns for each of them. Once training has been performed, prognosis can be predicted for any new tumor on the basis of gene-chip measurements.

Neuroblastoma is a cancer of early childhood, and is the most common extracranial solid tumor in pediatrics. However, it is still a rare disease with approximately 140 new cases per year in Germany. Neuroblastoma shows an extraordinarily heterogeneous course of disease, ranging from spontaneous regression to incurable metastatic progress. Therapy options thus differ, too. While localized tumors may be removed by surgery only, or even disappear without any treatment, metastatic neuroblastoma requires intense high-dose chemotherapy complemented by stem cell transplantation to rescue the child's hematopoietic system. A particular feature of neuroblastoma, and still a "clinical enigma"[3], is the fact that even metastatic tumors sometimes regress spontaneously and disappear by themselves, leaving no trace in the body. Clinical markers to predict the course of the disease exist, but are accurate in only about 80% of the cases. Thus, there is room for improvement based on novel diagnostic methods.

PROJECT STATUS/RESULTS

A neuroblastoma-specific gene chip

In a close collaborations of researchers from the NGFN neuroblastoma network (Dr. Fischer and Dr. Oberthür, University Hospital Cologne, and Dr. Westermann, DKFZ) and the bioinformatics platform in NGFN (Dr. Brors and Dr. König, DKFZ), a neuroblastoma-specific gene chip has been designed. This device covers all transcripts that were known to be indicative of clinically important subgroups at the time of construction. Probes, i.e. short oligonucleotides with gene-specific sequences, are present for about 10 000 different transcripts. The neuroblastoma-specific array is produced by the Agilent company. Tumor RNA is isolated and labeled with a fluorescent dye. This is mixed with reference RNA which serves as an internal standard and is labeled with a different dye. After hybridization to the gene chip, fluorescence of each dye is measured and recorded (Fig. 1). Data are stored in the standardized NGFN database iCHIP (http://www.ichip.de).

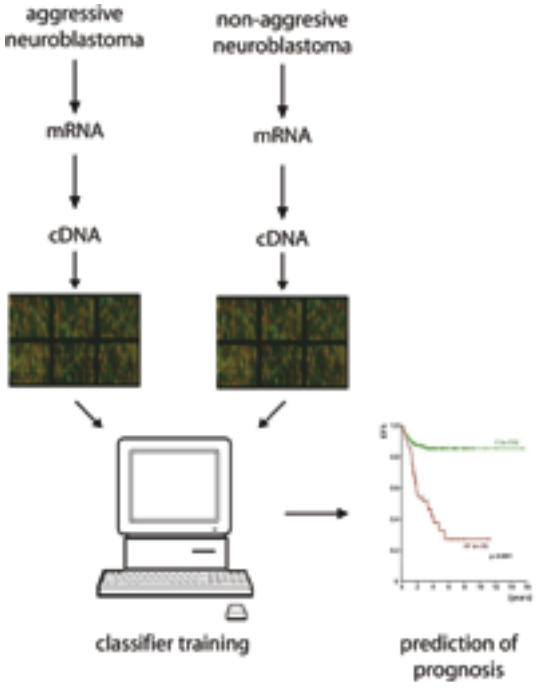


Fig. 1: Workflow for prognosis prediction of cancer based on gene expression profiling by DNA microarrays.

COORDINATOR
Professor Roland Eils
German Cancer Research Center, Heidelberg
r.eils@dkfz-heidelberg.de

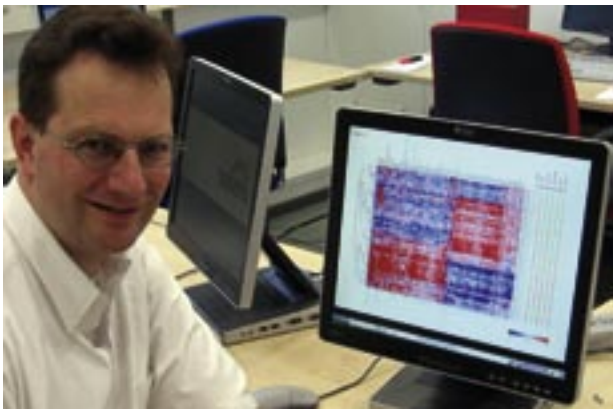


Fig. 2 : Predictive gene signature for neuroblastoma. Principal investigator: Benedikt Brors.

The researchers from the NGFN have used such data to construct a classifier that predicts prognosis of neuroblastoma with unprecedented accuracy.[4] An important step in this process was bioinformatic analysis covering preprocessing and quality control. Subsequent training was performed with 77 tumors of good course of disease and tumors with rapid progress. The trained classifiers were then used to predict prognosis in a set of 174 tumors which were more unclear to classify by classical clinical markers. The gene chip-based classifier was superior to classification by other clinical markers and accurately predicted the course of disease for most of the samples from the validation set. Through bioinformatic analysis of gene-chip data it also revealed that only two different subtypes exist – one very benign form of neuroblastoma which tends to regress spontaneously, and a very aggressive form that rapidly metastasizes to different organs. Validation of the result in a large set of tumors collected internationally is currently under way. It is hoped that the new method will soon be included in clinical trials to prove its capabilities also prospectively, and that it will become a standard method at least for certain neuroblastoma subgroups.

Deciphering the molecular traits of tumors

Gene-chip based analysis not only helps in diagnostics and prognosis prediction, it can also reveal which cellular processes are disturbed in cancer. In neuroblastoma, it is well known that genetic amplification of a certain oncogene, MYCN, occurs which is associated with a very poor prognosis. However, certain tumors have similar behavior but show absolutely normal levels of that oncogene. In a collaboration between Dr. Westermann (DKFZ), Dr. König (University of Heidelberg) and Dr. Brors (DKFZ), gene expression of cultured neuroblastoma cells was investigated that have been manipulated to express increased levels of the MYCN oncogene.

Analysis of gene chip data showed certain groups of genes that were either immediately induced in response to MYCN, or only after a delay. The group of delayed genes was highly associated with poor prognosis even in tumors that had normal levels of MYCN. In these tumors a genetic program is active that is downstream of MYCN, and which is activated by mechanisms independent of MYCN[5]. Thus, this analysis helps to understand why certain neuroblastomas are so aggressive, and reveals oncogenic pathways that may be targeted by novel therapeutics.

OUTLOOK

Analyses of gene chip data may soon become standard methods in molecular diagnostics, at least for certain cancers. Prerequisites are that information related to prognosis or therapy response can be obtained from such data, and that the knowledge of prognosis will have clinical consequences. This requires a broad range of therapy options, which is not given for all kinds of cancer. However, by further analysis of the “molecular portraits“ of tumors, novel targets for drug research can be found which may lead to development of novel therapeutics. Ultimately, the novel methods – in combination with powerful bioinformatics methods – may lead to a “personalized medicine“ that chooses a medication based on the very individual features of a disease as well as those of the patient.

References

[1] *Golub TR et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286:531-537, 1999.*

[2] *van ‘t Veer LJ et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415:530-536, 2002.*

[3] *Brodeur GM. Neuroblastoma: biological insights into a clinical enigma. Nat Rev Cancer 3:203-216, 2003.*

[4] *Oberthür A et al. Gene-expression based classification of neuroblastoma patients using a customized oligonucleotide-microarray outperforms current clinical risk stratification. J Clin Oncol 24:5070-5078, 2006.*

[5] *Westermann F et al. High Skp2 expression characterizes high-risk neuroblastomas independent of MYCN status. Clin Cancer Res 13:4695-4703, 2007.*

GENOME-WIDE ASSOCIATION STUDIES (GWAS) IN NGFN

MICHAEL STEFFENS[1], MAX P. BAUR[1], STEFAN SCHREIBER[2], THOMAS F. WIENKER[1]

[1] INSTITUTE FOR MEDICAL BIOMETRY, INFORMATICS AND EPIDEMIOLOGY, IMBIE, UNIVERSITY OF BONN
[2] INSTITUTE FOR CLINICAL MOLECULAR BIOLOGY, UNIVERSITY HOSPITAL OF SCHLESWIG-HOLSTEIN, KIEL

INTRODUCTION

"Are we ready for genome-wide association studies?" asks an editorial in *Cancer Epidemiology Biomarkers & Prevention* in April 2006.[2] This editorial arose from a critical conceptual and methodological inventory among workers in the field in 2006 (Duncan Thomas, personal communication). Now, not much later than a year, a number of large-scale genome-wide association studies (GWAS) has been published in highest ranking scientific journals (e.g. WTCCC, 2007), and more are to come. The problems succinctly addressed in this editorial, have they all found a solution since then? In the following, we will turn to this question and try to give a preliminary answer, with special reference to the ongoing GWAS within the NGFN consortium.

The GWAS initiative of the NGFN has been launched at the end of 2006; genotype data are coming in since April 2007, and results are expected within short time. Altogether, large patient cohorts of 28 different disease phenotypes with a total of 26 000 DNA samples and approx. 3500 control samples are involved. Each sample is genotyped for 0.55 to 1.0 x 10^6 SNPs. Genotyping is done on either one of the two established high-throughput platforms, Affymetrix® and Illumina®, using advanced microarray-based technology. Controls are typed on both platforms. Phenotypes are from frequent multifactorial diseases which (i) pose a high burden to medical services and thus are relevant under a public health perspective, and (ii) are selected in a way to group into (possibly) pathogenetically related nosological entities. This will allow for the analysis of broader disease entities beyond current concepts of pathology, and eventually will unravel common etiologically relevant genetic factors. It is expected, that an advanced understanding of critical disease processes will be obtained, which conceivably cannot be achieved by other approaches.

Bioinformatics

For a typical GWAS project of 1000 DNA samples, SNP genotype data amount to 0.6 GiB (gigabyte) to 1 GiB, and corollary data for identification and basic characterization of SNP markers and DNA samples will take up another 6 GiB to 8 GiB. Raw data of fluorescence signals on the microarray chips will amount to a total of 40 GiB to 80 GiB of data. Thus, the entire NGFN GWAS initiative will generate and handle more than 5 terabyte of data. Transfer, archiving, maintenance, storage and retrieval of data for fast access is a challenge. Legal requirements of data protection and data privacy are strictly observed.

Data quality, error detection and error filtration

Genotyping costs enjoyed an enormous drop within a short period. A few years ago, cost of a single genotype generated by a typical process of low to medium plex-grade was about 30 cents. Current microarray-based chip technology produces SNP genotypes at approx. 0.05 cents per genotype. Obviously, this made the large-scale GWAS just feasible! High-throughput processes operating under economic constraints are error prone. Even with very low error rates, a sizeable number of genotype errors will accrue, due to high volume. Detection and elimination of these errors is a major task, which has to precede data analysis. A comprehensive and rigorous error filtration protocol is executed on all data by IMBIE Bonn.

Population genetic artefacts deserve a closer look, too. Population stratification and admixture, cryptic relatedness, and others, may, if going unrecognized, exert detrimental effects. Consequently, these confounding artefacts are observed with scrutiny.[3]

Association analysis using single markers

A battery of tests is applied to the data under a single marker approach. Omnibus tests (chi-square goodness-of-fit tests), trend tests, and other tests addressing specified modes of inheritance are routinely employed.

Association analysis using multiple markers

Using more than one marker at a time will result in haplotypic analysis if the markers are closely linked, and in interaction (epi/hypostasis) analysis for unlinked markers. If carried out in a systematic fashion, the latter will involve ½ N^2 tests, where N is the number of markers. This number of tests has been regarded prohibitive until recently. There are good reasons to make human geneticist believe that gene x gene interactions are at the heart of the etiology of many, if not most, complex diseases. The GEM initiative of the NGFN is taking up the challenge.

RESULTS AND INFERENTIAL STATISTICS

The extreme data volume of input data is almost matched by an abundant volume of results, which, in their most compact form,

COORDINATOR
Professor Max P. Baur
University of Bonn
max.baur@ukb.uni-bonn.de

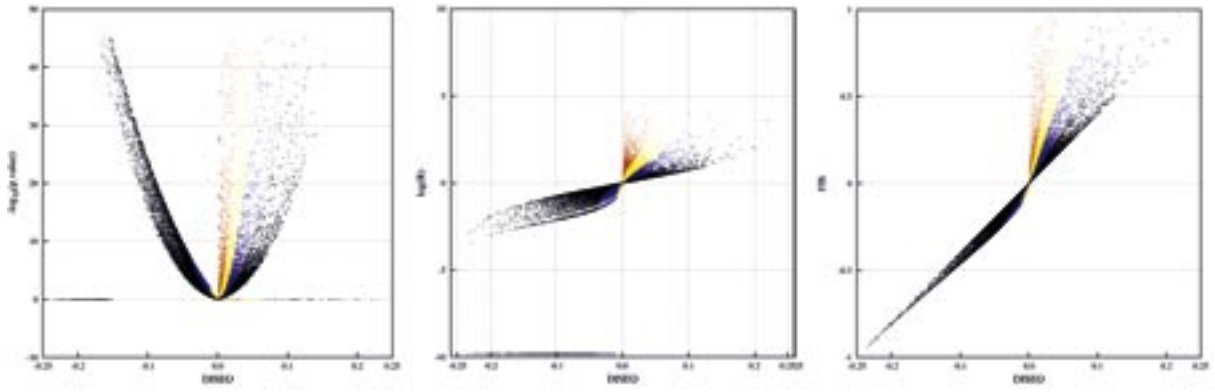


Fig.1. Deviation from Hardy-Weinberg equilibrium: Different effect size measures (DISEQ, FIS, GOR) and negative decadic logarithm of *P*-values. Minor allele frequencies are colour coded (red: <0.05; yellow: 0.05-0.1; blue: 0.1-0.2; black: 0.2-0.5). Outliers are clearly separated.

show up as *P*-values for each test. The number of tests calls for an adjustment of *P*-values for multiple testing, and there is no generally accepted way. Approaches based on the concept of the "false discovery rate" (FDR) have pros and cons, and will be explorated.

Graphical representations of huge amounts of data (correlograms, QQ-plots, bagplots, etc.) are highly useful in this context (see for example figs. 1 and 2). It is a desideratum, to develop **s**tandards **a**nd **r**ecommended **p**rocedures (SARPs) to this end.

CONCLUSION AND OUTLOOK

"Are we ready for genome-wide association studies?" was the initial (mostly rhetoric) question. The answer is "yes"; otherwise the NGFN initiative would have a moot point. However, there are some reservations, and with respect to a few topics briefly touched upon, the answer has to be a shy "not fully yet". Clearly, the rapid and stupendous progress in high-throughput microarray-based genotyping technology overtook the methodological capabilities in genetic biometry, bioinformatics, and genetic epidemiology. These closely related disciplines are striving for a catch-up, and there are notable achievements. Obviously, the full benefit of the NGFN GWAS collaborative effort will be achieved if methodologies in molecular genetics, epidemiology, and biostatistics are at the same level. The road map is clearly defined, but there is as well uncharted ground ahead: a challenge, rather an obstacle!

References

[1] *WTCC (The Wellcome Trust Case Control Consortium) (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661-678*

[2] *Thomas D (2006) Are we ready for genome-wide association studies? Cancer Epidemiol Biomarkers Prev 15(4):595-598*

[3] *Steffens M et al. (2006) SNP-based analysis of genetic substructure in the German population. Hum Hered 62(1):20-29*

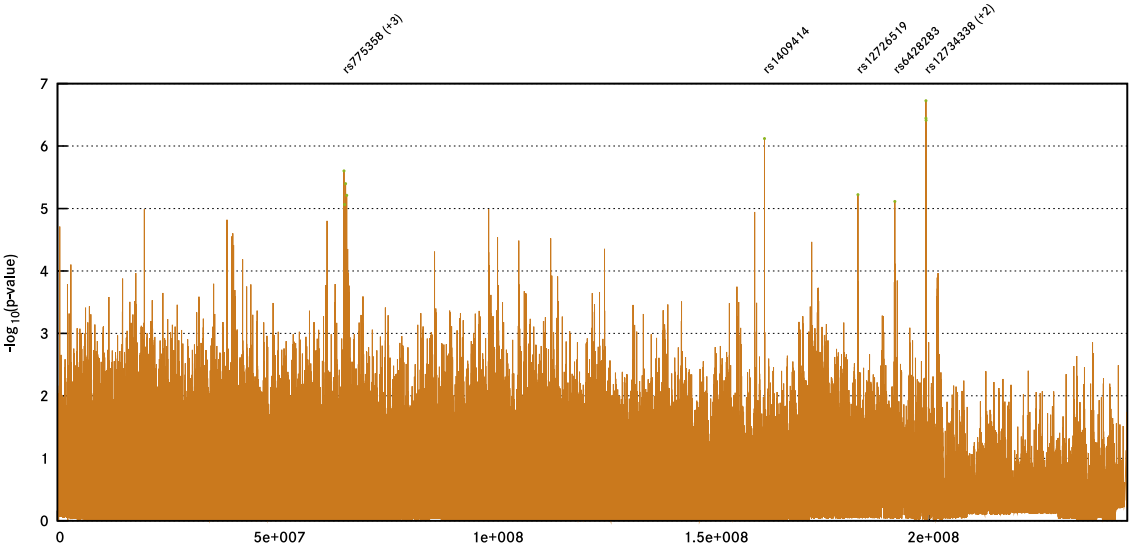


Fig.2. A typical GWAS experiment: Genomic position for each SNP on the X-axis (bp according to NCBI dbSNP build 126), and negative decadic logarithm of association *P*-values (Armitage's trend test) on the Y-axis. Putative hits which require closer scrutiny are marked by top dots and rs numbers.

POPGEN AND KORA-GEN: BIOBANKING FOR NGFN

MICHAEL KRAWCZAK[1], H.-ERICH WICHMANN[2], TIMOTHY T. LU[1]

[1] INSTITUTE OF MEDICAL INFORMATICS AND STATISTICS, UNIVERSITY HOSPITAL OF SCHLESWIG-HOLSTEIN, KIEL
[2] INSTITUTE OF EPIDEMIOLOGY AT THE GSF – RESEARCH CENTER FOR ENVIRONMENT AND HEALTH, MUNICH-NEUHERBERG

INTRODUCTION

Human populations exhibit substantial (and hitherto mostly unexplained) biological variability of profound importance to both the etiology of disease and the therapeutic prospects of their treatment. The scientific questions pertaining to this variability can only be fully addressed by patient-based research which, in turn, relies on access to high-quality collections of biomaterials and data. In this context, the medical research community is currently undergoing a cultural change. More and more national and international infrastructure is being created for long-term and interdisciplinary collaborations, which often entail the establishment of centralized data and sample collections, so-called 'biobanks'. Biobanks are not focused on single scientific projects but are intended to support a wide variety of future, possibly as yet unknown, research goals. In this respect, biobanks are different from individual, context-related clinical resources that may well be easier and cheaper to establish initially, but which because of their very multitude have inherent limitations which render them inefficient and inappropriate in the long term.

For most common human diseases, recruitment of samples for disease gene finding has been limited to carefully selected cases of particularly pronounced phenotype, either in terms of severity or familiarity. This strategy of 'extreme sampling' is based upon the expectation of easily identified so-called 'major genes', i.e. genes with strong effects regarding a particular disease risk. In contrast, unselected (i.e. 'normal') populations are used in the subsequent development of diagnostics and therapeutics which result from disease gene discoveries. For this to be feasible, genotype-specific disease risks in the unselected population must be determined, which can only be estimated from population-representative samples. Unbiased sampling of patients is best achieved at the population level by defining a confined geographical catchment area in which all clinically overt cases with the disease in question are recruited, either prospectively, retrospectively or in cross-sectional fashion. Typical examples of the successful pursuit of such a strategy are provided by PopGen and KORA-gen, two biobanks endorsed by the NGFN that serve different, geographically targeted recruitment areas in Northern and Southern Germany.

PROJECT STATUS/RESULTS

The PopGen project was initiated by clinical and non-clinical partners at the University of Kiel in 2002 to provide disease-orientated projects in the NGFN with a platform for the identification and cross-sectional recruitment of patients and controls. Since October 2004, PopGen has been funded as part of the GEM platform in NGFN-2. Reliable and efficient recruitment has been ensured by confining sampling activities to northern Schleswig-Holstein, using resources of the German public health care system, and by the quality-controlled generation and processing of clinical data. The biomaterial currently stored at PopGen comprises mainly whole blood and DNA. As of June 2007, more than 75 000 samples were available that corresponded to some 30 000 individuals (patients and controls). In addition, sera and other diverse biological samples have been included into PopGen for a number of individuals in disease-specific fashion. Depending on the phenotype in question, up to 250 different items of information are logged for each proband. SNP genotype data have been generated for several patient and control groups. The list of phenotypes currently under study at PopGen includes bipolar affective disorders, colon carcinoma, coronary heart disease, essential tremor, gallstones, inflammatory bowel disease and Parkinson's disease. In addition, PopGen hosts data and samples for a study on the genetic causes of healthy aging, and a variety of other control groups and cohorts.

KORA-gen was established by the GSF – National Research Center for Environment and Health in Munich-Neuherberg to facilitate genetic epidemiological research on the KORA platform ("Cooperative Health Research in the Region of Augsburg"). KORA-gen resources are available for collaborative research projects and have been used successfully in more than 100 national and international co-operations. The goal of KORA-gen is to provide access to quality-controlled genotypic and phenotypic data as well as biomaterials. KORA-gen includes 18 000 subjects from the general population, recruited in the age range of 25 to 74 years. In addition, follow-up investigations on more than 6000 subjects in the age range of 35 to 90 years have been performed. A repository has been established that contains samples (serum, plasma, DNA, urine) of all subjects (half of them at several time points), including immortalized cell lines of 2100 individuals. Clinical and preclinical data on the following disease pheno-

COORDINATOR
Professor Max P. Baur
University of Bonn
max.baur@ukb.uni-bonn.de

types, among others, are available in KORA-gen: diabetes, myocardial infarction, stroke, obesity, COPD, asthma, atopic dermatitis, thrombosis, and cancer. In addition, KORA provides a wealth of subject-related environmental and life style information (e.g. physical activity, smoking, nutrition, alcohol intake) as well as measurements of laboratory parameters (e.g. cholesterol, triglyceride, glucose). KORA-gen has been partly funded through the GEM platform in NGFN-2.

OUTLOOK

The availability of population-based patient and control samples will be of pivotal importance for clinical genetics research even after NGFN-2 funding ceases. In addition to continuing the provision of samples to their collaborative partners, PopGen and KORA-gen will endeavor to refine their existing resources. Expansions and additions will be made, for example, by including biopsies or other new tissue types, and by providing in-depth follow-up data to allow genetic information to be put into the context of whole disease courses. The two biobanks will also be closely linked with complex statistical genetics projects including, for example, the genome-wide disease association studies initiated during the late phase of NGFN-2.

Fig. 1: Scrutinizing proband data at the PopGen sample intake laboratory at the ICMB in Kiel. From left to right: Huberta von Eberstein (Project Manager), Stefan Schreiber and Michael Krawczak (Principal Investigators).

The samples and data in PopGen and KORA-gen will continue to promote the medical research community's understanding of the practical importance of new disease gene discoveries. They will therefore fulfill an important bridging function within

Fig. 2: KORA-gen is based upon the research platform "Cooperative Health Research in the Region of Augsburg" (Principal Investigator: H.-Erich Wichmann, GSF).

the NGFN successor programs. While PopGen is a population-based resource, providing phenotype information on diseases of low and moderate prevalence, KORA-gen is focused on the prospective evaluation of genetic risk factors and therefore has particular merit in the study of higher prevalence conditions. Together, both biobanks complement each other perfectly for the potential benefit of future genetic epidemiological research programs.

References

[1] *Krawczak M et al. (2006) PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. Community Genet 9: 55-61.*

[2] *Steffens M et al. (2006) SNP-based analysis of genetic substructure in the German population. Hum Hered 62: 20-29.*

[3] *Wichmann HE et al. for the MONICA/KORA Study Group (2005) KORA-gen - Resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen 67: S26-S30.*

In addition, data and samples from PopGen and KORA-gen formed the basis of more than 50 research articles published in high-ranking scientific journals.

RZPD – INFRASTRUCTURAL BASIS FOR GENOME RESEARCH

UWE RADELOF[1], BERND DRESCHER[1], ANDREAS HÖRLEIN[2], BERNHARD KORN[3], ANJELA KOLEGAI[2], DIRK SCHNEIDER[2], PETER NEUBERT, ANDREAS HEWELT, JOHANNES MAURER[2]

RZPD GERMAN RESOURCE CENTER FOR GENOME RESEARCH, BERLIN
CURRENT ADRESSES: [1] ATLAS BIOLABS GMBH, BERLIN; [2] IMAGENES GMBH, BERLIN; [3] DKFZ, HEIDELBERG

INTRODUCTION

Integrated genome research requires the costly generation and utilization of large molecular biology resources, e.g. clone libraries, protein expression clones, siRNA resources for the investigation of gene function etc. These are essential resources for biomedical research aiming at the improvement of human health. Today, they already consist of scores of millions of samples and their numbers are increasing steadily. Maintaining the long-term integrity and safety of this indispensable material, while guaranteeing public access for all researchers, is essential for the National Genome Research Network (NGFN) and molecular biology research in Germany in general. For this purpose, RZPD continuously generated and also imported relevant research material and applied stringent quality controls to assure a consistently high standard. RZPD ensured reliable and unrestricted access to these resources and all material that was generated and made publicly available by NGFN for a fee on a cost-recovery basis. These resources and related services were provided on the basis of a reliable, adequate laboratory infrastructure. In order to make sure that the experimental value of the biological material is not degraded over time, RZPD had built a unique system of logistics, maintenance, and quality control procedures. Within RZPD, all production, quality control and validation processes were performed and documented according to ISO9001 standard operating procedures (SOPs).

FROM FUNCTIONAL GENOMICS TO PROTEOMICS

A characteristic feature in molecular life sciences is the remarkable development of research and technologies in the proteome area. As an analogue to HUGO, the “Human Proteome Organization” (HUPO) was founded in 2001 in order to identify the most important problems in proteomics, to start appropriate initiatives to address these issues and to support the development of new technologies. Several large-scale proteomics projects have been launched since then, e.g. the Plasma Proteome Project[1], or the Human Antibody Initiative[2].

This development was reflected at RZPD by a coordinated palette of products and services for protein- and proteomics-related research, starting with various expression clones, via the expression and purification of proteins, to the point of functional studies. More than 34 000 full open reading frame (ORF) shuttle clones and nearly 430 000 full ORF expression clones derived thereof were available.

RZPD’s protein expression service included the choice of the protein of interest from the respective clone collection, including the full ORF expression clones. 5-10 mg of protein were expressed in an *E. coli* or a Baculovirus expression system and purified up to 90% by FPLC. Three protein array services were available, namely serum screening and antibody epitope mapping using RZPD’s protein arrays, and protein expression profiling using Becton-Dickinson (BD) antibody arrays. RZPD’s protein arrays were derived from several human tissues including fetal brain, testis, and T-lymphocytes. Each protein array consisted of up to 27 648 *E. coli*-expressed proteins, printed in duplicate on a PVDF membrane. The serum screening service on these arrays was designed to screen for autoantibodies in patient plasma or serum, which was of special interest for clinicians and researchers dealing with autoimmune diseases like diabetes type I or multiple sclerosis. An automated yeast two-hybrid (Y2H) interaction screening service was also available. In contrast to most traditional Y2H methods, a fluorogenic reporter was used which could be accurately quantified.[3]

SYNTHETIC BIOLOGY

Although RZPD offered the most comprehensive clone collection worldwide with 35 millions of clones and more than 1000 cDNA libraries, this collection was far from being complete, even for the most important model organisms. This gap could be filled with synthetic cDNA clones generated via gene synthesis, which guarantees 100% sequence identity to a database sequence and allows for sequence optimization to maximize protein yields in a certain expression system. On account of this, RZPD recently integrated synthetic clones into its portfolio, which allowed offering a virtually infinite clone collection. Synthetic clones could be identified and ordered by RZPD’s online search tool GenomeCube® and were delivered within 4-6 weeks. Gene synthesis also represents an ideal complement to next-generation sequencing. The classical way from the clone to the sequence is simply reverted: the sequence information is generated first – without classical cloning, directly starting from DNA – and then the clone needed, e.g. for protein expression, is synthesized with 100% compliance to the specified requirements.

COORDINATOR
Dr. Johannes Maurer
German Resource Center for Genome Research, Berlin
j.maurer@rzpd.de

Fig. 1: RZPD's GenomeCube. The GenomeCube offered direct access to a multitude of validated clone materials corresponding to the search term provided. Here, MAOA (monoamine oxidase A) was entered, for which a variety of different clones from several organisms were available.

BIOINFORMATICS AT RZPD – STRUCTURE OF RZPD'S PRIMARY DATABASE

During the last years, the structure of RZPD's primary database had been completely re-organized in order to meet the requirements of gene-centered queries for biological material, and to provide RZPD's customers with comprehensive and up-to-date annotations for these materials. With respect to different details and information, each major product group was stored in a product database of its own, e.g. clones, clone sets and libraries, tissue arrays, etc., where each product had one or more references to genes. These references were the basis to connect the product data with all kinds of annotations, e.g. sequence data, functional annotation, disease related information, etc. The annotation was retrieved from different sources. On the one hand, the submitters of biological material provided basic information with their submission. On the other hand, RZPD's bioinformatics group spent an enormous effort to improve the level of annotation with data from public resources. Data from Uniprot, Unigene, OMIM, Ensembl, Entrez, GeneOntology, KEGG, BioCarta, and many more had to be interlinked with each other, the product databases, and the RZPD services. The resource for up-to-date data annotations was MASI@RZPD. MASI (developed by Insilico Software GmbH), "Meta Annotated Sequence Investigation", is a compilation of public biological databases into a locally available relational database management system (RDBMS). The increasing spectrum of information provided covers all relevant sources of genetic, proteomic and metabolomic data. Therefore, particular biological entities cited at different sources become comparable and thus provide more comprehensive information. MASI was updated on a bi-weekly basis from the public resources.

PRODUCT SEARCH AND ORDERING WITH THE GENOMECUBE®

The GenomeCube™, an intuitive-to-use interface for gene-based material retrieval, was named after the three "dimensions" required to identify the most appropriate resources for the gene(s) of interest: gene, species, and clone/product type. Instant access was given to validated materials, e.g. sequence-verified cDNA clones, full length clones, full ORF clones, siRNA resources, etc. by entering a variety of search terms, e.g. gene symbols and descriptions, GenBank Accession numbers, Unigene Ids, Ensembl Ids, and Affymetrix Ids. The result of a GenomeCube™ query was a list of appropriate products (Fig. 1).

Comprehensive information could be obtained by a single mouse click, e.g. gene information, functional annotation details, disease related information, sequence data, and many more. Nevertheless, the search for appropriate biological material is not always straightforward. Therefore, to find biological material that is related to each other by the same genetic context, a new concept named "GeneContext" had been introduced into the GenomeCube®: genes were represented in (i) Positional Gene Sets, (ii) Motif Gene Sets, (iii) Curated Gene Sets, (iv) Computational Gene Sets provided by the Broad Institute laboratories (http://www.broad.mit.edu/gsea/msigdb/msigdb_index.html), (v) the controlled vocabulary used for indexing articles for MEDLINE/PubMed (MeSH Terms), and (vi) Disease Concepts provided by the Autoimmune Disease Database (http://www.sbi.uni-rostock.de/aidb/home.php).

OUTLOOK

RZPD was closed on July 31, 2007. However, most of the services have been continued or even expanded by two spin-off companies, ATLAS Biolabs GmbH (www.atlas-biolabs.de) and ImaGenes GmbH (www.imagenes-bio.de).

References

[1] *Omenn GS et al. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3,020 proteins and a publicly-available database. Proteomics 2005, 5: 3226-3245.*

[2] *Nilsson P et al. Towards a human proteome atlas: high-throughput generation of mono-specific antibodies for tissue profiling. Mol Cell Proteomics 2005, 5: 4327-37.*

[3] *Albers M et al. Automated yeast two-hybrid screening for nuclear receptor-interacting proteins. Mol Cell Proteomics 2005, 4: 205-13.*

RZPD'S EXPRESSION PROFILING PLATFORM

ANDREAS HÖRLEIN[1], MARTINA SCHAD[1], FLORIAN WAGNER[2], ROBERT WAGNER[1], KARSTEN R. HEIDTKE[2], BERND DRESCHER[2], STEFFEN HENNIG[1], UWE RADELOF[2], CHRISTIAN MÄRCKER[3] AND JOHANNES MAURER[1]

RZPD GERMAN RESOURCE CENTER FOR GENOME RESEARCH, BERLIN
CURRENT ADDRESSES: [1] IMAGENES GMBH, BERLIN; [2] ATLAS BIOLABS GMBH, BERLIN; [3] UNIVERSITY OF APPLIED SCIENCES, MANNHEIM

Based on its long-term experience as a service facility in human genome research, its comprehensive resource collection, a quality management system according to international ISO9001 standards and the continuous development of new resources for applied genome research according to the requirements of the researchers, RZPD has provided a "Central Unit for Services and Resources" to all researchers within NGFN-2. Centralized, standardized and quality-certified biological resources and laboratory services are the most time- and cost-efficient solution for both NGFN-researchers and funding agencies. This approach leads to an overall reduction in investments (less redundancy), and an optimal utilization of existing facilities is achieved.

RZPD has long-standing experience in generating and applying resources for gene expression profiling experiments. RZPD provides a wide variety of products and services for gene expression profiling, genotyping and microarray-based functional assays.

RZPD has been an approved Affymetrix Service Provider in Europe since August 2001. To date, several thousand expression analyses have been successfully performed using all currently available GeneChip™ expression arrays. This service is also provided to several groups of NGFN-funded networks (e.g. Infection/Inflammation Network: Dr. Thomas Häupl, Charité, Berlin; Dr. Michal Janitz, Max Planck Institute for Molecular Genetics, Berlin, and Assistant Professor Jörg Hoheisel, DKFZ Heidelberg).

A bioinformatics link between Affymetrix GeneChips™ and RZPD´s Unigene Sets or other validated clone resources at RZPD (e.g. full ORF expression clones) has been established. This link makes it possible to combine both resources in order to run large projects cost-efficiently: Initial genome-wide profiling using Affymetrix GeneChips™ is carried out in order to identify indication-specific subsets of genes of interest. Afterwards, customized microarrays or functional studies can easily be performed using these resources

During the NGFN-2 funding period two additional DNA Array-based platforms have been added: NimbleGen service and service using Agilent arrays. NimbleGen's array synthesis technology enables highest flexibility in array designs and is therefore predestined for small-scale custom experiments. This opens up the possibility of profiling gene expression in any organism with available sequence information. Therefore, RZPD's NimbleGen Service platform is favored for organisms where no Affymetrix catalog arrays are available, e.g. unusual prokaryotes.

The Agilent expression profiling platform complements RZPD's microarray platforms with an important feature while having the same flexibility in terms of the design: the availability of multiplex arrays. The parallel hybridization of up to eight samples on a single chip reduces the costs of expression profiling experiments enormously. Global gene expression array experiments can be performed to identify genes of interest. By resulting condensed multiplex arrays one can focus on the analysis of these genes in further experiments. The Agilent platform is furthermore employed for a two-step strategy to achieve high-quality multiplex arrays even for large genomes. This strategy comprises the generation of a first design for a high density array (e.g. 244K arrays) in which each gene was tiled by up to ten different oligos. The hybridization of a complex sample allowed the evaluation of the performance of each oligo. At RZPD a strategy was developed to select the best performing two to four oligos for a condensed design for multiplex arrays (e.g. 4x44K).

RZPD's extended data analysis pipeline for expression profiling experiments, initially developed for Affymetrix GeneChips™, could be successfully adapted to the requirements of Agilent and NimbleGen data outputs and has enjoyed great popularity. This software offers a standardized analysis and helps to control the quality of the experiments using well-established and tested statistical methods and clustering algorithms. A comprehensive up-to date annotation from publicly available data sources is also included. This analysis is not possible with standard tools provided by the chip manufacturers. Data formats from Affymetrix, Agilent or NimbleGen experiments, or gene/clone tables, which are formatted as hybridization expression values per column, are accepted. Scientists are provided with an extended analysis for gene expression profiling experiments that enables them to evaluate their experiments at a glance from a set of standardized parameters determined using RZPD's expertise.

In the postgenome era the platforms had to be extended from transcription (RNA) to protein expression profiling. Protein expression profiling was established on commercially available antibody arrays. Two protein extracts were labeled with

COORDINATOR
Dr. Johannes Maurer
German Resource Center for Genome Research, Berlin
j.maurer@rzpd.de

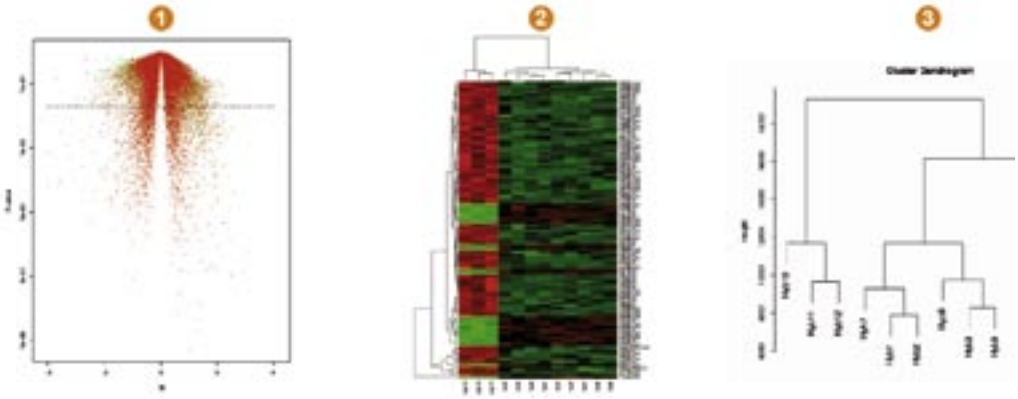


Fig. 1: Excerpt from the Microarray Analysis Report: (1) The volcano plot shows the relation between the p-value and log2-fold change based on the hybridization intensities. (2) The heat map shows a two-dimensional clustering of the top one hundred genes. Genes that fall into one cluster (vertical axis) have similar behaviors in the experiments. Hybridizations with similar behavior fall into one cluster (horizontal axis). It can be seen that in this experiment it suffices to focus on the top one hundred genes, to correctly differentiate between experiment conditions. Expression intensities are represented by red and green, for high and low intensities, respectively. Black indicates medium intensities. (3) Clustering hybridization experiments by their expression values means measuring the similarity of regulation between two experiments. The clustering is based on normalized expression values. The clusters are arranged along the x-axis and the dissimilarity values between the clusters form the y-axis.

Cy3 and Cy5, and the protein expression data were determined in accordance to the classic cDNA Array approaches. After successful protein profiling experiments on antibody arrays in several research projects, it was offered as a service. Moreover, a competent partner enabled us to integrate genome analyses on tissue microarrays (immunohistochemistry, DNA *in situ* hybridizations, RNA *in situ* hybridizations) in our molecular genome analysis platform.

However, we have concentrated our work on projects involving RZPD protein expression arrays. RZPD protein expression arrays (cooperation with Protagen AG) are based on libraries of tens of thousands of recombinant proteins derived from a variety of human tissues. The construction of RZPD's protein expression arrays uses an *E. coli*-based expression system for recombinant proteins. Since these recombinant proteins are expressed in *E. coli* and are extracted under denaturing conditions, it cannot be expected that the arrayed proteins are correctly folded and fully functional. Nevertheless, RZPD's protein arrays are very well suited for applications like patient serum profiling (screening for active auto-antibodies) or epitope mapping (screening for the epitope of antibodies with unknown specificity).

Each protein array consists of up to 27 648 proteins, which are printed in duplicate (total up to 55 296 protein spots) onto PVDF membranes of sizes up to 22 cm x 22 cm. These proteins are derived from expression-verified, full-length as well as shorter cDNA clones. Individual expression verified cDNA clones are available for in-depth follow-up experiments.

Fig. 2: Principle of protein detection: Each recombinant protein of the protein expression array is immobilized on a filter (PVDF) membrane. During a serum screening experiment the array is incubated with serum (primary antibody). In a second step the bound primary antibodies are detected with an anti-human IgG-Ig antibody, which is coupled to alkaline phosphatase (AP). The AP activity is then visualized using ECR® reagent.

In the first half of NGFN-2 this method was developed for use in a screening service facility of the RZPD. Since July 2006 approximately 60 serum screening experiments have been performed for NGFN-2 internal and other partners. One specific experiment series for Professor Hans-Peter Lenhof (University of Saarbrücken) led to the generation and production of a subarray of roughly 1000 clones, which will be used for the screening of 100 patient sera with the goal to identify novel biomarkers for tumor diagnostics. In addition numerous epitope mappings were performed during this period.

Further applications for protein expression arrays such as studying protein-protein interactions or the identification of kinase phosphorylation sites (cooperation with Tim Hucho, MPI Berlin) have been developed during the later phase of the funding period of NGFN-2.

With arrayed proteins derived from three different human tissues, testis, lymphocytes, and fetal brain, RZPD already offers one of the largest collections of arrayed proteins for screening experiments. In an effort to broaden the bases of RZPD's protein expression arrays and to enhance their quality, RZPD started two initiatives. First, the generation of a liver cDNA-library and a human ES-cell library was initiated. With these two additional tissues/cell types we will strongly enlarge the selection of proteins presented on RZPD protein expression arrays. Secondly, we addressed the challenge that not all proteins derived from tissue-based libraries can easily be identified by the DNA sequence of their corresponding cDNA clones. The Protein Expression Array proteins are expressed in *E. coli* as fusion proteins with an N-terminal His-Tag. To ensure that all determined sequences lead to a proper in frame protein, we sequenced all cDNA clones of the libraries in stock and assembled two new libraries, which are called the UNIPEX libraries. In these libraries we are aiming at a redundancy of three for each protein, i.e. the protein is represented by three or more cDNA clones.

NOVEL LINK BETWEEN GENOMICS AND FUNCTIONAL PROTEOMICS IMPACTS FUTURE CANCER IMMUNOTHERAPY

ANDREAS O. WEINZIERL, MATHIAS M. SCHULER, HANS-GEORG RAMMENSEE

UNIVERSITY OF TÜBINGEN

INTRODUCTION

Fig. 1: Prof. Dr. Hans-Georg Rammensee

Surgical removal of tumor tissue, as well as chemotherapy and radiation, are the conventional weapons in the struggle against cancer. Although these methods are routinely in use, they often do not lead to lasting success. Therefore, combating cancer by activating the immune system appears to be a promising strategy for the development of new cancer treatments. In this strategy, it is crucial to identify immune targets which are characteristic for tumor tissue and in turn bear only a minimal risk of immune reaction against normal tissue. These target structures, so-called tumor-associated antigens (TAAs), can be separated into several groups according to their basal expression level in normal tissues. Proteins expressed at low levels in normal tissue, but high in the tumor, form one class of TAAs. Regarding overexpression, the vital parameter for immunotherapies based on the effects of cytotoxic T cells (CTLs) is the epitope concentration on the cell surface. Epitope concentration refers to the relative frequency of one specific MHC-peptide complex on the tumor cell surface. Thus far, no extensive studies investigating frequency ratios of MHC ligands on tumors have been undertaken. To address this issue, we developed a method enabling us to measure exact quantitative differences of MHC-bound peptides between tumor and autologous normal tissue[1].

TAAs in tumor immunotherapy

Proteins overexpressed in cancer patients provide a source for potential tumor-associated CTL epitopes. Prominent examples include HER2/Neu, MUC1, CEA or p53. Despite their lack of absolute tumor-specificity, MHC-bound peptides from such proteins were shown to elicit T-cell responses directed against tumors *in vitro* as well as in clinical trials. This suggests that the search for overexpressed antigens could provide a potentially successful method of cancer immunotherapy. In most cases, overexpression of TAAs refers to the protein or mRNA level, usually measured with antibodies, PCR or microarrays. However, since CTLs recognize peptides bound to MHC molecules on the target cell surface, the density of such complexes bearing the peptide of interest – rather than mRNA or protein levels – will be the key determinant for defining overexpression in this system. Indeed, epitope density was reported to be an important parameter for T-cell priming as well as for target cell lysis in the CTL effector phase.

Commitment of our laboratory

For more than ten years we have been isolating and analyzing MHC-bound peptides from cell lines and tissue samples. This has led to the characterization of more than 50 allele-specific peptide motifs. By applying these peptide motifs, we have established prediction algorithms for T-cell epitopes that are accessible to the public via our Internet database “SYFPEITHI”. For the identification of peptides, we now separate MHC-derived peptides by liquid chromatography online coupled to a mass spectrometer. This online coupling enables sequence analysis by tandem mass spectrometry (MS/MS). Applying this technology, we have developed a technique to detect and verify predicted potential T-cell epitopes which resulted in the identification of a new epitope from the tumor antigen MAGE-A1[2].

PROJECT STATUS/RESULTS

The aim of this project, which combines genomics and functional proteomics, was to compare specific MHC-peptide complex densities on tumor and autologous normal tissue by relative quantification. Furthermore, we wanted to assess the extent of correlation between MHC-peptide complexes and the expression levels of their corresponding mRNAs.

In an initial step we started to pursue a strategy to identify as many MHC-presented peptides as possible because these are potential candidates for tumor-associated T-cell epitopes in individual patients. Currently, we are able to identify more than 100 MHC ligands from one tumor sample on a routine basis[3]. In this first approach, we additionally combined MHC-ligand identification with gene expression analysis of individual tumors to identify peptides corresponding to overexpressed or exclusively expressed genes in the tumor[4]. Thus, we established the Affymetrix GeneChip® technology in our laboratory five years ago[5]. Potential epitopes identified by this combined approach have already been successfully trans-

PROJECT LEADER
Professor Hans-Georg Rammensee
University of Tübingen
rammensee@uni-tuebingen.de

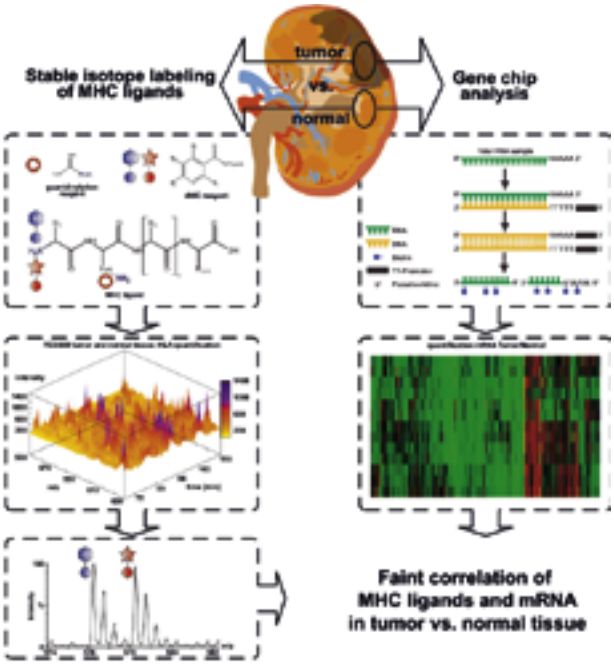


Fig. 2: Schematic drawing of the approach for a quantitative comparison of the MHC-bound peptides and their corresponding mRNA levels. From pairs of tumor and autologous normal tissues, both MHC peptides and mRNA were isolated and quantified. MHC derived peptides were labeled with tissue specific tags and afterwards quantified and sequenced by mass spectrometry. mRNA analysis was done by gene chip analysis.

ferred into clinical vaccination studies against renal cell carcinoma. Some peptides identified in this first step have already been verified as true T-cell epitopes mediating tumor cell lysis.

In a second step, the quantitative MHC-peptide complex analysis developed in our laboratory was used for a comparison of several renal cell carcinoma samples with their autologous normal tissues. In this approach, MHC-bound peptides from tumor and normal tissue were linked to a tag which was specific for the respective tissue from which the peptide was derived (see schematic illustration on the right). Subsequently, tumor and normal tissue-derived, labeled peptides were mixed in an average ratio of 1:1. In a detailed analysis, the introduced tag allowed a relative quantification of hundreds of individual peptides. In these experiments we could identify more than thirty MHC-peptide complexes as being highly overpresented on tumor cells.

These peptides are excellent candidate TAAs which are specifically overpresented on tumor tissue. After validation in T-cell experiments, they will be used directly in future vaccination trials. Furthermore, we could show in our experiments that the correlation of MHC-peptide complex density with the corresponding level of their mRNA is only marginal[6]. As a direct consequence, this lack of correlation indicates that application solely of the very efficient, relevant DNA microarray technology is unfortunately not a very promising strategy in identifying potential strong candidates of tumor-associated CTL epitopes. As these results confirmed that our method is easily applicable in individual patient sample analysis, we have started to cooperate with Professor Oliver Kohlbacher, a specialist in applied bioinformatics and especially mass spectrometry. In a joint project we attempted to automate the analysis of individual peptides, which is still very time-consuming and laborious. The first version of this software enabling high-throughput quantification of tagged MHC peptides was released in July 2007. Thus, we have provided both a theoretical and a practical basis for routine identification of overpresented MHC peptides which is a prerequisite for successful tumor immunotherapy.

References

[1] *Lemmel et al. Differential quantitative analysis of MHC ligands by mass spectrometry using stable isotope labeling. Nat Biotechnol. 2004 Apr; 22(4):450-4.*

[2] *Pascolo S et al, A MAGE-A1 HLA-A A*0201 epitope identified by mass spectrometry. Cancer Res. 61, 4072-4077 (2001).*

[3] *Krüger et al. Lessons to be learned from primary renal cell carcinoma: novel tumor antigens and HLA ligands for immunotherapy. Cancer Immunol Immunother. 2005 Sep; 54(9):826-36.*

[4] *Weinschenk T et al. Integrated functional genomics approach for the design of patient-individual antitumor vaccines. Cancer Res. 62, 5818-5827 (2002).*

[5] *Kramer et al. MAGED4-Expression in Renal Cell Carcinoma and identification of an HLA-A*25-restricted MHC class I ligand from solid tumor tissue. Cancer Biol Ther. 2005 Sep;4(0):943-8.*

[6] *Weinzierl et al. Distorted relation between mRNA copy number and corresponding major histocompatibility complex ligand densitiy on the cell surface. Mol Cell Proteomics. 2007 Jan; 6(1):102-13.*

REGULATION OF MESSENGER RNA SPLICING IN THE HUMAN SYSTEM: A COMBINED RNA INTERFERENCE AND MICROARRAY APPROACH

ALBRECHT BINDEREIF

INSTITUTE OF BIOCHEMISTRY – FACULTY BIOLOGY AND CHEMISTRY – UNIVERSITY OF GIESSEN

INTRODUCTION

This explorative project focuses on a novel approach to describe messenger RNA (mRNA) splicing on a genome-wide level, in particular its regulation. Typically, eukaryotic protein-coding genes are first transcribed as precursor mRNAs (pre-mRNAs) that contain their protein-coding information in exon regions, interrupted by long, non-coding intron sequences. This exon-intron organization can be very extensive. For example, in the human dystrophin gene the protein-coding information for this important muscle protein is spread over 78 exons, with interspersed intron regions of sometimes more than 100 000 nucleotides! The accurate recognition of the splice sites is essential for correct gene expression and requires splice signals, in particular at the exon-intron junctions. Each of the introns is excised by the spliceosome, a large RNA-protein complex assembled on the pre-mRNA and composed of more than 100 proteins and five small nuclear RNAs.

RNA splicing is even more complex because a single primary transcript often generates more than one mRNA product by using different splicing patterns, a process called alternative splicing (Fig. 1). Alternative splicing includes five basic patterns: Exons can be extended or shortened by the use of alternative splice sites (types A and B), skipped or included (type C), or included in a mutually exclusive manner (type D); in addition, introns can be either removed or retained (intron retention: type E). Several splicing patterns are often combined and occur in the same cell, their ratio being characteristic for a certain cell type or tissue. Protein isoforms arising from alternative splicing display functional differences, for example in their protein-protein interaction properties, cellular localization, protein folding, and enzymatic activity.

Such alternative splicing processes that produce functionally different protein products provide an important mechanism for controlling gene expression in a cell-type-, tissue-, and development-specific manner. Unexpectedly, the human genome project revealed that only 30 000 genes or even fewer suffice to specify development of the complex human system. Alternative splicing is certainly a very important source of protein complexity, in addition to alternative use of promoters and 3' end processing sites, RNA editing, and post-translational protein modifications. Estimates of how commonly alternative splicing occurs now reach more than 70% of the human protein-coding genes: Alternative splicing is more the rule than the exception.

How are these various splicing patterns regulated in different tissues and developmental stages? This brings us back to the machinery, because in addition to the basic splicing apparatus (the spliceosome), regulatory proteins can modify the splicing patterns by binding to splice-regulatory sequences on the pre-mRNA, either in exon or intron regions. These splice regulators usually occur only in certain tissues or developmental stages and can interact – either as activators or repressors – with the general spliceosome, which is present in all cell types. Our general model of splicing regulation is that relatively few activators or repressors (on the order of dozens) recognize many target genes in the human genome, using conserved binding sites; it is the combinatorial action that brings about a complex regulatory network between the regulators and the target genes.

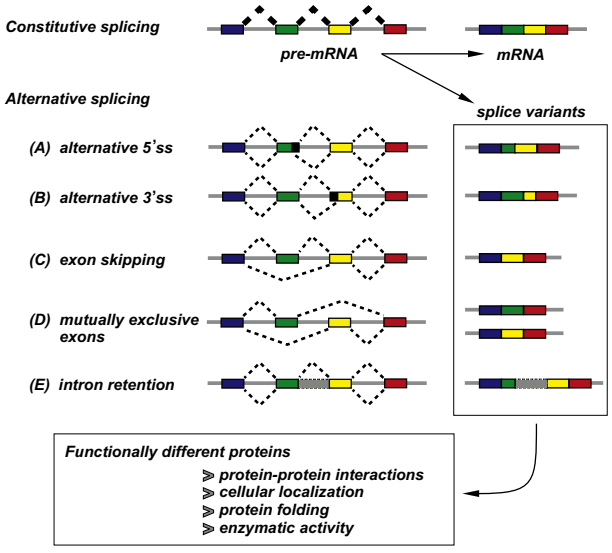


Fig. 1: Constitutive and alternative messenger RNA splicing. Alternative splicing patterns (A–E) generate from a single precursor (pre-)mRNA splice variants coding for functionally different proteins (for details, see text).

PROJECT STATUS/RESULTS

To identify entire networks between splicing regulator proteins and their target genes, we have focused our efforts initially on a specific factor called heterogeneous ribonucleoprotein L (hnRNP L), an abundant nuclear protein,

PROJECT LEADER
Professor Albrecht Bindereif
University of Giessen
albrecht.bindereif@chemie.bio.uni-giessen.de

which functions as a global regulator on the level of mRNA splicing.[4,3,2] Our first genome-wide search for alternative splicing targets of hnRNP L, based on its binding specificity and available EST data, yielded only few target genes.[3] Considering the wide abundance of CA-repeat and C/A-rich sequences as well as the high abundance of hnRNP L protein, it seemed likely there are many more targets in the human genome.

To effectively identify more hnRNP L targets that are regulated in their alternative splicing pattern, we have developed a combination of an exon-specific microarray platform (Affymetrix Human Exon 1.0 ST array) and RNAi-mediated knockdown, with the aim of searching for alternative splicing changes in response to hnRNP L depletion (Fig. 2).

The Affymetrix Human Exon 1.0 ST array (released in September 2005) covers all human genes on a single chip, carrying more than 5 million probes specific for all known expressed sequences. How do we detect all RNP L targets on a genome-wide basis? HnRNP L (and corresponding controls) were downregulated in HeLa cell culture by RNA intereference, RNA samples were prepared from three biological replicates, followed by hybridizing and scanning according to Affymetrix standard protocols (in collaboration with Dr. Vladimir Benes, Gene Core Facility, EMBL, Heidelberg). Most importantly, we have developed the data analysis methods and algorithms to test the validity of the splice array. As a result, we have identified 12 new target genes of hnRNP L, that were validated by RT-PCR and that represent several new modes of hnRNP L–dependent splicing regulation, involving both activator and repressor functions: first, intron retention; second, inclusion or skipping of cassette-type exons; third, suppression of multiple cryptic exons, and fourth, alternative poly(A) site selection. In sum, this approach revealed a surprising diversity of splicing-regulatory processes as well as poly(A) site selection, in which hnRNP L is involved.[1]

OUTLOOK

In many cases alternative splicing is essential for understanding human genetic diseases: Many mutations result in splicing defects and map to splicing-relevant sequences, often to the splice-site regions; moreover, mutations affecting exonic splicing enhancers or silencers have been suggested as a novel disease mechanism on the level of pre-mRNA splicing. Examples include the human SMN locus, where mutations cause spinal muscular atrophy (SMA), alternative splicing of the breast cancer susceptibility gene BRCA1, or various mutations in genes for general splicing factors, causing *Retinitis Pigmentosa*.

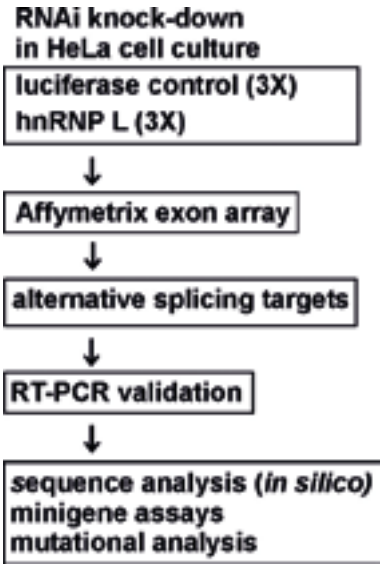


Fig. 2: Combined RNAi knock-down and splice-sensitive microarray analysis: Basic strategy of our genome-wide approach to detect networks between splicing regulators and target genes.

The overall goal of our project is to assess on a genome-wide level and in exon resolution the expression of all human splice variants. This is clearly of great importance for an in-depth understanding of human disease and will soon provide an additional dimension to the standard expression profiling approach. The “all-exon approach” promises to yield novel information on disease-relevant splice variants and thereby to open up new diagnostic avenues and possibly new targets for therapeutic strategies. The medical community is beginning to realize that we not only need to assess global gene expression patterns in profiling studies, but also have to proceed to the level of splice variant resolution. We anticipate that including this will soon become an essential component in the molecular diagnosis of human disease.

References

[1] *Hung L-H et al. (2007). Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis. Submitted for publication.*

[2] *Hui J and Bindereif A (2005). Alternative pre-mRNA splicing in the human system: Unexpected role of repetitive sequences as regulatory elements. Biological Chemistry 386, 1265-1271.*

[3] *Hui J et al. (2005). Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. EMBO J. 24, 1988-1998.*

[4] *Hui J et al. (2003). HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. Nature Struct. Biol. 10, 33-37.*

QUALITY MANAGEMENT IN THE NGFN

STEFAN WIEMANN

GERMAN CANCER RESEARCH CENTER, HEIDELBERG

WHY QUALITY MANAGEMENT?

Quality management (QM) strives to optimize procedures and processes in order to ensure the consistent high quality of products and/or data. Process descriptions need to be detailed enough to be unambiguous and to make results and their interpretation comprehensible for others. Moreover, apart from standardizing production or experimental processes, it is necessary to standardize documentation of the materials used and the results and data. Within the National Genome Research Network, data collection is mostly pursued in high-throughput experiments where large datasets are generated and analyzed. The exchange and comparability of data and information strictly depends on common standards that need to be established and adopted in the different projects. Tested and standardized protocols increase efficiency when different laboratories work in the same biomedical field or when different laboratories utilize and exploit the experimental results.

The need for guidelines and standards in the many fields of biomedical research pursued in NGFN was recognized right at the beginning of the project, and consequently, a quality management group was formed that comprehensively covers the different fields of research within NGFN. This QM group develops guidelines to enhance and maintain high quality standards for experimentation and data in the different NGFN projects.

NGFN'S QM WORK GROUP

Researchers representing the different clinical and research areas of the NGFN and who are involved in QM there have met on a regular basis since 2003 to form the NGFN work group QM & Standards. This group works out and discusses strategies to optimize the quality and reproducibility of research. The consensus is then transferred into the individual NGFN projects and implemented there. Quality management is a continuous process. Hence, standard protocols as well as the emerging standards for genomics experiments are published and updated on the web pages of NGFN's quality management platform (http://www.science.ngfn.de/509.htm), where documents, information and links to the responsible researchers are posted.

The work group is currently focusing on seven areas, which are worked on in dedicated teams: management of clinical and patient data, SNP genotyping, expression profiling microarrays, the generation of Open Reading Frame (ORF) resources, functional assays and cell culture, protein analysis, and animal disease models.

The respective quality standards follow international guidelines, if they exist. These standards are indispensable for improving the quality of results, and they ensure the transparency of experimental procedures. In the following, three examples are provided to give an impression of the successful work of the QM work group.

MANAGEMENT OF CLINICAL DATA

The confidence and trust of patients in the faithful handling of their personal data and tissue samples is prerequisite for the credibility and ultimately for the success of biomedical research. This applies in particular to the protection of patient rights and privacy and to the consistent compliance with the rules and regulations. Patient data must be either anonymous or pseudonymous in relation to the specimen in clinical studies to prevent unauthorized persons from establishing a direct link to individuals. Hence, NGFN database systems are developed in strict compliance with the recommendations of respective ethics commissions e.g., of the German Research Foundation (DFG) and the TMF Common Platform for Medical Research Networks. Work is in progress to establish compliant solutions for the transmission, storage and description of patient data:

1. Patient data, information about biomaterials (e.g. tissue, blood) and experimental data (e.g. genome data, protein data, SNPs) must be linked for analysis without potential re-identification of a single patient. This is realized by decentralized storage of that data.
2. The patient data are pseudonymized in the case of prospective studies. Only the medicating doctor is granted access to all personal data. Hence, this doctor is also the only person who can communicate new research findings to the patients and who is able to collect additional samples of these patients for the study.
3. The NGFN bioinformatics network and clinicians of the NGFN disease-oriented genome network have collaborated to establish specifications for the description of clinical data. When these standardized annotations are applied, complex clinical classifications become feasible. Participants are required to use a common vocabulary to ensure smooth data transfer and data compatibility. This NGFN ontology is based on the Web Ontology Language (OWL), which is also used in other initiatives worldwide and allows clinicians, biologists and bioinformaticians to 'speak' in one language. Through this effort, the QM work group has contributed substantially to the establishment of standards for a responsible and consistent handling of patient data.

Fig. 1: Standardized procedures for high quality data

STANDARD OPERATING PROCEDURES (SOPs)

SOPs are a basic element in quality management and standardization. Workflows and processes need to be described in sufficient detail to allow for an evaluation of the experiments and to permit their reproduction in other laboratories. While such procedures have long been established in industry, the strong need for the development and adoption of standards in research experiments was realized only some time ago, first in the gene expression analysis community. Since then, SOPs have been formulated and implemented successfully in other areas as well. These SOPs are essential for following and controlling experimental design and quality. Here the QM work group has made substantial contributions with a collection of SOPs that have been developed and are utilized in different NGFN projects, and that are made freely available via the Internet. Since research must be innovative and creative, SOPs need to be updated regularly to take novel technologies into account and to adapt to changing needs and specifications. The unambiguous tracking of individual data to the original protocols used in the respective experiments requires establishing SOPs in different versions and also ensuring the long-time availability of superseded versions. With its SOPs and registry of up-to-date and previous versions of SOPs, the QM work group contributes to the international visibility of NGFN.

"REPORTING" STANDARDS

Data are collected, analyzed and interpreted in NGFN projects mostly using high-throughput settings. The development and adaptation of minimum standards for the description of processes and data are prerequisite for the efficient dissemination and utilization of results. Without proper information, for example, about "why" and "how" a particular dataset was collected and analyzed, integrative studies where datasets from different experiments or laboratories shall be related would be unreliable and hence not really feasible. To this end, minimum information guidelines as they were first developed for the systematical analysis of gene expression (Minimum Information About a Microarray Experiment – MIAME) are important contributions. The MIAME standard has led to a true comparability of datasets and has also enhanced the quality of experiments and data. A large number of scientists have contributed with their experience to work out a short and yet precise collection of the most important parameters. Now, the description of research data must cover all the critical points.

Another main task of the QM work group within NGFN is to actively participate in the development of this and similar standardization projects, or even to take charge of such a project. In fact, similar standards are in development, covering different high-throughput science areas. The NGFN and the AGQM have been particularly active in the initiative for minimal information of the description about cellular functional assays (MIACA-Minimum Information About a Cellular Assay). This guideline consists of a description of these minimum requirements and of a consensus object model that integrates this information for efficient data exchange. Several NGFN projects have pooled their competence to develop and foster this new standard. The initiative has meanwhile developed into an international collaboration of scientists from over 30 institutes and companies and currently contributes to the development of a central data repository for data from cellular assays.

OUTLOOK

The QM work group deals with a number of initiatives besides the projects that are described above, and thereby contributes substantially to the success of the NGFN projects in general. For example, an initiative for integrated data management is ongoing in SMP models, and round-robin tests are carried out in the genotyping platform. Further information on the QM work group and the range of its activities is available at the webpage of the NGFN quality management (http://www.science.ngfn.de/509.htm).

Despite the many successes of the QM work group and NGFN, the optimization and implementation of quality standards is an uphill task in many areas. A stronger involvement of companies in the work of QM work group would be advantageous so that expertise from academia and industry could be combined in order to work out innovative concepts in this highly relevant area.

NGFN has demonstrated its ability to advance German genome research to achieve the highest quality of international standards. The QM work group has had a major part in this success story and will continue to improve the research that is carried out to unravel disease mechanisms and to develop novel strategies for the benefit of patients.

WRAP-UP: TECHNOLOGY TRANSFER COORDINATION IN NGFN-2

ISABEL VON KORFF

ASCENION GMBH

KICK-OFF WAS IN EARLY 2005

In the beginning of 2005, Ascenion GmbH and Max Planck Innovation GmbH were awarded a contract by the Federal Ministry of Education and Research (BMBF) to support technology transfer in the National Genome Research Network (NGFN). The goal was to coordinate all players in NGFN-2 as well as all associated technology transfer offices in order to translate NGFN findings into new product development more quickly and efficiently. “Quite a challenge”, Isabel von Korff comments, “as the NGFN-2 involves 54 grant holders from all over Germany, about 500 scientists of various disciplines and around 20 technology transfer offices.” In early 2005, she was appointed manager of the Technology Transfer Coordination Office (German acronym: KTT) within NGFN-2 that was newly established as part of Ascenion’s and Max Planck Innovation’s approach. One of the very first tasks of KTT was a kick-off event in April 2005 to touch base with major NGFN-2 TT-representatives and introduce them into the new concept.

Fig. 1: Dr Isabel von Korff (right), Project Manager of KTT and Claudia Keller (left), Assistant of KTT

WHAT WAS NEW?

A new set of rules was established and agreed upon in individual contracts between KTT and each grant holder within the NGFN-2. Major innovations were:

- All NGFN grant holders had to commit themselves to a technology transfer agency at the start of their projects.
- The agency of choice could request additional support from Ascenion or Max Planck Innovation should particular life-science expertise or experience be required during the commercialization process.
- A new patent fund totaling 350 000 € was available to grant holders to support priority-claiming patent applications.
- KTT would centrally screen all NGFN-related publications for IP-relevant contents.
- KTT would make all commercially relevant findings accessible and searchable on an internet platform.
- KTT would monitor progress of all technology transfer projects in the NGFN and report combined results to the BMBF.

WHAT HAS BEEN ACHIEVED?

In the two and a half years since its inception, KTT has screened over 1850 publications or abstracts and helped coordinate the different players within NGFN-2 in many ways. For instance, it has helped draft, negotiate and close over a dozen consortium agreements between NGFN-2 participants, supported the work of the NGFN Project Committee and provided opportunities for those involved with technology transfer to learn and share their views and experiences.

Total patent applications resulting from NGFN-2 research amount to 48 so far, representing 23 patentfamilies. 22 thereof were partly financed by the patent fund. In terms of marketing, KTT has been very proactive in increasing the visibility of these inventions to industry. The “Genome Marketplace”, a new Internet platform for the commercialization of findings from the NGFN-2 was quickly launched, and 28 technology offers have been posted so far. These cover different fields in the life sciences such as bioinformatics, diagnostics, devices or therapeutics and also include animal models and other commercially viable research tools that can well be marketed without patent protection. In parallel, all technology offers were made accessible through major data bases such as Pharmatransfer, Tech-Ex and Life Science Link. Moreover, KTT has engaged in direct dialogue with the private sector. Isabel von Korff joined dedicated partnering conferences such as BIO or BIO Europe and presented selected technology offers in the context of international technology transfer meetings such as this year’s ATUM meeting in San Francisco. On a national level, an alliance was closed with BIO Deutschland, Germany’s biotechnology industry association. Its members, who are ideally positioned to acquire and progress technologies and product ideas from public research, now receive accelerated access to NGFN-2 findings and can confidentially post tailored “wish lists” of in-licensing needs with KTT.

Thanks to these efforts and with the help of the various technology transfer agencies, many projects have already been advanced to the stage of commercialization. Two examples:

Fig. 2: Marketing NGFN-2 inventions

POTENTIAL NEW TREATMENTS FOR ADVANCED NEUROBLASTOMA

NGFN-2 researchers from the German Cancer Research Center (DKFZ) and Heidelberg University discovered a fungal toxin as potential treatment of neuroblastoma, a cancer of the nervous system which is particularly common in children. The data from *in-vitro* studies which suggest that the compound is superior to established treatments were identified as highly IP relevant by KTT during its regular publication screen. Following an in-depth market analysis by Ascenion, a patent lawyer was assigned and just two months after invention disclosure a first medical use claim was filed for the fungal toxin on behalf of the inventors. Marketing activities were initiated in 2006 and Ascenion rapidly attracted two potential partners, a Swiss and a German biotech company. Both signed confidentiality agreements in the second half of 2006 to further evaluate the data and a potential fit of the project with their own development programs. Negotiations are still ongoing and should be further accelerated by new data from animal models which are expected by the end of this year.

NEW DIAGNOSTICS FOR PARKINSON'S DISEASE

Another exciting project relates to a novel gene called LRRK2 that is associated with a particular heritable form of Parkinson-related syndromes. It was discovered by an international team of inventors including NGFN-2 grant holders from GSF – National Research Center for Environment and Health and Tübingen University Hospital. Following a joint patent application, marketing activities were started by Ascenion on behalf of its partners to exploit LRRK2's strong commercial potential – for therapeutic as well as diagnostic applications. A royalty-bearing license agreement is to be signed shortly with a US- based diagnostics company that has already developed a PCR test for LRRK2. The test will help physicians to decide on an appropriate therapeutic strategy for patients.

OVERALL: ABOVE AVERAGE RESULTS

Although no commercialization contract has been reported to KTT so far, it should be considered a success that a range of NGFN-2 projects have met with considerable interest from industry and are now being negotiated with potential partners – in particular, when taking into account that KTT just started two and a half years ago. In terms of patent applications, KTT can report impressive results: in 2005, for instance, 13 priority filings resulted from the work of about 500 scientists in NGFN-2, i.e. 26 per 1000 researchers, compared to an average of 9 filings per 1000 researchers in European public research institutions as was reported by the Association of European Science & Technology Transfer Professionals (ASTP).

The KTT project will end with the second phase of the NGFN project. As a final event Ascenion will organize a premium showcasing conference in Munich in May 2008. The aim of this conference is to present NGFN-2 inventions with high commercial potential to leading industry partners. A third phase of NGFN will follow, but so far it is unclear how the coordination of the technology transfer will exactly look like. The experience of KTT, however, will be extremely valuable. "We have learned a lot," Isabel von Korff says, "and this will certainly help to create an adapted approach for successful technology transfer in NGFN^{plus} and NGFN^{transfer} once its structure has been defined."

PUBLISHED BY
National Genome Research Network (NGFN)
Project Management NGFN at PT-DLR
Heinrich-Konen-Straße 1
53227 Bonn
Germany
Phone: +49 228 3821-331
Fax: +49 228 3821-332
E-mail: pm-ngfn@dlr.de
Internet: www.ngfn.de

PROJECT MANAGEMENT NGFN
Dr. Markus Albertini, Dr. Uta Straßer

EDITED BY
National Genome Research Network
MasterMedia GmbH, Hamburg

LAYOUT BY
MasterMedia GmbH, Hamburg

PRINTED BY
msp druck und medien gmbh, Mudersbach

2007

National Genome
Research Network

SPONSORED BY THE

Federal Ministry
of Education
and Research

Project Management NGFN at PT-DLR
Heinrich-Konen-Straße 1
53227 Bonn
Germany

Phone: +49 228 3821-331
Fax: +49 228 3821-332
E-mail: pm-ngfn@dlr.de
Internet: www.ngfn.de