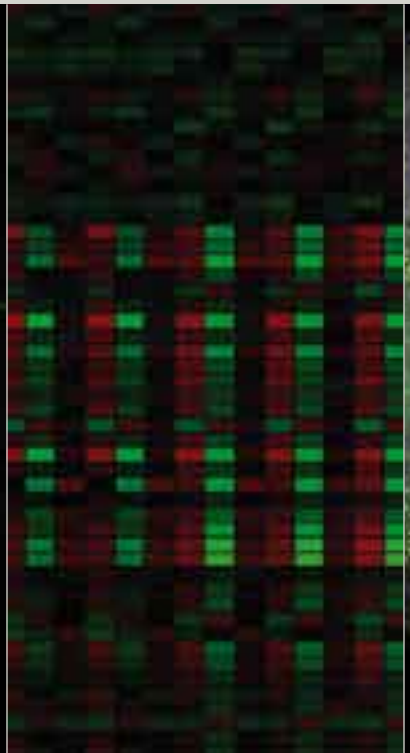
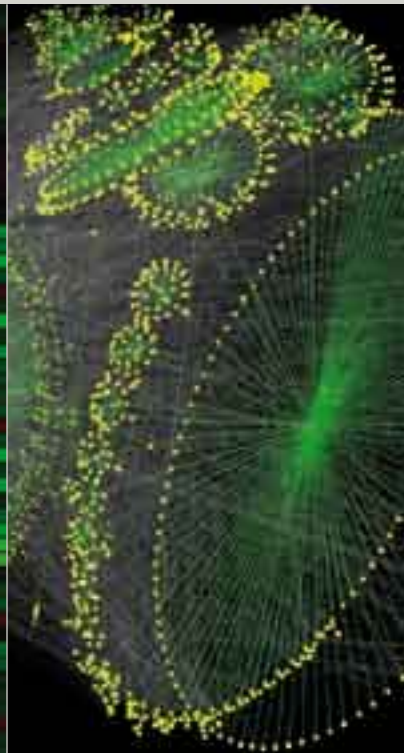
MAX-PLANCK-GESELLSCHAFT

MAX PLANCK INSTITUTE
FOR MOLECULAR GENETICS

Research Report 2006

Max Planck Institute for Molecular Genetics, Berlin



Imprint | Research Report 2006

Published by the Max Planck Institute for Molecular Genetics (MPIMG), Berlin, Germany, August 2006

Editorial Board Bernhard Herrmann, Hans Lehrach, H.-Hilger Ropers, Martin Vingron

Coordination Claudia Falter, Ingrid Stark

Design & Production UNICOM Werbeagentur GmbH, Berlin
Number of copies: 1,500

Photos Katrin Ullrich, MPIMG; David Ausserhofer

Contact Max Planck Institute for Molecular Genetics
Ihnestr. 63–73
14195 Berlin, Germany
Phone: +49 (0)30-8413 - 0
Fax: +49 (0)30-8413 - 1207
Email: info@molgen.mpg.de

For further information about the MPIMG please see our website: **www.molgen.mpg.de**

Table of Contents

Organisational Structure

Dept. H.-Hilger Ropers
Human Molecular Genetics

- **Tim Hucho** — Molecular pathways in mental retardation and pain
- **Vera Kalscheuer** — Chromosome rearrangement & disease
- **Andreas Kuß** — Familial mental retardation
- **Constance Scharff** — Neurobiology of bird song and speech development
- **Susann Schweiger** — Biochemistry of inherited brain disorders
- **Andreas Tzschach** — Clinical genetics
- **Reinhard Ullmann** — Molecular cytogenetics
- **Diego J. Walther** — Serotonin metabolism and disease; Mouse lab

Dept. Hans Lehrach
Vertebrate Genomics

- **James Adjaye** — Molecular Embryology and Aging
- **Konrad Büssow** — Protein Structure Factory
- **Johan Gobom** — Mass spectrometry
- **Ralf Herwig** — Bioinformatics
- **Heinz Himmelbauer** — Comparative and Functional Genomics
- **Margret Hoehe** — Genetic variation, haplotypes & genetics of complex diseases
- **Michal Janitz** — Oligofingerprinting / cell arrays
- **Edda Klipp** — Kinetic modeling
- **Zoltán Konthur** — *In vitro* ligand screening
- **Sylvia Krobitsch** — Neurodegenerative disorders
- **Bodo Lange, Thorsten Mielke** — Protein complexes & cell organelle assembly/ USN
- **Hans Lehrach** — Automation & Technology Development
- **Georgia Panopoulou / Albert Poustka** — Evolution & development
- **Silke Sperling** — Cardiovascular genetics
- **Marie-Laure Yaspo** — Human chromosome 21

Dept. Martin Vingron
Computational Molecular Biology
Managing Director

- **Peter Arndt** — Evolutionary Genomics
- **Stefan Haas** — Gene structure & array design
- **Roland Krause** — Microbial Virulence (joint group with the MPI for Infection Biology)
- **Alexander Schliep** — Algorithms
- **Rainer Spang** — Computational diagnostics
- **Martin Vingron** — Transcriptional regulation

Dept. Bernhard Herrmann
Developmental Genetics

- **Hermann Bauer** — Transmission ratio distortion
- **Markus Morkel** — Signal Transduction in Embryogenesis & Tumor Progression
- **Heiner Schrewe** — Development of endodermal organs
- **Ralf Spörle** — Expression analysis & 3D imaging
- **Lars Wittler** — Somitogenesis

Institute of Genetics,
Charité, University Medicine Berlin
Bernhard Herrmann

Research Group Stefan Mundlos
Development & Disease

Institute of Medical Genetics,
Charité, University Medicine Berlin
Stefan Mundlos

Administration / Research Support
Manuela B. Urban / Martin Vingron

- **Manuela Urban** — Administration
- **Ulf Bornemann** — Technical Management & Workshops
- Library
- **Rudi Lurz** — Imaging
- **Richard Reinhardt** — Computing
- **Richard Reinhardt** — Analytics
- **Ludger Hartmann** — Animal Facility / Transgene Unit
- **Bernhard Herrmann** — Stem-cell Facility
- **Stefan Mundlos** — Histology

Ribosome Group

- **Paola Fucini** — Ribosome Crystallography
- **Knud H. Nierhaus** — Ribosomal function

- **Rudi Lurz** — Microscopy
- **Richard Reinhardt** — High-throughput technologies

Otto Warburg Laboratory
(Independent Junior Research Groups)

- **Michael Lappe** — Bioinformatics

Official women's representative
Melanie Wendehack

Workscouncil
Chairman: Klaus von Knoblauch

Representative of the scientific staff at the Biology and Medicine Section
Stefan Haas

The Max Planck Institute for Molecular Genetics

Mission

Research at the Max Planck Institute for Molecular Genetics (MPIMG) concentrates on genome analysis of humans and other organisms to elucidate cellular processes and genetic diseases. It is the overall goal of the combined efforts of all MPIMG groups to gain new insights into the development of diseases on a molecular level, thus contributing to the development of cause-related new medical treatments.

Development of the Institute

The Max Planck Institute for Molecular Genetics (MPIMG) was founded in 1964 with the appointment of Heinz-Günther Wittmann and Heinz Schuster as heads of department, followed by the appointment of Thomas Trautner in 1965. At this time, the research of the institute was focussing on DNA replication and gene regulation in bacteria, bacterial phage and fungi (departments Schuster and Trautner) and on the structure, function and evolution of ribosomes which were central to the work of H.-G. Wittmann. In 1970, the three departments, as well as four independent junior research groups (the future Otto Warburg Laboratories) moved into the new premises of the institute situated in the Ihnestraße, Berlin-Dahlem. After the sudden death of H.-G. Wittmann in 1990 and the retirement of H. Schuster in 1995, the appointments of Hans Lehrach (1994, Dept. of Vertebrate Genomics), and Hans-Hilger Ropers (Dept. of Human Molecular Genetics, full-time since 1997) induced a major shift in the scientific orientation of the institute. Following the retirement of T. Trautner in 2000, Martin Vingron was appointed as head of the new Department for Computational Molecular Biology. At the same time, Stefan Mundlos was jointly appointed by the Humboldt University of Berlin and the Max Planck Society as head of the Institute for Medical Genetics and of an independent research group at the MPIMG. Together with the Free University of Berlin, Bernhard Herrmann has been appointed professor at the university and director at the institute in 2003, forming the fourth departement. In 2004 a new independent junior research group in Bioinformatics took up its work.

Research Concept

Genome research, the systematic study of genes and genomes, has changed the way in which research in molecular genetics is pursued. The focus and composition of the MPI for Molecular Genetics reflects this development. Large scale genome research (Dept. Lehrach) applying a variety of technologies supplies us with the technology and the data on genome sequences, genes, and their function. Human molecular genetics (Dept. Ropers) searches for disease genes and their biological function. Computational molecular biology (Dept. Vingron) exploits the generated data to better understand biological and disease processes. The Developmental Genetics Department (Dept. Herrmann) uses systematic functional analysis for understanding developmental mechanisms.

The institute pursues a number of large scale projects. Probably the most prominent national project is the German National Genome Research Network (NGFN), where all departments of the institute participate and collaborate with each other. Other prominent projects include a number of EU projects, participation in several projects of the German Ministry of Science as well as DFG "Sonderforschungsbereiche". Several international initiatives collecting patient data are pursued by the Human Molecular Genetics Department. During recent years, systems biology has become another major activity in the institute.

With this involvement in national and international research projects as well as by virtue of the research output of the institute, MPIMG is perceived internationally as a stronghold of genome and genetics research in Germany. The publications coming from the institute document the international competitiveness of the institute. Maintaining this status in the future will require continuing technological innovation, close cooperation with the universities and, in particular, their medical schools. Integration between genome research and genetics, as well as between experimental and computational biological research are key in this effort. These are the means by which research excellence shall be maintained and further strengthened in the future.

Prof. Dr. Bernhard Herrmann
Prof. Dr. Hans Lehrach
Prof. Dr. H.-Hilger Ropers
Prof. Dr. Martin Vingron

Department of Developmental Genetics

Head
Prof. Dr. Bernhard G. Herrmann
Phone: +49 (0)30-8413 1409
Fax: +49 (0)30-8413 1229
Email: herrmann@molgen.mpg.de

Secretary
Gabriele Eder
Phone: +49 (0)30-8413 1344
Fax: +49 (0)30-8413 1130
Email: eder@molgen.mpg.de

Scientists*
Dr. Phillip Grote (since 9/2005; funded by Charité)
Lorenz Neidhardt (since 3/2005; funded by NGFN)
Dr. Martin Werber (part time; since 1/2005; funded by NGFN)

Technicians*
Conny Kreschel (1/2005 - 3/2006; funded by Charité)
Karol Macura (since 1/2005; funded by NGFN)
Uta Marchfelder (since 10/2005)

Graduate Students*
Solveig Müller (since 5/2005)
Arnold Schröder (since 10/2003)

Student*
Helena Jambor (12/2003 - 8/2004)

* not named in group reports

Department of Developmental Genetics

Introduction

The formation of complex organisms from a single cell requires tight control of multiple reiterative steps of cell proliferation, patterning and differentiation, comprising frequent cell interactions and changes of the cellular readout. Trunk development involves a large repertoire of cellular responses controlled by competing signaling pathways, which employ different sets of transcriptional regulators forcing cells into various differentiation pathways. Developmental geneticists in general use mutagenesis tools to remove or alter the function of single genes to be able to analyse their roles in embryogenesis (provided a phenotype is observed). Often the knowledge gained from such analyses ends at the description of the phenotype. However, advances in genomics research have set the stage for gaining deeper insight into the genetic control of developmental processes, even of such complexity as trunk formation. To take that quantum leap, however, developmental biology research has to advance from the single gene analysis to the genomics level. This is certainly a real challenge, but promises to provide a deeper understanding of the regulatory networks controlling complex cellular interactions and responses leading to tissue and organ formation. Such knowledge will not only be of academic value, but generate additional benefit for medical research.

Scientific Overview

My department is engaged in two main topics, the molecular analysis of trunk development, with a focus on presomitic mesoderm formation, and transmission ratio distortion. The former involves a number of groups and scientists, investigating different aspects of this process with the common aim of deciphering the regulatory networks controlling it. The latter is carried out by a single group (for details see the research report by the "Transmission Ratio Distortion" group). Here I will focus on explaining our strategies of investigating trunk formation in the mouse at the whole-genome level.
In principle, our strategy is based on the simple assumption that differentiation processes require differential gene activity. If cells follow different fates there must be genes forcing them to do so, and their expression should correspond to the different cellular fates. If we knew those regulators, it should be possible to determine their function in this process and determine their targets. Regulator-target interactions can then be merged to regulatory networks controlling modules of cell behaviour and fate changes, which, in concert, give rise to differentiated tissues.

The first steps of trunk development involve cell proliferation, the induction of cells to different fates, changes in cell shape and cell behaviour (epithelial-mesenchymal transition, cell motility), maintenance of the induced cell fate, differentiation and patterning. In parallel the cells are endowed with positional information and organized by several organizing centers ensuring correct 3-D topology of the developing tissues.

We have started the following approaches to decipher regulatory networks on the transcriptional level:

Gene expression analysis in whole mount mouse embryos

Genome-wide gene expression analysis by whole-mount *in situ* hybridization allows assignment of genes to the cells and processes in which they act. Co-expression and syn-expression groups identifying genes acting in parallel or even in the same pathway are derived from such data. To date over 6,500 genes have been analysed (from over 12,500 cDNA clones; for details see the research report by the "Gene Expression and 3-D Reconstruction" group). Gene expression and annotation data are stored in the MAMEP database, providing a rich gene resource for systematic functional analysis of embryonic processes. From these data we have derived a co-expression group of over 600 genes showing tissue-restricted expression in the caudal region (caudal gene list). This list forms the genetic resource for regulatory network construction, involving approaches described below.

Promoter analysis *in silico*

The caudal gene list provides a set of functionally related genes, which are most likely commonly regulated by (interacting and competing) pathways. Bioinformatic tools are applied to identify the transcriptional regulators and their putative targets among the caudal gene list, providing (unconfirmed) *in silico* regulatory networks. Phylogenetic footprinting and other tools are utilized to further improve the predictions of *in silico* transcriptional regulatory netwoks (collaboration with Dr. Manke, Dept. Vingron).

Promoter analysis *in vitro*

Putative promoters identified by gene annotation available in public databases and by various bioinformatic analyses are tested on large scale *in vitro* using transactivation assays in cell arrays (collab. with Dr. Janitz, Dept. Lehrach). Proof of principle has been obtained for the *Dll1* promoter, which is controlled by Lef1/Tcf and Tbx transcription factors. We plan to test a set of about 160 promoters derived from genes primarily expressed in the presomitic mesoderm, for transactivation by several control factors known to be essential for presomitic mesoderm formation and differentiation. These assays will provide functional data on transcription factor/target interactions *in vitro* for a selected set of genes and control factors, which may be extended as network construction proceeds.
Transcription factor binding sites in the promoters can be analysed on large scale with dsDNA chips. This test provides data similar to the common EMSA assays with the important difference that many sites can be analysed in parallel and quantitation is possible, allowing to distinguish high and low affinity binding sites. Proof of principle has been obtained for this assay system; the experiments will be carried in the department Lehrach.

Functional promoter analyses *in vivo*

The specific activity of a putative promoter, in terms of tissue restriction, can only be determined directly in the embryo. Such assays are cost intensive and thus limited to a small selection of genes. We have set up assay systems involving standard cloning of (wild type and mutant) reporter constructs and transient expression analysis in chick and mouse embryos, using electroporation and pronuclear injection techniques, respectively. In addition, to overcome position effects of transgene integration and to increase efficiency we are setting up an ES cell based assay system, involving integration of reporter constructs into a defined locus via recombinase mediated cassette exchange and reporter assays in tetraploid embryo/ ES cell chimera.
In combination with *in vitro* binding site analyses, mutagenesis of predicted functional sites, *in vitro* transactivation assays and *in vivo* reporter assays of mutated promoter fragments the cis regulatory elements responsible for tissue specific gene expression in the embryo can be identified unambiguously. Promoters identified and verified in this manner are extremely useful tools for functional analyses *in vivo*, and for tissue engineering.

Functional analysis of control factors *in vivo* by gene knock-down

We are planning to use RNA interference mediated gene knock-down as "high-throughput" tool for functional analyses of genes in the embryo. The versatile miR-vector system in combination with tetraploid embryo chimera technology is utilized for knock-down of genes involved in EMT and early presomitic mesoderm formation. Proof of principle has been obtained so far on the *Brachyury* gene, which is essential in these processes. A temporally and spatially controlled (conditional) knock-down system is being established. Mutant embryos are characterized phenotypically, by expression profiling on Illumina microarrays, and by whole mount *in*

situ hybridization using various cell fate markers as probes. We focus the analyses on three major pathways controlling EMT and mesoderm formation in the embryo, Wnt, FGF and BMP. Microarray profiling data derived from comparisons of mutant and wild type tissue will identify sets of genes acting downstream of the regulators. New factors identified in this manner are verified by whole mount *in situ* hybridization. *In vitro* and *in silico* promoter data are matched with the *in vivo* data.
In parallel we are investigating genes identified as candidate control factors of EMT in the embryo, for their possible role in EMT in cancer progression using human colon carcinoma lines. In this manner, our knowledge of the control of EMT in the embryo is applied to cancer research directly. For more details see the research report of the "Signal Transduction in Embryogenesis and Tumor Progression" group.

Identification of direct target genes of transcriptional regulators

Chromatin immunoprecipitation (ChIP) assays are utilized to identify functional binding sites, occupied by transcriptional regulators in the genome. We focus on a selection of transcriptional regulators and chromatin control factors involved in EMT and presomitic mesoderm formation. We have designed custom-made Nimblegen chips covering about 25 kb of genomic DNA of appr. 600 genes expressed in the tailbud (caudal gene list). We plan to examine functional binding sites on a genome-wide scale once affordable whole genome tiling arrays become available. ChIP-chip data show physical interaction of regulators with targets, which, in combination with chromatin status assays, allow to distinguish active from repressed genes. The availability of chip-grade antibodies is still limiting. Therefore, we plan to tag genes of high interest by BAC recombineering, and to transfer engineered BACs to ES cells for carrying out ChIP-chip analyses in differentiating stem cells and in embryos.

I am confident that the combination of approaches outlined above will provide sufficiently reliable data allowing computation of robust regulatory networks, comprising hundreds of genes involved in regulatory modules controlling various mechanisms of cell diversification. This knowledge will substantially advance our understanding of developmental processes and aid tissue engineering and cancer research.

The approaches outlined above are complemented by two projects investigating embryonic development on different levels.

Molecular mechanisms of segmentation

We utilize a combination of genetic and microsurgical tools to investigate the molecular mechanisms controlling the segmentation process giving rise to somites, the precursors of vertebrae and skeletal muscle. Segmentation is the final step in a cascade of events leading from presomitic mesoderm formation to somites. It requires the interaction of a morphogen gradient with an oscillator. The Wnt, FGF and Notch signaling pathways are involved, with Wnt3a taking the role of a master control factor. We are investigating the modes of interaction between these pathways in controlling the segmentation process. A computer simulation of the segmentation clock has been worked out in collaboration with Christoph Wierling (Dept. Lehrach). For more details on this project see the research report by the "Segmentation" group.

In vitro differentiation in embryonic stem cells and tissue engineering

This project is placed at the earliest events of mesoderm formation, epithelial-mesenchymal transition and early patterning. In the embryo these important events take place in a small number of cells. This circumstance makes it extremely difficult to investigate the early transition steps from a resident proliferating epithelial stem cell to a mesenchymal, non-proliferating, motile cell. To overcome this limitation we are trying to copy this process in cell culture, using embryonic stem cells. We work on establishing culture conditions and treatments allowing to drive ES cells en bloc into presomitic mesoderm formation. Once this is achieved, we are able to generate sufficient material for investigating EMT and presomitic mesoderm differentiation *in vitro*. This technology will form the basis for engineering differentiated cell types, such as skeletal muscle and cartilage, *in vitro*. Application to human embryonic stem cells is planned.

Selected Information

Publications

- J. K. Dale, P. Malapert, J. Chal, G.Vilhais-Neto, M. Maroto, T. Johnson, S. Jayasinghe, P. Trainor, B. Herrmann, and O. Pourquie´. (2006). Oscillations of the Snail Genes in the Presomitic Mesoderm Coordinate Segmental Patterning and Morphogenesis in Vertebrate Somitogenesis. Developmental Cell 10, 355–366
- H. Bauer, J. Willert, B. Koschorz & B.G. Herrmann (2005). The t complex-encoded GTPase-activating protein Tagap1 acts as a transmission ratio distorter in mice. Nature Genetics 37, 969-973
- M. Hofmann, K. Schuster-Gossler, M. Watabe-Rudolph, A. Aulehla, B. G. Herrmann and A. Gossler (2004). WNT signaling, in synergy with T/TBX6, controls Notch signaling by regulating Dll1 expression in the presomitic mesoderm of mouse embryos. Genes Dev 18, 2712-2717
- A. Aulehla and B. G. Herrmann (2004). Segmentation in vertebrates: clock and gradient finally joined. Genes Dev 18, 2060-2067
- H. B. Abdelkhalek, A. Beckers, K. Schuster-Gossler, M. N. Pavlova, H. Burkhardt, H. Lickert, J. Rossant, R. Reinhardt, L. C.Schalkwyk, I. Müller, B. G. Herrmann, M. Ceolin, R. Rivera-Pomar, and A. Gossler. (2004) The mouse homeobox gene Not is required for caudal notochord development and affected by the truncate mutation. Genes Dev 18, 1725-1736
- A. Aulehla, C. Wehrle, B. Brand-Saberi, R. Kemler, A. Gossler, B. Kanzler, and B. G. Herrmann. (2003) Wnt3a plays a major role in the segmentation clock controlling somitogenesis. Dev. Cell 4, 395-406

Non-reviewed Publications:

- N. Veron, H. Bauer und Bernhard G. Herrmann. (2005) Spermien-Ein Wettlauf der Gene. GenomX-Press 4.05

Patent application:

- B.G. Herrmann, H. Bauer. (2005) Isolation of the t-complex distorters an applications thereof. Registration No: EP 05 01 7651.0

Transmission ratio distortion

Head
Dr. Hermann Bauer (since 10/2003)
Phone: +49 (0)30-8413 1329
Fax: +49 (0)30-8413 1229
Email: bauer_h@molgen.mpg.de

Scientists
Dr. Sigrid Schaper (since 9/2004; funded by Charité)
Dr. Yves Charron (since 5/2006)

Graduate Student
Nathalie Véron (since 9/2004)

Student
Sabrina Schindler (3/2006 - 7/2006)

Technician
Jürgen Willert

Scientific Overview

A considerable proportion of wild mice carry two variant forms of chromosome 17, the wild type and the *t*-form. Males heterozygous for the *t*-haplotype transmit this chromosome at a high ratio to the offspring, at the expense of the wild type chromosome. Mouse geneticists have identified several closely linked mutant loci involved in this phenomenon. The central factor is the *t-complex responder* (Tcr or R). Transmission of the chromosome carrying Tcr, is promoted by loci encoding *t-complex distorters* (Tcd (or D)1 to 4). Tcr acts in cis while the distorters act in trans. Absence of Tcd factors is disadvantageous for Tcr; its transmission ratio drops to about 20%, whereas expression of all *Tcds* may enhance the transmission of *Tcr* to 99% (Fig 1a).

Since its discovery in 1936 this phenomenon, termed transmission ratio distortion (TRD) has remained an enigma for almost 7 decades. A model explains the genetic observations as follows: Tcds are distributed to all sperm during meiosis and exert a harmful effect on both, t and wild type meiotic partners. Tcr is able to rescue the harmful effect of the distorters but remains restricted to the cells carrying the gene, the future *t*-sperm. Wild type sperm thus remain compromised while t-sperm is rescued (Fig 1b). This leads to preferential fertilization of the egg cells by t-sperm and prevalence of t-animals in the next generation (Fig 1c).

Access to understanding the molecular basis of TRD was first obtained by cloning of the central factor Tcr and the first distorter gene (Tcd1a) by our group (Herrmann *et al.*, Nature 402, 141-146 (1999), Bauer *et al.*, Nature Genetics 37, 969-973 (2005)). Tcr encodes a dominant negative variant of a novel Ser/Thr protein kinase, termed *Smok*, which is expressed in spermatids. The identified distorter gene, Tagap1 is a GTPase activating protein (GAP) for small G-proteins of the Rho subfamily.

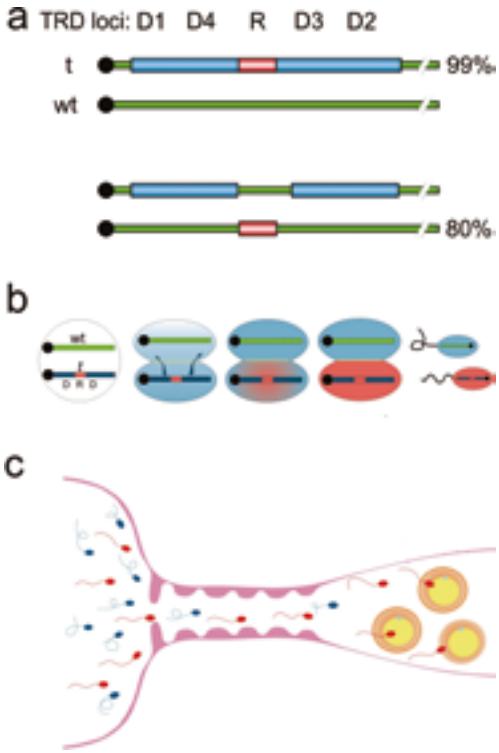


Figure 1
Genetics of the mouse t-complex and a model of cellular processes during transmission ration distortion (TRD).

a) Genetics of the t-complex. The transmission rate of the chromosome carrying the t-complex-responder (R) is enhanced by the distorters (D). The distorters can be located on the same chromosome as the Responder (natural condition, upper part) or on a different chromosome (experimental set up, lower part).
b) Proposed cellular mechanism of TRD. Harmful products (blue) of the distorter genes are distributed between t- and wild type meiotic partners across cytoplasmic bridges. Therefore both, t and wild-type meiotic partners are compromised. In contrast, the product of the responder gene (red), which counteracts the harmful distorter effect, is restricted to t-sperm and not shared. Consequently only t-sperm are rescued. They show a more progressive swimming behaviour as compared to their wild type competitors and have an advantage in fertilizing the egg cells (c).

Rho small G-proteins cycle between an active, GTP bound state and an inactive conformation in which they have bound GDP. GAPs enhance GTP hydrolysis to GDP and thus inactivate G-proteins while GEFs (guanosine nucleotide exchange factors) activate Rho proteins by promoting the exchange of GDP for GTP. Recently, we have identified a GEF molecule as another transmission ratio distorter. This gene is located in the genetic interval of distorter 2 (manuscript submitted).

Since spermatozoa derived from *t*/+ males show impaired flagellar function, we suggest that Tcr, Tagap1 (Tcd1a), the newly identified GEF and other distorters act in a signalling pathway controlling sperm motility. We propose a model, in which the distorters hyper-activate Smok, compromising sperm function. Tcr acts as an "antidote", bringing the signalling pathway back to a level favourable for normal flagellar function (Fig 2).

The t-forms of Tagap1 and the newly identified GEF-encoding distorter represent hypermorphic alleles, and both promote the transmission ratio of the t-haplotype. Since the two factors have a contrary effect on their respective G-protein, but act additively, we conclude that they act on two different G-proteins, which exert opposite effects on Smok (Fig 2). We are further analyzing this pathway using genetic and molecular methods.

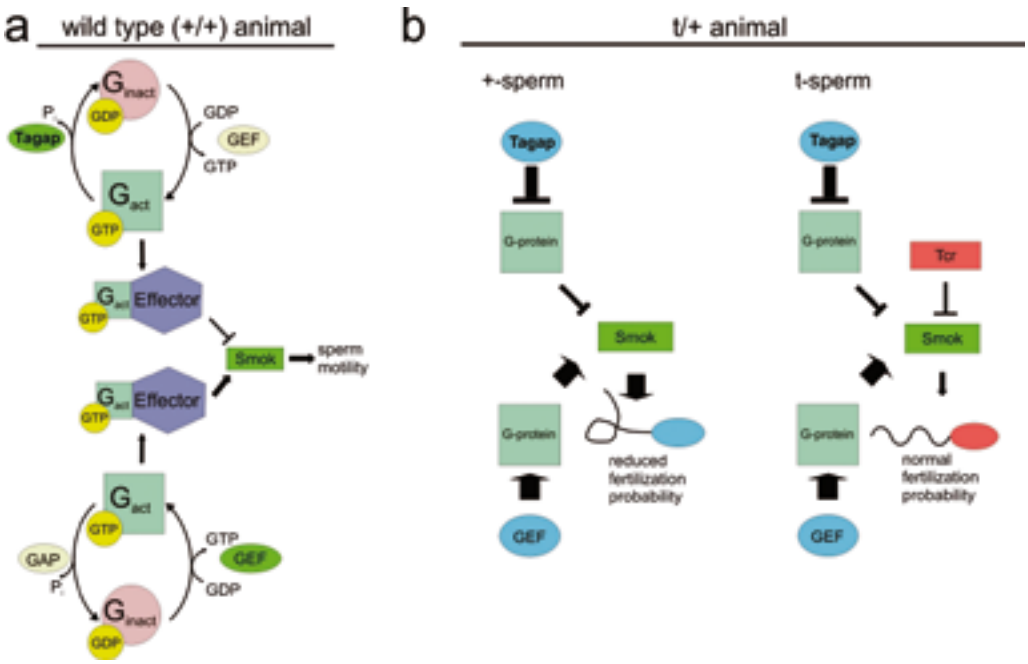


Figure 2:

Molecular mechanism of TRD. a) Distorter gene products (blue) activate the Smok-kinase signaling pathway. b) In t/+ animals, the t-complex alleles of the distorters additively hyperactivate this pathway which leads to compromised sperm function (left). Tcr, a dominant negative version of the Smok kinase brings signaling back to a normal level and thus rescues t-sperm (right).

Future Perspectives

Genetics of TRD (Yves Charron)

We employ positional candidate gene cloning approaches to identify additional distorter genes, which are tested in mouse models for transmission distortion effects. While also applying conventional transgenic- and knock-out-technology we are establishing an RNAi-based strategy to functionally characterize candidate factors more rapidly.

Biochemical analysis of the Smok signalling pathway (Sigrid Schaper)

In addition to this genetic approach we characterize molecules acting in TRD biochemically. Using protein-protein interaction assays and screening systems, we aim to identify new molecules acting in this process.

Elucidating the molecular mechanism of Tcr function (Nathalie Véron, Sabrina Schindler)

A particularly interesting question is how and where Tcr functions in the cell. Using transgenic approaches we want to find out how the products of *Tcr* are restricted to the spermatids carrying the gene, which is unexpected since spermatogenic cells are connected in a syncytium and sharing of gene products between meiotic partners is common.

Sex ratio distortion in mammals (Jürgen Willert, Hermann Bauer)

We have shown, that *Tcr* can distort the transmission rate of genetic traits linked to it, even outside its natural genomic environment. Therefore, Tcr and Tcds can be utilized for manipulating the transmission rate of genetic traits, such as sex. We are presently establishing knock-in mouse models with the goal to achieve efficient sex ratio distortion. In collaboration with the Institute for Molecular Animal Breeding and Biotechnology at the LMU Munich, headed by Prof. Eckhard Wolf, we are also investigating the functionality of Tcr and Tcds in a farm animal, the pig.

Selected Information

Publications

- Bauer, H., Willert, J., Koschorz, B. & Herrmann, B.G (2005). The t complex-encoded GTPase-activating protein Tagap1 acts as a transmission ratio distorter in mice. Nat Genet 37, 969-73.

Patent

- Herrmann B.G., Bauer, H. Isolation of the t-complex distorters and applications thereof. No. EP0501 7651.0 August 2005.

Guest Scientists

- Stéphan Gasca 04-05/2004.

Teaching

- Practical course: “Biology for medical students”. Charité medical University, Berlin; Each term.
- Practical course: “Molecular Medicine”. Charité medical University.

Internships

- Sylvia Möckel
- Jonas Saloum

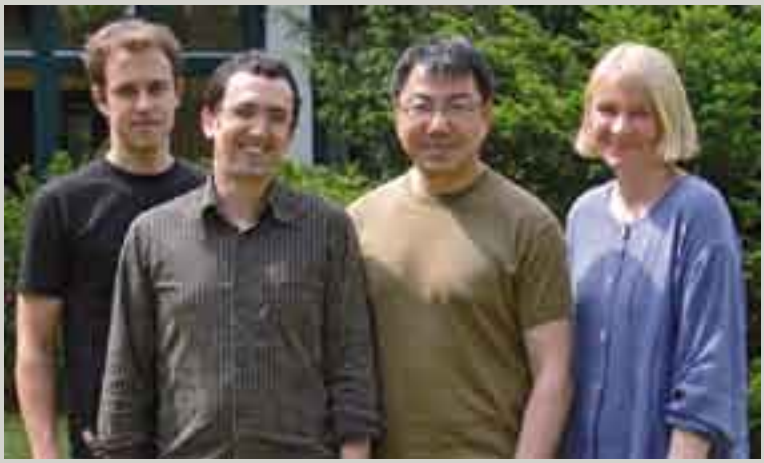
Public relations

- Science Fair 2005.
- Lange Nacht der Wissenschaften 2003, 2004, 2005.

Cooperations

- Rudi Lurz (Research Group Microscopy): Localization of the Tcr protein and transcript in testis and sperm
- Zoltán Konthur (Department Lehrach): Selection of phage displayed antibodies for specific detection of Smok/ Tcr.
- Sylvia Krobitsch (Department Lehrach): Yeast-two-hybrid screening.
- Prof. Dr. Reinhard Fässler (MPI for Biochemistry): Development of Tcr- and Smok-specific peptide antibodies.
- Prof. Dr. Eckard Wolf (Institute of Molecular Animal Breeding and Biotechnology, LMU München): Transmission ratio distortion in the pig.

Signal Transduction in Embryogenesis and Tumor Progression



Head
Dr. Markus Morkel
(since 6/2005, funded by NGFN)
Phone: +49 (0)30-8413 1288
Fax: +49 (0)30-8413 1229
Email: morkel@molgen.mpg.de

Scientist
Forrest Liu, PhD, MD (since 1/2005; funded by NGFN)

Technician
Antje Brouwer-Lehmitz (since 8/2005; funded by NGFN)

Graduate Student
Marc Leushacke (since 11/2005)

Scientific Overview

During embryonic development, epithelial cells can aquire properties of mesenchymal cells, e.g. the cells lose apical-basal polarity, re-organize their cytoskeleton, down-regulate cell adhesion, and become motile. This phenotypic switch is termed epithelial-to-mesenchymal transition (EMT). In the mouse embryo, EMT is required for a multitude of important processes during early steps of morphogenesis and organ formation. In the tailbud, for instance, cells under the influence of signals such as Wnt (wingless integration), Fgf (fibroblast growth factor) and BMP (Bone morphogenetic protein) detach from the epithelium, become motile and subsequently differentiate into different types of mesoderm.
In the adult organism, EMT is associated with a number of pathological states. Most significantly, EMT is an important process during tumor progression, and provides cancer cells with invasive and metastatic properties. Apparently, EMT during tumor progression employs very similar signalling pathways (such as Wnt and BMP) and key molecules (such as Twist and Snail) compared to such processes in the embryo. Our group is interested in defining signalling networks that are essential for EMT in the mouse embryo as well as in tumor cells.

Projects

EMT signalling networks in the embryo

During the elongation of the body axis in the mouse embryo, mesoderm forms from epithelial precursors in the tail bud. A number of key players for epithelial-to-mesenchymal transition that accompanies body axis elongation have been identified by gene ablation: embryos lacking Wnt3a, Fgf receptor I or brachyury do not elongate their body axis and do not properly form mesoderm. Moreover, in a high-throughput gene expression screen, we have identified approx. 600 genes specifically involved in this process (in co-operation with Ralf Spörle, Gene Expression and 3D Reconstruction Group, MPIMG). However, it remains to be tested, by which signalling pathways and transcription factors these genes are regulated. We therefore employ inducible RNAi technology *in vivo* to silence and/or hyperactivate the Wnt, Fgf, and BMP signalling pathways, plus selected downstream transcription factors, during EMT in the tailbud of the mouse embryo. To this end, we have constructed and integrated a novel transgenic expression system into the mouse genome that offers tight spacial and temporal control of microRNA-based RNA interference. The system combines appropriate tissue-specific RNA polymerase II promoters with a tetracycline-inducible system. From transgenic embryonic stem cells, mutant embryos are generated using the tetraploid aggregation technique. We have now generated the first knock-down embryonic stem cell line for Brachyury, and embryos derived from these cells phenocopy Brachyury-/- embryos (Fig. 1). We will establish knockdown embryos for a range of key factors, and, using microarrays and in-situ hybridisation, establish a hierarchical network of signalling pathways during EMT.

Figure 1: Phenotypes of d9.5 mouse embryos that are derived from ES cells expressing a control microRNA (a) and microRNA directed against Brachyury (b), using the tetraploid aggregation technique. Arrow indicates axis truncation, arrowhead indicates allantois malformation, two hallmarks of the Brachyury phenotype.

A loss of function phenotypic screen for EMT and metastasis genes in tumor cells

Epithelial-to-mesenchymal transitions during embryonic development and in tumor cells likely employ similar signalling pathways. However, while EMT is activated in a timely and controlled manner in the embryo, it is triggered in a variable subset of tumor cells during tumor progression. To overcome experimental limitations associated with the study of EMT in tumors,we now apply our knowledge on EMT in embryos to define key factors of EMT, motility, and metastasis in tumor cells. High-throughput in-situ screening has already revealed a set of candidate genes for EMT in the mouse embryo, i.e. genes that are specifically activated at sites of epithelial-to-mesenchymal transition (Ralf Spörle, Gene Expression and 3D Reconstruction Group, MPIMG). We now utilize esiRNA technology (in co-operation with F. Buchholz, MPI-CBG, Dresden) to assess the impact of these candidate genes on EMT, cell adhesion, motility and metastasis in tumor cells. Presently, we employ human colon cancer cell lines which are of epithelial origin, but display a mesenchymal phenotype, to knock down candidate genes and score mesenchmal-to-epithelial transition (Fig. 2). This approach allows for the first time to screen a large number of EMT candidate genes in a simple functional assay and thus to define important regulators of EMT and metastasis and potential drug targets that may modulate progression of human tumors. A high number of positive candidates highlights the similaritiy between EMT in embryogenesis and tumor metastasis.

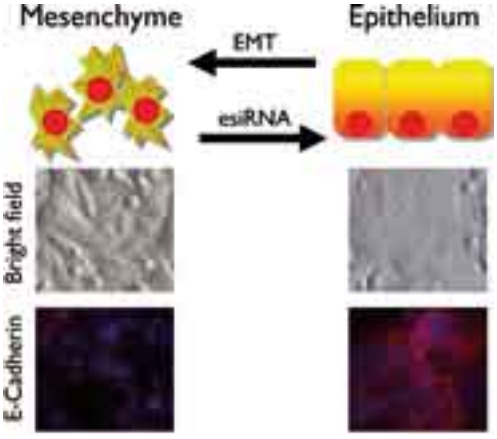


Figure 2: Loss of function phenotypic screen for EMT effector genes. Upper panel: schematic representation of epithelial and mesenchymal cell phenotypes. Middle panels: mesenchymal phenotype of control transfected SW480 cells (to the left), and epithelial phenotype of SW480 cells transfected with esiRNA directed against an EMT effector (to the right). Lower panels: immunolocalisation of E-cadherin in SW480 cells, transfected as in middle panels.

Perspectives

We are undertaking a systematic functional analysis of genes involved in epithelial-to-mesenchymal transition and mesoderm formation during embryonic development. Moreover, since epithelial-to-mesenchymal transition and tumor metastasis are similar on the molecular level, we use our data to study tumor metastasis. Today, little is known about molecular events that are crucial in metastasis formation, which is the fatal step in the disease. We currently use high-throughput methods to screen EMT candidate genes for phenotypic effects in human colon tumor cells. In the future, we will study expression of EMT effector genes in mouse and human tumors, and define their function in the metastatic process. This work will lead to the identification of novel prognostic markers as well as drug target candidates for tumor metastasis.

Selected Information

Publications

- Pohlers M., Truss M., Frede U., Stehle M., Kuban R.J., Hoffmann B., Morkel M., Birchmeier C., Hagemeier C. (2005). A role for E2F6 in the restriction of male germ cell specific gene expression. Current Biology 2005 Jun 7;15(11):1051-7
- Morkel M., Huelsken J., Wakamiya M., Ding J., van de Wetering M., Clevers H., Taketo M.M., Behringer R.R., Shen M.M., Birchmeier W. (2003). Beta-catenin regulates Cripto- and Wnt3-dependent gene expression programs in mouse axis and mesoderm formation. Development 2003 Dec;130(25):6283-9

Teaching

- Lecture and Lab courses *Molecular and Cellular Biology* (2001-2005), for students of Biology and Biochemistry of the Free University and the Humboldt-University, Berlin
- Practical Course *Biologie für Mediziner*, since SS 2005. Free University Berlin

Cooperations

- Ralf Spörle, Molecular Imaging Group, MPIMG Berlin
- Frank Buchholz, MPI-CBG Dresden, Screen for EMT candidate genes using esiRNA

Development of Endodermal Organs

Head
Dr. Heinrich Schrewe (since 9/04, funded by Charité)
Phone: +49 (0)30-8413 1302/ 1518
Fax: +49 (0)30-8413 1229
Email: schrewe@molgen.mpg.de

Graduate Students
Michaela Mayer (since 7/2005)
Jesper Bramsen (University of Aarhus, 3/2005 - 7/2005)

Undergraduate Students
Ewa Ratajczak (University of Perugia, 1/2005 - 6/2005)
Michael Plötz (since 10/2005)
Pedro Rocha (Leonardo da Vinci Fellow, since 10/2005)

Technician
Manuela Scholze (funded by Charité)

Scientific Overview

Organogenesis is the process by which complex and highly specialized structures develop from a small population of undifferentiated embryonic cells. Another essential feature of this process is cell migration, and migration defects have been shown to cause a number of human diseases. The understanding of the role of signaling pathways and cellular interaction in cell differentiation, cell migration and development of organs is important for understanding tissue regeneration and repair, and ultimately growing organs in culture.

Nuclear factor One B (Nfib) and lung development

Nuclear Factor One B (*Nfib*) mutant mice provide a model to study late lung morphogenesis. Molecular embryology of the lung is still a young field, but progress has been made recently in the identification of factors involved in lung morphogenesis and a variety of mouse models have been described that result in defects in respiratory organogenesis at different stages of development. To study the function of the *Nfib* gene we have generated Nfib deficient mice. *Nfib* null mice die early postnatally of respiratory failure and display severe lung hypoplasia. Heterozygotes do survive, but exhibit delayed pulmonary differentiation. Expression of TGF-β1 and sonic hedgehog is not down regulated in mutant lung epithelium at late stages of morphogenesis, which may result in incomplete lung maturation. The *Nfib* mutant is the first and so far only described mouse mutant with a lung phenotype that affects the late stages. Detailed gene expression analyses using the Illumina microarray system, combined with large scale *in situ* hybridization and immunohistochemical techniques we will identify more effected genes and their corresponding signalling pathways as well as time points and specific cell types responsible for the late lung defects caused by the *Nfib* mutation. Transgenic phenotype rescue approaches using a lung epithelium specific promoter, as well as lung organ cultures combined with a retroviral TVA/RCAS system approach will help to unravel the developmental pathways involved in mouse lung morphogenesis.

Functional analysis of Slit-like 2 (Slitl-2) in organogenesis and cell migration

The motility of a migrating cell can be changed by extracellular molecules that attract or repel a cell. This involves a complex crosstalk between signal transduction pathways. We have recently identified the novel membrane-bound putative repellent factor Slit-like 2 (Slitl2) that shows structural and functional similarities to Slit, a secreted cell guidance molecule, isolated and characterized first in Drosophila. Slitl2 is expressed in a variety of adult tissues, mainly in kidney and brain, but also in lung and pancreas. During early embryonic development expression can be detected in organ rudiments and developing neural structures. In order to study the role of Slitl2 during mouse development, we have successfully generated *Slitl2* conditional and null alleles in mouse embryonic stem cells and Slitl2 deficient mice will be generated. In addition, we have established a tetracycline-inducible siRNA gene knock-down strategy in tissue culture and will now apply this system to generating transgenic knock-down mouse lines. Both systems will be very powerful to study Slitl2 function in embryonic development, especially in processes of cell migration, tissue branching and cell differentiation.

Perspectives

Using the described activities, we expect to identify affected signalling pathways and genes leading to the pathological organogenesis. We will apply this information to additional newly generated mouse models and organ culture approaches to gain more insight into the molecular origins of organogenesis.

Selected Information

Publications

- Finkenzeller, D., Fischer, B., Lutz, S., Schrewe, H., Shimizu, T., and Zimmermann, W. (2003). The carcinoembryonic antigen-related cell adhesion molecule 10 (CEACAM10) expressed specifically early in pregnancy in the decidua is dispensable for normal murine development. Mol. Cell Biol. 23, 272–279.
- Seeliger, M. W., Weber, B. H. F., Besch, D., Zrenner, E., Schrewe, H., and Mayser, H. (2003). mfERG waveform characteristics in the RS1h mouse model featuring a 'negative' ERG. Doc. Ophthalmol. 107, 37–44.
- Soboleva, G., Geis, B., Schrewe, H., and Weber, B. H. (2003). Sorsby fundus dystrophy mutation Timp3S156C affects the morphological and biochemical phenotype but not metalloproteinase homeostasis. J. Cell Physiol. 197, 149–156.

Student thesis

- Ewa Ratajczak, Phenotype rescue in Nfib deficient mice via transgenesis and RNA interference. BSc Thesis, University of Perugia, Italy, 2005

Teaching

- Praktikum Biologie für Mediziner, Charité Berlin (WS 2004/2005, SS 2005, WS 2005/2006, SS 2006)
- Developmental Genetics Course, School of Biosciences, University of Birmingham, United Kingdom (February 2005, February 2006)
- Master Program Molecular Medicine, Module Developmental Genetics/Model Organisms, Charité Berlin (March 2006)

Internal Cooperation

- Dr. Ralf Spörle, Department of Developmental Genetics: OPT reconstruction of mouse mutant phenotypes

External Cooperation

- Dr. Christopher Bunce, School of Biosciences, University of Birmingham, United Kingdom
- Prof. Otmar Huber, Institut für Klinische Chemie und Pathobiochemie, Charité Berlin
- Prof. Annette Schürmann, German Institute of Human Nutrition, Department of Pharmacology, Potsdam
- Prof. Michael Wakelam, Institute for Cancer Studies University of Birmingham, United Kingdom

External Funding

- Medical Research Council, United Kingdom (October 2004–September 2007)
- NucSys, Marie Curie Research Training Network (January 2006–December 2009)

Gene Expression and 3D-Reconstruction

Head
Dr. Ralf Spörle (since 11/2004; funded by NGFN)
Phone: +49 (0)30-8413-1328
Fax: +49 (0)30-8413-1130
EMail: spoerle@molgen.mpg.de

Scientists
Dr. Martin Werber (since 1/2005; part-time; funded by NGFN)
Dr. Marta Caparros-Rodriguez (since 1/2006; part-time)

Technicians
Silvia Kietzmann (since 1/2005; funded by NGFN)
Manuela Scholze (part-time; funded by Charité)

Students
Jan-Ole Christian (9/2004 - 3/2006; computer scientist)

Introduction

The group aims to understand the molecular mechanisms underlying embryonic pattern formation, e.g. the anlagen and differentiation of body regions and organs during vertebrate embryonic development. We are focusing on epithelial-mesenchymal transition (EMT), and mesoderm formation, taking place during gastrulation and in the tail bud/caudal region. Our approach includes the high-throughput temporal and spatial analysis of the activity of genes during mid-gestation davelopment (characterization of gene expression patterns and their 3D-reconstruction). Currently more than 6,500 genes have been analyzed (L. Neidhardt *et al.*, unpubl.). Based on the similarity of gene expression patterns, co- and syn-expression groups of genes are established with bio-informatic tools. Predictions of genetic interactions based on syn-expression and binding-site analysis are investigated in molecular-genetic experiments performed by collaborating groups (P. Grote, M. Morkel, L. Neidhardt, L. Wittler). The results of these studies will substantially contribute to the elucidation of genetic regulatory networks in mesoderm formation and cancer progression.

Scientific Overview

Embryonic gene expression patterns are detected in different stages of mouse embryos (whole mount mRNA *in situ* hybridization, WISH). Experiments are performed in a semi-automated high-throughput set-up with hundred genes in parallel (Neidhardt *et al.* 2000, Mech. Devel.), i.e. 373 clones of potential chromatin regulators and 688 clones from tail buds were screened recently, resulting in 10 and 20% of specific pattern genes (Fig.1). Expression patterns are subsequently documented and analyzed (photos, anatomical description following EMAP standard, database entries (R. Spörle, M. Caparros), storage/MAMEP-database, bio-informatics (M. Werber), 3D-reconstruction (R. Spörle). Currently, 12,500 clones, corresponding to 6,500 genes, with 15,000 pictures, and 1,000 gene expression descriptions have been stored in the MAMEP database.
Curation of the database content and description of gene expression patterns are the limiting steps in the procedure. Moreover, the most interesting (and complex) patterns are chosen for a more detailed and flexible analysis of expression than is possible by means of photos and nomenclature. Thus, we perform 3D-reconstructions of the stained embryos using Optical Projection Tomography (OPT, Sharpe *et al.* 2002, Science). These deliver detailed coverage of morphology and gene expression in virtual, freely orientable sections and 3D-models of selected, i.e. the tail bud-expressed, genes (Fig.2).
Gene expression patterns are attributed bio-informatically to co- and syn-expression groups. Syn-expressed genes, which are co-expressed at several places at a time, are potentially commonly regulated and/or may interact with each other. Genes coding for transcription factors can be assumed to regulate the activity of other group members. Because only genes with common sites and time of expression are included, we achieve a higher "hit rate" in predicting interactions for experimental validation.

We have generated a vast set of around 600 genes expressed in the tail bud and/or somitic mesoderm formation based on our WISHs and the literature ('caudal gene list'). This set is currently tested in different functional approaches by collaborators, including bio-informatics (T. Manke, Dept. Vingron), promoter analyses in mouse embryos (L. Neidhardt), DNA-arrays (P. Grote), and functional screens in cancer cells (M. Morkel).

Future Development

Complete systematic gene expression analysis in mouse embryos and MAMEP-database

Finalising the public section of the MAMEP-database will be furthered with priority. This includes the completion of the running screens, as well as completion and adding of descriptions and pictures of gene expression patterns. Further analyses will focus on expression profiling of differentially expressed genes. Gene expression in mouse mutants of genes implicated in tail bud and mesodern formation are compared with wild-type, in the caudal embryonic body region are profiled against the rostral region, and in different stages of mouse development are compared to each other. Further screening will cover genes missing in the MAMEP-database.

Further, an extensive set of genes activated in epaxial EMT has been collected and will be further extended based on new entries of screened genes. In mouse epaxial EMT, both the epithelial dermomyotome cells, and the mesenchymal cells arising from them, express components of major signalling pathways (*Bmps, Wnts, Fgfs* and their regulators, Spörle 2001, Dev. Genes Evol.). This set will be used to feed candidate genes into the functional assays of collaborating groups (i.e. cancer progression, M. Morkel).

We will extend our analysis to micro RNAs (miRNAs) that are short endogenous non-coding RNAs, function in the post-transcriptional regulation of gene expression, have been implicated in embryonic development, and apparently can be expressed (and regulated) together with the translated genes in which or close to which they reside (Zhao *et al.* 2005, Nature). In the mouse, their number is estimated at around 1000, and 20 to 30 % of all mRNAs are thought to be miRNA regulated (*http://microrna.sanger.ac.uk*, Bentwich et la. 2005, Nat. Genet.). Recently, 80 miRNAs have been hybridised using 'locked nucleic acids' (LNAs) in the mouse, thereof 10% exhibited a pattern (Kloosterman *et al.* 2006, Nat. Methods). We will focus on miRNAs which reside in/close to genes known to be expressed during tail bud development, mesoderm formation and heart formation (collab. with C. Grimm/Sperling group, Lehrach Dept.)

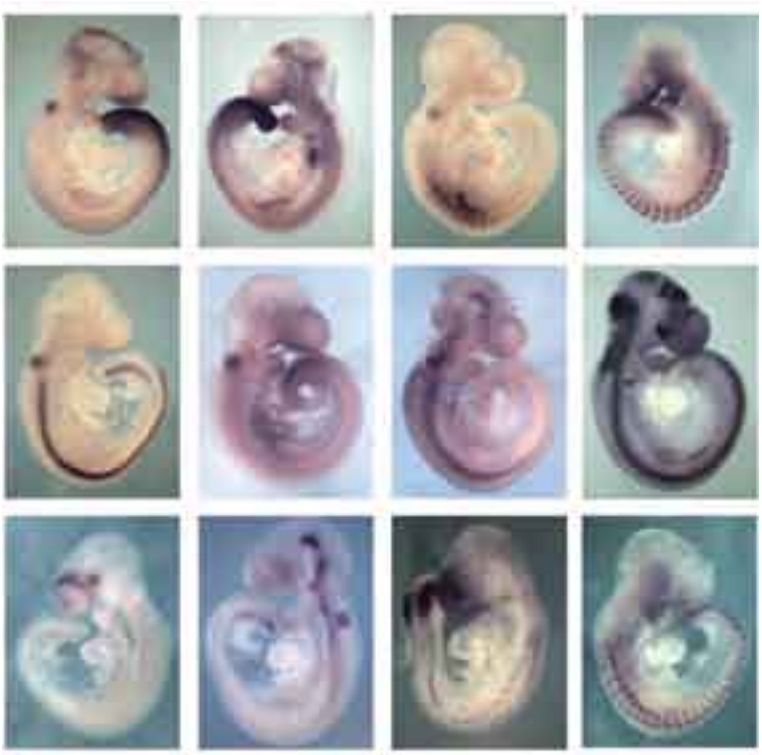


Figure 1: Examples of gene expression patterns from recent screens (chromatin regulator and tail bud screen) in whole mouse embryos.

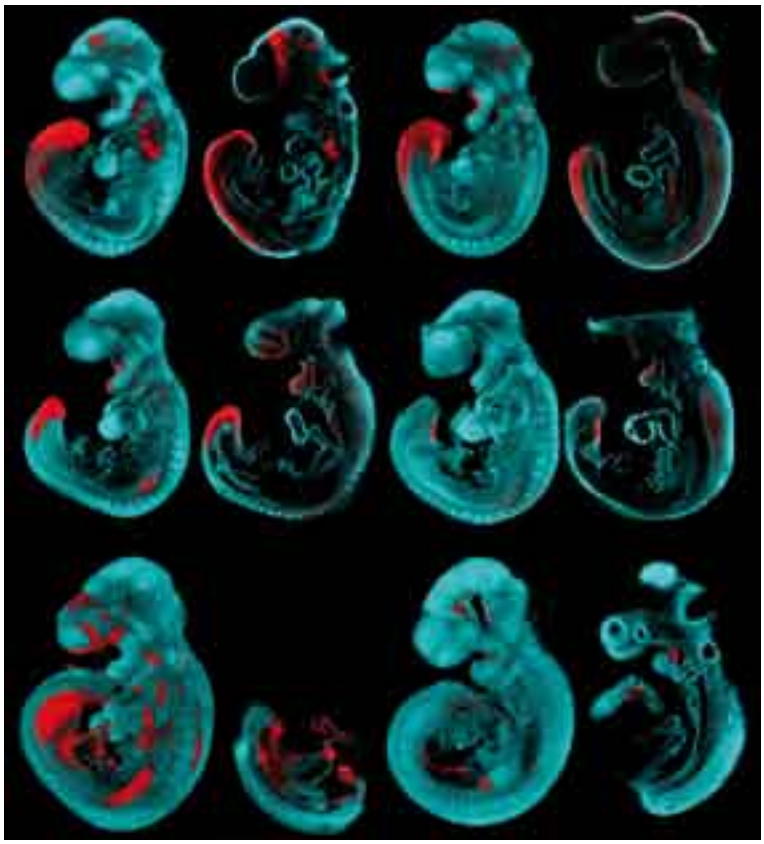


Figure 2: Six examples of 3D-representations and virtual sections after OPT-reconstruction of genes expressed during tail bud development of the mouse embryo (embryo in blue, staining in red).

Standard mouse embryo 3D-models for automated annotation of gene expression

OPT-3D-representations are ideal material to generate virtual sections, and represent a fast, flexible way to access the scientific potential of gene expression data. Moreover, a variety of re-reconstructions (rendering, mapping, and morphing) becomes possible. Both virtual sections and rendering can be analysed with unprecedented levels of freedom, and can be adopted to future needs.
Albeit surface rendering gives a plastic impression of expression shown as colourful rotatable blocks, volume rendering conserves information on expression levels and embryonic structures in reasonable detail. Mapping of such rendered 3D-representations of stained embryos should allow for a widely automated annotation of anatomical structures and expression patterns. This could be achieved by performing a "statistical" analysis of a bigger set of OPT-reconstructions from specific developmental stages, and further "mapping" the result into the annotated 3D-models of the EMAP atlas. This project requires support by informaticians involving advanced mathematics, hardware and software (collab. with A. Schliep, Vingron Dept.).

Selected Information

Publications

- Gurok U, Bork K, Nuber U, Spörle R, Nöhring S, Horstkorte R. (2006) Expression of Ndufb11 encoding the neuronal protein 15.6 during development. Gene Expression Patterns, Accepted

Student theses

- Jan-Ole Christian (Bachelor, MAMEP-database): Concept and Realisation of a Database for the Collection, Analysis and Visualisation of Gene Expression in the Mouse Embryo. Supervisors: Prof. Dr. M. Vingron, Dr. R. Spörle.

Teaching

- Biology for Medical Students, design of devel. genetics module with Drs. Schrewe, Wittler, lecture, tutorials; 6-12 half day courses, twice/year since 2004, Charité.
- Molecular Medicine (in English, design of one week's course programme with L. Neidhardt, lectures, tutorials, examinations, since 2006, Charité).
- Human Genetics weekend workshops for medical students, with Dr. Wittler, since 2006, Charité.

Public Relations

- Lange Nacht der Wissenschaften/Long Night of Sciences (since 2004, organised programme for the Dept. of Developmental Genetics in 2005 and 2006, with Dr. Wittler)
- Science Fair (public science exhibition, one week; organised programme for the Dept. of Developmental Genetics in 2005, with Dr. Wittler).
- Science Tunnel (exhibition, concept/materials on OPT for the Max-Planck-Society).

Work as scientific referee

- Development / Developmental Dynamics / Development, Genes and Evolution

Cooperations

Department of Developmental Genetics:

- Marc Leushacke/Dr. Markus Morkel: caudal gene list, knock-down of candidate genes implicated in EMT
- Lorenz Neidhardt: caudal gene list, candidate genes regulating expression in tail bud
- Dr. Phillip Grote: OPT of chromatin regulator expression, caudal gene list (DNA array)
- Michael Plötz/Dr. Heiner Schrewe: OPT of Lphn2 mutant mouse embryos
- Dr. Lars Wittler: OPT of segmentation gene expression, and of T/Brachyury mutant mouse embryos

MPI for Molecular Genetics (other departments):

- Dr. Mateusz Kolanczyk/Prof. Dr. S. Mundlos: OPT of mouse NF1 mutant cartilage and bone, publication in prep.
- Dr. Alexander Schliep (Algorithmics group)/Prof. Dr. M. Vingron: Development of standard 3D-models of mouse embryos for automated annotation of morphology and gene expression, exchange of staff planned for October 2006 (diploma student of informatics)
- Dr. Thomas Manke/Prof. Dr. M. Vingron, Dr. L. Wittler: caudal gene list: reconstruction/prediction of genetic regulatory networks downstream of T/Brachyury
- Dr. Christina Grimm/Dr. S. Sperling (Lehrach Dept.): screening for genes expressed in heart, OPT of mutant hearts, establishing LNA-micro RNA-WISH
- Daniel Birker/Dr. Georg Schwabe (Mundlos Dept., Charité): OPT of Dsh (Shh) mutant mouse embryos
- Dr. Sigmar Stricker/Prof. Dr. S. Mundlos: OPT of chick fetuses over-expressing muscle regulatory genes

Somitogenesis

Head
Dr. Lars Wittler (since 5/2004, funded by Charité)
Phone: +49 (0)30-8413 1329
Fax: +49 (0)30-8413 1229
Email: wittler@molgen.mpg.de

Engineer
Andrea König (since 9/2004)

Undergraduate Student
Eun-ha Shin (since 11/2005)

Student
Matthias Krause (since 1/2006)

Scientific Overview

Somitogenesis, the segmentation of the vertebrate body axis is an excellent model to study the molecular mechanisms of morphogenesis and differentiation. The formation of somites underlies a tight temporal control, resulting in the generation of a new pair of somites from the presomitic mesoderm (psm) every two hours in the mouse embryo. To understand the nature of this process, models were proposed involving an oscillator (clock) and a morphogen gradient (wavefront). The discovery of genes with an oscillating expression such as lunatic fringe (Fig. 1) and gradients of signalling molecules like Wnt3a and FGF8 in the psm gave access to understanding the process in its molecular detail. The search for novel genes involved in segmentation by high throughput gene expression analysis performed in our department resulted in the identification of Axin2 as a negative feedback regulator of Wnt signalling with an oscillating gene expresson in the psm of the mouse embryo. These findings integrate the clock and the wavefront, by placing the Wnt signalling cascade into the center of the process (Fig. 2). The main effort of our group focuses on further deciphering the molecular nature of the segmentation clock. Thereby we combine the advantages of two model organisms, the chick and the mouse. While the chick embryo serves as an excellent organism for microsurgical manipulation of the psm, the mouse allows to genetically dissect the segmentation process by overexpression, knock down and knock out approaches.

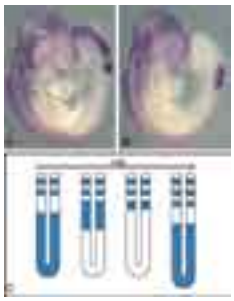


Figure 1: Dynamic expression of Lunatic fringe in the presomitic mesoderm of the midgestation mouse embryo. During one cycle of somite formation Lunatic fringe mRNA starts to be active over the entire presomitic mesoderm (A) and becomes focussed and stabilized at the forming somite (B). C Schematic representation of the Lunatic fringe expression domain over one cycle of somite formation.

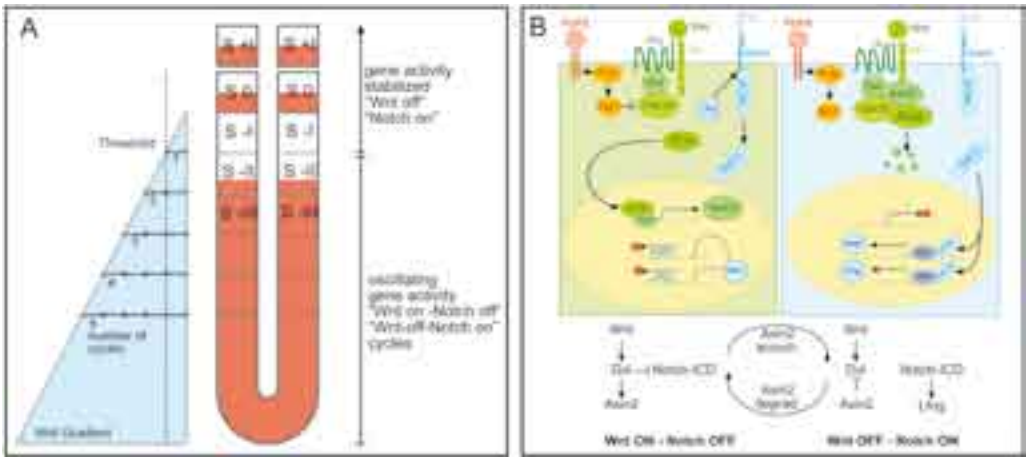


Figure 2: Control of the segmentation clock by Wnt signalling.
A: Schematic illustration of the gradient of Wnt activity over the psm. Cells exhibit oscillating gene expression along the psm until they fall under a certain threshold of Wnt activity. When the oscillation stops, gene expression becomes stabilized and a somite is formed.
B: Summary of the signalling pathways involved in the molecular oscillator. The alternating activation and downregulation of the signalling cascades and their target genes is controlled by negative feedback loops.

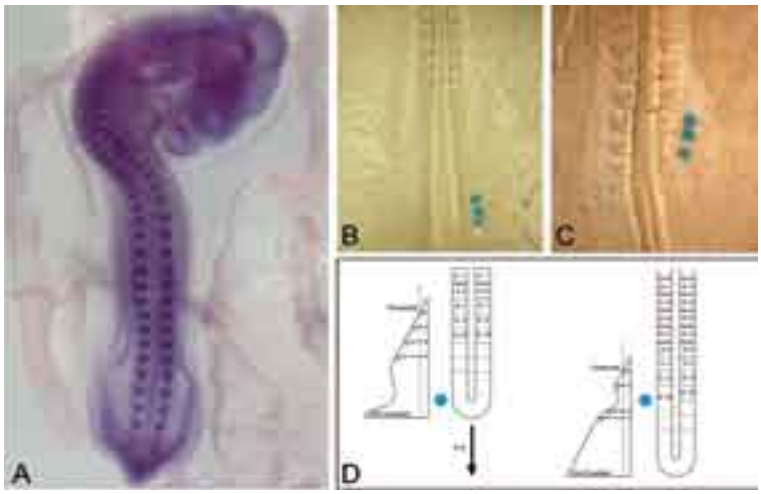


Figure 3: Manipulation of somitogenesis.
A: Chick embryo with Uncx4.1 epression detected by whole mount in-situ hybridization. The Uncx4.1 expression domain demarcates the posterior somite boundaries and serves as a marker gene for somite size analysis.
B: psm of chick embryo cultures after application of microcarrier beads loaded with inhibitors of Wnt signalling.
C: Same embryo as depicted in B after 12hrs incubation. Under the influence of the Wnt inhibiting factor the formation of the somite boundary is delayed, resultuing in an enlarged somite.
D: Model showing the action of the Wnt inhibitor on the Wnt morphogen gradient that explains the altered somite size.

Projects

Integrating the roles of the Wnt, FGF and Notch-Delta signalling pathways in somitogenesis

In the recent years it became apparent that the major signalling pathways controlling development are tightly interconnected. This makes it difficult to address the functional hierarchy of the signalling systems. In somitogenesis the connections between the three major signalling cascades, Notch, FGF and Wnt remain elusive. We use the presomitic mesoderm of chicken embryos to study the crosstalk between the signalling pathways on the organismic level. The existing broad variety of chicken embryo techniques like micromanipulation, gene overexpression by electroporation or application of drugs or protein factors that inhibit or stimulate components of the signalling cascades locally or systemically, make it an important experimental model (Fig. 3).
In cooperation with Dr. Zoltan Konthur (*in vitro* Ligand Screening Group, Department Vertebrate Genomics) and Dr. Sylvia Krobitsch (Neurogenerative Disorders Group, Department of Vertebrate Genomics), we are currently selecting aptamers and intrabodies against specific protein-protein interactions of components of the signalling pathways. The *in vivo* application of such inhibitor molecules will constitute a powerful tool to study the intracellular interconnections of the signalling cascades. The combination of these experimental approaches allows a stepwise deciphering of epistatic interactions and the hierarchical order of the pathways controlling somitogenesis.
When studying cellular differentiation processes in development, chromatin regulation and modification is a topic of increasing importance. The establishment of positional information in the anterior-posterior axis is controlled by chromatin modifying factors. In a cooperation with Dr. Phillip Grote of the Department of Developmental Genetics we elucidate the role of chromatin modifiers during somitogenesis and psm patterning in our *in vivo* model systems.

Detection and analysis of novel components of the segmentation clock

The ongoing high throughput expression analysis performed in the Department of Developmental Genetics revealed a number of genes, which show an oscillating expression pattern. Selected candidates are analysed for their potential role as regulators of the segementation clock. The functional analysis is facilitated by the comprehensive range of transgene techniques established in the department. Together with Dr. Christoph Wierling (Bioinformatics Group, Department of Vertebrate Genomics) we are currently constructing a mathematical model describing the function of the Wnt signalling cascade in the segmentation clock. The integration of novel components of the segmentation clock into mathematical models and the mathematical description of experimental alterations of the clock opens a potent interface between theoretical and experimental biology.
Additionally we perform a genomwide comparison of the psm in different phases of the segmentation clock with the Illumina microarray system established in the institute. The Illumina microarray analysis is accomplished in a cooperation with Dr. Phillip Grote and Dr. Martin Werber (both Department of Developmental Genetics).

Functional analysis of gene regulatory networks involved in the formation and patterning of the presomitic mesoderm

Based on an expression profiling analysis for factors dependent on the transcriptional activity of Brachyury/T, the regulatory elements of selected genes are being subjected to *in vitro* and *in vivo* analyses to integrate them into a regulatory network controlling mesoderm formation and somitogenesis. Binding site predictions and analysis of conserved elements is performed in cooperation with Dr. Martin Werber (Department of Developmental Genetics). We established a fast and efficient *in vivo* promoter analysis platform, based on ex ovo chick electroporation (Spieler *et al.*, 2004). This enables a preselection of promoter elements, which are then further characterised in the mouse and in cell culture transactivation assays. So far we were able to show that the regulation of Mesogenin1 is dependent on a synergistic combination of Wnt signalling and T-box transcription factors. The project is associated with the large scale promoter analysis platforms of the institute.

Future Perspectives

Our present work established and strengthened the central role of the Wnt signalling pathway in the control of the somitic clock. In the future we will apply the multitude of embryological and transgene techniques available to manipulate and control the process. By genetic modification of the oscillating components of the clock, we should be able to extend or accelerate the oscillation cycle, a direct proof for our model.
The psm arises from a group of cells in the taibud, which comprise stem cell properties and serve as a pacemaker for the segmentation clock. A major future challenge is the search for these pacemaker cells. Our genomic approaches will help to track down this stem cell population and possibly enable us to visualize the molecular oscillator also on a cellular level.

Selected Information

Publications

- Richter, U., Wittler, L. and Kessel, M. (2004). Restricted expression domains of Ezrin in developing epithelia of the chick. Gene Expr Patterns 4, 199-204.
- Spieler, D., Baumer, N., Stebler, J., Koprunner, M., Reichman-Fried, M., Teichmann, U., Raz, E., Kessel, M. and Wittler, L. (2004). Involvement of Pax6 and Otx2 in the forebrain-specific regulation of the vertebrate homeobox gene ANF/Hesx1. Dev Biol 269, 567-79.
- Wittler, L. and Kessel, M. (2004). The acquisition of neural fate in the chick. Mech Dev 121, 1031-42.

Book Chapter

- Wittler, L., Spieler, D. and Kessel, M. (2003). Hensens Node: The embryonic Organizer of the chick. In: Grunz, H. (ed), The vertebrate Organizer, Springer Verlag 395-409.

Teaching

- Biology for Medical Students, design of module on developmental genetics with Dr. Heinrich Schrewe and Dr. Ralf Spörle. Summer and winter terms since 2004; Charité – Medical Faculty, University Berlin
- Masters Programme Molecular Medicine, design of the patterning and organogenesis course of the module on Developmental Genetics/ Model Organisms with Dr. Hermann Bauer and Dr. Heinrich Schrewe. Since 2006, Charité – Medical Faculty, University Berlin
- Human Genetics weekend workshops for medical students; winter term 2006 with Dr. Ralf Spörle, Charité – Medical Faculty, University Berlin

Public Relations

- Lange Nacht der Wissenschaften; organization of the programme for the Department of Developmental Genetics, together with Dr. Ralf Spörle; 2005 and 2006
- Science Fair 2005 (public science exhibition organized by the Free University Berlin); presentation of principles and model organisms of developmental genetics; organized together with Dr. Ralf Spörle

Work as scientific referee

- for Development, Genes and Evolution

Memberships

- Member of the German Society for Developmental Biology

Cooperations

Department of Developmental Genetics:

- Dr. Phillip Grote: Microarray analysis on oscillating genes; *In vivo* analysis of chromatin modifiers
- Dr. Martin Werber: Bioinformatical predictions for gene regulatory elements involved in the segmentation process; bioinformatical evaluation of microarray results
- Dr. Ralf Spörle: OPT reconstruction of segmentation candidate genes; High throughput expression analysis of the midgestation mouse embryo
- Lorenz Neidhardt: *In vivo* analysis of promoter/enhancer elements

MPI for Molecular Genetics (other Departments):

- Dr. Zoltan Konthur, Dr. Sylvia Krobitsch (Department of Vertebrate Genomics): Selection of Aptamers/Intrabodies affecting signal transduction pathways
- Dr. Christoph Wierling (Department of Vertebrate Genomics): Mathematical modelling of the segmentation clock
- Dr. Thomas Manke (Department of Computational Molecular Biology): Caudal list: reconstruction/prediction of genetic regulatory networks downstream of Brachyury/T

External Cooperations

- Prof. Dr. Michael Kessel, Department of Molecular Cell Biology, Max-Planck Institute for Biophysical Cemistry, Göttingen
- Prof. Dr. Achim Gossler, Institute for Molecular Biology, Medizinische Hochschule Hannover

Department of Vertebrate Genomics

Head
Prof. Dr. Hans Lehrach
Phone: +49 (0)30-8413-1220
Fax: +49 (0)30-8413-1380
Email: lehrach@molgen.mpg.de

Secretary
Johanna Belart
Phone: +49 (0)30-8413-1221
Fax: +49 (0)30-8413-1380
Email: belart@molgen.mpg.de

Scientific Management
Dr. Claudia Falter
Phone: +49 (0)30-8413-1411
Email: falter@molgen.mpg.de

Dr. Ulrich Barenbrock
Phone: +49 (0)30-8413-1201
Email: barenbrock@molgen.mpg.de

Department of Vertebrate Genomics

Introduction

Since its establishment in 1994, the department of "Vertebrate Genomics" has focused on trying to understand the flow of information from the genome into the phenotype of the organisms, to understand, as far as possible, the complex networks of processes which transform the genomic information into the phenotype of the organism, and/or to develop systems able to predict aspects of this phenotype from molecular data (systems biology). For this, we have, for many years, been involved in the development of many of the techniques needed to achieve this goal, but have also contributed to many of the large "genome" data sets being developed world wide. Over the last few years, we have, in close collaboration with the departments of Bernhard Herrmann and Martin Vingron, increasingly focused on the next big challenge in genomics, the development of a systems biology of regulatory networks in different developmental processes as well as a number of disease areas.

Scientific Overview

We consider life as a computational process, translating the linear information in the genome into the phenotype of the organism, given a specific environment. This process is at the core of many key problems in basic research (e.g. development), but has also an immense practical implication, since disturbances are likely to be the basis of most or all serious common diseases.
It is however not clear at this point, if we will ever be able to 'understand' these processes in the classical sense, just as we are really not able to 'understand' the function of a neural net in detail, based on a detailed analysis of its components. Since there is an obvious analogy between such neural nets and the complex regulatory networks in biology, it is quite possible, that it will be similarly difficult to 'understand' such complex networks by studying individual components in isolation. Current problems in translating progress in basic research into clinical applications could therefore at least in part be due to this complexity, which can not adequately be addressed by the tools we have available at the moment. The fact, that the cure rates of most common cancers have hardly improved, in spite of massive investments in cancer research over the last few decades, might therefore reflect inherent limitations of current strategies.
In our view, the only long term solution to this problem is to combine functional genomics, able to identify and characterise the components of the complex networks, with systems biology, designed to develop predictive models of complex processes.
A major goal of the work of the department over the last few years can therefore be defined as the development of a systems biology of complex biological networks, with a particular focus on processes in development (early development in deuterostomes, employing human and mouse embryonic stem cells as *in vitro* models) as well as a number of disease areas (Down syndrome, heart disease, metabolic syndrome, cancer).

Tool development

Functional genomics

In spite of major progress in the tools available for a systematic analysis of structure and function of the genome, and its genes, we still have to maintain a significant effort in the development of new technologies funded to a very large extent from external sources. In many cases techniques, which we need, do not exist, do not have sufficient throughput, or are too costly to apply on large scale. Much of this technology development effort has been carried out in the context of an EU project coordinated by Ulf Landegren (Moltools), with components for the development of new techniques for (re)sequencing and genotyping (Nyarsik, Soldatov, Sauer), expression analysis by nanowell RT-PCR (Nyarsik), protein-DNA interactions (Seitz), and cell arrays (Janitz). In some cases, related work has also been funded by other sources (e.g. the development of plastic nanowell chips for genotyping and expression analysis under company funding, Nyarsik, application of such chips to understand heart disease, Bioprofile, Nyarsik, Sperling). Other areas of technology development have been in proteomics (Gobom, NGFN funding), in the analysis of structure and composition of protein complexes (Lange, Gobom, funded by EFRE and NGFN grants), the analysis of regulatory networks (Nietfeld, Janitz, Yaspo, Seitz, Hultschig, NGFN and funding), identification of protein binders by mass spectrometry (Büssow, Sauer, Bioprofil funding), new imaging techniques based on Quantum dots (Nietfeld, BMBF) and many others.

Database development

A key element in the work is the development of databases integrating the many different types of data generated within the different projects, but also available through many different external sources worldwide. For this we are involved in a number of different data base efforts (e.g. the data-base for the NGFN Protein platform (Herwig), a Laboratory Information Management System for the NGFN Antibody Factory (Konthur)). Our main tool for data integration has however been the Genome-Matrix (*http://www.genome-matrix.org*), a database/database interface system originally developed in collaboration with the RZPD. To overcome inherent technical limitations in this first implementation of the concept, we have now redeveloped this system, as part of the EU EMBRACE Consortium (Herwig), to allow more flexible storage and retrieval of information. Among other new features, this second generation system does allow the establishment of group specific sections, able to share confidential information in the context of the overall system (e.g. subsections for the groups within the institute, specific NGFN platforms, or other collaborative projects).

Systems biology

A key development for our future work is however the development of systems biology as a pragmatic response to the enormous complexity of biological networks. In this, we are following two parallel strategies. On one hand, the group of Edda Klipp, funded by the Berlin Centre for Genome based Bioinformatics (BCB), as well as by different EU grants, has been focusing on the modelling of specific pathways in *S. cerevisiae*.
As a second, more global approach, we have, over the last years, developed PyBioS, a systems biology engine based on a Python/Zope technology base. This system, developed mostly under EU funding (EMI-CD, Herwig), provides the option of carrying out modelling on the whole genome level, and has been tested in a number of different applications. This system is unique in being able to handle genome wide modelling, due to tight integration with external databases, but also due to a different focus compared to most other systems biology projects. While most of these systems attempt to revoke small segments of complex networks in isolation, our system focuses on modelling the effect of perturbations, taking advantage of a Monte Carlo strategy to compensate for the uncertainty in the kinetic data required to carry out modelling. Among other projects, this system provides the basis for our plans to develop cancer systems biology, funded at this stage as a collaborative network named ESBIC-D (Herwig), linking cancer centres and informatics groups throughout Europe.

Systems biology of regulatory networks

Genome evolution

The analysis of the evolution of genomes provides key information on the function of the different components of regulatory processes, genes (with their regulatory sequences), and the RNAs and proteins coded by these genes, particularly important in organisms, in which genetic tools are inherently limited. In an NGFN funded project, we have proceeded to complete the sequence of chimpanzee chromosome 22, as part of an international cooperation (Watanabe *et al.* 2004). In extension of this project, again under NGFN funding, we are now in the process of completing the sequence of parts of the chimpanzee X chromosome (Yaspo, Sudbrak, Reinhardt), of special interest due to the large number of phenotypes which have been mapped to this chromosome, but also due to the limited coverage of this chromosome in the shotgun sequencing strategy pursued in the US.
Our interest in the early evolution of the deuterostome genome, and in particular the genome of chordates (Panopoulou et al. 2003, Panopoulou and Poustka 2005), has led us to generating initial sequences of Oikopleura, a chordate with an exceptionally small genome. This work has been continued at Genoscope, resulting in a 5-fold genome coverage (Seo *et al.* 2004, Edvardsen *et al.* 2004, Volff *et al.* 2004, Edvardsen *et al.* 2005)
Additional contributions to genomic sequences relevant to the understanding of the evolution of the human genome has been the completion of the rat MHC sequence (Hurt *et al.* 2004, Sudbrak *et al.* 2003), contributions to the rat genomic sequence (Gibbs *et al.* 2004), as well as smaller scale sequencing projects on rhesus, macaque (NGFN funding), and mouse maki genomes.
Extending this work we are currently carrying out a detailed analysis of the evolution of promoter sequences through evolutionary shadowing the sequencing of promoters (predominantly from human chromosome 21, in 25 different monkey species (EU APES project Yaspo, coordinated by Sudbrak).

Gene regulation

The analysis of regulatory networks has become a key component of our work, carried out in close collaboration with the departments of Bernhard Herrmann and Martin Vingron. In this work, predominantly funded through the NGFN (DNA, RNAi, Epigenomics, Protein platforms) we have focused to a large extent on the analysis of genes and promoters from chromosome 21, due to the inherent medical interest (Trisomy 21 causes Down syndrome), but also due to the large amount of information already available for genes from this chromosome (Gitton *et al.*, 2002; Kapranov *et al.*, 2002; Watanabe *et al.*, 2003; Couley *et al.*, 2004).
For this, we have established a data generation and analysis pipeline to identify promoter elements, based on a detailed informatics analysis carried out in Martin Vingron's department. These segments are being amplified, and cloned into gateway vectors, allowing a functional analysis in cell arrays (Janitz, Nietfeld, Yaspo). To identify possible epigenetic effects, we are carrying out a systematic analysis of promoter methylation patterns in human blood and HEK cells, as part of the NGFN Epigenomics platform (Goyal *et al.* 2006), by ChIP on chip experiments (Yaspo), as well as by the systematic analysis of transcription factors binding to promoter arrays, using a number of different techniques (direct or indirect fluorescence labelling of TFs, Biacore surface plasmon resonance analysis) (NGFN Protein platform, Seitz, Hultschig). The complementary question, the downstream targets of regulatory genes is being pursued by inhibiting transcription factor expression by RNAi, followed by analysing the changes in gene expression patterns (NGFN RNAi platform, Yaspo, Krobitsch, Adjaye), as well as by an identification of transcription factor targets by chromatin immunoprecipitation (Sperling, Yaspo, Adjaye). Early in these projects, we have already gained much insight into patterns of gene regulation involving chromosome 21, which could help to understand some of the downstream effects of Down syndrome (Yaspo), have gained information on regulatory networks in heart development (Sperling), and have identified downstream targets and associated regulatory networks of the core transcription factors OCT4, SOX2 and NANOG which are required to maintain pluripotency in embryonic stem cells (Adjaye).

Protein interaction networks

Since many regulatory processes take place post-transcriptionally, the analysis of protein-protein interaction networks will contribute essential information, e.g. in providing molecular mechanisms for genetically identified regulatory processes. For this, we have continued to contribute to the development of large-scale 2-hybrid analysis, originally developed within the department by Erich Wanker, and extended in his position at the MDC in Berlin-Buch. This work has resulted in large-scale protein-protein interaction networks (Stelzl *et al.* 2005, Goehler *et al.* 2004).

To complement this work, we have, with funding from EFRE and NGFN, started to establish a pipeline for the systematic analysis of the composition and, in some cases, structure of large protein complexes (Lange, Gobom, Mielke). This strategy has been used extensively in the analysis of the *D. melanogaster* centrosome (Lange, Gobom), and is now being applied to a large-scale analysis of protein complexes isolated via a Tandem affinity purification (TAP)-tag strategy from HEK cells.
Technological innovations and novel applications have greatly advanced the field of protein microarrays. Over the past two years, different types of protein microarrays have been used to characterize protein-protein, protein-DNA interactions and protein-kinases (Kersten *et al.* 2004, Feilner *et al.* 2005, Seitz *et al.* in press).
To be able to identify the function of such protein-protein interaction, we have been working, next to efficient RNAi knock-down approaches (Lange), on new strategies to selectively inhibit such interactions, based on either RNA aptamers or intrabodies (Krobitsch, Konthur)

Genetics

As a new global strategy to unravel the regulatory networks in the vertebrate genome, we have started to take advantage of a new powerful genetic tool, a set of chromosome substitution strains, developed by Jiri Forejt (IMG, Prague, CZECH), carrying single chromosomes (or subchromosomal segments from PWD, a *Mus musculus musculus* inbred strain. In pilot experiments we have started to use this set of strains to map genes controlling transcript levels in the brain and other tissues (Dohrmann, Soldatov, Yaspo).

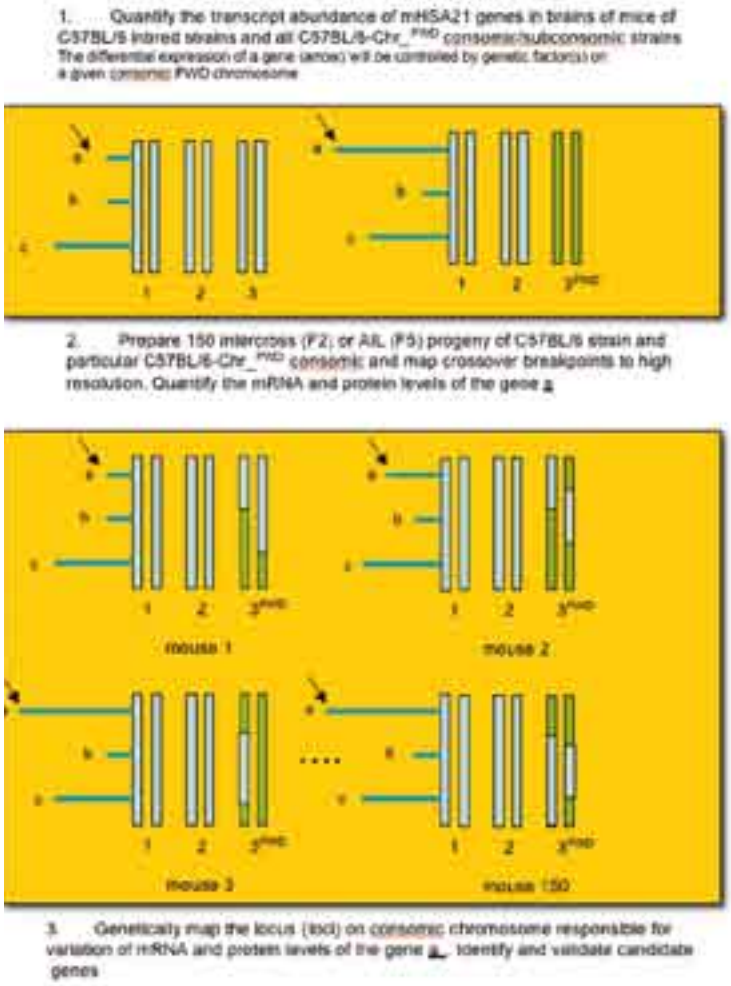


Figure 1:
Preliminary experiments to extend this analysis to factors controlling many other parameter (e. g. genomic methylation patterns, splice variants, protein abundance and modification, metabolites) have started.

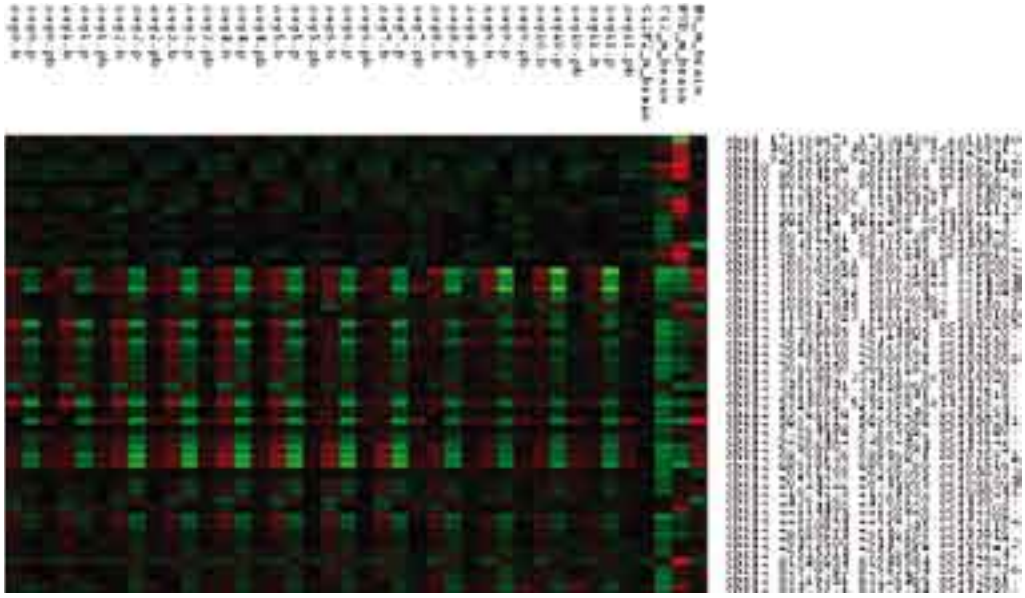


Figure 2: The image shows experimental results indicating map positions for factors regulating the expression of a subset of the genes likely to be regulated by factors on chromosome 12, based on the analysis of the first 18 intercross animals. The intensity peaks indicate the position of the regulatory factor on chromosome 12 (x axis). Differentially expressed genes and their chromosomal positions are on the y-axis. Red cells in the heatmap correspond to up-regulations, green cells to down-regulations.

This, and other related genetic tools, potentially spanning the genetic variability of all *Mus musculus* subspecies, will provide a powerful, systematic approach to unravel the entire network of regulatory processes in different tissues, and will provide a key resource in the establishment of predictive models of these regulatory networks.
In addition, our department has continued to play a key role in the German gene trap consortium, the first large scale public gene trap project in the world (Ruiz, funded through NGFN).

Systems biology of development

An important application, uniting much of the work in the department, is directed at the analysis of developmental processes. One major focus of this work is the analysis of early development in deuterostomes, the focus of the work in the group of Georgia Panoupoulou and Albert Poustka (Max Planck funding, EU Marine genomics), as well as in embryonic stem cells, analysed in the groups of James Adjaye (human stem cells, DFG and BMBF), and Heinz Himmelbauer (mouse stem cells, EU FW6 FunGenes). A similar analysis is being carried out on the development of cardiac myocytes (Sperling, funded through EU HeartRepair).
The systems biology of later developmental processes is being developed in close collaboration with the department of Bernhard Herrmann, expected to result in an increasingly accurate model of somitogenesis. EUREXpress, an EU project for the systematic high resolution analysis of the expression of all mouse genes by *in situ* hybridisation to 14.5 day mouse embryo sections will provide essential information on patterns of gene expression at this stage of development (Yaspo) (*http://genex.hgu.mrc.ac.uk/Protected/eurexpress/index.html*), complementing the work carried out in Bernhard Herrmanns department at earlier developmental stages using whole mount *in situ* hybridisation.

Systems biology of disease

Much of the justification for funding levels in biomedical research is based on the promise of translating advances in basic research into improved diagnosis and treatment for patients. With the combination of functional genomics and systems biology, we might, for the first time, have the tools in hand to fulfil this promise.
A number of projects in the department are developing in this direction. As described before, for a number of historic reasons, Trisomy 21 has been a particular focus of this work, with many aspects funded through a new EU grant, Aneuploidy (Yaspo). Similarly, work on heart disease (Sperling, Ruiz) ageing (Adjaye, Gerisch, Krobitsch) and Typell Diabetes (Herwig) can be expected to increasingly take advantage of these new tools. The analysis of neurodegenerative diseases (Krobitsch) and stroke (Nietfeld) will be in a particularly good position to benefit from the genetic crosses, which have been started. The work in a number of other disease areas is being carried out as well, either as part of a genetic strategy based on a detailed analysis of the haplotype structure of the human genome (Hoehe, funding from NGFN, Bioprofile and GIF), or as part of a number of EU grants, in which we will provide systems biology expertise and tools (Herwig).

An increasingly important focus of our work can however expected to be in the area of the systems biology of cancer, currently funded through an EU collaborative action, as well as a pilot project on the systems biology of prostate cancer from Austria (Herwig). In this, we hope to be able to develop clinically relevant models of the disease in individual patients, able to predict therapy response in specific patients.

Future developments

In the future, we anticipate an increasing focus on the analysis of regulatory networks, providing a common focus for much of the work at the institute. The genetic tools, which are becoming available, will link the work in our department, and the department of Bernhard Herrmann.
In the future, we expect to be able to shift more resources into the problem of systems biology of cancer and other common diseases, likely to provide a first target for the successful application of systems biology in a clinical setting.

Competitive position

We see the competitive position on an international level as very good. Technology development continues to provide new tools to approach complex problems. The analysis of regulatory networks integrates many techniques and resources available in this, as well as in other departments of the institute. The analysis of mouse inter-species chromosome substitution strains (consomics), which we are analysing in collaboration with Jiri Forejt, provide a unique new resource to unravel regulatory networks. The systems biology tools, which we have established, are probably unique worldwide and – with their implementation in several latest EU grants – we will have the possibility to maintain and further develop them.
Most of the work carried out within the department is however funded by external sources. Approximately 85 % of our total funding is covered by external grants, and most groups are either funded completely from the outside, or have only a very small component of their funding provided through the institute. This large external funding component is, on one hand, absolutely essential to be able to fulfil our research goals, to maintain the wide range of expertise needed, and even to establish and maintain much of the core infrastructure needed for our research, a fact, which had been clear before moving the group from London to Berlin. Acquiring and maintaining this high degree of external funding does however involve a quite significant effort on the part of the groups in question, with unavoidable effects on the work.
In spite of this, we have in our view established the basis for rapid progress in a number of important areas over the next few years. Much of the genetic, molecular and informatic tools needed to approach some of the most complex problems are now in place and we are finally in a position to approach many of the key problems, for which this infrastructure has had to be developed.

Molecular Embryology and Aging

Head
Dr. James Adjaye*
Phone: +49 (0)30-8413-1237
Fax: +49 (0)30-8413-1128
Email: adjaye@molgen.mpg.de

Scientist
Dr. Boris Greber*

Graduate Students
Thore Brink*
Marc Jung
Justyna Jagodzinska*
Doreen Janke (Practical Student, starts her diploma in 11/2006)

* external project funding

Scientific Overview

Research themes

- Deciphering the plethora of genes and complex transcriptional regulatory networks and signalling pathways operative in hES (human embryonic stem cells) and EC cells (embryonic carcinoma cells) that are required to appropriately maintain pluripotency and self-renewal.
- Oogenesis and preimplantation development
- Genetic etiology of human longevity

Ongoing research is divided into five inter-related areas

1) Transcriptional and signal transduction mechanisms regulating self-renewal and pluripotency of hES (embryonic stem cells) and EC (embryonic carcinoma cells).
2) Epigenetic and chromatin-mediated maintenance of pluripotency.
3) Derivation of functional hepatocytes from hES cells.
4) Transition from maternal to embryonic control of gene expression during preimplantation development.
5) Cellular and molecular alterations that occur in cells during aging and elucidation of related signalling and metabolic pathways

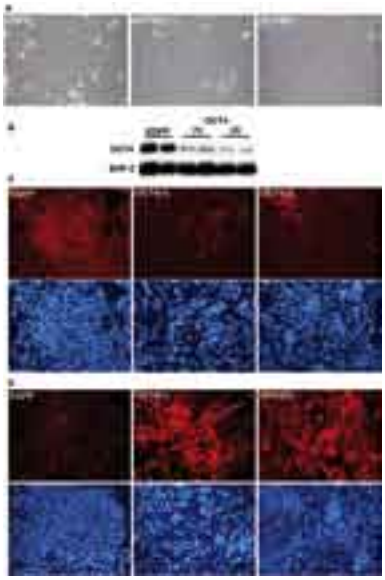


Figure 1: Down-regulation of OCT4 mRNA and protein levels in H1 hES cells induces trophoblast differentiation. (A) Morphology of hES cells 72 hours after transfection with EGFP, OCT4 (1) and OCT4 (2) siRNA. (B) Western blot analysis showing a reduction of OCT4 protein levels in siRNA transfected ES cells probed with OCT4 and SHP-2 (control) antibodies. (C) Reduction in OCT4 protein levels in siRNA transfected cells immunostained for OCT4 and counter stained with DAPI. (D) Up-regulated expression of the trophectoderm marker – Cytokeratin 18.

Background

Embryonic stem cells (ES cells) derived from the Inner Cell Mass (ICM) of the blastocyst can self-renew and are pluripotent, meaning they can be made to differentiate into cell types representative of the three embryonic germ layers, ectoderm, mesoderm and endoderm. These cells can be exploited *in vitro* to increase our meagre understanding of early developmental programmes and also for biomedical applications, disease modelling, pharmaceutical screening, and regenerative medicine.
Our research on aging is based firmly on the concept that the same signalling mechanisms that regulate development and plasticity of ES and EC (embryonic carcinoma) cells are altered during aging. Accordingly, an understanding of molecular and signalling pathways underlying the aging process is likely to lead to novel approaches to preventing and treating age-related disorders.

Current state of research

In earlier studies we uncovered gene expression patterns which suggest that the delineation of the pluripotent ICM and TE in the blastocyst is controlled by signalling pathways which are also crucial for determining cell fates later in embryonic development, and the maintenance of human ES and trophoblast stem cells *in vitro* (Adjaye *et al.* 2005).
We are now carrying out functional genomic studies based on interrogations of crucial signalling pathways employing chemical inhibitors and also RNAi coupled to whole genome transcriptome (Herwig) and proteome analysis. Additionally, chromatin immunoprecipitation assays in combination with microarray analysis (ChIP-chip) of selected transcription factors (e.g. OCT4) are employed to map transcriptional regulatory networks essential for maintaining pluripotency.

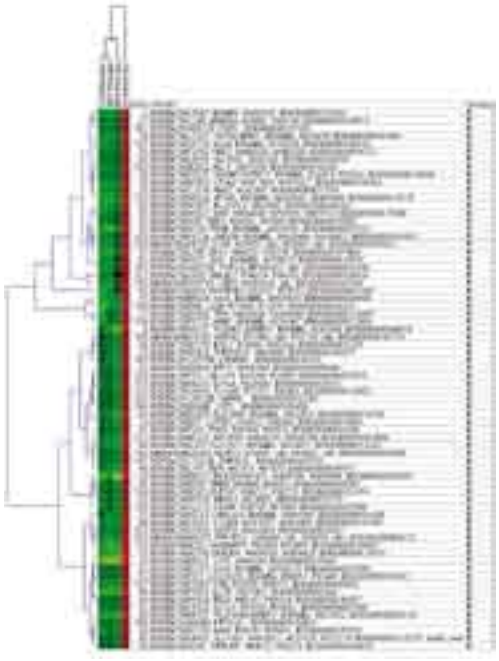


Figure 2: Hierarchical cluster analysis showing up-regulated expression of trophoblast markers such as EOMES upon loss of OCT4 function at 72hrs post transfection.

Potential pluripotency-controlling genes under investigation with the view of providing a framework for understanding the biological properties of ES cells include, transcription factors (OCT4, NANOG, SOX2), intracellular signalling factors (FGF2, CRIM1, DAB2), chromatin remodelling and epigenetic regulators (EZH2, EED, DNMT3B, JADE1).

For example, we have shown that RNAi-mediated suppression of OCT4 function induces differentiation of human ES to the trophoblast lineage, therefore recapitulating the primary differentiation event seen in the blastocyst (Figures 1 and 2; Babaie *et al.*-under review, collaboration with Ralf Herwig).

We have identified and further investigated a previously undescribed role of FGF2 in promoting self-renewal of hES cells mediated through MEFS (mouse embryonic fibroblasts) which are routinely used as feeder cells for ES cell culture. In line with our microarray data, hES cells grown in media conditioned in the absence of FGF2 differentiate significantly within 1-3 passages (Fig. 3a; Greber *et al.*, under review)

Additionally, RNAi-mediated suppression of the activity of the transcription factors OCT4 and SOX2 in hEC cells induces differentiation, which correlates with down-regulation of FGF2 as seen in the hES study. Accordingly, addition of SU5402- an inhibitor of FGF receptors induces differentiation and inhibits proliferation of ES and EC cells (Fig. 3b; Greber *et al.*, under review). Both experiments establish the essential role of FGF signalling in maintaining self-renewal and pluripotency.

As initial studies on the underlying molecular mechanism of aging, we devised an *in vitro* model of aging by employing age-related hormonal treatment of primary fibroblast and sebocyte cells (Makrantonaki *et al.*, 2006). For analysis *in vivo*, we compare epigenetic, transcriptional and proteomic changes in skin biopsies derived from young (20) and aged (60 plus) individuals and also several tissues of young (10 weeks) and old mice (14 months).

Future perspectives

Our studies on preimplantation development, embryonic stem cells and aging highlight the importance of epigenetics and chromatin remodelling in regulating differentiation programmes. We are therefore additionally pursuing research into this mode of transcriptional regulation during early development and aging.

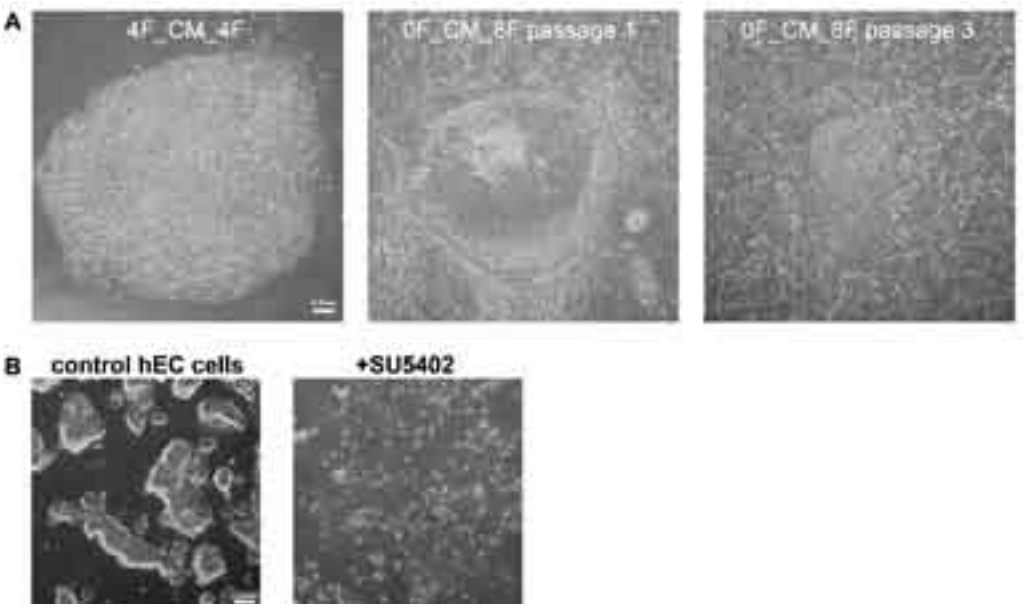


Figure 3: Effects of indirect and autocrine FGF2 signalling in human ES and EC cells respectively.
A: hES cells are maintained in medium conditioned by MEFs in the presence of FGF2 (4F_CM_4F). Omitting FGF2 in the conditioning step yields a medium (0F_CM_8F) that is less supportive, thus inducing differentiation of hES cells within 1-3 passages.
B: Autocrine FGF2 signalling is required for hEC cell self-renewal. Left panel: undifferentiated hEC cells.
Right panel: differentiation and proliferation phenotype caused by the FGF signalling inhibitor-SU5402.

Selected Information

Publications (2004–2006)

- Adjaye J (2006) Generation of amplified RNAs and cDNA libraries from single mammalian cells. In Single cell diagnostics: Methods and protocols. (Ed. Alan Thornhill) Humana Press Inc. (In press).
- Makrantonaki, E., Adjaye, J., Herwig, R., Brink, T.C., Groth, D., Hultschig, C., Lehrach, H., and Zouboulis, Ch.C. (2006). Age-specific hormonal decline is accompanied by transcriptional changes in human sebocytes *in vitro*. Aging cell 5: 331-344.
- Adjaye, J. (2005). Whole-genome approaches for large-scale gene identification and expression analysis in mammalian preimplantation embryos. Reprod., Fert., and Dev. 17: 37-45.
- Adjaye, J., Huntriss, J., Herwig, R., BenKahla, A., Brink, T.C., Wierling, C., Hultschig C., Groth, D., Yaspo, M-L., Picton, H.M., Lanzendorf, S.E., Gosden, R.G., and Lehrach, H. (2005). Primary differentiation in the human blastocyst: Comparative molecular portraits of inner cell mass and trophectoderm cells. STEM CELLS. 23:1513-1525.
- Adjaye J., Herwig R., Herrmann D., Wruck W., BenKahla A., Brink T.C., Nowak M., Carnwath JW, Hultschig C., Niemann H and Lehrach H. (2004). Cross-species hybridisations of human and bovine orthologous genes on high density cDNA microarrays. BMC Genomics 28;5(1):83.
- Papakonstantinou Eleni, Aletras Alexios J., Glass Evelyn, Tsogas Panagiotis, Dionyssopoulos Alexander, Adjaye James, Fimmel Sabine, Herwig Ralf, Lehrach Hans, Zouboulis Christos C. and Karakiulakis George (2004). Matrix Metalloproteinases of Epithelial Origin in Sebum of Facial Lesions and The Regulation by Isotretinoin during Treatment of Patients with Acne. J Invest Dermatol. 125(4):673-684.

Collaborations

- Prof. Heiner Niemann
  Transcriptomics of bovine preimplantation development
  Institute for Animal Science, (Department of Biotechnology) Germany.
- Prof Oliver Brustle – University of Bonn/ Life and Brain GmbH
  Epigenetic basis of transcriptional diversity in human ES cells.
- Prof. Zouboulis- Departments of Dermatology and Immunology, Dessau Medical Center
- Prof. Stephan Schreiber – University of Kiel
  Genetic etiology of human longevity
- Prof. Jeorg Gerlach – Charité medical school-Berlin
  Stem-cell based liver regeneration

Supervised/completed Diploma Thesis

- Thore Brink
  Analysis of gene expression during skin aging in human.

Invited Lectures

- University of Pavia – Italy (2005, 2006)
- Roslin Institute – Edinburgh (2005)
- Second international meeting of the Stem Cell Network, NRW, Bonn (2004)
- International embryo transfer society meeting. Denmark (2005)
- Second Leopoldina Conference – Embryonic and somatic stem cells – regenerative systems for cell and tissue repair. Dresden (2006)
- Royan international congress on reproductive biomedicine, stem cell biology and technology. Tehran (2006)

Prize

- Ist place winner of the 7th Royan International Research Award for reproductive biomedicine.

Public relations

- Articles about our work have been covered in the Frankfurter Allgemeine Zeitung, Süddeutsche Zeitung and the French newspaper Le Figaro. Additionally we had a press release from the institute when our manuscript Adjaye *et al.* 2005 was published.

Work as scientific Referee

- Nucleic Acids Research
- Human Molecular Genetics
- Human Molecular Reproduction
- Reproduction
- BMC Genomics
- FEBS Letters
- Physiological Genomics

External Funding

- NGFN II-EP: Genetic etiology of human longevity
- AD 184/4-1: Pluripotency controlling genes

Protein Expression and Protein Structure

Head
Dr. Konrad Büssow*
Phone: +49 (0)30-9406 2983
Fax: +49 (0)30-9406-2925
Email: buessow@molgen.mpg.de

Scientist
Dr. Christoph Scheich*

Graduate Student
Asgar Ergin

Undergraduate Student
Dimitri Soumailakakis

Engineer
Andrea Polewacz*

Technician
Janett Tischer*

* external project funding

Previous group members (2003-2006)
Dr. Hendrik Weiner, scientist
Dr. Volker Sievert, scientist
Claudia Quedenau, engineer
Thomas Faupel, Graduate student
Brigitte Hieke, technician

Scientific Overview

The group's research goal is the systematic expression, purification and analysis of eukaryotic proteins. We are taking part in projects that study protein protein interactions and solve protein structures in a systematic, high-throughput way.

Protein Structure Factory

Until 2004, the group was part of the Protein Structure Factory, PSF. The PSF was established as a structural genomics project for the systematic structural analysis of human proteins. Our task is to systematically select suitable human proteins by sequence analysis and express the proteins in recombinant form for structural analysis by crystallography and NMR spectroscopy.

To circumvent the labour-intensive cloning of single targets by PCR, we established the screening of a large cDNA expression library for human genes that can be readily expressed in *E. coli* in recombinant form (Büssow 2004). This screening approach let to a database of *E. coli* expression clones of human proteins (*http://www.proteinstrukturfabrik.de/hex1*) and to three novel protein structures (Manjasetty 2004; Manjasetty 2005).

A recent publication describes the strategies for target selection, high throughput cloning and expression clone characterisation and protein production of the PSF (Büssow 2005). A target list of about a thousand genes was compiled by sequence analysis. Systematic cloning of the targets and protein expression led to a set of clones for 139 human proteins that can be expressed in large amounts and at high yield in *E. coli* (Figure 1). The structures of eleven of these proteins were solved by the PSF (references in Büssow 2005; Manjasetty 2005 and Figure 2; Manjasetty 2006). All expression clones generated by the protein structure factory have been described in detail (Büssow 2005) and are available from the German Resource Center (RZPD).

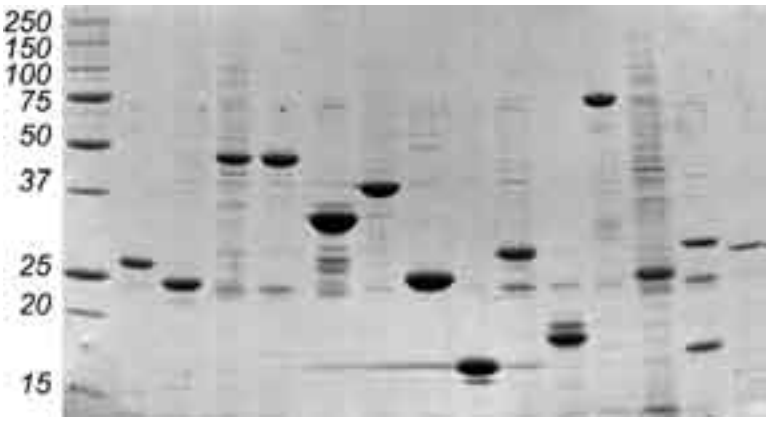


Figure 1: 14 Human proteins purified in soluble form from *E. coli* cells. Lane 1: size marker. SDS-PAGE and Coomassie staining.

Protein refolding

Only a small proportion of human proteins can be expressed in bacteria in full length and with correct folding. Typically, obtaining a human protein from *E. coli* host cells in the correctly folded form requires that the protein is either refolded *in vitro*, or that only parts of the native polypeptide are expressed. We have developed novel approaches towards both of these strategies. A novel detection method was developed for *in vitro* protein folding that integrates a number of highly sensitive biophysical and chromatographic detection methods to evaluate a microscale refolding screen for optimal conditions (Scheich 2004a). By this method, refolding can be achieved even if no biological assay is available for the protein of interest.

Fast identification of domain expression constructs for NMR

If one decides to express only a part of a protein, e.g. a single domain, rather than the complete molecule, researchers often end up with large sets of constructs representing different start and end sequence positions. We have developed a fast, small scale way of identifying constructs that are (a) solubly expressed and (b) suitable for NMR spectroscopy (Scheich 2004b).

Study of protein-protein interactions with high content protein arrays

Protein arrays that were developed in our department are now available at the RZPD. We have developed technologies for screening arrays of ten thousands of human proteins for interaction partners of proteins involved in neurodegenerative disease and signalling (Weiner 2004). We have also been involved in the identification of auto-antigens and kinase substrates using the same array platform (de Graaf 2004; Cepok 2005).

Current work

NGFN Protein Platform

The Protein Platform is a collaboration of groups of our department with the MDC Berlin, the GBF Braunschweig and the University of Rostock. The platform studies protein-protein interactions in a high-throughput, systematic way. Novel interactions are discovered in the groups of Erich Wanker, MDC and Bodo Lange/Johan Gobom, MPI-MG (Goehler 2004). Our task is to produce targets of the platform and protein complexes for structural analysis and antibody production. Targets and complexes suitable for crystallisation were selected from the large number of novel interactions discovered by yeast-two-hybrid screens at the MDC. For each target, several constructs are designed and expressed in *E. coli* and by the Baculovirus/insect cell system. We have developed a novel co-expression system, that is compatible with the Gateway cloning system used in the platform (Figure 3). Further advantages of the system are the unlimited number of subunits that can be co-expressed and the simple, PCR-free cloning procedure. See *http://www.ngfn.de*, *http://www.smp-protein.de*.

BioProfile Nutrigenomics

We are developing a new screening technology for natural products that bind to target proteins. This project is a collaboration with A. H. Pfeiffer (DiFE and Charité), AnalytiCon Discovery GmbH and Sascha Sauer of our department. We have developed a mass spectrometric binding assay for natural products. We are screening natural products provided by AnalytiCon against drug targets of the metabolic syndrome selected by A. H. Pfeiffer. The targets are cloned and produced for screening by us. Screening with the initial target, PPAR-gamma, is currently under way and five additional targets are under development. Natural products binding to target proteins will be validated by A. H. Pfeiffer's group with cellular assays. See *http://www.nutrigenomik.de*.

SFB633

We are taking part in a systematic screening project for bacterial proteins that play a role in the aetiology of Crohn's disease and related disorders. The high-throughput protein production technology developed by the PSF enables us to produce a large set of *E. coli* proteins (~300) for immunological assays. With these assays we want to identify immunodominant proteins towards T lymphocytes of Crohn's disease patients. See *http://www.sfb633.de*.

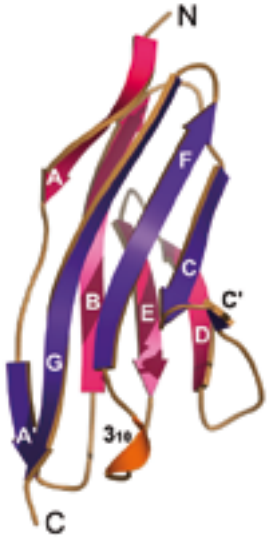


Figure 2: Structure of the human APEG-1 protein (Manjasetty 2005).

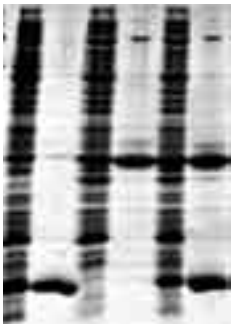


Figure 3: Protein co-expression. Cell extracts and purified proteins are shown for protein A, protein B and a clones co-expression A and B. SDS-PAGE/Coomassie staining.

Selected information

Publications

- Manjasetty, B. A., Büssow K., Fieber-Erdmann M., Roske Y., Gobom J., Scheich C., Götz F., Niesen F. H. and Heinemann U. (2006). "Crystal structure of Homo sapiens PTD012 reveals a zinc containing hydrolase fold." Protein Science 15: 914-920.
- Büssow, K., Scheich C., Sievert V., Harttig U., Schultz J., Simon B., Bork P., Lehrach H. and Heinemann U. (2005). "Structural Genomics of human proteins – target selection and generation of a public catalog of expression clones." Microbial Cell Factories 4: 21.
- Büssow, K., Quedenau C., Sievert V., Tischer J., Scheich C., Seitz H., Hieke B., Niesen F. H., Götz F., Harttig U. and Lehrach H. (2004). "A catalogue of human cDNA expression clones and its application to structural genomics." Genome Biology 5: R71.
- de Graaf, K., Hekerman P., Spelten O., Herrmann A., Packman L. C., Büssow K., Müller-Newen G. and Becker W. (2004). "Characterization of cyclin L2, a novel cyclin with an arginine/serine-rich (RS) domain: Phosphorylation by DYRK1A and colocalization with splicing factors." Journal of Biological Chemistry 279: 4612-4624.
- Scheich, C., Niesen F., Seckler R. and Büssow K. (2004). "An automated *in vitro* protein folding screen applied to a human dynactin subunit." Protein Science 13: 370-380.
- Scheich, C., Leitner D., Sievert V., Leidert M., Schlegel B., Simon B., Letunic I., Büssow K. and Diehl A. (2004). "Fast identification of folded human protein domains expressed in E. coli suitable for structural analysis." BMC Structural Biology 4: 4.
- Weiner, H., Faupel T. and Büssow K. (2004). Protein arrays from cDNA expression libraries. Protein Arrays Methods and Protocols. E. Fung. Totowa, Humana Press Inc. 278: 1-14.
- Heinemann, U., Büssow K., Mueller U. and Umbach P. (2003). "Facilities and methods for the high throughput crystal structure analysis of human proteins." Accounts of Chemical Research 36(3): 157 163.
- Scheich, C., Sievert V. and Büssow K. (2003). "An automated method for high-throughput protein purification applied to a comparison of His-tag and GST-tag affinity chromatography." BMC Biotechnology 3: 12.

External funding

- BMBF Protein Structure Factory (until 2004)
- NGFN1 Core Facility (until 2004)
- NGFN2 SMP Protein
- SFB 633
- BMBF BioProfil Nutrigenomik

Mass Spectrometry

Head
Johan Gobom, PhD*
Phone: +49 (0)30-8413-1542
Fax: +49 (0)30-8413-1380
Email: gobom@molgen.mpg.de

Scientists
Klaus-Dieter Klöppel, PhD.
Ekaterina Mirgorodskaya, PhD.*
Eckhard Nordhoff, PhD.*

Technicians
Beata Lukaszewska-McGreal*
Dorothea Theiss*

*external project funding

Scientific Overview

The Mass Spectrometry Group conducts proteome research based on mass spectrometry (MS) and associated protein- and peptide-separation techniques. The work encompasses both the development of new experimental methods and data interpretation routines, as well as their application to biological questions. The technology developments of the group since 2003 reflect the current general trend in many proteomics laboratories of moving away from two-dimensional gel electrophoresis (2DE) as the main technique for protein separation and quantitative/comparative analysis, in favour of liquid chromatography (LC) coupled to mass spectrometry. Another important focus is quantification and comparative analysis, which has become key issues in current proteome research. Two new analytical strategies based on stable isotope labelling and MS have been developed and implemented in the current research projects.

LC-MALDI MS

The technology development of the group includes a LC-MALDI interface that allows collection of LC-fractions of volumes as low as 80 nano-litres onto MALDI sample supports, prepared with microcrystalline layers of CHCA matrix (Mirgorodskaya *et al.* 2005a). Thus prepared MALDI samples are suitable for automated MS- and MS/MS-analysis. This development enabled the establishment of LC-MALDI MS in the lab as a routine analysis technique for protein mixtures that, for many applications, has replaced 2DE.

While electrospray (ESI) has been used as a detector for LC separations for a long time, the use of MALDI is relatively new. For protein identification both techniques employ the same general strategy: first a mass spectrum is acquired from the LC effluent. Based on the *m/z* values of the contained analytes, determined from this spectrum, selected peptide ions are then isolated in the mass spectrometer and subjected to fragment ion analysis. The recorded fragment ion spectra are used for protein identification by database searching. A distinct advantage with MALDI MS is that the acquisition of MS- and MS/MS- data can be separated from each other, which permits the implementation of efficient data-dependent analysis. With ESI, the time available for MS/MS data acquisition of a given peptide is limited to the time-width of its chromatographic peak, which is typically only a few seconds. For complex peptide mixtures, in which several analytes co-elute, the time available for analysis is often too short for acquisition of fragment ion spectra of all compounds. For MALDI, in contrast, all LC fractions can first be analysed in MS mode (which only consumes a minor part of the samples). Then, based on the entire MS data set, decision can be taken which peptides have to be analysed in MS/MS mode. This can be done without time-restraint, since the remainder of the sample can be stored on the MALDI sample plate for several days without degradation of the sample. Information obtained by a first set of MS/MS analyses can be used to guide the selection of precursor ions for further MS/MS analysis.

Below is shown the LC-MALDI-MS peptide profile obtained from a centrosome preparation of Drosophila, following tryptic digestion. Each detected compound (indicated by a black dot) is characterized by its m/z value (x-axis) and LC-retention time (y-axis). A total of 67 proteins were identified from this LC-separation.

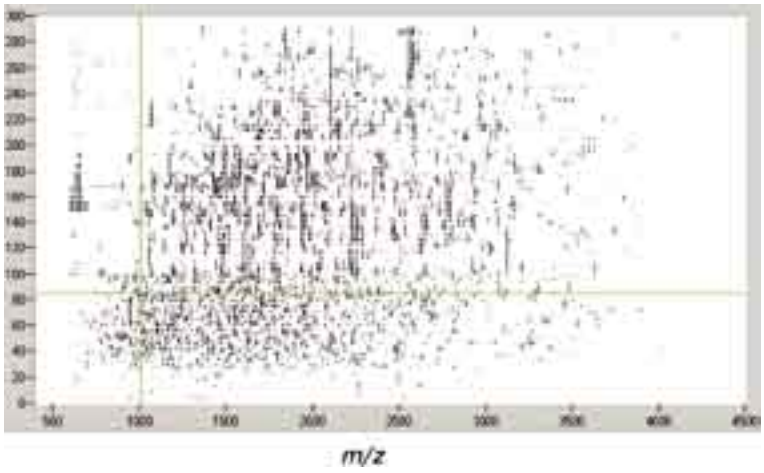


Figure 1: LC-MALDI MS peptide profile of a centrosome preparation of Drosophila following tryptic digestion. Each detected peptide is characterized by its determined molecular mass and LC retention time.

Proteomic studies based on 2DE-MS

While the ongoing studies performed in the group are mainly based on LC-MALDI MS, most proteomic investigations in the period 2003 – 2006 relied on the well-established analytical platform of 2DE-MS. Below, a few of these studies are briefly described.

Proteomic investigation in Crohn's disease:
The proteomic changes reflecting inflammatory signaling were analysed by 2DE of protein extracts of human cultured cells stably transfected with the NOD2 wild type gene or the truncated NOD2 (SNP13) gene. The two cell lines and the mock -transfected control cell line were subjected to stimulation with muramyl dipeptide (MDP), and the proteome changes over a time course were monitored. Differentially regulated proteins were identified by MALDI-TOF MS peptide mass fingerprinting and MALDI MS/MS. A large number of differences in the protein profiles of the different NOD2 genotypes were observed both under normal growth conditions and after stimulation with MDP. These findings are of relevance to our understanding of the influence of the NOD2 genotype on the etiopathogenesis of chronic inflammatory bowel disease (Weichart *et al.* 2006).

Proteome analysis of Arabidopsis thaliana:
A large-scale proteome analysis of *A. thaliana* was performed. Protein extracts were analysed by 2DE and MS, resulting in 2DE reference maps of eight different tissues and identification of approx. 3,000 proteins (Givavlisco *et al.* 2005).

Proteome analysis of Rhodopirellula baltica:
A systematic proteome analysis of the marine bacterium *Rhodopirellula baltica* using 2DE-MS was performed, resulting in 2DE reference maps and identification of more than 1,000 proteins (Gade *et al.* 2005a). Building on this study, the carbohydrate metabolism of the bacterium was investigated by a comparative proteomic study using differential fluorescence labeling and 2DE (Gade *et al.* 2005b).

Selected Information

Publications

- Weichart, D., Gobom, J., Klopfleisch, S., Hasler, R., Gustavsson, N., Billmann, S., Lehrach, H., Seegert, D., Schreiber, S., and Rosenstiel, P., (2006) Analysis of NOD2-mediated proteome response to muramyl-dipeptide in HEK293 cells. J Biol Chem, 281: p. 2380 - 2389.
- Giavalisco, P., Nordhoff, E., Kreitler, T., Kloppel, K.D., Lehrach, H., Klose, J., and Gobom, J., (2005) Proteome analysis of Arabidopsis thaliana by two-dimensional gel electrophoresis and matrix-assisted laser desorption/ionisation-time of flight mass spectrometry. Proteomics, 5(7): p. 1902-13.
- Giavalisco, P., Wilson, D., Kreitler, T., Lehrach, H., Klose, J., Gobom, J., and Fucini, P., (2005) High heterogeneity within the ribosomal proteins of the Arabidopsis thaliana 80S ribosome. Plant Mol. Biol., 57: p. 577-591.
- Gustavsson, N., Greber, B., Kreitler, T., Himmelbauer, H., Lehrach, H., and Gobom, J., (2005) A proteomic method for the analysis of changes in protein concentrations in response to systemic perturbations using metabolic incorporation of stable isotopes and mass spectrometry. Proteomics, 5(14): p. 3563-70.
- Mirgorodskaya, E., Wanker, E., Otto, A., Lehrach, H., and Gobom, J., (2005) Method for qualitative comparisons of protein mixtures based on enzyme-catalyzed stable-isotope incorporation. J Proteome Res, 4(6): p. 2109-16.
- Mirgorodskaya, E., Braeuer, C., Fucini, P., Lehrach, H., and Gobom, J., (2005) Nanoflow liquid chromatography coupled to matrix-assisted laser desorption/ionization mass spectrometry: Sample preparation, data analysis, and application to the analysis of complex peptide mixtures. Proteomics, 5(2): p. 399-408.
- Sauer, S., Lange, B.M., Gobom, J., Nyarsik, L., Seitz, H., and Lehrach, H., (2005) Miniaturization in functional genomics and proteomics. Nat Rev Genet, 6(6): p. 465-76.
- Kersten, B., Possling, A., Blaesing, F., Mirgorodskaya, E., Gobom, J., and Seitz, H., (2004) Protein microarray technology and ultraviolet crosslinking combined with mass spectrometry for the analysis of protein-DNA interactions. Anal Biochem, 331(2): p. 303-13.
- Muller, H., Lehmann, V., Gustavsson, N., Gobom, J., and Lange, B.M.H., (2004) Identification and molecular characterisation of centrosomal proteins in Drosophila melanogaster. European Journal of Cell Biology, 83: p. 76-76.

Invited lectures

- NanoLC-MALDI-TOF MS- and MS/MS in proteome research. Conference: "Functional Genomics: From Bacteria to Man", September 7 – 10, 2003, Greifswald.

Student theses

- "Proteomanalyse von Pflanzen", Fachbereich Biologie Chemie Pharmazie der Freien Universität Berlin. Patrick Giavalisco, Berlin 2003.

External Cooperations

- Bruker Daltonics GmbH, Bremen: development of LC-MALDI MS, hardware and data analysis software.
- The workgroup Algorithms in Bioinformatics, Free University, Berlin (Prof. Reinert): Development of MS data analysis algorithms.
- The Neuroproteomics Group, Max Delbrück Center, Berlin-Buch (Prof. Wanker): analysis of protein interactions in human brain.
- Max Planck Institute for Marine Microbiology, Bremen (Prof. Rabus): Proteomic analysis of Rhodopirellula baltica.
- Institute for Clinical Molecular Biology, University Hospital Schleswig Holstein, Campus Kiel (Prof Schreiber S): proteomic study of chronic inflammatory bowel disorder.
- Protein Structure Factory, Berlin: Characterization of recombinantly expressed proteins prior to structure determination.

External Funding

- IBB-ZF 2D/3D Protein Chips
- NGFN II SMP Protein

Bioinformatics

Head
Dr. Ralf Herwig
Phone: +49 (0)30-8413 1265
Fax: +49 (0)30-8413 1380
Email: herwig@molgen.mpg.de

Scientists/Developers
Thomas Kreitler* (since 11/2004)
Christoph Wierling* (since 3/2001)
Detlef Groth*, PhD (since 3/2001)
Wasco Wruck* (since 3/2001)
Elisabeth Maschke-Dutz* (since 1/2003)
Raffaello Galli* (since 3/2001)
Thomas Meinel* (since 11/2005)
Andriani Daskalaki* (since 3/2006)

Graduate Students
Hendrik Hache*(since 4/2003)
Axel Rasche* (since 11/2004)
Mario Drungowski* (since 3/2004)
Lukas Chavez Wurm* (since 4/2006)

Undergraduate Students
Atanas Kamburov (since 6/2005)
Marcus Albrecht* (since 4/2003)
Andre Rauschenbach* (student Bioinformatics, since 9/2005)

*external project funding

Former Members
Steffen Hennig (now RZPD Berlin)
Matthias Steinfath (now University Potsdam)
Claudia Schepers (now Merck Darmstadt)

Scientific Overview

The information that we can gain about a biological system (for example a disease process) appears in practice as an experimental observation, and research is restricted to the granularity and the precision of the experimental techniques in use. **Bioinformatics** research has seen a plethora of methods and tools published in recent years for gene expression, sequence and other analyses each targeting a specific bit of information in the puzzle (for example the transcriptome state). We contribute to these developments, with focus on non-parametric and information theoretic methods for gene expression and proteomics data. Data analysis components were applied to research projects with various disease backgrounds such as chromosome 21 (Kahlem *et al.* 2004), ageing (Makrantonakis *et al.* 2006) and stem cells development (Adjaye *et al.* 2005). Currently, we set up large data resources for cancer and type-2 diabetes.

Disease processes are complex and the systematic study of such processes involves typically a number of different labs working together at the same problem domain with the use of different technologies. Here, **data integration** is a major challenge. Research focuses on the correlation and comparative analysis of data. Our group performs such data integration analysis within different consortia, for example the SMP Protein of the National Genome Research Network (NGFN), the DFG research group DRIP, the EU Network of Excellence EMBRACE among others. In the SMP Protein we developed the *pLuXI* system – an XML-based data integration system for state-of-the-art proteomics technologies such as MS, Y2D, gels and protein structure data. Furthermore, we have integrated schemes for gene expression data, literature and meta-data and patient information to the system. *pLuXI* is a common platform which allows storing the experimental results in a very flexible way. We used the XML technology to set up the main database. Like SQL-based databases, our database can be easily maintained and is widely hardware independent. For access and visualization of data the XML schemata were embedded in a TWIKI infrastructure, which allows all members of the platform to make extensions and to add data. Using proper Java-tools data can be up- and downloaded and analyzed. For tracking genomic and disease-related information we have developed the *GenomeMatrix* system. For annotation of genes we have developed the *Goblet* server (Hennig *et al.* 2003; Groth *et al.* 2004). This system allows the annotation of gene ontology information via a comprehensive set of databases.

Systems biology aims at the explanation of physiology and disease from the level of interacting components such as molecular pathways, regulatory networks, cells, organs and ultimately the entire organism. With the use of computer models for such processes *in silico* predictions can be generated on the state of the disease or the effect of the individual therapy. The new approaches are about to revolutionize our knowledge on disease mechanisms and on the interpretation of data from high-throughput technologies. Crucial for the step from qualitative, explorative data analysis to quantitative, predictive analysis is the combination of experimental data with the knowledge of the underlying biological reaction system. This approach makes it possible to come up with conclusions about new properties of the system, even those that are not subject to experiments or are not

even amenable by any experimental approach. For that purpose we have developed the modeling and simulation system PyBioS (Wierling 2006; patent application pending). Our group has initiated several national and international systems biology projects in recent years, such as the EMI-CD project, one of Europe's first funded SB projects, and the ESBIC-D project.

Statistical data analysis, Data integration and database development and Systems Biology

Generally, our efforts can be subdivided in three parts: For each part we list a recent example of work.

Statistical data analysis

Genes do not act as individual units, they collaborate in overlapping pathways, the deregulation of which is a hallmark for the disease under study. We developed new bioinformatics tools that judge gene expression changes in the context of such pathway analysis. Our method has been applied recently for the identification of pathways altered upon differentiation of inner cell mass and trophectoderm in the human blastocyst (Adjaye *et al.*, 2005) and upon hormone-induced aging of the human skin (Makrantonaki *et al.*, 2006). The procedure is based on a priori pathway annotation (for example provided by public pathway databases). This information is then translated into a two-dimensional statistical test problem that involves robust non-parametric statistical testing to compute for each pathway a Z-score that quantifies the degree of alteration across the different experimental conditions. The results of the pathway analysis in the latter study, for example, implicate the involvement of several metabolic pathways in the aging process, such as C21-steroid hormone metabolism, phospholipid degradation, prostaglandin and leukotriene metabolism, 2,4-dichlorobenzoate degradation and fatty acid biosynthesis. Interestingly, pathways operative in neurodegenerative disease such as HD, DRPLA and ALS showed also significant age-dependent expression changes.

Data integration and database development

For visualization and information management of functional genomics data we have developed the *GenomeMatrix* system. In a rectangular view the user has the possibility to select genes according to several criteria (e.g. "kinase", chromosomal position, gene expression state) and to view comprehensive information on this set of genes. The GenomeMatrix is part of the European Network of Excellence EMBRACE that coordinates the integration of Europe's most important databases. The GenomeMatrix contains currently data for 21 species. The data are based on the latest Ensembl genome annotations. Large datasets have been integrated, for example approx. 350 microarray expression datasets taken from recent publications either for normal tissues or for disease related data sets. Furthermore links to web databases were added via virtual datasets.

Systems Biology

We have developed an experimental driven modelling system in recent years (Wierling 2006). With the modelling and simulation system PyBioS it is possible to construct models that are based on the topology of a cellular reaction network and adequate reaction kinetics. Based on this information the system can automatically construct a mathematical model of differential equations that can be used for subsequent simulation of the temporal behaviour and model analysis. Particularly, information on the topology of biological systems is available from several databases. PyBioS provides interfaces to these databases that can be used for the construction of appropriate model prototypes. Models include metabolic pathways, signal transduction pathways, transport processes, gene regulatory networks, among others, and can be accessed via a web- interface. The PyBioS software has won the 3rd price of the Heinz Billing Award for Scientific Computation of the Max Planck Society. It is integrated in seven different national and international consortia.

Future perspectives

The main sociological and economical impact of genome research is the molecular understanding of major human diseases and the development of new therapies and medicals for the combat of these diseases. These diseases, like many others, are caused by disturbances in the complex networks of biological processes in the organisms. Prevention, diagnosis and therapy of these diseases require a detailed understanding of those processes in health and disease, based on the application of techniques from the area of functional genomics on the individual patient, combined with the development of computational systems, that are able to model the disease process, based on available information as well as the information gained on the individual (genetic factors, chromosomal changes, expression patterns on the RNA and protein levels etc.).
Major future goals will be to integrate data and analysis resources available at clinical and functional genomics centres to develop and use new mechanisms to identify pathways likely to be most relevant to the problem domain (for example cancer), to incorporate these into the models, and to carry out benchmarking studies on the pilot scale to identify strength and weaknesses of current approaches. This translational component will be the driving element in future computational biology and systems biology research.
Towards this perspective, the group has been successful in maintaining its national and international funding. Current contract negotiations in Framework 6 encompass the EU integrated projects CARCINOGENOMICS and ANEU where it carries out data integration and modeling analysis with respect to carcinogenic and genotoxic responses to chemical compounds and the effect of aneuploidies (in particular trisomy 21). Furthermore, two EU STREPS will allow us to apply and expand our technology in the context of type-2-diabetes (SysProt) and host-pathogen interactions upon infection (SysCo). In the BMBF QuantPro program the group successfully applied for metabolic modeling in the context of stem cells research. All projects are supposed to start 2007. A further collaborative project started in April 2006 with the University of Innsbruck on the systemic analysis of prostate cancer patients will enforce our current research on cancer systems biology.

Selected Information

Publications

- Makrantonaki, E., Adjaye, J., Herwig, R., Brink, T.C., Groth, D., Hultschig, C., Lehrach, H., and Zouboulis, C.Z. (2006) Age-specific hormonal decline is accompanied by transcriptional changes in human sebocytes in vitro. Aging Cell 5(4):331-44.
- Herwig, R., Lehrach, H. (2006) Expression profiling of drug response – from genes to pathways. Dialogues in Clinical Neuroscience, in press.
- Adjaye, J., Huntriss, J., Herwig, R., BenKahla, A., Brink, T.C., Wierling, C., Hultschig, C., Groth, D., Yaspo, M.-L., Picton, H.M., Lanzendorf, S., Gosden, R.G., and Lehrach, H. (2005) Primary differentiation in the human blastocyst: Comparative molecular portraits of inner cell mass and trophectoderm cells. Stem Cells 23:1514-1525.
- Groth, D., Lehrach, H., Hennig, S. (2004) GOblet: a platform for Gene Ontology annotation of anonymous sequence data. Nucleic Acids Research 32 (Web Server issue):W313-W317.
- Kahlem, P., Sultan, M., Herwig, R., Steinfath, M., Balzereit, D., Eppens, B., Saran, N.G., Pletcher, M.T., South, S.T., Stetten, G., Lehrach, H., Reeves. R.H., and Yaspo, M.-L. (2004) Transcript Level Alterations Reflect Gene Dosage Effects Across Multiple Tissues in a Mouse Model of Down Syndrome. Genome Research 14:1258-1267.
- Dickmeis, T., Plessy, C., Rastegar, S., Aanstad, P., Herwig, R., Chalmel, F., Fischer, N., and Strähle, U. (2004) Expression profiling and comparative genomics identify a conserved regulatory region controlling midline expression in the zebrafish embryo. Genome Research 14:228-238.
- Dieterich, C., Herwig, R., Vingron, M. (2003) Exploring potential target genes of signaling pathways by predicting conserved transcription factor binding sites. Bioinformatics 19 (Suppl. 2):ii50-ii56.

Books/Book chapters

- Klipp, E., Herwig, R., Kowald, A., Wierling, C., and Lehrach, H. (2005) Systems Biology in Practice. Wiley-VCH, Weinheim.
- Herwig, R., Schuchhardt, J., Eickhoff, H., Herzel, H., and Lehrach, H. (2003). Datenanalyse von Biochips: Von der Sequenz zum System. In Grundlagen der Molekularen Medizin (D. Ganten and K. Ruckpaul (eds.)), Neuauflage, Springer Verlag, Berlin, pp. 360-387.

Invited plenary lectures (selection)

- 14.03.2003 27th Annual Conference of the Gesellschaft für Klassifikation (GfKl).
- 10.-11.09.2003 Computational Systems Biology (CSB) - Its future in Europe, EU workshop Brussels.
- 20.09.2005 GBM Conference, Berlin.
- 26.09.2005 Berlin meets Asia, workshop, Berlin.
- 27.04.2006 Spring Academy Systems Biology, Magdeburg.
- 30.06.2006 ICT for biomedical sciences, Brussels.
- 04.05.2006 OGSB Ontology, Grid and Semantic Integration for Biology, Workshop Bordeaux.

Prizes

- Christoph Wierling (2005) PyBioS – ein Modellierungs- und Simulationssystem für komplexe biologische Prozesse. Heinz-Billing Award for the Advancement of Scientific Computation of the Max-Planck Society. 3rd Price.

Patents

- Patent No.: EP 2005/005357, Prediction of the dynamic behavior of a biological system (PyBioS).

Selected International Cooperations

- Dr. Ewan Birney, EBI, ENSEMBL group, Cambridge, UK.
- Prof. Dr. Graham Cameron, EBI, Cambridge, UK.
- Prof. Dr. John Williams, Roslin Institute, Edinburgh, UK.
- Dr. Crispin Miller, Paterson Institute for Cancer Research, Manchester, UK.
- Dr. Chris Dodge BioWisdom SRS, Cambridge, UK.
- Dr. Christophe Blanchet, CNRS, Lyon, France.
- Dr. Emmanuel Barillot, Institut Curie, Paris, France.
- Dr. Jean-Philippe Vert, ARMINES, Paris, France.
- Prof. Dr. Ron Shamir, Department of Computer Science, Tel-Aviv University, Israel.
- Prof. Dr. Koussay Dellaghi, Institut Pasteur, Tunis, Tunesia.
- Prof. Dr. Kurt Zatloukal, Medizinische Universität Graz, Austria.
- Prof. Dr. Bartsch / Prof. Dr. Günther Bonn, Universität Innsbruck, Austria.
- Dr. Adam Antebi, University of Texas, Houston, USA.
- Dr. Winston Hide, South African National Bioinformatics Institute, Capetown.

External Funding

- EU Projects: EMI-CD, ESBIC-D, EMBRACE
- NGFN II SMP Protein
- BMBF FUGATO IRAS

Comparative and Functional Genomics

Head
Dr. Heinz Himmelbauer
Phone: +49 (0)30-8413-1354
Fax: +49 (0)30-8413-1128
Email: himmelbauer@molgen.mpg.de

Scientists
Dr. Katarina Bilikova (8/2003 - 3/2004; MPG Partnering Group)
Dr. Andreas Ludewig*

Engineers, Programmers
Helena Tandara* (until 3/2006)
Thomas Meinel* (until 12/2005)
Jan Wolfertz*

Technicians
Stefanie Palczewski*
Marion Klein*
Steve Hermann*

Graduate Students
Anja Berger* (until 3/2005)
Boris Greber* (until 10/2004)
Maryam Zadeh Khorasani* (until 12/2004)
Peter Hurt (until 2/2004)
Tobias Nolden*
Cornelia Lange*
Juliane Dohm*
Florian Mertes*
Ruben Rosenkranz*

Undergraduate Students
Gabriele Imre* (until 6/2003)
Christoph Campregher* (until 4/2003)

Students
Elham Alishirkonhi* (until 6/2003)
Anja Feldner*

* external project funding

Scientific Overview

Current state of research and future perspectives

With the sequence of the human genome completed and the majority of human genes known, the focus shifts to model organisms, for comparative genome analysis and for the analysis of gene functions. Our model system since 1995 has been the mouse. Complementing work on medaka (rice fish) was initiated in 2001. For a deeper understanding of major histocompatibility complex (MHC) structure and evolution we first determined the genomic sequence of the rat MHC, and presently analyse the MHC of platypus, a monotreme. A new project is the physical mapping of the sugar beet genome. This, for once, allows us to refine the physical mapping technologies developed by us in the course of previous mapping projects (mouse and medaka), and, secondly, provides us with an excellent entry point for the comparative analysis of plant genomes, which are far more plastic and evolve much more rapidly than animal genomes.

Mouse functional genomics

One focus of the lab is to create a resource of mutagenised ES cells that can be screened for mutations within any gene, with particular focus on genes with biomedical relevance, such that appropriate mouse models can be generated. Ethylnitrosourea (ENU) is the most potent mutagen. Using ENU, we created two libraries encompassing in total 80,000 mutagenised ES cell clones, so far the largest library of mutagenised ES cells worldwide. In parallel, mutation detection technologies were evaluated for their power to detect point mutations and deletions in pools of mutagenised cells, e.g. heteroduplex analysis (WAVE), capillary electrophoresis, Cel I mismatch cleavage for the detection of point mutations and exon-skipping PCR for the detection of deletions in cDNA, caused by aberrant ENU-induced splicing. So far, we have successfully identified mutations within several genes of interest, and, as proof of principle, transferred a splicing defect in the mouse Kit gene to the germline (Fig. 1). Further work to parallelize mutation screening is presently ongoing, to assemble a catalogue of mutations, to have rapid access to mutations of potential interest.

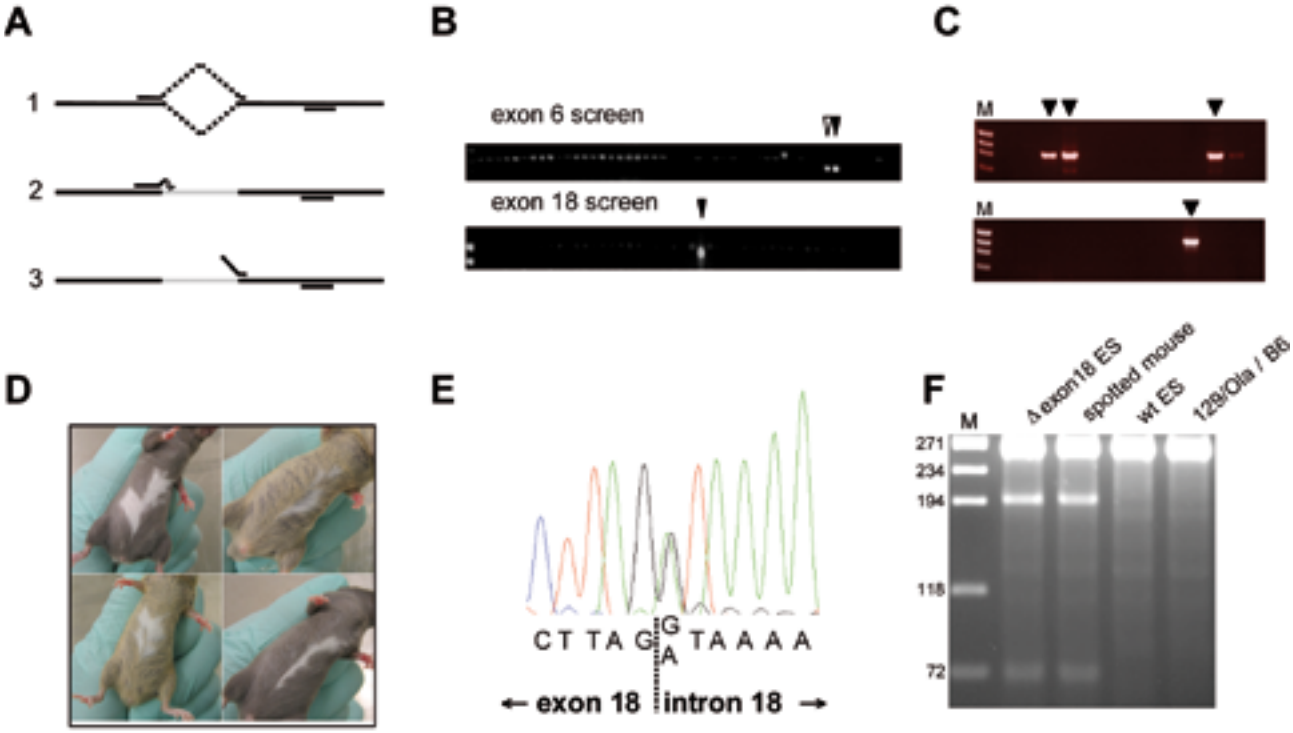


Figure 1: Screening of an archive of ES cell archive for ENU-induced mutations for splicing defects in the mouse Kit gene. **A:** Screening strategy based on the use of exon skipping primers that will selectively amplify mutants (1) but will fail to amplify wildtype (1,2). **B:** Identified mutant cell clones in primary screen. **C:** Confirmation of skipping events on plated clones retrieved from archive. Delta exon 6 (top), Delta exon 18 (bottom). **D:** Heterozygous mutant offspring obtained after germline transmission of the Delta exon 18 mutation. **E:** Molecular characterisation of defect in Delta exon 18 reveals splice donor mutation. **F:** Cel I – based confirmation of presence of mutation in heterozygous mutant.

Structure and evolution of the mammalian major histocompatibility complex (MHC)

We have previously reported the genomic sequence of the rat MHC, spanning almost 4 Mb. The rat MHC shares a common architecture with other characterised mammalian MHCs, namely the presence of class I and class II gene containing segments, embedded within framework gene regions. The MHCs of fish and birds display a much simple architecture and it is is unclear which evolutionary events lead the increased complexity observed in mammals. To close the gap, we identified, sequenced and annotated several BACs from the platypus MHC. Platypus and its twin species echidna are the only extant members of the ancient group of monotremes, which separated from placental mammals 180-210 Mio years ago. Monotremes are famous for their combination of reptilian and mammalian features, i.e. they lay eggs, but nurse their offspring with milk. Two BACs were characterised in detail. One BAC encompassed part of a core MHC with class I and class II genes, as well as genes required for antigen presentation. The other BAC contained a class I pseudogene in a genomic context reminiscent of RT1-CE or HLA B, C, respectively. However, the BACs mapped to different chromosomes, as determined by FISH mapping (collaboration with F. Grützner, Adelaide). Our conclusion is that the structure of the mammalian MHC is ancient and predates the split between monotremes and modern mammals. As a relatively recent event, the MHC of platypus was broken up, in the course of chromosomal rearrangements.

Medaka UniGene cDNA set and BAC map

The medaka (Oryzias latipes) is an ‚old" model system for genetic research with tradition going back almost 100 years. Medaka is closely related to Fugu and Tetraodon (divergence 60-80 Myr). Medaka and zebrafish are only distantly related. Medaka has advantages, e.g. perfectly inbred strains and a small 800 Mb genome (zebrafish: 1700 Mb). We have followed two approaches, first, the generation of a UniGene cDNA set and, second, the generation of a medaka BAC map. Libraries encompassing 119,000 cDNA clones were normalised by fingerprinting (OFP). 19,000 OFP clusters and singletons were sequenced (collaboration with N. Shimizu, Keio University). Subsequent EST clustering indicated > 10,000 unique entities. 3500 of our sequences were novel. For establishing a medaka BAC map, a data set was assembled from hybridising 35mer probes against arrayed BACs from three medaka strains. The integrated map contained ~ 3500 markers, most of them gene-derived, and > 11,000 BAC clones, hit at least twice. Our map was utilised for several positional cloning projects

(e.g. cloning of hirame, collaboration with M. Furutani-Seiki). Most importantly, our map served as backbone for the sequencing of chrom. LG22. Presently ongoing is the integration of BAC map data with whole genome shotgun scaffolds.

Sugar Beet (Beta vulgaris) mapping

The outline of the project follows the generation of the medaka BAC map. > 10,000 35mer probes from sugar beet expressed sequences were selected on the basis of conservation in sequenced plant genomes (Arabidopsis, poplar and rice) and hybridised against BAC resources (55.000 clones, 9-fold coverage of genome). The data set is presently utilised for the generation of a first-pass BAC map. In a pilot project, two BACs representing haplotypes from different cultivars were sequenced, to estimate the degree of intraspecific variation and the extent of synteny in comparison with other plant genomes (Fig. 2).

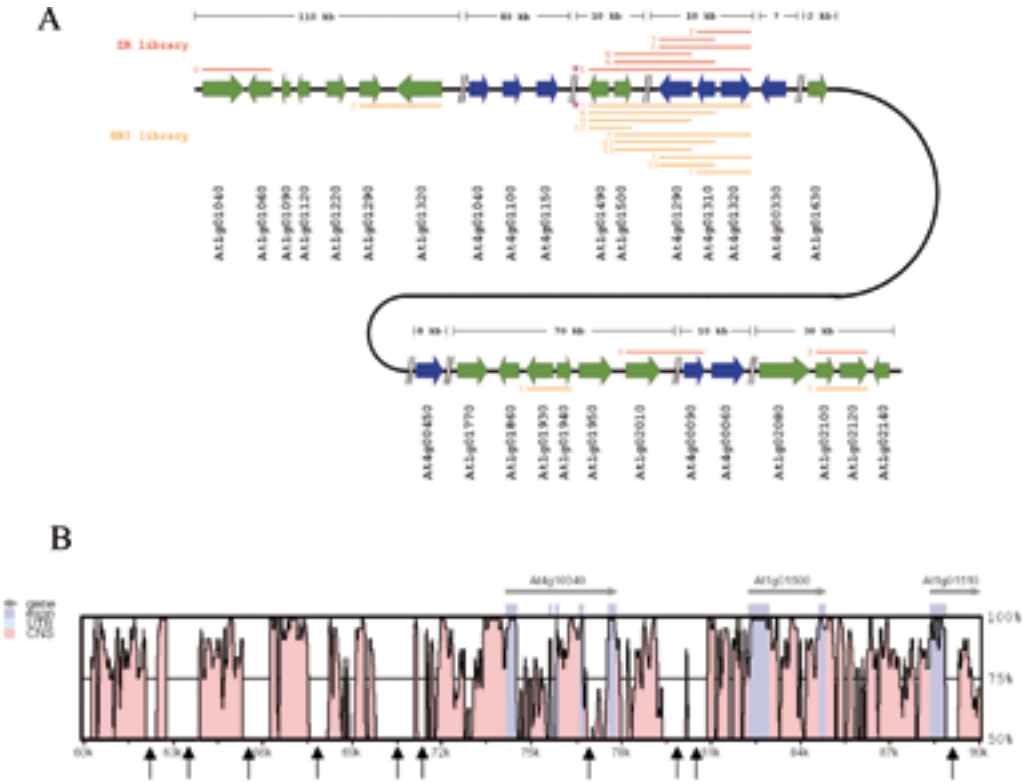


Figure 2: Microsynteny and intraspecific variation in sugar beet. A, Selected region of the Arabidopsis (At) genome. Gene order follows the ancestral gene order, predating the most recent genome duplication in At (see Blanc and Wolfe, 2004). Green: Genes on present-day At chrom 1. Blue: Genes on At chrom. 4. Sugar beet BAC clones (from two libraries, ZR and SBI) spanning more than one gene are indicated by lines. Asterisks indicate two BACs that were selected for genomic sequencing. B, Detail of Vista-alignment from sugar beet ZR and SBI BAC clone finished sequences. Average difference is 1% in coding and 3% in non-coding regions. Indels occur frequently (arrows).

Selected Information

Publications

- Wendeler M. W., Jung R., Himmelbauer H. and Geßner R. (2006) Unique gene structure and paralogy define 7D-cadherin family. CMLS 63, 1564-1573
- Traut W., Weichenhan D., Himmelbauer H., and Winking H. (2006). New members of the neurexin superfamily: multiple rodent homologues of the human CASPR5 gene. Mamm Genome 17, 723-731
- Greber B., Tandara H., Lehrach H. and Himmelbauer H. (2005). Comparison of PCR-based mutation detection methods and application for identification of mouse Sult1a1 mutant ES cell clones using pooled templates. Hum Mut 25, 483-490.
- Greber B., Lehrach H. and Himmelbauer H. (2005). Mouse splice mutant generation from embryonic stem cells – a gene driven approach. Genomics 85, 557-562.
- Henrich T., Ramialison M., Wittbrodt B., Assouline B., Bourrat F., Berger A., Himmelbauer H., Sasaki T., Shimizu N., Westerfield M., Kondoh H. and Wittbrodt.J. (2005). MEPD: A resource for medaka gene expression patterns. Bioinformatics 21, 3195-3197.
- Churchill G.A., *et al.* (2004). The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat. Genet. 36, 1133-1137.
- Zadeh Khorasani M., Hennig S., Imre G., Asakawa S., Palczewski S., Berger A., Hori H., Naruse K., Mitani H., Shima A., Lehrach H., Wittbrodt J., Kondoh H., Shimizu N. and H. Himmelbauer H. (2004). A first generation physical map of the medaka genome in BACs essential for positional cloning and clone-by-clone based genomic sequencing. Mech Dev 124, 903-913.
- Kittler R., Putz G., Pelletier L., Poser I., Heninger A.-K., Drechsel D., Fischer S., Konstantinova I., Habermann B., Grabner H., Yaspo M.-L., Himmelbauer, H., Korn B., Neugebauer K., Pisabarro M. T. and Buchholz F. (2004). A large scale esiRNA screen identifies genes essential for cell division. Nature 432,1036-1040.
- Zechner U., Shi W., Hemberger M., Himmelbauer H., Otto S., Orth A., Kalscheuer V., Fischer U., Elango R., Reis A., Vogel W., Ropers H., Rüschendorf F. and Fundele R. (2004). Divergent genetic and epigenetic post-zygotic isolation mechanisms in Mus and Peromyscus. J. Evol. Biol. 17, 453-460.
- Rat Genome Sequencing Project Consortium (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428, 493-521.
- Krzywinski M., *et al.* (2004). Integrated and sequence-ordered BAC and YAC-based physical maps for the rat genome. Genome Res 14, 766-779.
- Hurt P., Walter L., Sudbrak R., Klages S., Müller I., Shiina T., Inoko H., Lehrach H., Günther E., Reinhardt R. and Himmelbauer H. (2004). The genomic sequence and comparative analysis of the rat major histocompatibility complex. Genome Res 14, 631-639.
- Abiola O., *et al.* (2003). A community's view on the nature and identification of quantitative trait loci. Nature Reviews Genetics 4, 911-916.
- Greber B., Lehrach H. and Himmelbauer H. (2003). Characterization of trimethylpsoralen as mutagen for mouse embryonic stem cells. Mutation Research 525, 67-76.

PhD Thesis

- Maryam Zadeh Khorasani (2006). Physical mapping of medaka (Oryzias latipes) genome. Free University Berlin.
- Boris Greber (2004). Chemical mutagenesis of mouse embryonic stem cells. Free University Berlin.

Student Theses

- Christoph Campregher (2003). Chemical mutagenesis of mouse embryonic stem cells and development of a screening procedure for the detection of nonsense mutations. Faculty of Natural Sciences, University of Salzburg, Austria.
- Gabriele Hebenstreit (2003). Physical mapping of chromosome 22 in medaka. Faculty of Natural Sciences, University of Salzburg, Austria.

External Funding

- HFSP "Molecular und genetic analysis of lens-retina interactions in vertebrates" (until 2004)
- NGFN 1 "ENU mutagenesis of mouse ES cells" (until 2005)
- Bioprofile Nutrigenomik "Obesity candidate genes" and "Mouse models for xenobiotics metabolising enzymes"
- EU FP6 IP "Functional genomics in embryonic stem cells" (FunGenES)
- GABI 2 "GABI-Beet Physical Map"

Genetic Variation

Head
Dr. Margret Hoehe*
(Adj. Professor, The Rockefeller University, NYC)
Phone: +49 (0)30-8413 1468
Fax: +49 (0)30-8413 1462
Email: hoehe@molgen.mpg.de

Scientists
PD Dr. Werner Terhalle*
Dr. Eun-Kyung Suk*
Dr. Dmitri Parkhomchuk*
Dr. Thomas Kroslak*
Dipl. biol. Karolina Janitz*
Dipl. biol. Bernd Timmermann* (until 4/2004)

Technicians
Dipl. biochem. Annett Neubert*
Dipl. biol Sandy Vadakkadathu*
Pamela Kepper* (until 2/2006)
Mario Sontag* (until 6/2004)

*External project funding

Scientific Overview

Major focus is the systematic analysis of human genetic variation. The following aspects are being addressed:

The systematic analysis of genetic variation in candidate genes by application of high throughput (HT) medical resequencing and genotyping technologies

Based on longstanding lines of research, technology development and production, a substantial data set has been generated by in-depth resequencing, applying HT 'Multiplex PCR Sequencing' (up to 40 plex) and capillary sequencing technologies. More than 30 candidate gene loci of functional and/or positional relevance for major common diseases have been analyzed in an average of 333 (up to 1050) individuals, an unprecedented depth, up to 20 times deeper than previous resequencing studies in comparable populations (more than 35 Megabases of finished sequence). These data are currently being expanded by resequencing and large-scale genotyping of additional candidate genes with a focus on obesity. These variation data provide a test-bed, in conjunction with publicly available variation data, to address important aspects of genetic variation.

The application and development of bioinformatic tools to the analysis of the amount, nature and organization of genetic variation

The systematic analysis of genetic variation structure and its major components in genes and the genome provides an essential basis for the design of valid, efficient and powerful association studies, the identification of genetic risk patterns, and ultimately all molecular and functional analyses. Central theme is the analysis of the specific combinations of given sequence variants for each of the chromosomes defined as haplotypes; these may reflect units of organization structuring the human genome, or individually different forms of a gene. In this context, we have applied and developed an arsenal of methods, programs and resources that allow comprehensive analysis of both our high-resolution genetic variation data and publicly available data (*www.molgen.mpg.de/~genetic-variation/index.html*). Moreover, we have incorporated the information-theoretical concept of entropy (Shannon 1948) and developed programs that allow analysis of the information contained in variation data.

We have applied these programs and approaches to a comprehensive comparative evaluation of high-resolution genetic variation data providing essentially complete information on DNA sequence variation in genes or genomic regions of interest vs. the most advanced publicly available data base of common variation, the HapMap. Genetic variation structure in its major aspects was analysed at a total of 28 loci both under conditions of complete and incomplete information. Our data revealed substantial differences and demonstrate that the selection and use of subsets of SNPs without complete knowledge of variation may have significant implications on the power and validity of indirect approaches to disease gene discovery. Future successful strategies must rely on maximum (or at least sufficient) resolution of genetic variation by resequencing genes and genomes (Hoehe *et al.*, 2006, submitted).

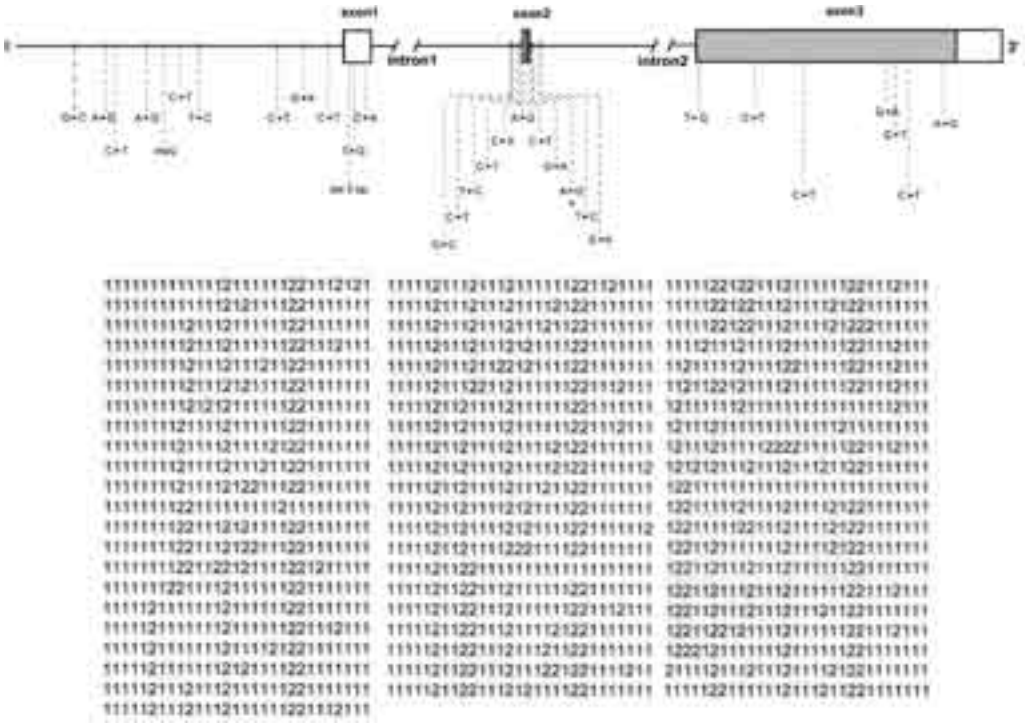


Figure 1: Example of a candidate gene region of interest; multiple haplotypes are inferred.

To achieve rapid and efficient data processing and analysis, the need for automation and integration has become evident. This spurred development of a software package that integrates all the distinct programs, both the publicly available ones and those developed by us to allow semi-automated modular analysis of genetic variation data. This software toolbox will be made available to the scientific community via a website.

Complementary investigations include a) the application of information theory to optimise SNP selection genome-wide to increase the probability to detect associations few-fold, b) evolutionary aspects related to variation.

The development and application of *in silico* approaches to the analysis of complex genotype/haplotype-phenotype relationships

The analysis of genetic variation at the ultimate level of resolution has provided evidence for abundant genome sequence diversity in genes and genomes. Numerous individually different forms of a gene/target region may exist. This allelic complexity poses tremendous challenges to the analysis of genotype-phenotype relationships (Hoehe *et al.*, 2003). A key question of increasing relevance is, how to filter the important variants from the unimportant ones. Approaches to reduce this complexity and condense information on genetic variation will be required. We have pursued successfully approaches to haploype classification based on sequence-structure-function similarity (Hoehe, 2003). We explore the decomposition of candidate genes into haplotype blocks as the basis for association studies (cooperation R. Shamir). We have shown the importance of considering spectra of rare mutations (Branson *et al.*, 2003). Experimental and informatic approaches are being undertaken to group mutations, haplotypes and genes with reference to common (functional) output signals, e.g. triggered by pathways. We have developed a heuristic approach based on Boolean Algebra to examine combinations of variants (patterns) in genotypes or haplotypes. First applications have identified patterns of SNPs that are significantly associated with obesity, each SNP on its own not showing any association. Moreover, our obesity database including variation data on currently 22 genes as well as extensive phenotypic data is being subjected to the analysis of haplotype-haplotype interactions (cooperation J. Ott).

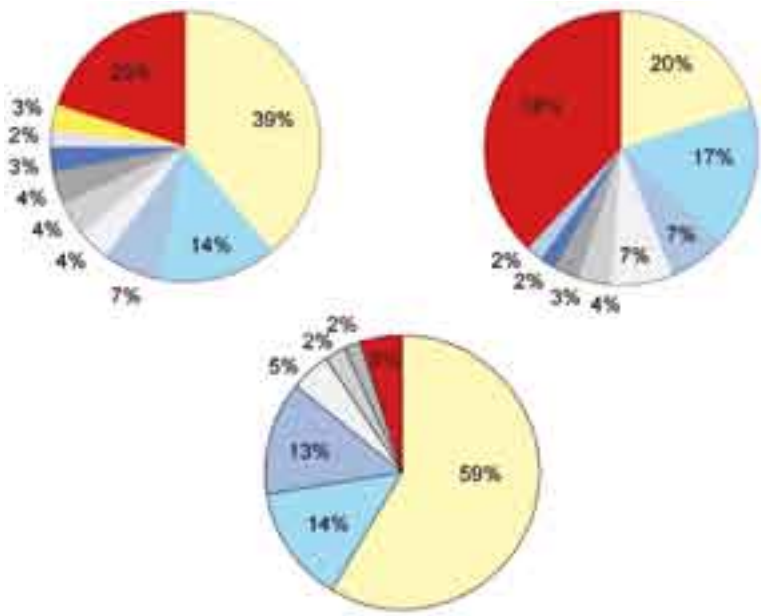


Figure 2: Distribution of haplotype frequencies illustrated for three candidate genes. Each color-coded segment represents proportionately the frequency (in percent) of one specific haplotype in a population, with the exception of the red segment, that represents the total of all rare haplotypes with a frequeny < 1%. For a fraction of genes, there exists no major predominant gene form.

Molecular genetic approaches to haplotype analysis

The validation/complementation of *in silico* analyses by determination of the naturally existing, molecular haplotypes still represents a major challenge. We are currently establishing a reference resource (RR) of haploid clone pools from up to 250 individuals (500 haploid genomes) from a representative German population (POPGEN, collaboration S. Schreiber) and selected disease samples (KORA, collaboration H-E. Wichmann). This RR allows genome-wide determination of molecular haplotype structures of any gene or chromosomal region of interest. The technology, clone-based systematic haplotyping (CSH), has been developed by Burgtorf *et al.* (2003) at the MPI-MG and relies on the generation of pooled fosmid/cosmid libraries and subsequent SNP typing of PCR products. CSH is currently optimised to establish a 'semi-HT' technology platform. The RR will prepare the ground to 1) validate haplotype structures, in particular in the case of potential disease genes or genes of relevance for pharmacogenomics; 2) generate a molecular reference system for optimisation of algorithms; 3) conduct first functional analyses of molecular haplotypes. This will ultimately provide the necessary molecular basis for disease gene characterization, pharmacogenomics and an individualized medicine.

Our genetic variation data have shown that the Mendelian concept of a 'wild type' may not apply to a substantial fraction of genes. This requires novel approaches to evaluate the impact of genetic variation on function and phenotype, the units of analyses no longer being single mutations, but combinations of variants, as they characterize promoter and coding haplotypes. The RR, representing a 'molecular catalogue of all haplotypes of all genes', will serve as a key resource to facilitate proof-of-principle experiments. First analyses of promoter haplotype diversity will be conducted in cell-based assays and HT functional testing systems (collaboration M. Janitz, MPI-MG).

Perspectives

Development of approaches and tools to analyze DNA sequence diversity and its implications on gene function and phenotype; in this context, development of functional approaches to haplotype analysis as a contribution to 'individual functional genomics' and an individualized medicine.

Selected Information

Publications

- Zhang J, Vingron M, Hoehe MR (2005) Haplotype reconstruction for diploid populations. Hum Hered 59: 144-156.
- Hoehe MR, Kroslak T (2004) Genetic variation and pharmacogenomics: Concepts, facts, and challenges. Dialogues Clin Neurosci 6: 5-25 (Lead Article).
- Patkar AA, Berrettini WH, Mannelli P, Gopalakrishnan R, Hoehe MR, Bilal L, Weinstein S, Vergare MJ (2004) Relationship between serotonin transporter gene polymorphisms and platelet serotonin transporter sites among African-American cocaine-dependent individuals and healthy volunteers. Psychiat Genet 14: 25-32.
- Potoczna N, Branson R, Kral JG, Piec G, Steffen R, Ricklin T, Hoehe MR, Lentes K-U, Horber FF (2004) Gene variants and binge eating as predictors of comorbidity and outcome of treatment in severe obesity. J Gastrointest Surg 8: 971-982.
- Branson R, Potoczna N, Kral JG, Lentes K-U, Hoehe MR, Horber FF (2003) Binge eating is a major phenotypic characteristic of melanocortin-4 receptor gene mutations. New Engl J Med 348: 1096-1103.
- Burgtorf C, Kepper P, Hoehe M, Schmitt C, Reinhardt R, Lehrach H, Sauer S (2003) Clone-based Systematic Haplotyping (CSH) – a procedure for physical haplotyping of whole genomes. Genome Res 13: 2717-2724.
- Franke P, Wendel B, Knapp M, Schwab SG, Neef D, Maier W, Wildenauer DB, Hoehe MR (2003) Introducing a new recruitment approach for genetic association studies in opioid dependence. Eur Psychiat 18: 18-22.
- Hoehe MR (2003) Haplotypes and the systematic analysis of genetic variation in genes and genomes. Pharmacogenomics 4: 547-570 (Invited Review, Paper of the Month).
- Hoehe MR (2003) Individuelle Genomanalyse als Basis neuer Therapiekonzepte. (Individual genome analysis as the basis for novel therapeutic concepts.) In: Das genetische Wissen und die Zukunft des Menschen. De Gruyter, Berlin, New York, p. 301-317 (Invitation by the BMBF).
- Hoehe MR, Timmermann B, Lehrach H (2003) Human inter-individual DNA sequence variation in candidate genes, drug targets, the importance of haplotypes and implications for pharmacogenomics. Curr Pharm Biotechnol 4: 351-378 (Invited Article, selected as Lead Article among 40 articles).
- Wenzel K, Felix SB, Flachmeier C, Heere P, Schulze W, Grunewald I, Pankow H, Gläser C, Scherneck S, Bauer D and Hoehe MR (2003): Identification and characterization of KAT, a novel gene preferentially expressed in several human cancer cell lines. Biol Chem 384: 763-775.

Appointments, honors

- Appointment as Adjunct Professor at The Rockefeller University, NYC, effective June 1st, 2005
- ranked among the 50 most cited researchers in Germany, 2000-2002, in human genetics by the Laborjournal.

Public relations

- Research results reported in the New York Times, Wall Street Journal and all other major US media, 2003

Organization of scientific events, work as scientific referee

- Chairperson and organizer of the Weimar Conference of the German Society of Genetics on 'Genetic Variation in Man'
- reviewer of grants for major national and international funding agencies
- reviewer for international journals

Cooperations

- J. Ott, The Rockefeller University, NYC
- G. M. Church, Harvard Medical School, Boston
- R. Shamir, Tel Aviv University, Tel Aviv
- K. Kidd, H. Zhao, Yale University, New Haven
- J. Reich, K. Rohde, Max-Delbrück-Center, Berlin
- S. Schreiber, Christian-Albrechts-University, Kiel
- A. F. H. Pfeiffer, DIfE & Charité, Berlin
- F. F. Horber, Clinic Hirslanden, Zürich
- H. E. Wichmann, GSF, Neuherberg
- R. Wambutt, AGOWA, Berlin

External funding

- BMBF/NGFN1 Core Comparative Candidate Gene Sequencing, Haplotype Analysis and Genetic Risk Profile Identification.
- BMBF BioProfile Potsdam Berlin: Verbundvorhaben 'Innovation des Therapiekonzeptes für das Metabolische Syndrom' – Teilprojekt: Haplotype analysis.
- BMBF/NGFN2: Haplotype approaches to disease gene discovery: A systematic investigation and establishment of reference resources.
- BMBF BioProfile Potsdam Berlin: Genetics and Pharmacogenomics of Obesity.
- German-Israeli Foundation (GIF): Haplotyping and Association Algorithms and their Applications to Model Disease Genes.
- GlaxoSmithKline Award: Analysis of high-resolution genetic variation data with particular emphasis on haplotype structures and LD patterns.

Cell Arrays/Oligofingerprinting

Head
Dr. Michal Janitz*
Phone: +49 (0)30-8413 1463
Fax: +49 (0)30-8413 1380
Email: janitz@molgen.mpg.de

Scientists
Dr. Anna Guerasimova*
Dr. Dominique Vanhecke*

Graduate Students
Andrea Fiebitz
Xi Cheng*
Jianping Liu
Yuhui Hu

Technicians
Irina Girnus*
Sabine Thamm

* external project funding

Overview of the scientific work

The main research focus of the group comprises high-throughput functional genomics and genome-wide expression analysis using transfected-cell array (TCA) and oligonucleotide fingerprinting (ONF) technologies, respectively.

Cell array-based functional genomics

In the field of functional genomics we optimised and further developed a cell array platform based on the reverse transfection process. Briefly, full-length open reading frames of genes inserted in expression vectors are printed at a high density on a glass slide along with a lipid transfection reagent using a robotic arrayer. Densities of up to 8000 cell clusters per standard slide can be achieved. When the microarray of DNA constructs are covered with a layer of adherent cells only the cells growing on top of the DNA spots became transfected, resulting in the expression of specific proteins in spatially distinctive groups of cells. The phenotypic effects of this 'reverse transfection' of hundreds or thousands of genes can be detected using specific cell-based bioassays, e.g. immunofluorescence or induction of apoptosis. Our current TCA-based research projects comprise following topics:

Intracellular protein localization

Taking advantage of transfection of hundreds different cDNAs in parallel using the transfection array we are able to determine localisation of the expressed proteins in a high-throughput manner using a microscope (Fig. 1). By evaluation of the transfected cells forming a single spot we can determine whether the protein expressed from the transfected cDNA localises in the nucleus, cytoplasm or is transported to the cellular membrane. We recently accomplished screening of human chromosome 21 proteins using the cell array platform. For 34 of these proteins, their localization is described for the first time. Furthermore, the alteration in cell morphology and growth as a result of protein over-expression for claudin-8 and claudin-14 genes has been characterized.

CAPPIA – Cell-Array-Protein-Protein-Interaction-Assay

We established a new and cost-effective method for the high-throughput detection of protein-protein interactions in mammalian cells that combines the advantages of mammalian two-hybrid systems with those of microarrays. In this assay, mixtures of bait and prey expression plasmids together with an auto-fluorescent reporter are immobilized on glass slides in defined array formats. Adherent cells that grow on top of the microarray will become fluorescent only if the expressed proteins interact and subsequently trans-activate the reporter. Using known interacting partners and by screening of a small prey library against the human androgen receptor ligand-binding domain we demonstrated that this assay allows the quantitative detection of protein interactions in different types of mammalian cells and under influence of different compounds. The high capacity of preys that can be tested per slide together with the flexibility to interrogate any bait of interest and the small amounts of reagents that are required makes this assay currently the most economical high-throughput detection assay for protein-protein interactions in mammalian cells.

RITA – RNA Interference Transfection Array.

In this project we developed a high-throughput approach, an RNA Interference Transfection Array (RITA), for loss-of-function studies in mammalian cells. RITA combines reverse transfection array technique with recently developed RNA interference technology based on a usage of short interfering RNA for transient or stable gene silencing (Fig.2). We optimized the RITA technique for a number of cell lines representing different physiological compartments of the organism, in which genes comprising for example entire signal transduction pathways can be silenced in a single experiment. Thus, RITA represents a powerful alternative tool to *in vivo* knockout studies allowing for substantial reduction of experimental animal usage. Recently we established the cell array-based apoptosis detection assay for evaluation of anti-apoptotic genes whose silencing by means of RNAi results in induction of apoptosis in particular cell clusters. Moreover we developed an assay for multiplex gene silencing in an array format which can be applied for detection of synthetic lethal interactions in a high-throughput manner.

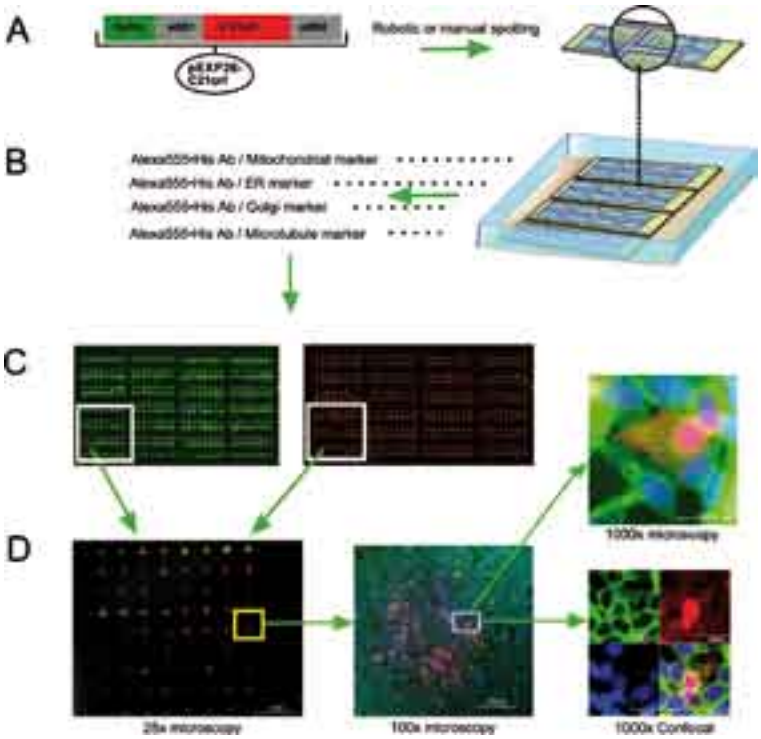


Figure 1: Schematic presentation of the cell array-based protein localization procedure. (A) Open reading frames (ORFs) derived from human chromosome 21 were cloned into a mammalian expression vector containing an amino-terminal His6 tag, spotted in an array format and reverse transfected into an HEK293T cell line. (B) Proteins expressed on the cell arrays were detected using anti-His antibody and a set of organelle counterstaining was performed. (C) The efficiency of protein expression and labeling was monitored using a BioCCD laser scanning system. The green signal indicates the expression of enhanced green fluorescent protein (EGFP). The red signal indicates His6-tagged protein. (D) For single-cell analysis, fluorescent and confocal microscopy was used. The red signal in the 1000x magnification images represents expressed C21orf25 protein, whereas the green and blue signals represent labeled microtubule and nucleus, respectively.

Functional promoter analysis

Accomplishment of the human and mouse genome projects resulted in identification of large number of new genes. Information derived from sequence analysis provided insights into the structure of the genes whereas gene expression analysis provided a lot of hints about gene activities on RNA message level. As a consequence of these advances another challenge is the deciphering of the network of regulatory mechanisms inducing genes activity. We use cell arrays to evaluate in a high throughput manner activity of non-coding regions of the genome. In applied strategy thousands upstream sequences of known genes will be evaluated for transcription activation (i.e. promoter activity) of the reporter genes in different cell lines, in order to assess the mechanisms of regulation of the respective genes *in vivo*. The promoter candidates were selected according to *in silico* predictions based on evolutionary conserved non-coding regions upstream to orthologous in mouse and human genes. These regions were cloned into the plasmid constructs upstream of fluorescent reporter. The level of reporter gene expression is normalised to the co-transfected reference reporter. The proposed strategy allows to evaluate influence of sequence polymorphisms in promoters regions on the transcription inducing activity and hence their implication in disease mechanisms. Combined with siRNA methods the role of particular transcription factors in the gene regulation can also be investigated

Oligonucleotide fingerprinting

We took an effort to increase the robustness and cost-effectiveness of the ONF approach. In cooperation with the group of Dr. Nyarsik we developed an assay for the fluorescent detection of short oligonucleotide probe hybridization in miniaturized high-density array platforms. It combines hybridization in solution with real-time fluorescent detection, which involves measurement of fluorescence increase by means of an induced fluorescence resonance energy transfer. The feasibility of this approach using DNA or RNA as a target, and short DNA- as well as LNA (locked nucleic acid)-modified oligonucleotides as probes has been shown. This approach could potentially contribute to a significant increase in the throughput of large-scale genomic applications, such as oligofingerprinting and genotyping, and also reduce material consumption.
Furthermore, a robust method for the detection of hybridization events using a microarray-based assay on a nanoporous membrane platform was developed. The technique is characterized by a hybridization time of only 1 hour and uses Cy5-labeled, 7-mer oligonucleotide probes modified with locked nucleic acid (LNA) bases. We could show that the volume of the DNA spotted onto a nanomembrane can be reduced to ~4 nl with detectable signal intensity. Moreover, the amount of the DNA target could be reduced to 4 fmol. The described approach could dramatically increase the throughput of techniques based on sequencing by hybridization, such as oligofingerprinting, by decreasing the total number of probes that are needed for analysis of large clone sets and reduction of the sample/reagent consumption. The method is particularly advantageous when numerous hybridization-based assays must be performed for characterization of sample sets of 100,000 or more.

Perspectives

Future research will be focused on establishment of the cell array for primary cells based on the lentiviral transduction towards analysis of promoter activities within different mammalian tissues. We will also further develop the siRNA multiplexing towards discovery of the new signal tranduction and apoptosis induction pathways using relatively small siRNA libraries of several hundred genes spotted simultaneously in different combinations and reverse transfected to variety of cell types.

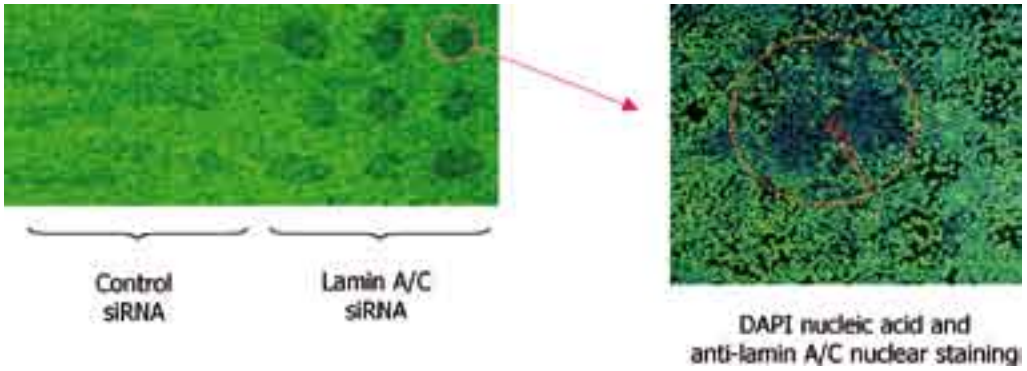


Figure 2: Downregulation of lamin A/C gene expression using the cell array platform. Silencing of the lamin A/C gene in cells covering siRNA spots can be visualized with an array scanner (left) as black holes in the monolayer of cells stained with anti-lamin A/C FITC-labeled antibody. Red circle indicates a single reverse transfected cell cluster shown with 40x magnification using Zeiss Axiocam microscope (right). Cellular nuclei were counterstained with DAPI.

Selected Information

Publications

- Hu Y-H, Warnatz H-J, Vanhecke D, Wagner F, Fiebitz A, Thamm S, Kahlem P, Lehrach H, Yaspo M-L, Janitz M (2006) Cell array-based intracellular localization screening reveals novel functional features of human chromosome 21 proteins. BMC Genomics, in press.
- Janitz M, Vanhecke D, Lehrach H. (2006) High-throughput RNA interference in functional genomics. Handb Exp Pharmacol 173:97-104.
- Guerasimova A, Nyarsik L, Liu JP, Schwartz R, Lange M, Lehrach H, Janitz M. (2006) Liquid-based hybridization assay with real-time detection in miniaturized array platforms. Biomol Eng 23:35-40.
- Hu YH, Vanhecke D, Lehrach H, Janitz M. (2005) High-throughput subcellular protein localization using cell arrays. Biochem Soc Trans 33:1407-1408.
- Gutjahr C, Murphy D, Lueking A, Koenig A, Janitz M, O'Brien J, Korn B, Horn S, Lehrach H, Cahill DJ. (2005) Mouse protein arrays from a TH1 cell cDNA library for antibody screening and serum profiling. Genomics 85:285-296.
- Vanhecke D, Janitz M (2005) Functional genomics using high-throughput RNA interference. Drug Discov Today 10:205-212.
- Illiger J, Herwig R, Steinfath M, Przewieslik T, Elge T, Bull Ch, Radelof U, Lehrach H, Janitz M (2004) Establishment of T cell-specific and natural killer cell-specific unigene sets: towards high-throughput genomics of leukaemia. Eur J Immunogenet 31:253-257.
- Vanhecke D, Janitz M (2004) High-throughput gene silencing using cell arrays. Oncogene 23:8353-8358.
- Bauer O, Guerasimova A, Sauer S, Thamm S, Steinfath M, Herwig R, Janitz M, Lehrach H, Radelof U. (2004) Multiplexed hybridizations of positively charge-tagged peptide nucleic acids detected by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 18:1821-1829.

Academic cooperations

- Andreas Radbruch, Deutsches Rheumaforschungszentrum Berlin, Germany
- Gerd Burmester, Rheumatology Clinic, Charité, Humboldt University Berlin, Germany
- John Williams, Roslin Institute, UK
- Jörn Koch, Aarhus University, Denmark
- Anthony Brookes, University of Leicester, UK
- Dolores Cahill, Royal College of Surgeons, Ireland
- Olli Kallioniemi, VTT Technical Research Centre, Finland

Industrial cooperations

- Cooperation with Qiagen for development of the RNA Interference Transfected-Cell Array

Organization of scientific events

- EU-funded (Integrated Project MolTools) training course on transfected-cell array technology, March 2005

External Funding

- BMBF Gene expression profiling (until 2004)
- BMBF Oligonucleotid Fingerprinting (until 2004)
- EU: BOVGEN (until 2005), MOLTOOLS,
- BMBF Cell transfection arrays
- NGFN II EP None Messenger RNA
- KWS Oligofingerprinting

Kinetic Modeling

Head
Dr. Edda Klipp*
Phone: +49 (0)30-804093 16
Fax: +49 (0)30-804093 22
Email: klipp@molgen.mpg.de

Scientists
Dr. Axel Kowald*
Dr. Jörg Schaber* (since 6/2004)
Dr. Wolfram Liebermeister* (since 3/2004)

Graduate Students
Wolfram Liebermeister (until 4/2003)
Simon Borger* (since 1/2004)
Marija Cvijovic* (since 6/2005)
Zhike Zi* (since 8/2005)

Undergraduate Students
Bente Kofahl
Sebastian Schmeier*
Susanne Gerber
Christof Dehmel
Stephan Menz*
Anselm Helbig*
Judith Wodke*
Marvin Schulz*
Jannis Uhlendorf*

*external project funding

Scientific Overview

Focus of the Kinetic Modeling Group is the development of dynamic models of biochemical reaction networks for selected cellular processes. These models shall represent the available experimental knowledge and have a high explanatory and predictive value. Their development is based on close collaboration with experimental groups. Considered processes involve signaling pathways for stress response, gene expression regulation, metabolism and cell cycle. The main model organism is the yeast *Saccharomyces cerevisiae*, but there are also models under development for *E. coli*, mouse and human cells, and for sea urchin.

The main problems in kinetic modeling are the uncertainty of information and the lack of specific data. Therefore, we apply and combine three strategies: (i) information retrieval and integration, (ii) model handling, and (iii) actual model development.

Information Retrieval and Integration

Kinetic models deserve kinetic data to allow simulations and quantitative predictions. Since much biological information gathered in the past is available in the overwhelming scientific literature only, we developed a text-mining approach for kinetic data to perform advanced, domain specific search queries on text. We analyzed 15 million PubMed abstracts for ten different kinetic parameters, such as Michaelis-Menten constants or maximal enzyme velocities and extracted the numerical values and units together with the respective reactions, enzymes, or compounds. The extracted data are stored in our database KMedDB, freely accessible through a web interface, and provide a useful resource for building large-scale mathematical models in systems biology.
We develop a scheme for the integration of kinetic information from heterogeneous sources that allows the assignment of parameters to kinetic models. To this end, we combine data from various databases for different species, enzymes and compounds. The integration is performed by an ANOVA method. Assignment of kinetics to individual reactions based on newly developed convenience kinetics takes advantage of Bayesian approaches.
Information about the structure of the biochemical networks is obtained in the classical way by literature search and communication with experienced experimentalists.

Model Handling – Strategies and Tools

To making use of the experimental data from literature or project partners (qualitative information, but also time-courses of protein levels, mRNA abundance, and metabolite concentrations) necessitates development and implementation of tools and strategies to estimate network parameters and network structure from available data.

Parameter estimation: For estimating parameters from time-course data together with the appropriate network structure for signaling pathway models, we use an evolutionary strategy. In addition, we developed the tool SBML-PET that estimates parameters for models in SBML format using any kind of experimental data and supporting experiment-related events.

Handling uncertain parameters: Modeling of biochemical networks becomes delicate if kinetic parameters are varying, uncertain or unknown. Facing this situation, we quantified uncertain knowledge or beliefs about parameters by probability distributions and infered probabilistic statements about dynamic network properties, such as steady-state fluxes and concentrations, signal characteristics or control coefficients. We proposed a graphical scheme, the 'dependence graph', to bring out known dependencies between parameters, e.g. due to equilibrium constants. For a narrow parameter distribution the resulting variable distribution can be computed by expansion around a set of mean parameter values. We computed distributions of concentrations, fluxes and probabilities for qualitative variables such as flux directions.

Model reduction: For the case that a comprehensive model is given, but recent focus is only on a subsystem, we developed a model reduction approach based on balanced truncation, such that the system in focus remains and the dynamics of the reduced system still resemble the dynamics of the original system.

Model integration: More and more individual models are developed for different cellular pathways. These models are based on sound considerations and compared with experiments. To arrive at a comprehensive view on cellular networks it becomes necessary to integrate such models in an intelligent way. We developed a scheme for integration of models based on their description in Systems Biology Markup Language (SBML). The pipeline involves a check on the validity of the SBML model, a guide for the annotation of model components and finally a guide to actually merge the respective models.

Model analysis: We apply standard methods to test quality and predictions of our models, like parameter sensitivity and control analysis. Since the quantitative and qualitative information about the structure and dynamics of signaling pathways is often not sufficient to construct a unique model according to a set of criteria, we tested various measures for the quality of a model that takes the number of parameters and the goodness of fit into account. This was applied to the integration of two branches of the HOG pathway (see below). Signaling pathways are not sharply separated, but use common components and may exchange information. We newly develop a measure to assess the crosstalk between pathways upon different stimuli and to quantify signal specificity and signal integration.

The webpage *http://sysbio.molgen.mpg.de/* provides links to the tools developed in our group:
Kinetikon – Database for kinetic parameters for specific enzymes, compounds, and reactions
KMedDB – Search Medline abstracts for key words related to enzyme kinetic parameters.
SBMLmerge – Computer-assisted combination of biochemical models in SBML format.
Metabolic network toolbox – Collection of MATLAB files for modeling, simulation, analysis, and visualization of biochemical networks.
SBML-PET – Systems Biology Markup Language (SBML) based Parameter Estimation Tool.

Actual Model Development

While many groups try to decipher cellular stress response with global approaches such as microarray analyses to record the transcriptional response of all genes upon a stimulus or yeast-2-hybrid methods, we focus on small- and medium scale modeling to understand how cells process the response to environmental changes. Such models are usually sets of ordinary differential equations involving kinetic expressions for every individual reaction.

For yeast as model organism, we developed in close collaboration with experimental groups computational models for the following signaling and stress response processes:

- stimulation and regulation of the pheromone pathway activity
- crosstalk between pheromone pathway and filamentous growth pathway,
- osmotic stress response including the activation of the HOG pathway, regulation of gene expression, alteration of metabolism and glycerol production, and regulation of volume, turgor pressure and osmotic pressure,
- the TOR pathway (under development),
- the cell cycle (with focus on G1 phase and transition to S phase)
- the processing of information from the various signaling pathways by the cell cycle machinery

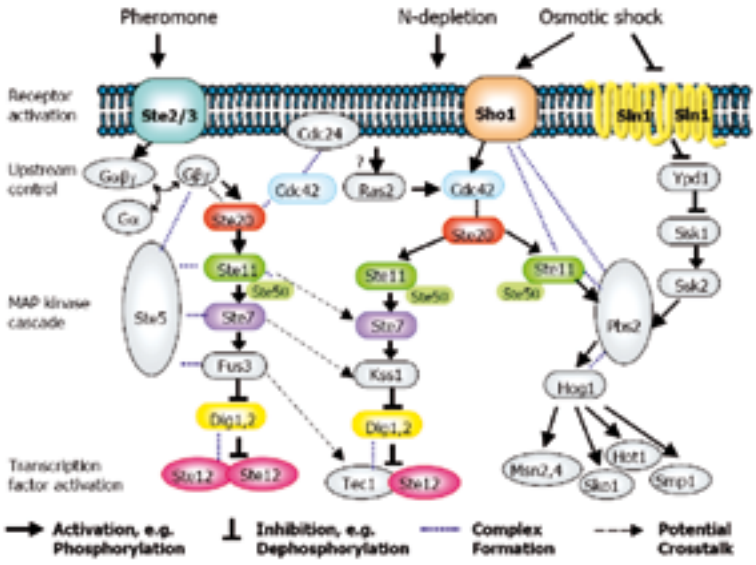


Figure 1: Yeast Signaling Pathways

With respect to mammalian cells, we developed a model that can explain the counterintuitive toxicity of superoxide dismutase (SOD), which is one of the most important antioxidant enzymes and which catalyzes the dismutation of superoxide radicals to peroxide.

Contribution to standards

There is need for exchange, reusability and connection of biochemical pathway models. This demands for standards in the formulation of models, in their actual encoding and in the information accompanying the model publication. We use and support the Systems Biology Markup Language (SBML) for model encoding, we contributed to the MIRIAM initiative that formulated standards for model annotation and publication, and we develop tools that enable to check whether models are SBML-compliant. Currently, we perform a survey on model standards and requirements within the Systems Biology scientific community.

Selected Information

Publications

- Klipp E & Liebermeister W. (2006). Mathematical modeling of intracellular signaling pathways. BMC Neuroinformatics. In press
- Schaber J, Kofahl B, Kowald A & Klipp E. (2006). A modelling approach to quantify dynamic crosstalk between the pheromone and the starvation pathway in baker's yeast. FEBS Journal. In press.
- Klipp E & Schaber J. (2006). Modelling of Signal Transduction in Yeast – Sensitivity and Model Analysis. In "Understanding and Exploiting Systems Biology in Bioprocesses and Biomedicine" (M. Canovas, ed.). In press.
- Liebermeister, W. & Klipp, E. (2005). Biochemical networks with uncertain parameters. IEE Systems Biology, 152 (3), 97-107.
- Schaber J, *et al.* (2005). Putative gene expression levels influence amino acid usage and evolutionary rates in endosymbiontic bacteria. Gene 352:109-117.
- Klipp, E, Nordlander, B, Krüger, R, Gennemark, P & Hohmann, S. (2005). The dynamic response of yeast cells to osmotic shock – a systems biology approach. Nature Biotechnology, 23, 975-982.
- Liebermeister, W, Baur, U, and Klipp, E. 2005. Complexity reduction of biochemical network models by balanced truncation. FEBS Journal, 272(16):4034-43.
- Klipp, E., Nordlander, B., Kofahl, B., and Hohmann, S. (2004). Shutting the MAP off – and on again? Current Genomics, 5, 637-647.
- Liebermeister, W., Klipp, E., Schuster, S., and Heinrich, R. (2004). A theory of optimal differential gene expression. BioSystems, 76, 261-278
- Kowald A., and Klipp E. 2004. Alternative pathways might mediate toxicity of high concentrations of superoxide dismutase. Annals of the New York Academy of Sciences,1019, 370-374.
- Klipp E., Liebermeister W., and Wierling C. (2004). Inferring Dynamic Properties of Biochemical Reaction Networks from Structural Knowledge. Genome Informatics 15(1): 125-137.
- Klipp E, (2003). Mathematical modeling of signal transduction in yeast. YEAST 20: S286-S286 Suppl.

Textbook

- Klipp, E., Herwig, R., Kowald, A., Wierling, C. and Lehrach, H. 2005. Systems Biology in Practice: Concepts, Implementation and Application. Wiley-VCH, Weinheim. ISBN 3-527-31078-9

Selected invited international plenary lectures

- Klipp, E. Computational Modeling of Yeast Cell Stress Response, First International Symposium on Systems Biology – From genomes to *In silico* and back, June 1-2, 2006, Murcia, Spain.
- Kowald E. Introduction to Systems Biology: Experiments, Models & Examples. Lisbon, Portugal, May 2006, "Advanced Workshop on Regulatory Networks".
- Klipp, E. Dynamic Modeling of Yeast Cell Stress Response. Keynote talk at the First Annual RECOMB Satellite Workshop on Systems Biology, December 2nd-4th, 2005, University of California, San Diego, U.S.A.
- Kowald A. On the relevance of mitochondrial fusions for the accumulation of mitochondrial deletion mutants. Newcastle, England, June 2005. Biogerontology Research Symposium
- Klipp, E. Modeling of signal transduction using different mathematical approaches. Workshop on Mathematical Aspects of Systems Biology, November 13-15, 2003, Göteborg, Schweden.

Prizes /Appointments /Scientific honors

- Kowald A. Habilitation (teaching certificate) at the Humboldt-University of Berlin. June 2005, "Fundamental aspects of Aging"; Title of the presentation: "The origin of the genetic code"

PhD Theses

- Wolfram Liebermeister (co-advisor, with Prof. R. Heinrich, Humboldt University) "Optimal Regulation of Gene Expression after Small Perturbations" Promotion HU Berlin, 2004.
- Sandra Pohl "Experimentelle Prüfung mitochondrienbasierter Alternstheorien" PhD thesis, Goethe University of Frankfurt, 2005.

External Funding

- BMFF NGFN 1 BCB (until 06/06)
- EU: QUASI, YSBN, System Biology, ENFIN

In Vitro Ligand Screening

Head
Dr. Zoltán Konthur*
Phone: +49 (0)30-8413-1586
Fax: +49 (0)30-8413-1365
Email: konthur@molgen.mpg.de

Scientists
Dr. Jörn Glökler (until 6/2004)
Dr. Jürgen Kreutzberger
Dr. Josmar Langner* (since 2/2005)
Dr. Volker Sievert* (since 10/2005)
Dr. Axel Wettich (10/2003 - 12/2003)

Engineer
Jeannine Wilde* (Dipl Ing. (FH))

Technicians
Annette Poch-Hasnek* (since 7/2005)
Carola Stoschek

Graduate Students
Gisela Burguera-Hoek* (B.Sc.; 5/2003 - 4/2004)
Chenna Reddy Galiveti* (M.Sc.; since 11/2004)

Undergraduate Students
Yvonne Baermann (8/2004 - 10/2004)
Jörn Hanusch* (10/2005 - 7/2006)
Alexander Sternjak (until 1/2003)
Franziska Wegerich (since 3/2006)

* external project funding

Scientific Overview

The major interest of the group centres around the application of *in vitro* screening techniques for the discovery of protein-protein interaction pairs. These include the use of phage display libraries, nucleic acid aptamer technology and protein array screening methods. In recent years, we have been able to develop semi-automated protocols for the *in vitro* generation of ligand binders (recombinant antibodies) and have successfully merged selection techniques such as phage display with robotics. Additional interest of the group is the use of *in vitro* screening techniques for the discovery of biologically relevant and naturally occurring interactions, ranging from antibody-protein, protein-protein to RNA-protein interactions.

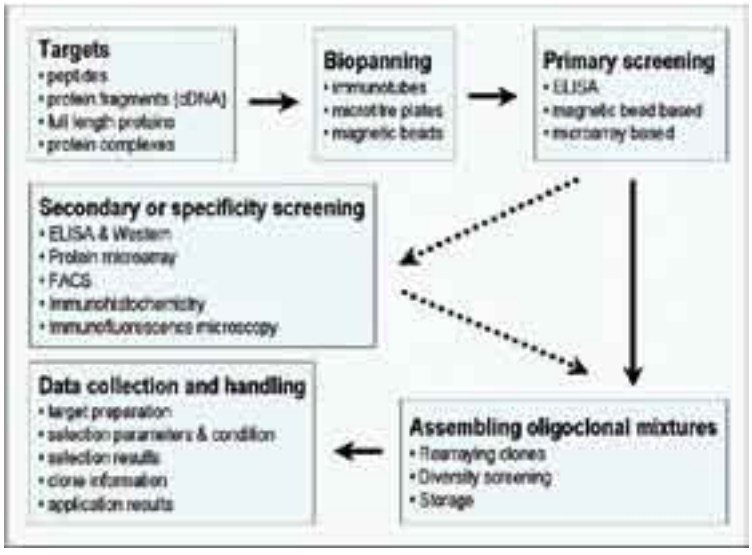


Figure 1: Schematic representation of the antibody selection pipeline. Due to its open architecture both, the individual modules and the pipeline itself can easily be modified.

Projects

Generation of phage display-derived antibodies

Within the National Genome Research Network (NGFN), the overall aim of the SMP "Antibody Factory" is the development of technological tools for the generation of recombinant antibody molecules by phage display in a high-throughput and systematic fashion against human gene products. Our subproject is concerned with the semi-automation of panning and evaluation procedures and is dedicated to the assembly of a selection pipeline resulting in oligoclonal mixtures of phage display derived antibody fragments for proteomic research applications. The pipeline (Figure 1) is concerned with all necessary steps in antibody generation starting from antigen cloning and expression, the selection procedure, oligo- and monoclonal binder screening and evaluation, as well as data handling and experiment documentation. Some examples of our efforts are e.g. the inves-

tigation of *in vitro* transcription/translation coupled protein expression for target generation (in collaboration with RiNA GmbH), protein microarray based screening methods for fast determination of target-specificity applying the Multiple Spotting Technique, and the development of a laboratory information management system (LIMS) to store all our relevant experimental data. The current list of candidate proteins applicable for antibody selection contains ~1200 proteins, of which many result from inter-departmental collaborations (Marie-Laure Yaspo, Sylvia Krobitsch, Konrad Büssow).

Generation of binder molecules with inhibitory properties

This project is exploring the possibility to select binder molecules, which allow specific blocking of protein-protein interactions. The work is carried out in close collaboration with the group of Dr. Sylvia Krobitsch (Neurodegenerative disorders). We are developing a method for the selection of inhibitory ssRNA aptamers / intrabodies using a combined *in vitro* / *in vivo* approach. First, ssRNA aptamer or phage display libraries are enriched *in vitro* on immobilised target proteins and once the diversity of the libraries are reduced to a complexity amenable to reverse yeast-2-hybrid screens, the selection is continued *in vivo*, where the selection is carried out with interaction pairs of the target proteins. Resulting binders should not only bind their respective target proteins, but also inhibit defined protein-protein interaction of the targets. The method is established using protein-protein interaction pairs involved in neurodegenerative disorders as well as the Wnt, FGF and Notch-Delta signaling pathways (Dr. Lars Wittler, Department of Developmental Genetics).

Evaluation and functional characterization of non-protein-coding RNA molecules

Recent evidence indicates that the majority of sequences in eukaryotic genomes are transcribed and that the proportion of non-protein-coding RNAs (npcRNAs) increases with developmental complexity. Within an Exploratory Project of the NGFN, we have set out to identify novel npcRNAs encoded in the human genome by *in silico* prediction and subsequent proof of existence by experimental methods. We have developed an EST-filtering strategy (in collaboration with Dr. Steffen Hennig, RZPD), which yielded around 17,000 potential npcRNA candidates in the human genome (Figure 2). Experimental analysis of a subset of introns from 63 genes suggests 91 independent npcRNA candidates that are represented by 266 individual EST's. The results of RT-PCR and/or quantitative real-time PCR suggest that 74 candidates (81.4%) are transcriptionally active with varying expression levels in a variety of tissues and cells. Northern hybridisation with probes to 22 npcRNA candidates gave signals for 12 of the candidates (54.5%) in at least two different tissues. For the functional analysis of the candidates, we are currently setting up *in vitro* and *in vivo* screening techniques to find protein interaction partners by phage display and cell culture/transfected cell array approaches (in collaboration with Dr. Michal Janitz's group), respectively.

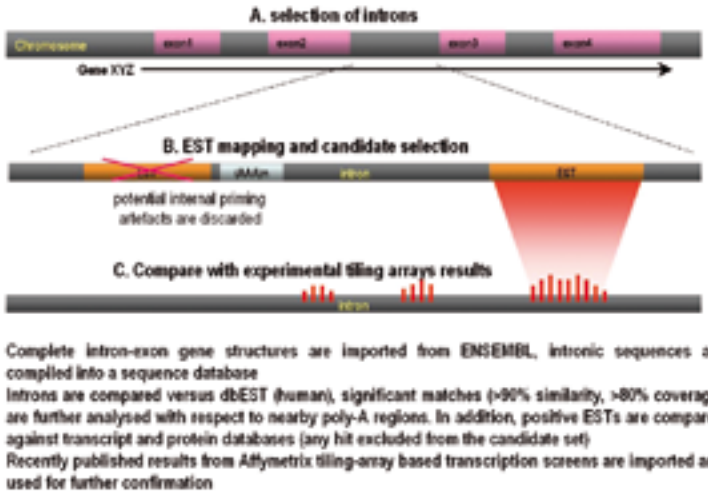


Figure 2: *In silico* selection of npcRNA candidate regions

Identification of autoantigenic markers in human diseases

Many human diseases are characterised by self-antibodies towards a variety of "self"-antigens. However, the diversity of the antigens and the specificity of the autoantibody response as well as the role of the antigens/autoantibodies in disease development and progression are only poorly understood. For the elucidation of these complex mechanisms, we investigate general autoantigenic patterns of large patient cohorts and try to correlate the observed immune responses with disease progression.

In our current studies with various external collaboration partners, we are investigating (a) the autoantibody profiles in autoimmune disorders within a BioProfile funded project, and (b) in Dilated Cardiomyopathy, a myocardial disease characterized by progressive depression of myocardial contractile function and by ventricular dilatation (SFB Transregio 19). For this purpose, high-density protein macroarrays consisting of >10,000 different human recombinant proteins are screened with clinically well-defined patient sera. Subsequently, selected proteins are expressed, purified and used to generate protein microarrays to perform qualitative and quantitative validation of these putative autoantigens (Figure 3).

Next to protein arrays, screening for autoantigens is additionally performed using cDNA expression libraries cloned in M13 or T7 display vectors presenting the recombinant proteins on the bacteriophage surface. Selection is carried out in an iterative process – essentially based on affinity enrichment – using patient-derived immunoglobulin fractions as select targets and finally, sequencing the cDNA inserts of individual bacteriophage molecules identifies the putative autoantigens.

Identification of potential diagnostic markers in Meningitis

Neisseria meningitidis is the most common cause of meningitis. We are investigating the immune responses to phase-variable expressed proteins of the most common N. meningitis serogroup B, as this group of proteins seems to play an important role in virulence determination. Using protein microarrays containing 67 recombinant proteins resulted in the identification of the phase-variable opacity protein OpaV (NMB0442) being potentially a good diagnostic and vaccine candidate.

Another subproject is focusing on the use of protein microarrays to monitor immune reactions following vaccination, and – in the case of using outer membrane vesicle vaccines – to identify the immunogenic proteins.

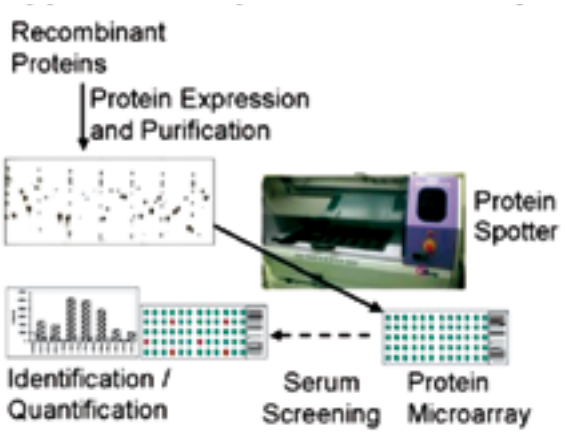


Figure 3: Overview for generations and application of protein microarrays

Selected Information

Publications

- Angenendt P, Kreutzberger J, Glokler J, Hoheisel JD. (2006) Generation of high-density protein microarrays by cell-free in situ expression of unpurified PCR products. Mol Cell Proteomics. Jul 5; [Epub ahead of print]
- Hultschig, C., Kreutzberger, J., Seitz, H., Konthur, Z., Büssow, K., Lehrach, H. (2006). Recent advances in protein microarrays. Current Opinion in Chemical Biology 10,
- Steller, S., Angenendt, P., Cahill, D.J., Heuberger, S., Lehrach, H., Kreutzberger, J. (2005). Bacterial protein microarrays for identification of new potential diagnostic markers for Neisseria meningitidis infections. Proteomics 5, 2048-2055.
- Konthur, Z., Hust, M., Dübel, S. (2005). Perspectives for systematic *in vitro* antibody generation. Gene 364, 19-29.
- Angenendt, P., Wilde, J., Kijanka, G., Baars, S., Cahill, D.J., Kreutzberger, J., Lehrach, H., Konthur, Z. & Glökler, J. (2004). Seeing better through a MIST: Evaluation of monoclonal recombinant antibody fragments on microarrays. Analytical Chemistry 74, 2916-2921.
- Kodzius, R., Rhyner C., Konthur, Z., Buczek, D., Lehrach, H., Walter, G. & Crameri, R. (2003). Rapid identification of allergen-encoding cDNA clones by phage display and high density arrays. Combinatorial Chemistry & High Throughput Screening 6, 143-151.
- Angenendt, P., Glökler, J., Konthur, Z., Lehrach, H., Cahill, D.J. (2003). 3D protein microarrays: performing multiplex immunoassays on a single chip. Analytical Chemistry 75, 4368-4372.
- Rhyner, C., Konthur, Z., Blaser, K., Crameri, R. (2003). High-throughput isolation of recombinant antibodies against recombinant allergens, Biotechniques 35, 672-676.
- Konthur, Z. & Crameri, R. (2003). High-throughput applications of phage display in proteomic analyses. DDT:TARGETS 2, 261-270.

Conference Lectures

- Galiveti, C. R. RNomics – Prediction and analysis of *in silico*-derived non-protein coding RNAs in the human genome. HUGO's 11th Human Genome Meeting, Helsinki, Finland, June 2006
- Konthur, Z. Modularised solutions towards automated selection and screening of recombinant antibodies. 4th International Congress on Recombinant Antibodies, IBC Conference, Berlin, Germany, June 2005

Diploma Theses

- Yvonne Baermann: Molekulare Charakterisierung von zentrosomalen Proteinen auf frühen embryonalen Stadien der Drosophila melanogaster (2005; Diploma Thesis, Humboldt University Berlin)
- Alexander Sternjak: Search for Autoantigens in Rheumatoid Arthritis Utilising Automated T7 Phage Display (2003; Diploma thesis, Vienna University).
- Franziska Wegerich: Protein Microarray Analysen der Immunantwort gegen Neisseria meningitidis (2006; Master thesis, University of Applied Sciences, Wildau, Germany)

Cooperations

External academic cooperations besides funded network activities

- Prof. Dr. Gerd-Rüdiger Burmester & Prof. Dr. Falk Hiepe, Department of Rheumatology and Clinical Immunology, Charité – University Medicine, Berlin
- Prof. Dr. Dolores J. Cahill, Conway Institute, University College Dublin, Ireland
- Prof. Dr. Reto Crameri, Swiss Institute of Asthma and Allergy Research, Davos, Switzerland
- Prof. Dr. Horst Dürkop, Institute for Pathology, CBF, Charité – University Medicine, Berlin
- Dr. Sigrid Heuberger, National Reference Centre for Meningococci, Graz, Austria
- Prof. Dr. Wolfram Saenger, Institute for Chemistry and Biochemistry, Department of Crystallography, Free University Berlin
- Dr. Jozef Simuth, Institute of Molecular Biology, Slovak Academy of Sciences, Bratislava, Slovakia.
- Prof. Dr. Martin Steup, Institute for Biochemistry and Biology, University Potsdam
- Dr. Elisabeth Wedege, National Institute of Public Health, Oslo, Norway

Industrial Cooperations

- IMTEC Immundiagnostika GmbH, Berlin
- In.vent Diagnostica GmbH, Hennigsdorf
- RiNA GmbH, Berlin

External Funding

- NGFN2 SMP Antibody
- Bioprofile Nutrigenomics

Neurodegenerative disorders

Head
Dr. Sylvia Krobitsch*
Phone: +49 (0)30-8413-1351
Fax: +49 (0)30-8413-1380
Email: krobitsc@molgen.mpg.de

Scientist
Dr. Josmar Langner*

Graduate Students
Linda Hallen*
Karolin Huckauf*
Ute Nonhoff*
Markus Ralser

Diploma/Master Students
Paul Bartels (until 9/2004)
Naresh Kumar Desagani
Norbert Mehlmer (until 7/2003)
Franziska Welzel

Technician
Silke Wehrmeyer*

Students
Annika Quast (until 9/2004)
Dr. Anja Westram (until 8/2004)

* external project funding

Scientific Overview

Late-onset neurodegenerative diseases, such as Alzheimer´s disease, Parkinson´s disease or the polyglutamine disorders, represent an enormous health care problem, since millions of people worldwide suffer from these disorders or are at risk. Genes that are causative for these disorders have been identified, and over the last few years, worldwide research has clearly demonstrated that aberrant folding of the respective gene products is responsible for the formation of intranuclear/cytoplasmic aggregates and/or for abnormal protein protein interactions. Despite this, the cellular pathways affected in the disease state are yet unclear for most of these disorders, and consequently, further explorations of the global cellular network of the disease proteins are fundamental for the discovery of the pathomechanisms underlying these so far untreatable disorders.

Deciphering molecular mechanisms in neurodegenerative disorders

The group aims in particular to elucidate molecular mechanisms contributing to the pathogenesis of Alzheimer´s disease and the polyglutamine disorders Huntington´s disease (HD) and spinocerebellar ataxia type 2 (SCA2) by combining yeast genetics as well as functional and bioinformatic approaches. By means of a comparative proteome analysis for the SCA2 gene product, ataxin-2 (ATX2), we substantiated a function of human ATX2 in the cellular RNA metabolism exploiting the protein network of its yeast homolog Pbp1. Within these studies, we discovered that ATX2 associates with cellular components regulating and controlling mRNA stability and translation, and moreover, that ATX2 is a component of so-called stress granules, cellular structures representing sites of mRNA triage. Additionally, we have broadened the protein interaction network for ATX2 by continuously performing large-scale yeast-2-hybrid screens. These analyses suggest that ATX2 – besides its role in the cellular RNA metabolism – seems to function in a wide range of cellular processes such as transcriptional regulation, protein stabilisation, or actin filament formation and organization. To further elucidate the physiological implications of these interactions, we started to address the aspect what the resulting functional consequences of alterations in the intracellular ATX2 concentration on its known cellular network are. This is a very important task in view of the fact that variations in the cellular ATX2 concentration seem to be critical in the pathogenesis of SCA2. Therefore, such sets of experiments shall provide valuable informations about the pathways that might contribute to SCA2. Indeed, we discovered a molecular pathway potentially contributing to the pathogenesis of SCA2 with such an approach. We observed that increased cellular concentrations of ATX2 leads to elevated levels of its interaction partner T-plastin, a protein that is involved in actin cross-link formation and filament stabilization. This finding is very promising in the light that morphological changes of the primarily affected neurons may well be based on disturbances in actin metabolism as demonstrated in the mouse SCA2 disease model. Strikingly, by performing comparative yeast genetics we discovered that interference with plastin-related pathways could be of importance for the pathogenesis of HD as well, since an interplay between ATX2 and huntingtin, which are in addition functionally connected through their interaction with endophilin-A complexes, was observed in this related pathway.

Transcriptional regulatory networks

Since transcriptional dysregulations has often been discussed as potential pathomechanism underlying several neurodegenerative disorders, we are covering this aspect by global analyses of transcription factors potentially involved in polyglutamine disorders and Alzheimer´s disease using RNAi technology in combination with microarray analysis to gain insights into the nature and complexity of the regulatory networks on the transcriptional level. This approach will help to identify novel target proteins as well as common or related pathways underlying these disorders.

Blocking of protein-protein interactions

Another challenge we are still facing is the fact that the function of most protein-protein interactions identified so far is not fully understood and their role in pathogenesis still has to be validated. In general, to comprehend and verify the significance of protein interactions in certain cellular processes, it would be an advantage to study the cellular role of individual protein-protein interactions in the natural milieu. This is complicated by the fact that knock-out technologies as well as sole over expression often do not comply with the natural environment. For obtaining appropriate tools allowing characterizing the cellular relationship of protein-protein interactions and investigating their involvement in certain pathways, we are developing in collaboration with Dr. Zoltán Konthur´s group, a novel screening procedure that is based on a combination of in vitro/in vivo ligand screening methods using nucleic acid aptamer/phage display technology and the reverse yeast-2-hybrid system. This approach will result in a number of highly specific ssRNA aptamers/intrabodies with specific inhibitory activity that will be used to validate the protein-protein interactions in focus. Importantly, this approach is generally applicable for all protein-protein interactions happening in the yeast-2-hybrid system. In an ongoing cooperation with Dr. Lars Wittler (Dep. Developmental Genetics) we are currently selecting aptamers/intrabodies against specific protein-protein interactions implicated in specific signalling pathways allowing analyzing intracellular interconnections of these signalling cascades. In the future, the accessibility of molecules with blocking properties not only promise to be valuable tools in functional proteomics but may also be very useful in structural biology for basic research and drug discovery – perspectives that are of central relevance for medicine.

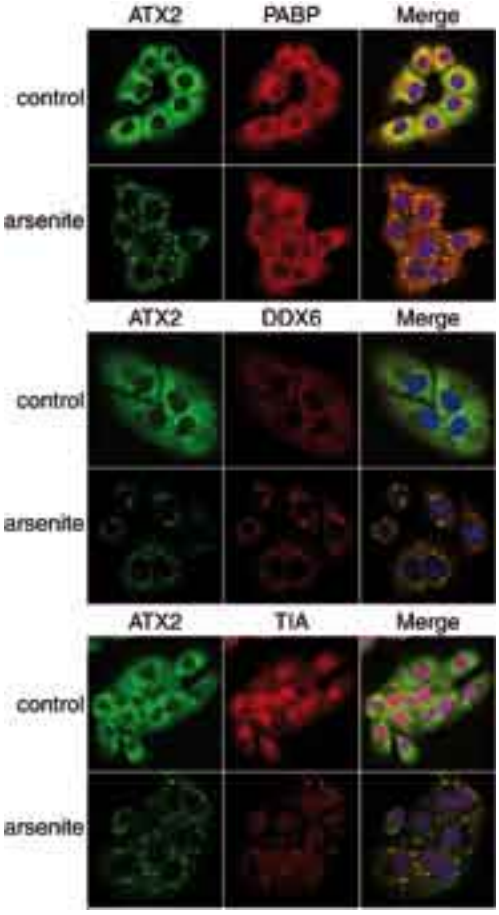


Figure 1: ATX2 is a component of stress granules DU145 cells were treated for 1 h with 0,5 mM arsenite or left untreated. Cells were stained for ATX2 and its interaction partners PABP or DDX6. TIA, as marker protein for stress granules, was visualized as well. Nuclei were stained with Hoechst. Pictrues were taken with an LSM510meta (Zeiss).

Future perspective

Since molecular pathways implicated in neurodegenerative disorders potentially play a role in human ageing as well, the group has recently identified a promising candidate gene implicated in human ageing. Currently, we are characterizing this candidate gene and its gene product in collaboration with Dr. Bernd Timmermann (group Dr. Richard Reinhardt). Moreover, by combining general yeast genetics and systems biology approaches in collaboration with Dr. Edda Klipp, we set up to model the respective pathway.

Selected Information

Publications

- Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., Timm, J., Mintzlaff, S., Abraham, C., Bock, N., Kietzmann, S., Goedde, A., Toksoz, E., Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehrach, H., and Wanker, E.E. (2005). A human protein-protein interaction network: a resource for annotating the proteome. Cell 122 (6): 957-968
- Ralser, M., Nonhoff, U., Albrecht, M., Lengauer, T., Wanker, E.E., Lehrach, H., and Krobitsch, S. (2005). Ataxin-2 and huntingtin interact with endophilin-A complexes to function in plastin-associated pathways. Hum. Mol. Genet. 14 (19): 2893-2909
- Ralser, M., Goehler, H., Wanker, E.E., Lehrach, H., and Krobitsch, S. (2005). Generation of a yeast-two-hybrid strain suitable for competitive protein binding analysis. Biotechniques 39 (2): 165-166, 168
- Ralser, M., Albrecht, M., Nonhoff, U., Lengauer, T., Lehrach, H., and Krobitsch, S. (2005). An integrative approach to gain insights into the cellular function of human Ataxin-2. J. Mol. Biol. 346 (1): 203-214

Invited plenary lectures

- 2005 Gordon Conference, CAG Triplet Repeat Disorders, Mount Holyoke, USA
- 2006 Center of Molecular Medicine, Seminar series, University of Cologne, Germany

Diploma Theses

- Markus Ralser, Identifizierung von Interaktionspartnern von Ataxin-2, University of Salzburg, Austria
- Norbert Mehlmer, Identifizierung der Interaktionspartner der β- und γ-Synucleine, University of Salzburg, Austria
- Paul Bartels, Identifizierung von Interaktionspartnern des Proteins Ataxin-2, Free University of Berlin

PhD Thesis

- Markus Ralser, Exploring molecular pathways contributing to spinocerebellar ataxia type 2, University of Salzburg

Teaching

- 2006 Practical course in Physics, University of Hamburg

Work as scientific referee

- Biotechniques
- Human Molecular Genetics

External Cooperations

- Prof. Gillian Bates, GKT School of Medicine, King´s College, London, UK
- Prof. Dr. Dr. Christian Betzel, University of Hamburg, Germany
- Prof. Michael Breitenbach, University of Salzburg, Austria
- Prof. Bernd Groner, Georg Speyer Haus, Institute for Biomedical Research, Frankfurt, Germany
- Prof. Dr. Susan Hollán, National Institute of Blood Transfusion, Budapest, Hungary
- Prof. Thomas Lengauer, Mario Albrecht, Max Planck Institute for Informatics, Saarbruecken, Germany
- Prof. Dr. Stefan Schreiber, Institute for Clinical Molecular Biology, University Hospital Schleswig-Holstein, Kiel, Germany
- Prof. Erich Wanker, Max Delbrueck Center for Molecular Medicine, Berlin, Germany

External Funding

- Ataxia UK: Investigations into the cellular function of ataxin-2 by the use of ssRNA aptamers
- EU-FP6: A European Infrastructure of Ligand Binding Molecules against the human proteome
- NGFN: Analysis of transcription regulatory networks

Protein Complexes & Cell Organelle Assembly/ USN

Heads
PD Dr. Bodo M.H. Lange*
Phone: +49 (0)30-8413 1645
Fax: +49 (0)30-8413 1128
Email: lange_b@molgen.mpg.de

Dr. Thorsten Mielke*
Phone: +49 (0)30-8413 1644 / 1141
Fax: +49 (0)30-8413 1128
Email: mielke@molgen.mpg.de

Scientists
Marie-Laure Fogeron*
Justus Loerke*
Dr. Christian Regenbrecht

Technicians
Verena Lehmann
Eileen Bastian

Graduate Students
Karin Habermann
Hannah Müller
Anne-Kathrin Scholz

Engineer
Tanja Kurtz

Students
Maik Hintze
Karolin Pollok
Reza Fard
Lars Hildebrandt

* external project funding

Scientific Overview

Sequencing the genomes of multiple organisms including *Drosophila*, zebra fish and human has revealed the total number of genes and the sequence of the encoded proteins. The next steps towards understanding biological processes including diseases require the characterization of the protein composition in a given cell, the protein function and protein protein-interactions. Consequently, understanding protein function at a molecular level requires profound knowledge of the structure of a protein and its different functional states. However, many key processes in biology are performed by large protein complexes such as multi-enzyme-complexes, ribosomal complexes and signaling complexes. These complexes act as highly organized molecular machines. Importantly, these complexes show a variable assembly and are subject to dynamical regulation, both making structure determination difficult. Our group aim for a systematic analysis of macromolecular complexes combining complex isolation, analysis of their protein composition, protein function and finally structure determination by cryo-electron microscopy.

Protein Complexes and Cell Organelle Assembly

(group leader B. Lange)

Key questions

How do molecular interactions regulate or mediate biological processes and how do changes in these interactions lead to diseases?

Approach

We have established an efficient platform to isolate cell organelles and large protein complexes using magnetic bead-based affinity purification protocols, to isolate microgram quantities of native protein complexes. Using this technology in parallel with tandem-affinity purification (TAP), protein complexes are isolated and then characterised by mass spectrometry (in collaboration with the mass spectrometry group Gobom). Protein complexes are analysed for their suitability for cryo-electron microscopy using negative staining electron microscopy. This work is followed by systematic RNA interference (RNAi) of the identified proteins to uncover their function on a cellular level. The data obtained by mass spectrometry and cryo-EM will be highly complementary, with the former establishing the composition of a complex, and the latter the structure and interactions between the components.

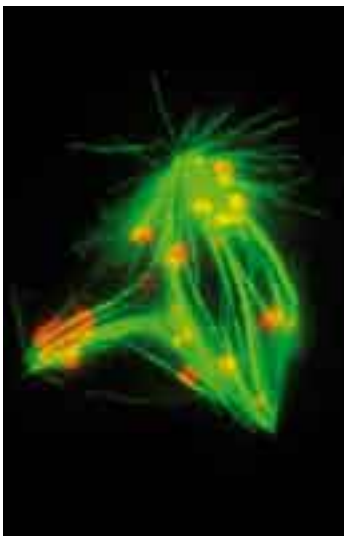


Figure 1: Immunofluorescence microscopy staining of a Drosophila SL2 cell uncovering the function of an identified centrosomal component from our RNA interferences functional screen. Chromosomes (red) are displaced from the metaphase plate, while microtubuli filaments (green) have formed an extra half spindle in vicinity to the main spindle.

Results

We applied large-quantity-cell extracts ($5x10^9$ cells) for TAP and immunoprecipitation experiments for the isolation of native protein complexes from human cells. So far, over 50 bait proteins were expressed and are now in the immediate pipeline targeted for protein-protein interaction determination. Applying the TAP method to isolate key regulatory signalling molecules from exponentially growing cell cultures we identified a series of novel protein-protein interactions in cell division and checkpoint control pathways. In addition we affinity purified the large protein complex the centrosome (the microtubule organising centre in higher eukaryotic cells) from the early syncytial stages of the *Drosophila* embryo. We identified over 250 candidate protein components of this cell organelle and characterised their function via RNAi for centrosome structure, cell cycle progression and mitosis (Fig.1). Hence, this project provides molecular insight into some of the major pathways that regulate cell proliferation, growth and stress signalling occurring in both healthy and diseased tissue.

Figure 2: The Tecnai G2 Polara cryo-electron microscope of the USN.

Cryo-EM: Structure determination of macromolecular complexes by cryo-electron microscopy – Ultrastructure Network

(group leader T. Mielke)

The approach

Cryo-electron microscopy in combination with the single-particle approach is a powerful method to gain structural information on macromolecular protein complexes without the need to crystallize these complex and dynamic systems. The molecules of interest are embedded in a thin layer of vitreous ice under near physiological conditions and imaged by transmission cryo-electron microscopy under low-dose conditions. Image processing then allows the reconstruction of the three-dimensional structure at sub-nanometer resolution.

Results

Starting in 2004, we have established a state-of-the-art cryo-electron microscopy facility providing latest techniques such as semi-automated sample vitrification and intense computing resources. The core instrument of the USN is a helium-cooled 300 kV Tecnai G2 Polara cryo-electron microscope (FEI) equipped with a 4kx4k CCD camera (TVIPS). The leginon system (AMI group, The Scripps Research Institute, La Jolla, USA) has been implemented for automated routine data collection. Acquisition of tilt series for cryo-electron tomography has been established using the FEI tomography package.

During the installation of the facility, we already recorded cryo-data from more than 50 mainly ribosomal complexes provided by the groups of R. Beckmann (Charité Berlin, now University of Munich – LMU), C. Spahn (Charité, Berlin), K. Nierhaus (MPIMG) and other partners of the USN. Further target complexes amongst others are RNA polymerases and components from the visual signal transduction cascade. The data is currently processed at different stages with some data sets clearly showing a resolution better than 10 Å. For example, we solved the structure of the active ribosome in complex with the signal recognition particle (SRP) and the SRP receptor at 8 Å resolution (Halic *et al.* (2006) Science 312, 745-747). This complex is a key player in protein translocation. SRP binds to the N-terminal signal sequence of a new polypeptide chain emerging from the ribosome and transfers the ribosome to the translocon complex in the membrane of the endoplasmic reticulum. Interestingly, upon SRP binding, the ribosome is unable to bind directly to the translocon complex. This step instead requires the additional interaction with the SRP receptor. The obtained structure of the complex reveals how interaction of the SRP receptor with both, the ribosome and the SRP, displaces parts of the SRP molecule, which leads to the exposition of ribosomal translocon binding sites.

Future perspectives

Rapid developments in the field of cryo-electron microscopy including modern microscopes such as the G2 Polara allow for the very first time structure determination of macromolecular complexes at sub-nanometer resolution. After establishing the high-end cryo-electron microscopy facility of the USN, we are now aiming for developing new tools and methods e.g. in the field of image processing, which are required to fully exploit the capacity of cryo-electron microscopy in combination with the single particle approach. This involves new strategies for classifying and sorting particles from different sub-states of a complex, implementation of cryo-electron tomography for initial structure determination and specific problems of high-resolution projects. For this purpose we initiated the project "Anwenderzentrum". This project also addresses questions in the field of biomedical structural research, which so far were out of reach by cryo-electron microscopy techniques.

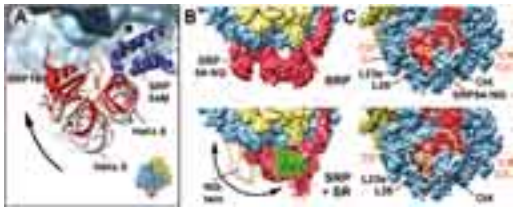


Figure 3: Cryo-EM reconstruction of the 80S RNC-SRP-SR complex. Binding of the SRP receptor (SR) to the ribosomal nascent chain-SRP complex leads to the rearrangement of SRP S domain and the exposure of the translocon binding site. (Halic *et al.* (2006) Science 312, 745-747)

Selected Information

Publications

- Halic, M., Gartmann, M., Schlenker, O., Mielke, T., Pool, M.R., Sinning, I. and Beckmann, R. (2006) Signal Recognition Particle Receptor Exposes the Ribosomal Translocon Binding Site. Science 312, 745-747
- Lange, B.M.H., Kirfel, G., Gestmann, I., Herzog, V., González, C. (2005). Structure and Microtubule-Nucleation Activity of Isolated Drosophila Embryo Centrosomes Characterized by Whole Mount Scanning and Transmission Electron Microscopy. Histochemistry and Cell Biology, 2005. 124(3-4):325-334.
- Lehmann, V.; Müller, H. and Lange, B.M.H. (2005) Immunoisolation of centrosomes from Drosophila melanogaster. Current Protocols in Cell Biology, supplement 29, Dec 2005
- Ruprecht, J., Mielke, T., Vogel, R., Villa, C. & Schertler G.F.X (2004) Electron crystallography reveals the structure of metarhodopsin I. EMBO J. 23, 3609–3620
- Sauer, S.; Lange, B. M. H.; Gobom, J.; Nyarsik, L.; Seitz, Harald and Lehrach, H. (2005) Miniaturization in functional genomics and proteomics. Nature Genetics Review, 6, 6(6):465-476

Teaching

- T. Mielke, Lecture as part of the program of study "Medizinische Physik", Charité
- B. Lange, Lecture as part of the FU series "Gene and Genomes", Practical course in Drosophila Developmental Biology as part of the Module 3 of the Molecular Medicine Master Course, Charité

Cooperations (B. Lange & T. Mielke)

- Prof. K.P. Hofmann, Prof. C.M.T. Spahn, Institut für Medizinische Physik und Biophysik, Charite – Universitätsmedizin Berlin, Germany
- Prof. R. Beckman, Prof. P. Cramer, Genzentrum, Universität München, Germany
- Prof. E. Wanker, Max Delbrück Centrum für Molekulare Medizin, Berlin-Buch, Germany
- Dr. G.F.X. Schertler, MRC Laboratory of Molecular Biology, Hills Road, Cambridge, UK
- Dr. M. Reuter, Institut für Virologie, Charite – Universitätsmedizin Berlin, Germany
- Dr. Christiane Richter-Ehrenstein/Prof. Dr. A. Schneider, Campus Charité, Berlin, Germany
- Dr. Aram Prokop, Arbeitsgruppe Experimentelle Onkologie, Klinik für Pädiatrie mit Schwerpunkt Onkologie, Campus Virchow-Klinikum, Berlin, Germany
- Dr. Michael Boutros, German Cancer Research Center, Heidelberg, Germany
- Prof Mathias Uhlen, Department of Biotechnology, AlbaNova University Center, Royal Institute of Technology (KTH), Stockholm, Sweden
- Prof. Cayetano González, ICREA-IRB-PCB, Cell Division Laboratory, Barcelona, Spain

Student Theses (B. Lange)

Diploma Theses

- H. Müller: Identifizierung und Molekulare Charakterisierung von Zentrosomalen Proteinen in Drosophila melanogaster, 2004
- Y. Baermann: Molekulare Charakterisierung von zentrosomalen Proteinen aus frühen embryonalen Stadien der Drosophila melanogaster, 2005

Doctoral Thesis

- C. Regenbrecht: Identifikation epigenetisch regulierter Gene in Gliomen 2005

Work as Scientific Referee (B. Lange)

Journals

- Cancer Research
- Cell Motility and the Cytoskeleton
- Contributing Member of Faculty of 1000
- Trends in Cell Biology
- Journal of Cell Science

Grant Evaluations

- GEN-AU (Genome Research in Austria)

External Funding (B. Lange & T. Mielke)

- Ultrastructure Network Grant, EFRE (June 2003 – June 2006)
- Anwenderzentrum, EFRE (May 2006 – November 2008)

B. Lange

- SMP Protein (NGFN2) (November 2004 – October 2007)
- Thyssen Project Grant (June 2005 – June 2007)

Automation & Technology Development

Head
Prof. Dr. Hans Lehrach
Phone: +49 (0)30-8413 1220
Fax: +49 (0)30-8413 1380
Email: lehrach@molgen.mpg.de

Senior Scientists
Dr. Claus Hultschig*
Dr. Wilfried Nietfeld* (Guest)
Dr. Lajos Nyarsik*
Dr. Sascha Sauer*
Dr. Harald Seitz*
Dr. Aleksey Soldatov*
Dr. Ralf Sudbrak* (Guest)
Dr. Ulrike Dohrmann*
Dr. Birgit Gerisch*

Postdoc
Dr. Tatiana Borodina

PhD-Students
Andreas Dahl*
Gina Ziegler*
Scholastica Baek*
Anika Andersson*

Diploma Student
Anna Politis

Engineers
Thomas Przewieslik*
Thomas Nitsche
Matthias Lange*
Regine Schwartz*
Tillmann Wegwerth*

Technicians
Nicole Greiner
Kerstin Zurth
Ulrike Borgmeier*
Karl Rak

*external project funding

Scientific Overview

The main direction of this group has been to carry out exploratory projects, mostly in the direction of technology development, and to provide some of the essential infrastructure needed for high throughput genome analysis. The group has been at the forefront of technology development for genome research for many years (e.g. first development of automated array spotters and high density filter formats in 1987) and continues to carry out high level technology development for some of the problems in functional genomics and systems biology, needed as the basis of e.g. a systems biology of cancer and other diseases. Automation and miniaturization of biological methods, e.g. gene expression analysis, genotyping or sequencing provide the basis for whole-genome, transcriptome and proteome analysis in high-throughput. They are essential prerequisites for a functional understanding of biological processes. In cooperation with partners from industry and academia, the development of novel detection devices, protocols, robots, robot control software, and automation concepts has been pioneered. The importance of such research is reflected by scientific cooperations with many groups worldwide.

Due to the nature of the work, and the diverse funding sources, the group contains a number of small subgroups, each focused on a particular area of technology development. In addition, we describe here two still small (pilot) projects, the analysis of ageing in *C. elegans* (B. Gerisch, NGFN ageing grant, in collaboration with A. Antebi), and the systematic analysis of regulatory networks in the mouse (U. Dohrmann, A. Soldatov, H. Herwig, M.-L. Yaspo) based on mouse chromosome substitution strains (in collaboration with J. Forejt, Prague)

Current state of research

The team of L. Nyarsik is developing technologies for high throughput genome and transcriptome analysis. A novel nanowell technology platform for versatile multifunctional applications has been developed e.g. for PCRs in volumes below 50 nl. New sophisticated equipments are established for liquid handling in nanoliter range or processing and detection in nanowell formats. The main aim of the miniaturization concepts is the reduction of costs and sample requirements for large-scale applications. The nanowell technology will focus on gene expression profiling by RT-PCR (S. Sperling, M.-L. Yaspo), oligonucleotide fingerprinting (M. Janitz) and genotyping. For this, a novel, high throughput protocol for SNP scoring based on optical detection has been developed.

The research topic of the group of S. Sauer is technology development within life science in conjunction with biological variation studies. Over the past years they developed procedures for genotyping single nucleotide polymorphisms and molecular haplotyping, recently created a powerful methodology for the detection of small molecules binding to large molecules such as proteins and nucleic acids. In the frame of the EU MolTools project they are currently working on novel methods for array-based mass spectrometric detection of nucleic acids and on miniaturization of assays in nanowells using optical detection. Using these technologies genotyping and haplotyping studies in the frame of NGFN-2 and GABI-Génoplante are performed. Furthermore they are screening protein-binding partners of pharmacologically relevant proteins.

The group of A. Soldatov is specialized in the area of SNP genotyping and sequencing. A new SNP-genotyping method was developed, patented (EP03019521) and successfully applied for genotyping of *Arabidopsis thaliana* and human DNA samples. Now the work is concentrated on the development of new sequencing technologies in the frame of the EU MolTools project: (I) microbead based sequencing (EP05018950.5) and (II) ligation based sequencing (U. Landegren).

The team of C. Hultschig is active in the generation of high quality microarrays to be used in multitude of different applications ranging from the determination of chromosomal imbalances, mRNA expression profiling, identification of protein targets of kinases, and analyzing protein-protein and protein-DNA interactions. To this end, robots for the generation of microarrays, have been modified and significantly improved in both hard and software resulting in a quality controlled arraying unit. Microarrays have been used for analyzing relative abundances of coding transcripts and micro-RNAs in different murine, human, avian, and amphibian samples and deciphering protein-protein interactions and identifying kinase targets. During the last two years more than 2300 arrays each presenting more than 40,000 PCR products were produced for analysis of chromosomal imbalances in patient samples (R. Ullmann, Dept. Ropers). Currently our scientific focus is on the highly parallel investigation of protein-DNA interactions using DNA microarrays and recombinant proteins.

The focus of the group of H. Seitz is the developing and establishing of microarray based methods to characterize and identify protein-protein and protein-DNA interactions. Using the developed technique, it was possible to identify cation dependent interactions between members of the S100 protein family and putative interaction partners. Beside already know interactions several new once were verified with an independent method. In addition functional protein microarrays for the analysis of protein-DNA interactions have been developed. Together with the companies febit, Scienion and Biacore different array based techniques to characterize protein-DNA interactions were established.

The team of R. Sudbrak is involved in genomic sequencing of human chromosomes 3 and X and other medically important regions of the human genome. A gene catalogue of the human chrX was established which provided the basis for the first chromosome-specific cDNA microarray. Another array project aimed to improve the understanding of the interaction between humans and the bacterium *Tropheryma whipplei* in *Whipple's Disease* (EU grant Whipple Disease, R. Sudbrak). Since 2003 this group contributed to the completion of the human genome sequence, especially to the completion and analysis of human chrX and chr3. In collaboration with various clinical partners the disease genes BSND, NPHP4, ATPC1, DNAH5, and TM4SF2 were successfully cloned as well as NPHP3, and NPHP5. Based on this know-how a resequencing platform was established within the NGFN. With the near completion of the human genome the attention has shifted to the sequencing of model organisms (rat MHC, chimpanzee). The chimpanzee chromosome 22 (PTR22 – equivalent of human chr21) was sequenced and analyzed under coordination of RIKEN. Sequencing of chimpanzee PTRX is under way (in co-operation with R. Reinhardt and M-L. Yaspo (coordination)). These comparative genomic approaches are supplemented by the NGFN project Pathway mapping, where they concentrate on evolutionary aspects of disease genes and the EU funded project APES which will reveal deeper insights to the question "What makes us human" (co-operation with Jean-Louis Nahon, CNRS, Leo Schalkwyk, KCL, and Lutz Walter, DPZ, project co-ordination by R. Sudbrak).

The research activity of W. Nietfeld's team is focused on the development of technologies for functional genomics, e.g. generation of gene specific tags in *Arabidopsis* (DNA arrays and RNAi) and human promoter-reporter constructs. In collaboration with clinicians (S. Schreiber, U. Dirnagl and H. Wekerle) inflammatory processes in different diseases (inflammatory bowl disease, brain ischemia and experimental autoimmune encephalitis (EAE)) are studied.. A further focus is to establish nanoparticels (Quantum Dots) for labeling RNA/cDNA and proteins (M. Seydack), these particles are stable to photo bleaching and differently labeled samples can be analyzed in parallel (multiplexing).
Birgit Gerisch is following up work on the regulation of development and life span in *C. elegans* initiated by Adam Antebi during his time as leader of an independent junior group (funded by NGFN, EP on aging). The life span of *C. elegans* can be extended by different mechanisms including calorie restriction, reduced Insulin/IGF-1 signaling, germline ablation, ablation of food sensing, neurones, mitochondrial deficiency, as well as decreased temperature. Genes and pathways involved in survival in respond to challenging environmental conditions are mostly conserved from worm to man. The decision between arrest at a long-lived alternate dauer stage, or reproductive development in respond to environmental signals is ultimately mediated by daf-9, a cytochrome P450 related to steroidogenic hydroxylases and its cognate nuclear receptor daf-12, implying organism-wide coordination by lipophilic hormones. A conserved Rieske-like oxygenase, DAF-36, was identified and characterized as another component in this hormone metabolism. Larval phenotypes of daf-9 and daf-36 are reversed by 7-dehydrocholesterol (Montola *et al.*, 2006). Its effect on life span regulation and stress resistance is being analysed (paper in preparation).
To identify additional genetic factors, which contribute to a long and healthy life as well as to identify molecular pathways that are associated with the physiology of ageing and/or age-related diseases, this group is using the correlation between longevity and stress resistance in *C. elegans* to identify genes that are involved in the process of life span regulation. So far, they have analysed around 500 phosphatases and kinases using RNAi experiments for heat and oxidative stress resistance, and tested candidates provided by other partners. We identified about 150 candidate genes for stress resistance, which we are testing in longevity experiments. Among those, an evolutionary conserved phosphatase involved in cell division, shows effects in resistance to heat stress and in life span extension.
As a new, powerful approach to unravel the network of regulatory processes translating genotype (and environment) into the phenotype of the organisms, we have started a pilot project taking advantage of a set of chromosome substitution strains constructed by Jiri Forejt (Prague) (U. Dohrmann, A. Soldatov, W. Nietfeld, M-L.Yaspo). This system offers the chance to globally analyse regulatory networks in the mouse, and to serve as the basis of a systems biology of regulatory networks, with impact on most or all of the work within the department, as well as the other departments in the institute.

Future perspectives

The work of the group, together with many of the other groups within the department, is focussed on the development of many of the tools needed for a systems biology of development (including aging) and disease, with components from physics, chemistry, engineering, biology, genetics and medicine in mouse and model organisms.

Selected Information

Publications

- Seitz H, Hutschenreiter S, Hultschig C, Zeilinger C, Zimmermann B, Kleinjung F, Schuchhardt J, Eickhoff H, Herberg F. (2006) Identification of new S100 proteins interaction partners combining functional protein microarrays and surface plasmon resonance, Proteomics, in press
- Kersten B, Possling A, Blaesing F, Mirgorodskaya E, Gobom J and Seitz H (2004) Application of protein microarray technology and UV crosslinking combined with mass spectrometry for the analysis of protein-DNA interactions, Anal Biochem15;331(2):303-13
- Muzny DM, Scherer SE, Kaul R, Wang J, Yu J, Sudbrak R, Buhay CJ, Chen R, Cree A, Ding Y, Dugan-Rocha S, Gill R, Gunaratne P, Harris RA, Hawes AC, Hernandez J, Hodgson AV, Hume J, Jackson A, Khan ZM, Kovar-Smith C, Lewis LR, Lozado RJ, Metzker ML, Milosavljevic A, Miner GR, Morgan MB, Nazareth LV, Scott G, Sodergren E, Song XZ, Steffen D, Wei S, Wheeler DA, Wright MW, Worley KC, Yuan Y, Zhang Z, Adams CQ, Ansari-Lari MA, Ayele M, Brown MJ, Chen G, Chen Z, Clendenning J, Clerc-Blankenburg KP, Chen R, Chen Z, Davis C, Delgado O, Dinh HH, Dong W, Draper H, Ernst S, Fu G, Gonzalez-Garay ML, Garcia DK, Gillett W, Gu J, Hao B, Haugen E, Havlak P, He X, Hennig S, Hu S, Huang W, Jackson LR, Jacob LS, Kelly SH, Kube M, Levy R, Li Z, Liu B, Liu J, Liu W, Lu J, Maheshwari M, Nguyen BV, Okwuonu GO, Palmeiri A, Pasternak S, Perez LM, Phelps KA, Plopper FJ, Qiang B, Raymond C, Rodriguez R, Saenphimmachak C, Santibanez J, Shen H, Shen Y, Subramanian S, Tabor PE, Verduzco D, Waldron L, Wang J, Wang J, Wang Q, Williams GA, Wong GK, Yao Z, Zhang J, Zhang X, Zhao G, Zhou J, Zhou Y, Nelson D, Lehrach H, Reinhardt R, Naylor SL, Yang H, Olson M, Weinstock G, Gibbs RA.(2006) The DNA sequence, annotation and analysis of human chromosome 3. Nature 440: 1194-1198
- Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, Platzer M, Howell GR, Burrows C, Bird CP, Frankish A, Lovell FL, Howe KL, Ashurst JL, Fulton RS, Sudbrak R *et al.* (2005) The DNA sequence of the human X chromosome. Nature 434: 325-337.
- Horvath J, Fliegauf M, Olbrich H, Kispert A, King SM, Mitchison H, Zariwala MA, Knowles MR, Sudbrak R, Fekete G, Neesen J, Reinhardt R, Omran H. (2005) Identification and analysis of axonemal dynein light chain 1 in primary ciliary dyskinesia patients. Am J Respir Cell Mol Biol. 33 (1): 41-47.
- Otto EA, Loeys B, Khanna H, Hellemans J, Sudbrak R, Fan S, Muerb U *et al.*. (2005) Nephrocystin-5, a ciliary IQ domain protein, is mutated in Senior-Loken syndrome and interacts with RPGR and calmodulin. Nat Genet. 37:282-288
- International Human Genome Sequencing Consortium. (2004) Finishing the euchromatic sequence of the human genome Nature 431: 931-945
- Watanabe H, Fujiyama A, Hattori M, Taylor TD, Toyoda A, Kuroki Y, Noguchi H, BenKahla A, Lehrach H, Sudbrak R *et al.* (2004) DNA sequence and comparative analysis of chimpanzee chromosome 22. Nature 429: 382-388.
- P. Hurt, L. Walter, R. Sudbrak, S. Klages, I. Müller, T. Shiina, H. Inoko, H. Lehrach, E. Günther, R. Reinhardt and H. Himmelbauer (2004) The genomic sequence and comparative analysis of the rat major histocompatibility complex. Genome Res 14: 631-639.
- Olbrich H, Fliegauf M, Hoefele J, Kispert A, Otto E, Volz A, Wolf MT, Sasmaz G, Trauer U, Reinhardt R, Sudbrak R, Antignac C, Gretz N, Walz G, Schermer B, Benzing T, Hildebrandt F, Omran H. (2003) Mutations in a novel gene, NPHP3, cause adolescent nephronophthisis, tapeto-retinal degeneration and hepatic fibrosis. Nature Genetics 34: 455-459.
- Sudbrak R, Reinhardt R, Hennig S, Lehrach H, Gunther E, Walter L. (2003) Comparative and evolutionary analysis of the rhesus macaque extended MHC class II region. Immunogenetics 54:699-704.

Reviews

- Weigel C., Seitz H. (2006) Bacteriophage replication modules, FEMS Microbiol. Rev. 30(3):321-81
- S. Sauer, B.M. Lange, J. Gobom, L. Nyarsik, H. Seitz, and H. Lehrach (2005) "Miniaturization in functional genomics and proteomics" Nature Reviews Genetics. 6: 465-476

Organization of scientific events

- Organisation of the workshop "Moleculare Interaction", Berlin 2005 and 2006 (C. Hultschig. H. Seitz and W. Nietfeld)
- Local organizer of HGM2004 (Human Genome Meeting) in Berlin

External Funding

- EU Genetics of IBD, INFRAQTL (until 2004), CAGE (until 04/06) MOLTOOLS, Whipples Disease
- NGFN2 SMP DNA SMP Protein
- Applera

Honors (Hans Lehrach)

- The Ján Jessenius SAS Medal of Honour for outstanding achievements in medical sciences, Slovak Academy of Sciences, Bratislava (2003)
- Karl Heinz Beckurts Award (2004)

Evolution and Development

Heads
Dr. Georgia Panopoulou*
Dr. Albert J. Poustka
Phone: +49 (0)30-84131235
Fax: +49 (0)30-84131380
Email: panopoul@molgen.mpg.de
poustka@molgen.mpg.de

Postdoc
Dr. Maryam Khorasani*

Informatics
Florian Thiel
Clemens Kühn

Phd Student
Alexander Kühn

Technician
Vesna Weisse

* external project funding

Scientific Overview

The aim of our group is to analyse at the sequence level the changes of genes during evolution and the impact of these changes on gene function. The divergence of protostomes and deuterostomes and the transition from invertebrates to vertebrates were accompanied by novel structures and an increase in gene number. Invertebrates have approximately half the number of genes than vertebrates. Nevertheless, genome comparisons have revealed that 83% of *C.elegans* genes have human orthologs and reverse only 20% of the human genes are vertebrate specific. This indicates that these extra vertebrate genes are not novel but were amplified from already existing gene families.

It has been argued that this increase in gene number was the result of complete genome duplications at the origin of vertebrates (2R hypothesis). In a previous project, we constructed a gene catalogue from amphioxus, the closest invertebrate chordate to vertebrates. Via the phylogenetic analysis of the amphioxus genes and their vertebrate orthologs we estimated that this duplication event took place between 300 and 680 Mya. We then developed an approach for detecting duplicated segments in the human genome that contain these "old" duplicates. Based on the distribution of these segments we showed that this duplication event was at least one complete genome duplication.

Studies mainly from yeast genomes have shown that the fate of the duplicates is to get silenced shortly after duplication. Indeed, even though at least one complete genome duplication took place at the origin of vertebrates, the human genome contains less than twice the number of invertebrate genes. We intend to understand how this decision on which duplicated genes are maintained is made (see below).

Genes function within the context of regulatory networks. The fate of duplicated genes is probably determined to some extent as a result of the network they belong to. To understand how gene regulatory networks (GNR) function and evolve, we concentrate on the study of the endomesoderm network in sea urchin (*S.purpuratus*) and *Ciona intestinalis*.

Due to the fast development and robustness of embryos in combination with having their complete genome sequenced both these organisms are unique model systems for studying GRN.

As the first few components of the sea urchin endomesoderm network start becoming known, it becomes apparent the complexity as well as the accuracy of this network in controlling and reinforcing appropriate cell fate at specific times. Considering that sea urchins have only 20,000 genes, the complexity that vertebrate GNRs might have becomes obvious as well as the necessity to turn to a less complex system such as sea urchins before we attempt to understand vertebrate GRN.

The sea urchin endomesoderm gene regulatory netwok (GRN)

In sea urchins, two array screens that targeted endomesoderm genes have been carried out: a subtractive screen where RNA from lithium treated embryos was subtracted with RNA isolated from cadherin injected embryos (Ransick *et al.*, 2002) and a Brachyury target gene screen (Rast *et al.*, 2002). Lithium activates the Wnt signaling pathway and enhances endomesoderm formation. Cadherin blocks endomesoderm formation.

We followed an alternative approach consisting of separate array hybridisations of LiCl, ZnCl2 and NiCl2 treated and normal embryos followed by *in silico* subtraction of the genes regulated differently in those treatments. As lithium leads to expanded endomesodermal domain and zinc eliminates endomesoderm, selecting for genes that are upregulated in the first treatment and the same genes downregulated in the second, we select for endomesoderm specific genes. Through this approach, we avoid the amplification or subtraction of probes that can lead to the distortion of the representation of transcripts. Finally, our screens were carried out using unigene arrays (Poustka *et al.*, 2003) that represent 70% of all sea urchin genes. All genes found to be regulated in our screens as well as being expressed in the prospective endomesoderm domain are currently added in the endomesoderm network (Fig. 1). We have also identified a sea urchin homolog of dickkopf (SpDkk) which is a known inhibitor of Wnt signalling in vertebrates. Whole mount *in situ* hybridisation revealed that SpDkk is indeed expressed in prospective endomesoderm. SpDKK knockout results in expanded endomesoderm. We are currently testing if genes upregulated in the LiCl-ZnCl2 screen are targets of SpDKK.

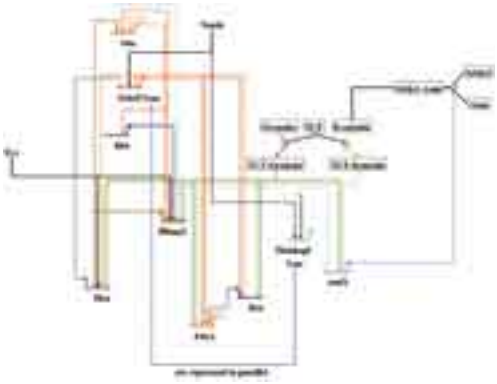


Figure 1: The sea urchin *(S. purpuratus)* endomesoderm gene regulatory network (GNR).

The Ciona Intestinalis endomesoderm gene regulatory network (GRN)

In collaboration with the group of Patrick Lemaire (IBDM, Marseille) we intend to identify genes involved in endomesoderm development through gene knockout and array hybridization experiments. During the first phase of this project, we have constructed a *Ciona Intestinalis* oligonucleotide microarray that carries all transcription factors, signal transduction and cell cycle genes (for project description see *http://goblet.molgen.mpg.de/cgi-bin/webapps/ciona.cgi*). This project is funded within the EU funded Marine Genomics Network of Excellence.

Evolution of gene function duplication

We have performed a whole mount *in situ* hybridisation screen of 50 gene families in both amphioxus and zebrafish (Fig. 2). We observe that: i) neofunctionalisation is very frequent, ii) the expression domain of zebrafish duplicates becomes more restricted as embryonic development progresses thus probably achieving less overlap.

Generating Critical Resources for Marine Organisms

During the past years we have generated significant genomic resources for sea urchin, amphioxus and zebrafish. Embryonic libraries from comparable developmental stages were normalised by oligonucleotide fingerprinting. These libraries and normalisation results were the basis for subsequent EST sequencing projects in USA and locally at the MPI (Poustka *et al.* 1999, Clark *et al.* 2001, Panopoulou *et. al.* 2003, Poustka et. al 2003).

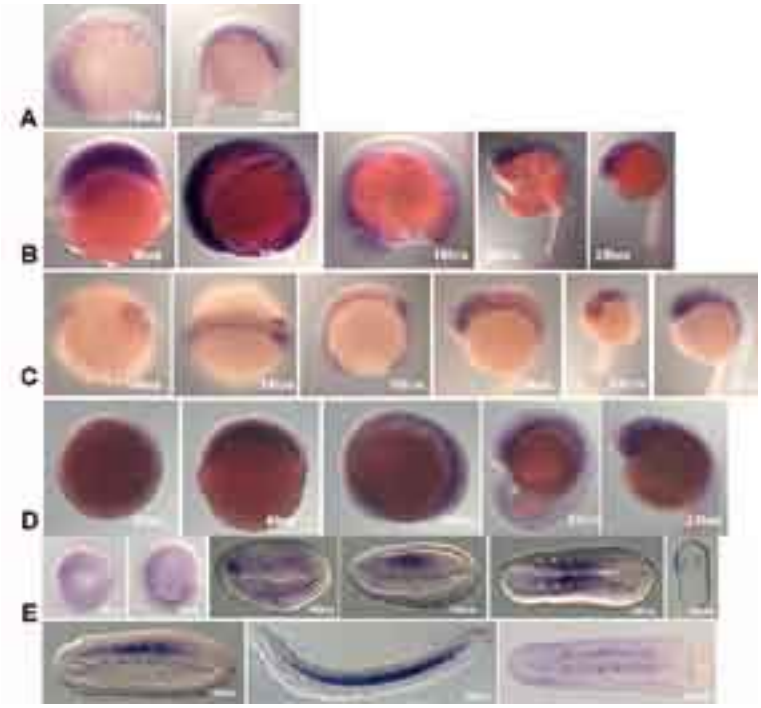


Figure 2: The zebrafish discs large genes (Dlg) (A-D) are expressed in cephalic mesoderm and in multiple regions of the central nervous system. The amphioxus single Dlg ortholog (E) is initially expressed in ectoderm (a region that includes the prospective neural ectoderm region) while later it switches into being expressed only in the somites (paraxial mesoderm).

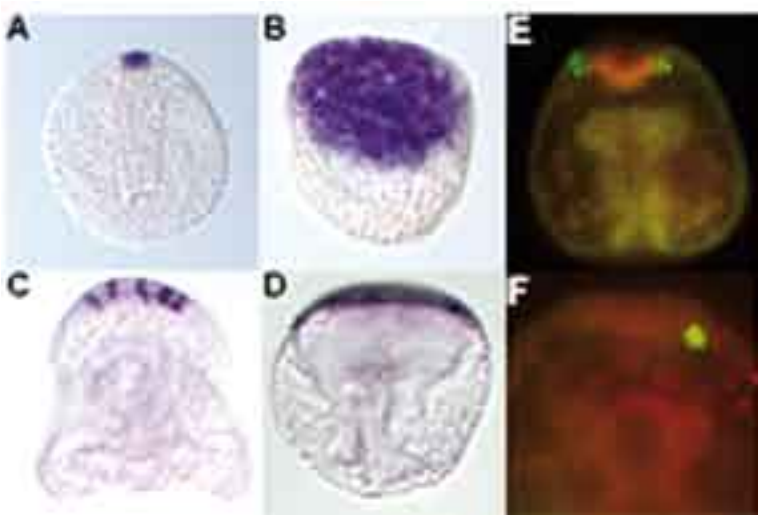


Figure 3: WMISH of the transcription factors FoxQ (A,B) and Mox (CD) in normal (A,C) and zinc treated (B,D) sea urchin embryos. Fox Q is normally expressed in about 20 cells of the apical organ, which harbours the neurons of the early embryo. Zinc treatment expands this neural tissue to more 200 neurons. Mox localisation in a normal late gastrula stage embryo, shows expression in serotonergic neurons (C). The gene was identified in a screen of Zinc treated embryos. WMISH of Mox in zinc treated embryos (D) surprisingly reveals that the treatment causes a 10 fold multiplication of serotonergic neurons in the embryo (E) Costaining with an anti-serotonin Ab (green) and WMISH (red) verifies that the transcription factor Mox is exclusively expressed in serotonergic neuroblasts whereas in (F) the transcription factor FoxQ marks the neural non-serotonergic cells. We are using this system now to analyse the gene regulatory network underlying the differentiation of serotonergic neurons.

During the last two years we have continued this effort of generating resources for marine evolutionary important organisms. In a project funded through the EU Marine Genomics Network of Excellence, we have generated 4 full length cDNA libraries from the mediterranean sea urchin *Paracentrotus lividus* 45,000 clones of these libraries have been already EST sequenced by Genoscope and another 100,000 are planned to be sequenced during this year. Finally, full length cDNA libraries have been constructed from the octopus, hagfish brain and ophoroid regenerating arms. Each of these libraries is in the size of 110,00 clones. All of them have been arrayed and printed on macroarrays.

Finally, we have invested a significant effort in annotating and organising our ESTs, WMISH and array screen information in databases to allow the maximum exploitation of results locally and by external users. The links to individuals databases are:

Ciona Intestinalis microarrays: *http://goblet.molgen.mpg.de/cgi-bin/webapps/ciona.cgi*

Sea urchin WMISH, arrays, ESTs: *http://goblet.molgen.mpg.de/cgi-bin/seaurchin-database.cgi*

Amphioxus ESTs, WMISH: *http://goblet.molgen.mpg.de/cgi-bin/Blast-amphioxus.cgi*

Future projects

Gene duplication and the evolution of regulatory elements: In collaboration with the department of Prof. M. Vingron we intend to detect conserved non-coding (candidate regulatory) elements between members of vertebrate gene families. To investigate whether the number of shared elements can be correlated with neo- subfunctionalisation we intend to initially concentrate on the 50 gene families for which we have studied their expression patterns in zebrafish.

Subsequently we intend to test whether these elements already existed in amphioxus and *Ciona Intestinalis*. Identified conserved non coding elements will be tested whether they are functional in *Ciona Intestinalis*.

More than 40 components of the sea urchin endomesoderm gene network are known. We plan identifying regulatory elements shared by the genes involved in this network.

Sea urchin apical plate: The apical organ contains the serotonergic nervous system of the sea urchin pluteus larva. The evolutionary relationship between the apical organ of dipleura type larvae and the dorsal central nervous system of chordates remains unresolved. The Garstang theory proposes that the ciliated band and the apical organ have moved dorsally during evolution in vertebrates and hence these cells are proposed to be of common origin. We have identified apical organ specific genes in sea urchin (Fig.3) and want to study the GRN defining this new territory in the sea urchin embryo and compare it to a GRN operating in the development of a vertebrate CNS in order to evaluate the homology of these neuronal domains.

Selected Information

Publications

- Panopoulou, G. and Poustka, A.J. (2005) Timing and mechanism of ancient vertebrate genome duplications – the adventure of a hypothesis. Trends Genet 21, 559-67.
- Panopoulou, G. and Poustka, A.J. (2005) Evolutionary analysis of Genes in "Encyclopedic Reference of Genomics and Proteomics in Molecular Medecine" (Ganten, D. editor) Springer Verlag
- Poustka, A.J., Kuhn, A., Radosavljevic, V., Wellenreuther, R., Lehrach, H. and Panopoulou, G. (2004) On the origin of the chordate central nervous system: expression of onecut in the sea urchin embryo. Evol Dev 6, 227-36.
- Poustka AJ, Groth D, Hennig S, Thamm S, Cameron A, Beck A, Reinhardt R, Herwig R, Panopoulou G and Lehrach H (2003). Generation, annotation, evolutionary analysis and database integration of 20,000 unique sea urchin EST clusters. Genome Res 13, 2736-46.
- Panopoulou G, Hennig S, Groth D, Krause A, Poustka AJ, Herwig R, Vingron M and Lehrach H. (2003). New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes. Genome Res 13: 1056-66.
- Minguillon, C., Jimenez-Delgado, S., Panopoulou, G. and Garcia-Fernandez, J. (2003) The amphioxus Hairy family: differential fate after duplication. Development 130, 5903-14.
- Stricker S, Poustka AJ, Wiecha U, Stiege A, Hecht J, Panopoulou G, Vilcinskas A, Mundlos S, Seitz V (2003). A single amphioxus and sea urchin runt-gene suggests that runt-gene duplications occurred in early chordate evolution. Dev Comp Immunol 27(8):673-84.

Theses

- Christoph Wiegreffe, Cloning and expression analysis of the BMP antagonist Chordin in amphioxus, Diploma thesis, Free University Berlin, 4/2005-12/2005
- Alexander Kühn, Cloning and Characterisation of Dickkopf in sea urchins: the head organizer gene in a headless animal, Diploma thesis, Free University Berlin, 10/2004-4/2005

Cooperations within the institute

- Prof. M.Vingron, Depart. Computational Molecular Biology (Identification of conserved non-coding elements)
- Prof. S. Mundlos, Development & Disease Group (Evolution of the Runt family of transcription factors)
- Dr. Claus Hultschig, Depart. Vertebrate Genomics (Ciona cDNA microarray construction)

External academic cooperations

- Prof. P. Holland, Dr. Seb Shimeld, University of Oxford, England (evolution of the zinc finger genes)
- Dr. Patrick Lemaire, LGPD, IBDM, Marseille, France (Ciona endomesoderm network)
- Prof. Eric Davidson, Division of Biology, California Institute of Technology, U.S.A. (sea urchin endomesoderm network)
- Prof. D. McClay, Department of Biology, Duke University, Durham, U.S.A. (sea urchin endomesoderm network)
- Mike Thorndyke, Kristineberg Marine Research Station, Sweden (Xenoturbella project)
- Prof. Thomas Hankeln, Johannes Gutenberg Universität Mainz, Institut für Molekulargenetik (evolution of the globin gene family)

Scientific referee for the journals

- Trends in Genetics
- Genome Biology
- BMC Evolutionary Biology

External Funding

- Marine Genomics, EU Network of Excellence

Cardiovascular Genetics

Head
Dr. Silke Sperling
Phone: +49 (0)30-8413 1232
Fax: +49 (0)30-8413 1699
Email: sperling@molgen.mpg.de

Scientists
Dr. Christina Grimm* (since 12/2002)
Dr. Andreas M. Rickert* (4/2003 - 12/2003)
Dr. Stefanie Hammer* (3/2004 - 3/2006)
Dr. Martin Hillebrand* (7/2003 - 11/2003)

Engineers
Raffaello Galli* (1/2004 - 12/2005)
Tammo Krüger* (since 3/2006)

Technician
Ilona Dunkel (since 1/2002)

PhD Students
Bogac Kaynak (1/2001 - 6/2005)
Dominik Seelow (11/2000 - 4/2004)
Martin Lange* (since 4/2003)
Jenny J. Fischer* (since 1/2004)
Martje Tönjes* (since 3/2006)

Diploma-/Master Students
Katrin Bach (5/2003 - 12/2003)
Jan Vogel (6/2003 - 1/2004)
Silke Grabow (10/2004 - 5/2005)
Antje Purmann (2/2005 - 9/2005)
Philip Tomann (1/2006 - 8/2006)
Markus Schüler* (since 5/2006)

* external project funding

Scientific Overview

The group consists of an interdisciplinary team with biologists, physicians and bioinformaticians and aims to analyse the molecular mechanisms underlying the cardiac developmental process in a systems biology approach. We focus on the transcriptional regulation process, which plays a key role for normal and abnormal cardiac development leading in the latter case to congenital heart diseases (CHD). Through the combinatorial nature of transcriptional regulation, a relatively small number of regulatory elements can cause a broad range of overlapping phenotypes, a characteristic of cardiac malformations. A large number of known cardiac transcription factors (TF) remain active in adult myocardium providing the opportunity to study the potential effect of their dysfunction through the analysis of myocardial samples in patients.

Collection of clinical and molecular data

Since the majority of CHDs can be traced to abnormalities in specific developmental milestones, the detailed phenotypical description of analysed heart defects remains the first step for their association with molecular disturbances. Therefore, we collected 600 patient samples (coop. German Heart Center Berlin), developed a taxonomy for CHDs, set-up a Cardiovascular Genetics database (CVGdb) and a software *d-matrix* for data visualisation and analysis. These tools are currently transformed to a European CVGdb within FP6-IP "Heart Repair".

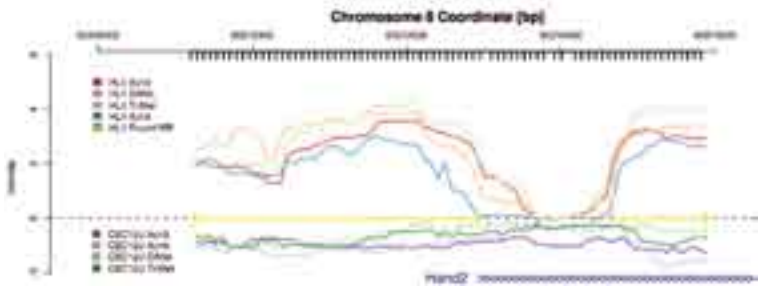


Figure 1: ChIP-chip of mouse HL1 and undifferentiated C2C12 cells. Shown are histone modifications of the promoter and first intron region of Hand2 gene. Hand2 is specifically expressed in HL1 where acetylation of histone 3 and 4 as well as histone 3 tri- and dimethylation are present, whereas it is not expressed in C2C12U where non of these modifications could be observed.

Transcriptional networks

In 2003, we have been able to show for the first time, disease specific gene expression profiles for Tetralogy of Fallot and single ventricular septal defects using genome-wide expression arrays. We are now analysing a broader phenotype collection to gain further insights into the transcriptional network of the heart. The underlying idea is to identify regulatory relationships of TFs and target genes based on gene expression profile disturbances and TF binding sites (coop. Dep. Vingron, U. Pape). We were able to extract transcriptional networks for atria and ventricle based on expression profiles of 50 TFs and targets obtained from 200 samples, and could confirm known relationships and furthermore suggest several new ones. Currently the expansion of these networks by implementing additional TFs and targets as well as experimental validation is in progress.

These data lead to our particular interest in the transcription factors Cited-2, TBX20 and DPF3. Mutation analyses in patients with CHDs as well as cell culture and mouse model experiments were performed.
To gain insights into the dynamics of histone modifications in the process of muscle cell differentiation and specification, and to combine the epigenetic modulation with gene expression and regulation through specific transcription factors, we established chromatin immunoprecipitation (ChIP) for skeletal and cardiac mouse muscle cells (C2C12, HL-1), set-up a muscle cell specific promoter array and are currently performing ChIP-chip experiments for several histone modifications and transcription factors (NimbleGen, coop. W. Huber at EBI).

CITED2

CITED2 encodes a CBREBBP/EP300 interacting transcriptional modulator of HIF1A and TFAP2. We showed for the first time the causative impact of CITED2 in the development of human CHDs, in particular of septal defects, studying 392 patients and 192 control individuals. We identified mutations leading to alterations of the amino acid sequence, which significantly reduce the capacity of CITED2 to transrepress HIF1A and diminish TFAP2C coactivation.

DPF3

DPF3 is a transcription factor containing C2H2-zinc and double PHD fingers and belongs to the family of D4 proteins. DPF3 knockout mice were analyzed and show a dilated cardiomyopathy phenotype (coop. V. Buchmann, Cardiff).
Expression analysis during embryonic development revealed that Dpf3 is expressed in the mesodermal stripes of the cardiac crescent at ED 7.5 and throughout the heart at later stages. In order to identify the upstream and downstream transcriptional pathway, we analyzed the promoter and enhancer of *Dpf3* and performed ChIP-chip studies. We show that *Dpf3* is regulated by cardiogenic TFs Nkx2.5, NR2F2 and Sall4 during development. LacZ-reportergene studies in mouse revealed an evolutionary conserved enhancer element of the human *DPF3* with an onset of activity tracing back to the first establishment of left-right asymmetry during heart development at E8.5 (Fig. 2). A single *Nkx2.5* binding site is necessary for specific interventricular septum (IVS) specific expression *in vivo*. In addition, the transcription factor *Sall4* could be identified as a repressor of the promoter. As this is the first enhancer element with specific IVS expression during cardiac development, it will be a very useful tool to express transgenes, perform fate mapping, and establish knockout models. The enhancer is now distributed to several other laboratories and we are currently setting up a *Cre*-line.

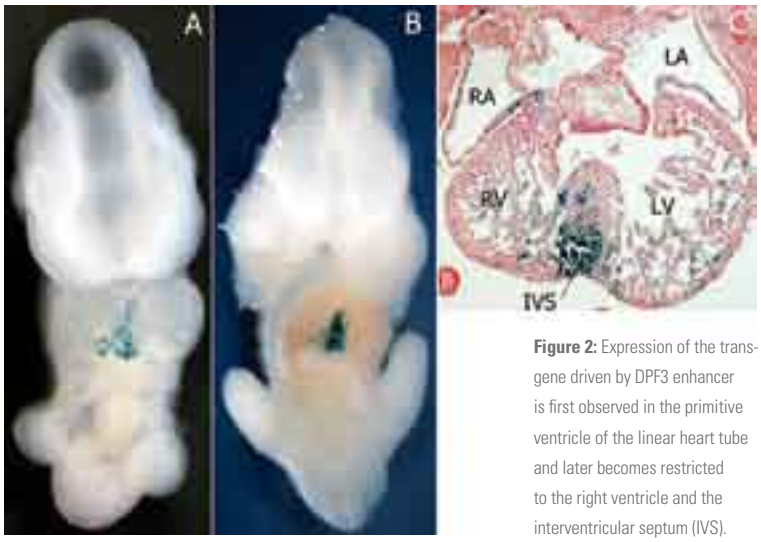


Figure 2: Expression of the transgene driven by DPF3 enhancer is first observed in the primitive ventricle of the linear heart tube and later becomes restricted to the right ventricle and the interventricular septum (IVS). X-gal stained transgenic embryos of Line1001.
A: E8.5
B: E11.5
C: transverse section of E12.5

Genomic clustering of mammalian transcriptomes

There has been much evidence recently for a link between transcriptional regulation and chromosomal gene order, but the relationship between genomic organization, regulation and gene function in higher eukaryotes remains to be precisely defined. Analysing the cardiac transcriptome identified by our genome-wide array analysis, FANTOM and GNF Symatlas datasets, we found striking evidence for such relationship. Highly co-expressed gene clusters are phylogenetically conserved, have a length limit and mainly consist of non-paralogous genes, and show a weaker functional and similar regulatory relationship to each other than general genomic neighbours. Thus, our data point to so far unknown *cis*-acting units and reject co-functionality as a driving force. We hypothesize that highly co-expressed gene clusters are essential for a higher order of transcriptional regulation, while specific TFs are likely to handle the fine-tuning of transcription on shorter time scales.

Further perspectives

Beside the short-term plans, a major goal will be the integration of the various data (e.g. genomic clustering and histone code) and the modelling of regulatory pathways in the heart. Further experiments of systematic knock-down/over-expression of TFs and target genes and the dissection of pathways, as for DPF3, are envisaged.

Selected Information

Publications

- Liu, H., H. Wang, Z. Shi, H. Wang, C. Yang, S. Sperling, W. Tan, and Z. Lu. (2006). TaqMan probe array for quantitative detection of DNA targets. Nucleic Acids Research 34: e4.
- Westhoff, T.H., S. Scheid, M. Tolle, B. Kaynak, S. Schmidt, W. Zidek, S. Sperling, and M. van der Giet*. (2005). A physiogenomic approach to study the regulation of blood pressure. Physiol Genomics.
- Vogel, J., A. von Heydebreck, A. Purmann, and S. Sperling. (2005). Chromosomal Clustering of a human transcriptome reveals regulatory background. BMC Bioinformatics 6: 230.
- Sperling, S., C.H. Grimm, I. Dunkel, S. Mebus, H.P. Sperling, A. Ebner, R. Galli, H. Lehrach, C. Fusch, F. Berger, and S. Hammer. (2005). Identification and functional analysis of CITED2 mutations in patients with congenital heart defects. Hum Mutat 26: 575-582.
- Carninci, P. T. Kasukawa S. Katayama J. Gough M.C. Frith N. Maeda R. Oyama T. Ravasi B. Lenhard C. Wells R. *et al.* (2005). The transcriptional landscape of the mammalian genome. Science 309: 1559-1563.
- von Heydebreck, A., S. Sperling, B. Kaynak, H. Lehrach, and M. Vingron. (2004). Genexpressionsanalyse komplexer klinischer Phenotypen mittels cDNS-Arrays - Gene Expression Profiling of Complex Clinical Phenotypes using cDNA-Arrays. it - Information Technology: 26-30.
- Seelow, D., R. Galli, S. Mebus, H.P. Sperling, H. Lehrach, and S. Sperling. (2004). d-matrix - database exploration, visualisation and analysis. BMC Bioinformatics 5: 168.
- Rickert, A.M., H. Lehrach, and S. Sperling. (2004). Multiplexed real-time PCR using universal reporters. Clin Chem 50: 1680-1683.
- Rickert, A.M., T.A. Borodina, E.J. Kuhn, H. Lehrach, and S. Sperling. (2004). Refinement of single-nucleotide polymorphism genotyping methods on human genomic DNA: amplifluor allele-specific polymerase chain reaction versus ligation detection reaction-TaqMan. Anal Biochem 330: 288-297.
- Kaynak, B., A. von Heydebreck, S. Mebus, D. Seelow, S. Hennig, J. Vogel, H.P. Sperling, R. Pregla, V. Alexi-Meskishvili, R. Hetzer, P.E. Lange, M. Vingron, H. Lehrach, and S. Sperling. (2003). Genome-wide array analysis of normal and malformed human hearts. Circulation 107: 2467-2474.

Book chapters

- Sperling, S. (2006). Array analysis applied to maformed hearts: Molecular Dissection of Tetralogy of Fallot. In Congenital Heart Disease: Molecular Diagnostics (ed. M. Kearns), pp. 233-246. Humana Press.
- Hammer, S. and S. Sperling. (2006). Microarrays for the transcriptional regulation process. In Biochips Nanotechnology (ed. H. Nongyue), pp. to appear. American Scientific Publishers.

Memberships

- Member of the FANTOM Consortium (since 2004)
- Board Member of the European Society of Human Genetics (2005-2010)
- Board Member and Workpackage Coordinator of the FP6-IP "Heart Repair" (2005-2009)

Patents

- Deutsches Patent # 103 35 359.3: Verfahren und Vorrichtung zur graphischen Darstellung

Teaching

- Free University of Berlin, Department of Bioinformatics, Series of lectures: Transcriptional regulation: theory and methods (2004 and 2005)
- Otto Warburg International Summer School and Workshop 2005 on Networks and Regulation, August 2005, Berlin, Germany

Cooperations

- Richard P. Harvey, Victor Chang Cardiac Research Institute, Darlinghurst, Australia
- Benoit G. Bruneau, The Hospital for Sick Children, Toronto, Canada
- Marthy L. Bulyk, Harvard Medical School, Boston, USA
- Wolfgang Huber, EMBL-EBI, Hinxton, UK
- Vladimir Buchmann, Cardiff University, UK
- Felix Berger, German Heart Center, Berlin, Germany

Work as Scientific Referee

Journals:

- NAr, BMC Bioinformatics, Molecular Medicine, Circulation Research

Funding agencies:

- GIF, Estonia Genome Project

External Funding

- BMBF BioProfile: Metabolisches Syndrom Nanowell PCR
- NGFN1: Entwicklung einer Patientenmatrix
- EU FW6: Moltools, HeartRepair

Chromosome 21, Gene Expression and Regulation

Head
Dr. Marie-Laure Yaspo
Phone: +49 (0)30-8413-1356
Fax: +49 (0)30-8413-1128
Email: yaspo@molgen.mpg.de

Scientists
Dr. Ilaria Piccini*
Dr. Alia Ben Kahla (until 31-10-2004)

Technician
Cindy Springer*

Student
Dominic Schmidt*

Engineers
Daniela Balzereit*
Emilie Dagand*
Asja Nürnberger*
Sabine Schrinner*
Reha Yildirimann

Graduate Students
Alon Magen
Marc Sultan
Robert Querfurth*
Hans-Jörg Warnatz*

* external project funding

Scientific Overview

Following up our contribution to the DNA sequencing and annotation of human chr.21 (Hattori *et al.* Nature 405. 2000), we focused on primate genomics, proteomics and expression analysis of chr.21 genes, and on molecular genetics of trisomy 21 or Down syndrome (DS). DS is a complex congenital disorder affecting 1:700 live births, and a leading genetic cause of mental retardation. The identification of genes and molecular mechanisms responsible for the DS phenotypes has therefore been a high priority for genome research, motivating a large fraction of the molecular analysis of the approx. 300 protein-coding genes on chr.21. More recently, we have expanded our interest to the analysis of genome-wide gene expression patterns and regulation networks.

Comparative sequencing in primates (NGFN1/NGFN2)

We participated to the DNA sequencing and analysis of the equivalent of chr.21 in chimpanzee in a consortium coordinated by RIKEN. Data showed different expansion of subfamilies of retrotransposons between the lineages, suggesting different impacts of retrotranspositions on human and chimp evolution. 1.44% of the chromosome consists of single-base substitutions in addition to 68,000 insertions or deletions, generating changes in most of the proteins (Watanabe *et al.* 2003). Present projects include 1) identification of chimp/human-specific splice variants, which may contribute to the gene expression differences seen between the two species and 2) finishing 15 Mb of sequence of chimp chr. X (coop. R. Sudbrak, R. Reinhardt). We are also using phylogenetic shadowing in apes and monkeys for identifying primate/human specific regulatory elements in promoter regions of chr.21 genes that might contribute to differential gene expression in distinct clades. This work will be integrated into the EU project APES, starting end of 2006 (in co-operation with R. Sudbrak)

Proteomics of chr.21 (EU-T21targets)

In order to get new insights into the molecular function of chr.21 proteins, we identified protein-protein interactions (PPIs), forming the basis of a chr.21 protein interactome by 1) using an automated yeast two-hybrid (Y2H) mating-array set-up (coop. E. Wanker) and 2) screening the literature and protein interaction databases, BIND and HPRD, and two Y2H studies (Rual *et al.* 2005; Stelzl *et al.* 2005). We tested most of the chr.21 open reading frames by Y2H, identifying 57 novel PPIs for 23 chr.21ORFs, and we retrieved 634 additional PPIs from public databases. We built a chr.21 interactome network by computationally integrating these 691 direct interactions for 108 HSA21 proteins into a meshwork comprising 14,949 PPIs between 5,655 proteins (indirect PPIs) retrieved from databases (coop. M. Lappe). Data enabled to propose a number of novel functions for chr.21 proteins and their integration into larger biochemical networks (Warnatz *et al.* submitted) (Fig. 1). Data are made available in a database and offer a resource for studying pathways dysregulated in DS.

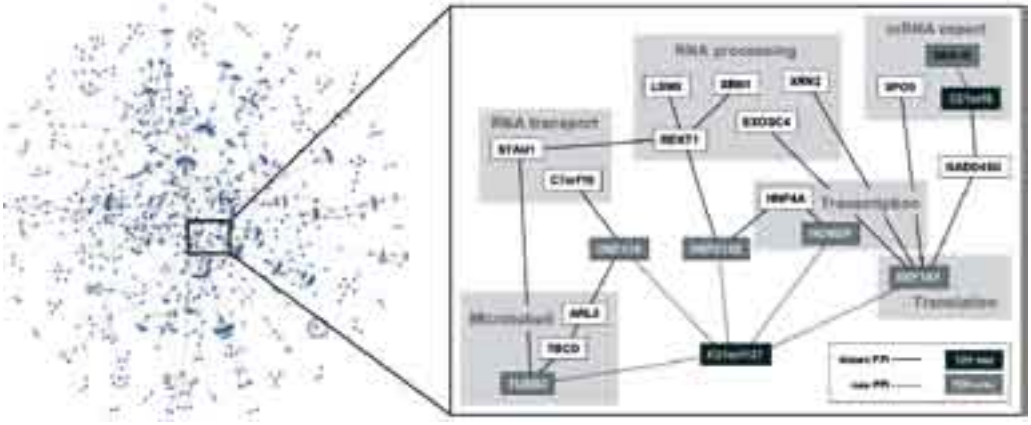


Figure 1: Chr21 interaction network and a subnetwork involving unknown proteins C21orf127 and C21orf6 in mRNA processing and transport.

Analysis of gene regulatory networks by RNA interference and ChIP-on-chip (NGFN2)

We plan to identify gene targets and associated genetic networks for ca. 150 human transcription factors (Tfs) involved in key developmental processes and pathologies. In this, we combine RNAi and expression profiling on microarrays (Fig. 2). We are comparing data obtained with different silencing molecules: synthetic short interfering RNA (siRNAs) and esiRNA (endoribonuclease-prepared siRNA, coop. with F. Buchholz) for minimizing off-target effects. More than 100 Tfs have been knocked-down in Hek 293T cells. For example, we identified a high-confidence set of 77 target genes for BACH1 (BTB and CNC homology 1, basic leucine zipper Tf 1) that includes HMOX1 (Heme oxygenase 1), a known target of this Tf. Focusing on the Tfs mapping to chr.21, we are identifying DNA binding sites in Hek 293T cells by chromatin immunoprecipitation coupled with DNA microarrays (ChIP-on-chip). Known targets of BACH1 and GABPA (GA binding protein Tf, α subunit), e.g. HMOX1 and TFAM (mitochondrial TfA) respectively, were enriched by ca. 20 fold, and the promoter chip analysis is ongoing. We are analysing Tf binding motifs in coop. with the department of M. Vingron.

Figure 2: Deciphering gene regulation networks

Expression profiling in a mouse model of Down Syndrome (NGFN1 and EU-T21targets)

Using Ts65Dn, a well-established mouse model of DS, we carried out gene expression profiling by cDNA arrays and quantitative real-time PCR (qRT-PCR) (coop. R. Reeves). We focused on the chr.21 genes orthologs because they are the primary contributors of trisomy. We analysed RNAs from several tissues of Ts65Dn and controls, either as pools or as individual samples. With pools, we observed a trend of 1.5 fold overexpression for the majority of trisomic genes with however exceptions to this rule for a few genes (Kahlem *et al.* 2004). For many genes, it is unlikely that such a modest change in gene expression levels will have drastic effects on the fitness of the organism. To select genes, where slight changes in gene activity could have a more dramatic phenotypic effect, we focused on the normal variation of gene expression levels. Genes whose level of gene expression is critical are more likely to be tightly controlled than those, for which slight variation of expression will not have a detrimental effect. Our working hypothesis is that genes relevant for DS reach an expression threshold in trisomic individuals that is not or rarely encountered in controls. Hence, we determined inter-individual expression differences for 50 chr.21 gene orthologs in the cortex, midbrain and cerebellum of individual Ts65Dn mice and controls by qRT-PCR (TaqMan). Our study enabled the identification of a short list of potential key contributor genes of the trisomy phenotypes in the brain. Among those, we found App, the amyloid precursor protein (Sultan *et al.* submitted).

Gene expression patterns in the mouse embryo by large-scale *in situ* hybridisation (EU- coop. with . S. Mundlos)

EURExpress is a EU-integrated project, (FP6, start. Jan. 2005, coord. by A. Ballabio, see list of partners in the cooperations list) for generating expression data for > 20,000 mouse genes by RNA *in situ* hybridization (ISH) on sagittal sections from E14.5 wild type C57Bl6 embryos. A "transcriptome atlas" describing gene expression patterns is being generated using an automated RNA *in situ* hybridization system (GenePaint), where experimental procedures, data collection and display have been standardized. Images are scanned automatically and annotated centrally (Fig. 3). Our laboratory is involved in:

1) Selection of genes and clones
2) Template generation, quality controls, distribution
3) Establishing infrastructure for template data management and storage
4) ISH pipeline

We set up automatic routines for the selection of cDNA clones for all mouse genes, matching ENSEMBL and REFSEQ transcript sequences to mouse ESTs. So far, clones for 9,358 mouse genes have been selected and picked (RZPD).

We established standard operating procedures for template distribution. 7,000 templates have been distributed among the five ISH units. Our lab coordinates the flow of templates in coop. with the TIGEM. We generate the genomic metadata files, which are then loaded to the tracking database and in our own database (*http://goblet.molgen.mpg.de/cgi-bin/eurexpress*) for data storage and visualization.

We aim at producing expression patterns for ~4,000 genes. We process routinely 30 genes/week and we have generated high-resolution expression data for 850 genes. 62% of the genes show a predominant or restricted expression pattern. Data are disseminated via web interface (*www.eurexpress.org; www.genepaint.org*), allowing integration with existing European Projects (e.g. mouse mutagenesis and phenotyping).

Future directions

In the next integrated project (ANEU grant, start Fall 2006, coord. by S. Antonarakis), we will continue our DS studies on larger populations in man, mouse and cellular models surveying the global transcriptome. We are also involved in the analysis of gene regulation networks project using consomic strains, a component of the ANEU grant for the chr.21 genes. Integration of complex gene expression data together with RNAi will give novel insights into genetic networks involving chr.21 genes. We are using systems biology in coop with the group of R. Herwig, to exploit the various datasets for charting and predicting regulatory networks dysregulated in trisomy.

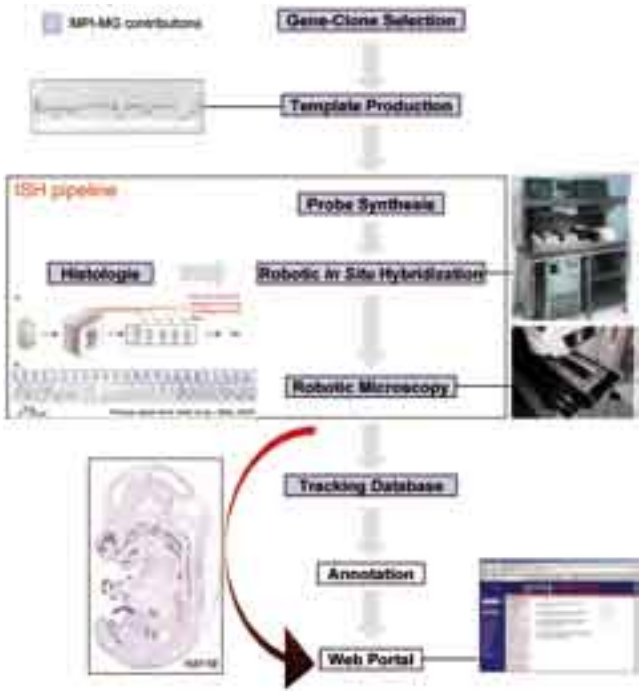


Figure 3: EURExpress workflow

Selected Information

Publications

- Adjaye J, Huntriss J, Herwig R, BenKahla A, Brink TC, Wierling C, Hultschig C, Groth D, Yaspo ML, Picton HM, Gosden RG, Lehrach H. Primary differentiation in the human blastocyst: comparative molecular portraits of inner cell mass and trophectoderm cells. Stem Cells. 2005 (10):1514-25.
- Hu, Y-H., Warnatz, H-J., Vanhecke, D., Wagner, F., Fiebitz, A., Thamm, S., Kahlem, P., Lehrach, H., Yaspo M-L., Janitz, M. Cell array-based intracellular localization screening reveals novel functional features of human chromosome 21 proteins. BMC Genomics. 2006 7(1):155.
- Kahlem P, Sultan M, Herwig R, Steinfath M, Balzereit D, Eppens B, Saran NG, Pletcher MT, South ST, Stetten G, Lehrach H, Reeves RH, Yaspo ML. Transcript level alterations reflect gene dosage effects across multiple tissues in a mouse model of down syndrome. Genome Res. 2004 (7):1258-67.
- Kittler R, Putz G, Pelletier L, Poser I, Heninger AK, Drechsel D, Fischer S, Konstantinova I, Habermann B, Grabner H, Yaspo ML, Himmelbauer H, Korn B, Neugebauer K, Pisabarro MT, Buchholz F. An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division.. Nature. 2004 432 (7020):1036-40.
- Watanabe H, Fujiyama A, Hattori M, Taylor TD, Toyoda A, Kuroki Y, Noguchi H, BenKahla A, Lehrach H, Sudbrak R, Kube M, Taenzer S, Galgoczy P, Platzer M, Scharfe M, Nordsiek G, Blocker H, Hellmann I, Khaitovich P, Paabo S, Reinhardt R, Zheng HJ, Zhang XL, Zhu GF, Wang BF, Fu G, Ren SX, Zhao GP, Chen Z, Lee YS, Cheong JE, Choi SH, Wu KM, Liu TT, Hsiao KJ, Tsai SF, Kim CG, OOta S, Kitano T, Kohara Y, Saitou N, Park HS, Wang SY, Yaspo ML, Sakaki Y. DNA sequence and comparative analysis of chimpanzee chromosome 22. Nature. 2004 429(6990):382-8

Invited Plenary lectures

- M-L Yaspo, Human Genome Meeting, April 2004, Berlin, Germany, "Chromosome 21 and Down Syndrome"

Membership

- HUGO member (Human Genome Organization)

Public Relations

- Contribution to the article "Der Kleine Unterschied" Geo Kompakt no. 7, 2006, "Der Mensch und seine Gene"

Organization of Scientific events

- NGFN-RNAi Workshop (coop. with RiNA, Berlin) May 15-18th, 2006 MPI Molecular Genetics, Berlin, Germany, M. Sultan, I. Piccini and M-L Yaspo, Analysis of transcription regulatory networks by RNA interference

Work as Scientific referee

- Referee for journals (Science, Genome Biology, Genomics, Trends in molecular medicine) and funding agencies (Austrian Science Fund)

External Funding

- NGFN 1: Core Facility (until 2004)
- NGFN 2: SMP RNAi, SMP DNA
- EU: T 21 Targets, EURExpress

Associated Group: Gene Traps and Microarrays – Molecular Analysis of Heart Failure

Head
Prof. Dr. Patricia Ruiz
Phone: +49 (0)30-450578 744
Fax: +49 (0)30-450525 901
Email: ruiz@molgen.mpg.de

Scientists
Dr. Karin Effertz
Dr. Carsten Horn
Dr. Markus Irgang*

Scientific Manager
Dian Michel

Technicians
Carola Geiler
Judith Grosser
Julia Pfitzner
Aydah Sabah
Beata Thalke*
Carsta Werner

Graduate Students
Jörg Isensee
Cathrin Manefeld
Luca Meoli
Henning Witt
Safak Yalcin
Valeria Zazzu

Students
Lena Oertzen
Christoph Nabzdyk
Amaya Vaz Perez

Secretrary
Jutta Bebla

* external project funding

Scientific Overview

Functional analysis of the mammalian genome by large-scale gene trap mutagenesis

A major challenge of the post-genomic era is the functional characterization of every single gene within the mammalian genome. In an effort to address this challenge, we have assembled a collection of mutations in mouse embryonic stem (ES). Using four different gene trap vectors, we have generated sequences adjacent to the gene trap integration sites (gene trap sequence tags, GTSTs) from over 19,242 ES cell clones and trapped 3,723 unique genes and ESTs. A detailed analysis led to the conclusion that the most effective way to tag most of the mouse genome with gene trap insertions is by using a combination of gene trap vectors. Therefore and to overcome embryonic lethal phenotypes we have developed and optimized vectors that preferentially trap secreted proteins as well as conditional vectors. Moreover we have established PCR-based semi-automated strategies to identify the insertion site of the vectors at transcriptional (5` RACE) as well as at genomic level (splinkerrette PCR; Figure 1).

Altogether we have provided evidence that gene trapping allows a large-scale and cost-effective production of ES cell clones with mutations distributed throughout the genome. In particular, we are currently analysing the phenotype of two lines of mutant mice with gene trap insertions into the plakophilin 4 (PKP4) and the ect-2 gene, respectively. Whereas the latter leads to an embryonic lethal phenotype during early embryogenesis due to a failure in mesoderm development, mice with a mutated PKP4 gene are viable and fertile, but develop cardiac hypertrophy with age.

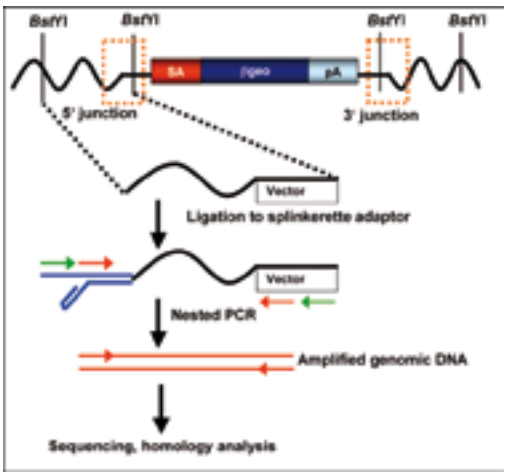


Figure 1

Identification and characterization of genetic modifiers of heart failure

Cardiomyopathies are heart muscle disorders characterised by chamber dilation often accompanied by wall thinning, severe systolic and diastolic dysfunction and frequently heart failure. Although several genes were found to be differentially expressed the cellular and molecular basis of the disease still remains poorly understood. We therefore performed expression profiling on cardiac samples from human idiopathic cardiomyopathy patients and from five different mouse disease models: muscle LIM protein- (MLP), ErbB2-, ErbB4- and plakoglobin-defective mice and doxorubicin treated ones. Notably, model-, i.e. pathway-specific gene

expression signatures could be identified. More importantly, the gene expression pattern of the MLP-knock-out mice most closely resembled the gene expression profiles obtained from human end-stage disease samples. Upon these results we speculate that the disruption of cytoarchitecture, cell junctions and/or signalling in cardiomyocytes may independently generate stress conditions via deregulation of different sets of genes. Consequently, myocardial dilation might develop as a passive reaction, as opposed to resulting from an active remodelling process of the heart.

Furthermore, we aimed at identifying genes involved in cardiac disease and regeneration processes. Therefore, the newt as a model organism for regeneration processes was chosen, since it is the evolutionary youngest organism still capable of regenerating different tissues. Moreover, its four chambered heart resembles better the mammalian heart anatomy than other less developed organisms, such as the zebrafish. We hybridized a newt cardiac cDNA library (Prof. T. Braun, Bad Nauheim) against complex probes generated from limb, tail and cardiac tissues at different time points of regeneration (in collaboration with Claus Hultschig). Indeed, we were able to distinguish different genetic programs during the regeneration processes in the different organs (Fig.2).

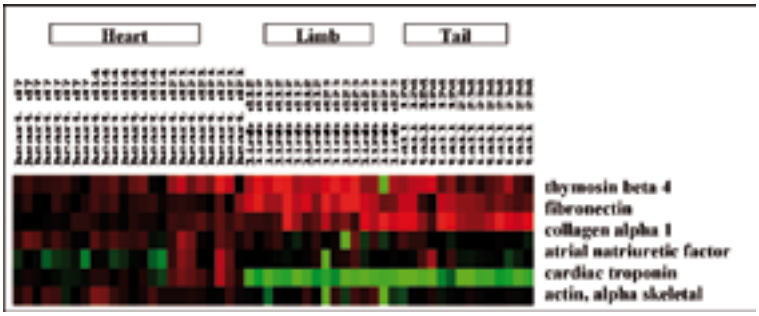


Figure 2

To analyse protein cascades and to identify protein interaction partners of ErbB2 we screened a mouse embryonic and a human cardiac library using a yeast-2-hybrid system. Hereby we identified a novel ErbB2 adaptor protein (Nck2), confirmed and characterized their interaction in mammalian cells and have functionally analysed the relevance of the ErbB2-Nck2 interaction in cell migration assays.

The molecular basis of gender differences in cardiovascular diseases

There are numerous health problems that are affected by gender. Women are more susceptible than men to depression, osteoporosis, asthma, lung cancer due to smoking and autoimmune disease. Still little is known about the molecular basis for this difference in e.g. cardiovascular disease. To tackle this question we have generated age- and gender-specific gene expression profiles of cardiac tissue of mice and human samples. Therefore probes from young and old male and female mice were hybridized to microarrays containing 8,600 cDNA clones, respectively. Furthermore, probes generated from young (<40 years) and older (50-60 years) men and women were hybridized to microarrays containing 11,000 cDNA clones, respectively. Several differentially regulated genes were validated using RT-PCR approaches. Data analysis was performed grouping the corresponding genes into functional categories and age- and gender-specific genes could be identified (e.g. osteoprotegerin and carbonic anhydrase).

Estrogens play a major role in molecular processes of gender-specific pathomechanisms and represent one of the main causes for differences in etiology of cardiovascular disease between men and women. They regulate gene expression via nuclear receptors (ERs) acting as ligand-activated transcription factors. However, estrogens also exert rapid non-genomic effects mediated by membrane-bound variants of ERs or receptors of other type. Recently, it was shown that Gpr30 binds estradiol with high affinity and mediates estrogen signals in tissue culture. In order to decipher the *in vivo* function of Gpr30, we are currently analyzing Gpr30 knock-out mice generated by targeted gene disruption. Reporter gene expression shows predominant Gpr30 expression in blood vessels of several organs. To further unravel the signaling pathway of Gpr30 we performed yeast-2-hybrid screening and identified several interaction candidates.

Selected Information

Publications

- Ruiz P and Witt H. (2006). Microarray analysis to evaluate different animal models for human heart failure. J Mol Cell Cardiol. 40: 13-15.
- Schnutgen F, De-Zolt S, Van Sloun P, Hollatz M, Floss T, Hansen J, Altschmied J, Seisenberger C, Ghyselinck NB, Ruiz P, Chambon P, Wurst W, von Melchner H. (2005). Genomewide production of multipurpose alleles for the functional analysis of the mouse genome. Proc. Natl. Acad. Sci. USA. 102: 7221-7226.
- Schupp M, Clemenz M, Gineste R, Witt H, Janke J, Helleboid S, Hennuyer N, Ruiz P, Unger T, Staels B, Kintscher U. (2005). Molecular characterization of new selective peroxisome proliferator-activated receptor gamma modulators with angiotensin receptor blocking activity. Diabetes. 54:3442-52.
- Skarnes WC, von Melchner H, Wurst W, Hicks G, Nord AS, Cox T, Young SG, Ruiz P, Soriano P, Tessier-Lavigne M, Conklin BR, Stanford WL and Rossant J (2004). A public gene trap resource for mouse functional genomics. Nature Genetics, 36: 543-544.
- Hansen J, Floss T, van Sloun P, Füchtbauer EM, Vauti F, Arnold HH, Schnütgen F, Wurst W, von Melchner H, and Ruiz P. (2003)..A large scale, gene-driven mutagenesis approach for the functional analysis of the mouse genome. Proc. Natl. Acad. Sci. USA, 100: 9918-9922.

Work as a scientific referee

- Reviewer of "The Journal of Molecular Medicine" and of "Cardiovascular Research".
- Evaluator activities for the World Cancer Research Fund International, the Swiss Cancer League and the European Commission (6th Framework Programme).
- Communicating editor of the journal "Molecular Genetics and Genomics" (MGG).

Patents

- Application/Patent No. PCT/EP02/12522 "Novel Markers for Cardiopathies"
- Application/Patent No. 050870ep/JH/PCH/ml "Gene Trap Cassettes for Random and
- Targeted Conditional Gene Inactivation"

Teaching

Since 2001

- Paris-Lodron University Salzburg (Austria) and at the Charité Univ. Medizin Berlin
- Lectures on Mouse Models in the Analysis of Human Diseases

Department of Human Molecular Genetics

Head
Prof. Dr. H.-Hilger Ropers
Phone: +49 (0)30-8413 1240
Fax: +49 (0)30-8413 1383
Email: ropers@molgen.mpg.de

Secretary
Hannelore Markert
Phone: +49 (0)30-8413-1241
Fax: +49 (0)30-8413 1383
Email: markert@molgen.mpg.de

Department of Human Molecular Genetics

Introduction

During the past 5 to 10 years, national and international funding agencies as well as the pharmaceutical industry have invested massively into the search for genetic risk factors in complex disorders, but so far with very limited success. It is becoming increasingly clear that common risk factors with a high effect on the morbidity are the exception rather than the rule, and that very few of the ~5 million known frequent single nucleotide polymorphisms (SNPs) predispose to disease.

Moreover, it has turned out that common disorders are often not multifactorial, but consist of a heterogeneous group of different diseases, and that many of these are due to defects of single genes. Indeed, the elucidation of Mendelian disorders has contributed very significantly to our current knowledge about the pathogenesis of various common disorders, such as Alzheimer, Parkinson, epilepsy, diabetes and cancer. Together, this has led to a renewed interest in monogenic disorders, as documented by recent publications in leading journals (e.g., see Beckman and Antonarakis, Nature Rev. Genet. 2006) and by the recent decision of the NIH to fund research into their etiology, pathogenesis and therapy.

In view of the predictable (and predicted) conceptual and practical difficulties complicating the genetic investigation of multifactorial disorders and the limited diagnostic and prognostic relevance of such studies, we have decided early on not to jump on that bandwagon but to scale up our research into monogenic disorders instead. In the Eighties and early Nineties, my group was one of the first in Europe to successfully employ positional cloning strategies for the molecular elucidation of single gene disorders, with a focus on X-linked gene defects causing blindness, deafness and mental retardation (MR). In the mid-Nineties, our decision to concentrate on MR was largely motivated by the overriding socio-economic importance of this common disorder (e.g., see Ropers and Hamel, Nature Rev.Genet. 2005) and by the expectation that its molecular elucidation would provide new insight into the function of the human brain. Presently, much of the research of our department focuses on the identification of genetic factors for mental disability and related disorders, and on the function of these genes in health and disease.

Scientific Overview

Recruitment of patients and families

As discussed previously (see Research Report 2003, p. 76), the recruitment of patients and families as well as their clinical examination has become the most important limiting factors for the elucidation of hereditary diseases. Already in 1995, we have therefore joined forces with other groups from France, Belgium and the Netherlands to found the European MRX Consortium. This has been instrumental in recruiting and investigating more than 600 families with X-linked MR, which is still the largest cohort of its kind. Moreover, close collaboration with a Danish group led to the establishment of the Mendelian Cytogenetic Network, which provided the basis for recruiting large cohorts of patients with disease-associated balanced translocations (DBCRs) and paved the way for their systematic clinical, cytogenetic and molecular characterization.

In 2003, contacts with a leading group from Tehran have led to a formalized collaboration aiming at the systematic elucidation of familial forms of MR. So far, our Iranian partner has collected more than 200 consanguineous families with two or more affected children. This continuously expanding collection is particularly valuable for studying the molecular causes of autosomal recessive MR, which are still largely unexplored. Also in 2003, a Max Planck Partner Group was established at the University Department of Medical Genetics in Poznan (Poland). This group has also been actively recruiting patients with DBCRs, families with XLMR and various other genetic disorders. Our fruitful, mutually beneficial collaboration with this group will persist even though funding of this group has recently expired.

More recently, through collaboration with another group from Copenhagen, we have obtained access to several hundred patients with MR and dysmorphic signs, but apparently normal karyotype, which are being screened for submicroscopic deletions and duplications by high-resolution array CGH (see below). In parallel, our array CGH technology has been transferred to the Department of Medical Genetics at the Charité, headed by Stefan Mundlos, for analogous collaborative studies into a variety of dysmorphic conditions. Encouraging preliminary results have prompted us to approach several other groups to look for genomic imbalances in a variety of different, not necessarily monogenic conditions, and several of these studies are already well underway.

The local recruitment and clinical characterization of patients and families with MR and other hereditary disorders is in the hands of two clinical geneticists who are partly supported by the Max Planck Society through a Tandem Project grant (see report Tzschach). Experienced clinical geneticists are also indispensable as competent local counterparts for external collaborations, as discussed above.

Identification of gene defects and genetic risk factors

Most mild forms of mental retardation (IQ 70 to 50) are thought to represent the lower end of the normal intelligence spectrum, which means that they are often multifactorial. In contrast, severe forms (IQ<50) are generally due to catastrophic events, and most of these are caused by specific genetic defects.

To identify these defects, we have employed a variety of complementary strategies. Apart from the systematic characterization of disease-associated balanced chromosome rearrangements (DBCRs), which has been very successful in identifying autosomal dominant and X-linked candidate genes for MR and other conditions (see report Kalscheuer) and the equally productive search for mutations in defined X-chromosomal intervals to identify molecular causes of non-syndromic X-linked MR, our Iranian contacts and support from the Max Planck Society enabled us for the first time to search systematically for autosomal recessive causes of MR (see report Kuss/Ropers). Another milestone was the generation of very dense BAC contig arrays comprising 36.000 individual clones, which has provided the basis for detecting submicroscopic deletions and duplications as small as 100 kb in the genome of patients and tumors (see report Ullmann).

Between 2003 and 2006, our group contributed significantly to the identification of novel genes for non-syndromic X-linked MR (NS-XLMR), and as a whole, the Euro-MRX Consortium has been involved in the identification of more than 60 percent of the NS-XLMR genes known to date. There is reason to believe that mutations in these 25 genes account for about half of the molecular defects underlying non-syndromic XLMR (reviewed by Ropers, Curr. Opin. Genet. Dev. 2006). Breakpoint mapping and sequencing in patients with DBCRs has been streamlined by preparative sorting of derivative chromosomes followed by hybridization to a high resolution BAC array, and genome-wide array CGH screening has been introduced to rule out submicroscopic imbalances which are not uncommon in these cases (see reports Ullmann, Kalscheuer and Tzschach). SNP array-based homozygosity mapping in >100 consanguineous families with autosomal recessive MR (ARMR) from Iran has enabled us to quadruplicate the number of known loci for this condition and to identify at least one novel ARMR gene. However, the most unexpected results were achieved by high-resolution array CGH (see report Kuss/Ropers). Specific genomic imbalances were not only found in patients with syndromic forms of MR, but also in a wide variety of other disorders such as autism and amyotrophic lateral sclerosis. Several of these changes were confined to patients and arose apparently de novo (see report Ullmann). This argues against the assumption that most risk factors for multifactorial disorders are evolutionarily old and may explain why in many of these conditions, association studies have failed.

Functional studies

For various aspects of our research, functional expertise is required. Knowing the relevant metabolic and regulatory pathways is a major asset when selecting candidate genes from linkage or deletion intervals for mutation screening. Choosing the right candidates for MR is a particular challenge, given the complexity of the brain and our limited knowledge about its function. Moreover, functional testing is often necessary to prove the pathogenetic relevance of sequence variants found in candidate genes, and finally, understanding the function of these genes in health and disease is a prerequisite for any form of therapy.

To expand and complement the already existing know-how in this field, e.g. in the groups of S. Schweiger, D. Walther , V. Kalscheuer, C. Scharff and H. Scherthan (see Group Reports as well as Scientific Report 2003, pg. 78), we have attracted another group leader with a broad spectrum of relevant techniques (see report of T. Hucho)). However, given its multifaceted character, research into the function of disease genes can never keep track with their identification. Therefore we have also established several extramural collaborations, e.g. with S. Sigrist (ENI, Göttingen) and H. Kawabe from the group of N Brose (MPI for Experimental Medicine, Göttingen) for the generation and study of Drosophila and mouse models, respectively. Moreover, we are collaborating with potent groups that have a primary interest in specific disorders and both the know-how and the infrastructure to perform follow-up studies on candidate genes identified in our laboratory.

Outlook

For many aspects of our work, such as the handling of large bodies of data resulting from whole-genome array CGH experiments and other approaches, bioinformatics is the key. Another future challenge for bioinformaticians will be high speed-low cost sequencing, which provides a fascinating new perspective for mutation screening and genome analysis in general. In view of the rapidly increasing knowledge about human genes and proteins, bioinformatics will also be decisive for functional studies, e.g. for the identification of functional candidate genes in defined genomic intervals, and large, comprehensive databases will be required to integrate the molecular and clinical data generated by the various large-scale approaches. Therefore, we have advanced plans to establish an Applied Bioinformatics group in our department, which will also strengthen our ties with the Department of Computational Biology.
Based on our own experience, e.g. the direct interaction between the products of two genes for Rett syndrome (Tao *et al.*, 2004; see report Kalscheuer), we expect that in the long run, large-scale gene finding will identify series of interacting genes and proteins that play a role in MR and related disorders, and that this will diminish the need for 'wet' functional studies. Finally, there is reason to believe that the application of our various strategies to common, apparently multifactorial disorders such as autism, schizophrenia and congenital malformations will dissect many of these into separate monogenic entities, and/or lead to the identification of genetic risk factors and modifiers that association studies have failed to detect.

Selected Information

Scientific honors and memberships

- Ropers, H.-H.: Elected member of HUGO Council 2003; re-elected 2006 and member of HUGO Scientific Program Committee
- Ropers, H.-H.: Elected member of the Royal Netherlands Academy of Arts and Sciences and the Berlin-Brandenburg Academy of Sciences
- Kuss, A.W.: German Genetics Society (GfG), German Society for Animal Breeding (DGfZ), German Society for Immunology (DGfI)
- Kuss, A.W.: International Society for Animal Genetics (ISAG)

Prizes

- Shoichet, S.: Promotionspreis der Berliner Wissenschaftlichen Gesellschaft für Erforschung der genetischen Ursachen von Hirnfunktionsstörungen (2005)
- Haesler, S.: Schloessman-Preis; Beitrag eines Gens (FoxP2) zur Entwicklungsregulation des Gesanglernens bei Zebrafinken (2005)
- Shoichet, S.: Otto-Hahn-Medaille für Untersuchungen zur Aufklärung der molekularen Ursachen von erblichen Hirnfunktionsstörungen beim Menschen (2004)
- L. Musante et al.: The European Society of Human Genetics / Nature Publishing Group Award (2003)

Professorships accepted by former department members

- Ulrike Nuber, C4 Professorship at University Lund, 2005
- Susann Schweiger, C3 Professorship at Humboldt University, 2005
- Constanze Scharff, C4 Professorship at Free University Berlin, 2004
- Harry Scherthan, C3 Professorship at Sanitätsakademie der Bundeswehr, 2003

State doctorate (Habilitation)

- Ulrike Nuber: *DNA Mikroarrays: von der Genetik zur Zellbiologie.* Humboldt Universität Berlin, 2005.

PhD Theses

- Budny, B.: *Molecular background of X-linked mental retardation.* PhD Thesis, Karol Marcinkowski University, Poznan, 2006
- Peter, J.-U.: *Molekularbiologische und pharmakologische Manipulation der Tryptophan-Hydroxylasen.* PhD Thesis, Freie Universität Berlin, 2006
- Dlugaszewska, B.: *Molecular studies on HOXD genes: new insights into the mechanisms of limb development and pathogenesis.* PhD Thesis, Freie Universität Berlin, 2005
- Freude, K.: *Identifizierung, molekulare Analyse und funktionelle Charakterisierung von Genen, die an neuronaler Entwicklung und geistiger Behinderung beteiligt sind.* PhD Thesis, Freie Universität Berlin, 2005
- Gurok, U.: *Gene expression changes in the course of neural progenitor cell differentiation.* PhD Thesis, Freie Universität Berlin, 2005
- Hartmann, N.: *Untersuchungen zur Chromosomenevolution und Charakterisierung des Telomerkomplexes von Muntjakhirschen.* PhD Thesis, Universität Kaiserslautern, 2005
- Herr, A.: *Hochauflösende CGH mit Hilfe von DNS-Mikrorastern.* PhD Thesis, Freie Universität Berlin, 2005
- Krauß, S.: *Charakterisierung der Mikrotubulus-assoziierten PP2A und ihrer Zielproteine.* PhD Thesis, Freie Universität Berlin, 2005
- Liebe, B.: *Molekularzytologische Studien zur Chromosomendynamik in der meiotischen Prophase von Knockout- und transgenen Modellen der Maus (Mus muscullus).* PhD Thesis, Technische Universität Kaiserslautern, 2005
- Roloff, T.: *Identification of MECP2 target genes and of proteins related to MCP2.* PhD Thesis, Freie Universität Berlin, 2005
- Tao, J.: *Identification and characterization of serine/threonine kinase 9 as a causative gene for X-linked mental retardation.* PhD Thesis, Humboldt Universität Berlin 2005
- Shoichet, S.: *Identification and characterization of genes involved in cognitive function.* PhD Thesis, Freie Universität Berlin, 2004
- Wieczorek, G.M.: *Expressionsanalysen muriner Knochenmarkstromazellen und ihrer Differenzierung in einen neuronalen Phänotypen mittels cDNS-Mikroarrays.* PhD Thesis, Freie Universität Berlin, 2004
- Erdogan, F.: *Typisierung biallelischer Marker (SNPs) mit DNS-Mikrorastern.* PhD Thesis, Freie Universität Berlin, 2003
- Lehmann, T.: *Isolierung und Charakterisierung exprimierter Sequenzen im Bereich des MID1-Gens.* PhD Thesis, Humboldt Universität Berlin, 2003
- Meunier, D.: *Functional analysis of the mouse G90 gene.* PhD Thesis, Albert-Ludwigs-Universität Freiburg, 2003
- Musante, L.: *Molecular Characterization of Noonan Syndrome.* PhD Thesis, Universita' Degli Studi Di Torino, 2003
- Prietz, S.: *Expressionsanalysen mit cDNS-Mikroarrays – Aufklärung der Pathogenesemechanismen einer seltenen Augenkrankheit.* PhD Thesis, Freie Universität Berlin, 2003
- Winter, J.: *Molekulare Charakterisierung des MID1-Gens.* PhD Thesis, Freie Universität Berlin, 2003

Signal Transduction in Mental Retardation and Pain

Head
Dr. Tim Hucho
(since 9/2005)
Phone: +49 (0)30-8413 1243
Fax: +49 (0)30-8413 1383
Email: hucho@molgen.mpg.de

Scientist
Dr. Chandan Goswami

Technician
Vanessa Suckow

Scientific Overview

Molecular mechanisms of pain

Sensitization toward physical stimulation as well as chronic pain constitutes a widespread medical problem. New approaches especially for interference with pain are urgently needed. We focus on the elucidation of intracellular signaling events underlying sensitization, as the knowledge about this aspect is especially sparse.
A large variety of extracellular mediators have been identified to modulate nociceptive neuron sensitivity. They bind to many different receptor classes, which are coupled to multiple intracellular signaling cascades. Parallel as well as converging signaling can be hypothesized. We find indications of conversion onto a "sensitization module". To analyze the participating intracellular events and the components of the hypothetical sensitization module we take two very different approaches.

Cell biology of Pain

We center the search for a sensitization module on a catecholamine induced signaling pathway involving the protein kinase C epsilon (PKCε). So far, we characterized upstream events leading to the activation of this kinase. This gave proof of principal, that a_s/cAMP can activate the PKC signaling pathway mediated by the GDP/GTP exchange factor, Epac. Even though PKCε plays an important role also in heart attack and stroke, nearly no downstream targets are known. Currently we employ protein arrays as well as 2-D gel analysis of membrane proteins for the identification of PKCε downstream targets. Identified substrates will then be investigated *in vitro* as well as through lentivirus-mediated shRNA downregulation *in vivo*.
In parallel we take a candidate approach: the Capsaicin receptor TRPV1 (formerly called VR1) appears to be a downstream target of PKCε. We analyze the interplay of TRPV1 with the cytoskeleton. Indeed, TRPV1 binds to tubulin *in vitro* as well as *in vivo*. The interaction site constitutes a new tubulin-binding motive. A PKCε phosphorylation site is right next to it. Preliminary results indicate a surprising relation between tubulin binding and TRPV1 phosphorylation.
So far, the opiate and cannabinoid receptors are the only endogenous mechanism known to interfere with nociceptor sensitization in the peripheral sensory neuron. Recently, three new mechanisms could be identified: the interaction of sensory neurons with the surrounding extracellular matrix, concentration changes of hormones, and the successive stimulation of the PKCε signaling pathway. All three lead to blocked sensitization *in vivo*, and all three depend on the activity of PKCε. The molecular mechanisms are currently under investigation.

Genetics of Pain

There is no systematic genetic approach to pain. This is partially due to the still challenging task to exactly identify mutations underlying diseases. Currently we recruit patients with a common painful neuropathy for genetic analysis. We chose a disease that has been associated with changes in the catecholamine signaling system as well as the serotonergic system. The former is central to our ongoing research laid out above. The latter aims for synergistic effects with Diego Walther in the department whose research very successfully characterizes this transmitter system.

Functional analysis of mutations in mental retardation

TRPV1 as synaptic protein

Changes in neuronal morphology and synapse function are cellular correlates of mental retardation. Calcium channels are involved in the establishment and tuning of these structures. Recently one mentally retarded patient has been identified with two calcium channels located in and directly next to a chromosomal breakage area. We currently investigate one of these, TRPV1, a protein nearly exclusively studied in the context of peripheral pain sensitization. A member of this class of calcium channels has not been associated with a synaptic function so far.

We find TRPV1 in biochemical preparations of synaptic structures as well as in growth cones of primary neurons. The resting channel stabilizes pioneering microtubule, an important cytoskeletal component involved in growth cone guidance. In contrast, activation of this channel results in microtubule destabilization and retraction of the growth cone. Our research substantiates the role of TRPV1 as synaptic protein even further. Exogenous expression of TRPV1 induces filopodial structure development in various cell lines. The induced filopodia show features of synaptic cytoskeleton components. Currently we investigate the transport of TRPV1 to the synapse via synaptic transport packages.

Punctual analysis

The department identifies disease-causing mutations in increasing numbers. Not for all a detailed analysis such as for TRPV1 can be performed. For most we therefore narrow the focus to a small number of essential aspects. E.g. the group of Andreas Kuss found in a patient a deletion in the extracellular domain of a glutamate gated ion channel. In an internal collaboration we perform ligand-binding assays. These are complemented by tests for changes in the functionality of the channels such as calcium-imaging and electrophysiological recording.

We are currently also in the process of evaluating recent mutations found by Reinhard Ullman via array-CGH for future functional analysis. As a rather large number has to be analyzed, collaborations with well established neurochemical laboratories will be central to this work.

The technical stretch

In addition to external collaborations it is crucial to also have a broad spectrum of techniques established in-house. Our expertise includes the biochemical purification and characterization of proteins and protein complexes with a special focus on membrane proteins (differential centrifugation, immuno-precipitation, denaturing and native 1&2D-gel, BAC-gels, overlay blots, binding assays, …), cell-biological methods (immunfluorescent subcellular localization, time laps microscopy, *in vitro* bioassays, …) as well as molecular-biological standard methods (transfection, RT-PCR, …). As the functional properties of proteins depend to a great extend on the cellular environment, an investigation in primary neurons is preferred. To circumvent the notoriously low transfection efficiency of primary neurons, we are currently adapting a new generation of HIV-based lentiviruses to our needs. One powerful approach will be the combination of virus-mediated expression of proteins under the control of lox-P site flanked sequences with classical mouse genetics, namely mouse lines, which express Cre-recombinase in a tissue specific manner.

Selected Information

Publications

- Goswami C., M. Dreger, H. Otto, B. Schwappach, F. Hucho (2006). Rapid disassembly of dynamic microtubules upon activation of the capsaicin receptor TRPV1. J Neurochem 96, 254-266
- Hucho T., O. Dina; J.D. Levine (2005). Epac mediates a cAMP to PKC signaling in inflammatory pain: an IB4(+)-neuron specific mechanism. Journal of Neuroscience, 25(26), 6119-26
- Dina O., T. Hucho, J. Yeh, M. Malik-Hall, D. Reichling and J.D. Levine (2005). Primary Afferent Second Messenger Cascades Interact with Specific Integrin Subunits in Producing Inflammatory Hyperalgesia. Pain, 115(1-2), 191-203
- Goswami C., M. Dreger, R. Jahnel, O. Bogen, C. Gillen, F. Hucho (2005). Identification and characterization of a Ca2+-sensitive interaction of the vanilloid receptor TRPV1 with tubulin. J Neurochem 91, 1092-1103
- Gschwendtner A., Z. Liu, T. Hucho, M. Bohatschek, R. Kalla, G. Dechant, G. Raivich (2003). Regulation, cellular localization, and function of the p75 neurotrophin receptor (p75NTR) during the regeneration of facial motoneurons. Mol Cell Neurosci, 24(2), 307-22

Chromosome Rearrangements and Disease

Head
Dr. Vera M. Kalscheuer
Phone: +49 (0)30-8413 1293
Fax: +49 (0)30-8413 1383
Email: kalscheu@molgen.mpg.de

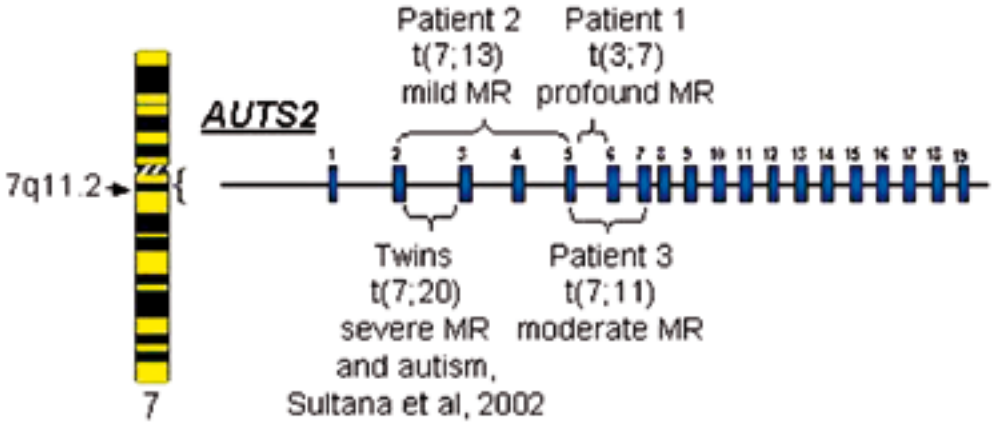
Scientist
Dr. Luciana Musante

Graduate Students
Nils Rademacher
Stella-Amrei Kunde
Sarah Shoichet (-9/2005)
Jiong Tao (-8/2004)
Kristine Freude (-12/2004)
Barbara Dlugaszewska (-1/2004)
Oliver Hagens (-8/2005)

Technicians
Kirsten Hoffmann
Ute Fischer
Astrid Grimme
Corinna Menzel

Scientific Overview

In our group, research focuses on the genetic causes of human disorders of the brain, and on the pathogenetic mechanisms underlying known gene mutations.
During the past few years, we have been very successful in identifying numerous novel disease genes through the molecular characterisation of disease-associated balanced chromosome rearrangements. Examples of truncated genes on autosomes in patients with mental retardation (MR) include *Netrin G1, FOXG1B* and *JNK3* (Freude *et al.*, 2005; Shoichet *et al.*, 2005; 2006). Using the same approach we have investigated three patients with limb malformations, of which one also exhibited MR. Analysis of the combined results suggested that the three rearrangements most likely disturbed normal *HOXD* gene regulation by position effects (Dlugaszewska *et al.*, 2006). More recently, in three unrelated patients with similar breakpoints in 7q11.2, we have found that the *AUTS2* gene is truncated. Two of the patients show profound mental retardation, whereas the third patient is less severely affected (Kalscheuer *et al.*, in preparation; collaboration with R. Ullmann and A. Tzschach; N. Tommerup, Copenhagen, D. Fitzpatrick, Edinburgh;). The function of *AUTS2* is presently unknown, but the gene is also truncated in a monozygotic twin pair with autism and mental retardation (Sultana *et al.*, 2002), which supports our conclusion that *AUTS2* is a novel MR-related gene.



On the X chromosome, truncation of *KIAA1202*, *ZNF41* and *CDKL5/STK9* caused mental retardation, due to preferential inactivation of the normal X chromosome (Hagens *et al.*, 2006; Shoichet *et al.*, 2003; Kalscheuer *et al.*, 2003). Cyclin-dependent kinase-like 5 (*CDKL5/STK9*) was disrupted in two female patients with balanced X;autosome translocations, both exhibiting severe infantile spasms and profound mental retardation. More recently, we and others have shown that mutations in *CDKL5* are responsible for the X-linked atypical Rett syndrome variant that is characterised by an early onset of seizures, with attacks starting in the first months of life (Tao *et al.*, 2004; Kalscheuer *et al.*, unpublished results, collaboration with F. Laccone, Göttingen). We are interested in the function of CDKL5 protein and if there is a relationship between CDKL5 and MeCP2, the latter being responsible for classic Rett syndrome.

We have also recently found that *ARHGEF9* is disrupted in a mentally retarded female patient who suffers from epilepsy. A truncated transcript can be amplified from the patient lymphoblastoid cell line by RT-PCR, suggesting that a mutant protein likely exists. We are currently investigating how the mutant protein might cause the disorder (Kalscheuer *et al.*, in preparation; collaboration with R. Harvey, London). Interestingly, another mentally retarded female patient with a breakpoint in *ARHGEF9* has been presented (E. Marco, ASHG 2005).
In addition, we have analysed the first known gene for Noonan syndrome (*PTPN11*) and novel candidate genes for mutations in a cohort of clinically well-characterised Noonan syndrome patients (Musante *et al.*, 2003; Schubbert *et al.*, 2006; collaboration with R. Ullmann, R. Reinhardt and M. Zenker, Erlangen).
As spin-off from studying patients with balanced rearrangements and amyotrophic lateral sclerosis (collaboration with T. Meyer, Berlin), we have investigated the breakpoints and the origin of the phenotypically silent inv(10)(p11.2q21.2), which is considered to be a polymorphic variant. Cloning and sequencing of the junction fragments on 10p11 and 10q21 revealed that neither inversion breakpoint directly involved any genes or repetitive sequences. All 20 apparently unrelated inv(10) families included in our study had identical breakpoints, and detailed haplotype analysis showed that the inversions were identical by descent. Thus, although considered a common variant, inv(10)(p11.2q21.2) has a single ancestral founder among northern Europeans (Gilling and Dullinger *et al.*, 2006).
In a parallel approach for identifying novel X-linked mental retardation (XLMR) genes, we systematically searched for mutations in selected candidate genes located on proximal Xp11. This endeavour led to the discovery of three novel XLMR genes; (Freude *et al.*, 2003, Jensen *et al.*, 2004); one of these is the polyglutamine-binding protein 1 (*PQBP1*) gene (Kalscheuer *et al.*, 2003, Cossee *et al.*, 2006). Mutations in *PQBP1* most often result in a syndromic form of XLMR where mental retardation is frequently accompanied by microcephaly and short stature. Remarkably, significant clinical variability was associated with similar and even identical frameshift mutations. In addition to MR, some of the patients presented with spasticity, heart malformations and/or other midline defects. We are interested in the causes of this clinical variability and the pathomechanism of *PQBP1* mutations. To shed more light on these issues, we are currently establishing a transgenic PQBP1 mouse model overexpressing mutant human PQBP1 protein (in collaboration with D. Walther). In addition, we have used numerous approaches to identify novel PQBP1 interaction partners (collaboration with S. Schweiger, E. Müller, Berlin), and among others, the splicing factor SIPP1 – already known to interact with PQBP1 – has been isolated. Interestingly, novel PQBP1 candidate interaction partners are predominantly expressed in brain, which helps to explain the restricted disease phenotype associated with the ubiquitously expressed *PQBP1* gene. We have also isolated proteins that are components of the same complex or sub-cellular compartment with an established role in transcription, splicing, or post-transcriptional regulation of gene expression. We have selected some of these proteins for more detailed investigations and have been able to confirm their interaction with PQBP1 in mammalian cells by co-immunoprecipitation experiments. The role of normal and mutant PQBP1 protein in these cellular processes and the possible involvement of these pathways in the pathomechanism of the disease will be a primary focus of our future research.

Selected Information

Publications

- Gilling M, Dullinger JS, Gesk S, Metzke-Heidemann S, Siebert R, Meyer T, Brondum-Nielsen K, Tommerup N, Ropers H-H, Tumer Z, Kalscheuer VM, Thomas NS (2006) Breakpoint cloning and haplotype analysis indicate a single origin of the common Inv(10)(p11.2q21.2) mutation among northern Europeans. Am J Hum Genet 78: 878-83
- Hagens O, Dubos A, Abidi F, Barbi G, Van Zutven L, Hoeltzenbein M, Tommerup N, Moraine C, Fryns JP, Chelly J, van Bokhoven H, Gecz J, Dollfus H, Ropers H-H, Schwartz CE, de Cassia Stocco Dos Santos R, Kalscheuer V, Hanauer A (2006) Disruptions of the novel KIAA1202 gene are associated with X-linked mental retardation. Hum Genet 118: 578-9
- Shoichet SA, Duprez L, Hagens O, Waetzig V, Menzel C, Herdegen T, Schweiger S, Dan B, Vamos E, Ropers H-H, Kalscheuer VM (2006) Truncation of the CNS-expressed JNK3 in a patient with a severe developmental epileptic encephalopathy. Hum Genet 118: 559-67
- Dlugaszewska B, Silahtaroglu A, Menzel C, Kubart S, Cohen M, Mundlos S, Tumer Z, Kjaer K, Friedrich U, Ropers H-H, Tommerup N, Neitzel H, Kalscheuer VM (2006) Breakpoints around the HOXD cluster result in various limb malformations. J Med Genet 43: 111-8
- Shoichet SA, Kunde SA, Viertel P, Schell-Apacik C, von Voss H, Tommerup N, Ropers H-H, Kalscheuer VM (2005) Haploinsufficiency of novel FOXG1B variants in a patient with severe mental retardation, brain malformations and microcephaly. Hum Genet 117: 536-44

- Borg I, Freude K, Kubart S, Hoffmann K, Menzel C, Laccone F, Firth H, Ferguson-Smith MA, Tommerup N, Ropers H-H, Sargan D, Kalscheuer VM (2005) Disruption of Netrin G1 by a balanced chromosome translocation in a girl with Rett syndrome. Eur J Hum Genet 13: 921-7
- Tao J, Van Esch H, Hagedorn-Greiwe M, Hoffmann K, Moser B, Raynaud M, Sperner J, Fryns JP, Schwinger E, Gecz J, Ropers H-H, Kalscheuer VM (2004) Mutations in the X-linked cyclin-dependent kinase-like 5 (CDKL5/STK9) gene are associated with severe neurodevelopmental retardation. Am J Hum Genet 75: 1149-54
- Freude K, Hoffmann K, Jensen LR, Delatycki MB, des Portes V, Moser B, Hamel B, van Bokhoven H, Moraine C, Fryns JP, Chelly J, Gecz J, Lenzner S, Kalscheuer VM, Ropers H-H (2004) Mutations in the FTSJ1 gene coding for a novel S-adenosylmethionine-binding protein cause nonsyndromic X-linked mental retardation. Am J Hum Genet 75: 305-9
- Shoichet SA, Hoffmann K, Menzel C, Trautmann U, Moser B, Hoeltzenbein M, Echenne B, Partington M, Van Bokhoven H, Moraine C, Fryns JP, Chelly J, Rott HD, Ropers H-H, Kalscheuer VM (2003) Mutations in the ZNF41 gene are associated with cognitive deficits: identification of a new candidate for X-linked mental retardation. Am J Hum Genet 73: 1341-54
- Kalscheuer VM, Freude K, Musante L, Jensen LR, Yntema HG, Gecz J, Sefiani A, Hoffmann K, Moser B, Haas S, Gurok U, Haesler S, Aranda B, Nshedjan A, Tzschach A, Hartmann N, Roloff TC, Shoichet S, Hagens O, Tao J, Van Bokhoven H, Turner G, Chelly J, Moraine C, Fryns JP, Nuber U, Hoeltzenbein M, Scharff C, Scherthan H, Lenzner S, Hamel BC, Schweiger S, Ropers H-H (2003) Mutations in the polyglutamine binding protein 1 gene cause X-linked mental retardation. Nat Genet 35: 313-5
- Kalscheuer VM, Tao J, Donnelly A, Hollway G, Schwinger E, Kubart S, Menzel C, Hoeltzenbein M, Tommerup N, Eyre H, Harbord M, Haan E, Sutherland GR, Ropers H-H, Gecz J (2003) Disruption of the serine/threonine kinase 9 gene causes severe X-linked infantile spasms and mental retardation. Am J Hum Genet 72: 1401-11

Selected invited plenary lectures

- 3rd Iranian Congress of Genetic Disorders & Disabilities, Tehran Iran, 2004, Balanced chromosome rearrangements and disease
- Annual meeting of the clinical geneticists of Quebec, Montreal Canada, 2004, Progress in X-linked mental retardation
- European Society of Human Genetics Meeting, Munich, 2004, Chromosome rearrangements and disease genes
- Symposium on Molecular Medicine and Health, Hyderabad India, 2005, Balanced chromosome rearrangements and disease

Prizes

- Shoichet, S.: Promotionspreis der Berliner Wissenschaftlichen Gesellschaft für Erforschung der genetischen Ursachen von Hirnfunktionsstörungen (2005).
- Shoichet, S.: Otto-Hahn-Medaille für Untersuchungen zur Aufklärung der molekularen Ursachen von erblichen Hirnfunktionsstörungen beim Menschen (2004).
- L. Musante *et al.*: The European Society of Human Genetics / Nature Publishing Group Award (2003).

PhD Theses

- Luciana Musante: Molecular Characterization of Noonan Syndrome. Universita' degli studi Di Torino 2003.
- Sarah Shoichet: Identification and characterisation of genes involved in cognitive function, Charite-Universitätsmedizin Berlin 2004.
- Jiong Tao: Identification and characterization of serine/threonine kinase 9 as a causative gene for X-linked mental retardation, Charite-Universitätsmedizin Berlin 2005.
- Kristine Freude: Identifizierung, molekulare Analyse und funktionelle Charakterisierung von Genen, die an neuronaler Entwicklung und geistiger Behinderung beteiligt sind, Freie Universität Berlin 2005.
- Barbara Dlugaszewska: Molecular studies on HOXD genes: New insights into the mechanisms in limb development and pathogenesis 2005.

Student Theses

- Jens Ruschmann, Funktionelle Untersuchungen zum KIAA1202 Protein, einem neuen Kandidaten für X-chromosomal vererbte geistige Behinderung, Freie Universität Berlin, 2003
- Stella-Amrei Kunde, Molekulargenetische Studien an dem Gen FOXG1B, Freie Universität Berlin, 2004
- Dejan Ninkovic, Untersuchungen zu PQBP1 und einigen wahrscheinlichen Proteininteraktionspartnern, Freie Universität Berlin, 2006

Teaching

- Lecture Genetik für Bioinformatiker, Free University Berlin, SS 2006, 2 SWS

External Funding

- NGFN2, Identification of genetic risk factors for complex disorders by studying patients with associated balanced chromosomal rearrangements. Joint with Dr. A. Tzschach, MPIMG
- DFG, SFB 577 "Analysis of Clinical Variability in Mendelian Disorders", subproject Search for modifier genes in X-linked mental retardation
- DAAD-DST (D0209618): Identification and characterisation of new genes for X-linked mental retardation. Joint with Prof. Dr. B.K. Thelma, University of Delhi South Campus, New Delhi, India

Familial Cognitive Disorders

Heads
Dr. Andreas W. Kuss
(since 01-06-2005)
Phone: +49 (0)30-8413 1253
Fax: +49 (0)30-8413 1383
Email: kuss_a@molgen.mpg.de

Prof. Dr. H.-Hilger Ropers
Phone: +49 (0)30-8413 1240
Fax: +49 (0)30-8413 1383
Email: ropers@molgen.mpg.de

Dr. Steffen Lenzner
(until 31-12-2004)

Scientist
Dr. Lars Riff Jensen

Graduate Students
Bartlomiej Budny
Lia Abbasi-Moheb
Wei Chen
Masoud Garshasbi
Mohammad Mahdi Motazacker
Rui Tian (supervised by S. Sigrist, Göttingen)
Joanna Walczak

Technicians
Melanie Wendehack
Marion Amende-Acar
Bettina Lipkowitz
Marianne Schlicht

Undergraduate Students
Achim Salamon (Diploma Student)
Katharina Albers (Bachelor Student)
Franziska Welzel (Student)
Annika Quast (Student)

Scientific Overview

X-linked Mental Retardation (XLMR)

Two main forms of XLMR – syndromic XLMR (S-XLMR), which is associated with additional phenotypes, and non-syndromic XLMR (NS-XLMR) – are being distinguished. NS-XLMR is the more common form, affecting 2/3 of the patients. However, while the majority of the genetic defects that underlie S-XLMR is either known or has been mapped to small chromosomal regions, less than 50% of those that underlie NS-XLMR have been identified, which is why part of our efforts is still focussed on the identification of new causative genetic defects in NS-XLMR. In addition to that we strive to facilitate molecular diagnostics of mutations in known XLMR genes in order to promote options for genetic counselling and prenatal diagnosis.

A major effort of this group during the past years was the screening of an XLMR candidate region on Xp11 between *ELK1* and *ALAS2* for causative mutations, which so far led to the identification of four mental retardation (MR) genes but also numerous sequence variants with so far unknown function. These investigations are still ongoing and for several genes that were identified in this screen we are also pursuing functional studies in order to characterize their role in brain function and/or development. Here we are using a variety of methods, working not only on immortalised lymphoblastoid patient cell lines but also in model organisms like mouse and Drosophila. Our efforts in this respect focus on *JARID1C* {Jensen, 2005 Coiled-coil domain-containing protein 22).

JARID1C, based on the presence of several DNA binding domains, is a putative transcription factor and for the identification of putative target genes we are employing Illumina applications for genomwide expression analysis. In parallel, *in situ* hybridisation experiments on mouse embryos have already yielded first results concerning *JARID1C* tissue distribution and we are presently carrying out cloning experiments in order to study e.g. its subcellular localization.

In a close collaboration with Dr. S. Sigrist (see below) we are investigating the function of the so far scarcely characterized *JM1* gene by generating GFP-transgenes and by performing mutagenesis experiments in *Drosophila*.

In addition, we have succeeded in generating a custom-resequencing array based on the Affymetrix 50k platform. This chip enables us to resequence the exonic sequences and splice sites of 17 known XLMR genes. It is presently being employed for mutation screening in a large number of XLMR families.

In an attempt to carry our investigations beyond protein coding regions we investigated all 13 known, brain-expressed X-chromosomal micro RNAs (miRNAs) in a cohort of 464 patients with NS-XLMR and found 4 nucleotide changes in 3 different pre-miRNA hairpins {Chen 29 /id}, submitted). MiRNAs are small non-coding RNAs that control the expression of target genes at the posttranscriptional level and have been reported to modulate various biological processes. Their function as regulatory factors in gene expression renders them attractive candidates for harbouring genetic variants with subtle effects on IQ. MiRNAs exert their regulatory function by binding to specific elements in the 3'UTR of their respective target genes. Therefore we are also interested in changes affecting known or putative miRNA target sites in the 3'UTR of XLMR genes and are presently screening the 3'UTRs of candidate genes in a number of selected patients where mutations in the coding regions of XLMR genes have been excluded.

Autosomal Recessive Mental Retardation (ARMR)

Autosomal recessive causes of MR are thought to be significantly more common than X-linked causes but they are still largely unknown and to date causative mutations genes have only been identified in three genes {Higgins, 2004 14 /id;Molinari, 2002 13 /id;Basel-Vanagaite, 2006 44 /id}.

In an effort to find (or exclude) major genetic causes of non syndromic ARMR (NS-ARMR) we have genotyped more than 100 consanguineous Iranian MR families, using the Affymetrix 10K SNP panel. Linkage analysis of these pedigrees has enabled us to identify 8 new loci harbouring candidate genes for NS-ARMR (Najmabadi *et al.*, in preparation) thus arguing against a genetic major player for this disease. We are now screening the identified intervals for causative mutations. This has already led to the discovery of a deletion affecting a prominent synaptic protein, and we are at present engaged in functional analyses of the mutant molecule.

Besides focussing on NS-ARMR we are also considering syndromic forms of ARMR as well as microcephaly. In this context we were able to define linkage intervals with a LOD-Score > 3 for a number of families that show different features accompanying a MR phenotype, and in a family with several microcephalic patients we could identify a 150-200 kb deletion, encompassing the promoter and the first six exons of the *MCPH1* gene {Garshasbi, 2006 1 /id}.

Focussing on very large pedigrees, the collection of further families in Iran and Oman is ongoing, and we have now established the Illumina GoldenGate genotyping assay in order to perform linkage analysis based on the Illumina Linkage IVb Panel (http://www.illumina.com/products/snp/snp_linkage_analysis.ilmn).

Prospects in XLMR and ARMR research

Continuing family recruitment, genotyping and linkage analysis will lead to the discovery of further MR loci. And, by relying on proven strategies in combination with new technologies (e.g. Illumina applications and novel sequencing technologies) we expect to facilitate not only the identification of linkage intervals, but also the discovery of further causative mutations for MR. These studies and ongoing efforts to elucidate the role of the relevant genes in health and disease should greatly improve the prospects for diagnosis, prevention and even therapy of MR. At the same time they will broaden our understanding of the function and development of the human brain.

Selected Information

Publications

XLMR

- Budny B, Chen W, Omran H, Fliegauf M, Tzschach A , Wisniewska M , Jensen L R, Raynaud M, Shoichet S A, Badura M, Lenzner S, Latos-Bielenska A, Ropers H-H. (2006). A novel X-linked recessive mental retardation syndrome comprising macrocephaly and ciliary dysfunction is allelic to oral-facial-digital type I syndrome. Human Genet 2006, accepted.
- Tzschach A, Lenzner S, Moser B, Reinhardt R, Chelly J, Fryns J P, Kleefstra T, Raynaud M, Turner G, Ropers H-H, Kuss A W, Jensen L R. (2006). Novel JARID1C/SMCX mutations in patients with X-linked mental retardation. Hum Mutat; (27): 389.
- Jensen L R, Amende M, Gurok U, Moser B, Gimmel V, Tzschach A, Janecke A R, Tariverdian G, Chelly J, Fryns J P, Van E H, Kleefstra T, Hamel B, Moraine C, Gecz J, Turner G, Reinhardt R, Kalscheuer V M, Ropers H-H, Lenzner S. (2005). Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. Am J Hum Genet; (76): 227-236.

ARMR

- Garshasbi M, Motazacker M M, Kahrizi K, Behjati F, Abedini S S, Nieh S E, Firouzabadi S G, Becker C, Ruschendorf F, Nurnberg P, Tzschach A, Vazifehmand R, Erdogan F, Ullmann R, Lenzner S, Kuss A W, Ropers H-H, Najmabadi H. (2006). SNP array-based homozygosity mapping reveals MCPH1 deletion in family with autosomal recessive mental retardation and mild microcephaly. Hum Genet; (118): 708-715.

Neurobiology

Head
Prof. Dr. Constance Scharff
Phone: +49 (0)30-8413 1214
Fax: +49 (0)30-8413 1383
Email: scharff@molgen.mpg.de

Scientists
Dr. Christelle Rochefort, Postdoc
Dr. Sophie Scotto-Lomassese, Postdoc

Graduate Students
Sebastian Haesler
Christina Roeske
Jana Petri
Kirill Tokarev

Diploma Student
Ute Eggers

Student
Susanne Finger, HiWi

Technician
Arpik Nshdejan

This group was established in October 2001 as part of the MPG's program to promote exceptional female scientists. In October 2004 I accepted the C4 Professorship for Animal Behavior at the Freie Universtität Berlin (FU). Through a generous cooperation agreement with the MPG my group is continuing its work at the MPI through August 2006, after which time we are moving into renovated FU space. My association with the MPI will continue through strong scientific cooperations with the Departments Ropers and Vingron.

Molecular and cellular neuroethology of vocal learning in songbirds and humans

We aim to understand how the brain brings about behavior and how in turn behavior influences the brain. To study this complex relationship, we focus on learned acoustic communication. Although most vertebrates communicate acoustically, but only a few, among them humans, ocean mammals, bats and three orders of songbirds, learn this trait. For our ethological, neuroanatomical, cell biological and molecular experiments we are using songbirds as a model, because the brain pathways controlling their acoustic behavior are well characterized. Moreover, the neuroanatomical and functional similarities between avian and mammalian brains strongly suggests the relevance of research on songbirds for understanding mammalian brain function. In addition, bird brains retain the capacity for regeneration of nerve cells in adulthood. This is in contrast to the much more limited capacity for neurogenesis in adult mammalian brains. Therefore understanding how neuronal replacement is controlled in birds and how it relates to the behavioral output is vital for both basic and medical research.

Songbirds as a model to study the FoxP2 gene, which is implicated in a human speech and language deficit

Human speech and birdsong share behavioral and neural similarities. Both are learned during a critical period via the interaction of auditory and motor centers and require a set of specialized cerebral structures. While innate dispositions to learn and produce species-appropriate sounds are present in both humans and birds, until recently no genes had been linked to learned vocalizations. FOXP2 is the first gene implicated in the human specific traits speech and language because mutations of FOXP2 (a member of the winged-helix/forkhead box transcription factor gene family) were identified in individuals with developmental verbal dyspraxia (DVD). DVD is characterised by a severe difficulty articulating speech and both expressive and receptive language deficits. Morphological and functional analysis of patients' brain implicate the basal ganglia in the etiology of the impairment. Since song learning in birds critically depends on the basal ganglia, we are interested in the role of FoxP2 in the development and function of song-relevant neural circuitry in zebra finches.

Expression patterns, molecular evolution and function of FoxP2
Sebastian Haesler, Ph.D. Student
In a previous study we found that FoxP2 is expressed in the songbird forebrain, predominantly in neurons of the basal ganglia, the brain region most affected in patients with FOXP2 mutations. Strikingly, within the male zebra finch striatum, FoxP2 was upregulated specifically in Area X, a nucleus essential for vocal learning, at the age when vocal learning occurs. Thus, place and time of FoxP2 expression in the avian Area X are compatible with a role in learned vocal communication.

To investigate if FoxP2 expression in Area X is causally related to song learning in the zebra finch we used lentivirus-mediated RNAi to knockdown FoxP2. Birds with reduced FoxP2 levels in Area X learned their song incompletely and inaccurately. Moreover, the acoustic structure and the duration of song syllables was abnormally variable, similar to word production in children with DVD. These findings establish that FoxP2 mediates auditory-guided vocal motor learning in songbird basal ganglia. Accordingly, FoxP2 mutations in patients with DVD might compromise the acquisition of basal ganglia-dependant auditory-guided motor programs during speech learning. The fact that FoxP2 is essential for both birdsong and speech suggests that the molecular mechanisms underlying the uniquely human capacity of language are not exclusive to the hominid lineage.

I collaborate with research groups Vingron (with Benjamin Georgi) on quantitative analysis of song syntax, i.e. the sequential arrangement of sound elements within in a song. Current experiments aim to establish methods for clustering song data using mixture-models to allow the categorization of song elements.

Search for interaction partners of FoxP2
Jana Petri, Ph.D Student
This project aims to identify molecular interaction partners of the zebra finch FoxP2 that are relevant for song learning, using *in vitro* and *in vivo* approaches. These include Chromatin Immunoprecipitation and zebra finch c-DNA microarrays developed in Eric Jarvis' lab at Duke University. Towards both ends we are combining lentivirally mediated FoxP2-overexpression and knock-down approaches within striatal nucleus Area X of zebra finches, where FoxP2 is also endogenously expressed. In addition, we are using neural cell lines to complement our studies and establish protocols.

Adult neurogenesis: molecular control and behavioral function

Adult neurogenesis is a widespread phenomenon in vertebrates. In songbirds, neural progenitors proliferate along the lateral ventricles during development and in adulthood. After migration, the post-embryonically formed neuroblasts integrate into many forebrain regions, among them song nuclei HVC and Area X that participate in song learning and production. Recruitment of post-embryonically generated neurons into HVC, into rodent hippocampus and olfactory bulb has been linked to learning and memory.

Regulation and function of adult neurogenesis in Area X, a nucleus of the striatum involved in song learning.
Dr. Christelle Rochefort, Postdoctoral Fellow.
The cellular identity and the role of post-embryonically generated neurons in Area X are unknown. My recent experiments demonstrate that the majority of new neurons in post-embryonic Area X of male zebra finches express DARPP32 (dopamine-and-cAMP-regulated phosphoprotein of 32 kDa) but not acholine acetyltransferase or parvalbumin. Retrogradely labelled neurons projecting to thalamic nucleus DLM are not renewed. These results suggest that post-embryonically recruited neurons in Area X are medium spiny neurons. The medium spiny neurons in Area X express FoxP2, a transcription factor critical for normal speech and language development in humans. Since FoxP2 expression is upregulated during periods of vocal plasticity we also investigated whether developmental upregulation of FoxP2 in Area X and neuronal recruitment during song development might be associated. Consistent with their medium spiny phenotype new neurons in Area X did express FoxP2 and recruitment increased transiently during the juvenile song learning period. Moreover we found that FoxP2 was expressed in the ventricular zone of adult songbirds and in migrating neuroblasts but was absent from the germinal zones in adult mouse brains, the hippocampus and the subventricular zone. Together these results raise the possibility that neuronal recruitment and FoxP2 expression in Area X are associated and mediate vocal learning.

What role does the signaling molecule retinoic acid play during song learning in the zebra finch brain?
Christina Roeske, Ph.D Student
Retinoic acid is primarily known for controlling pattern formation processes during the embryonic development of vertebrates. It is crucial for the development of numerous structures, among them limbs, face, and hindbrain. The idea that retinoic acid also plays a role in learning is relatively new. First evidence came from knockout mice lacking certain retinoic acid receptors. These mice showed poorer spatial learning than their receptor equipped relatives.

That retinoic acid is also important for song learning in the bird brain was suggested for the first time in the course of a gene expression study on the zebra finch brain. A retinoic acid synthesizing enzyme, RalDH is expressed specifically in parts of the song control system. Interestingly, pharmacological blocking of this enzyme in the song nucleus HVC in juvenile birds lead to impaired song learning. This shows that the retinoic acid signaling pathway is indispensable for normal song learning; however, the mechanisms by which this is achieved remain unclear. I am studying the expression patterns of the different players of the retinoic acid signaling pathway to get a more detailed idea of where exactly retinoic acid is exerting its effect.

Analysis of neuronal cluster in the adult zebra finch song nucleus HVC

Dr. Sophie Scotto Lomassese, Postdoctoral Fellow.

New HVC neurons are incorporated into an existing neural network where they replace neurons that have died. The HVC is composed of 3 populations of neurons: the projection neurons to nucleus RA (HVC->RA) that control the production of learned song, the projection neurons to Area X (HVC->X) that are involved in the auditory dependant-song plasticity and inhibitory interneurons (HVC-IN). Interestingly, many neurons of these three populations are packed tightly together, with extensive soma-soma appositions, forming cell clusters in which the newly arriving neurons can be recruited.

We are investigating the hypothesis that the incorporation of new neurons into the existing functional HVC network depends on correct 'cluster' association with particular neuron types. Could the clusters constitute functional units where immature neurons are instructed, perhaps through gap junctions, before replacing the neurons that acted as instructors? Before investigating whether recruitment of new neurons follows specific rules, we first determined which population(s) in the HVC is renewed in adults. Although HVC-IN were initially considered as the only population to be replaced in adult, further studies demonstrated the renewal of the HVC->RA. We found that none of the new neurons express calbindin, parvalbumin and calretinin that identify the different HVC-IN subpopulations. Moreover, we verified that all the HVC-IN express at least one of these calcium-binding proteins. Together, this demonstrates that HVC-IN are not replaced in adult.

Analysis of the frequency and composition of clusters using confocal laser microscopy of immunocytochemically marked cell populations (cell birth and cell phenotype markers) in conjunction with stereotaxic injections of retrograde tracers shows that 80% of the neurons have one or more partners. The cluster distribution is not randomly distributed but those composed of two or three neurons were observed most often. Interestingly, one month after their birth, 70% of the new neurons are already found in clusters. Moreover, these newly arriving cells do not incorporate into the most frequently encountered cluster types but apparently associate preferentially with the relatively rare single neurons or join clusters of two cells. In fact, new neurons seem to participate in maintenance of the overall cluster distribution, suggesting the existence of a mechanism that guides new cells toward the right place. We also analyzed the phenotype of the cell in direct contact with a new neuron and found that they are also not distributed according to frequency of occurrence of the different cell classes. Together, these results suggest that integration of new neurons into the HVC is not random and that clusters might participate to the mechanisms controlling the incorporation of the new neurons into the existing network. Ultrastructural studies are currently underway to clarify the presence of gap junctions between new and already existing neurons.

Selected Information

Publications

- Scharff C, White SA (2004) Genetic components of vocal learning. Ann N Y Acad Sci 1016: 325-47
- Gurok U., Steinhoff C., Lipkowitz B, Ropers H-H, Scharff C, Nuber UA (2004) Gene expression changes in the course of neural progenitor cell differentiation. J. Neurosci. 24(26):5982-6002
- Scharff C, Haesler S (2005) An evolutionary perspective on FoxP2: strictly for the birds? Curr Opin Neurobiol. 15(6):694-703
- Haesler, S. Wada, K., Nshdejan A., Morrisey, E.. Lints, E. K. T.,. Jarvis, E.D., Scharff, C.(2004) FoxP2 expression in avian vocal learners and non-learners. J Neurosci. 24 (13):3164-3175
- Kalscheuer VM, Freude K, Musante L, Jensen LR, Yntema HG, Gécz J, Sefiani A, Hoffmann K, Moser B, Haas S, Gurok U, Haesler S, Aranda B, Nshedjan A, Tzschach A, Hartmann N, Roloff TC, Shoichet S, Hagens O, Tao J, van Bokhoven H, Turner G, Chelly J, Moraine C, Fryns JP, Nuber U, Hoeltzenbein M, Scharff C, Scherthan H, Lenzner S, Hamel BCJ, Schweiger S & Ropers H-H (2003) Mutations in the polyglutamine-binding protein 1 gene cause X-linked mental retardation. Nat Genet 35(4):313-5

Selected international invited plenary lectures

- 2004 April 6: Outfoxing the competition – the 'speech' gene FoxP2 in avian vocal learners and non-learners. Developmental Neuroscience Seminar, The Rockefeller University, New York, USA
- 2004 April 8: FoxP2 in avian learners and non-learners:association with vocal plasticity. Department of Biological Sciences, Columbia University, New York, USA
- 2004 April 30: Can birds contribute to understanding the role of FoxP2 in a human speech deficit? Dept.Human Anatomy and Psychobiology, Medical School, Univ. Murcia, Spain
- 2005 June 30: What does a speech gene do in a bird's brain? Gordon Research Conference Neural Circuits & Plasticity, Salve Regina University, Newport, RI, USA
- 2005 September 27: What is a speech gene doing in a bird's brain? Function of FoxP2 in songlearners: McDonnell Centre for Cognitive Neuroscience Autumn School Conference, University of Oxford, UK
- 2005 October 3: siRNA FoxP2 knockdown interferes with song learning in zebra finches, University Lecture. The Rockefeller University, New York, USA

Prizes

- Haesler, S.: Schloessman-Preis; Beitrag eines Gens (FoxP2) zur Entwicklungsregulation des Gesanglernens bei Zebrafinken (2005).

Organization of scientific events

- Standing Member of Dahlem Colloquium Organization Committee
- Science Frontiers – Birdsong in Behavioural and Neurobiological Research, International Symposium, Aug. 10-12 2006, Berlin.
- Member of the steering committee of the 'Zebra Finch Genome Biology Consortium"
- International Symposium :"Developmental Disturbances in the Nervous System" to be held Nov 15-17, 2007

Work as scientific referee

Journals

- Neuron
- Journal of Neuroscience
- Journal of Neurobiology
- Nature Neuroscience Reviews
- Experimental Neurology
- Current Biology

Institutions

- Leibniz Gemeinschaft, Primatenzentrum Göttingen

Cooperations

- Prof. Erich Jarvis, Duke University USA
- Prof. Richard Mooney, Duke University USA
- Prof. Fernando Nottebohm, The Rockefeller University USA
- Prof. Luis Puelles, University of Murcia, Spain

Clinical Genetics and Biochemistry

Head
Prof. Dr. Susann Schweiger
Phone: +49 (0)30-8413 1254
Fax: +49 (0)30-8413 1383
Email: schweiger@molgen.mpg.de

full-time until March 31, 2005; from April 1, 2005, Associate Professor at the Charité, part-time (20%) group leader at the MPIMG

Scientists
Dr. Jennifer Winter
Dr. Sybille Krauss
Dr. Beatriz Aranda-Orgilles

Graduate Students
Ewa Jastrzebska
Malgorzata Romanowska
Eva Kickstein

Technicians
Melanie Kunath
Nathalia Haritonow

Previous Group Member
Vanessa Suckow (Technician)

Scientific Overview

For a large number of monogenic disorders, the underlying gene defects have been identified during the past two decades, but the physiological roles of most of these genes remain unknown. Our group undertakes a strategy that systematically combines clinical phenotype analysis and biochemical analysis to maximize exploitation of this as-yet largely unused database. Molecular signaling networks that we identify with this system then form the basis for an understanding of complex disorders and the establishment of novel therapeutic approaches.

Our previous work has demonstrated the potential of this approach. Starting from Opitz BBB/G syndrome (OS), a rare monogenic disorder affecting ventral midline development, the underlying genetic defect was identified as the gene coding for the microtubule-associated MID1 protein. By means of detailed clinical analysis of related monogenic phenotypes, a previously unknown role of this protein complex in the sonic hedgehog (shh) pathway could be uncovered (Krauss *et al.*, manuscript submitted)..

Furthermore, we found that the MID1/α4 complex associates with active polyribosomes, participating in the formation of a novel microtubule-associated translation complex in conjunction with several other proteins and mRNA. We also identified G-quartet- like RNA structures that act as determinants for specific binding of mRNA to this complex. Finally, we have shown that mRNAs of both ephrinB ligands and receptors, central mediators of midline pathfinding, are integrated into this complex via their G-quartets. Thus, our data suggest a molecular basis for the phenotypic overlap between OS and craniofrontonasal dysplasia, ventral midline disorders resulting from mutations in *MID1* and *ephrin B1*, respectively. Moreover, they provide a novel regulatory mechanism for the sequence-specific control of mRNA translation that may be dysregulated in a wide range of ventral midline defects, including polygenic malformations such as congenital heart defects, and cleft lip and palate (Aranda *et al.*, manuscript submitted).

The MID1/α4/PP2A complex for the therapy of cancer, Alzheimer's disease and Huntington's disease

Involvement of the MID1/α4/PP2A complex in at least two central oncogenic signaling cascades and the previously described tumor suppressive function of PP2A renders this protein complex a particularly interesting target for the development of a novel **tumor therapy**. Biological validation of a possible influence of the complex on tumor cell progression was achieved by analysis of the effects of an RNAi-based α4 knock-down on a panel of different tumor cell lines, demonstrating that cells with α4 knock-down are led into G1 arrest and subsequent apoptosis. α4 knock-down leads to a loss of PP2Ac ubiquitination and degradation, resulting in an increase of PP2A activity. Real-time PCR of cyclins in α4 knock-down cells revealed that, shortly before the observed G1 arrest, cyclin D1, a target gene of shh signaling, is down-regulated 5-fold (Jastrzebska *et al.*, manuscript in preparation). Furthermore, we can show that the MID1/α4 ubiquitin ligase counteracts CYLD, a microtubule-associated ubiquitin hydrolase that stabilizes the IKK complex, thereby inactivating NF-kappaB pro-apoptotic signalling. In summary, our data show that exogenous manipulation of the interaction between MID1, α4 and PP2A leads to cyclin D1-mediated G1 arrest and, subsequently, to stabilization of IKK, inactivation of NFkappaB and apoptosis (So *et al.*, manuscript in preparation).

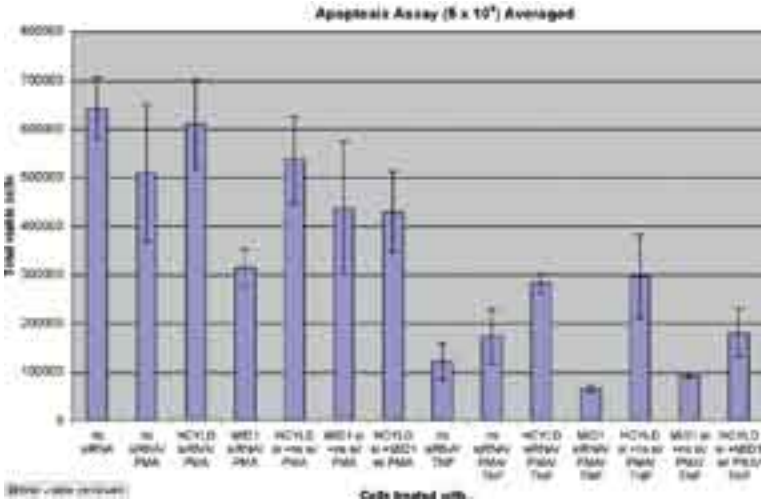


Figure 1: MID1 counteracts CYLD. Cell viability was analysed after NF-kappaB induction (with PMA) with (left 7 columns) and without (right 7 columns) apoptosis stimulation by TNFalpha. Cells were transfected with non-silencing siRNA (ns), CYLD-specific siRNA (HCYLD) or MID1-specific siRNA (MID1) or double-transfected.

Hyperphosphorylation of the microtubule-associated tau protein leads to accumulation of this protein as the major component of paired helical filaments found in brains of **Alzheimer** patients. Although the pathomechanism of Alzheimer's disease is incompletely understood, these filaments appear to play a key role in the induction of neuronal cell death. Microtubule-associated PP2A has been identified as the key enzyme dephosphorylating tau, which makes exogenous inhibition of PP2Ac degradation by manipulation of the MID1/α4/PP2A interaction a highly interesting tool for the development of an Alzheimer therapy.
In order to biologically validate the possible influence of the MID1/α4/PP2A complex on tau phosphorylation, we have narrowed down the α4-binding region of the MID1 protein to the 60 amino acids comprising the B-Box1 domain and have expressed this peptide in primary mouse neurons using gene-gun technology. Transfected cells were then analysed by immunofluorescence for tau phosphorylation using an antibody specifically detecting dephosphorylated tau. As controls, non-transfected cells, only EGFP-transfected cells and cells transfected with a B-Box1 lacking α4 binding affinity due to an A to T mutation were used. Quantification of the detected signals revealed significantly ($p<0,0001$) higher fluorescence in the wild-type B-Box1-transfected cells than in the controls.

analysed cells	**mean avergae %**	**n**
non-tranfected	100,00+/-6,15	223
Bbox1-transfected*	158,34+/-9,40	45
Bbox1mut-transfected	97,39+/-7,72	23
only EGFP-transfected	98,14+/-7,5	33

phosphorylation-dependent quantification of tau

analysed cells	**mean avergae %**	**n**
non-tranfected	100,00+/-7,9	75
Bbox1-transfected*	110,80+/-11,70	12
Bbox1mut-transfected	nd	nd
only EGFP-transfected	90,00+/-5,76	34

phosphorylation-independent quantification of tau

Huntington's disease is an autosomal dominant neurodegenerative disorder with a poor prognosis. It is caused by a triple-repeat expansion of a CAG codon in exon 1 of the *Huntingtin* gene, leading to its translation into Huntingtin protein with an expanded glutamine stretch in the N-terminus and subsequent formation of intracytoplasmic aggregates in brains of Huntington patients.

Using an RNA-protein pull-down assay, we have now seen that mRNA with expanded CAG repeats specifically bind to the MID1/α4/ PP2A translation unit, leading to mTOR-coupled induction of protein translation of the respective mRNA. Using a filter assay that allows quantification of insoluble Huntingtin protein, we can show that mTOR inhibition by rapamycin or α4 knock-down leads to a significant reduction of the pathognomonic Huntingtin protein in a cell line expressing exon1 of Huntingtin with a pathological number of CAGs (exon1+83Glu). Exogenous interference with the MID1/α4/PP2A protein complex is therefore also a promising strategy for treatment of Huntingtin's disease.

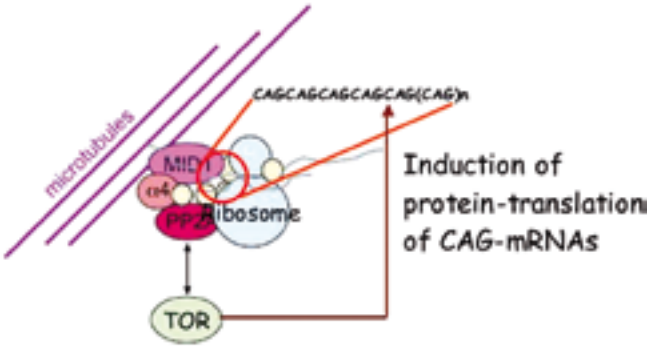


Figure 2: Model. CAG repeats of exon1 of Huntingtin mRNA bind to a microtubule-associated MID1/α4/PP2A translation unit in a length-dependent manner. Low PP2A activity, maintained by the negative PP2A regulators MID1 and α4, leads to a TOR-dependent increase of Huntingtin mRNA translation and the production of more (pathological) protein.

Selected Information

Publications

- So J., Suckow V., Kijas Z., Kalscheuer V., Moser B., Baars M., Firth H., Lunt P., Hamel B., Moraine C., Odent S., Schinzel A., van der Smagt J.J., Devriendt K., Passarge E., van der Burgt I., Cox T., Opitz J., Ropers H.-H. & Schweiger S. (2005) Mild phenotypic spectrum in a series of Opitz BBB/G syndrome patients with MID1 mutations. Am. J. Med. Genet. 132A:1-7
- Winter J., Lehmann T., Krauß S., Trockenbacher A., Kijas Z., Foerster J., Suckow V., Yaspo M.L., Kulozik A., Kalscheuer V., Ropers H.-H., Schneider R., Schweiger S. (2004) Regulation of the MID1 protein function is fine-tuned by a complex pattern of alternative splicing. Human Genetics 114:541-552
- Schweiger S. and Schneider R. (2003) The MID1/PP2A complex: a key to the pathogenesis of Opitz BBB/G syndrome. Bioassays 25, 356-366. invited review

Clinical Genetics

Head
Andreas Tzschach, MD
Phone: +49 (0)30-8413 1416
Fax: +49 (0)30-8413 1383
Email: tzschach@molgen.mpg.de

Scientist
Maria Hoeltzenbein, MD

Technicians
Hannelore Madle
Susanne Freier

Scientific Overview

The main task of the Clinical Genetics group is the recruitment and clinical characterization of patients and families for the various research groups of our department. As previously costly and laborious molecular genetic investigations get faster and cheaper thanks to technological advances, the recruitment of patients and families with specific phenotypic features increasingly constitutes the bottleneck for all projects in human genetic research.

Disease associated chromosome aberrations

Breakpoint analysis of disease-associated balanced chromosome rearrangements combining array CGH (R. Ullmann's group), FISH and other molecular genetic techniques (V. Kalscheuer's group) is a fast and efficient strategy to identify novel disease-causing genes. Comprehensive clinical characterisation is a prerequisite for the selection of patients who are suitable for analysis. New questions concerning specific phenotypic details often arise after the identification of a disrupted gene or genes at the breakpoints. The Clinical Genetics group obtains these data through collaboration with referring doctors or by contacting the patients or their families. Another major task of our group is the recruitment of patients with *de novo* disease associated balanced chromosome rearrangements by maintaining and extending a network of clinical geneticists and other specialists.

Late-onset disorders

Apart from disorders that are usually present at birth or in early childhood, we also perform research on balanced chromosome rearrangements in patients with late-onset diseases. This approach has also the potential to elucidate genes involved in the etiology of complex disorders, as has been shown e.g. for schizophrenia or Tourette's syndrome before. We have performed a questionnaire survey among adult carriers of balanced chromosome rearrangements, which resulted in the identification of several patients and families in whom an association of the chromosome rearrangement with the respective disease was likely. Subsequent breakpoint analysis led to the identification of candidate genes for psoriasis, dyslexia, hyper-IgE syndrome and haematological malignancies.

Unbalanced chromosome aberrations

Both large, cytogenetically visible unbalanced chromosome aberrations and small, submicroscopic aberrations which are only detectable by the novel very high resolution array CGH technique (R. Ullmann's group) are an important cause of congenital malformations and mental retardation, and they can point to single genes responsible for a specific phenotype. Our group characterizes these patients clinically and establishes genotype-phenotype correlations.

X-linked mental retardation

The elucidation of novel genes involved in X-linked mental retardation is a major research focus of our department. Our group continues to recruit families with probable X-linked MR by collaboration with clinical geneticists in Germany and abroad. We also complete clinical data of formerly submitted families and communicate the results of mutation analysis to the respective physicians or families. Apart from families with non-syndromic XLMR, we also investigate families with syndromic forms of mental retardation. Linkage analysis and mutation analysis in these families is being performed in the "Familial Mental Retardation" group (A. Kuss, H.-H. Ropers).

Autosomal recessive mental retardation

In a large collaboration with the Genetic Research Centre in Tehran/Iran (Dr H Najmabadi) we received DNA and clinical data, including photographic documentation, of more than 100 consanguineous families with multiple mentally retarded children. Whereas the majority of these families suffer from unspecific (non-syndromic) mental retardation, several families have additional clinical problems. Many of these syndromes are apparently novel and do not resemble other published MR syndromes. Beside the clinical evaluation (frequently including proposals for specific investigations such as MRI scans, ophthalmologic examinations and others), we screen the regions of homozygosity for promising functional candidate genes according to morphological similarity or (putative) functional links to other syndromes based on several databases.
In a collaborative project with Dr A Rajab, the leading clinical geneticist in Oman, we obtained blood samples from several consanguineous families with syndromic forms of MR. The spectrum of disorders includes a form of albinism, ectodermal dysplasia, a muscle disorder and a large family with MR and epilepsy. We also received material from families with syndromic forms of mental retardation from our partners in Poznan/Poland (Prof A Latos-Bielenska).

Cell culture facility

We establish permanent cell lines from peripheral blood lymphocytes by EBV transformation after obtaining informed consent from patients in whom molecular cytogenetic or molecular genetic investigations are planned. Our cell culture lab performs EBV transformation, stores the cell lines and provides ready-to-use DNA, RNA or metaphase chromosome spreads for FISH investigations to the respective research groups.

Genetic counselling

In their function as members of the Institute of Clinical Genetics and the Social Pediatric Centre at the Humboldt University/Charité-Universitätsmedizin Berlin, Andreas Tzschach and Maria Hoeltzenbein actively participate in genetic counselling clinics and the regular dysmorphology, clinical genetics and cytogenetics meetings.

Selected Information

Publications

- Tzschach, A., I. Schulzke, C. Menzel, V. Kalscheuer, H. Tönnies, H. Scherthan, M. Radke, A. Knoblauch, H.-H. Ropers and M. Hoeltzenbein (2006).Molecular cytogenetic analysis of a de novo 5q23.3q31.2 deletion and further delineation of a del5q22q31 syndrome. American Journal of Medical Genetics 140(5), 496-502.
- Tzschach, A., K. Hoffmann, M. Hoeltzenbein, I. Bache, N. Tommerup, C. Bommer, H. Körner, V. Kalscheuer and H.-H. Ropers (2006). Molecular characterization of a balanced chromosome translocation in psoriasis vulgaris. Clinical Genetics 69(2), 189-193.
- Yan, K.L., X-J. Zhang, Z-M. Wang, S. Yang, G-L. Zhang, J. Wang, F-L. Xiao, M. Gao, Y. Cui, J-J. Chen, X. Fan, L-D. Sun, Q. Xia, K-Y. Zhang, Z-M. Niu, S-J. Xu, A. Tzschach, H.-H. Ropers, J-J. Liu and W. Huang (2006). A novel MGST2 non-synonymous mutation in a Chinese pedigree with psoriasis vulgaris. Journal of Investigative Dermatology 126(5), 1003-1005.
- Tzschach, A., I. Krause-Plonka, C. Menzel, A. Knoblauch, H. Toennies, M. Hoeltzenbein, M. Radke, H.-H. Ropers and V. Kalscheuer (2006). Molecular cytogenetic analysis of a de novo interstitial chromosome 10q22 deletion. American Journal of Medical Genetics 140(10), 1108-1110.
- Tzschach, A., S. Lenzner, R. Reinhardt, B. Moser, J. Chelly, J-P. Fryns, T. Kleefstra, M. Raynaud, J. Gécz, H.-H. Ropers, A. Kuss and L.R. Jensen (2006). Novel JARID1C mutations in patients with X-linked mental retardation. Human Mutation 27(4), 389.
- Tzschach, A. and H.-H. Ropers (2006). X-chromosomale mentale Retardierung. medgen 18, 187-193.
- Budny, B., W. Chen, H. Omran, M. Fliegauf, A. Tzschach, M. Wisniewska, L.R. Jensen, M. Raynaud, S.A. Shoichet, M. Badura, S. Lenzner, A. Latos-Bielenska and H.-H. Ropers (2006). A novel X-linked recessive mental retardation syndrome comprising macrocephaly and ciliary dysfunction is allelic to oral-facial-digital type I syndrome. Human Genetics Online (2006).
- Tzschach, A., M. Hoeltzenbein, K. Hoffmann, C. Menzel, A. Beyer, V. Ocker, G. Wurster, H.-H. Ropers, V. Kalscheuer and H. Heilbronner (2006). Situs ambiguus and cardiac defect in a girl with chromosome translocation t(X;1)(q26;p13.1) and involvement of ZIC3. European Journal of Human Genetics in press.
- Garshasbi, M., M.M. Motazacker, K. Kahrizi, F. Behjati, S.S. Abedini, S.E. Nieh, S.G. Firouzabadi, C. Becker, F. Ruschendorf, , P. Nurnberg, A. Tzschach, R. Vazifehmand, F. Erdogan, R. Ullmann, S. Lenzner, A.W. Kuss, H.-H. Ropers and H. Najmabadi (2006). SNP array-based homozygosity mapping reveals MCPH1 deletion in family with autosomal recessive mental retardation and mild microcephaly. Human Genetics 118(6), 708-715.
- Jensen, L.R., M. Amende, U. Gurok, B. Moser, V. Gimmel, A. Tzschach, A.R. Janecke, G. Tariverdian, J. Chelly, J.-P. Fryns, H. Van Esch, T. Kleefstra, B. Hamel, C. Moraine, J. Gecz, G. Turner, R. Reinhardt, V.M. Kalscheuer, H.-H. Ropers and S. Lenzner (2005). Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. American Journal of Human Genetics 76, 227-236.
- Kalscheuer, V.M., K. Freude, L. Musante, L.R. Jensen, H.G. Yntema, J. Gécz, A. Sefiani, K. Hoffmann, B. Moser, S. Haas, U. Gurok, S. Haesler, B. Aranda, A. Nshdejan, A. Tzschach, N. Hartmann, T.C. Roloff, S. Shoichet, O. Hagens, J. Tao, H. van Bokhoven, G. Turner, J. Chelly, C. Moraine, J.P. Fryns, U. Nuber, M. Hoeltzenbein, C. Scharff, H. Scherthan, S. Lenzner, B.C.J. Hamel, S. Schweiger and H.-H. Ropers (2003). Mutations in the polyglutamine-binding protein 1 gene cause X-linked mental retardation. Nature Genetics 35(4), 313-315.

Cooperations

Autosomal recessive mental retardation

- Hossein Najmabadi, Genetic Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran
- Anna Rajab, MD, Genetic Unit, DGHA, Ministry of Health, Muscat, Sultanate of Oman

Hyper-IgE syndrome

- Ellen Renner, MD, Children's hospital, University of Washington, Seattle, USA
- Prof Bernd Belohradsky, Children's hospital, University of Munich

Psoriasis vulgaris

- Jian-Jun Liu, Department of Dermatology, Anhui Medical University, Hefei, China

Balanced chromosome rearrangements

- Niels Tommerup, Wilhelm Johannsen Centre, Copenhagen, Denmark

X-linked mental retardation

- EUROMRX consortium (*www.euromrx.com*) and numerous clinical geneticists in Germany and abroad

Molecular Cytogenetics

Head
Dr. Reinhard Ullmann
Phone: +49 (0)30-8413 1251
Fax: +49 (0)30-8413 1383
Email: ullmann@molgen.mpg.de

Scientist
Dr. Fikret Erdogan

Technicians
Ines Müller
Melanie Wendehack
Marei Schubert
Jan Jurkatis

Graduate Student
Artur Muradyan

Students
Vivien Boldt
Anne Steininger
Britta Vorwerk
Stephanie Stabentheiner

Scientific Overview

The Molecular Cytogenetics Group was founded in February 2004. Within the first year of its existence, the group generated a sub-megabase resolution whole genome tiling path BAC array, consisting of a human 32k Re-Array set (*http://bacpac.chori.org/pHumanMinSet.htm*; clones and DNA kindly provided by Pieter de Jong), the 1Mb Sanger set (clones kindly provided by Nigel Carter, Wellcome Trust Sanger Centre) and a set of 390 subtelomeric clones (assembled by members of the COST B19 initiative: Molecular Cytogenetics of Solid Tumors). DNA of these clones was amplified by linker adapter PCR, and high density arrays are printed by the in-house array facility headed by Claus Hultschig (Dept. Lehrach). Moreover, in cooperation with Wei Chen (WG Kuss) we have developed a superior novel tool for the analysis and visualization of array CGH data, the software package CGHPRO, which we have made freely available to the scientific community (W. Chen *et al.*, BMC Bioinformatics 2005).

Application of BAC arrays in human genetics

Array CGH-based screening for disease-associated genomic imbalances is complicated by the existence of functionally neutral DNA copy polymorphisms, which are also found in the normal population. To identify at least the most common of these polymorphisms, hundreds or even thousands of healthy individuals have to be studied. To avoid this extremely costly and time-consuming effort, and in view of related activities of other laboratories, we have decided to bypass this step by studying various large cohorts of patients with different disorders and to infer the clinical relevance of copy number changes found in one cohort by comparing it to the entire data set and to relevant public databases.

Meanwhile, this data pool already consists of several hundreds of patients, which we have analysed in cooperation with clinical experts in the field. Examples are the set of 150 patients with mental retardation, which we have studied in parallel to about 100 autistic patients (a set of further 120 autistic patients was used to screen for de novo DNA copy number changes employing quantitative PCR and to sequence selected candidate genes). Tourette syndrome, schizophrenia, attention deficit/hyperactivity disorder (ADHD), and amyotrophic lateral sclerosis (ALS) are further diseases that we have studied at this scale. Hybrizations of patients have been completed and we are in the process of finishing the verification and testing of the parents. In all of these cohorts, we have detected a considerable number of apparently relevant aberrations, which were either de novo or co-segregated with the disease.

Apart from these external cooperations, we have strong research ties with other groups of the department. In cooperation with Vera Kalscheuer, we have screened 28 patients with Noonan syndrome with unknown cause and identified a de novo deletion at 10q24 in one patient. Together with Susann Schweiger, we have searched for new candidate genes in 15 patients with Opitz syndrome. Five DNA copy number changes have been identified and we are now waiting for the parental DNA. Additionally our group is part of the project "Balanced Chromosome Rearrangements and Disease", where we are responsible for the array-based fine mapping of chromosomal breakpoints and the exclusion of DNA copy number changes in patients with balanced rearrangements. In addition, several single cases have been analysed for the Clinical Genetics Group and the Familial Mental Retardation Group.

Many of these efforts were motivated by their implications on Health Care. In this context and in collaboration with the Mundlos Group we have transferred the high-resolution array CGH technology to the Institute of Human Genetics, Charité, Berlin and agreed to provide this laboratory with BAC arrays for non-commercial, science-driven diagnostics.

Application of BAC arrays in tumor biology

In tumor biology array CGH is employed for both improvement of therapy and better understanding of tumor formation. Examples for therapy-oriented projects are a study of dendritic cell tumours for illustrating the potential of array CGH to assist in the classification of rare tumors, and an ongoing project aiming at the definition of patterns of chromosomal aberrations that could help to distinguish therapy responders from non-responders in pediatric ALL. In another study, we have combined laser-microdissection and array CGH to demonstrate that aberrations are not confined to the tumor itself, but are also present in normal-looking adjacent epithelium and stroma, thereby explaining the frequent local relapse of apparently complete tumor extirpation thus implicating the necessity to integrate these parts into therapeutic considerations too.

Our long-term aim is to improve our understanding of tumorigenesis and to optimize tumor therapy. To this end, we are going to integrate array CGH, methylation analysis, ChIP on Chip, gene expression analysis and tissue micro array data derived from lung carcinomas and will supplement these data with insights derived from experimentally induced chromosomal rearrangements. On one hand, these data will be used to create a comprehensive map of oncogenes, representing promising candidates for defect-targeted therapies in lung carcinomas. On the other hand, we expect insights into the mechanisms that predispose chromosomes to break.

Outlook

BAC arrays have turned out to be an optimal platform for the detection of DNA copy number changes, which, once having set up the facility, provides good resolution at relatively low costs and which is not as sensitive to variation of DNA quality as oligo nucleotide arrays. Therefore, for the foreseeable future we will continue to use BAC arrays for screening large cohorts of monogenic and complex disorders to identify causative gene defects for disease. Following the strategy that we have successfully employed in the past, these studies will be largely performed by guest scientists from collaborating clinical partners. Further projects will result from the fact that BAC arrays still are the only platform that satisfies the special requirements of formalin-fixed, paraffin embedded tissue, which puts us into a privileged position and will greatly facilitate our access to archived pathological material.

The array technology is an extremely dynamic field, which forces us to actively participate in the development and implementation of new techniques. We already have established gene expression analysis and MeDIP, an array based technique for the detection of methylation patterns, and we are going to implement ChIP on Chip. Furthermore, we will expand our possibilities by introducing customized oligo nucleotide arrays to our laboratory.

In the future, the most important challenge, however, will be the analysis and meta-analysis of data. The Molecular Cytogenetics group will join the department´s initiative to set up a central database, which aims at cross-linking all clinical, genetic, cytogenetic and molecular data generated within the department, which we consider as indispensable for the study of complex traits such as mental retardation. Additionally, in cooperation with Wei Chen, we will integrate new analysis tools into our software package CGHPRO. We expect that the growing importance of meta-analysis will also influence the distribution of manpower within the group, leading to a reduction of the number of technicians, as much of the screening will be done by invited guest scientists, and increasing the input of bioinformaticians and programmers.

Selected Information

Publications

- Erdogan, F., W. Chen, M. Kirchhoff, V.M. Kalscheuer, C. Hultschig, I. Müller, R. Schulz, C. Menzel, T. Bryndorf, H.-H. Ropers and R. Ullmann (2006). Impact of low copy repeats on the generation of balanced and unbalanced chromosomal aberrations in mental retardation. Cytogenetics and Genome Research; accepted.
- Schwarzbraun, T., R. Ullmann, M. Schubert, M. Ledinegg, L. Ofner, C. Windpassinger, K. Wagner, P.M. Kroisel and E. Petek (2006). Characterisation of a de novo complex chromosome rearrangement (CCR) involving chromosomes 2 and 12, associated with mental retardation and impaired speech development. Cytogenetics and Genome Research, accepted.
- Klopocki, E., L. M. Neumann, H. Tönnies, H.-H. Ropers, S. Mundlos and R. Ullmann (2006). Ulnar-mammary syndrome with dysmorphic facies and mental retardation caused by a novel 0.8 Mb deletion encompassing the TBX3 gene. European Journal of Human Genetics, accepted.
- Müller D., E. Klopocki, L.M. Neumann, S. Mundlos, M. Taupitz, I. Schulze, H.-H. Ropers, U. Querfeld and R. Ullmann (2006). A complex phenotype with cystic renal disease. Kidney International, accepted.
- Chen, W., F. Erdogan, H.-H. Ropers, S. Lenzer and R. Ullmann (2005). CGHPRO – a comprehensive data analysis tool for array CGH. BMC Bioinformatics 6(1), 85.
- Wieser, R., B. Fritz, R. Ullmann, I. Müller, M. Galhuber, C.T. Storlazzi, A. Ramaswamy, H. Christiansen, N. Shimizu and H. Rehder (2005). Novel rearrangement of chromosome band 22q11.2 causing 22q11 microdeletion syndrome-like phenotype and rhabdoid tumor of the kidney. Human Mutation 26(2), 78-83.

Neurochemistry Group & Mouse Lab

Head
Dr. Diego J. Walther
Phone: +49 (0)30-8413 1664
Fax: +49 (0)30-8413 1383
Email: dwalther@molgen.mpg.de

Graduate Students
Maik Grohmann
Nils Paulmann
Jens-Uwe Peter

Technicians
Monika Dopatka
Angela Lüttges
Sabine Otto

Undergraduate Students
Tunc Döser
Anja Göbel
Stephanie Grabow
Paul Hammer
Sylvana Henschen
Elisa Lusian
Andreas Naumann
Stephan Pohl
Silke Stahlberg
Yong-joon Suh
Anyess von Bock
Jakob Vowinckel
Maria Walther

Scientific Overview

Our group is focused on the elucidation of molecular causes of human diseases by the generation, analysis, and rescue experiments of transgenic and knockout mouse models. We generate transgenic and knockout mice using classical methods and novel retroviral procedures that allow a rapid generation of knockdown animal models using either ribozymes or shRNA expression vectors. Currently we work on animals with defined genetic dysfunctions in monoaminergic systems and related biochemical pathways aiming the elucidation of the numerous hormonal and neurotransmitter effects of serotonin, histamine, and catecholamines. Furthermore, we study the uptake and release mechanisms from vesicles and the involved signal transduction, using platelets, amongst others, as model system for neuronal vesicular processes.

Tryptophan hydroxylases and serotonin

Tryptophan hydroxylase 1 (TPH1) catalyses the rate-limiting step of serotonin biosynthesis in extraneuronal tissues. Extraneuronal serotonin is involved in primary hemostasis [**Cell** 115 (2003) 851], mammary gland involution [**Dev. Cell** 6 (2004) 193], liver regeneration [**Science** 312 (2006) 104], T cell-mediated immune responses, gastrointestinal function, water and electrolyte homeostasis, and insulin secretion. Our collaborators and we are working on these items and also on the analysis of tissue-specific expression of splicing isoforms of TPH1.
Tryptophan hydroxylase 2 (TPH2) catalyses the rate-limiting step of serotonin biosynthesis in neurons [**Science** 299 (2003) 76; **Biochem. Pharmacol.** 66 (2003) 1673]. The neurotransmitter serotonin is involved in multiple facets of mood control and the regulation of sleep, anxiety, alcoholism, drug abuse, food intake, and sexual behaviour. We and our collaborators are working on the elucidation of TPH2-dependent human psychiatric disorders **[Mol. Psychiatry** 9 (2004) 980].
Platelets can be easily obtained from peripheral blood. Washed platelets are an accepted model for synaptic vesicle metabolism mechanisms. Therefore, the platelets of our *Tph1-/-* mice deliver the first opportunity to study transmitter-devoid vesicles. We are cooperating with G. Ahnert-Hilger to elucidate vesicular trafficking mechanisms. We have recently reported that the vesicular monoamine content regulates VMAT2 activity through Gαq in mouse platelets based on evidence for autoregulation of vesicular transmitter uptake [**J. Biol. Chem.** 278 (2003) 15850]. Similar autoregulation is also given in other neurotransmitter storage systems [**J. Neurosci.** 25 (2005) 4672]. In addition, under physiological conditions, platelets are crucial players in delivering serotonin to a variety of target organs and play a central role for instance in the immune response and in liver regeneration [**Science** 312 (2006) 104].

Monoaminylation of signalling proteins

Recently, we have identified a novel signalling mechanism in platelets, the `serotonylation´ of small GTPases. This novel signal transduction pathway triggers platelet α-granule release [**Cell** 115 (2003) 851]. In analogy to serotonylation, we have found `catecholaminylation´ and `hisaminylation´ to occur as well, and therefore coined the generic term `monoaminylation´ for all of these signalling processes.

This novel role of monoamines in signal transduction consists in a covalent, transglutaminase (TG-)-mediated binding of the monoamines to small GTPases, thereby leading to their constitutive activation in signalling [**Cell** 115 (2003) 851]. Abnormal regulation of small GTPases is an important cause of mental retardation and other disorders; moreover, TGs have been implicated in Huntington's disease. Therefore, we study serotonylation or more generally, monoaminylation, as an attractive novel modifying (and possibly causative) principle in the pathogenesis of genetic diseases.

Transgenesis and ES cell culture

The use of transgenic mice is one of the most straightforward tools to study gene function. Our facility offers the generation of transgenic and knockout mice to all groups in the Department, the Max-Planck-Institute and other interested laboratories. Our facility constitutes in part the Central Facility for Animal Model Generation of the SFB577: "Molecular Basis of Clinical Variability in Mendelian Disorders" providing the techniques pronuclear microinjection, retrovirus-mediated transgenesis, ES cell culture (knockout/knockin), blastocyst microinjection, *in vitro* fertilization, cryopreservation of embryos.
Furthermore, totipotent ES cells can be differentiated *in vitro* giving rise to high yields of neuronal precursor cells. These cells can be differentiated further, delivering for instance serotonergic neurons. The lack of a serotonergic neuronal cell line has hampered studies into TPH analysis for decades. We are establishing such cell lines with the goal to characterize TPH2 expression in a living system.
Another field is the use of harmless prodrugs that are enzyme-specifically toxified in tissues expressing either endogenous tryptophan hydroxylase or transgenic nitroreductase of *E. coli* to induce defined lesions in tissues of interest. The first approach is based on the finding that tryptophan hydroxylase metabolises 7-hydroxytryptophan to 5,7-dihydroxytryptophan, which is rapidly decarboxylated to 5,7-dihydroxytryptamine (5,7-DHT) only in tryptophan hydroxylase-expressing cells. 5,7-DHT is a potent cytotoxic agent in the cytoplasm but not in the extracellular milieu. Therefore, its synthesis *in situ* can be used to specifically target serotonin-producing tumours, such as small cell lung carcinomas and carcinoids [**Cancer** 94 (2002) 3135]. Moreover, this prodrug can be used as an experimental tool for the induction of specific lesions of tryptophan hydroxylase-expressing tissues, when applied at high dosage.
Ecstasy abuse leads to the degeneration and death of serotonergic neurons, thereby causing anxiety and depressive syndromes. Using transgenic animals expressing *E. coli* nitroreductase under control of the *TPH2* promotor and the prodrug CB 1954 we are able to analyse the effect of the loss of serotonergic neurons in the adult brain. This model allows testing therapeutical approaches after the loss of serotonergic neurons. Moreover, we can use our NTR-cassette also to target other neuroendocrine systems, thereby obtaining models for parkinsonism and other neurodegenerative diseases.

Selected Information

Publications

- Lesurtel, M., Graf, R., Aleil, B., Walther, D.J., Tian, Y., Jochum, W., Gachet, C., Bader, M., and Clavien, P.-A. (2006). Platelet-derived serotonin mediates liver regeneration. Science, 312, 104-107.
- Matsuda, M., Imaoka, T., Vomachka, A.J., Gudelsky, G.A., Hou, Z., Mistry, M., Bailey, J.P., Nieport, K.M., Walther, D.J., Bader, M., and Horseman, N.D. (2004). Serotonin regulates mammary gland development via a novel autocrine-paracrine loop. Dev. Cell, 6, 193-203.
- Walther,D.J., Peter, J.-U., Winter, S., Höltje, M., Paulmann, N., Grohmann, M., Vowinckel, J., Alamo-Bethencourt, V., Wilhelm, C.S., Ahnert-Hilger, G., and Bader, M. (2003). Serotonylation of Small GTPases is a Signal Transduction Pathway that Triggers Platelet alpha-Granule Release. Cell, 115, 851-862.
- Walther, D.J., Peter, J. U., Bashammakh, S., Hörtnagl, H., Voits, M., Fink, H., and Bader, M. (2003). Synthesis of serotonin by a second tryptophan hydroxylase isoform, Science, 299, 76.

Associated Group: Molecular Cytology

Head
apl. Prof. Dr. Harry Scherthan
(since 12/2003; guest scientist)
Phone: +49 (0)30-8413 1242
+49 (0)89 3168 2636
Fax: +49 (0)30-8413 1383
Email: schertha@molgen.mpg.de

Scientists
Dr. Edgar Trelles-Sticken (-12/2004)
Dr. Caroline Adelfalk (-5/2005; since then guest scientist)

Graduate Students
Bodo Liebe (-3/2005)
Nils Hartmann (2/2002 - 4/2005)

Technician
Barbara Meinck (-11/2003)

Scientific Overview

Meiosis and recombination reshape the genome and lead to genetic diversity in the offspring. Karyotype reshuffling is a driving force of evolution. In humans, chromosome rearrangements and segregation errors in meiotic cells contribute to aneuploidy, a major cause of pregnancy loss. Our work thus focuses on the understanding of the regulation of chromosome dynamics in germ cell differentiation. Moreover, we are interested in the mechanisms of karyotypic evolution.

Meiosis

Fertilization and embryogenesis depend on halving and doubling of the chromosome complement during gameto/sporogenesis and fertilization, respectively. Chromosome number reduction requires pairing of homologues and a dramatic reorganization of nuclear architecture. Paring and recombination then lead to physical connections between homologues, chiasmata, which are requisite for the segregation of homologues in the meiosis I division. Recombination at illegitimate sites is thought to fuel chromosome rearrangements, pathological ones and those that are seen in evolution. It is important to understand the make-up of break points, fragile sites and recombinogenic sequences in the context of nuclear and genome topology. Since many of the genes involved in meiotic differentiation are required for fertility, we have established tools to fine-stage prophase I progression in the model systems budding yeast and mouse, which allow to detect errors in this complex differentiation progress.

During the time span covered by this report we investigated meiotic chromosome dynamics and telomere behavior in mutants of budding yeast und mouse. In collaboration with R. Benavente and C. Höög we were able to show that chromosome architecture and telomere dynamics are altered in *Sycp3*-/- KO mice, a model for maternal age effect. We could also show that mutations in A-type lamins do lead to fertility problems in male mice, while nuclear architecture in prophase I is largely intact, except for a XY pairing defect. In a collaboration with R. Jessberger we could identify the *Smc1ß* cohesin mutant mouse as a first model for telomere/ NE attachment in mammalian meiosis.
In an attempt to understand the role of other factors determining the architecture of the nuclear envelope for the pairing of meiotic chromosomes and telomere attachment, we are currently finalizing the investigation the spermatogenesis of *Gmcl-1*, *Asm1*mice. We also noted that a chromatin and DSB-repair modifying factor, the histone 2 variant H2AX is required for regulating meiotic telomere dynamics. Currently, we investigate the impact of the DSB repair pathway on meiotic telomere behavior by use of mice deficient for genes like *Spo11*, *Dmc1*, *Hop2*, *Ku80* and *Hr6b*.
In yeast we found that the Mec1 kinase, an ortholog of the human ATR kinase that responds to genotoxic insults and regulates cell cycle progression, regulates telomere clustering at the onset of prophase I. The phenotype of the Set1 histone methyl transferase mutant suggests that epigenetic chromatin modifications play an important role in regulating chromosome dynamics. Using live imaging of a Rap1-GFP strain, we were the first to show a detailed time map of telomere and nuclear motility in live meiocytes of a synaptic organisms and obtained first and conclusive evidence that actin is the major motor protein driving telomere clustering and nuclear dynamics during prophase I of budding yeast (Fig.1). Furthermore, we obtained novel insights in the genetics of telomere cluster exit and showed that functional cohesin is required for bouquet stage exit.

Altogether, we identified major regulators for meiotic telomere dynamics in synaptic meiosis. Also, we further developed our map of prophase I arrest phenotypes and their impact on meiotic chromosome behavior in mammals, which may aid analysis of germ cell differentiation failure in infertile patients. Furthermore, our data will add to the understanding of the generation of aneuploid gametes, which contribute to developmental failure and pregnancy loss.

Karyotype evolution

In this field we worked with model species of Muntjac deer (*Muntiacus muntjac vag.*, MMV) which has the lowest chromosome number (2n=6) among all mammals. During the past 3 years we were able to isolate sequences located at evolutionary breakpoints in the 3 MMV chromosomes, revealing that telomere and GC-rich satellite sequences reside at the breakpoints. We also cloned MMV *Terf1* and *Terf2* genes, which encode telomere-binding proteins and confer karyotypic stability, from two muntjac genomes and detected *Terf* expression and novel muntjac- and human-specific alternatively spliced transcripts. We also determined the expression profile of telomere-relevant genes in Muntjac tissues and cell lines.
Finally, we provided expertise in advanced immunofluorescence and FISH methods as well as microscopy to the department and its students. As guest researchers in the department, we are currently continuing our live cell imaging analysis of yeast mutants and want to investigate the role of Set1and Ndj1 proteins in the regulation of live chromosome dynamics.

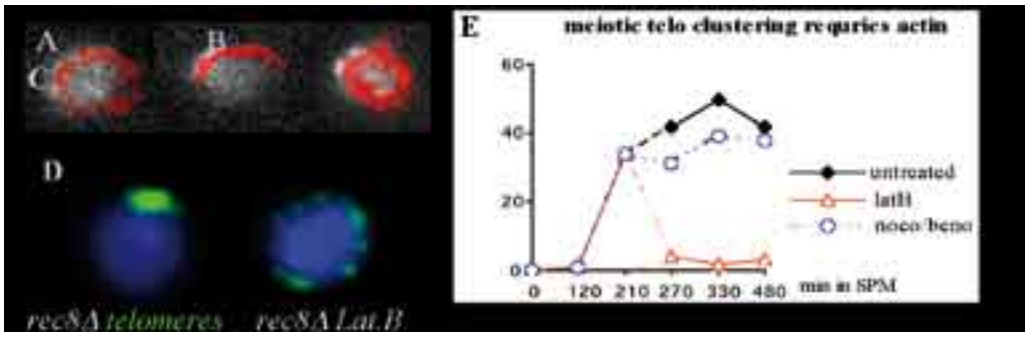


Figure 1: Rap1-GFP telomere movements in live S. cerevisiae meiocytes. (A-C) Trajectories (red) of tracked Rap1-GFP dots marking telomere movements derived of time-lapse series at a fixed focal plane. (A) Lateral mobility around the leptotene nuclear periphery leads to a web of trajectories around the nucleus. (B) Telomeres motility is restricted to nuclear periphery sector of a rec8Δ bouquet nucleus. (C) The broad rim of trajectories around the nucleus reflects the numerous nuclear deformations at pachynema. (D) Telomeres (green) form a single cluster in a rec8Δ meiocyte. Inhibition of actin polymerization (rec8Δ Lat) disperses telomeres. (E) The frequency of rec8Δ meiocytes with a single telomere cluster in the presence of no (untreated), microtubule (noco/beno) or actin (latB) poisons during sporulation. Telomere clustering is sensitive to inhibition of actin polymerization (see Trelles-Sticken *et al.* 2005 for details).

Selected Information

Publications

- Scherthan, H. (2006) Meiotic telomeres. In "Telomeres, 2nd ed.". T. de Lange, V. Zakian and E. Blackburn, eds. CSH Press, Cold Spring Harbor, pp 225-259.
- Trelles-Sticken E., Adelfalk C., Loidl J. and Scherthan H. (2005) Meiotic telomere clustering requires actin for its formation and cohesin for its resolution. J Cell Biol. 170: 213-223.
- Hartmann N. and Scherthan H. (2005) Characterization of the telomere complex, Terf1 and Terf2 genes in muntjac species with fusion karyotypes. Exp. Cell Res., 306:64-74.
- Weiss-Schneeweiss H., Riha K., Gee Jang C., Puizina J., Scherthan H., and Schweizer D (2004) Chromosome termini of the monocot plant Othocallis siberica are maintained by telomerase which specifically synthesises vertebrate-type telomere sequences. Plant J. 37:484-493.
- Hartmann N. and Scherthan H. (2004) Characterization of ancestral chromosome fusion points in the Indian muntjac deer. Chromosoma 112:213-20.
- Liebe B., Alsheimer M., Höög C., Benavente R., Scherthan H. (2004) Telomere attachment, meiotic chromosome condensation, pairing and bouquet stage duration are modified in spermatocytes lacking axial elements. Mol. Biol. Cell, 15:827-837.
- Revenkova E., Eijpe M., Heyting C., Hodges C.A., Hunt P.A., Liebe B., Scherthan H. and Jessberger R. (2004) SMC1ß is required for meiotic chromosome dynamics, sister chromatid cohesion, and DNA recombination. Nat. Cell Biol. 6:555-562.
- Roig I., Liebe B., Egozcue J., Cabero L, Garcia M., and Scherthan H. (2004) Female-specific features of recombinational double-stranded DNA repair in relation to synapsis and telomere dynamics in human oocytes. Chromosoma 113:22-33.
- Voet, T., Liebe, B., Labaere C., Marynen P., and Scherthan H. (2003) Telomere-independent homologue pairing and checkpoint escape of accessory ring chromosomes in male mouse meiosis. J. Cell Biol., 162:795-808.

Selected Invited Plenary Lectures

- 3rd European Telomerase Meeting, Ladenburg, Ger. Plenary lecture 11/2004.
- Gordon Research Conference on Meiosis, Chair and plenary lecture, NH, USA, 6/2004
- Riga Meeting of Comprehensive Cell Biology. Riga, Latvia. 5/2005
- 7th European Meiosis Meeting / EMBO Lecture Course. Plenary lecture, Madrid, Spain. 7/2005.
- Joint meeting of the Biochem. Soc. and the Genet. Soc., Univ. of Warwick, UK. 3/2006

Teaching

- Course & Lecture Basics in molec. Cytology, each term; TU Kaiserslautern

External funding

- DFG, Sche 350/8-4: 1 Postdoc & 2 PhD students funded -5/2005

Work as scientific referee

- Chromosoma, Dev. Cell, Dev. Biol., J. Cell Biol., J. Cell Sci., Reprod., The Wellcome Trust., German Israeli Found., DFG; Editorial board member Cytog.Gen.Res., 2000-05

Cooperation

- R. Benavente, Univ. Würzburg, Biozentrum, D
- R. Jessberger, Dresden Univ. of Techn., D
- J. Loidl, Dept. of Chrom. Biol., Univ. Vienna, A
- V. Geli, CRNS, Marseille, F
- S. Keeney, Sloan Ket. Cancer Cent., NYC, USA;
- D. Camerini-Otero, NIH, Bethesda, USA;
- Dr. M. Digweed, Charité Berlin, D

Awards

- 41st rank of the 50 most frequently cited German Molecular Biologists (Lab J., 4/04)

Department of Computational Molecular Biology

Head
Prof. Dr. Martin Vingron
Phone: +49 (0)30-8413 1150
Fax: +49 (0)30-8413 1152
Email: vingron@molgen.mpg.de

Secretary
Birgit Löhmer
Phone: +49 (0)30-8413 1151
Fax: +49 (0)30-8413 1152
Email: vinoffic@molgen.mpg.de

Assistant
Dr. Patricia Beziat
Phone: +49 (0)30-8413 1716
Fax: +49 (0)30-8413 1671
Email: patricia.beziat@molgen.mpg.de

Computer Systems Administrator
Wilhelm Rüsing
Phone: +49 (0)30-8413 1160
Fax: +49 (0)30-8413 1152
Email: wilhelm.ruesing@molgen.mpg.de

tment of Computational Molecular Biology

Introduction

The research of the Computational Molecular Biology Department focuses on the analysis of the data generated by today's sequencing and functional genomics programs. Numerous challenging questions can be posed based on these data concerning, e.g., whole genome comparison, gene and genome evolution, gene structure, gene regulation, the analysis of large scale gene expression data, and their utilization for disease diagnosis. The department is structured into several **research groups**, the largest of which is the Transcriptional Regulation group headed by Martin Vingron. The work of this group focuses on theoretical concepts in the prediction of cis-regulatory elements, gene regulatory networks and epigenetic aspects of regulation. Rainer Spang heads the Computational Diagnostics group which deals with microarray data analysis in the context of cancer profiling. The Algorithmics group is headed by Alexander Schliep and develops machine learning algorithms for pattern recognition in a number of biological problems. Peter Arndt heads the Evolutionary Genomics group which works on developing models how the DNA in primates has evolved. Stefan Haas is heading the Gene Structure and Array Design group. The focus of this group lies on exploiting the genomic information available today in order to derive the structure of genes with respect to alternative splicing and alternative transcriptional start sites. Roland Krause has started to build up a group on "Microbial Virulence" linking the MPI for Molecular Genetics and the MPI for Infection Biology. This cooperation shall apply bioinformatics methods on question that are studied in infection biology. The head of the Protein Families and Evolution group, Eike Staub recently left the institute. We report below on the results which he achieved, although it is not planned to continue this effort on the level of a separate group. The projects that still fall under this topic are currently supervised by Martin Vingron.

The **computer equipment** of the department comprises PCs under Linux, a cluster of 32-bit processors largely for sequence analysis and typical bioinformatics applications, two new 4-processor 64-bit computers with large memory (64 GB) for numerical computations, and several RAID arrays with together more than 10 TB of disk space. We are in the process of acquiring a new cluster of 64-bit processors. The 16 processor compute server from COMPAQ that was acquired in 2001 is still largely functional and will remain in use as long it does not incur any repair cost. The whole set-up is maintained by the department system administrator, Wilhelm Rüsing, in close cooperation with the institute computing unit.

Department members contribute substantially to the **bioinformatics curriculum** at Free University of Berlin. We teach a number of courses and offer students to do internships, practical courses, and thesis work with us. This brings many bright, young students to the department and at the same time allows the university to show the students a much larger spectrum of bioinformatics than would normally be possible in the university framework. In cooperation with the university we have established an **International Max Planck Research School on Computational Biology and Scientific Computing** (IMPRS-CBSC, *http://www.imprs-cbsc.mpg.de/*). In fall 2006 the second generation of students who have been recruited through an international selection procedure have taken up their work. With funding from the Max Planck Research Award we have initiated the International Otto Warburg Summer Schools (*http://ows.molgen.mpg.de/*). The first such school was held in summer 2005 in Berlin and was devoted to "Networks and Regulation", the second one will be on "Evolutionary Genomics" in early September 2006. The schools bring together international lecturers with a select group of national and international students and combine lecture-style teaching and research seminars to give an overview of a new important area in computational biology.

During the report period we have been involved in a number of **national and international projects** and collaborations. Most prominently, we are part of the Berlin Center for Genome Based Bioinformatics (BCB), a large network made up of several Berlin bioinformatics groups. This project is funded by the German Minstry of Science (BMBF) and will end in December 2006. The "Computational Diagnostics" group headed by R. Spang is part of BCB and is also active in the German National Genome Research Network (NGFN), providing education and know-how in microarray data analysis. We are part of the EU Network of Excellence in Bioinformatics "BioSapiens", provide promoter informatics within NGFN, and are part of an SFB on theoretical biology.

Selected Information

Honors and awards

- Max Planck Research Award for Bioinformatics, 2004, to Martin Vingron and Gene Myers.
- Martin Vingron elected to German Academy of Natural Scientists Leopoldina

Professorships accepted by former department members

- Jörg Schultz, C3, Universität Würzburg, 2003
- Antje Krause, C3, Fachhochschule Wildau, 2004

Evolutionary Genomics

Head
Dr. Peter F. Arndt
Phone: +49 (0)30-8413 1162
Fax: +49 (0)30-8413 1152
Email: arndt@molgen.mpg.de

Graduate Students
Philipp W. Messer (since 2/2005)
Paz Polak (since 7/2006)
Federico Squartini (since 9/2005)

The focus of the evolutionary genomics group lies in the analysis of genomic data with respect to evolutionary aspects. Today many completed genomes are available, which gives us the opportunity to elucidate the evolutionary processes that acted upon these genomes. For example, we study the emergence of the compartmentalization or so-called isochore structure of vertebrate genomes. Both comparative methods and the study of genomic fossils (e.g. retroviral sequences, pseudo genes) are used to learn more about the particular processes that have shaped vertebrate genomes. Processes on short length scales, e.g. nucleotide substitutions, insertions and deletions and on longer length scales, e.g. recombination and segmental duplications, are investigated. Implications to functional evolution are also addressed.

Comparative Analysis – Phylogeny

With several complete genomes at hand a comparative analysis of genomic sequence data is the method of choice to study evolutionary processes. Variations of these processes along chromosomes and back in time into the deeper branches of the phylogeny can be mapped. In contrast to the commonly used assumptions our analysis does not assume that the underlying processes are time reversible or that a stationary state has been reached. We also include neighbor dependent substitution processes as for instance the CpG methylation deamination, which turns out to be the predominant nucleotide substitution process in vertebrates. Therefore, this process needs to be incorporated into the modeling beyond the usual scheme of independently evolving nucleotides. Taking into account the non-reversibility of the underlying processes a much richer picture of genomic sequence evolution emerges revealing more details on the underlying processes of molecular evolution across different times in our evolutionary history as well as across different regions along chromosomes.

Recently the just completed HapMap project provides ample new data complementing the genomic data with information on single nucleotide polymorphisms (SNPs). Comparing fixed and still polymorphic differences will allow us to analyze the processes involved in fixating evolutionary innovations in a population.

Models for genomic evolution

The presence of long-range correlations in the nucleotide composition along the genomic sequences of vertebrates is still poorly understood. This so-called isochore structure was first recognized more than 30 years ago, however no convincing model of their emergence has been established so far. Recently, it had been shown that simple expansion randomization systems (ERS) are able to generate such long-range correlated sequences. A wide range of such systems fall within one universality class and the characteristic decay exponent of the correlation function can easily be calculated from the rates of the underlying processes. This result gives us also a simple method to simulate long-range correlated sequences and recently we were able to quantify the influence of such correlations on the alignment statistics of sequence, which turned out to be quite substantial. Corresponding corrections should be taken into account when calculating p-values for the alignment of genomic sequences.

The key ingredients of these expansion randomization systems are processes that duplicate, insert, or delete segments of a sequence. These processes need to be investigated and their rates to be quantified in the genomic context. The insertion of repetitive elements (300-6000 bp) into the vertebrate genomes is one important such process, but duplication processes can also be observed on much smaller length scales. At the end we will generate a much richer null model for genomic evolution.
This will allow us to distinguish neutral changes from potentially adaptive ones and will help us to understand the evolution and design of promoter regions.

Evolution of Transcription Factor Binding Sites

The evolution of transcriptional regulatory elements is hard to track on a genome wide scale. On the other hand, from observations on one or two genes one may not be able to deduce a general principle. Out of this reason we studied an intermediate sized gene set - the cytosolic ribosomal protein genes – in several mammalian and vertebrate lineages. These 79 genes are tightly regulated since at least one of each of the encode proteins is needed to assemble one ribosome. These genes are an ideal set to study common feature of the promoters in one species and to compare them with features observed for other species. This way we were already able to describe a novel binding site in the proximal promoter for 31 of these genes in the mammalian lineages. Corresponding sites are not present in fishes. In-house wet lab biologists (Diego J. Walther, Department Ropers) are right now analyzing this system to identify the binding factor and its role in the regulation of ribosomal proteins.

Selected Information

Publications

- Stefan Roepcke, Degui Zhi, Martin Vingron, Peter F Arndt (2006). Identification of highly specific localized sequence motifs in human ribosomal protein gene promoters. Gene 365 (2006) 48-56
- Philipp W. Messer, Ralf Bundschuh, Martin Vingron, and Peter F. Arndt (2006). Alignment Statistics for Long-Range Correlated Genomic Sequences A. Apostolico *et al.* (Eds.): RECOMB 2006, LNBI 3909, pp. 426-440.
- Peter F Arndt, Terence Hwa (2005). Identification and measurement of neighbor dependent nucleotide substitution processes. Bioinformatics 21 (2005) 2322-2328
- Philipp W Messer, Peter F Arndt, Michael Lassig. Solvable sequence evolution models and genomic correlations. Physical Review Letters 94 (2005) 138103

Invited talks

- Arndt P. F. Substitution pattern of mammalian transposable elements, 1st International Conference on the Genomic Impact of Eukaryotic Transposable Elements, Asilomar, 31.3. – 4.4.2006
- Arndt P. F. Reconstructing the evolutionary history of mammalian genomes, Symposium on Structural and Evolutionary Genomics, Puntarenas, Costa Rica, 26.2. –2.3.2005
- Arndt P. F. Regional and Time-resolved Base Substitution Pattern of the Human Genome, Stazione Zoologica Anton Dohrn, Naples, 26.03.2004

Diploma thesis

- Philipp Messer „Stochastic Sequence Evolution and Genomic Correlations", 2005, Universität Köln

Gene Structure and Array Design

Head
Dr. Stefan Haas
Phone: +49 (0)30-8413 1164
Fax: +49 (0)30-8413 1152
Email: stefan.haas@molgen.mpg.de

Graduate Students
Shobhit Gupta (-1/2006)
Helge Roider

Alternative splicing

Despite the seemingly small number of genes in higher eukaryotes a much larger diversity is observed on the level of transcripts or proteins. This is mainly due to the process of alternative splicing that is expected to generate alternative transcripts per gene in at least 60-70% of the genes studied so far.
Our work focuses on the prediction of alternative splicing events by analysing the exon-intron structure of transcripts based on EST sequences. Since the annotation and sequence quality of ESTs varies a lot we extensively filter the set of EST sequences in order to provide a robust basis for subsequent exon structure evaluation. After aligning EST consensus sequences to the genome sequence we assign a confidence value to each splice donor/acceptor site, finally allowing us to reliably predict alternatively spliced exons. In tight collaboration with the RZPD we could show that the vast majority of our prediction tested experimentally is in fact expressed in one or more tissues.
In a more detailed analysis we use the annotations attached to ESTs to categorize transcripts according to tissue type, developmental stage, disease state, or methodological differences like normalization. This way we now additionally predict in which of these categories certain transcripts are over represented. Our experiments demonstrated that ESTs can successfully be used to predict the expression of a transcript in a discrete tissue. However, for alternative transcripts predicted to be exclusively expressed in tumours our experimental results suggested a low-level expression of these isoforms in a variety of healthy tissues.

Alternative promoters

Recent evidence from large-scale datasets tailored to determine transcriptional start sites (TSS) suggests that a significant number of alternative transcripts are caused by alternative promoters rather than by alternative splicing. The regulation of alternative splicing and alternative promoter usage involves signals specific to the underlying mechanisms. While alternative splicing is regulated via exonic/intronic splicing enhancers or silencers, the use of alternative promoters is mainly defined by the binding of distinct transcription factors, the methylation status of the promoter, or the overall chromatin structure. Functionally, alternative promoters more likely influence the expression pattern or cellular localization of the respective gene product rather than changing the encoded protein since alternative first exons usually cover only the very C-terminal part of a protein.
The computational analysis of overrepresentation of regulatory elements, as performed in our CORG project, relies on the correct definition of the transcriptional start sites sincea major fraction of functional transcription factor binding sites is expected to be in close proximity to the TSS. We therefore recently developed methods to reliably differentiate between alternative first exons and alternatively spliced exons. Preliminary investigation of our EST data emphasizes that at least 50% of the tissue-specifically expressed isoforms may be caused by alternative promoters. By the integration of our EST-based predictions with full-length clone data (DBTSS) and CAGE data we generated a comprehensive set of TSSs that will be used for an improved detection of functional regulatory elements.

Tissue-specific expression of genes and transcripts

A basic requirement for the analysis of gene regulatory networks is to group genes acting within a common context. Intuitively, the expression pattern of genes with respect to tissues constitutes a biologically reasonable constrain. Genes part of such an anatomical category are likely to share regulatory features like common binding of certain transcription factors, or modification of chromatin structure.
Using statistical methods, we therefore define sets genes specifically expressed in a certain tissue or cell type based on the expression information derived from our GeneNest database. The subsequent analysis of potential transcription factor binding sites common to promoter regions of genes within one category together with those that are significantly different in promoters of genes from categories, will facilitate the detection of key regulators of gene expression in the respective tissue.

Chip design

The design of PCR primers or oligos is an essential step when developing DNA-microarrays since the selection of primers will affect the potential of successfully interpreting the experimental resultsobtained from such an array. We implemented a primer design software (GenomePRIDE) for the design of whole transcriptome arrays for S. pombe, Drosophila, etc. Recently, we extended GenomePRIDE to the design of splice isoform specific primers as well as to generate primers for the amplification of genomic DNA fragments for ChIP experiments in the context of large-scale promoter analysis project.

Selected Information

Publications

- Hecht, H., Kuhl, H., Haas, S.A., Bauer, S., Poustka, A.J., Lienau, J., Schell, H., Stiege, V., Seitz, V., Reinhardt, R., Duda, G.N., Mundlos, S. and Robinson, P.N. (2006) Gene Identification and Analysis of Transcripts Differentially Regulated in Fracture Healing by EST Sequencing in the Domestic Sheep. BMC Genomics, 7:172
- Hui, J., Hung, L.H., Heiner, M., Schreiner, S., Neumuller, N., Reither, G., Haas, S.A, Bindereif, A (2005) Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. EMBO J., 24(11), 1988-1998
- Gupta, S., Zink, D., Korn, B., Vingron, M. and Haas, S.A. (2004) Strengths and weaknesses of EST-based prediction of tissue-specific alternative splicing. BMC Genomics, 5:72
- Gupta, S., Zink, D., Korn, B., Vingron, M. and Haas, S.A. (2004) Genome wide identification and classification of alternative splicing based on EST data. Bioinformatics, 20(16), 2579-2585
- Boutros, M., Kiger, A., Armknecht, S., Kerr, K., Hild, M., Koch, B., Haas, S.A., Heidelberg FlyArray, Paro, R. and Perrimon, N. (2004) Genome-wide RNAi Analysis of Cell Growth and Viability in Drosophila Cells. Science, 303 (5659), 832-835
- Hild, M., Beckmann, B., Haas, S.A., Koch, B., Solovyev, V., Busold, C., Fellenberg, K., Boutros, M., Vingron, M., Sauer, F., Hoheisel, J.D. and Paro, R. (2003) An integrated gene annotation and transcriptional profiling approach towards the full gene content of the Drosophila genome. GenomeBiology, 5:R3
- Kalscheuer, V., Freude, K., Musante, L., Jensen, L.R., Yntema, H.G., Gecz, J., Sefiani, A., Hoffmann, K., Moser, B., Haas, S., Gurok, U., Haesler, S., Aranda, B., Nshedjan, A., Tzschach, A., Hartmann, N., Roloff, T.-C., Shoichet, S., Hagens, O., Tao, J., van Bokhoven, H., Turner, G., Chelly, J., Moraine, C., Fryns, J.-P., Nuber, U., Hoeltzenbein, M., Scharff, C., Scherthan, H., Lenzner, S., Hamel, B.C.J., Schweiger, S. and Ropers, H.-H. (2003) Mutations in the polyglutamine-binding protein 1 gene cause X-linked mental retardation. Nature Genet., 35 (4), 313-315
- Haas, S.A., Hild, M., Wright, A.P.H., Hain, T., Talibi, D., Vingron, M. (2003) Genome-scale design of PCR primers and long oligomers for DNA microarrays. Nucleic Acids Res., 31 (19), 5576-5581

Invited talks

- Analysis of alternative splicing, Molecular Interactions Workshop, Berlin, Germany, 2005
- EST-based prediction of alternative promoters, University of Frankfurt/Main, Germany, 2006
- Prediction of alternative promoters, ATD conference, EMBL, Heidelberg, Germany, 2006

PhD Thesis

- Shobhit Gupta
  EST-based detection and analysis of mammalian transcripts, 2005

Cooperations

- Prediction and experimental analysis of alternative splice variants based on ESTs with B. Korn, Resource Centre, Heidelberg (until 2005)
- Analysis of conserved intronic sequences in the context of alternative splicing with A. Bindereif, University of Gießen
- Design of a whole genome microarray of Drosophila melanogaster with M. Hild, R. Paro, J. Hoheisel, ZMBH+DKFZ, Heidelberg (2001-2004)
- Design of DNA fragments for large-scale RNAi experiments with M. Boutros, DKFZ Heidelberg (2004)
- Design of a whole genome microarray of Schizosaccharomyces pombe with A. Wright, Södertörns Hökskola, Huddinge, Sweden (2002-2003)
- Design of a flexible DNA-microarray for parallel use of PCR fragments in expression cloning and RNAi (Anopheles gambiae) with G. Christophides, F. Kafatos, EMBL, Heidelberg (2003-2004)
- Representing genes on DNA-chips by a minimal set of short oligonucleotides with M Beier, FeBit AG, Mannheim
- Analysis of a genomic region on chromosome X related to X-linked mental retardation with H.-H. Ropers, Dept. of Human Molecular Genetics, MPIMG (2003)
- Generation of a gene index database of the domestic sheep (Ovis aries) with P. Robinson and S. Mundlos, Charité, Berlin (2005-2006)

Microbial Virulence

Head
Dr. Roland Krause
(since 2005)
Phone: +49 (0)30-8413 1158
Fax: +49 (0)30-8413 1152
Email: roland.krause@molgen.mpg.de

The group is set up as permanent collaboration with the Department of Cellular Microbiology of Prof. Zychlinsky at the Max-Planck-Institute for Infection Biology. The department studies the different strategies of the innate immune system used by higher organisms to defend themselves against various microbial agents. We are interested in the interactions of microorganisms and their hosts and aim to understand why some interactions lead to human disease. Our group uses computational approaches to determine the function of genes involved in the pathomechanisms, both in the microorganisms and the host organisms. Our integrated setup allows us to take bioinformatics predictions to the bench quickly and experimental feasibility is considered at the level of bioinformatics analysis. We focus our experimental work on the Enterobacteriaceae *Salmonella typhimurium* and *Shigella flexneri* but the computational techniques can be easily applied to other pathogens. The genomes of the Enterobacteriaceae, including *Shigella* and related pathogenic and non-pathogenic strains, are particularly well represented amongst the more than 300 prokaryote genomes sequenced.

Protein-protein interaction

When the group was started in January 2005, I was invited to collaborate with Cellzome AG and the EMBL, Heidelberg in the analysis of protein complexes. Biochemical purification of protein complexes and mass spectrometric identification of their components were carried out for the whole genome of *Saccharomyces cerevisiae*. I had participated in an earlier analysis and supported the screen during the time of my graduate work from 2000-2003.
The information comprised more than 2000 individual purifications. Important steps in the analysis were the comparison of the protein-protein interactions to previously generated information with respect to localization, function and membership in established protein complexes. The work led to a publication in early 2006 (Gavin *et al.*).

Host-pathogen interaction is governed by the same biochemical principles as interspecies protein-protein interaction. We want to apply the insights gained to the research in infectious diseases. One computational technique – phylogenetic profiling – that was initially used to predict interacting proteins can be extended to predict novel virulence factors.

Identification of virulence factors by comparative genomics

About 5% of the genes of *Salmonella* are estimated to be involved in the mechanisms of infection. While many factors have been identified, several features in the virulence are still to be discovered. We are using comparative genomics to identify and assess novel or weakly characterized virulence factors.
Genes involved in microbial pathogenesis have been identified based on their distribution in the genome. Genes coding for virulence factors are often located in pathogenicity islands (PAI) in *Salmonella* and *Shigella*. While PAIs are of great importance, several factors have been found recently that are not encoded in adjacency to other virulence genes and it is believed that a substantial number of factors have not been discovered.

Virulence factors are likely to be present in virulent, but absent in avirulent strains. This notion of phylogenetic profiling was used to delineate genes found enriched in pathogens. We generated a gene list that includes the known virulence factors as well as about 100 uncharacterized candidates. We assessed candidate genes by integrating additional data from large scale screens of mRNA-transcription, gene knock outs, and protein-protein interactions. We are currently testing our predictions in *Salmonella typhimurium*.

Prediction of virulence factors in *Mycobacterium tuberculosis*

Presence and absence of genes is an effective yet rather coarse approach to characterize microbial virulence factors. Single mutations could render a strain avirulent while the gene content would be identical to a pathovar. Such features can be studied by several techniques, typically based on microarrays. Recently, whole genome re-sequencing of closely related strains has become a feasible and powerful approach to the problem.
In collaboration with the Department of Immunology of Prof. Kaufmann at the Max Planck Institute for Infection Biology, we compared an established virulent strain and a derived, avirulent lab strain of *Mycobacterium tuberculosis*. Both strains are well characterized *in vivo* and *in vitro*. The complete genome of the avirulent strain was sequenced and was subjected to close inspection by bioinformatics methods. The analysis suggested that most of single changed features were errors in the genome sequence of the virulent strain that was already published. This was confirmed by re-sequencing of the virulent strain and finally suggests that the loss of virulence was caused by a single mutation.

Outlook

Other organisms studied at the Department of Cellular Microbiolgy include the yeast *Candida albicans* and gram positive bacteria such as *Staphylococcus aureus*. Ultimately, we want to build integrated models that integrates all information from genomic, proteomic and other high-throughput techniques and refine the notion of virulence factor.

Selected Information

Publications

- Chu, W., Ghahramani, R., Krause, R., Wild, D.L., (2006) Identifying Protein Complexes in High-Throughput Protein Interaction Screens Using an Infinite Latent Feature Model. (Pacific Symposium on Biocomputing, 2006)
- Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B., Edelmann, A., Heurtier, M.A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A.M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J.M., Kuster, B., Bork, P., Russell, R.B., Superti-Furga, G. (2006) Proteome survey reveals modularity of the yeast cell machinery. Nature, 440,631-6.

Cooperations

- Protein-protein interaction with Cellzome AG, Heidelberg, and EMBL, Heidelberg
- Mycobacterium tuberculosis with H.E. Kaufmann, Dept. of Immunology, Max Planck Institute for Infection Biology, Berlin

Algorithmics

Head
Dr. Alexander Schliep
Phone: +49 (0)30-8413 1166
Fax: +49 (0)30-8413 1152
Email: alexander.schliep@molgen.mpg.de

Graduate Students
Ivan Gesteira Costa Filho
Wasinee Rungsarityotin
Benjamin Georgi

Research in the Algorithmics group focuses on methods from mathematics and statistics which are crucial for answering relevant biological questions. An emphasis is put on analyzing high-dimensional and heterogeneous data, such as time-courses, and on optimizing experimental designs.

Optimizing experimental designs

DNA-Microarrays are best known for their use in gene expression analysis. The same underlying hybridization reactions can also be used to infer presence and absence of biological agents, say viruses or bacteria, in a sample from the hybridization pattern of oligonucleotide probes to genomic DNA of the agents. This detection is crucial for epidemiological studies and vaccine design when used for detection of viral subtypes in Influenza or HIV, in food and safety control, in studies of microbial diversity or, particularly in the USA, in bio-threat reduction.

Due to close evolutionary relationships between agents oligonucleotide probes uniquely identifying agents cannot be found in many applications; consider identifying all Influenza virus subtypes. Nevertheless, even non-unique probes can be used for the detection with great success as there is a connection to the mathematical field of statistical group testing. It bridges across combinatorial design theory, Bayesian statistics and Markov Chain Monte Carlo methods.

Based on prior work at Los Alamos National Laboratory we have implemented a method to select oligos for DNA chips in situations where due to a high degree of sequence similarity unique oligos cannot be found. This will be applied to the analysis of meiobenthos samples, as well as to HIV subtyping, where the very high incidence of multi-viral HIV infections such as in populations in Southern Africa make analysis difficult. Further work on the theoretical side includes optimization of the underlying combinatorial designs and modeling more of the underlying biology – e.g. phylogenetic information in the analysis step. We have proposed the first method which can, with a success rate of up to 70%, correctly identify previously unknown agents in biological samples.

The same general theory has been very successfully used for example for the first physical map of Human chromosome 16. Other applications elsewhere include screening blood donations and optimizing yeast2hybrid experiments; we actively investigate further experimental techniques. We closely collaborate with David Torney (LANL), who first proposed and implemented group testing for molecular biology applications.

Clustering Heterogeneous Data

Detecting relevant groups in data is still one of the central unsolved problems. The massive amounts of data created in molecular biology, the high error rates and the question of how to combine several heterogeneous data sets, pose challenges for research. We have contributed clustering based on mixture models, novel component models for specific data types and the framework of partially supervised learning for the integration of sparse heterogeneous data. A complimentary technique for fusing heterogeneous datasets is based on a a naive Bayes approach coupled with an innovative way of minimizing model complexity. Furthermore, we also developed a methodology for validating clusterings of genes with functional annotation. The results of our research are successfully used in the following areas:

– Analyzing gene expression time-courses
– Analyzing gene expression during development of blood cells
– Joint analysis of in-situ hybridization and gene expression time-courses for Drosophila
– Inference of protein complexes from purification data
– Joint analysis of genotype and phenotype data from ADHD-patients
– Detection of functional residues in protein sequences
– Clustering protein sequences for detection of remote homologues

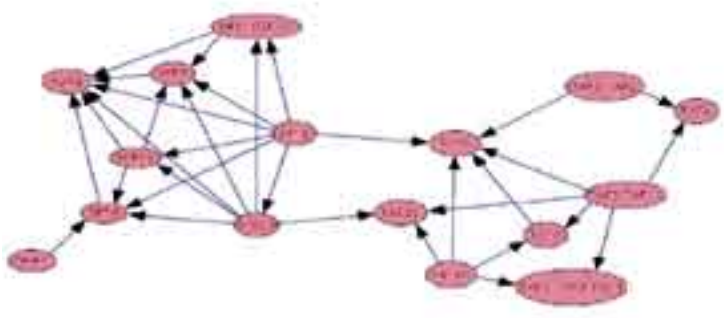


Figure 1: Clustering of Protein copurification data: Collaborative work of Algorithmics

Hidden-Markov-Models (HMMs)

Hidden-Markov-Models, originally developed for speaker-independent speech recognition, have been widely used in their basic form as so-called Profile HMMs for the detection of remote homologs, or in the slightly more complex form of labeled HMMs, for finding eukaryotic genes. The basic framework supports a number of extensions; they can also be used for either classification or clustering.

On one hand our work with HMMs concerned itself with learning HMM topology and different training methods. On the other hand, we investigated novel applications using HMMs as qualitative time-series models and, among others, non-Markovian HMM extensions. We proposed the framework of partially supervised learning for both clustering and mixture modeling. Furthermore, we develop the only free (licensed under the LGPL) library for HMMs, the General Hidden Markov Model Library (GHMM), which is widely used both in industry and academia. Our group uses HMMs in the following applications:

– Analyzing gene expression time-series data,
– Detecting chromosomal aberrations from ArrayCGH and gene expression data
– Ab-initio gene prediction
– Analysis of bird-song data

Visualization

Both teaching and research in algorithms are accelerated by computer tools which allow to experience the dynamic nature in a rich multi-medial environment. Gato, the Graph animation toolbox, provides such an environment. Due to its flexibility and (semi-) automated visualization of user-implemented graph algorithms, it surpasses the capabilities of existing products. A Springer textbook, covering an introduction to combinatorial optimization, is forthcoming. As an extension, visualization of bioinformatics algorithms is under research as well as a graphical tool for working with Hidden-Markov-Models.

Selected Information

Publications

- A. Schliep, S. Rahmann (2006). Decoding Non-Unique Oligonucleotide Hybridization Experiments of Targets Related by a Phylogenetic Tree, accepted to proceedings of ISMB 2006
- G. W. Klau, S. Rahmann, A. Schliep, M. Vingron, K. Reinert. (2006). Integer Linear Programming Approaches for Non-Unique Probe Selection. To appear in: Discrete Applied Mathematics
- A. Schliep, C. Steinhoff, A. Schönhuth (2004). Robust inference of groups in gene expression time-courses using mixtures of HMMs Bioinformatics. Aug 4;20 Suppl 1:I283-I289 (Proceedings of the ISMB 2004).
- A. Schliep, A. Schönhuth, C. Steinhoff (2003). Using Hidden Markov Models to Analyze Gene Expression Time Course Data. Bioinformatics. 19 Suppl 1: i255-63 (Proceedings of the ISMB 2003).
- A. Schliep, D. C. Torney, S. Rahmann (2003). Group Testing With DNA Chips: Generating Designs and Decoding Experiments. Proceedings of the 2nd IEEE Computer Society Bioinformatics Conference (CSB 2003).

Invited plenary lectures

- 24. Mai 06: Treeprobes: Grouptesting biological agents related by phylogenetic trees. Los Alamos National Laboratory.
- 19. Mai 06: Group testing DNA Microarrays for detection of biological agents. DIMACS workshop on combinatorial group testing.
- 26. Oktober 05: Analyzing ArrayCGH Data using Non-Homogenous Hidden Markov Models. Royal Institute of Technology (KTH), Stockholm
- 9. Juni 05: Analyzing Gene Expression Time-Courses. Deutsches Krebsforschungszentrum, Heidelberg
- 21. August 04: Designing Microarrays for Detection of Biological Agents. The Center for Advancement of Genomics (TCAG), Rockville, MD
- 19. Juli 04: Inference of Groups in Gene Expression. The 2004 Meeting of the International Federation of Classification Societies, Chicago, IL
- 3. October 03: Analyzing Gene-Expression Time-Series Data. Center for Bioinformatics and Molecular Biostatistics, University of California, San Francisco, CA
- 25. September 03: Mixture models of Heterogenous Data. T-10 Group Seminar, Los Alamos National Laboratory, NM

Diploma theses

- Michael Seifert Analyzing Microarray Data Using Homogenous and Inhomogenous Hidden Markov Models. Diplomarbeit im Studiengang Bioinformatik, Martin-Luther-Universität Halle (2006)
- Matthias Heinig. Development of a Pair HMM based Gene Finder for the Paramecium Genome. M.Sc. Bioinformatics, FU Berlin (2005)
- Benjamin Georgi. Mixture Modeling and Group Inference in Fused Genotype and Phenotype Data M.Sc. Bioinformatics, FU Berlin (2005)
- Wasinee Rungsarityotin. Graph-based clustering for biological data. M.Sc. Bioinformatics, FU Berlin (2004)

Cooperations

- Analysis of ADHD-Data with M. A. Spence, College of Medicine, Dept. Human Genetics, University of California at Irvine
- DNA-Microarrays for the analysis of microbial communities with D. Tautz, Institut für Genetik, Universität zu Köln
- Analysis of functional sites in protein sequences with J. Schultz, Biozentrum, Universität Würzburg
- Development of a Pair Gene Finder with V. Schachter, M. Heinig, Genoscope, Paris
- Group testing with D.C. Torney, Los Alamos National Laboratory
- Analysis of ArrayCGH data from BAC-clone arrays with H.-H. Ropers, Wei Chen, Dept. of Human Molecular Genetics, MPIMG
- Algorithm visualisation with W. Hochstättler, Institut für Mathematik, Fernuniversität Hagen
- Analysis of in-situ gene expression during Drosophila development with S. Posch, Institut für Informatik, Martin-Luther-Universität Halle-Wittenberg
- Predicting protein homology with support vector machines with L. Arvestad, Stockholm Bioinformatics Center, KTH, Stockholm
- Integer linear programming for selecting DNA-microarray probes with K. Reinert, G. Klau, Institut für Informatik, Freie Universität Berlin
- Spectral methods in clustering with P. Deuflhard, M. Weber, Zuse Institut Berlin
- Analysis of gene expression time-courses with A. Schönhuth, CUBIC, Universität zu Köln
- Design of DNA-Miocrarrays with S. Rahmann, Technische Fakultät, Universität Bielefeld
- Meta-learning for analysis of gene expression time-courses with T.B. Ludemir, F. de Assis, T. de Carvalho, M.C. Pereira de Souto, R. Bastos, C. Prudencio, Dept. of Computer Science, Federal University of Pernambuco and Dept. of Informatics and Applied Mathematics, Federal University of Rio Grande do Norte
- Analysis of in-situ gene expression during mouse development with R. Spörle, Dept. of Developmental Genetics, MPIMG
- Analysis og gene expression during Hematopoiesis with F. Melchers, Lymphocyte Development Group, MPI for Infection Biology

Computational Diagnostics

Head
Dr. Rainer Spang
Phone: +49 (0)30-8413 1175
Fax: +49 (0)30-8413 1176
Email: spang@molgen.mpg.de

Scientists
Dr. Claudio Lottaz
Dr. Juby Jacob (since 12/2005)

Graduate Students
Florian Markowetz
Stefanie Scheid
Jochen Jäger
Dennis Kostka
Stefan Bentink
Xinan Yang

Scientific Overview

The focus of this group is to develop novel computational methodology for the analysis of molecular disease mechanisms in cancer. Modern array technologies produce high dimensional readouts of molecular activities in patient specimens. Characteristic patterns in this data shed light on previously unclear cancer classifications and can predict clinical outcome or treatment response. We characterize disease related molecular disorders from array data including microarrays, RNAi screening and cytometric readouts. Our research aims to improve molecular diagnosis and to set the stage for personalized medicine. The work of the computational diagnostics group includes both theoretical projects in which we aim to develop novel analysis methodology and applied data analysis projects with clinical cooperation partners.

Reconstructing non-transcriptional signaling pathways

We are currently pioneering an algorithm to infer non-transcriptional pathway features based on differential gene expression in silencing assays. In this approach, we distinguish two kinds of genes: the candidate pathway genes, which are silenced by RNAi, and the genes, which show effects of such interventions in expression profiles. We call the first *S-genes* (S for "silenced" or "signaling") and the second *E-genes* (E for "effects"). Because large parts of signaling pathways are nontranscriptional, there will be little or no overlap between S-genes and E-genes. Elucidating relationships between S-genes is the focus of our analysis; the E-genes are only needed as reporters for signal flow in the pathway. E-genes can be considered as transcriptional phenotypes. S-genes have to be chosen depending on the specific question and pathway of interest. E-genes are identified by comparing measurements of the stimulated and non-stimulated pathway; genes with a high expression change are taken as E-genes. Our approach models how interventions interrupt the information flow through the pathway. Thus, S-genes are silenced while the pathway is stimulated to see which E-genes are still reached by the signal. In our publication (Markowetz F. and Spang R. Bioinformatics 21, 4026-32, 2005) we developed a scoring function, which measures how well hypotheses about pathway topology are supported by experimental data. Input to our algorithm is a list of hypotheses about the candidate pathway genes. A hypothesis is characterized by (1) a directed graph with S-genes as nodes, and (2) the entry point(s) of signal into the pathway. This setting is summarized in Fig. 1. Our model is based on the expected response of an intervention given a candidate topology of S-genes and the position of the intervention in the topology. Pathways with different topology can show the same downstream response to interventions. We identify all pathways, which make the same predictions of intervention effects on downstream genes, by one so called *silencing scheme*. Sorting silencing schemes by our score shows how well candidate pathways agree with experimental data. Output of the algorithm is a strongly reduced list of candidate pathways. The algorithm is a filter, which helps to direct further research. Fig1 gives an example for a expression pattern, which clearly characterizes the pathway topology shown to its right.

Compensating for confounding variables in the analysis of microarray data

A major obstacle in the analysis of microarray data are hidden confounding variables like the genetic background of patients or undiscovered experimental artefacts, which leave traces in the expression data, contaminating score distributions, and hindering a meaningful significance analysis of differential gene expression. While the effects of know confounders can be compensated using established Methodology, so fare little was known on how to deal with unknown confounders. We have developed a computational method called *permutation filtering*, which exploits the information across genes to detect and compensate the effects of hidden confounders. Typically, significance analysis of microarrays is based on permutation scores, which are derived from shuffling sample labels. If by chance a single permutation correlates with a hidden confounder, the resulting score distribution has larger tails. This can be detected by transforming the random scores into p-values. By theory, these p-values should be uniformly distributed. Any deviation from uniformity indicates that a permutation resembles a hidden covariate thus introducing confounding signal into the permutation test. In applications, we commonly observe confounding permutations. A permutation might even introduce more signal than the actually covariate of interest. In MDS representation in the figure below, these permutations appear as outlying points. In Scheid and Spang (2006) we propose permutation filtering to address this problem. The underlying idea is that the p-value distributions across genes of random class labels carry information, whether a candidate permutation is correlated to an unknown confounding variable. If this is the case and there exist a sufficiently large number of genes that are differentially expressed with respect to the unknown confounder, a deviation from uniformity can be detected. The rationale of permutation filtering is simply to avoid such permutations. We could show that permutation filtering leads to a more accurate screening for differentially-expressed genes, and to more accurate estimates of false discovery rates.

The molecular Burkitt-Lymphoma

Burkitt's lymphoma (BL) and diffuse large-B-cell lymphoma (DLBCL) are subtypes of mature aggressive B-cell lymphoma. The diagnostic distinction of BL from DLBCL is important, since BL differs in therapy and outcome. The problem is that the two entities are hard to distinguish morphologically. As part of the Joint Research Project "Molecular Mechanisms in Malignant Lymphoma" (MMML), which is supported by the Deutsche Krebshilfe, we have investigate the potential of large scale expression profiling to sharpen the definition of Burkitt lymphomas. To this end, we have developed a novel computational method called core group extension. Starting from a small set of samples of known diagnosis it searches additional cases within a set of samples of unknown diagnosis and simultaneously delivers a diagnostic signature. We used this algorithm to define a novel lymphoma entity called molecular *Burkitt-Lymphoma* (mBL). A molecular signature was generated, which significantly extends the spectrum of Burkitt lymphomas. In total 44 mBL cases were identified. Of these, 11 had the morphologic appearance of diffuse large-B-cell lymphomas while only 29 displayed the morphology of a classic Burkitt lymphoma. These results will be the basis for the design of future clinical studies addressing treatment efficiency in aggressive lymphomas. Ongoing work on lymphomas includes the dissection of DLBCL and BL into further clinical subtypes by hierarchical network models of mixed large scale genomic data (array-CGH, FISH, immunophenotypes and mRNA-profiles)

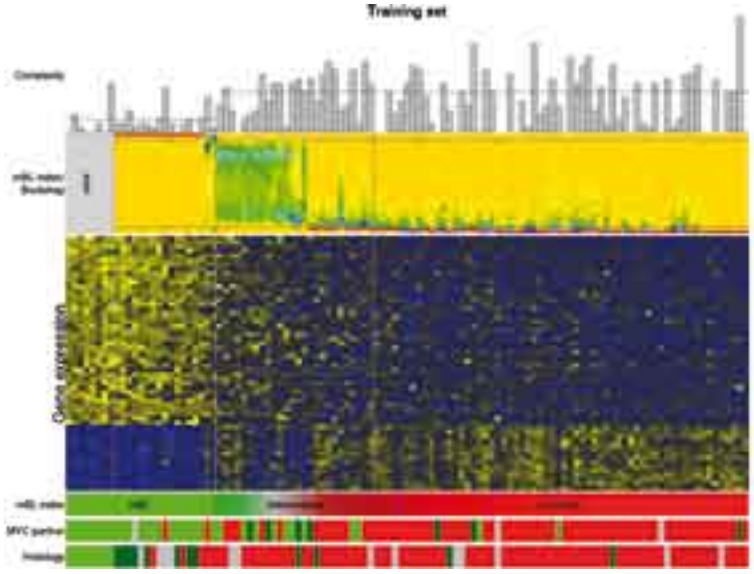


Figure1: The diagnostic signature of molecular Burkitt lymphomas. Shown are expression values of the signature genes as well as additional genetic and morphologicall patient characterisations.

Early relapse in childhood ALL

In childhood acute lymphoblastic leukemia (ALL), approximately 25% of children suffer from a relapse. In recurrent disease, despite intensified therapy, overall cure rates of 40% remain unsatisfactory and survival rates are particularly poor in certain subgroups. The probability of long-term survival after relapse is currently predicted from prognostic factors like the time and site of relapse. We have engaged in a collaboration with several experimental and clinical partners in the Acute Leukemias Consortium of the NGFN2 to investigate the molecular causes of poor treatment outcome. Aiming at identifying molecular pathways we analyzed gene

expression profiling from 60 prospectively collected samples of first relapse patients. We performed differential gene expression analysis, in order to characterize high and low risk patients on a molecular level. In addition, we measured the fraction of proliferating cells by cell sorting and used least angle regression to predict this fraction from gene expression patterns. Very early relapse of ALL are characterized by a distinctive gene expression pattern. When comparing very early relapsed ALL to late relapses, we discovered 83 differential genes, the vast majority of which are up-regulated and many were late cell cycle genes with a function in mitosis.

Future perspectives

We expect that the near future will bring a multitude of platforms for highly parallel molecular intervention experiments. Genome wide RNAi screening in lower model organisms is already established. At present, little is known about how to analyze this data. We plan to use data from these technologies to understand alterations of signaling pathways in tumors. We plan to transform expression patterns, which are associated to specified signaling defects into predictive expression signatures allowing for a diagnosis of pathway alterations in patient specific expression profiles. The resulting signatures will be applied to clinical expression studies in various cancer entities, leading to a novel tumor classification according to aberrations in signalling pathways. Particularly, we intend to develop predictive parameters for the susceptibility toward therapeutic agents that target specific signalling molecules.

Selected Information

Publications

- Scheid S., Spang R. (2006). Permutation filtering: A novel concept for significance analysis of large-scale genomic data. In: Apostolico A, Guerra C, Istrail S, Pevzner P, and Waterman M (Eds.): Research in Computational Molecular Biology: 10th Annual International Conference, Proceedings of RECOMB 2006, Venice, Italy, April 2-5, 2006, Lecture Notes in Computer Science vol. 3909, Springer, Heidelberg, pp. 338-347
- Hummel, M., Bentink, S., Berger,H., Klapper, W., Wessendorf,S., Barth, .F.E., Bernd,H.-W., Cogliatti, S.B., Dierlamm, J., Feller,A.C., Hansmann, M.L., Haralambieva, E., Harder, L., Hasenclever, D., Kühn, M., Lenze, D., Lichter, P., Martin-Subero, J.I., Möller, P., Müller-Hermelink, H.K., Ott, G., Parwaresch, R.M., Pott, C., Rosenwald, A., Rosolowski, M., Schwaenen, C., Stürzenhofecker, B., Szczepanowski, M., Trautmann, H., Wacker, H.-H., Spang, R., Loeffler, M., Trümper, L., Stein, H., Siebert, R. (2006) A Biologic Definition of Burkitt's Lymphoma from Transcriptional and Genomic Profiling, New England Journal of Medicine, 354:2419-2430.
- Markowetz F., Bloch J., Spang R. (2005). Non-transcriptional pathway features reconstructed from secondary effects of RNA interference. Bioinformatics 21, 4026-4032
- Lottaz C., Spang R. (2005). stam - a Bioconductor compliant R package for structured analysis of microarray data. BMC Bioinformatics 6(1):21
- Lottaz C., Spang R. (2005). Molecular Decomposition of Complex Clinical Phenotypes using Biologically Structured Analysis of Microarray Data. Bioinformatics 21(5): 1971-1978

Invited plenary lectures

- DIA EuroMeeting, Rome (3/03)
- ISI 2003, Berlin (8/03)
- Human Genome Meeting (HUGO), Berlin, (4/04)
- University of Freiburg, (5/04)
- ETH Zurich (5/04)
- Analytica 2004, Munich (5/04)
- Johns Hopkins University, Baltimore (10/04)
- University of Wuerzburg (2/05)
- Oxford University (6/05)
- ISCB, Szeged (8/05)
- 27. Deutscher Krebskongress, Berlin (3/06)

Diploma Theses

- Gene Ontology as a tool for the systematic analysis of large-scale gene-expression data Stefan Bentink, Technische Fachhochschule Berlin
- Multivariate Approaches to the Detection of Leukemic Cells in Flow Cytometry Readouts Jörn Tödling, Bioinformatics Department, FU Berlin

PhD Theses

- Florian Markowetz, Probabilistic Models for Gene Silencing Data, 2006
- Jochen Jäger, Deriving small diagnostic biomarker panels from genome wide, clinical microarray studies, 2006
- Stefanie Christina Scheid, Novel Concepts for the Significance Analysis of Microarray Data, 2006

Transcriptional Regulation

Head
Prof. Dr. Martin Vingron
Phone: +49 (0)30-8413 1150
Fax: +49 (0)30-8413 1152
Email: vingron@molgen.mpg.de

Scientists
Dr. Thomas Manke
Dr. Stefan Röpcke
Dr. Christine Steinhoff
Dr. Ho-Ryun Chung
Dr. Tomasz Zemojtel
Dr. Szymon Kielbasa
Dr. Lloyd Demetrius
Dr. Steffen Grossmann

Graduate Students
Sven Rahmann (-2004)
Christoph Dieterich (-2005)
Utz Pape (IMPRS)
Holger Klein
Abha Singh Bais

Visitors
Dr. Ricardo Bringas
Dr. Homayoun Bagheri
Dr. Jerzy Dyzkowski

Scientific Overview

The theoretical study of transcriptional regulation has entered a new era with the availability of many fully sequenced genomes in conjunction with a number of new experimental techniques for the study of protein-DNA binding. The gene regulation group focuses on the delineation of regulatory motifs and interactions based on the integration of this variety of information sources. In yeast, this involves not only theoretical considerations but also the analysis of the chromatin immuno precipitation data, which provide large scale information about transcription factor binding. In mammalia, comparison of non-coding, upstream sequences of orthologous genes can pinpoint regions that are likely to play a regulatory role. This can be extended by comparing evolutionarily conserved sequences to binding site descriptors that have been collected in publicly available or commercial databases. Microarray generated gene expression data may further serve to understand regulatory interactions between genes. The group engages both in theoretical studies and in collaborative efforts studying particular genes and systems.

Modularity in Regulatory Networks

The focus of this project is on the transcriptional control of gene expression through regulatory networks in yeast. For this model organism much data has been accumulated to support individual interactions between transcription factors and specific DNA regions to which they can bind. However, the actual response mechanisms of the cell to specific environmental signals are still far from being understood. One challenging aspect in the interpretation of regulatory networks is their dynamical organisation into "transcriptional modules" that involve the synergistic action of groups of transcription factors which regulate groups of genes.
In this effort we use a range of different computational methods to identify such coherent groups of regulators and their respective targets. We apply data mining techniques and graph-theoretical methods to define synergistic pairs and associations of transcription factors based on genome-wide information on regulatory networks. These methods have first been applied to yeast, where protein-protein interaction data often provide additional support for our predicted associations between transcription factors. The identification of synergistic transcription factors also provides a starting point for further experimental work. In the context of regulation of replication, our work has provided concrete suggestions for possible target regions of both Orc1 and Sum1, many of which have been confirmed experimentally in a collaboration with the group of Ann Ehrenhofer-Murray (formerly MPIMG, now University of Essen).
To evaluate our findings we often rely on curated information from the Gene Ontology database, which assigns one or more biological processes to a given gene. A significant over-representation of certain functional categories (e.g., "DNA replication") in a putative transcriptional module can be taken as an indication that the identified module is utilized in those biological processes. However, the traditional enrichment analysis is not well-suited for testing many overlapping categories. To this end we continue to develop statistical tests for structured background models (see also the methodological work of Steffen Grossmann).

More recently we have complemented this functional analysis through integration of gene expression data. In yeast, we draw upon a large compendium of expression data from more than 1000 different experimental condition. This approach allows us to resolve some of the dynamical organisation of transcriptional modules which cannot be inferred from the static representation of interaction networks alone. Specifically, we identified groups of condition-invariant as will as condition-specific transcriptional modules. Many of the derived modules share certain components which lend themselves to a more detailed investigation of transcriptional control mechanisms. In the above work we have performed a comprehensive analysis of such individual associations to determine those regulator-target pairs which show a high level of co-expression.

Enrichment analysis

Data analysis in the area of expression and regulation frequently follows a certain scheme: First, groups of genes are delineated based on either clustering or biclustering. Clustering means grouping together genes that behave similarly across conditions or time-points. Biclustering groups different descriptors, e.g., genes and conditions, in such a way that a "module" comprises a subgroup of genes which behaves similarly in a subgroup of conditions. Either of the two methods results in groups of genes which then need to be further interpreted. This is usually done through the inclusion of additional information, typically the gene function annotation provided be the Gene Ontology (GO). In general, we call this step enrichement analysis, because we search for attributes that are particularly enriched among the genes of interest. For GO categories this boils down to the question of identifying a functional category which is statistically significantly enriched among the genes in a given group. This is traditionally studied with a statistic based on quantifying the overlap between the given group and all categories in GO. To remedy the shortcomings of this approach (in particular, strong dependencies between the categories) we have developed an alternative statistic. This is a collaboration with P.N. Robinson from the group of Stefan Mundlos at Charité.

CORG, PROMOTION, and T-Reg databases

Over the last years, considerable effort of group members has gone into the development of a database and user interface to study transcriptional regulation and to allow for integration of theoretical analysis and laboratory work. When it started, this was the CORG (Comparative Regulatory Genomics) database, originally developed by Christoph Dieterich. In 2004, under the umbrella of the National Genome Research Network (NGFN, funded by BMBF) several groups at the MPI started to collaborate on questions of transcriptional regulation. This lead us to develop CORG into a common platform for promoter analysis. In this context we adopted the GenomeBrowser environment as our front end and use it to serve gene structures, transcriptional start sites, comparative information, transcription factor motif annotation, as well as the BACs, on which a promoters can be found and suggestions for primers to use in order to amplify a certain promoter region. This information is now used in ongoing experiments in the departments of Hans Lehrach and Bernhard Herrmann.
We also developed the T-Reg database to obtain representations of transcription factors, their DNA binding specificities of transcription factors and regulated genes. In addition, we developed T-Reg Comparator which delivers a detailed report on similarities between user-supplied sequence motifs and known motifs that specify the typical binding site of a transcription factor. Apart from simple one-to-one relationships, T-Reg Comparator is also able to detect similarities between sub matrices. Typical areas of application for T-Reg Comparator are motif identification or regulatory module finding and annotation of regulatory genomic regions.

Networks

With large scale data on protein-protein interactions or transcription factor-target gene relationships, a global view of biological networks has become possible. Lloyd Demetrius, a senior researcher in mathematical biology has joined the department and cooperates with Thomas Manke to apply Demetrius' theory of entropy on graphs. This has led to a new generative model for scale-free graphs, which is based on an evolutionary fitness function, and to an alternative model for crucial nodes in a graph based not simply on the degree but rather on the entropy contribution of a node. In another line of work, Martin Vingron has cooperated with Alexander Mikhailov from Fritz Haber Institute on the study of a toy models for the evolution of signal transduction networks. Lloyd Demtrius has also developed mathematical models for aging.

The proximal promoter of human ribosomal protein genes (Roepcke, with D. Walther)

Ribosomal protein genes as part of an essential apparatus possess many interesting properties: They are highly and co-ordinately expressed and highly conserved. This set allows us gain insight into some fundamental features of eukaryotic promoters. To date promoter recognition and transcription initiation is only partially understood, which is reflected in the poor performance of promoter prediction algorithms. This project is joint work with Peter Arndt and Diego Walther from the department of Hilger Ropers. We started with the human cytosolic RP genes [Roepcke *et al.*, Gene] and extended our analyses to zebrafish, yeast (with Jing Zhang) and to RP genes of the human mitochondrion (Marcel Schulz, an IMPRS student). Currently, we are extending our project by investigating two over-represented motifs experimentally in the lab of Diego Walther.

Epigenetics and other projects

Several smaller projects look at transcriptional regulation from a variety of standpoints. Christine Steinhoff works on the role of methylation in gene regulation. Ho-Ryun Chung studies nucleosomal positioning and competition of transcription factors among each other and with nucleosomes. Tomasz Zemojtel studies the role of LINE repeats in gene regulation. Holger Klein develops statistical methods to detect pairs of transcription factors that tend to co-occur on the DNA sequence. Abha Singh Bais works on a program to integrate the alignment of regulatory regions with the identification of binding motifs.

Collaborations

The department collaborates with a number of other groups. On the computational side, we closely interact with Ewan Birney with respect to transcription factor annotation. With him and also with Alvis Brazma we collaborate in the context of the EU Network of Exellence "BioSapiens". Various members of the department also maintain collaborations with experimental scientists. This comprises a collaboration with Alfred Nordheim from Tübingen on prediction of SRF target genes. Similar questions are dealt with in an project with Christine Sers from Charité (funded by DFG in an SFB). Stefan Röpcke collaborates with Diego Walther from the department of Hilger Ropers on the analysis of the ribosomal promoters. The long-standing collaboration on heart development with Silke Sperling (Dept. Lehrach) has been continued by Utz Pape from our department.

Selected Information

Publications

- Kowald, A., Demetrius, L. (2005). Directionality theory: a computational study of an entropic principle in evolutionary processes. Proc Biol Sci Royal Soc. 272:741-749
- Manke T, Dieterich C, Vingron M. (2005). Detecting Functional Modules of Transcription Factor Binding Sites in the Human Genome. Lecture Notes in Computer Science 3318, Springer.
- Dieterich, C., Grossmann, S., Tanzer, A., Ropcke, S., Arndt, P. F., Stadler, P. F., Vingron, M. (2005). Comparative promoter region analysis powered by CORG. BMC Genomics 6: 24
- Schmid, M., Davison, T. S., Henz, S. R., Pape, U. J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D., Lohmann, J. U. (2005). A gene expression map of Arabidopsis thaliana development. Nature Genetics 37: 501-506
- Irlbacher, H., Franke, J., Manke, T., Vingron, M., Ehrenhofer-Murray, A.E. (2005). Control of replication initiation and heterochromatin formation in Saccharomyces cerevisiae by a regulator of meiotic gene expression. Genes Dev. 19(15):1811-22
- Philippar, U., Schratt, G., Dieterich, C., Muller, J. M., Galgoczy, P., Engel, F. B., Keating, M. T., Gertler, F., Schule, R., Vingron, M., Nordheim, A. (2004). The SRF target gene Fhl2 antagonizes RhoA/MAL-dependent activation of SRF. Mol Cell 16:867-80
- Dieterich, C., Rahmann, S., Vingron, M. (2004). Functional inference from non-random distributions of conserved predicted transcription factor binding sites. Bioinformatics 20:I109-I115
- Manke, T., Bringas, R., Vingron, M. (2003). Correlating Protein-DNA and protein-protein interactions. J Mol Biol. 333(1):75-85

Selected Information

Invited Seminars, Lectures

- Computational Biology Basel Conference [BC]², 'Comparative Genomics', "Computational Methods in Regulatory Genomics", März 2006
- SIB – Swiss Institute of Bioinformatics, Leysin, Januar 2006
- MPI für Immunbiologie, Freiburg, Juni 2005
- INSERM, Workshop on Epigenomics: 'Large scale analysis of chromatin modifications and genome-transcription factor interactions', Marseille, "Functional Associations in Regulatory Networks", Mai 2005
- ETH Institut, Workshop on Statistics in Functional Genomics, Ascona, Juni 2004
- Linnaeus Centre for Bioinformatics, Uppsala, Mai 2004
- NIBR Pharma Novartis, Basel, "Computational Methods in Gene Expression an Gene Regulation", Mai 2004
- Plant Systems Biology, VIB – Ghent University, Februar 2004
- UC San Diego, "Linking Gene Expression and Regulation by Correspondence Analysis", August 2003
- Joint Statistics Meeting, San Francisco, "Bioinformatics for functional genomics: Transcriptional control and microarray data", August 2003
- Royal Statistical Society Topic Meeting "Genetics and Statistics", Belgium, "Gene Regulation and the interpretation of microarray data", August 2003

PhD Theses

- Sven Rahmann
  Algorithms for Probe Selection and Microarray Design, 2004
- Christoph Dieterich
  Comparative sequence analysis and association mining in gene regulation, 2005
- Hannes Luz
  Family Specific Rates of Protein Evolution, 2005

Cooperations

- Genome annotation with E. Birney, A. Brazma, EBI, Hinxton, UK
- SRF target prediction with A. Nordheim, Universität Tübingen
- Algorithm development with J. Tiuryn, University of Warsaw
- Signal transduction cascades with Christine Sers, Charité Berlin
- Evolution of biological networks with A. Mikhailov, Fritz Haber Institute, Berlin
- Microarray data analysis with D. Weigel, MPI für Entwicklungsbiologie, Tübingen
- Genome annotation with P. Stadler, Universität Leipzig

Protein Families and Evolution

Head
Dr. Eike Staub
(-Feb 2006)

Prof. Dr. Martin Vingron
Phone: +49 (0)30-8413 1150
Fax: +49 (0)30-8413 1152
Email: vingron@molgen.mpg.de

Scientist
Thomas Meinel (-2005)

Graduate Student
Hannes Luz

Visitor
Dr. Behshad Behzahdi

This group was originally headed by Antje Krause (until early 2004) and Eike Staub was supposed to take over the group. After his leave in 2006, Martin Vingron has supervised the ongoing projects, but it is not planned to maintain this as a separate group.

Protein families, phylogenetic profiles, and horizontal gene transfer

The SYSTERS project provides a partitioning of the whole protein sequence space by a fully automatic procedure. A refined two-step algorithm assigns each protein to a family and a superfamily. The results were made available via the SYSTERS web server *http://systers.molgen.mpg.de*). In addition to the Swiss-Prot and TREMBL databases, the sequence data underlying SYSTERS release 4 now comprise several protein sequence databases derived from completely sequenced genomes (ENSEMBL, TAIR, SGD and GeneDB). The SYSTERS web server provides access to 158 153 SYSTERS protein families each of which is represented by a consensus sequence. Submitting a query sequence and searching against the consensus sequences provides fast access to the SYSTERS cluster set. To augment the automatically derived results, information from external databases like Pfam and GeneOntology are added to the web server. Furthermore, users can retrieve pre-processed analyses of families like multiple alignments and phylogenetic trees. New query options comprise a batch retrieval tool for functional inference about families based on automatic keyword extraction from sequence annotations.
The taxonomic system as maintained by the NCBI was integrated in SYSTERS. The user now can assess the taxonomical complexity of protein family and select protein families of a specific taxonomic internal taxon, for example, all protein families that exclusively occur in mammals. A special taxonomic view of a protein family focuses on the presence/absence patterns of member proteins across organisms. Such presence/absence patterns are called phylogenetic profiles. Orthology-based phylogenetic profiles can be valuable indicators of horizontal gene transfer (HGT) which is the basis of our study on the role of HGT in tRNA synthetases. By phylogenetic profiling and tree reconstruction we found that also leucyl-tRNA synthetases, the subfamily that was long thought to unaffected by HGT, has experienced at least one HGT event between bacteria and archaea.

Evolution of eukaryotes

The SYSTERS project provided a comprehensive overview about the gene content of an organism. This data is used by us and others to reconstruct the evolution of organisms from all domains of life. We focus on the influence of endosymbiotic events for phylogeny reconstruction based on gene content. Endosymbiotic events lead to massive transfer of symbiont genes into the host genome. The evolution of eukaryotes has at least seen two endosymbiotic events leading to plastidial and mitochondrial genes in the nuclear genome of eukaroytes. It is estimated that large fractions of the eukaryotic nuclear genes actually stem from endosymbionts. Therefore, one can not assume that gene content data is tree-like, but rather has to be modelled by a network. However, we found that reconstruction of organismal phylogeny by gene content is astonishingly robust against influences from endosymbiotic gene transfer.
Another line of our research is the evolution of the eukaryotic nucleolar proteome from prokaryotic building blocks. The nucleolus is a dense nuclear compartmnt which houses the ribosome factory of the eukaryotic cell. Consequently, a large fraction of nucleolar proteins is of archaeal origin. Surprisingly, we discovered that a substantial fraction of nucleolar proteins stems from bacteria. Is seems as if these proteins were integrated into the nucleolar machinery subsequent to the integration of mitochondrial genes into the proto-eukaroyte archaeal-type genome. We also found that a majority of nucleolar proteins does not have prokaryotic counterparts, highlighting that the evolution of the nucleolar compartment occured during an extended time span in early eukaryote evolution and is probably still ongoing. Very recently we compiled a compre-

hensive inventory of the yeast nucleolar proteome using high throughput data on from mass spectrometry studies and protein-protein interaction data that serves us as a data source for further evolutionary studies. Our interest in nucleolar biology overlaps with topics from the ribosome groups in our institute which led to a fruitful collaboration on the evolution and function of small GTPases involved in translation, in particular the LepA protein.

Selected Information

Publications

- Dohm, J., Vingron, M., Staub, E. (2006): Horizontal gene transfer in aminoacyl-tRNA synthetases including leucine-specific subtypes. J Mol Evol. In press.
- Luz, H. and Vingron, M. (2006). "Family specific rates of protein evolution." Bioinformatics 22(10): 1166-71.
- Krause, A., Stoye, J., Vingron, M. (2005). "Large scale hierarchical clustering of protein sequences." BMC Bioinformatics 6:15.
- Staub, E., Fiziev, P., Rosenthal, A., Hinzmann, B. (2004). Insights into the evolution of the nucleolus by an analysis of its protein domain repertoire. Bioessays 26: 567-81.

Research Group Development & Disease

Head
Prof. Dr. Stefan Mundlos
Phone: +49 (0)30-8413 1263/ 1449
Fax: +49 (0)30-8413 1385
Email: mundlos@molgen.mpg.de

Scientists
Dr. rer. nat. Sigmar Stricker
Dr. rer. nat. Petra Seemann
Dr. rer. nat. Uwe Kornak
Dr. med. Katrin Süring
Dr. rer. nat. Mateusz Kolanzcyk
Dr. med. Pablo Villavicencio-Lorini (2/2005-)
Dr. rer. nat. Volkhard Seitz
Dr. med. Georg Schwabe (-6/2004)

Technicians
Norbert Brieske
Britta Hofmann (-8/2005)
Kathrin Seidel (8/2005-)
Asita Stiege
Christine Zwingmann

Graduate Students
Andrea Albrecht (-12/2005)
Sebastian Bauer (5/2005-)
Paulina Chudobska (-2/2005)
Michael Niedermaier (-6/2005)
Nicole Verhey van Wijk (-6/2006)
Jochen Hecht (1/2003-)
Barbara Dlugaszewska (-9/2005)
Daniel Birker (11/2005-)
Anja Klutzny (10/2005-)
Nadine Kossler (3/2005-)
Pia Kuss(11/2004-)
Andreas Ratzka (10/2004-)
Ulrich Wilkening (6/2004-)
Haikuo Zhang (8/2004-)
Wibke Schwarzer (5/2006-)

Diploma Students
Cindy Ast
Anne Baude
Jan Börgermann (-1/2005)
Julia Haupt
Manuela Magarin (-8/2003)
Maren Schmidt von Kegler (-6/2005)
Maren Urban
Ulrike Wiecha (-6/2003)

Research Group Development & Disease

Introduction

The research group Developmental & Disease is interested in normal and pathological mechanisms of vertebrate development and their relationship to disease. The group is part of and works in close collaboration with the Institute for Medical Genetics (IMG), which is located at the Campus Virchow of the Charité, Universitätsmedizin, Berlin. The IMG provides clinical and diagnostic service for the Charité and the Berlin/Brandenburg area. Research at the Institute covers a spectrum of clinical and molecular analysis of genetic syndromes with a focus on conditions with skeletal involvement. The Institute harnesses and integrates clinical medicine and molecular techniques to define disease-causing gene mutations and genetic variations that increase disease risk. The research group at the MPIMG develops and analyzes genetically engineered *in vitro* and *in vivo* models to elucidate pathogenetic mechanisms for human disease. Thus, our approach synergistically combines research on human genetic disorders with gene function analysis *in vitro* and in animal models. The combination of the basic science oriented research at the MPIMG with the more clinically oriented work at the IMG provides a unique opportunity for the translation of basic biology into clinical application and vice versa.
The genetic basis by which form and structure are regulated during vertebrate development is a fundamental question in modern biology and medicine. Dysregulations of this process manifest in humans as congenital malformations, one of the leading causes for childhood morbidity in Western countries. In the vast majority of cases, the cause for such aberrant development is still unknown, but, propagated by the Human Genome Project, major advances have been made over the past years. A large number of genetically determined conditions have been characterized and the disease causing mutation(s) identified. Functional studies using *in vitro* and *in vivo* approaches are beginning to unravel the molecular pathology of certain mutations as well as the natural function of the mutated genes. However, in spite of our increasing knowledge about genetic diseases and the causative genes involved, we are frequently unable to predict the outcome, i.e. the phenotype or the course of the condition. Phenotypic variability, reduced penetrance and genetic heterogeneity are generally accepted concepts to explain this phenomenon, but their molecular basis is poorly understood. For a better understanding we need to know the function of the genes involved as well as the interactive network in which they work. Research on the composition of regulatory networks that control certain developmental processes

will help us to understand the molecular basis of distinct disease groups and, at the same time, offers an opportunity to understand variability. Progress in this direction will have an impact not only on our understanding of biological processes, but will also be important for treatment decisions and counseling of patients.
Our research focuses on normal and abnormal mechanisms that influence development, growth, and regeneration of the skeleton. Patterning genes such as the Hox, Hedgehog, and BMP genes control the overall bauplan of the skeleton and instruct mesenchymal cells where and how to differentiate into the skeletal anlagen. Defects during this stage of development result in an abnormal number or shape of skeletal elements. Differentiation of mesenchymal progenitor cells into cartilage producing chondrocytes and/or bone producing osteoblasts is controlled by numerous factors. Searching for key players in this process, we identified Runx2 as an essential factor for osteoblast differentiation. Chondrocyte differentiation follows a genetically determined set of events beginning with the formation of a cartilaginous anlage and the differentiation of cells within this anlage to form a growth plate. Factors that control proliferation and differentiation in the growth plate such as the FGFs and their receptors and extracellular matrix proteins such as the proteoglycans or the collagens are essential for proper growth of the skeleton before and after birth. The individual skeletal elements are connected by joints, cartilage-covered areas at the ends of the bones that permitting an almost friction-free movement. By studying human diseases and various mouse models, we have identified BMPs, GDFs, and their inhibitors as major components of joint development. Bone homeostasis, i.e. the balance between osteoblast derived production of new bone and osteoclast driven resorption of old bone, is controlled by an elaborate network of molecules that are intensely studied due to their association with one of the most common age-related human diseases, osteoporosis.
After injury bone is able to efficiently regenerate through the recruitment and differentiation of stem cells. Newly generated bone tissue undergoes extensive remodeling ultimately producing a tissue that is equivalent in structure and mechanical properties to the adjacent old bone. The biology of this remarkable process is only slowly unraveling but developmental pathways, as they are used during organogenesis of bone and cartilage, have been shown to play a role. Thus, the study of development mechanisms are likely to be helpful for our understanding of regeneration. This knowledge can then be translated into the guided regeneration of skeletal tissues for example by the administration of growth factors or the guided manipulation of certain pathways.

Results and Significance

To develop a more comprehensive understanding of the molecular mechanisms that govern skeletal development and regeneration we use three major approaches. First, we concentrate on human diseases with an abnormal phenotypes in digit and joint development and use a classic human genetics approach to identify novel disease related genes. Second, we have established several *in vitro* and vivo systems to evaluate the function of selected genes and third, we use a set of screening methodologies to identify new major players in the field.

Strategies for disease gene identification

Genetic conditions with skeletal involvement comprise a large group of conditions of which many have been characterized. However, a large number of conditions remain in which the disease causing mechanisms are unknown. The most important factor limiting a large scale approach to comprehensively cover the majority of all conditions is in the availability of well characterized patients and families that are accessible for genetic analysis. For this purpose, the extensive exchange with the clinical genetics unit at the IMG (headed by D. Horn) and the Children's hospital at the Charité is essential and has proven to be very successful. In addition, we have established an intense collaboration with the Clinical Genetics Unit of the Sultanate of Oman. This collaboration provides us with a large number of clinical conditions that are continuously being investigated. Using this approach we have been successful in the characterization of new syndromes (for ex. SED Omani type), clinical delineation of known conditions (Escobar syndrome, wrinkly skin syndrome, geroderma osteodysplastica a.o.) and the molecular characterization of these conditions (mutations in CHST3 in SED Omani type).

In many cases, however, classical genetic approaches are not possible because of small family size. CGH-array technology, developed at the Dept. Ropers, was used by us to detect small rearrangements in the genome in sporadic cases without family history. Together with the Dept. Ropers we have successfully used this approach and have identified a number of disease causing deletions. For example, we have identified deletions in Nievergelt syndrome, a condition with severe limb malformations. Furthermore, we have solved the riddle of apparently non-mendelian inheritance in TAR syndrome.

Molecular pathology of skeletal disease

To investigate the molecular pathology of certain disease genes we have established a set of technologies that enable us to elucidate the molecular mechanisms governing cell differentiation along the chondrocytic and osteoblastic lineage. We analyze gene function by retroviral mediated overexpression in mesenchymal stem cells derived from chicken limb buds or bone and use a range of biological parameters as the read out. This system is particularly useful because of high infections rates and the possibility to use the virus for *in vivo* overexpression studies in chick embryos.

To investigate gene function *in vivo* we take advantage of the new mouse house. The presence of a now working facility has enabled us to pursue a number of new projects that were not possible before. One major project has been the elucidation of the role of neurofibromin 1 (Nf1) in skeletal development. Nf1 is mutated in neurofibromatosis type 1 (NF1), a prevalent genetic conditions with a high incidence of malignant and begin tumors mainly derived from neural tissues. In addition to the increased risk for tumor formation, many individuals with NF1 have skeletal diseases such as short stature, osteoporosis, bending of the spine (scoliosis), penciling of ribs, and, as the most severe complication, bending, fracture and formation of a "false joint" (pseudarthrosis) of the tibia. To investigate the role of Nf1 in skeletal development, we selectively inactivated the gene in mesenchymal progenitor cells in the limb. Similar to NF1 affected individuals, Nf1-/- mice show a reduced bone density, diminished growth, fusions of joints, and, as a more specific sign, bowing of the tibia (Fig.1). These multiple effects on the skeleton are due to a differentiation defect in osteoblasts, reduced proliferation rates and abnormal differentiation of chondrocytes, and abnormal vascularization of skeletal tissues. The latter effect is likely to be induced by activated Ras-signaling, as shown by retrovirus mediated overexpression of mutated Ras. Importantly, many aspects of the phenotype resemble changes associated with increased FGF signaling thus linking FGF and Nf1 pathways in skeletal development. Using this approach we have thus been able to identify a completely new role of Nf1 in skeletal development.

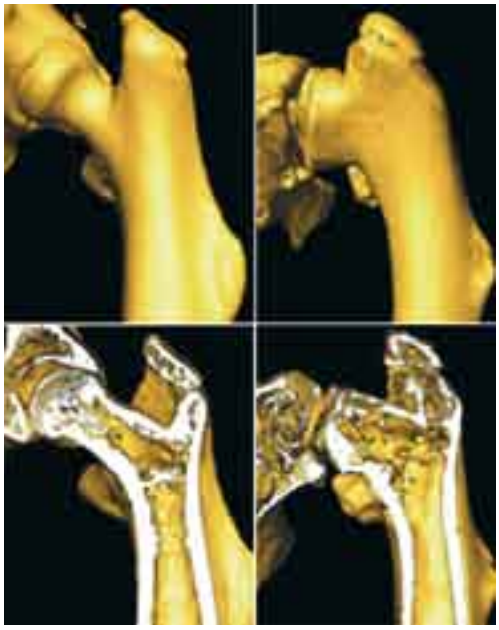


Figure 1: Abnormal bone and joint formation in mice with inactivate neurofibromin 1. µCT of normal (left) and Nf1-/- (right) proximal femur with cross section shown below. Seletive inactivation of Nf1 in limb mesenchymal cells results in abnormal formation of the femoral head and the hip joint.

Osteoporosis is one of the most frequent age-related diseases caused by a mix of genetic and environmental factors. To learn more about the genetic mechanisms that control the regulation of bone density we have collected families with a low bone density phenotype and have genetically characterized them. Two conditions are of particular interest in this respect: wrinkly skin syndrome (WSS) and geroderma osteodysplastica (GO). Both show diminished bone formation, but only GO is characterized by a progressive age-related bone loss similar to the naturally occurring condition, but accelerated. We have identified the gene responsible for GO and are currently investigating it by *in vitro* and transgenic mouse models.

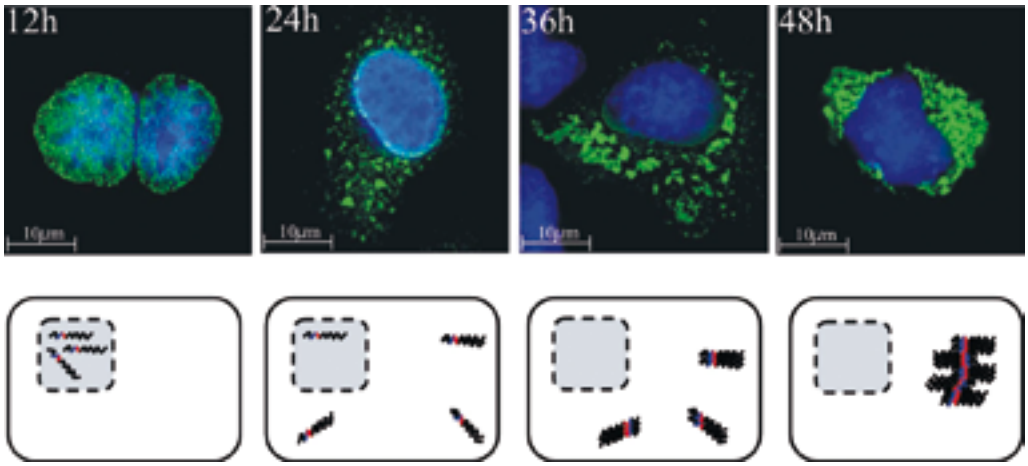


Figure 2: Dynamics of aggregate formation associated with alanine repeat expansions. In cells overexpressing Hoxd13 with expanded alanine repeats (+14) aggregate formation takes place over a time period of 24–48 hours. 12h after transfection the mutant protein (green) localizes to the nucleus (blue). After 24h the protein starts to form small aggregates in the periphery of the cell, and is less abundant in the nucleus. The aggregates become bigger with time and finally form large clumps after 48h. The lower panel gives a schematic of aggregate formation over time.

Expansions of trinucleotide repeats are a relatively common disease mechanism primarily occurring in late onset neurological conditions. We have identified a novel mutational mechanism by expansions of trinucleotide repeats that code for alanines (Fig.2). In contrast to previously known repeats, alanine expansions are meiotically and mitotically stable, occur only in transcription factors, and results in developmental defects. Using alanine expansions in HOXD13 that cause the human hand malformation synpolydactyly, we have identified a common molecular pathology that unifies these conditions. We were able to show that alanine expansions beyond a certain threshold invariably result in protein aggregation and degradation and thus a loss of protein function . Variables that control protein degradation are currently being investigated.

Brachydactylies comprise a Molecular Disease Family

One major approach in understanding genetic variability is to unravel the regulatory networks that control gene function and thus the interaction of one gene product with its partners. In principle, this can be approached from two angles: the identification and analysis of all genes that interact with a central molecule of the pathway, and/or the identification of all related phenotypes and their genetic characterization. We have investigated the disease group of brachydactylies (small hands/feet due to shortening of digits) as a model system to investigate the relationship between genotype and phenotype in a well studied pathway. Based on our results we have promoted the term "molecular disease family" to describe the close relation between overlapping phenotypes that originated from defects in a pathway or in closely interacting pathways. The brachydactlies have been clinically defined in the groups A, with subgroups A1-A3, B, C, D, and symphalangism (condition with fusion of joints). Over the last years we were able to identify the genetic cause for several of these conditions and have established a tight interactive network between its individuals components. Key players within this network appear to be the BMPs and GDF5 together with their corresponding receptors BMPR1A and BMPR1B, as well as the inhibitor Noggin. We have identified a number of activating and deactivating mutations that disturb the balance between the signals by altering ligand binding specificity either to the receptor or the inhibitor, changing receptor activity, or by changing inhibitor activity. In addition, we have linked the Hedgehog pathway to the brachydactyly phenotype by unraveling the molecular cause of the short digit (Dsh) mutation in the mouse. Further work has shown that the receptor tyrosine kinase Ror2, which is mutated in brachydactyly type B, modulates the Gdf5 pathway. A finding which is substantiated by our recent results that mutations in NOGGIN can cause a B-like phenotype. All brachydactyly-associated genes are part of a regulatory pathway controlling the development of digits and joints. To investigate the developmental pathology of these conditions *in vivo* we have established mouse models for brachydactylies types A1, B, and C. The pathology of individuals mutations is tested in *in vitro* systems using chicken limb bud micromass cultures.

Mechanisms of muscle-bone interaction

Bone is profoundly sensitive to mechanical forces. Without mechanical stimuli bone is degraded loosing much of its stability. We are interested in the mechanisms that induce and control the translation of mechanical force into a molecular signal. For this purpose, we have established a cell strain system which allows a precise mechanical stimulation of cells. We are establishing a time curve of responsive genes using whole genome microarray technology. The long term goal of these studies is to identify the molecular pathways that translate mechanical stimuli into bone inducing signals.

Mechanical force is transmitted to the bone through muscles and tendons. Muscle disease is frequently associated with congenital bone and joint pathology indicating that prenatal movement is important for skeletal development. This is particularly evident by the frequent association of congenital muscle disease with abnormal joints. Taking advantage of a large Omani family, we have identified the molecular cause for Escobar syndrome, a condition with muscle webs and stiff joints. We found mutations in the γ-subunit of the acetylcholine receptor. This subunit was shown to be expressed only during muscle development and is replaced with the δ-subunit by birth, thus explaining the developmental phenotype and the non-progressive character of the condition. We are investigating a number of related conditions with abnormal muscle development and joint phenotypes to investigate this association further.

To learn more about the mechanisms of muscle development, we are currently analyzing the role of the transcription factor Odd-skipped related 1 (Osr1) in the musculoskeletal development of the chicken limb. Overexpression of Osr1 in chick embryos results in loss of musculature and an increase in connective tissue which ultimately results in the bending of bones. Osr1 shows exclusive expression in all types of connective tissue fibroblasts and also in joint interzone cells (Fig.3). We found that Osr1 is sufficient to repress the formation of mesenchymal tissues such as cartilage, tendon and muscle while it promotes the differentiation of connective tissue fibroblasts. Furthermore, we could show that Osr1 *in vitro* and *in vivo* suppresses muscle formation in a non cell-autonomous manner by modulating Wnt signalling in muscle connective tissue.

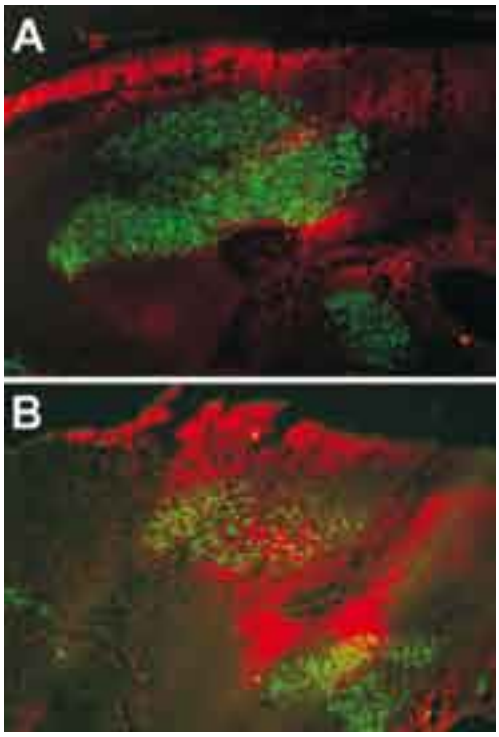


Figure 3: Expression of Osr1 in subcutaneous and muscle connective tissue of chicken embryonic limbs. A) Fluorescent double in-situ hybridisation for Osr1 (red) and the myogenic marker MyoD (green) showing expression of Osr1 in mesenchymal cells between the muscle precursors and beneath the ectoderm. B) Fluorescent in-situ hybridisation for Osr1 (red) followed by immunhistochemistry for myosin heavy chain (green) showing expression of Osr1 in cells surrounding differentiated myotubes.

Identification and Characterization of genes expressed in developing and healing bone

In order to reach a comprehensive understanding of bone and joint development we need to identify all genes that are relevant for this process. We have used several approaches to tackle this problem. RNA-profiling techniques have been used to isolate regulated genes that are differentially expressed in the E14.5 humerus of Runx2–/– and other mutant mice vs wt mouse embryos. To identify genes that are regulated during the process of chondrocyte condensation and joint formation we compared expression patterns of limbs between stages E11.5 and E14.5 in mouse embryos. Regulated genes are evaluated for their expression pattern in sections of E15.5 forelimbs using an automated in situ hybridization system (GenePaint). Analysis of these expression studies revealed multiple unique expression patterns that can easily be linked to certain cells types and differentiation steps, thus giving first ideas about gene function. In a collaboration with the Dept. Lehrach, this system is also used for a large scale in situ hybridization analysis of the genome in a EU-collaborative project. Genes that are expressed in the limbs are documented and downloaded to our BoneExpress database. Interesting candidates are being evaluated for their function and regulation. Using this system we have identified several candidate genes such as Osr1 and Col1a22, that are expressed in the developing joints and that are now investigated for their function.

In the event of injury, bones heal by generating new bone rather than by scar tissue. Recent studies have provided evidence that skeletal regeneration as it occurs in fracture repair is similar to embryonic bone development. In this project we intend to systematically evaluate and categorize genes that are expressed during the early phase of fracture repair. We have used the controlled fracture of the sheep tibia as a model and callus from several stages of fracture healing as a source of material. In collaboration with the Depts. Lehrach and Vingron we have established the first sheep EST library of bone and callus. The genes identified in this library will be an important source to identify new induce and control bone healing.

Future development and goals

Pivotal to our understanding of skeletal biology is the comprehensive identification and characterization of regulatory networks that control key events in this complex process. It is our goal to expand our knowledge in this field by the systematic identification of all genes that are relevant for the formation of the skeleton. For the first part we will focus on the events of early joint development, the condensation of chondrocytes, and the formation of digits. Based on the data collected during the last years we intend to perform large scale in situ hybridzation analysis in order to get an overview of gene expression patterns. This approach will be complemented by our continuous efforts to identify new disease-related genes of conditions with joint and/or cartilage defects. Part of our long term goal is to translate our increasing knowledge about the basic mechanisms of cell differentiation and organogenesis into clinical applications for the treatment of heritable bone diseases and for the development of novel strategies for bone and cartilage regeneration. To reach these goals we have started several initiatives to strengthen the field of bone biology in the Berlin area. This includes, for example, the successful application for the Berlin Center for Regenerative Therapies which will be located at the Charité Campus Virchow.
Future Developments in molecular genetic technologies are expected to have great impact on medical practice. It is our goal to translate new developments into the clinics thereby improving the diagnostics, counseling and treatment of heritable diseases. Whole genome analysis for small deletions by array CGH has already shown to greatly expand our diagnostic capabilities. Novel sequencing technologies as they soon will be available are likely to have a major impact as well. The close cooperation between the MPIMG and the Charité offers a good opportunity to accomplish this goal.

Selected Information

Publications

- Woods CG, Stricker S, Seemann P, Stern R, Cox J, Sherridan E, Roberts E, Springell K, Scott S, Karbani G, Sharif SM, Toomes C, Bond J, Kumar D, Al-Gazali L, Mundlos S. (2006) Mutations in WNT7A Cause a Range of Limb Malformations, Including Fuhrmann Syndrome and Al-Awadi/Raas-Rothschild/Schinzel Phocomelia Syndrome. Am J Hum Genet 79(2):402-8.
- Hoffmann K, Muller JS, Stricker S, Megarbane A, Rajab A, Lindner TH, CohenM, Chouery E, Adaimy L, Ghanem I, Delague V, Boltshauser E, Talim B, Horvath R, Robinson PN, Lochmuller H, Hubner C, Mundlos S. (2006) Escobar Syndrome Is a Prenatal Myasthenia Caused by Disruption of the Acetylcholine Receptor Fetal gamma Subunit. Am J Hum Genet 79(2):303-12.
- Hecht J, Kuhl H, Haas SA, Bauer S, Poustka AJ, Lienau J, Schell H, Stiege AC, Seitz V, Reinhardt R, Duda GN, Mundlos S, Robinson PN. (2006) Gene identification and analysis of transcripts differentially regulated in fracture healing by EST sequencing in the domestic sheep. BMC Genomics. 7(1):172
- Seemann P, Schwappacher R, Kjaer KW, Krakow D, Lehmann K, Dawson K, Stricker S, Pohl J, Ploger F, Staub E, Nickel J, Sebald W, Knaus P, Mundlos S. (2005) Activating and deactivating mutations in the receptor interaction site of GDF5 cause symphalangism or brachydactyly type A2. J Clin Invest 115(9):2373-81.
- Albrecht A, Mundlos S. (2005) The other trinucleotide repeat: polyalanine expansion disorders. Curr Opin Genet Dev 15(3):285-93.
- Niedermaier M, Schwabe GC, Fees S, Helmrich A, Brieske N, Seemann P, Hecht J, Seitz V, Stricker S, Leschik G, Schrock E, Selby PB, Mundlos S. (2005) An inversion involving the mouse Shh locus results in brachydactyly through dysregulation of Shh expression.J Clin Invest 115(4):900-9.
- Albrecht AN, Kornak U, Boddrich A, Suring K, Robinson PN, Stiege AC, Lurz R, Stricker S, Wanker EE, Mundlos S. (2004) A molecular pathogenesis for transcription factor associated poly-alanine tract expansions. Hum Mol Genet 13(20):2351-9.
- Thiele H, Sakano M, Kitagawa H, Sugahara K, Rajab A, Hohne W, Ritter H, Leschik G, Nurnberg P, Mundlos S. (2004) Loss of chondroitin 6-O-sulfotransferase-1 function results in severe human chondrodysplasia with progressive spinal involvement.Proc Natl Acad Sci USA 101(27):10155-60.
- Lehmann K, Seemann P, Stricker S, Sammar M, Meyer B, Suring K, Majewski F, Tinschert S, Grzeschik KH, Muller D, Knaus P, Nurnberg P, Mundlos S. (2003) Mutations in bone morphogenetic protein receptor 1B cause brachydactyly type A2. Proc Natl Acad Sci USA 100(21):12277-82. Epub 2003 Oct 1.
- Kornak U, Mundlos S. (2003) Genetic disorders of the skeleton: a developmental approach. Am J Hum Genet 73(3):447-74.

Selected Cooperations

- Prof. P. Fratzl (MPI für Kolloid- und Grenzflächenforschung, Potsdam). Study of bone material quality in development and disease.
- Prof. G. Duda (Muskuloskeletales Forschungszentrum, Charité, Berlin). Fracture healing.
- Prof. Sebald, (Institut für Physiologische Chemie, Universität Würzburg). BMP-signaling.
- Dr. A. Rajab (Genetic Unit, DGHA, Ministry of Health, Muscat, Sultanate of Oman). Genetic analysis of inbred populations.
- Dr. M. Warman (Dept. of Orthopedic Research, Harvard Medical School, Boston, USA). Genetics of skeletal dysplasias.
- Dr. D. Chan (Dept. of Biochemistry, Hong Kong, PRC). Role of Ihh in digit development.
- Prof. D. Rimoin (Cedar's Sinai Center, Los Angeles, USA). Heritable bone diseases.
- Dr. E. Zelzer (Weizman Institute, Israel). Role of angiogenesis in digit development.

PhD Theses 2005

- Michael Niedermaier
  Molekulargenetische Analyse der Mausmutante Short Digits (DSH).
- Andrea Nicole Albrecht
  Synpolydaktylie in Mensch und Maus.
- Barbara Dlugaszewska
  Molekulare Studien an HOXD-Genen.

Diploma Theses 2005

- Jan Börgermann
  Molekulare Mechanismen der Internalisierung von BMP-Rezeptoren.
- Mareen Schmidt von Kegler
  Charakterisierung der Heparinbindungsaktivität von GDF5. Ein Modell zur posttranslationalen Regulation von GDF5.
- Nadine Kossler
  Analyse potentieller Gene der Skelettogenese:
  a) durch Etablierung eines retroviralen siRNA Testsystems
  b) unter Verwendung biochemischer Methoden.

External Grants

- EFRE-Array CGH – Etablierung eines Zentrum für Microarray-gestützte CGH
  Europ. Kommission (EU), 1/03-12/05
- Klin. FG TP 9: Molecular mechanism of fracture healing
  DFG (German Research Foundation), 1/02-12/07
- SFB 577 TP A4: Craniometaphyseal dysplasia (CMD) – Clinical Variability and Pathogenic Pathways
  DFG (German Research Foundation), 7/01-6/08
- SFB 577 TP A6: HOXD-gene. Molecular Pathology and Embryology of HOXD-related Limb Malformations
  DFG (German Research Foundation), 7/01-6/08
- SFB 577 TP A8: Analysis of the receptor Tyrosine Kinase Ror2: A Paradigm for Phenotypic Variation in Allelic Genetic Disorders
  DFG (German Research Foundation), 7/04-6/08
- SFB 577 TP A9: Pelger Syndrom : Analysis of phenotypic and functional variances of lamin B receptor mutations
  DFG (German Research Foundation), 7/04-6/08
- SFB 577 TP Z 05: Central Facility for Animal Model Generation
  DFG (German Research Foundation), 7/04-6/08
- NF1-Neurofibromatosis as a Bone Dysplasia
  US-Army, 5/03-4/06
- ANABONOS, Molecular Mechanisms of Bone Formation and Anabolism
  European Union's FP6, 3/04 -2/07
- EURExpress, a European consortium for large-scale gene expression analysis
  European Union's FP6, 01/05 – 12/08
- Sonic hedgehog: SFB 665: Projekt A5 : Sonic hedgehog regulation during development and in disorders of the nervous system.
  DFG (German Research Foundation), 7/05-6/09
- GDF5-Signalweg: Klinische Korrelation und funktionelle Analyse von Mutationen im GDF5/BMPR1-Signalweg
  DFG (German Research Foundation), 7/05-6/07
- Charakterisierung von molekularen Pathomechanismen der Osteoporose bei Geroderma osteodysplastica
  Elsbeth Bonhoff Stiftung, 2/05-2/07

Ribosome Crystallography

Group Leader
Dr. Paola Fucini
Phone: +49 (0)30-8413 1691
Fax: +49 (0)30-8413 1747
Email: fucini@molgen.mpg.de

Scientists
Dr. Frank Schluenzen
Dr. Joerg Harms
Dr. Daniel Wilson

Technicians
Joerg Buerger
Renate Albrecht
Barbara Schmidt
Uwe Vogel

Diploma Students
Sabrina Schaefer
Christian Femmer

Scientific Overview

The main subject of the research conducted in our group is the ribosome, which is used as a model system to study molecular interactions. Molecular interactions are in fact fundamental to any type of communication that occurs inside and outside of the cell and they seem to represent the *modus operandi* of life – for at least how it is understood so far. Understanding the basis of these interactions is an essential step to fully comprehend how life expresses itself and how we can possibly intervene.
The aim of the specific projects that have been undertaken is to elucidate at the structural level how the various components of the translational apparatus interact with each other to guide and/or regulate the ribosome through the translational cycle. Moreover, investigations into the interaction between the translational apparatus and the primary product of its activity, namely the nascent polypeptide chain, have been initiated, with the aim to directly monitor the folding of *de novo* translated protein.
Despite the size of the ribosome, it is possible to study its structure by three different techniques that nicely complement each other namely x-ray crystallography, NMR and cryo-EM. The most powerful technique utilized in our group is x-ray crystallography, since it provides atomic details of the interaction between the ribosome and its ligands. Our laboratory is in the unique position of being able to prepare six distinct types of ribosomal crystals: the *Thermus thermophilus* 30S, 50S, and 70S particles and the 50S from three different organisms, *Deinococcus radiodurans*, *Bacillus stearothermophilus* and *Haloarcula marismortui*.

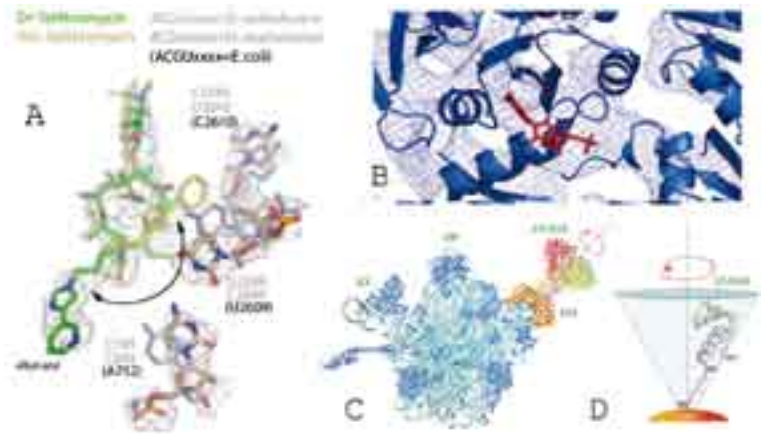


Figure 1: a) Comparison of the position and conformation of telithromycin bound to the D50S (green) and H50S (yellow) (reviewed in Wilson *et al.*, 2005). The electron density (meshed) for the D50S-telithromycin revealing the different orientation of the heterocycle sidechain of telithromycin (arrowed), as well as shifted positions of Dr2588 in domain IV and Dr765 in domain II of the 23S rRNA. The equivalent positions of these nucleotides (C2644 and C845, respectively) in the H50S (PDB1YIJ; Tu *et al.*, 2005b) are shown in grey.
b) Detail showing the high resolution obtained for a cryoEM reconstruction of a EFG-70S complex. The project was realized in a collaborative effort with the Group of Prof. C.M. Sphan. c and d) Representation of the orientation of L7/L12 on the ribosome as envisaged from the study conducted in collaboration with the group of Prof. Dobson (Christodoulou *et al.*, 2004).
c) The hinge and CTD regions of L7/L12 modeled into the structure of the 50S subunit with L11 (in orange indicating the base of the stalk), rRNA (cyan), and the remaining ribosomal proteins (dark blue). L7/L12 is drawn in various colors and positioned in six orientations representing a full rotation around the principal axis (gray line) as reported in (d).
d) The orientation of the CTD of an L7/L12 protein derived from rotational anisotropy analysis of relaxation data position L7/L12 within a cone in an extended conformation and with the hinge residues (43–51) modeled in a random coil conformation.

The ability to prepare these crystals forms the basis for a theoretically endless number of new functional complexes. Because of the (i) large size of the ribosome, (ii) the few contacts that the particle form within the crystal lattice and thus (iii) the large space available within the crystal, it is possible to soak preformed ribosomal crystals with small molecules that will diffuse into the array of packed ribosomes and find their correct binding position on the particles. This technique, initially used for the study of ribosome in complex with very small ligands as the antibiotics (range of 1 KDa), has been recently used for the preparation of ribosome in complexes with much larger ligands as the binding domain of translational factors (range of 100 KDa).
The analysis of the prepared crystals, previously done in collaboration with Prof. Yonath is now done completely independently in our group. In the last years, through the improvement of the crystallization conditions and handling of the crystals the resolution of the structure of the large ribosomal subunit of *Deinococcus radiodurans* has been improved from 3.4 Å to 2.8 Å (for a 90% completeness of the data), allowing the generation of a full atom model. This model is the first one available at relatively high resolution for the eubacteria large ribosomal subunit (the other available models being either at a lower resolution or from different organisms (archebacteria). The improved structure has been a key point for a more accurate investigation on the binding site of various antibiotic (Biol Chem. 2005, 386,1239-52), some of which showed in previous studies a discrepancy in their binding position on ribosome from *Deinococcus radiodurans* or *Haloarcula marismortui*.
A number of ribosomal crystals have also been investigated elucidating the mechanism of binding of various factors. Among them, the studies on the ribosome translational factors RRF (EMBO J. 2005, 24, 251-60) and TF (Structure 2005, 13, 1685-94) that elucidate the mechanism of binding and action of these factors respectively in the last phase of the translational cycle (ribosome recycling) and the co-translational folding acquisition of the polypeptide nascent chain. Both studies are the subject of a patent application since in both cases the factors are only present in the cytoplasm of prokaryotic cells and the details of their binding mode on the ribosome can be used for the development of new ribosome antibiotic.
In addition to the crystallography studies, relying on the expertise of Prof. R. Agrawal (USA) and Prof. C. Spahn (Germany), we have studied a number of different complexes by cryo-electron microscopy including the GTPase Era on the small ribosome subunit (Mol Cell. 2005, 18, 319-29) and the elongation factor G on the 70S ribosome (at the remarkable resolution of 7-8A).
Further collaborative studies are undergoing with different groups with the aim to investigate i) the composition and dynamic of the ribosomal stalk from different organisms by mass spectrometry (in collaboration with the group of Prof. C. Robinson, UK; Proc Natl Acad Sci USA. 2005, 102, 8192-7), ii) the folding acquisition of the newly synthesized polypeptide chain by nuclear magnetic resonance on the ribosome (in collaboration with the group of Prof. C. Dobson, UK; Proc Natl Acad Sci USA. 2004, 101, 10949-54; Mol Cell. 2004, 14, 57-66).

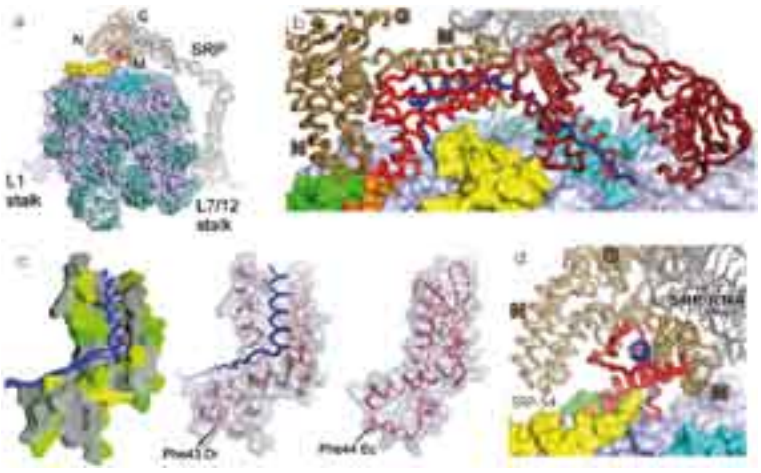


Figure 2: a) Overview of the D. radiodurans 50S in complex with TF (Schlünzen et al., 2005) revealing the possible simultaneous occupation of the TF (red) and SRP (Ffh in brown, RNA in grey). The N, G and M domains of SRP are indicated, as is an alpha-helical nascent chain (blue) located in the located in the hydrophobic cleft of the TF binding domain (TF-BD). b) Close-up of (a) showing the close proximity between the M domain of SRP54 (Ffh) and the alpha-helical nascent chain (blue) present in the hydrophobic crevice of the TF-BD (red). c) Surface representations revealing the hydrophobic cavity present in the TF-BD upon binding to the ribosome, with a putative alpha-helical nascent chain (blue) modelled into it. Left, the hydrophobic regions indicated on the opaque surface in grey, and polar regions are green/orange. Middle, same orientation but with transparent surface with the universally conserved Phe43 flipped into the hydrophobic loop region, as indicated. Right, similar representation of the TF-BD as the middle but for an unbound TF-BD. d) Alternative orientation of (b) showing a possible path for the nascent surrounded on one side by the open crevice on the TF and on the other side by the signal sequence binding groove, present in the SRP54 M domain.

Selected Information

Publications

- Schlünzen F., Takemoto C., Wilson D.N., Kaminishi T., Harms J., Hanawa-Suetsugu K., Szaflarski W., Kawazoe M., Nierhaus K.N., Yokoyama S., and Fucini P. (2006) The antibiotic kasugamycin mimics mRNA nucleotides to destabilize tRNA binding and inhibit canonical translation initiation. Nature Structural and Molecular Biology, in press.
- Wilson D.N., Harms J.M., Nierhaus K.H., Schluenzen F., Fucini P. (2005) Species-specific antibiotic-ribosome interactions: implications for drug development. Biol Chem. 386(12):1239-52.
- Schlünzen, F., Wilson D.N., Tian P., Harms J.M., McInnes S., Hansen H.A.S, Albrecht R., Buerger J., Wilbanks S. and Fucini P. (2005) The binding mode of the trigger factor on the ribosome: Implications for protein folding and SRP interaction. Structure 13(11):1685-94.
- Sharma M.R., Barat. C., Wilson D.N., Booth T.M, Kawazoe M., Hori-Takemoto C., Shirouzu M., Yokoyama S., Fucini P., and Agrawal RK. (2005) Interaction of Era with the 30S ribosomal subunit implications for 30S subunit assembly. Mol Cell. 18(3):319-29.
- Giavalisco, P., Wilson D., Lehrach H., Klose J., Gobom J., and P. Fucini. (2005) High heterogeneity within the proteins of the Arabidopsis thaliana 80S ribosome. Plant Mol Biol. 57(4):577-91.
- Wilson D.N., Schluenzen F., Harms J.M., Yoshida T., Ohkubo T., Albrecht R., Buerger J., Kobayashi Y., and Fucini P. (2005) X-ray crystallography study on ribosome recycling: the mechanism of binding and action of RRF on the 50S ribosomal subunit. EMBO J. 24(2):251-60.
- Christodoulou J., Larsson G., Fucini P., Lindorff-Larsen K., Connell S.R., Pertinhez T.A., Hanson C.L., Nierhaus K.H., Robinson C.V., Schleucher J., and C.M. Dobson. (2004) Heteronuclear NMR investigations of dynamic regions of intact E. coli ribosomes. Proc. Natl Acad. Sci. USA 101(30): 10949-10954.
- Gilbert R.J.C., Fucini P., Connell S., Fuller S.D., Nierhaus K.H., Robinson C.V., Dobson C.M., and D.I. Stuart. (2004) Three-dimensional structures of translating ribosomes by Cryo-EM. Mol Cell. 14(1): 57-66.

External Funding

- 'X-ray crystallography studies describing functional states of the ribosome in complex with regulatory factors and ligands'. DFG project FU-579.
- 'Characterization of the ribosomal protein of healthy and infected sugar cane plants'. DFG project 445 AGY-112/20/06, in collaboration with Dr. Elsayed Hafez, Mubarak City for Scientific Research & Biotechnology Application, Egypt.
- 'Improvement in the preparation and resolution of Deinococcus radiodurans ribosome crystals'. Pfizer Inc, New London, USA.
- 'Structural study on the binding site of a Novel Ribosome Inhibitor of translation'. Abbott Laboratories, Illinois, USA.

Patents Applications

- 'A crystal of a complex and use thereof', (US60/701,936), concerning the development of new antibiotics on the bases of the atomic details of the interaction between TF and the ribosome (Schlünzen *et al.*, 2005, Structure).
- 'Screening method for Ribosome Recycling Inhibitor', (US60/632,309), concerning the development of new antibiotics on the bases of the atomic details of the interaction between RRF and the ribosome (Wilson *et al.*, 2005, EMBO J.).

Collaborations

- Prof. A. Bogdanov, Moscow State University, Russia. 'Exploring the pathway and interaction of the polypeptide nascent chain within the ribosomal tunnel'.
- Prof. S. Wilbanks, Otago University, New Zealand. 'On the interaction between the polypeptide nascent chain and the TF chaperon'.
- Prof. Y. Kobayashi, Osaka University, Japan. 'X-ray crystallography studies on functional complex involving the small ribosomal subunit from T. thermophilus'.
- Prof. C.M. Dobson, Cambridge University, England. 'NMR studies on Co-translational folding'.
- Prof. C. Robinson, Cambridge University, England. 'Characterization of the ribosomal stalk from different organisms'.
- Prof. R. Agrawal, Wadsworth Center, USA. 'CryoEM analysis of small GTPases, regulatory elements of the translational apparatus'.
- Prof. C. Spahn, Universitätsmedizin Berlin, Germany. 'CryoEM analysis of the ribosome in complex with various translational factors'.

Translation, Structure and Functions of Ribosomes

Head
Prof. Dr. Knud Nierhaus
Phone: +49 (0)30-8413 1700
Fax: +49 (0)30-8413 1690
Email: nierhaus@molgen.mpg.de

Secretary
Gabriele Eder
Phone: +49 (0)30-8413 1344
Fax: +49 (0)30-8413 1690
Email: eder_g@molgen.mpg.de

Scientist
Dr. Madina Iskakova

Technicians
Edda Einfeldt
Detlev Kamp

Graduate Students
Witold Szaflarski
Oliver Vesper
Daniela Wittek (since 5/2006)
Romi Gupta (since 4/2006)
Zhala Karim (since 3/2006)

Students
Daniela Kaul (since 7/2006)
Daniela Wittek (8/2005 - 3/2006)

Ribosome Groups

Scientific Overview

The reference numbers refer to those of the reference list

Ribosome assembly

In contrast to reports in high-level journals we demonstrated that the chaperone DnaK is not involved in ribosome assembly at physiological temperatures (ref. 5).

Role of the E site, tRNA release from the E site

We identified an essential role of the E site, a universal tRNA binding site on ribosomes (ref.13), *viz*. the importance of this site for the maintenance of the reading frame. The step of the ribosomal elongation cycle, during which the deacylated tRNA is released from the E site has been identified as the binding of the ternary complex to the A site before EF-Tu dependent GTP cleavage (refs. 21, 25).

Tetracycline resistance mechanisms

One of the most common resistance mechanism against tetracycline in human and veterinary medicine has been elucidated, the factor involved is Tet(O) (together with the Taylor group in Edmonton, Canada), refs. 3, 8.

Localization of the chaperone trigger factor

The localization of the trigger factor was assessed by means of neutron scattering techniques (ref. 2).

The protein LepA is an elongation factor present in all bacteria, mitochondria and chloroplast

An outstanding finding of the last years was the identification of a new ribosomal elongation factor. We demonstrated that the extremely conserved LepA protein, present in all bacteria and mitochondria, is a third elongation factor required for accurate and efficient protein synthesis. LepA is essential for cell viability at high ionic strength, where – without LepA – translocation is impaired and translational misreading increased. LepA restores translational fidelity, whereas it cannot counteract decoding errors induced by antibiotics such as aminoglycosides and edeine. LepA is a back-translocator of post-translocational ribosomes, and the results suggest that it recognizes ribosomes after a defective translocation reaction and induces a back-translocation, thus giving EF-G a second chance to translocate the tRNAs correctly. We suggested renaming LepA elongation factor 4 (EF4). Submitted to Cell.

Further activities in cooperation with other groups

Mechanisms of erythromycin and related compounds (ref. 4) and of that of pactamycin and edeine together with the Kalpaxis group in Patras, Greece (ref. 9) and the Fucini group (ref. 23). Protein expression in nanowell chip format (together with the Lehrach department; ref. 10). Structure analysis of the ribosome by NMR techniques (refs. 1, 14), identification of hexameric (instead of the canonical tetrameric) L7/L12 in thermophile bacteria (together with the Robinson group in Cambridge; ref. 19). Identification of the protein Bip controlling the globular regulator Fis (together with the O'Connor group in Southampton; ref. 15). Work on the mechanism of the monster molecule tmRNA (together with the Shpanchenko group, Moscow, ref. 18).

Selected Information

Publications

1. C. L. Hanson, P. Fucini, L. L. Ilag, K. H. Nierhaus and C. V. Robinson: Dissociation of intact *Escherichia coli* ribosomes in a mass spectrometer – Evidence for conformational change in a ribosome elongation factor G complex. J. Biol. Chem. 278:1259-1267 (2003)

2. G. Blaha, D. N. Wilson, G. Stoller, G. Fischer, R. Willumeit and K. H. Nierhaus: Localization of the trigger factor binding site on the ribosomal 50S subunit. J. Mol. Biol. 326:887-897 (2003)

3. S. R. Connell, C. A. Trieber, G. P, Dinos, E. Einfeldt, D. E. Taylor and K. H. Nierhaus: Mechanism of Tet(O)-mediated tetracycline resistance. EMBO J. 22:945-953 (2003)

4. G. P. Dinos, S. R. Connell, K. H. Nierhaus, and D. L. Kalpaxis: Erythromycin, roxithromycin, and clarithromycin: Use of slow-binding kinetics to compare their *in vitro* interaction with a bacterial ribosomal complex active in peptide bond formation. Mol. Pharmacol. 63:617-623 (2003)

5. J.-H. Alix and K. H. Nierhaus: DnaK-facilitated ribosome assembly in *Escherichia coli* revisited. RNA 9:787-793 (2003)

6. U. Stelzl, J. M. Zengel, M. Tovbina, M. Walker, K. H.Nierhaus, L. Lindahl, and D. J. Patel: RNA-structural mimicry in Escherichia coli ribosomal protein L4-dependent regulation of the S10 operon.J. Biol. Chem. 278: 28237-28245 (2003)

7. D. N. Wilson and K. H.Nierhaus: The Ribosome through the looking glass. Angew. Chem. Int. Ed. Engl. 42: 3464-3486 (2003): Das Ribosom unter der Lupe. Angew. Chem. 115:3586–3610 (2003)

8. S. R. Connell, D. M. Tracz, K. H. Nierhaus and D. E. Taylor: Ribosomal protection proteins and their mechanism of tetracycline resistance. Antimicrob. Agents Chemother. 47:3675-3681 (2003)

9. G. Dinos, D. N. Wilson, Y. Teraoka, W. Szaflarski, P. Fucini, D. Kalpaxis and K. H. Nierhaus: Dissecting the ribosomal inhibition mechamisms of edeine and pactamycin: The universally conserved residues G693 and C795 regulate P-site RNA binding. Mol. Cell 13:113-124 (2004)

10. P. Angenendt, L. Nyarsik, W. Szaflarski, J. Glökler, K. H. Nierhaus, H. Lehrach, D. J. Cahill and A. Lueking: Cell-free protein expression and functional assay in nanowell chip format. Anal. Chem. 76:1844-1849 (2004)

11. R. J. C. Gilbert, P. Fucini, S. Connell, S. D. Fuller, K. H. Nierhaus, C. V. Robinson, C. M. Dobson and D. I. Stuart: Three-dimensional structures of translating ribosomes by Cryo-EM. Mol. Cell 14:57-66 (2004)

12. M. M. Anokhina, A. Barta, K. H. Nierhaus, V. A. Spiridova and A. M. Kopylov: Mapping of the second tetracycline binding site on the ribosomal small subunit of *E.coli.* Nucleic Acids Res. 32:2594-2597 (2004)

13. V. Márquez, D. N. Wilson, W. P. Tate, F. Alonso-Triana and K. H. Nierhaus: Maintaining the Ribosomal Reading Frame: The Influence of the E Site during Translational Regulation of Release Factor 2. Cell. 118:45-55 (2004)

14. J. Christodoulou, G. Larsson, P. Fucini, S. R. Connell, T. A. Pertinhez, C. L. Hanson, C. Redfield, K. H. Nierhaus, C. V. Robinson, J. Schleucher and C. M. Dobson: Heteronuclear NMR investigations of dynamic regions of intact *Escherichia coli* ribosomes. Proc. Natl. Acad. Sci. USA 101:10949-10954 (2004)

15. R. M. Owens, G. Pritchard, P. Skipp, M. Hodey, S. R. Connell, K. H. Nierhaus and C. D. O'Connor: A dedicated translation factor controls the synthesis of the global regulator Fis. EMBO J. 23:3375-3385 (2004)

16. D. N. Wilson and K. H. Nierhaus: The how and Y of cold shock. Nature Str. Mol. Biol. 11:1026-1028 (2004)

17. A. Bonincontro, C. Cametti, K. H. Nierhaus, M. G. Ortore and G. Risuleo: Ribosomes deprived of select proteins show similar structural alterations induced by thermal treatment of native particles. Cell Biochem Biophys. 42:55-60 (2005).

18. O. V. Shpanchenko, M. I. Zvereva, P. V. Ivanov, E. Y. Bugaeva, A. S. Rozov, A. A. Bogdanov, M. Kalkum, L. A. Isaksson, K. H. Nierhaus and O. A. Dontsova: Stepping tmRNA through the ribosome. J. Biol. Chem. 280:18368-18374 (2005)

19. L. L. Ilag, H. Videler, A. R. McKay, F. Sobott, P. Fucini, K. H. Nierhaus and . Robinson: Heptameric (L12)6/L10 rather than canonical pentameric complexes are found by tandem MS of intact ribosomes from thermophilic bacteria. Proc. Natl. Acad. Sci. U S A. 102:8192-8197 (2005)

20. D. N. Wilson K. H. Nierhaus: RelBE or not to be. Nature Str. Mol. Biol. 12:282-284 (2005)

21. G. Dinos, D. L. Kalpaxis, D. N. Wilson and K. H. Nierhaus: Deacylated tRNA is released from the E site upon A site occupation but before GTP is hydrolyzed by EF-Tu. Nucleic Acids Res. 33:5291-5296 (2005)

22. D. N. Wilson and K. H. Nierhaus: Ribosomal proteins in the spotlight. Crit. Rev. Biochem. Mol. Biol. 40:243-267, (2005)

23. D. N. Wilson, J. M. Harms, K. H. Nierhaus, F. Schlünzen, P. Fucini: Species-specific antibiotic-ribosome interactions: implications for drug development. Biol. Chem. 386:1239-1252 (2005)

24. H. S. Seo, S. Abedin, D. Kamp, D. N. Wilson, K. H. Nierhaus, B. S. Cooperman: EF-G-dependent GTPase on the ribosome. conformational change and fusidic acid inhibition. Biochemistry 45:2504-2514 (2006).

25. K. H. Nierhaus: Decoding errors and the involvement of the E-site. Biochimie Mar 23 (2006); [Epub ahead of print]

Otto Warburg Laboratory

Bioinformatics / Structural Proteomics

Head
Dr. Michael Lappe
Phone: +49 (0)30-8413 1717
Fax: +49 (0)30-8413 1130
Email: lappe@molgen.mpg.de

Scientist
Dr. Brijnesh Johannes Jain

Engineer
Jose Duarte

Graduate Students
Ioannis Filippis
HyeongJun An
Henning Stehr
ChangBum Park

Internships/Master Students
Azime Akman
Nadine Atwood
Tecla Lucca

Bachelor Student
Joachim von Eichborn

Independent Junior Research Group

Scientific Overview

The functional characterisation of all the genes and their gene products is the main challenge of the post-genomic era. Interaction information for every gene-product is a clean way to assemble the jigsaw puzzle of proteins into a functional map. Recent advances in experimental and computational techniques have enabled the study of interactions among proteins on a large-scale. Our present research interests revolve around novel algorithms for interaction networks, aiming at the inference of protein structure, function, signalling pathways and ultimately entire interactomes.

Interaction patterns operationalise protein function in a computationally clean and graspable way. Since the architecture of biological networks differs from classical random networks, the functional maps contain a signal that can be used for predictive purposes. Interestingly, this distinct architecture of biological networks is observed at various levels of organisation, ranging from the atomic level up to social networks. This allows us to transfer algorithmic principles and apply them to different networks.

We have already entered a stage where high-throughput protein interaction data is being generated at a rate that resembles the influx of sequence data a decade ago. Hence it will be crucial to be able to analyse these interactions in molecular detail. In this context we follow the working hypothesis that protein binding and folding adhere to the same principles. This is supported by the fact that at least a third of all human proteins are estimated to have intrinsically unstructured regions which fold upon binding. The intensive study of the protein-folding problem has generated detailed experimental data for a number of proteins, although a brake-through in de-novo structure prediction has not yet been achieved. The resemblance of residue interaction graphs and protein-protein interaction networks has motivated us to tackle structure prediction, docking and catalysis from a graph theoretic perspective, leading to novel methods to address sequence-structure relationships, conformational flexibility, disorder and co-operativity.

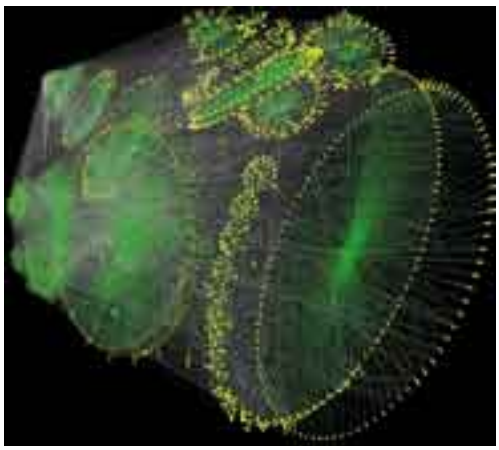


Figure 1: The Protein-Protein Interaction network in *S. cerevisae* visualised in 3D

Current Work

The current work on Protein-interaction data focuses on the quantification of noise levels in high-throughput datasets down to a verification of individual interactions, addressing the high levels of both false positives and negatives in the current data. For predictive purposes, mapping protein-protein interaction reliably down to the level of domain-domain interactions remains an important question. Also, the evolutionary aspects of the observed degree distributions are not very well understood. In practical terms, algorithms that integrate phylogenetic analysis with interaction networks should allow for very effective and accurate genome-wide prediction of protein function.

Besides the developing of novel algorithms for the analysis of protein interaction networks we concentrate on the prediction of protein function and structure. Using a graph-based approach, we tackle issues of conformational flexibility and unstructured proteins, especially in binding regions and in Protein-Protein interfaces. Obviously, the definition of relevant residue contacts has a profound impact on the resulting networks. We apply information theoretic criteria to address this central question.

For the analysis of protein structures the entire MSD (Macromolecular Structure Database) has been converted into a suitable graph-representation using both geometrical distance thresholds and Delauney triangulation. We are implementing the appropriate databases on our new cluster that has been designed to cope with the resulting amounts of data. The residue interaction graphs contain approximately 200 million residue-residue interactions and represent an empirical force field that is being refined using a variety of graph measures, additional experimental data and statistical methods.

Furthermore we have devised an algorithm that reliably predicts viable split-sites for the design of enzymes as molecular probes for use in high-throughput in-vitro interaction screening methods. For structure prediction and the analysis of interactions we concentrate on

- Contact map prediction as a crucial step towards structure prediction
- Novel methods for the generation of fragment libraries
- Classification of Protein-Protein Interfaces using geometric and graph based approaches
- Analysis of catalytic sites and in-silico ligand screening

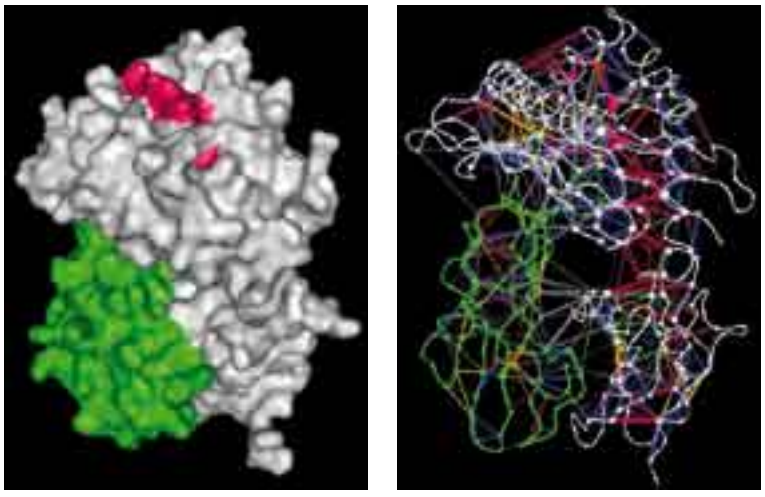


Figure 2: The Major Histocompatibility Complex MHC1
a) in surface representation b) as a residue interaction graph

Future directions

In collaboration with experimentalists sequence-structure relationships are investigated with the aim to highlight determinants of binding specificity and folding pathways. In this context existing NMR and x-ray data is analysed with respect to mutations and kinetic information. For interaction networks, we have developed algorithms that help to determine 80% of an interactome with less than 20% of the experimental effort needed in unguided strategies. This will contribute to focus TAP/MS experiments and yield a scaffold of the interactome with high coverage and reliability while minimizing time and resources. Currently we are working on an adaptation of this strategy to Yeast-two-Hybrid screens.
We are collaborating with experimentalists on novel reporter proteins and systems for high-throughput interaction detection (ProSA). In the near future, we will contribute in-silico structure and interaction predictions that go beyond the limits of established methods. This will be mutually beneficial as testing such predictions experimentally will help to focus experimental effort on promising targets while improving the prediction algorithms.

Selected Information

Publications

- Paszkiewicz, K. H., Sternberg, M. J. & Lappe, M. (2006). Prediction of viable circular permutants using a graph theoretic approach. Bioinformatics 22, 1353-1358.
- Park, D., Lee, S., Bolser, D., Schroeder, M., Lappe, M., Oh, D. & Bhak, J. (2005). Comparative interactomics analysis of protein family interaction networks using PSIMAP (protein structural interactome map). Bioinformatics 21, 3234-40.
- Gong, S., Yoon, G., Jang, I., Bolser, D., Dafas, P., Schroeder, M., Choi, H., Cho, Y., Han, K., Lee, S., Choi, H., Lappe, M., Holm, L., Kim, S., Oh, D. & Bhak, J. (2005). PSIbase: a database of Protein Structural Interactome map (PSIMAP). Bioinformatics 21, 2541-3.
- Lappe, M. & Holm, L. (2005). Algorithms for protein interaction networks. Biochem Soc Trans 33, 530-4.

Appointments

- Training Coordinator of European Network "ProSA"

Training and Teaching

- ProSA workshop on bioinformatics methods in Protein Structure and Interaction analysis. (9/2006)
- Public Relations: Berlin Long Night of Sciences, open house visit and presentation of the laboratory to the general public (6/2005 and 5/2006)

Organisation of scientific events

- Chairman of the Dahlem Colloquium committee, an institute wide seminar series (5/2006-present)

External Funding

- European network on selection and analysis of protein-protein interactions Marie-Curie Research & Training Network "ProSA" MRTN-CT-2005-019475

Collaborations

- "Sampling properties of protein interaction networks" with Dr. Michael Stumpf, Imperial College, London
- "Prediction of viable circular permutants" with Konrad H. Paszkiewicz and Dr. Michael Sternberg, Imperial College, London
- "Compartmentalised selection of protein-protein interactions using a split-protein sensor", with Pascale Mathonet and Hans Leemhuis, Research Group Hollfelder, University of Cambridge.
- "Binding sites in instrincially unstructured regions" Colin Kleanthous, University of York.
- "Covering strategies and domain interactions in Y2H" with Ulrich Stelzl, AG Wanker, MDC, Berlin
- "Analysis and visualisation of Chromosome 21 interactions" with Hans-Jörg Warnatz, Research Group Yaspo, Dept. Lehrach, MPIMG Berlin
- "Mutation Prediction at the GDF5/BMP Interface" with Petra Seemann, Research Group Mundlos, MPIMG Berlin
- "Phylogenetic analysis of residue interaction graphs" with Peter Arndt, Dept. Vingron, MPIMG Berlin

Work as a scientific referee

- Bioinformatics
- BMC Bioinformatics
- Serving on the Programme committee of Bioinformatics Research and Development conference BIRD'07

Miscellaneous Research Groups

Microscopy

Head
Dr. Rudi Lurz
Phone: +49 (0)30-8413 1644
Fax: +49 (0)30-8413 1385
Email: lurz@molgen.mpg.de

Technician
Gerhild Lüder

Scientific Overview

The microscopy group provides a central scientific service unit of the institute. The lab is running two transmission electron microscopes in close cooperation with the USN (Anwenderzentrum) project. A Philips CM100 is the "working horse" of the lab. It is equipped with a 1k CCD camera and is also used for screening vitrified samples. A new FEI Spirit TEM is replacing the almost 30 years old EM400. The Spirit is designed for routine and cryo work. In combination with the highly sensitive 2k Eagle camera we will be able to acquire data for image processing from negatively stained or vitrified samples which can be used for initial 3D-reconstructions. Additionally we will be able to acquire tomographic tilt series and do subsequent calculation of 3D-structure of sections, negatively stained objects or samples in vitreous ice.

A broad range of preparation methods is established. Our main focus is on:
– ultra-thin sections of embedded samples;
– specific labelling of sections or isolated structures;
– visualization of nucleic acid - protein interactions;
– fine structural analysis of protein complexes or viruses after negative staining;
– cryo preparation of the samples in vitreous ice.

Within the institute we perform projects with all departments having wet labs.
In collaboration with M. Kolancyk (Group Mundlos) we have localized by immuno-gold labeling a nitric oxide synthase in mitochondria that is a mammalian ortholog of ATNOS1 protein. This protein may contribute in mammals to multiple physiological processes during embryogenesis including liver haematopoesis and bone development.
A combination of EM and LM techniques is used to visualize expression patterns in sections of *in situ* labelled of sea urchin and amphioxus embryos (G. Panopoulou, Abt. Lehrach).
In the song-learning period of young males of zebra finch new neurons are formed which integrate into the high vocal center (HVC). After tracer (dextran amin biotin) injection in the RA we could detect the projecting neurons in the (HVC) by EM. Now we are looking for gap junctions between neurons in the HVC. Neurons are forming clusters and may communicate via gap junctions. (S. Scotto, Group C. Scharff, Abt. Ropers).
Male transmission ratio distortion (TRD) is a property of mouse t haplotypes requiring the t complex responder locus (Tcr). Within sperms from mouse epididymis we are localizing the responder protein by IEM. Now we want to localize the expression by *in situ* hybridization (N. Véron, Abt. Herrmann).

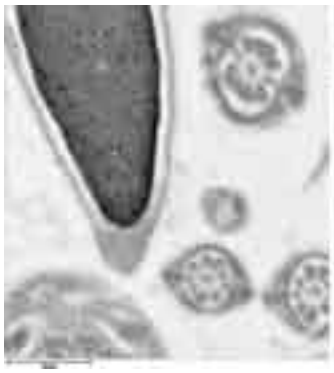


Figure 1:
IEM gold labelling of the responder in mouse sperm head

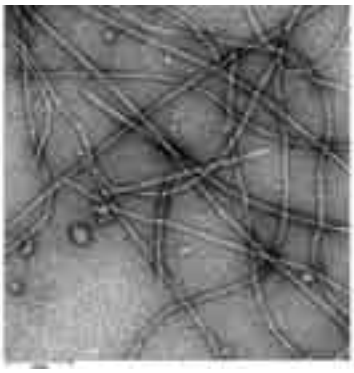


Figure 2:
β amyloid fibrils formed at low pH

Figure 3:
phage A511 from *Listeria monocytogenes*

Collaborations outside the institute are often in continuation of projects started during the stay of the people in this institute. One example is the group of E. Wanker (MDC, Berlin-Buch). He is focused on neurodegenerative diseases, which show formation of protein aggregates in the brain (amyloids) like Chorea Huntington, Parkinson's or Alzheimer's disease. The collaboration started about 10 year ago in Abt. Lehrach and up to now one dozen papers are published. Currently we are analysing the influence of drugs on the formation of huntingtin, α-synuclein and β-amyloid fibrils. EM visualizes the effects of the most promising drugs on fibril formation and structure in parallel.
A similar fruitful cooperation is with Paulo Tavares (CNRS, Gif-sur-Yvette). Our collaboration started during his Ph.D. time in the lab of T.A. Trautner on the structure of the portal protein (gp6) of phage SPP1. Over the years we analyzed by EM different aspects of infection, morphogenesis and packaging of this phage. At the time we do image processing to analyze the oligomerization of different gp6 mutants. Another project is the high resolution structure of the tail (with connector) in vitrified samples. The receptor (YueB) from the host bacterium (*B. subtilis*) is defined and purified (and visualized by EM). The binding of YueB to the phage triggers DNA ejection. We use binding of YueB under different conditions to study the DNA ejection process. An *in vitro* system of the procapsid formation is established which requires only three structural proteins: the portal protein (gp6), the scaffolding protein (gp11) and the major capsid protein (gp13). We are analyzing structure and symmetry of the proheads formed.

Selected Information

Publications

- Boeddrich, A., Gaumer, S., Haacke, A., Tzvetkov, N., Albrecht, M., Evert, B. O., Muller, E. C., Lurz, R., Breuer, P., Schugardt, N., Plassmann, S., Xu, K., Warrick, J. M., Suopanki, J., Wullner, U., Frank, R., Hartl, U. F., Bonini, N. M., and Wanker, E. E. (2006). An arginine/lysine-rich motif is crucial for VCP/p97-mediated modulation of ataxin-3 fibrillogenesis. Embo J 25(7), 1547-58.
- Sao-Jose C., Lhuillier S., Lurz R., Melki R., Lepault J., Santos MA. and Tavares P. (2006) The ectodomain of the viral receptor YueB forms a fiber that triggers ejection of bacteriophage SPP1 DNA. .J Biol Chem. 281(17):11464-70.
- Zemojtel, T., Kolanczyk, M., Kossler, N., Stricker, S., Lurz, R., Mikula, I., Duchniewicz, M., Schuelke, M., Ghafourifar, P., Martasek, P., Vingron, M., and Mundlos, S. (2006). Mammalian mitochondrial nitric oxide synthase: characterization of a novel candidate. FEBS Lett 580(2), 455-62.
- Carrasco, B., Ayora, S., Lurz, R., and Alonso, J. C. (2005). Bacillus subtilis RecU Holliday-junction resolvase modulates RecA activities. Nucleic Acids Res 33(12), 3942-52.
- Korn, C., Scholz, S. R., Gimadutdinow, O., Lurz, R., Pingoud, A., and Meiss, G. (2005). Interaction of DNA fragmentation factor (DFF) with DNA reveals an unprecedented mechanism for nuclease inhibition and suggests that DFF can be activated in a DNA-bound state. J Biol Chem 280(7), 6005-15.
- Ziegelin, G., Tegtmeyer, N., Lurz, R., Hertwig, S., Hammerl, J., Appel, B., and Lanka, E. (2005). The repA gene of the linear Yersinia enterocolitica prophage PY54 functions as a circular minimal replicon in Escherichia coli. J Bacteriol 187(10), 3445-54.
- Carrasco, B., Cozar, M. C., Lurz, R., Alonso, J. C., and Ayora, S. (2004). Genetic recombination in Bacillus subtilis 168: contribution of Holliday junction processing functions in chromosome segregation. J Bacteriol 186(17), 5557-66.

- Ayora, S., Carrasco, B., Doncel, E., Lurz, R., and Alonso, J. C. (2004). Bacillus subtilis RecU protein cleaves Holliday junctions and anneals single-stranded DNA. Proc Natl Acad Sci U S A 101(2), 452-7.
- Albrecht, A. N., Kornak, U., Böddrich, A., Süring, K., Robinson, P. N., Stiege, A. C., Lurz, R., Stricker, S., Wanker, E. E., and Mundlos, S. (2004). A molecular pathogenesis for transcription factor associated poly-alanine tract expansions. Hum Mol Genet 13(20), 2351-9.
- Pingoud, V., Sudina, A., Geyer, H., Bujnicki, J. M., Lurz, R., Lüder, G., Morgan, R., Kubareva, E., and Pingoud, A. (2004). Specificity changes in the evolution of type II restriction endonucleases: a biochemical and bioinformatic analysis of restriction enzymes that recognize unrelated sequences. J Biol Chem.

Selected international cooperations

- J.C. Alonso, Centro Nacional de Biotecnologia, Departamento de Biotecnologia Microbiana, Madrid: Replication and recombination in B. subtilis and phage SPP1.
- M. Espinosa, Centro de Investigaciones Biológicas (CSIC), Madrid: Replication and mobilization of the promiscuous plasmid pMV158.
- M. J. Loessner, Swiss Federal Institute of Technology (ETH), Zürich: DNA organization of Listeria phage A511.
- E. Orlova, Birkbeck College, Dept. of Crystallography, University of London: Image processing of data sets from components of phage SPP1.
- J.N. Reeve, Ohio State University, Columbus: Mapping of the archaeal transcription repressor (TrpY) from Methanothermobacter.
- P. Tavares, Laboratoire de Virology Moléculaire et Structurale, CNRS, Gif-sur-Yvette: Life cycle of B. subtilis bacteriophage SPP1 (infection, morphogenesis, packaging); fine structural analysis of the phage components.
- J. Zakrzewska-Czerwinska, Institute of Immunology and Experimental Therapy, Polish Academy of Sciences, Wroclaw: Architecture of the partitioning complexes (segregosomes) in Streptomyces coelicolor.

High throughput technology group (htpt)

Head
Dr. Richard Reinhardt
Phone: +49 (0)30-8413-1226
Fax: +49 (0)30-8413-1365
Email: reinhardt@molgen.mpg.de

Secretary
Tamara Safari
Phone: +49 (0)30-8413-1126
Fax: +49 (0)30-8413-1365
Email: safari@molgen.mpg.de

Technicians
Beatrice Baumann
Gabriele Bläß
Ilona Hauenschild
Katja Heitmann
Pamela Kepper
Anna Kosiura
Isabelle Kühndahl
Sonia Paturej
Ines Müller
Rönick Monique
Silvia Schmoger
Mario Sontag
Janina Thiel

Technicians/TFH
Patricia Klemmer
Bettina Moser
Grzegorz Wozniak

Scientists
Dr. Michael Kube
Dr. Bernd Timmermann

PhD-Student
Heiner Kuhl

Student
Steffen Scheer
Ute Trauer

Progress Report

The **htpt**-group is a co-operation partner of the international human genome sequencing project consortium, as described by H. Lehrach (Dept. Vertebrate genomics). The group has established a good infrastructure for large-scale genomic analysis projects such as sequencing, mutation analysis and mass spectrometry, as has been partially described in the section of the service group.

The recent publications of human chromosomes and of the completed human genome sequence were the ultimate successes of our major effort within the time period since 2000. Important steps on this successful way are the sequencing and final analysis of chromosome 21, the second finished human chromosome, still one of the most accurately analysed one, and our contribution to several regions of the human chromosomes 1, 3, 17, and X. Our latest finalised project along this line is the completed elucidation of chimpanzee chromosome 22 (recently to be renamed chromosome 21), the ortholog to human chromosome 21. The whole project was organised by an German-Asian consortium, wherein MPIMG was responsible for the German part. These results, being summarised within the present report period, are not only important because chimpanzee is our closest relative, it is the first time that a large genomic arrangement, two complete chromosomes of man and chimpanzee, are comparatively analysed. Therefore, not only genes and variations within the coding elements are comparable, but also intronic regions and even more important, promoter elements are accessible for any comparative analysis and elucidation.

In addition, we are involved in the analysis of model organisms such as mouse (chromosome 2 and 6), rhesus MHC and the complete analysis of the rat MHC (RT1) complex, which plays an important role in infectious diseases. The MHC region belongs to the most densely packed, gene rich regions and although it spans only over a 4 Mbases area, we have identified 220 genes, nearly as many as in the human chromosome 21 region, which is about 34 Mbases large.

Other launched projects concern contributions to the final sequence of chimpanzee chromosomes X and Y, with special interest to Xq28 and regions associated with mental retardation. We are also involved in national and international projects, from bacterial genomes to model organisms (together with Dr. Michael Kube) like the urochordate *Oikopleura*, as listed below (project grants and MPG projects).

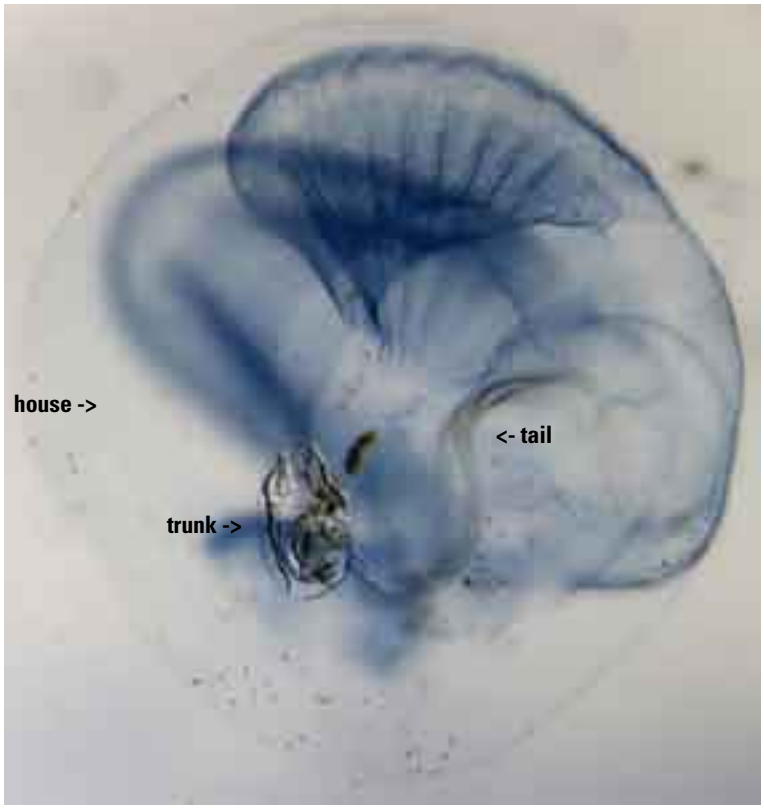


Figure1: Oikopleura in its house (E. Thompson, Norway)

The early scientific interest related to *Oikopleura dioica* was focused on questions of systematics, the phenomenon of "marine snow" and of bioluminousity, research in *Oikopleura´s* nervous system, and ecological questions, like the influence on picoplancton. With a genome size of only around 75 Mbases (estimated number of 15.000 genes), smaller than *C. elegans*, and less than half that of *D. melanogaster*, the genome of *Oikopleura dioca* gives the chance for a closer look inside an early chordate genome. In addition, this organism has also other interesting features, making it a key system to understand the functions of human/ vertebrate genome.

Within this line of interest is our scientific contribution to varies established EU project: for the NoE **M**arine **G**enomics **E**urope, where RR is also member of the scientific steering committee.

In addition, we are managing several NGFN-2 projects listed below, specifically should be mentioned the project for disease gene identification and systematic re-sequencing of candidate genes of genomic regions of interest, where data exchange between the clinical partners, also non-NGFN-2 partners, is organized by a Web interface (*http://www.resequencing.mpg.de/*). Finished genomic sequence data are submitted to public data bases or/and are presented on our project Web pages.

Infrastructure of htpt-group

The htpt-group has established a good infrastructure for large-scale genomic analysis projects such as sequencing, medically related re-sequencing, mutation analysis and mass spectrometry, using most advanced methods and hardware systems.

DNA samples are purified, using our patent related magn. beads methods and a novel PEG/org. solvent based precipitation methode (Heiner Kuhl). Our angular gel electrophoresis system (patented) is able to visualise 384 sample on a MTP-sized gel and to be automatly processed exploring a capacity of more than 15.000 samples per day. Besides these 'high-throughput highlights', our lab equipment involves all necessary items to run large scale projects, e.g. thermocycler, centrifuges, incubators, protein purification systems and sophisticated computer equipment.

Selected Information

Publications

- Burmeister T, Gökbuget N, Schwartz S, Reinhardt R, Hoelzer D, Thiel E «NUP214-ABL1 in adult T-ALL, the GMALL study group experience» Blood; 2006 Jul 27; [Epub ahead of print]
- Erkel C, Kube M, Reinhardt R, Liesack W, «Genome of Rice Cluster I Archaea – the Key Methane Producers in the Rice Rhizosphere» Science. 2006 Jul 21; 313(5785):370-2.
- Hecht J, Kuhl H, Haas SA, Bauer S, Poustka AJ, Lienau J, Schell H, Stiege AC, Seitz V, Reinhardt R, Duda GN, Mundlos S and Robinson PN. «Gene Identification and Analysis of Transcripts Differentially Regulated in Fracture Healing by EST Sequencing in the Domestic Sheep.» BMC Genomics; 206,7:172 [Epub ahead of print]
- Gebert C, Wrenzycki C, Herrmann D, Gröger D, Reinhardt R, Hajkova P, Lucas-Hahn A, Carnwath J, Lehrach H, Niemann H . «The bovine IGF2 gene is differentially methylated in oocyte and sperm DNA.» Genomics. 2006 Apr 24 [Epub ahead of print].
- International Human Chromosome 3 Consortium. (MPI contributors 9/109: Sudbrak R, Hennig S, Kube M, Lehrach H, Reinhardt R . «The DNA sequence, annotation and analysis of human chromosome 3.» Nature. 2006 Apr 27; 440(7088):1194-8.
- Tzschach A, Lenzner S, Moser B, Reinhardt R, Chelly J, Fryns JP, Kleefstrab T, Raynaud M, Turner G, Ropers H-H, Kuss A, Jensen LR . «Novel JARID1C/SMCX mutations in patients with X-linked mental retardation.» Hum Mutat. 2006 Apr;27(4):389.
- Goyal R, Reinhardt R, Jeltsch A. «Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase.» Nucleic Acids Res. 2006 Feb 25;34(4):1182-8.
- Ullrich S, Kube M, Schubbe S, Reinhardt R, Schuler D. «A hypervariable 130-kilobase genomic region of Magnetospirillum gryphiswaldense comprises a magnetosome island which undergoes frequent rearrangements during stationary growth.» J Bacteriol. 2005 Nov;187(21):7176-84
- Kube M, Beck A, Zinder SH, Kuhl H, Reinhardt R, Adrian L. «Genome sequence of the chlorinated compound-respiring bacterium Dehalococcoides species strain CBDB1.» Nature Biotechnol. 2005 Oct; 23(10):1269-73
- Ivens AC, et.al. (MPI contributors 6/100:Beck A, Borzym K, Klages S, Kube M, Reinhardt R,) «The genome of the kinetoplastid parasite, Leishmania major.» Science. 2005 Jul 15; 309(5733):436-42
- Hoefele J, Sudbrak R, Reinhardt R, Lehrack S, Hennig S, Imm A, Muerb U, Utsch B, Attanasio M, O'Toole JF, Otto E, Hildebrandt F. «Mutational analysis of the NPHP4 gene in 250 patients with nephronophthisis.» Hum Mutat. 2005 Apr; 25(4):411
- International Human Genome Sequencing Consortium. (MPI contributors 9/279: Sudbrak, R., Beck, A., Heitmann, K., Hennig, S., Klages, S., Kosiura, A., Müller, I., Reinhardt, R., Lehrach, H.,), (March 2005) «The DNA sequence of the human X chromosome.» Nature 434, 325-337
- Otto EA, Loeys B, Khanna H, Hellemans J, Sudbrak R, Fan S, Muerb U, O'Toole JF, Helou J, Attanasio M, Utsch B, Sayer JA, Lillo C, Jimeno D, Coucke P, De Paepe A, Reinhardt R, Klages S, Tsuda M, Kawakami I, Kusakabe T, Omran H, Imm A, Tippens M, Raymond PA, Hill J, Beales P, He S, Kispert A, Margolis B, Williams DS, Swaroop A, Hildebrandt F. «Nephrocystin-5, a ciliary IQ domain protein, is mutated in Senior-Loken syndrome and interacts with RPGR and calmodulin.» Nature Genet. 2005 Mar; 37(3):282-8
- Jensen LR, Amende M, Gurok U, Moser B, Gimmel V, Tzschach A, Janecke AR, Tariverdian G, Chelly J, Fryns JP, Van Esch H, Kleefstra T, Hamel B, Moraine C, Gecz J, Turner G, Reinhardt R, Kalscheuer VM, Ropers H-H, Lenzner S. «Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation.» Am J Hum Genet. 2005 Feb;76(2):227-36
- International Human Genome Sequencing Consortium. (MPI contributors: Sudbrak R, Beck A, Ben-Kahla A, Borzym K, Buczek D, Gimmel V, Heitmann K, Hennig S, Klages S, Kosiura A, Kube M, Müller I, Lehrach H, Schmoger S, Yaspo M-L, Reinhardt R) «Finishing the euchromatic sequence of the human genome.» Nature. 2004 Oct21;431(7011):931-45
- Tao J, Van Esch H, Hagedorn-Greiwe M, Hoffmann K, Moser B, Raynaud M, Sperner J, Fryns JP, Schwinger E, Gecz J, Ropers H-H, Kalscheuer VM. «Mutations in the X-Linked Cyclin-Dependent Kinase-Like 5 (CDKL5/ STK9) Gene Are Associated with Severe Neurodevelopmental Retardation.» Am J Hum Genet. 2004 Oct 21;75(6)
- Seo HC, Edvardsen RB, Maeland AD, Bjordal M, Jensen MF, Hansen A, Flaat M, Weissenbach J, Lehrach H, Wincker P, Reinhardt R, Chourrout D. «Hox cluster disintegration with persistent anteroposterior order of expression in the chordate Oikopleura dioica.» Nature, 2004 Sep 2; 431(7004):67-71

- Freude K, Hoffmann K, Jensen LR, Delatycki MB, des Portes V, Moser B, Hamel B, van Bokhoven H, Moraine C, Fryns JP, Chelly J, Gecz J, Lenzner S, Kalscheuer VM, Ropers H-H. «Mutations in the FTSJ1 gene coding for a novel S-adenosylmethionine-binding protein cause nonsyndromic X-linked mental retardation.» Am J Hum Genet. 2004 Aug; 75(2):305-9
- Abdelkhalek HB, Beckers A, Schuster-Gossler K, Pavlova MN, Burkhardt H, Lickert H, Rossant J, Reinhardt R, Schalkwyk LC, Muller I, Herrmann BG, Ceolin M, Rivera-Pomar R, Gossler A. «The mouse homeobox gene Not is required for caudal notochord development and affected by the truncate mutation.» Genes Dev. 2004 Jul 15; 18(14):1725-36
- International Chimpanzee Chromosom 22 Consortium (MPI contributors: Ben Kahla A, Lehrach H, Sudbrak R, Kube M, Reinhardt R, Yaspo M-L) «DNA sequence and comparative analysis of chimpanzee chromosome 22.» Nature 2004 May 27; 429(6990):382-8
- Kalscheuer VM, Freude K, Musante L, Jensen LR, Yntema HG, Gecz J, Sefiani A, Hoffmann K, Moser B, Haas S, Gurok U, Haesler S, Aranda B, Nshedjan A, Tzschach A, Hartmann N, Roloff TC, Shoichet S, Hagens O, Tao J, Van Bokhoven H, Turner G, Chelly J, Moraine C, Fryns JP, Nuber U, Hoeltzenbein M, Scharff C, Scherthan H, Lenzner S, Hamel BC, Schweiger S, Ropers H-H. «Mutations in the polyglutamine binding protein 1 gene cause X-linked mental retardation.» Nature Genet. 2003 Dec; 35(4):313-5
- Krüger M, Meyerdierks A, Glöckner FO, Amann R, Widdel F, Kube M, Reinhardt R, Kahnt J, Böcher R, Thauer RK and Shima S. «A conspicuous nickel protein in microbial mats that oxidise methane anaerobically.» Nature (426) 18/25 2003 Dec; 878-881
- Olbrich H, Fliegauf M, Hoefele J, Kispert A, Otto E, Volz A, Wolf MT, Sasmaz G, Trauer U, Reinhardt R, Sudbrak R, Antignac C, Gretz N, Walz G, Schermer B, Benzing T, Hildebrandt F and Omran H . «Mutations in a novel gene, NPHP3, cause adolescent nephronophthisis, tapeto-retinal degeneration and hepatic fibriosis.» Nature Genetics; 2003; 34: 455.459
- Glöckner FO, Kube M, Bauer M, Teeling H, Lombardot T, Ludwig W, Gade D, Beck A, Rabus R, Schlesner H, Amann R and Reinhardt R. «Complete genome sequence of the marine planctomycete Pirellula sp. strain 1.» PNAS. 2003 Jul; 100(14) 8298-8303
- Sauer S, Lehrach H and Reinhardt R. «MALDI mass spectrometry analysis of single nucleotide polymorphisms by photocleavage and charge-tagging.» Nucleic Acids Res., 2003 Jun 31, e63

Research Support & Central Units

Histology Unit

Head
Prof. Dr. Stefan Mundlos
Phone: +49 (0)30-8413 1263/ 1449
Fax: +49 (0)30-8413
Email: mundlos@molgen.mpg.de

Technician
Monika Oswald

The Core Histology Facility at the MPIMG has been established to assist the investigator in producing quality histologic slides. We have the equipment and expertise to produce sections from frozen, paraffin-embedded, and resin-embedded tissues. We can produce routine hematoxylin and eosin stained slides and can assist in performing special stains, immunohistochemistry, and in-situ hybridization. The facility covers the preparation and processing of tissues, their cutting/sectioning and staining. The lab houses modern equipment: much of which is fully or partially automated. This includes: Two automated cryostats, an automated tissue processor, a wax-embedding station, two automated microtomes for paraffin and plastic sectioning, a staining robot, and an automated slide coverslipping machine.

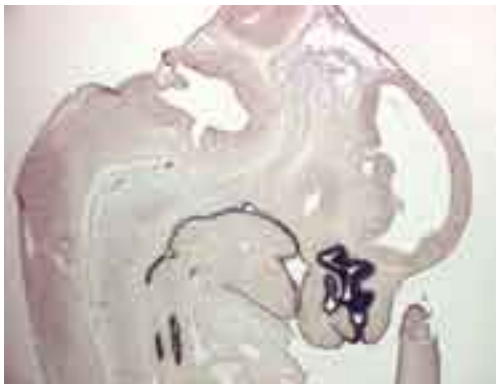


Figure 1: specific expression in the epithelium of the developing mouse E14.5 produced by robotic system.

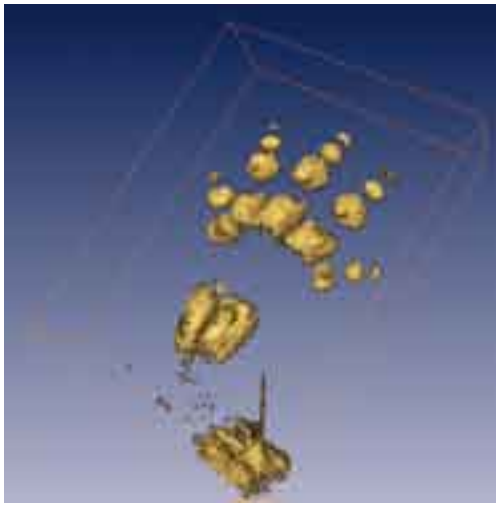


Figure 2: 3D reconstruction of ptch expression pattern in the developing mouse limb

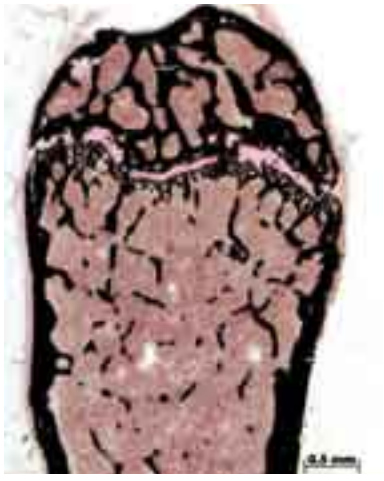


Figure 3: mouse tibia stained for mineralized tissue with von Kossa.

In situ hybridization on frozen and paraffin embedded tissues is established in a variety of techniques including a robotic system (GenePaint) allowing the processing of 96 slides in 2 days. For a throrough analysis of particular pathologies and gene functions it becomes more and more important to quantify the distribution of certain tissues or expression patterns. On a two-dimensional level this is achieved by histomorphometry techniques which can be provided by the Unit. During development and to assess complex expression patterns in an entire organ 3D images are necessary that can be reconstructed from stained serial sections.

Service Group Analytics and IT

Head
Richard Reinhardt
Phone: +49 (0)30-8413-1226
Fax: +49 (0)30-8413-1365
Email: reinhardt@molgen.mpg.de

Secretary
Tamara Safari
Phone: +49 (0)30-8413-1126
Fax: +49 (0)30-8413-1365
Email: safari@molgen.mpg.de

Analytics
Katja Borzym
Roman Pawlik
Sven Klages

Apprentice
Julia Cekanov
Student
Daniel Rhiel
Lab kitchen
Elisabeth Sahler

EDV
Alfred Beck
Donald Buczek
Peter Marquardt
Sven Püstow,
Frank Rippel

Apprentice
Stefan Heider
Josefine Mages
Students
Hermann Gebert
Marius Tolzmann

The scientific service group is active in the fields of DNA template preparation, purification, sequencing and sequence analysis, protein purification and analysis by Edman sequencing, MALDI-MS methods, enzyme preparation and purification as well as synthesis of highly specific oligo-nucleotides like Energy-Transfer primers etc.

Automation of procedures in any of these methods plays an important role. Another very important feature of our work is the miniaturisation, e.g. the down-scaling of reaction volumes and costs per reaction. The group has a good infrastructure for mutation analysis and DNA-sequencing, especially for genome sequencing, analysis and medically related re-sequencing. In parallel to conventional sequencing, we examine approaches designed for improving the efficiency of large-scale projects, like MS-MALDI methods for base determination and SNP detection based on the simplified GOOD assay and mini-sequencing.

The service costs for our main issues are calculated and those requesting the service are charged on a monthly basis of an individually assembled cost calculation. Besides the service aspects of our work, the group is a co-operation partner of the international HUGO project, European based projects, the national DFG and NGFN projects together with dept. Lehrach, Ropers and Vingron.

For this purpose several software tools were optimised or developed in close co-operation with the computing people of the group, using advanced LINUX/UNIX based clustered hardware. For further needs we will extend new clustering strategies (multi-processor PC-based LINUX-cluster) for the automated assembly of very large data sets, automated checking and editing steps and web based software tools for co-ordinating projects with external partners. Most of this will be done in close co-operation with dept. Vingron at the institute, the Sangercentre (UK) and the University of Washington (US).

Infrastructure analytic group

The analytic group has established a good infrastructure for large-scale genomic analysis projects such as sequencing, mutation analysis and mass spectrometry, using most advanced methods and systems for protein purification and sequencing (ABI 394 and Bruker 2-D-MALDI), SNP-detection and DNA-sequencing, mutation- and genetic variation analysis (WAVE-systems, MS-MALDI-Biflex, ABI 3730XL and MegaBace 4500, equipped with automated Caddy system). With the aim of miniaturisation, highly automated procedures including varies kinds of robotics are continuously developed, e.g. automated PCR, plasmide, fosmide and BAC template preparation and sequencing site, equipped with a CRS robotic arm on a track line, automated precipitation line incorporating a CRS arm, 96 and 384 Beckman Multimek pipetting systems, MTP centrifuge and a MTP-UV detection system, an automated clone hit picking device, designed to handle more than 200 MTP's in one run, incorporating a CRS arm, MTP centrifuge, Tecan freedom evo device to use 1536er MTP's and varies automated 384er Hydra/Twister mini systems to handle DNA samples to be purified, using our patent related magnetic beads methods and a novel PEG/org. solvent based fractionated precipitation method (Heiner Kuhl). Our angular gel electrophoresis system (patented) is able to visualise 384 sample on a MTP-sized gel and to be automatically processed exploring a capacity of more than 15.000 samples per day. A just recently installed 'Illumina platform' provides the latest generation of µ-array technology for expression profiling and genotyping for human and mouse samples. Besides these 'high throughput highlights', our lab equipment involves all necessary items to run large scale projects, e.g. thermocycler, centrifuges, incubators, protein purification systems and sophisticated computer equipment.

Infrastructure IT-group

The computer group is responsible for updating and servicing the biological databases and the corresponding software tools. It is also responsible for the operation and development of the whole IT-infrastructure of the institute, which includes workstation and server systems, wire based and wireless LAN, Internet access, Internet services and remote access via modem, ISDN and DSL, security devices (anti-virus and anti-SPAM software, data backup-, fire-wall). Our online storage capacity on disk-based file servers exceeds 85 TB of data, while the monthly backup volume has been dramatically increased due to the needs of new projects to about 25 TB, summing up to a total backup capacity of 200 TB on tape robot systems for back-up and additionally 30 TB for archived data. To manage and control the massive flow of data, our fibre based GigaBit-LAN, connecting laboratories in Fabeck- and Harnackstrasse to the campus Ihnestrasse, is segmented by about 100 manageable switches giving us the ultimate flexibility to control each segment and if necessary to configure each switch port individually. Presently we serve about 450 Window based PCs and 350 Linux/Unix systems with a variety of hard- and software components and about 80 MAC systems. A variety of WEB-server are protected by our fire-wall installation , 48 WEB-server are actively run and maintained by us, including hard- and software development and are serving the scientific departments as well as the service and administration groups.
Both, analytic- and IT-group are very active in the training and education of young technicians, students, trainees and apprentices .

Animal Facility & Transgene Unit

Head
Dr. Ludger Hartmann
Phone: +49 (0)30-8413-1189
Fax: +49 (0)30-8413-1197
Email: hartmann@molgen.mpg.de

Technicians
Ingo Voigt
Lara Mosch

Apprentice
Sina Ackermann
Angelina Döltsch
Stefan Gandecki
Viviane Haß
Maria Pohle
Eileen Jungnickel
Ramona Warnke

Care takers
Ulf Schroeder
Sonja Banko
Rebecca Kalsow
Mirjam Peetz
Sylvia Perkiewicz
Katja Reinsch
Robert Schilke
Janine Wetzel
Julia Wiesner
Carolin Willke

Service
Stefan Ilius

The Animal Facility with a new animal laboratory was completely brought into service in the year 2003. It provides an optimal research environment in the field of Laboratory Animal Science which includes the basic animal breeding and maintenance service for approximately 160 genetically modified and 25 wildtype mouse strains and technical services with a highly motivated staff. The mouse strains are kept under specified pathogen free (SPF) conditions in areas with restricted access. By using several physical barriers and standard working protocols we have been strongly committed ourselves to keep our rodent colony free of rodent pathogens. All strains are housed in individually ventilated caging systems and are handled under sterile conditions (approximately 6.000 cages).

The Animal Facility provides high standard services which includes:
– Experimental work and colony maintenance
– Cryopreservation of mouse embryos
– Sterile Embryotransfer
– Tissue biopsies
– Blood and organ collection
– Assistance in experimental design and techniques
– Training for researchers, caretakers and trainees
– Import/export of animals

For the management of these mouse strains and the offered services, a mouse-colony management software program (PyRAT®) was established. By using this software all mouse data are easily accessible for scientists.

The Transgene Unit of the Animal Facility was established in the years 2003/2004 to enable the successful and efficient generation of genetically modified mice for the scientific staff of the institute. It provides a centralised resource and state-of-the-art technology in generating knockout mice by injection of embryonic stem cells into mouse embryos or by aggregation of diploid and tetraploid embryos and transgenic mice by injection of DNA into the pronucleus of fertilised mouse oocytes. Above that the service provides cleaning, freezing and thawing of mouse embryos and in-vitro-fertilisation.

In the beginning the main work of this unit was focussed on rederivation of the more than 160 mouse strains of the institute. After that already more than 50 DNA-constructs were used for pronuclear microinjection to generate transgenic mice and 11 ES-cell clones were used for blastocyst microinjection to generate knockout mice.

In addition to this more than 60 mouse strains were cryopreservated for backup of important transgenic mouse lines of scientists of the institute and stored in liquid nitrogen (up to 250 embryos for each strain).
Other assisted reproductive technologies like sperm freezing, intracytoplasmatic sperm injection (ICSI) and ovary transplantation will be established in the near future.

The Fish Facility of our institute is set up to raise and keep up to 15.000 zebrafish (*Danio rerio*). The aquatic system is located in the animal house and consists of approximately 150 single tanks. It is available to all researchers at the institute, either to keep fish, or to provide eggs, embryos and larvae. For embryonic manipulations the pressure-driven microinjection technique for mRNA or DNA is available.

The Animal Facility also includes rooms outside of the animal house with bird cages and aviaries for breeding and maintaining of up to 500 zebra finches (AG Scharff).

Library

Librarians
Dipl.-Fachinf. Sylvia Elliger
Praxedis Leitner, M.A.
Phone: +49 (0)30-8413 1314
Fax: +49 (0)30-8413 1309
Email: library@molgen.mpg.de

Library Committee
Peter Arndt, Dept. Vingron
Vera Kalscheuer, Dept. Ropers
Enzo Russo, Dept. Lehrach
Silke Sperling, Dept. Lehrach
Ralf Sudbrak, Dept. Lehrach
Ralf Spörle, Dept. Herrmann
Sigmar Stricker, FG Mundlos
Knud Nierhaus, Ribosome Group
Michael Lappe, Otto Warburg Laboratory

The library covers all research areas of the institute and is organized as a reference library.It holds about 50,000 volumes and subscribes to over 120 scientific journals and series. The library still reduces the holdings of print journals but improves the electronic spectrum of scientific information and e-books. In addition to the web catalogue, increasing number of databases as well as electronic interlibrary loan service are offered. The library team undertakes searches in numerous online databases. The team also offers introduction courses in how to use the databases as well as courses in how to use other services of the library.

Seminars, with guest speakers, about recent changes in electronic information systems are offered for the scientists of the institute. All publications of the institute are edited and submitted into the yearbook of the Max Planck Society. The library team is part of the new pilot program within the Max-Planck-Society project “eSciDoc” and “OpenAcess”.

In autumn 2006 the library team will host the library conference of the biological medical section of the Max Planck Society.
The goal for further development of the library is still to improve the “Virtual Library”, a network of knowledge systems ensuring the delivery of information to researchers’ desktops wherever and whenever needed.

Service Group Imaging

Graphics
Monica Shevack (part-time)
Phone: +49 (0)30-8413 1313
Email: shevack@molgen.mpg.de

Photo
Katrin Ullrich (part-time)
Phone: +49 (0)30-8413 1312
Email: foto@molgen.mpg.de

Light Microscopes
Rudi Lurz
Phone: +49 (0)30-8413 1644
Email: lurz@molgen.mpg.de

The scientific illustrator cooperates closely with the scientists to prepare publications and presentations. Most of the artwork is computer generated (Pc and Mac) using programes such as Photoshop, Adobe Illustrator, Freehand or Dreamweaver. Additionally traditional hand drawings are prepared on demand. Recently the textbook "Systems Biology in Practice" was illustrated for the Kinetic Modeling group.

The classical wet darkroom is replaced by digital methods. The main issues are scans for presentations and photographic documentation including a broad range from reproduction of colonies on Petri dishes to the documentation of events of the institute. One major project is preparing pictures of the articles for the "e-Lager" store shop. Photographs and scans are finished in Photoshop.

Two confocal microscopes (LSM510 and LSM510 meta, purchased by Abt. Lehrach) and a fluorescence microscope (Z1 imager) are run as a core facility. Technical and research support are provided.

Administration and Research Support

Head
Dr. Manuela B. Urban, MBA
Phone: +49 (0)30-8413-1360
Fax: +49 (0)30-8413-1394
E-mail: urban@molgen.mpg.de

Secretary
Jeannine Moisel
Phone: +49 (0)30-8413-1399
Fax: +49 (0)30-8413-1394
E-mail: moisel@molgen.mpg.de

Personnel Department
Ruth Schäfer (Department Head)
Hannelore Feiks
Jeannette Bertone (part-time)
Jeanette Brylla
Hilke Wegwerth

Accounting Department
Angelika Brehmer (Department Head)
Petra Saporito
Malgorzata Klemm
Ursula Schulz (part-time)

External Project Funding
Anke Badrow
Joachim Gerlach

Guest Houses, Apartments
Rosemarie Wolniak (part-time)
Helena Netzer (part-time)
Sophie Hessel (part-time)

Purchasing Department
Jutta Roll (Department Head)
Karsten Krause
Ute Müller
Rita Röfke-Bohnau
Kerstin Steudtner

Stock Room
Jürgen Joch
Olaf Kischkat
Dirk Grönboldt-Santana

Reception, Post Office
Sebastian Klein (part-time)
Monika Schweizer-Annecke (part-time)

Driver
Claus Langrock

Apprentice
Kristina Bönisch

During the last years there have been several legal changes which are of importance for the institute. On October 1st, 2005, a new collective labour agreement for the federal public service (Tarifvertrag für den Öffentlichen Dienst, TVöD) became effective which the Max Planck Society is bound to. With this new agreement, cooperation and exchange of personnel with universities and other research institutions has become extremely difficult, since the salary scales may differ significantly depending on the individual situation. For instance, the salary for PhD students with family is shortened by about 10 %. Altogether, the employment conditions for new personnel have become less attractive. On the other hand, the trend to conclude individual bargaining agreements for the "Länder" (state) or individual universities is still going on.

In 2004, the German Federal Constitutional Court declared portions of the "Hochschulrahmengesetz", the law providing guidelines for German universities, as unconstitutional. As a result, for several months there has been no legal framework for employing scientists. The reforms of the "Hochschulrahmengesetz" (HRG) mean significant changes for the career perspectives of junior scientists.

An exceptionally important challenge in the last two years has been the cessation of the Nationales Genomforschungnetz 1 (NGFN-1, German national genome research network) and the Deutsches Humangenomprojekt (DHGP, German humane genome project). Since the NGFN-2 started with a delay of several months, the institute was forced to bridge the gap in order to retain personnel and continue work. This was made possible only by a supplementary allowance from the Max Planck Society.

A very positive recent development concerning the administrative framework is the increasingly flexible budgetary rules of the Max Planck Society. Now the institutional funds are allocated as a global budget, i.e. the funds may be spent totally flexible. The categories personnel, consumables, fellowships and equipment are used now only for planning procedures. The flexible rules enable the institute to rededicate considerable parts of the consumables in order to complement the short personnel and equipment budgets.

In 2004 the institute implemented a cost accounting system which now runs in a refined form. It now serves to improve budgetary control. However, since the planning procedures of the Max Planck Society are still cameralistic, the main focus in budgetary control is on expenditures and liquidity, not on costs.

The new personnel administration system (SAP R/3 HR), which was implemented in 2003, still needs further improvement in order to really facilitate work. It turned out that mapping the different employment conditions, salary and stipend scales as well as requirements resulting from externally funded projects is quite complex.

Recently, the storeroom has been reorganized so that goods can be ordered electronically from any work place in the lab or office, and the goods are delivered twice a day directly to the lab. The storeroom records are now connected to the integrated accounting system. Furthermore, the institute is considering taking part in the electronic procurement system (e-procurement) of the Max Planck Society.

In the last several years the institute has intensified its efforts in vocational education. Starting with one animal keeper in 1999, a total of 14 trainees have been educated at the institute as animal keepers, office clerks, laboratory assistants and IT specialists for system integration.

Technical Management and Workshops

Head
Dipl.-Ing. (FH) Ulf Bornemann
Phone: +49 (0)30-8413-1424
Fax: +49 (0)30-8413-1394
E-mail: bornemann@vw.molgen.mpg.de

Staff
Sebastian Klein, Secretary (part-time)

Electrical Engineering
Frank Michaelis (Department Head)
Klaus Krüger (retirement 11/2006)
Udo Abratis
Lars Radloff
Bernd Roßdeutscher
Bernd Zabka

Building and Technical Services
Reinhardt Strüver (Department Head)
Frank Kalaß
Manfred Lemke
Thomas Oster
Bernd Roehl

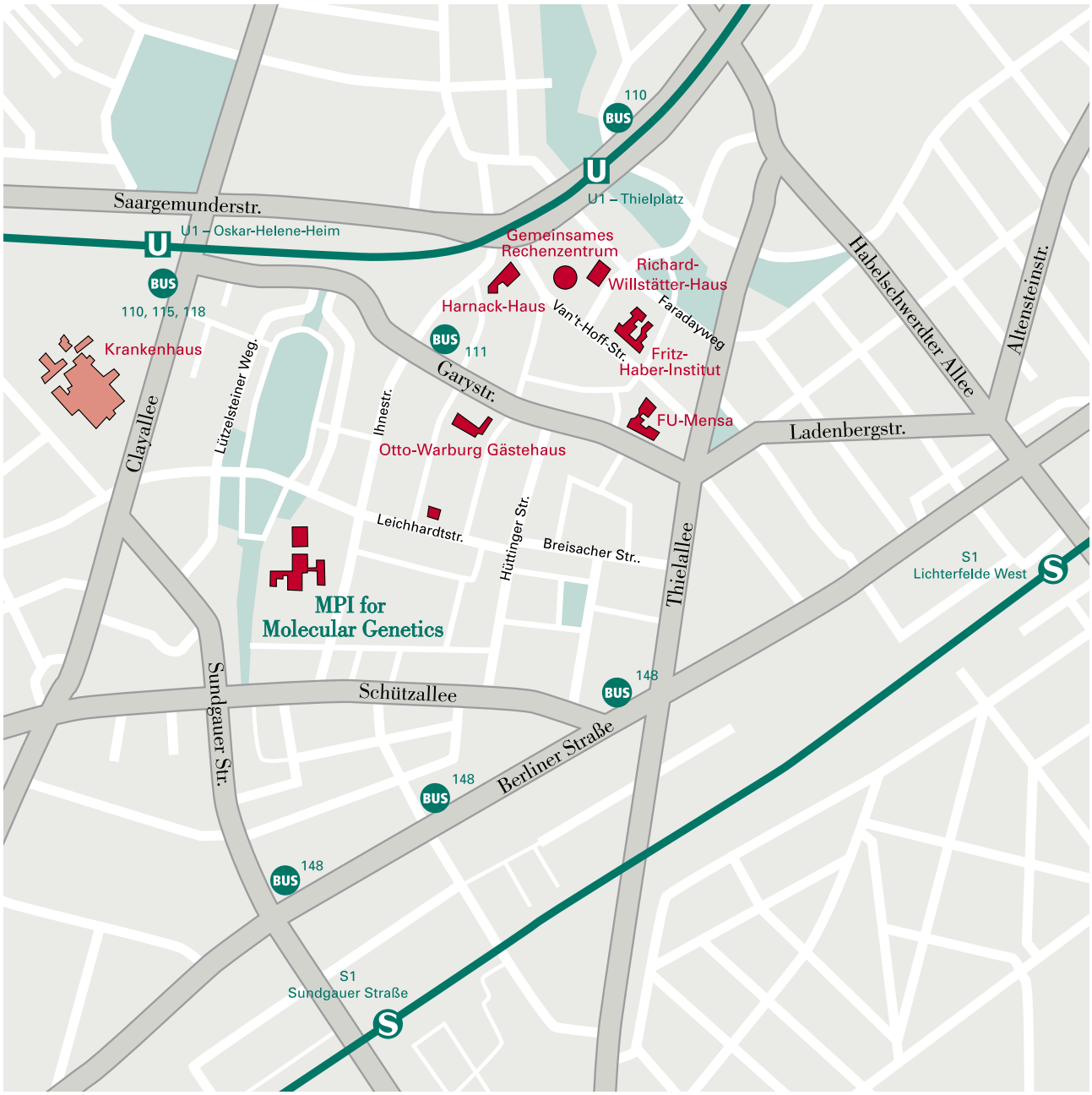
Electrical Mechanics
Detlef Becker (Department Head)
Carsten Arold
Gisela Bosse
Florian Zill

Glass Instruments Construction
Peter Ostendorf (part-time)

In 2003 the institute received the "go ahead" from the President of the Max Planck Society to build Tower 3 and to completely renovate Towers 1 and 2, built originally in 1968 and 1970 respectively. The necessary planning funds were approved in early 2006. By August of 2006 the planning stage should be completed after which the building application will be filed. Building on Tower 3 is predicted to begin in 2007.
To keep the effects on research operations to a minimum, an interval between the construction of Tower 3 and the renovation of Towers 1 and 2 will be taken. First Tower 3 will be constructed and then the two older towers will be renovated one after another.

Tower 3 will be used by the Vingron Department, the theoretical work groups of the other departments, and Junior Research Group Lappe, and the IT Service group. It will be the central entry foyer for the institute and the designated seminar rooms on the ground floor will be used for events. In addition a new server room will be built in Tower 3. The construction work is estimated to last from 2007 to 2009. Then the renovation of Tower 2, followed by Tower 1 will be carried out. The renovation includes the updating of the technical infrastructure such as air ventilation, electrical installations, waterworks and laboratory equipment, as well as the improvement of the structural state of Towers 1 and 2. This includes changing the size and layout of the rooms. Further measures include replacing the floor covering on all floors, renewing the façade, as well as taking measures to save energy.
Because the work can't take place while research is being conducted, it will be necessary to move into the freed up space in other towers and in our satellite laboratories in Fabeck Street.

How to get to the Institute



Max Planck Institute for Molecular Genetics
Ihnestr. 63–73
14195 Berlin, Germany
Phone: +49 (0)30-8413 - 0
Fax: +49 (0)30-8413 - 1207
Email: info@molgen.mpg.de

www.molgen.mpg.de