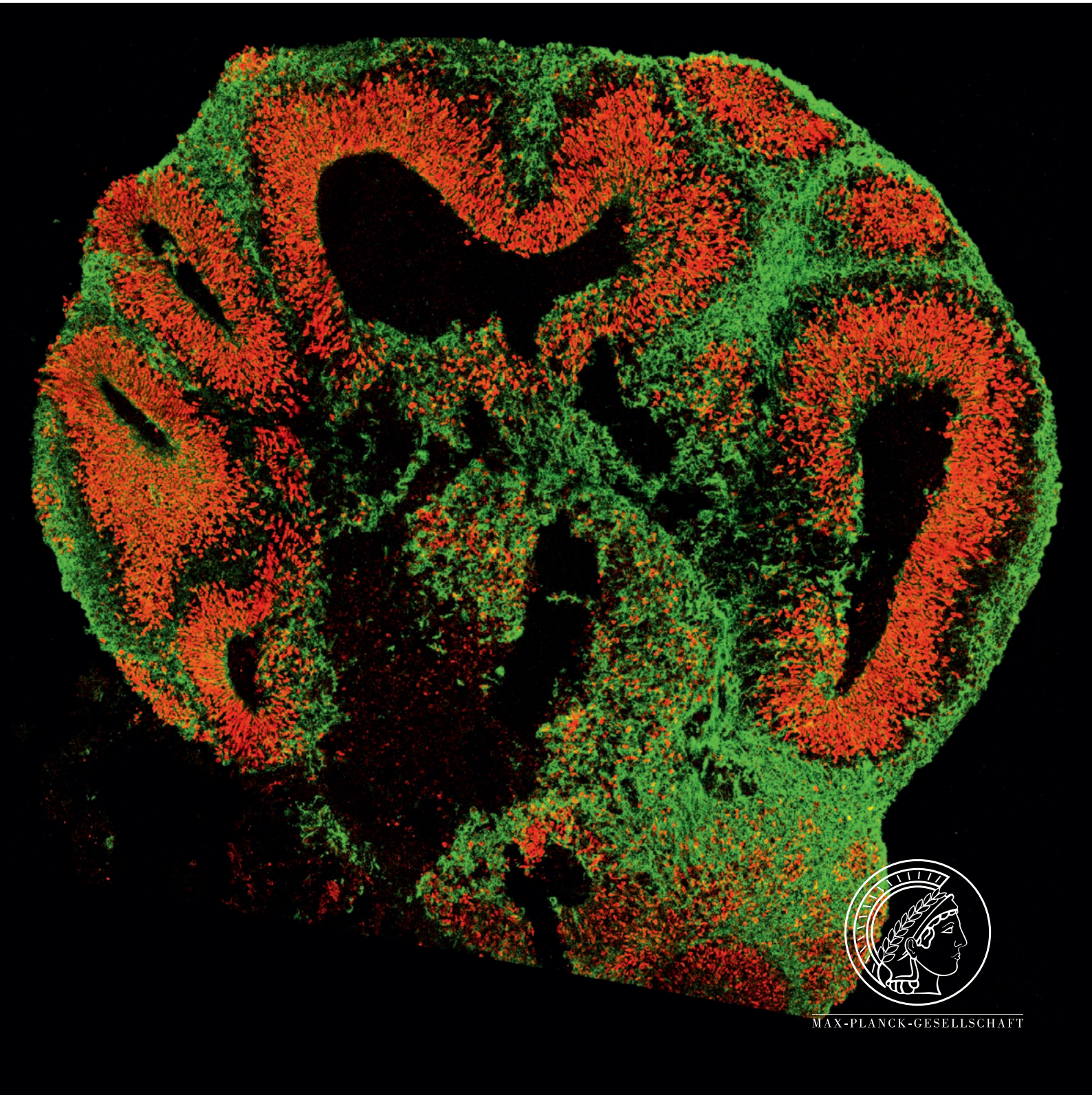
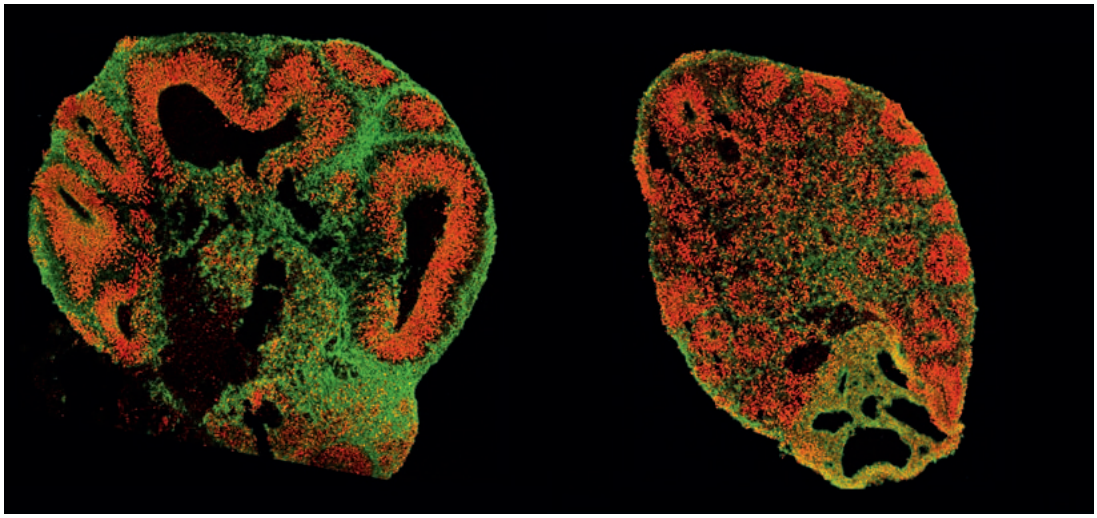
RESEARCH REPORT 2018

Max Planck Institute for Molecular Genetics



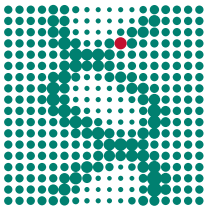
MAX-PLANCK-GESELLSCHAFT



Impairment of cerebral organoid development induced by microRNA overexpression

Shown are day 30 organoids following microRNA overexpression (**left**), and their matched day 30 untreated organoids (**right**), stained for the neural stem/progenitor cell marker SOX2 (red) and the neuronal differentiation marker DCX (green). The development of cerebral organoid cytoarchitecture can be clearly seen in control organoids, while it appears stalled following microRNA overexpression. The Elkabetz group is studying the mechanisms by which microRNAs govern cerebral development.

Picture: Elkabetz Lab / Human Brain and Neural Stem Cell Studies



RESEARCH REPORT 2018

Max Planck Institute for Molecular Genetics

MAX-PLANCK-GESELLSCHAFT

Michael Müller, Governing Mayor of Berlin, at a visit at the MPIMG during the Long Night of Sciences 2017 (from left to right: Hans Lehrach, Martin Vingron, Michael Müller, Bernhard Herrmann, Alexander Meissner)

Editorial

We are very pleased to present the 2018 Research Report of the Max Planck Institute for Molecular Genetics (MPIMG), covering the time period from 2015 to mid-2018. The Institute looks back on three challenging and productive years.

The first important step in the transition and scientific reorientation following the retirement of H.-Hilger Ropers and Hans Lehrach in 2014 has been completed successfully by the appointment of Alexander Meissner to director of the new department "Genome Regulation". This recruitment marks the beginning of a new important track of research at the MPIMG with a strong emphasis on gene regulatory mechanisms in stem cell differentiation, embryogenesis and human disease. In addition, the refurbishment of tower 2 has been finalized successfully, providing two departments and a large research group with state-of-the art laboratories. We are also looking forward to the refurbishment of tower 1, which after a long preparation phase is finally scheduled to start this year.

We would like to take the occasion to express our gratitude to the Max Planck Society and the many funding agencies that support our work, for providing scientific freedom and independence to pursue the research we burn for and to tackle the most interesting scientific questions. We are confident that the scientific results presented in this report have made a significant contribution to science.

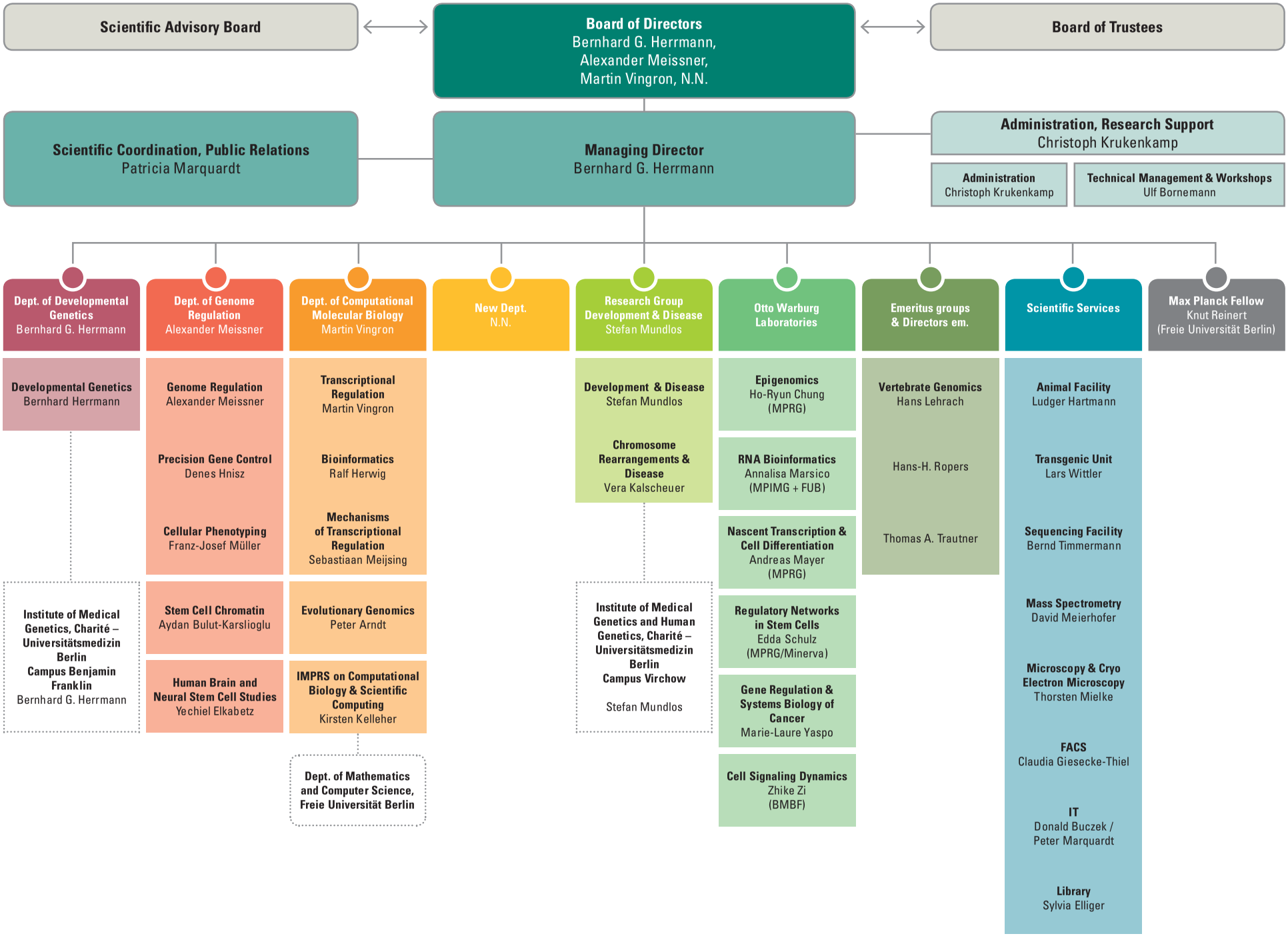
Compared to earlier reports, this year we have strived to produce a shorter, more concise report. Compact information on our research projects and research results can be found in the reports of the individual departments and groups. In addition, the introduction summarizes the main development of the MPIMG and gives a broader outline of the Institute. We hope that the Research Report 2018 will provide a clear impression of the Institute and the work we are doing.

Berlin, October 2018

Bernhard G. Herrmann, Alexander Meissner, Martin Vingron

Table of contents

Table of contents


Scientific Advisory Board
Board of Directors
Bernhard G. Herrmann,
Alexander Meissner,
Martin Vingron, N.N.
Board of Trustees
Scientific Coordination, Public Relations
Patricia Marquardt
Managing Director
Bernhard G. Herrmann
Administration, Research Support
Christoph Krukenkamp
Administration
Christoph Krukenkamp
Technical Management & Workshops
Ulf Bornemann
Dept. of Developmental Genetics
Bernhard G. Herrmann
Dept. of Genome Regulation
Alexander Meissner
Dept. of Computational Molecular Biology
Martin Vingron
New Dept.
N.N.
Research Group Development & Disease
Stefan Mundlos
Otto Warburg Laboratories
Emeritus groups & Directors em.
Scientific Services
Max Planck Fellow
Knut Reinert
(Freie Universität Berlin)
Developmental Genetics
Bernhard Herrmann
Institute of Medical Genetics, Charité – Universitätsmedizin Berlin
Campus Benjamin Franklin
Bernhard G. Herrmann
Genome Regulation
Alexander Meissner
Precision Gene Control
Denes Hnisz
Cellular Phenotyping
Franz-Josef Müller
Stem Cell Chromatin
Aydan Bulut-Karslioglu
Human Brain and Neural Stem Cell Studies
Yechiel Elkabetz
Transcriptional Regulation
Martin Vingron
Bioinformatics
Ralf Herwig
Mechanisms of Transcriptional Regulation
Sebastiaan Meijsing
Evolutionary Genomics
Peter Arndt
IMPRS on Computational Biology & Scientific Computing
Kirsten Kelleher
Dept. of Mathematics and Computer Science, Freie Universität Berlin
Development & Disease
Stefan Mundlos
Chromosome Rearrangements & Disease
Vera Kalscheuer
Institute of Medical Genetics and Human Genetics, Charité – Universitätsmedizin Berlin
Campus Virchow
Stefan Mundlos
Epigenomics
Ho-Ryun Chung
(MPRG)
RNA Bioinformatics
Annalisa Marsico
(MPIMG + FUB)
Nascent Transcription & Cell Differentiation
Andreas Mayer
(MPRG)
Regulatory Networks in Stem Cells
Edda Schulz
(MPRG/Minerva)
Gene Regulation & Systems Biology of Cancer
Marie-Laure Yaspo
Cell Signaling Dynamics
Zhike Zi
(BMBF)
Vertebrate Genomics
Hans Lehrach
Hans-H. Ropers
Thomas A. Trautner
Animal Facility
Ludger Hartmann
Transgenic Unit
Lars Wittler
Sequencing Facility
Bernd Timmermann
Mass Spectrometry
David Meierhofer
Microscopy & Cryo Electron Microscopy
Thorsten Mielke
FACS
Claudia Giesecke-Thiel
IT
Donald Buczek /
Peter Marquardt
Library
Sylvia Elliger

The Max Planck Institute for Molecular Genetics

Structure and organization of the Institute

Since 2015, the Max Planck Institute for Molecular Genetics (MPIMG) has undergone a major reorganization and scientific reorientation. Along with the retirement of two directors, H.-Hilger Ropers and Hans Lehrach, the Institute has faced a substantial reduction in personnel, paralleled by a decrease of its scientific output. The search for outstanding scientific leaders heading the vacant departments turned out to be more difficult and time consuming than anticipated, and has kept the Institute in a state of uncertainty for quite some time. However, this difficult phase has been overcome in the meantime, due to the recruitment of Alexander Meissner, formerly full professor at Harvard University and Broad Institute, Cambridge, MA. He has been appointed to Director and Head of the Department of Genome Regulation in 2016, and started in secondary employment the same year. In July 2017, he took up work at the MPIMG in principal employment and moved to Berlin with his family.

Alexander Meissner's research concept perfectly matches the scientific concept of the MPIMG, which focuses on investigating gene regulation mechanisms and networks in embryonic development, tissue differentiation, homeostasis, and disease. The new director has set up his department at a stunning pace. He managed to build up his own group and hired the heads of four additional independent research groups within a few months. Together, the new department contributes strong expertise in mouse and human stem cell models, in epigenomics, and state-of-the-art technology to the Institute. The recent year has already demonstrated the positive impact of the new department for scientific interactions and has brought a plethora of new ideas, which influences the work and scientific thinking of the other groups at the MPIMG.

Figure 1.
Martin Vingron, Alexander Meissner and Bernhard Herrmann (from left to right), the current directors of the MPIMG.

However, the transition in scientific leadership at the Institute could not be completed yet. The fourth department unfortunately is still vacant, and Max Planck Society (MPG) as well as the Institute are working hard on recruiting another outstanding scientist as director at the MPIMG.

With respect to lab and office space the Institute has improved substantially. Tower 2 has been refurbished, and in January 2017 has been occupied by groups moving in from tower 1. However, tower 1 is still awaiting refurbishment (see below). Unfortunately, lacking tower 1 for much longer than anticipated currently severely hampers the recruitment of new groups and the necessary growth in personnel, which is vital for reaching critical mass. We very much hope that this constriction of the MPIMG will be overcome before the next SAB meeting.

In autumn 2018, the MPIMG consists of three active departments, headed by Bernhard Herrmann (Dept. of Developmental Genetics), Alexander Meissner (Dept. of Genome Regulation) and Martin Vingron (Dept. of Computational Molecular Biology). The departments are complemented by the research group Development & Disease headed by Stefan Mundlos, who also holds the Chair of Medical Genetics and Human Genetics at the Charité – Universitätsmedizin Berlin. The MPIMG was very pleased that MPG has appointed Stefan Mundlos as External Scientific Member of the MPIMG in 2017. In that way, MPG has acknowledged the outstanding scientific achievements of Stefan Mundlos in recent years, encouraging him and the MPIMG to continue the long-standing and very fruitful cooperation between the Charité and the Institute.

The core structure of the MPIMG comprises a number of independent research groups, which are not integrated into departments, known as the "Otto Warburg Laboratories" (OWL). The group leaders receive funding from various sources and agencies, e.g., from Max Planck Society (MPG), the German Ministry of Education and Research (BMBF), the German Research Foundation (DFG), or the Alexander-von-Humboldt Foundation, and may obtain additional finances through grants. They have full responsibility for their research and publications and manage their budgets independently. Their appointment to the MPIMG usually is temporary (from five up to nine years).

The current group leaders of the OWL are Annalisa Marsico (RNA Bioinformatics), Andreas Mayer (Nascent Transcription & Cell Differentiation), Edda Schulz (Regulatory Networks in Stem Cells), Marie-Laure Yaspo (Gene Regulation & Systems Biology of Cancer), and Zhike Zi (Cell Signaling Dynamics).

In 2017, the former OWL group leader Ulf Ørom (Long non-coding RNAs) has accepted an appointment to associate professor at the University of Aarhus, Denmark. He left the MPIMG in November 2017, while keeping his lab over a transition period for finalizing some projects until summer 2018. In summer 2018, Ho-Ryun Chung accepted the appointment to a W3 professorship at the University of Marburg. His lab will be closed in November 2018. In December 2018, Tugce Aktas, now in the lab of Asifa Akhtar, MPI of Immunobiology and Epigenetics, Freiburg, will join the OWL to build up a new group working on "Quantitative RNA Biology".

In 2014, the Max Planck Society appointed Knut Reinert, Professor of Algorithmic Bioinformatics at Freie Universität (FU) Berlin, as Max Planck Fellow, following the proposal from the Institute. Reinert's research focuses on providing efficient tools for the analysis of genomic data; he cooperates with many groups at the MPIMG already since 2002.

In addition, the scientific groups of the MPIMG are supported by a number of scientific service groups that maintain a range of core technologies for all departments and research groups. In order to constantly keep up with technological advances and changing needs

of the MPIMG scientists, a new service group "FACS" dedicated to flow cytometry has been established in September 2018. For details about the service groups as well as about the general administration and the workshops of the MPIMG, please see the reports of the individual groups.

Research concept of the Institute

In the 1990s, the MPIMG was scientifically oriented towards human genetics and genome research. With this focus, it has played an important role in deciphering the human genome sequence and the genetic cause of many human diseases. This very successful phase has established the MPIMG as an internationally recognized institute for genome research, but with the retirement of the directors Hans Lehrach and H.-Hilger Ropers, its scientific concept had to be renewed. Based on its international reputation, the MPIMG has received support from MPG to continue working in the area of genomics/epigenomics. The board of directors decided to focus on the interface between genetics and genome research, however with a strong emphasis on gene regulatory processes involved in cell differentiation, embryogenesis, and the etiology of disease. The concept led to the recruitment of Alexander Meissner, whose research is focussing on epigenetic regulation mechanisms in stem cell differentiation, early embryogenesis, and in disease.

Genome regulation has many facets comprising for instance the epigenetic modifications effecting the activation or silencing of genes, 3-D organization of chromosomal subdomains, the control of transcription initiation or the splicing of transcripts, to name just a few important aspects. Many of these are topics of interest at the MPIMG, in normal development as well as with respect to the etiology of malformations and disease in humans.

The research concept of the Institute opens enormous opportunities for biomedical research on human disease. For instance, induced pluripotent stem cells (iPSC) can be generated from skin cells derived from human patients - a method that has been pioneered by Alexander Meissner – and be differentiated into almost any tissue in the culture dish. Thus iPSCs can be utilized to investigate the etiology of disease at the genomic and epigenomic level. The method can be utilized e.g. for investigating the contribution of genetic variants on the etiology of disease of individual patients. Understanding complex disease and the role of genetic variation in the etiology of disease can only succeed via genomics and epigenomics research, combined with strong bioinformatics. Therefore, the MPIMG is very well positioned for becoming a leading research centre for the investigation of cell differentiation, mammalian embryogenesis, and (complex) disease.

Scientific highlights

The main results of the scientific work of the MPIMG since 2015 are described in detail in the research reports of the individual departments and groups. Here, we present some of the most important and interesting results to give a general impression of the research performed at the Institute.

Alexander Meissner published a base pair resolution analysis of global remethylation from the hypomethylated state of the preimplantation embryo into the early epiblast and extraembryonic ectoderm. He could show that these two states acquire highly divergent genomic distributions and, in addition, that most cancer types share an epigenetic landscape that closely mirrors the early extraembryonic patterns, suggesting a possible developmental origin of the disease. The project was completed during Meissner's transition to the MPIMG and provides the foundation of numerous ongoing efforts in the current lab [Smith et al. Nature 2017]. Already with his Berlin team, Meissner developed a methylation depleted but maintenance competent mouse ES cell line (DKO^{zero} line) as a unique tool to track all dCas9-methyltransferase activity that can be cleanly separated from any endogenous *de novo*

methylation by Dnmt3a and Dnmt3b. Therefore, Meissner and his lab provided a first true account of the unexpected genome-wide DNA methylation footprints left by dCas9-coupled to the catalytic domain of Dnmt3a [Galonska et al. Nat Commun 2018].

Bernhard Herrmann and his team presented a global analysis of the development of neural and mesodermal tissues of the trunk from neuro-mesodermal progenitor cells (NMPs). They showed that the transcription factors Sox2 and Brachyury are in a balanced state in the NMPs, but antagonize each other in differentiating cells undergoing lineage choice prior to committing to the neural or mesodermal cell fate. They also showed that Brachyury is required for shaping the chromatin for the mesodermal lineage and acts upstream of the transcription factors Tbx6 and Msgn1 in forming paraxial mesoderm, the progenitors of the vertebral column and skeletal muscle. Their findings refine previous models and establish molecular principles underlying mammalian trunk development, comprising NMP maintenance, lineage choice, and mesoderm formation, thus providing a basis for the elucidation of disease processes and malformations resulting from dysfunction of these control mechanisms (Koch et al. Dev Cell 2017).

Stefan Mundlos and his team have been very successful in investigating the effect of large scale rearrangements, so called structural variants (SVs), on 3D genome folding, gene regulation and gene expression. Their studies are based on the seminal discovery of topologically associated domains (TADs) that were shown to be regions of high contact, separated by boundaries with low contact, thereby confining the region an enhancer can interact with. Based on this they developed a framework for the interpretation of SVs in regard to their pathogenic potential. They interpret SVs in regard to their effect on TAD configuration potential taking into account their effect on 3D genome architecture and gene regulation. Deletions and inversions, for example, can lead to the fusion of TADs thereby connecting two regulatory regions that were previously separated. Large duplications that include boundary elements, in contrast, can result in the formation of new chromatin domains with own regulatory domains. Increasing the dosage of enhancers in TADs can lead to gene mis- and overexpression. With these studies Stefan and his team describe a new mutational pathomechanism that is highly relevant for the interpretation of SVs in congenital disease and cancer (Lupianez et al. Cell 2015; Spielmann et al. Nat Rev Genet 2018).

Parts of this work, especially focusing on 3D chromatin structure, have been done in a close collaboration with the group of Martin Vingron. He and his team developed tools and provided data analysis support for analysing 3D chromatin structure data with respect to implications for gene regulation. The emphasis lies on enhancer-promoter interactions, which are affected due to genomic rearrangements with subsequent changes in chromatin interactions (Andrey et al. Genome Res 2017; Franke et al. Nature 2016).

Martin Vingron and his team, in collaboration with a Chinese plant biology group have sequenced a hexaploid plant, the sweet potato *Ipomoea batatas*. It was generally assumed that phasing such a complicated genome would be impossible, but based on Illumina sequencing, Vingron and his team managed to develop the necessary algorithms for phasing and succeeded in assigning almost half of the genome to haplotypes (Yang et al. Nature Plants 2017).

Annalisa Marsico and Ulf Ørom, both from the OWL, continued their very fruitful cooperation on long non-coding RNAs. After their analysis of the activity of the microprocessor complex during miRNA biogenesis (Conrad et al. Cell Rep 2014), they now proposed that the release of long ncRNAs from chromatin is a crucial functional aspect in transcription regulation of long ncRNA target genes. They functionally validated the long ncRNA A-ROD, which enhances *DKK1* transcription via its nascent spliced released form. Their data provide evidence that the regulatory interaction requires dissociation of A-ROD from chromatin, with target specificity ensured within the pre-established chromosomal prox-

imity. Thus, they propose that the post-transcriptional release of a subset of long ncRNAs from the chromatin-associated template plays an important role in their function as transcription regulators (Ntini et al. Nat Commun 2018).

Marie-Laure Yaspo, also from the OWL, continued her successful work deciphering mechanisms underlying tumorigenesis and drug sensitivity in several cancer types by deep omics profiling and imaging mass cytometry. After generating large data resources in ICGC and IMI projects, her group identified novel sensitivity markers to EGFR inhibitors in colon cancer (Schütte et al. Nat Commun 2017), targetable pathways in the fatal TCF3-HLF acute lymphoblastic leukemia (Fischer et al. Nat Genet 2015), and specific cell trajectories in early onset prostate cancer (Gerhauser et al. Cancer Cell, 2018, in press). She coordinates a precision oncology trial on refractory melanoma.

Altogether, MPIMG scientists have authored 361 peer reviewed publications between 2015 and autumn 2018. About 20% of these had contributions from more than one department or group, thus demonstrating the continuous cooperation between the individual MPIMG groups (figure 2). Nearly 50% of all publications have main contributions with at least one first or last author from the MPIMG, and about 30% were published in journals with impact factors >10 (see figures 3 and 4).

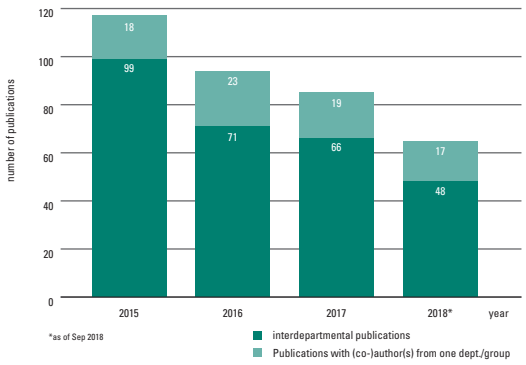


Figure 2.
Number of peer-reviewed publications from one department/group and with participation of at least two departments or groups (including service groups) of the MPIMG

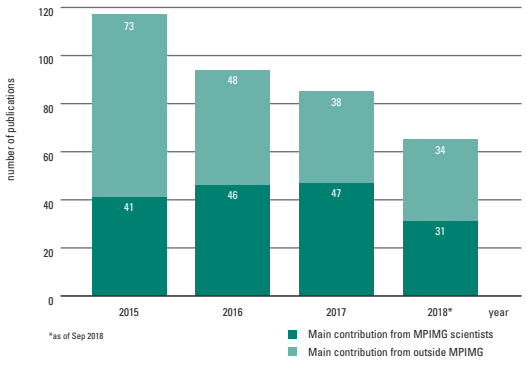


Figure 3.
Number of peer-reviewed publications with main contributions (at least first or last author) from MPIMG scientists

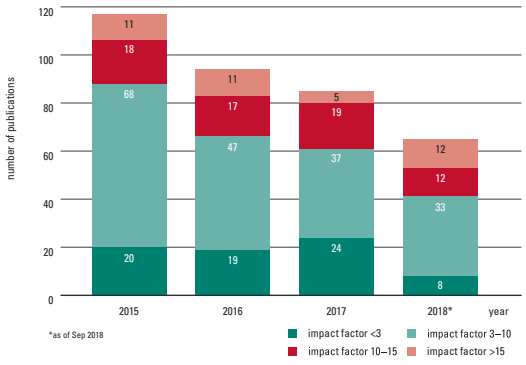


Figure 4.
Number of peer-reviewed publications sorted by impact factors

Scientific awards

In appreciation of their scientific achievements, members of the MPIMG have been awarded with several prizes. Amongst others,

- Aydan Bulut Karslioglu received a Sofja Kovalevskaja Award 2018 from the Alexander von Humboldt Foundation to build up her own research group at the MPIMG;
- Annita Louloupi has been invited to attend the 68th Lindau Nobel Laureate Meeting in 2018;
- Alexander Meissner has been elected as member of the European Molecular Biology Organization (EMBO) in 2018;
- Karin Moelling, retired member and visiting scholar of the MPIMG, has received the Officer's Cross (Verdienstkreuz 1. Klasse) of the Order of Merit of the Federal Republic of Germany in 2018;
- Olga Jasnovidova obtained a Long-Term Fellowship of the Federation of European Biochemical Societies (FEBS long-term postdoctoral fellowship, 2018);
- Edda Schulz received a HFSP Career Development Award from the Human Frontier Science Program in 2018;
- Jesse Veenvliet has been awarded for the best talk at the Berlin Postdoc Day in 2018;
- Anja Will obtained the PhD Award in the class "Natural Sciences" from the German Society for Human Genetics in 2018. In addition, she won a ESHG Young Investigator Award for Outstanding Science from the European Society of Human Genetics in 2017;
- Stefanie Schöne obtained the Elisabeth-Gateff-Preis 2017 from the German Genetics Society for her doctoral thesis. She also was awarded with a fellowship of the Christiane Nüsslein-Volhard Foundation and the L'Oreal-UNESCO for Women in Science Program in 2015;
- Stefan Mundlos has been elected as a member of the European Molecular Biology Organization (EMBO), in 2017. In addition, has been honored with the ESHG Award 2016 from the European Society of Human Genetics (ESHG);
- Magdalena Socha won a ESHG Young Investigator Award for Outstanding Science from the European Society of Human Genetics in 2017;
- Martin Mensah has been honored with an ESHG Young Investigator Award 2016 from the European Society of Human Genetics;
- Juliane Perner received the first BIH Award for Early-Career Women Scientists in Bioinformatics from the Berlin Institute of Health in 2016;
- Peter N. Robinson received the EURORDIS Scientific Award of the European Organization for Rare Diseases in 2016;
- Katerina Kraft has been honored with the Peter Hans Hofschneider Prize for Molecular Medicine of the Max Planck Society for the Advancement of Science in 2015;
- Dario Lupianez received an ESHG Young Scientist Award from the European Society of Human Genetics, as well as the best lecture award from the Deutsche Gesellschaft für Humangenetik [German Society of Human Genetics] in 2015;
- Bruno Pereira won a Young Researcher Poster Price at the EMBO Workshop: Embryonic-Extraembryonic Interfaces in Göttingen in 2015.

Appointments of former members of the MPIMG

Several people have left the MPIMG and took up positions in other institutions or universities in Germany and abroad. The most important appointments have been (ordered by year)

- Ho-Ryun Chung: W3 Professor for Medical Bioinformatics and Biostatics, Director of the Institute for Medical Bioinformatics and Biostatistics, Philipps-University of Marburg (2018);
- Darío Garcia Lupiañez: Head of Independent Junior Group "Epigenetics & Sex Development" at the Berlin Institute for Medical Systems Biology (BIMSB) – Max Delbrück Center for Molecular Medicine (MDC) Berlin-Buch (2017);
- Ulf Andersson Vang Ørom: Assistant Professor at Aarhus University, Department of Molecular Biology and Genetics, Denmark (2017);
- Sascha Sauer: Head of Scientific Genomics Platform and Laboratory of Functional Genomics, Nutrigenomics and Systems Biology at Max Delbrück Center for Molecular Medicine (MDC) Berlin-Buch (2016);
- Smita Sudheer: Assistant Professor at Central University of Kerala, Department of Genomic Science, India (2016);
- Hao Hu: Professor of Genetics at Zhongshan School of Medicine, Sun Yat-sen University, and Director of the Department of Molecular Diagnostics, Guangzhou Women and Children's Medical Center, Guangzhou, China (2015);
- Matthias Heinig: Junior Group Leader Genetic and Epigenetic Gene Regulation at the Institute of Computational Biology, Helmholtz Zentrum München, German Research Center for Environmental Research (2015);
- Ulrich Stelzl: (Full) Professorship on Biopharmaceutica und Proteomics at the University of Graz, Austria (2015).

Recruitment of new group leaders

A number of new group leaders have joined the Otto Warburg Laboratories, thus bringing not only new research topics, but also new ideas and impetus to the Institute.

In January 2015, Edda Schulz established a Max Planck Research Group on *Regulatory Networks in Stem Cells*. She and her team are working on the elucidation of regulatory principles of molecular networks that control the transition of embryonic stem cells to further differentiated cell types.

In January 2017, Andreas Mayer started to build up his Max Planck Research Group *Nascent transcription & Cell Differentiation*. His primary goal is to reveal the key mechanisms that underlie the regulation of chromatin-mediated nascent RNA polymerase II transcription, including non-coding transcription, in differentiated mammalian cells and during cell differentiation.

In December 2018, Tugce Aktas will start to build up her Max Planck Research Group *Quantitative RNA Biology*. The new group will work on several aspects of transposon-host interactions and their impact on the evolution and wiring of post-transcriptional RNA processing networks.

In addition, Claudia Giesecke-Thiel has been recruited as head of the *FACS service group*. The new group has started operation in September 2018.

Support of junior scientists at the MPIMG

In March 2018, 70 students pursued their PhD studies at the Institute. All students are part of an institute-wide education program (PhD program) that has been established in 2008. It encourages the exchange of knowledge and the development of skills throughout the disciplines. For one, this is achieved by intense interdisciplinary courses held in turn by the different departments, the so called PhD week. It offers insight and hands-on experience in fields that go beyond the scope of the respective PhD theses. Since 2008, twenty such courses have provided supplementary experience especially to junior students in their first and second year, and have initiated new cooperation and fresh perspectives for many PhD projects. In October 2018, the PhD week was organized by the new Department of Genome Regulation for the first time.

Another central aspect of the PhD program is an annual PhD retreat, organized by the Student Association (STA). This self-organized student representation organ was founded in 2001 to act for the interests of all students at the MPIMG by addressing the directors, group leaders, or the works council. Each department and the OWL groups are represented by a student in the STA steering committee. In recent years, the STA has grown in strength and impact and provides a valuable voice for junior scientists at the Institute. Since 2017, a STA representative is also invited as a regular guest to the extended meetings of the board of directors. The STA also provides a platform for organizing social and networking events to foster interdepartmental scientific discussion and exchange of technical and scientific knowledge between the students.

In addition to the PhD program, the implementation of a Thesis Advisory Committee (TAC) for each PhD student has been strongly recommended by the directors of the MPIMG. Each TAC consists of three scientists, the main supervisor at the MPIMG together with two other scientists at the Institute or another institution. Together they should cover all disciplines that the PhD research touches. The TAC meets once a year and guides the student in a scientific as well as in a practical way. It also provides an opportunity for the PhD student to discuss potential problems with the main supervisor with the two other TAC members.

Most PhD students working in the field of bioinformatics pursue their thesis work in the context of the International Max Planck Research School for Computational Biology and Scientific Computing (IMPRS-CBSC). This graduate program was established in 2005 by Martin Vingron and Knut Reinert, Faculty of Mathematics and Informatics at Freie Universität Berlin (FUB) and also Max Planck Fellow at the MPIMG since 2014. Its strength is to capitalize on the synergies of MPIMG and university in the area of biology-related computer science. Currently, eight group leaders from MPIMG, eight from FUB, and three from associated institutes are members of the IMPRS-CBSC faculty. To date, the IMPRS-CBSC has had 90 students, 49 of which have already successfully submitted and defended their thesis. The curriculum consists of scientific training and research, career support through education in transferable skills and tutoring. Excellent supervision is guaranteed through TACs and a supervision agreement that defines the cornerstones for training structure, funding and supervision quality. A PhD coordinator assists in issues of all PhD students of the Institute. The IMPRS-CBSC will reach its end-of-life at the end of 2018 and a new graduate school shall be established (pending a positive decision by the Max Planck Society) from 2019. This new school entitled International Max Planck Research School for Biology and Computation (IMPRS-BAC) is intended to bridge the gap between natural and formal sciences. In accordance, the faculty will grow to include not only bioinformaticians and computer scientists from MPIMG and FUB, but also biochemists and biologists from both institutions. All students of the MPIMG will have the chance to become part of the IMPRS-BAC and future students will be recruited through a structured process. This will unite existing structures to make cooperation, scientific and transferable training even more efficient and fruitful.

Figure 5.
Impression of the IMPRS-CBSC After School Special, the first alumni and career event of the IMPRS in September 2018

Equal opportunities

The MPIMG pays special attention to topics of gender equality and work-life balance. In December 2017, the Institute was one of the first in the Max Planck Society to give itself an equal opportunities policy and strives to reach the ambitious goals listed therein. These are divided into three categories: career development for female scientists, compatibility of work and family life, and promoting the visibility of gender equality issues.

In the area of career development the MPIMG promotes and participates in the initiatives of MPG. In addition, female scientists are strongly encouraged and supported when applying to career advancement or mentoring programs. An ongoing effort is made to increase the number of women in the areas of the MPIMG where they are underrepresented.

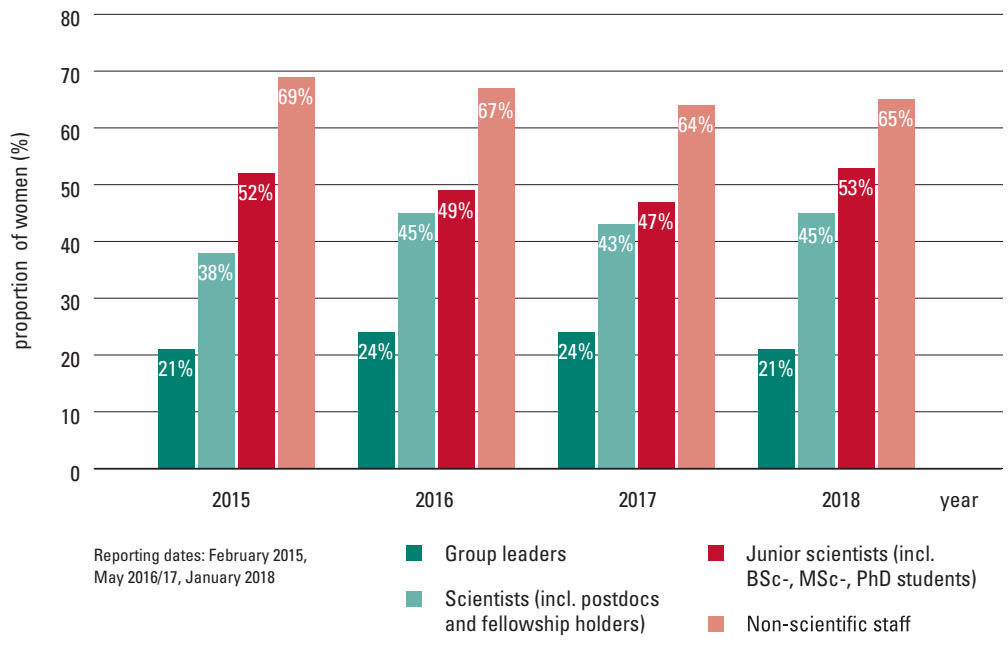


Figure 6.
Proportion of women on different career levels at the MPIMG

The MPIMG takes an active role in supporting the institute members who have children. The Institute has contracts with three nearby day-care facilities including a German-English and a German-French nursery to allow the admission of children on short notice. A childcare service is available that can be requested for seminars or events taking place at the Institute. To assist parents who need to take their children to work, the Institute offers a room dedicated to children that can be used by all MPIMG staff. Flexible working hours and teleworking is supported. According to the German law on temporary employment contracts in the scientific field (Wissenschaftszeitvertragsgesetz), temporary contracts during the so-called qualification period (Qualifizierungsphase) are extended automatically to make up for times of parental leave. Furthermore, the Institute is making efforts to offer this to externally funded employees, too, whenever possible.

In addition, the Max Planck Society closed a contract with the company "pme familienservice" [family service] that supports staff in finding a broad range of services from child- and elderly care to housekeeping to all Max Planck staff. Under the umbrella of the "pme Akademie", a wide range of seminars and further training events are also provided, most of which can be attended free-of-charge. In 2018, the Max Planck Society was awarded the certificate "Beruf und Familie" for the fifth time for its family-friendly human resources policy by the non-profit berufundfamilie Service gGMBH.

Finally, the MPIMG is raising awareness of the importance of gender equality issues amongst its members by a number of measures. The gender equality officer has been invited as a regular guest to the extended meetings of the board of directors and reports regularly at the works council meetings. There is one workshop offered per year that aims to offer solutions to problems in gender equality. Already since 2006, the Institute participates in the Girls' Day - Future prospects for Girls, a nation-wide event for girls aged ten and above to get insights to professions, in which women are underrepresented. Many employees engage by offering science and technology based projects for participating girls.

Material resources, equipment and spatial arrangements

A unifying feature for all MPIMG's research groups is the genomic approach to biology. The Institute houses a range of large-scale equipment like sequencing systems, mass spectrometers, a range of microscopes including wide field systems, confocal laser scanning microscopes, screening systems, and 3D-imaging systems, as well as some transmission electron microscopes, a cell analyser and cell sorter for flow cytometry, and a large IT infrastructure. Most of it is maintained by the Institute's service groups, who operate the equipment and provide high-level support for all in-house scientists and many external collaborators. Details can be found in the reports of the individual service groups.

A major challenge for the whole MPIMG during the last years has been the ongoing refurbishment of the old laboratories, built around 1970 and continuously used since then, into new, state-of-the-art laboratories. From October 2013 until October 2016, the structural and technical refurbishment of tower 2 took place. Details about it can be found in the report of the Technical Management / Workshops on the last pages of this report.

Figure 7.
View into one of the new laboratories in tower 2 before start of the experimental work.

The completely renewed tower 2 comprises three laboratory floors with labs of different size, equipment rooms, lab kitchens and offices for the departments of Bernhard Herrmann and Alexander Meissner and the research group of Stefan Mundlos. The ground floor contains infrastructural lab space for all scientific groups, including two labs for experiments according to biological safety level S2 (German Genetic Engineering Law), an area for working with radioactive nucleotides, a teaching lab for student courses, and a seminar room. Already in summer 2016, Alex Meissner began to set-up his new department on the second floor of tower 2. In winter 2016/2017, the department of Bernhard Herrmann as well as the research group of Stefan Mundlos moved to their new labs on the third and first floor of tower 2, respectively. This took place very smoothly and all groups could get on with their experimental work shortly after the move.

The next step will be the renovation of tower 1. Unfortunately, this had to be postponed initially after the handing over of tower 2, because MPG needed a new approval from the Joint Science Conference, the legal authority for major building measures of MPG, at first. By now, the approval is available and the dismantling of tower 1 will start in late autumn 2018. Tower 1 will contain further labs for the department Meissner as well as for two other scientific departments, and a specialized lab floor for the microscopy and sequencing service groups. In addition, the library as well as a staff canteen will be located there. We hope that the Institute will be able to move into tower 1 in 2021.

Public relations work

The MPIMG has continued its public relations activities to inform about its work and discuss the implications of modern genome research with the public. In the last years, special emphasis has been put on the communication about animal experiments. In 2016, the Max Planck Society published a White Paper – Animal Research in the Max Planck Society, summarizing the results of an internal discussion about animal research and the responsibility of each individual working with laboratory animals. The White Paper has been intensively discussed at the MPIMG and confirmed us in our ongoing engagement to communicate as openly and transparently as possible on this challenging topic. In this context, we publish information about animal experiments on our website and explain the use of laboratory animals in our press releases. Stefan Mundlos, whose group is working with animals intensively, has given numerous interviews for print, radio, and television and continues these activities. In addition, we present our work on public events like the Berlin-wide "Lange Nacht der Wissenschaften" (Long Night of Sciences) and discuss with the visitors about the need of laboratory animals for the type of research performed at the MPIMG. In 2018, in a joint effort with the press office of the Max Planck Society, a short film about our laboratory mouse work has been realized, which is now available on youtube.com ("Gene, die in die Knochen fahren") as well as on our own website. In order to support our staff, trainings in communication about animal experiments have been organized for scientists as well as for technicians and animal caretakers, which shall be repeated regularly.

The regular communication of the MPIMG with different target groups includes the publication of press releases about our work; the placement of scientific experts on topics of overall interest; a visiting program for school children and student groups to visit a lab, discuss with the scientists and perform simple experiments like DNA isolation, cell staining, or microscopy on their own; and the regular participation in the "Lange Nacht der Wissenschaften" (Long Night of Sciences). At this event, about 70-80 universities and research institutions all over Berlin and Potsdam open their doors for one night and invite the general public to visit their labs, learn about the work that is done here and discuss it with the scientists. In 2018, the MPIMG has also participated in the Max Planck Day, a nationwide, public MPG science event that took place in September throughout Germany and beyond. Together with the Max Planck Institutes for Human Development and for the History of Science, the MPIMG organized a workshop about science communication for junior journalists and junior scientists. Sixteen trainees from the ems – electronic media school, Potsdam, met ten junior scientists from the MPIs and developed journalistic concepts to communicate the respective research projects of the PhD students. As a second activity in the context of the Max Planck Day, a postdoc of the MPIMG participated in the Science goes Stage! Science Slam – Party – Chemie Live-Show, jointly organized by the MPIs in Berlin and Potsdam and the Berlin Office of Max Planck Society, and, to our great pleasure, won the first price at the Science Slam.

Figure 8.
Science experiments for kids during the Long Night of Sciences at the MPIMG

Figure 9.
Jesse Veenvliet, postdoc at the Dept. of Developmental Genetics, at the Max Planck Science Slam in September 2018

Department of Developmental Genetics

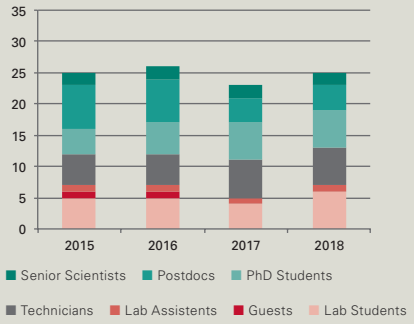
The department has been established in November 2003.

Prof. Dr. Bernhard G. Herrmann
herrmann@molgen.mpg.de
++49 30 8413-1409

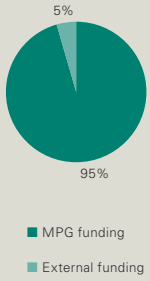
Scientific coordinator:
Dr. Marta Caparros Rodriguez
caparros@molgen.mpg.de
++49 30 8413-1344

Secretary:
Christel Siegers
siegers@molgen.mpg.de
++49 30 8413-1344

Team



Funding



Focus areas

- *Developmental genetics* – gene regulation networks controlling embryonic body formation
- *Non-Mendelian inheritance* – mechanisms causing transmission ratio distortion

Research concept

The Department of Developmental Genetics combines strong expertise in mouse genetics and embryology with state-of-the-art technology to investigate two major questions: i) how are trunk and tail formation during embryogenesis genetically controlled, and ii) what are the basic principles of non-Mendelian inheritance in mammals. We utilize a combination of classical embryology methods, Crispr/Cas mediated genetic engineering of embryonic stem cells, highly efficient techniques for the generation of transgenic chimeric embryos and mice, imaging, FACS, (single cell) transcriptome analysis, genomics approaches and bioinformatics.

Understanding mammalian development requires deciphering the mechanisms by which stem cells give rise to various cell lineages, and how they organize into functional 3-dimensional structures and organs. The latter is an immensely complex process requiring thousands of protein-coding as well as non-coding genes, which change activity states from cell type to cell type. Since Brachyury (T) is one of a small set of key control factors (together with Sox2, Foxa2) essential for trunk and tail development, a major focus of our work is concerned with the role of T in relation to these factors in shaping the embryonic trunk and tail. Bodies are formed from billions of cells, and therefore understanding development requires to unravel gene regulation mechanisms at the single cell level and deci-

phering the self-organization of cells taking different fates in response to signals secreted from their neighbors, thereby forming a functional assembly of organs and tissues.

More than a century ago, Gregor Mendel discovered the rules of genetic inheritance. But any rule has exceptions, and such phenomena are summarized under the term non-Mendelian inheritance. It describes the unequal transmission of alleles from the parents to their offspring. The classical model of non-Mendelian inheritance in mouse (first described in 1936) is the phenomenon of transmission ratio distortion caused by the mouse *t*-haplotype. The latter is transmitted almost exclusively from heterozygous (*t*/+) males to their offspring. We have worked out the molecular mechanisms promoting this selfish behavior of the *t*-haplotype over many years and have revealed the molecular principles, genes and pathways involved. Now we urge to find out if TRD is a peculiarity of the *t*-haplotype or if there's more to it.

Scientific highlights

- Brachyury directs histone acetylation to target loci during mesoderm development [Beisaw et al. EMBO Reports 2018]
- Antagonistic activities of *Sox2* and *Brachyury* control the fate choice of neuro-mesodermal progenitors [Koch et al. Dev Cell 2017]
- Pattering and gastrulation defects caused by the t^{w18} lethal are due to loss of *Ppp2r1a* [Lange et al. Biology Open 2017]
- Different concentrations of FGF Ligand, FGF2 or FGF8 determine distinct states of WNT-induced presomitic mesoderm [Sudheer et al. Stem Cells 2016]

Structure and organization of the department

The Department of Developmental Genetics is organized in several project groups and one research group. Scientists, students and technicians cooperate in various complementary projects supervised by the department head. Master and Bachelor students are supervised by post-doctoral or senior scientists. The infrastructure and resources are maintained mainly by the technicians.

Results of completed projects during reporting period

Gene regulation networks controlling trunk development in the mouse

The spinal cord and mesodermal tissues of the trunk such as the vertebral column and skeletal musculature derive from neuro-mesodermal progenitors (NMPs). *Sox2*, *Brachyury (T)*, and *Tbx6* have been correlated with NMP potency and lineage choice; however, their exact role and interaction in these processes have not yet been revealed. We have presented a global analysis of NMPs and their descending lineages performed on purified cells from embryonic day 8.5 wild-type and mutant embryos. We showed that *T*, cooperatively with WNT signaling, controls the progenitor state and the switch toward the mesodermal fate. *Sox2* acts antagonistically and promotes neural development. *T* is also involved in remodeling the chromatin for mesodermal development. *Tbx6* reinforces the mesodermal fate choice, represses the progenitor state, and confers paraxial fate commitment. Our findings refine previous models and establish molecular principles underlying mammalian trunk development, comprising NMP maintenance, lineage choice, and mesoderm formation.

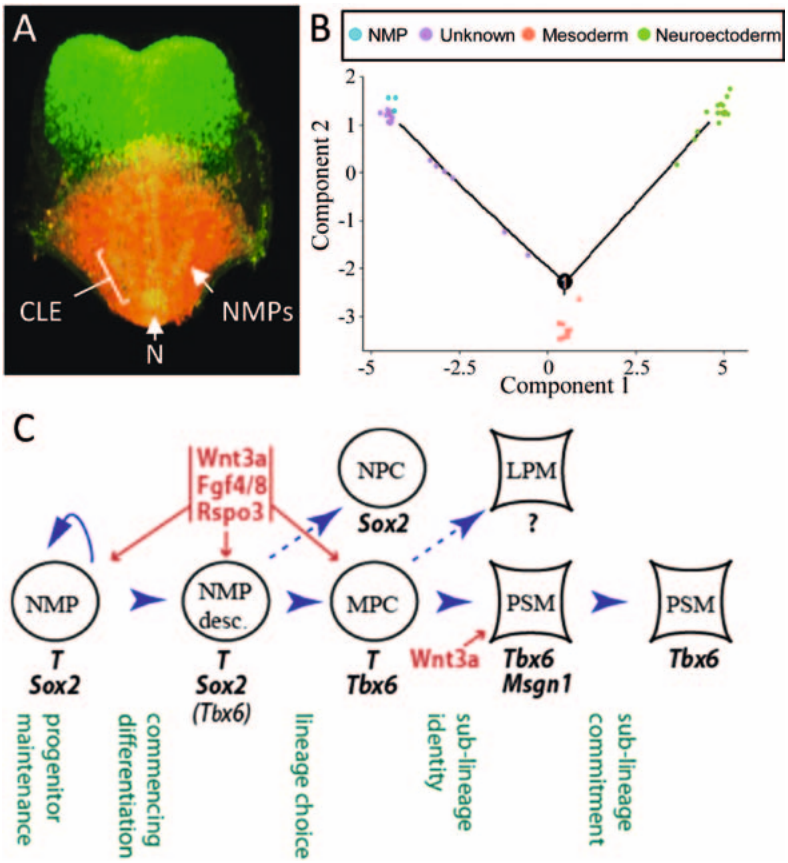


Figure 1.
Location of NMPs in the growth zone, lineage choice, and genetic control of paraxial mesoderm formation during trunk development.
(A) 3D model of a TS12 (E8.25) embryo expressing T:mCherry (red) and Sox2:Venus (green) reporters imaged by lightsheet microscopy. Caudal view onto the growth zone; NMPs are identified in the caudal lateral ectoderm (CLE) and double positive for *T* and *Sox2* (yellow), as is the node (N).
(B) PCA plot showing the cells trajectories based on differential gene expression between cell types. Note that cells classified as "unknown" are located next to NMPs and spread along the trajectory, indicating NMP descendants undergoing lineage choice. (C) Paraxial (pre-somitic) mesoderm (PSM) formation is a multi-step process controlled by the WNT targets *T*, *Tbx6*, and *Msgn1*. *Tbx6* supports T in the mesodermal fate choice; their target *Msgn1* confers PSM identity and *Tbx6* PSM commitment.

Koch F, Scholze M, Wittler L, Schifferl D, Sudheer S, Grote P, Timmermann B, Macura K & Herrmann BG. Antagonistic Activities of *Sox2* and *Brachyury* Control the Fate Choice of Neuro-Mesodermal Progenitors. Dev Cell 42: 514-526 e7 (2017)

Epigenetic control of gene expression in embryonic processes

T-box transcription factors play essential roles in multiple aspects of vertebrate development. We showed that the cooperative function of BRACHYURY (T) with histone-modifying enzymes is essential for mouse embryogenesis. A single point mutation (T^{Y88A}) results in decreased histone 3 lysine 27 acetylation (H3K27ac) at T target sites, including the *T* locus, suggesting that T autoregulates the maintenance of its expression and functions by recruiting permissive chromatin modifications to putative enhancers during mesoderm specification. Our data indicate that T mediates H3K27ac recruitment through a physical interaction with p300. In addition, we determine that T plays a prominent role in the specification of hematopoietic and endothelial cell types. Hematopoietic and endothelial gene expression programs are disrupted in T^{Y88A} mutant embryos, leading to a defect in the differentiation of hematopoietic progenitors. We show that this role of T is mediated, at least in part, through activation of a distal *Lmo2* enhancer.

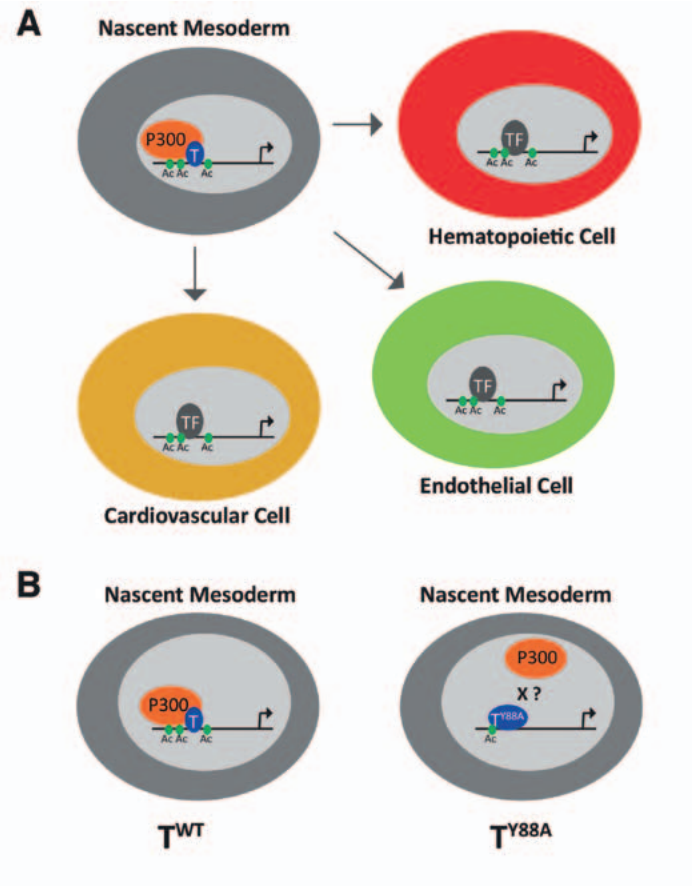


Figure 2.
Model of T function in early mesodermal development.
(A) Schematic illustrating T-mediated recruitment of p300 and H3K27 acetylation in nascent mesoderm to target loci involved in development of cell types derived from the mesoderm, including hematopoietic, endothelial, and cardiovascular cells.
(B) Schematic illustrating the proposed model that p300 recruitment to *T* target genes, including the *T* gene itself fails in mesodermal cells expressing the T^{Y88A} mutant, resulting in reduced H3K27 acetylation and lower target gene expression.

Beisaw A, Tsaytler P, Koch F, Schmitz S U, Melissari M-T, Senft A D, Wittler L, Pennimpede T, Macura K, Herrmann BG & Grote P. BRACHYURY directs histone acetylation to target loci during mesoderm development. EMBO Rep 19(1):118-134 (2018)

Cooperation and distinct activities of growth factors in mesoderm formation

Presomitic mesoderm (PSM) cells are the precursors of the somites, which flank both sides of the neural tube and give rise to the musculo-skeletal system shaping the vertebrate body. WNT and FGF signaling control the formation of both the PSM and the somites and show a graded distribution with highest levels in the posterior PSM. We have used reporters for the mesoderm/PSM control genes *T, Tbx6*, and *Msgn1* to investigate the differentiation of mouse ESCs from the native state via EpiSCs to PSM cells. We showed that the activation of WNT signaling by CHIR99021 (CH) in combination with FGF ligand induces embryo-like PSM at high efficiency. By varying the FGF ligand concentration, the state of PSM cells formed can be altered. High FGF concentration supports posterior PSM formation, whereas low FGF generates anterior/differentiating PSM, in line with *in vivo* data. Furthermore, the level of *Msgn1* expression depends on the FGF ligand concentration. We also show that Activin/Nodal signaling inhibits CH-mediated PSM induction in EpiSCs, without affecting *T*-expression. Inversely, Activin/Nodal inhibition enhances PSM induction by WNT/high FGF signaling. The ability to generate PSM cells of either posterior or anterior PSM identity with high efficiency *in vitro* will promote the investigation of the gene regulatory networks controlling the formation of nascent PSM cells and their switch to differentiating/somitic paraxial mesoderm.

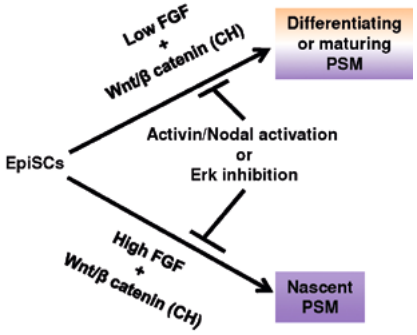


Figure 3.
The WNT (CH)-driven differentiation of EpiSCs can be directed toward distinct PSM cell fates, depending on the FGF ligand concentration. High FGF concentration supports the posterior, nascent PSM state, low FGF supports the anterior, maturing PSM and somitic (differentiating) cell state. ERK pathway inhibition or Activin/Nodal activation interfere with WNT (CH)-driven PSM formation. Abbreviations: CH, CHIR99021; EpiSCs, epiblast stem cells; FGF, fibroblast growth factor; PSM, presomitic mesoderm.

Sudheer S, Liu J, Marks M, Koch F, Anurin A, Scholze M, Senft AD, Wittler L, Macura K, Grote P & Herrmann B G. Different concentrations of FGF Ligand, FGF2 or FGF8 determine distinct states of WNT-induced presomitic mesoderm. Stem Cells 34: 1790-1800 (2016)

Cloning and embryonic phenotype analysis of a *t*-haplotype lethal from the mouse *T/t* complex

The mouse *t* haplotype, a variant 20cM genomic region on Chromosome 17, harbors 16 embryonic control genes identified by recessive lethal mutations isolated from wild mouse populations. Due to technical constraints so far only one of these, the t^{w5} lethal, had been cloned and molecularly characterized. We have reported the molecular isolation of the t^{w18} lethal. Embryos carrying the t^{w18} lethal die from major gastrulation defects commencing with primitive streak formation at E6.5. We have used transcriptome and marker gene analyses to describe the molecular etiology of the t^{w18} phenotype. We showed that both WNT and Nodal signal transduction are impaired in the mutant epiblast, causing embryonic patterning defects and failure of primitive streak and mesoderm formation. By using a candidate gene approach, gene knockout by homologous recombination and genetic rescue, we have identified the gene causing the t^{w18} phenotype as *Ppp2r1a*, encoding the PP2A scaffolding subunit PR65alpha. Our work highlights the importance of phosphatase 2A in embryonic patterning, primitive streak formation, gastrulation, and mesoderm formation downstream of WNT and Nodal signaling.

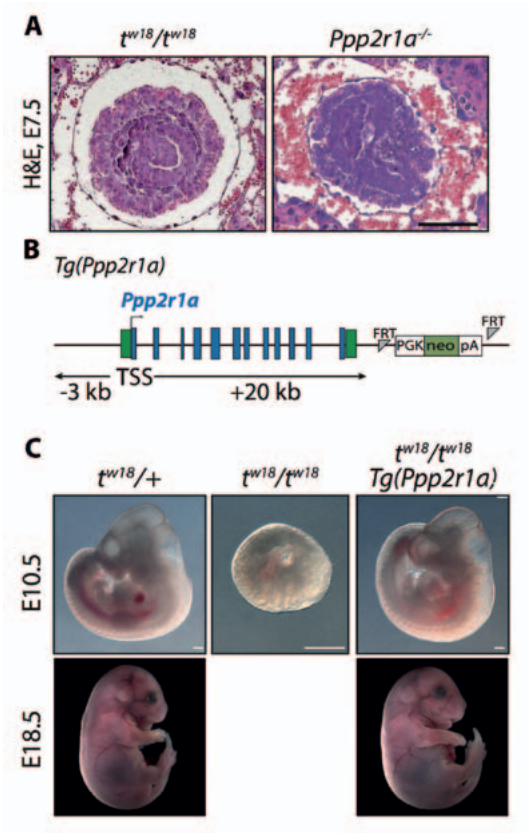


Figure 4.
Phenotypic rescue of t^{w18}/t^{w18} mutant embryos by a transgene construct demonstrates that *Ppp2r1a* encodes the t^{w18} lethal.
(A) Histological H&E-stained paraffin sections of E7.5 *Ppp2r1a* knockout and t^{w18}/t^{w18} embryos. The amniotic cavities of both embryos are filled with cells.
(B) Rescue construct containing the *Ppp2r1a* locus comprising 3 kb upstream and 20 kb downstream of the transcription start site (TSS).
(C) *Ppp2r1a* expression rescues the t^{w18} homozygous phenotype. Embryos from breedings of $t^{w18}/+$ females with $t^{w18}/+$, Tg(*Ppp2r1a*) males were analyzed at E10.5 and E18.5; homozygous t^{w18} embryos carrying the transgene Tg(*Ppp2r1a*) developed normally at least until E18.5.

Lange L, Marks M, Liu J, Wittler L, Bauer H, Piehl S, Bläß G, Timmermann B & Herrmann BG. Patterning and gastrulation defects caused by the t^{w18} lethal are due to loss of *Ppp2r1a*. Biology Open 6: 752-764 (2017)

Ongoing and future projects

Gene regulation networks involved in tissue formation, lineage choice and organization generating the trunk and tail of the mouse

[Frederic Koch, Jesse Veenvliet, Manuela Scholze-Wittler, Dennis Schifferl, Lars Wittler, Michael Gerloff, Sandra Piehl]

Trunk and tail formation in the mouse embryo occur consecutively by continued addition of cells, generated from stem or progenitor cells located in the caudal lateral ectoderm near the node and later at the caudo-neural hinge, to the growing body axis. The stem cells, via cell division generate descendants allowed to undergo differentiation, and the latter, under the influence of signals determining their fate, go through a number of lineage decisions, before their destination is finally fixed. Important embryonic structures controlling the fate choice are the node and its derivative, the notochord, which organize cells in the growth zone and the tissues and organs derived from it, into a dorso-ventral pattern determining, together with other signaling centers, the final body anlage.

We have shown that the neural tube and the mesodermal tissues derive from neuro-mesodermal progenitors giving rise to the trunk. We already understand in quite some detail, how the formation of paraxial mesoderm, which eventually gives rise to the axial skeleton and skeletal muscles, is genetically controlled. However, the origin of lateral mesoderm in the trunk is not well understood. In addition, the cellular origin of the node and notochord has not been defined at the genetic level, and the same is true for the tail. All these questions are very tricky to address, and require a toolbox of finely tuned genetic instruments and high-resolution microscopy as well as single cell transcriptomics. We have embarked on answering these questions and expect to come up with satisfactory answers soon.

A 3-D model for mid-gestational development in a dish

[Jesse Veenvliet, Frederic Koch, Manuela Scholze-Wittler, Lars Wittler]

Axial elongation leading to formation of the trunk and tail is a reiterative process of continued generation of mainly mesodermal and neural cell lineages from neuro-mesodermal progenitors (NMPs), followed by dorso-ventral patterning guided by the organizing centers, the node and notochord. These processes are characterized by extensive growth, multiple lineage decisions and morphogenetic changes. Naturally, in mammals the dynamics of these processes in space and time are difficult to study *in vivo*. A robust *in vitro* culture system that reliably and reproducibly mimics mid-gestational mouse development would overcome these limitations and help to reduce animal experiments. It was recently shown that mouse embryonic stem cells, aggregated under defined conditions, can self-organize into structures resembling the posterior region of the embryo [van den Brink et al., Development 2014; Turner et al., Development 2017]. Inspired by these reports, we have recently started to generate 3D structures that recapitulate trunk/tail development (tGastruloids). These structures appear to induce all three germ layers in a spatially organized manner reminiscent of their *in vivo* counterpart and thereby lay the blueprint for the body plan. Importantly, tGastruloids comprise cells co-expressing Brachyury and Sox2 at their posterior end, resembling neuro-mesodermal progenitors (NMPs), the multipotent cell core essential for trunk development. We work on further improving the tGastruloid system to develop it into a powerful platform for studying the morphogenetic changes and lineage decisions underlying mid-gestational development, in the culture dish.

Genetic and physiological basis of transmission ratio distortion in the mouse

[Hermann Bauer, Alexandra Amaral, Pedro Gamez, Jana Sie, Jürgen Willert, Bettina Lipkowitz, Karin Berktold]

The observation of non-Mendelian inheritance in the mouse was a fortuitous discovery published in 1936. It was observed that the recessive "allele" of the *T* locus, *t*, is transmitted at an unusually high rate from *T/t* males to their offspring. Five decades of genetic analysis of the *t* "allele" revealed that the latter consists of a chromosomal region of some 40 Mb, now called *t*-haplotype. It contains several distorters and a responder, which interact to achieve the high transmission ratio of the called *t*-haplotype. We have cloned the responder and four distorters and have shown that the latter act in Rho signaling cascades controlling the activity of the wild-type responder, the protein kinase SMOK, in spermatozoa swimming towards the egg cells. The distorters impair the regulation of SMOK, thereby interfering with the directional movement of sperm cells. However, the *t*-responder, $SMOK^{TCR}$, can rescue forward progression of sperm carrying the *t*-haplotype conferring an advantage to spermatozoa expressing the *t*-responder. Thus, the latter have an advantage in reaching the egg cells faster than the spermatozoa carrying the wild-type chromosome.

There is no doubt that the combination of gene variants locked into the *t*-haplotype able to cause transmission ratio distortion is the result of a long evolutionary process involving multiple steps of mutation, selection and fine tuning. However, a crucial question has not been answered yet: is the *t*-responder a singular mutant gene optimized for TRD, or are there more gene loci in the mammalian genome, which can distort the transmission ratio? This is a crucial question we want to address. In addition we are working on applying TRD to farm animal breeding.

The Mediator complex and transcriptional regulation

[Heinrich Schrewe, Bruno Perreira, Andrea König]

The Mediator complex, a conserved multi-subunit signal processor, is a pivotal co-activator of enhancer activity that serves as a molecular bridge between distal, enhancer-associated regulators and the RNA polymerase II machinery assembled on core promoters. Several approaches are used to decipher the complex functions of Mediator subunit Med12 in transcriptional regulation. Using a series of tissue- and cell type-specific cre-deleter and our conditional Med12 knock-out mouse line, we generated embryos with particular disease phenotypes. Studying these mutants, we aim to identify key Med12-linked signalling pathways and transcriptional regulators that control specific developmental processes.

Three human X-linked intellectual disability (XLID) syndromes have been associated with missense mutations in *MED12*. We have recently generated mouse models for two human MED12 and XLID syndromes, which express the Med12 mutant forms only. We are currently generating a comprehensive picture of the transcriptomes and proteomes of cells and tissues derived from these mutants to understand the biology of the developmental defects and of the disease.

Knowing that human MED12 not only binds transcriptional activators, but also lncRNAs directly or via RNA-binding proteins, our aim is to characterize all Med12 interaction partners, including lncRNAs. Using *in vivo* proximity- and epitope-biotinylation followed by affinity purification, immunoprecipitation and mass spectrometry we are currently identifying candidate interactors linked to developmental and neurological defects associated with Med12-related pathologies.

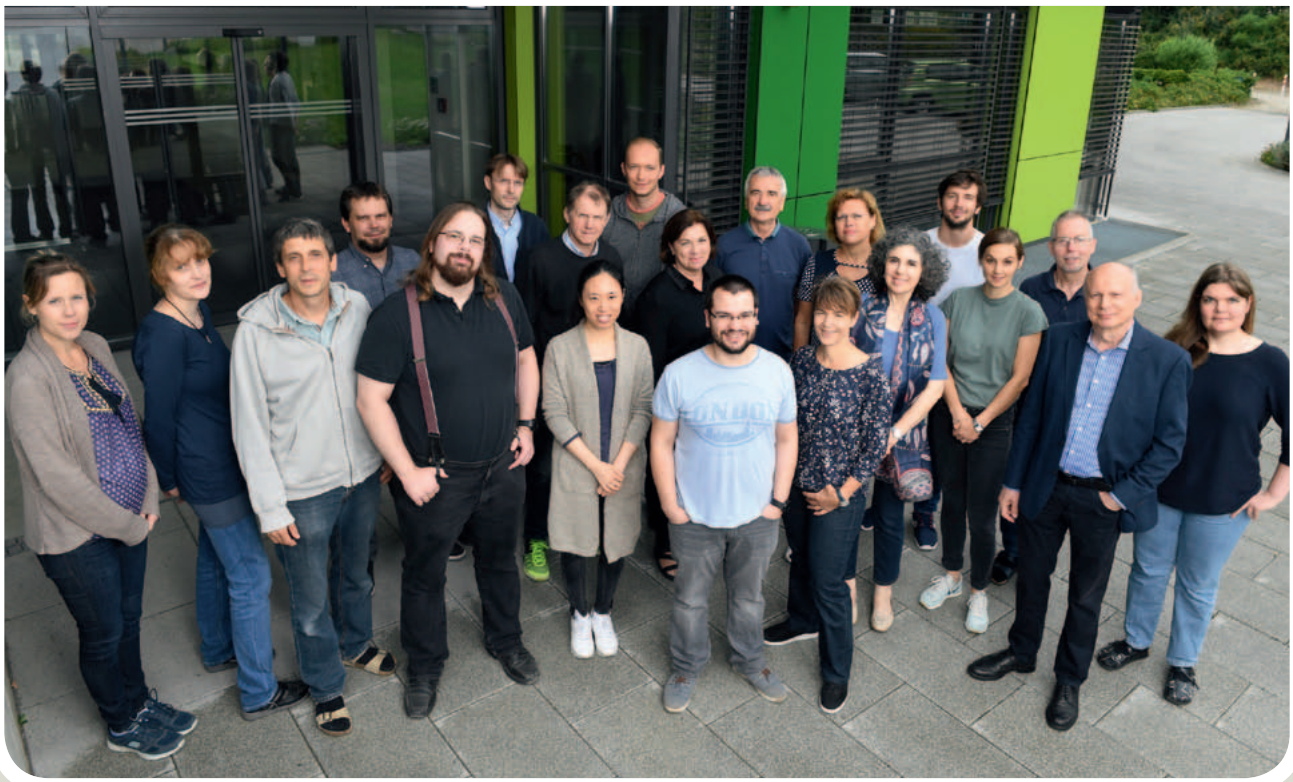
Scientific honours

- Smita Sudheer was appointed to Assistant Professor at the Central University of Kerala, Department of Genomic Science, India, after leaving the department in 2016.

Material resources, special facilities and equipment

The Department of Developmental Genetics provides personnel (one scientist, one engineer) and expertise to the *Transgenic Unit* of the Institute, which produces transgenic embryos and mice from genetically modified embryonic stem cells for all groups at the Institute.

Members of the department have had the supervision over the Fluorescence Activated Cell Sorter operation of the Institute from start. In June 2018, a new head of operation for the FACS has been employed by the Institute. The department continues the scientific supervision of the facility.



Department of Genome Regulation

The department has been established in July 2017.

Prof. Dr. Alexander Meissner
meissner@molgen.mpg.de
++49 30 8413-1880

Secretary:
Cordula Mancini
office-meissner@molgen.mpg.de
++49 30 8413-1881

Structure and organization of the department

The department of Genome Regulation is currently made up of five independent groups that include the Meissner and Elkabetz labs, which provide a more detailed report below as well as the more recent additions of the Bulut-Karslioglu, Hnisz and Müller groups, which are briefly introduced here. Aydan Bulut-Karslioglu is the latest member to join our department and has previously trained with Thomas Jenuwein, Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany, and Miguel Ramalho-Santos, University of California, San Francisco. She has published many high-profile papers and received the prestigious Emmy Noether (declined) and Sofja Kovalevskaja (accepted) awards. Aydan began to set up her lab (Bulut-Karslioglu lab - Stem Cell Chromatin Group) at the MPIMG in July 2018. Together with her team, she wants to study the molecular mechanisms of chromatin and gene regulation in stem cells and in the developing embryo. Denes Hnisz did his postdoctoral training with Richard A. Young at the Whitehead Institute for Biomedical Research, Cambridge MA, and transitioned to the MPIMG in early 2018. His group (Hnisz lab - Precision Gene Control group) seeks to understand how transcriptional programs are established during development, how their corruption leads to human pathologies, and how this knowledge can be used for therapeutic manipulation of human disease genes. Franz-Josef Müller trained at the Scripps Research Institute with Jeanne F. Loring before setting up his own lab in Kiel, Germany. He continues his work there and, in addition, runs a small group (Müller lab - Cellular Phenotyping Group) in a collaborative setting at the MPIMG. Along with his interest in defining cellular phenotypes, developmental states and disease processes through genome wide measurements, his team is developing applications based on nanopore sequencing.

Research concept of the whole department

The Department of Genome Regulation brings together strong expertise in mouse and human stem cell models, early mouse developmental biology and transcription biology with a range of genomic approaches. The overarching theme of the department is to obtain a comprehensive picture of how our genomes are regulated in normal development and which parts are derailed in disease states in particular cancer.

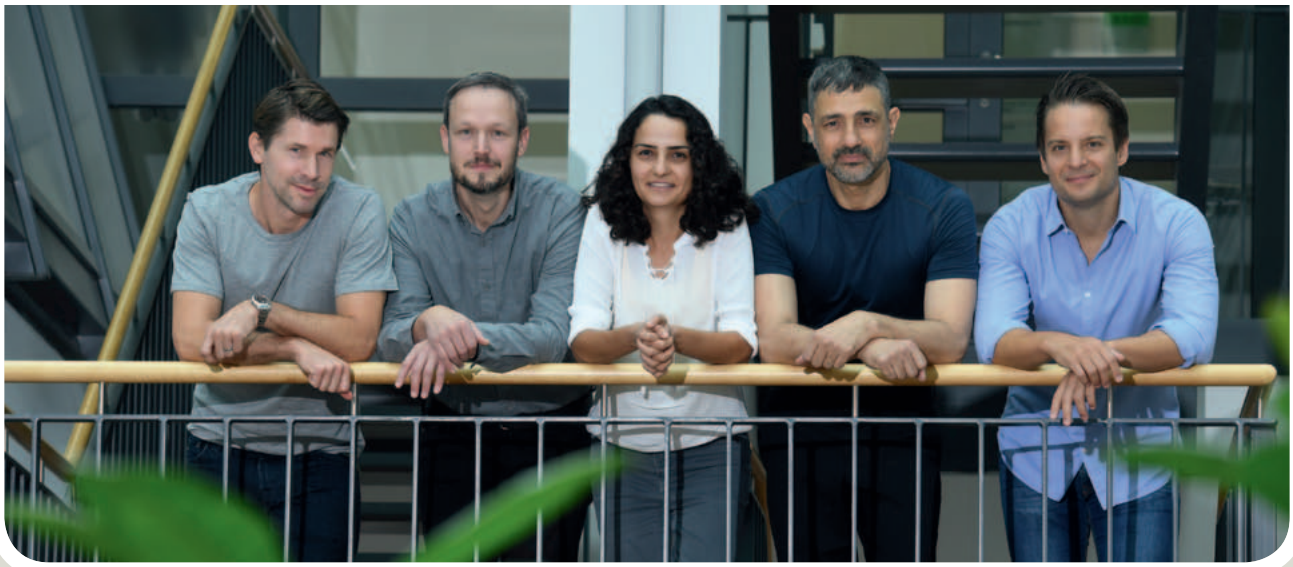
Material resources, equipment, and spatial arrangements

The department currently offers classic molecular biology bench space for around 25 people and has three tissue cultures (7 hoods and 12 incubators) with state of the art equipment. Despite its short existence we managed to fill all available space on the second floor of tower 2 and the plan is to further expand into the second floor of tower 1 upon the completion of its ongoing renovation. All five groups share the three tissue cultures to maximize interaction within the department and the Meissner, Bulut-Karslioglu, Hnisz and Müller labs have adjacent bench space. All molecular biology equipment is shared in central places and includes a BioRad ChemiDoc MP Imaging system, Nanodrop, 4D Nucleofector, Oxford Nanopore MinIONs, Nanostring, BD Celesta, Covaris E220 and ATT Fragment Analyzer and diverse high-end microscope systems for cell biology. Moreover, members of the department have full access to the MPIMG's extensive technological and computational resources, including the Illumina high throughput sequencing pipelines and advanced single cell genomic tools including a Fluidigm C1, a 10x Genomics Chromium, and FACS capabilities.

Ongoing work and planned developments

Since its inception little more than a year ago we have established a fully operational and lively department that has also successfully integrated into the larger institute. The department has already more than 35 members and actively participates in the Institute seminars as well as the IMPRS. We have recently established a faculty chalk talk series that aims to bring the different research groups closer together and promote collaboration within the Institute. Members of our department are working closely with the various core facilities and have begun to collaborate with the other groups at the MPIMG.

Details about the ongoing and planned developments for the Meissner lab and the Elkabetz lab can be found in the individual reports of the groups on pages 30 ff. The other groups started operation between spring and summer 2018 and more information can be found on their respective websites.

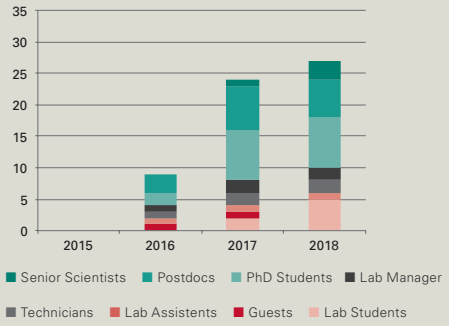


Meissner Lab – Genome Regulation Group

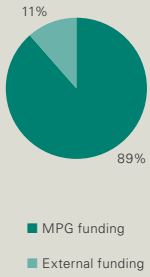
Alex Meissner started to build up his group in secondary employment in November 2016. His principal employment started in July 2017.

Prof. Dr. Alexander Meissner
meissner@molgen.mpg.de
++49 30 8413-1880

Team



Funding



Focus areas

- *Developmental biology* – genome regulation in early embryogenesis
- *Cancer epigenomics* – origin and relevance of the altered cancer methylome
- *DNA methylation* – regulatory dynamics in development and disease
- *Pluripotency and reprogramming* – mechanisms that regulate cell state transitions

Research concept

We are a mixed group of experimental and computational biologists that uses genomic tools to study developmental and stem cell biology. By applying a range of tools from zygotic Cas9 mediated knockouts, single cell RNA sequencing, Cas9 based lineage tracing and various epigenomic approaches, we focus in particular on early mammalian development. Some of this led to a more recent interest in cancer biology, where we explore links to early embryonic development and general mechanisms of epigenetic deregulation. DNA methylation also remains a central focus in our basic research program that helps us to study genome organization including annotation of regulatory features and their dynamics in development. Finally, we continue to study the underlying mechanisms that regulate cell state transitions in particular exit and entry to pluripotency.

Scientific highlights

- Most cancer types share an epigenetic landscape that closely mirrors the early extraembryonic patterns and suggests a possible developmental origin of the disease [Smith et al. Nature 2017]
- BrdU labeling of nascent DNA demonstrates a lag between copying of genetic and epigenetic information during DNA replication [Charlton et al. Nat Struct Mol Biol 2018]
- Pioneer transcription factors are sampling the genome at many future binding sites but stable target engagement appears regulated through co-factors [Donaghey et al. Nat Genet 2018]
- *De novo* methylation is required for the proper specification towards motor neurons and functionally linked to key gene regulatory elements [Ziller et al. Cell Stem Cell 2018]
- Generation and use of a methylation depleted but maintenance competent mouse ES cell line shows ubiquitous nuclear activity of dCas9-DNMTs [Galonska et al. Nat Commun 2018]

Scientific honours & selected invited talks

- Election as a Member to the European Molecular Biology Organization (EMBO), 2018
- Invited speaker Keystone Symposium "DNA and RNA methylation" Vancouver, Canada, 2018
- Invited speaker Ernst Klenk Symposium in Molecular Medicine, Cologne, Germany, 2018
- Invited speaker and organizer of the ISSCR, ESGCT, SFTCG meeting on "Changing the face of modern medicine: Stem cells and gene therapy", Lausanne, Switzerland, 2018
- Invited speaker Annual Meeting of the ISSCR, Boston, USA, 2017

Results of completed projects during reporting period

Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer

This study was completed during my transition to the MPIMG and provides the foundation of numerous ongoing projects in the current lab. Concerted efforts over past decades have established a thorough understanding of the canonical somatic DNA methylation landscape as well as its systematic misregulation across many cancers. In this study we present a base pair resolution analysis of global remethylation from the hypomethylated genome of the preimplantation embryo into the early epiblast and extraembryonic ectoderm. We show that these two states acquire highly divergent genomic distributions: while the epiblast establishes the canonical CpG density-dependent pattern found in somatic cells, the extraembryonic epigenome becomes substantially more mosaic. This alternate pattern includes specific *de novo* methylation of hundreds of CpG island (CGI) promoter containing genes that function in early embryonic development and are orthologously methylated across a large number of human cancer types. Based on these data, we propose a model where the evolutionary innovation of extraembryonic tissues in mammals required cooption of DNA methylation-based suppression as an alternate pathway to the embryonically utilized Polycomb group proteins, which otherwise coordinate germ layer formation in response to extraembryonic cues at the onset of gastrulation. Moreover, we establish that this decision is made deterministically downstream of promiscuously utilized, and frequently oncogenic, signaling pathways via a novel combination of epigenetic cofactors. Recruitment of this silencing mechanism to developmental genes during tumori genesis may therefore reflect the misappropriation of an innate regulatory pathway that may be spontaneously sampled as an alternate epigenetic landscape within somatic cells.

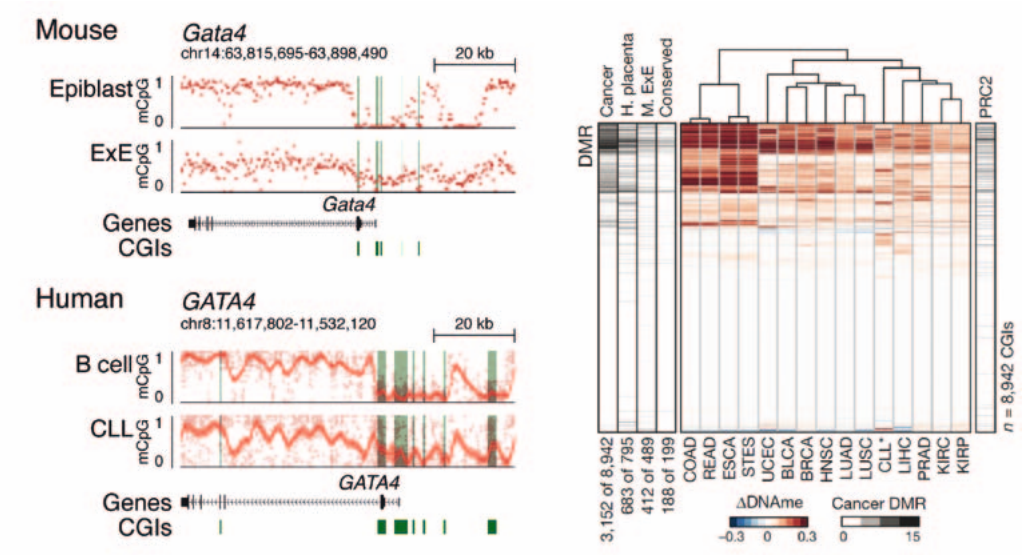


Figure 1:
Whole genome bisulfite sequencing (WGBS) data shows the striking resemblance of the DNA methylation landscape in mouse developmental samples (epiblast and extraembryonic ectoderm) compared to human cancer (B cell to chronic lymphocytic leukemia). Expansion of this representative example indicates that these CGIs are the most conserved targets across and within cancer types.

Smith ZD, Shi J, Gu H, Donaghey J, Clement K, Cacchiarelli D, Gnirke A, Michor F & Meissner A. Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. Nature 549(7673):543-547 (2017)

Genome-wide tracking of dCas9-methyltransferase footprints

Since the emergence of the CRISPR-Cas9 system for genome editing, it has evolved into a versatile toolbox for numerous applications, including the targeted manipulation of the epigenome using dCas9-coupled epigenetic effector proteins. Initial proof-of-concept studies have highlighted the general applicability of the system, thereby raising the possibility for functional dissection of individual epigenetic features and their direct impact on gene regulation. Our recent work offers a unique tool (the DKO^{zero} line) to track all dCas9-methyltransferase activity that can be cleanly separated from any endogenous *de novo* methylation by Dnmt3a and Dnmt3b. It provides a first true account of the unexpected genome-wide DNA methylation footprints left by dCas9-coupled to the catalytic domain of Dnmt3a.

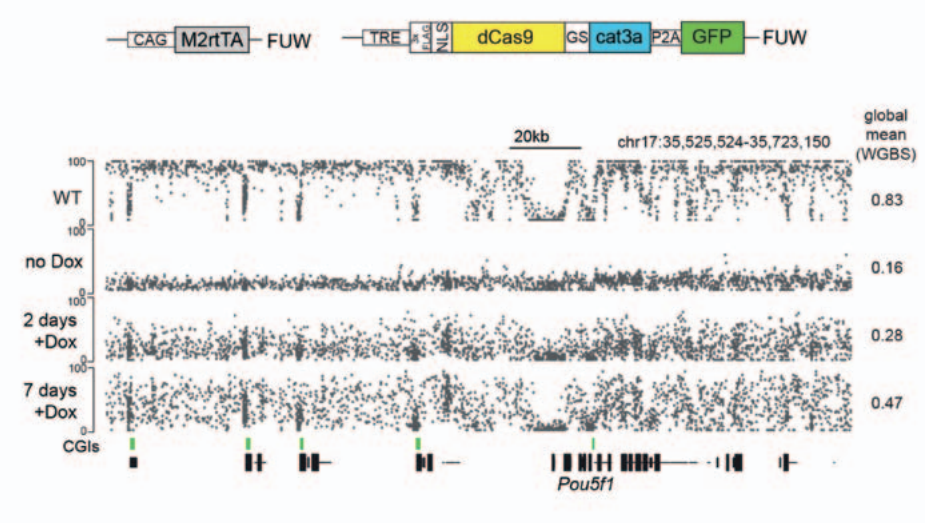


Figure 2:
Simplified schematics showing the M2rtTA and dCas9-cat3a FUW-based lentiviral expression constructs that were introduced into a DNA methylation depleted (Dnmt3a/b double knockout) maintenance competent ESC line (DKO^{zero}). Representative WGBS tracks displaying CpG methylation in WT as well as our engineered DNA methylation depleted cells with 0, 2, or 7 days of Dox induction (no guide RNA present) show the global off-target activity. Global mean CpG methylation levels are shown on the right. CGIs are highlighted beneath in green.

Galonska C, Charlton J, Mattei AL, Donaghey J, Clement K, Gu H, Mohammad AW, Stamenova EK, Cacchiarelli D, Klages S, Timmermann B, Cantz T, Schöler HR, Gnirke A, Ziller MJ & Meissner A. Genome-wide tracking of dCas9-methyltransferase footprints. Nat Commun 9(1):597 (2018)

Global delay in nascent strand DNA methylation

Cytosine methylation is widespread among organisms and essential for mammalian development. In line with early postulations of an epigenetic role in gene regulation, symmetric CpG methylation can be mitotically propagated over many generations with extraordinarily high fidelity. In our recent study we combined BrdU labeling and immunoprecipitation with bisulfite sequencing to explore the inheritance of cytosine methylation onto newly replicated DNA in human cells. Globally, we observe a pronounced lag between the copying of genetic and epigenetic information in ESCs that is reconsolidated within hours to accomplish faithful mitotic transmission. Populations of arrested cells show a global reduction of lag-induced intermediate CpG methylation when compared to proliferating cells, whereas sites of transcription factor engagement appear cell-cycle invariant. Alternatively, the cancer cell line HCT116 preserves global epigenetic heterogeneity independently of cell-cycle arrest. Taken together, our data suggest that heterogeneous methylation largely reflects asynchronous proliferation, but is intrinsic to actively engaged cis-regulatory elements and cancer. We are currently further exploring post replication DNA methylation maintenance using novel sequencing technologies (see below).

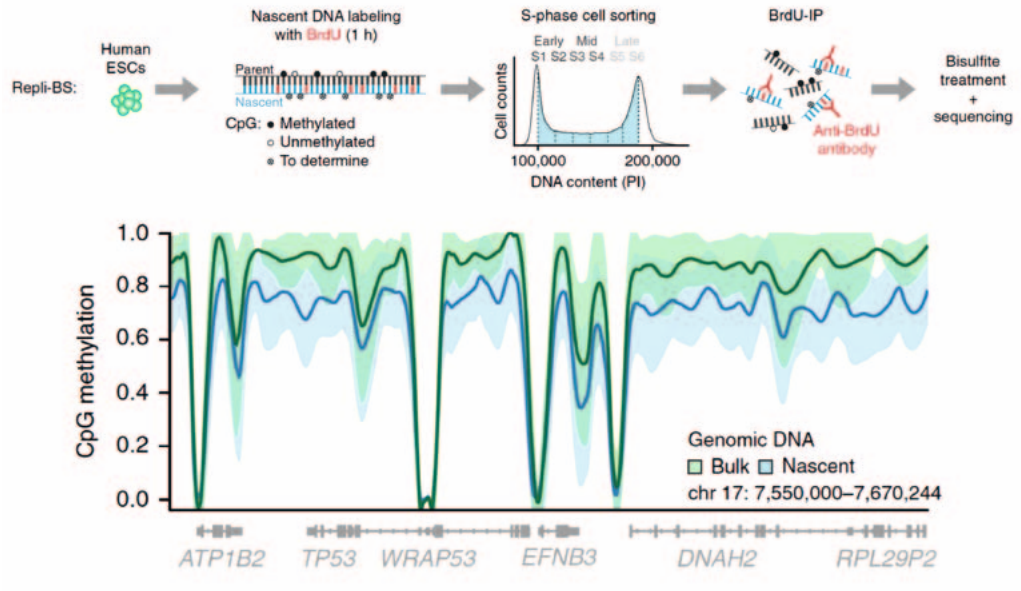


Figure 3: Schematic of our experimental design. Asynchronously growing human ESCs are labeled with BrdU for 1 h. During the pulse, BrdU (red) is incorporated into newly synthesized nascent DNA (blue) as it is copied from the parental strand (black). Empty circles marked with an 'x' ('to determine') indicate CpGs investigated by Repli-BS for methylation status. Six gates were used to sort S-phase cells according to DNA content. BrdU immunoprecipitation (BrdU-IP) of newly replicated DNA strands, which were treated with bisulfite before library generation and sequencing. LOESS-smoothed mean methylation levels are lower for nascent DNA compared to bulk (bold lines). Shaded areas show one s.d.

Charlton J, Downing TL, Smith ZD, Gu H, Clement K, Pop R, Akopian V, Klages S, Santos DP, Tsankov AM, Timmermann B, Ziller MJ, Kiskinis E, Gnirke A & Meissner A. Global delay in nascent strand DNA methylation. Nat Struct Mol Biol 25(4):327-332 (2018)

Ongoing and future projects

Nanopore sequencing enables direct DNA methylation measurements on long reads

[Pay Giesselman, Helene Kretzmer, Charles Haggerty, Nina Bailly, Alexandra Mattei, Jocelyn Charlton, Franz-Josef Müller]

Over the past year we developed and established an effective pipeline that uses the Oxford Nanopore technology to directly read DNA methylation (and selected base analogs) from long sequencing reads. Several ongoing projects are utilizing this for the precise mapping of pathological repeat expansions, dissection of phased DNA methylation patterns including during DNA replication, detailed mapping of repetitive feature across the genome as well as early cancer diagnostics.

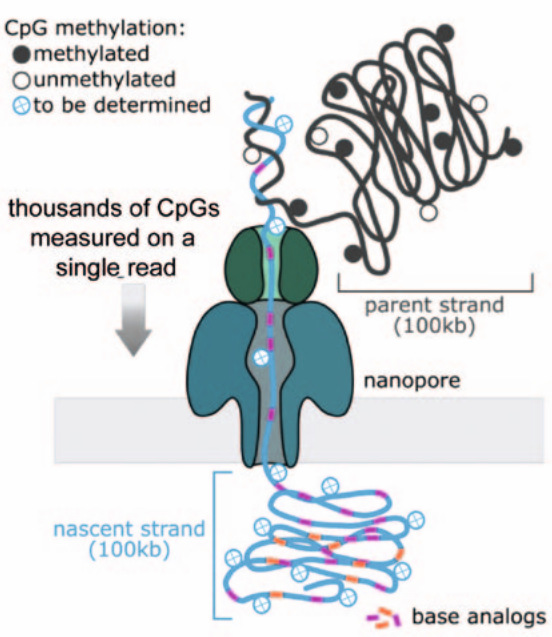


Figure 4:
Simplified schematic of a nanopore that can be used to determine DNA methylation of thousands of CpGs directly on a single read. Base analogs such as EdU can be used to confirm nascent versus parental strand identities during DNA replication.

Single cell RNA sequencing as a tool to dissect the role of epigenetic and transcriptional regulators *in vivo*

[Stefanie Grosswendt, Helene Kretzmer, Abhishek Kumar, Adriano Bolondi in collaboration with Zachary D. Smith, Harvard University]

We have established an effective experimental and computational pipeline to use zygotic Cas9/sgRNA injections to screen for the *in vivo* function of epigenetic and transcriptional regulators using a highly multiplexed single cell RNA sequencing readout. To date we generated a high-resolution wild-type reference data set, established efficient analysis pipelines and begun the systematic comparison to matched knockouts from DNMTs, PRC components and other early developmental regulators.

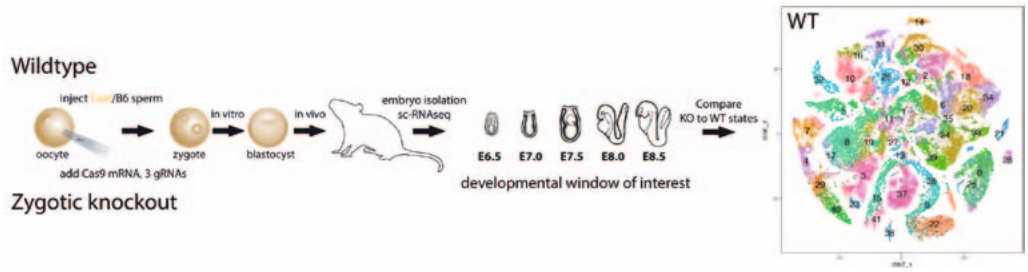


Figure 5:
A fast approach for generating dozens of knockouts was combined with single cell sequencing (10X Genomics) to study developmental roles of epigenetic and transcriptional regulators. Wildtype (WT) data were used to assign 42 states that serve as a reference for the knockout samples.

The altered DNA methylation landscape in B cell derived cancers emerges early and persists after treatment

[Helene Kretzmer, Sara Hetzel in collaboration with Kendell Clement, Harvard University, Michaela Gruber, Anat Biran and Cathy Wu, Dana Farber Cancer Institute]

The methylomes of most cancers show a striking departure from the normal bimodal somatic landscape, exhibiting a reduction of global methylation and higher-than-expected methylation levels at CGIs. Using the transition from normal B cells to monoclonal B cell lymphocytosis (MBL) and then chronic lymphocytic leukemia (CLL) as a model we are closely investigating the emergence of the alternate DNA methylation landscape as well as the post treatment molecular dynamics in a cohort of patients.

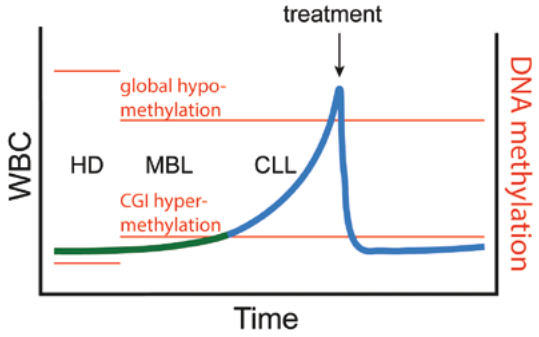


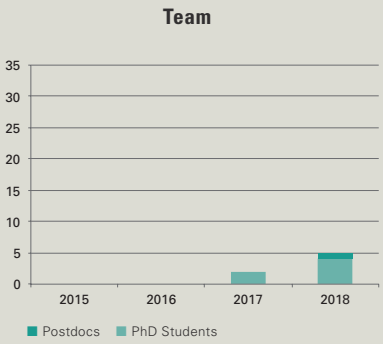
Figure 5:
Global hypomethylation and CGI hypermethylation compared to healthy donors (HD) arises early and remains stable over time including post treatment. Simplified schematic is based on our unpublished data.

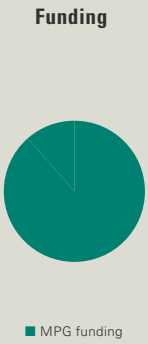
These are just some examples to illustrate the ongoing work of our lab. For more information, please have a look at our website on:
https://www.molgen.mpg.de/Genome-Regulation/Meissner-lab

Elkabetz Lab – Lab for Human Brain & Neural Stem Cell Studies

The group has been established in May 2017.

Dr. Yechiel Elkabetz
elkabetz@molgen.mpg.de
++49 30 8413-1886





Focus areas

- *Cortical development and disease* – neural stem cell patterning, heterogeneity and potential
- *Neural stem cell ontogeny* – programming and reprogramming towards neural stem cell fates
- *Fate determinants in lineage commitmen*t – lineage analysis of dividing stem cells
- *Origin of brain tumorigenesis* – molecular and cellular landmarks of medulloblastoma

Research concept

One of the major goals of the Elkabetz lab is to develop strategies for generating homogeneous sources of cultured neural stem cell (NSC) types. We are mainly motivated by the idea that only clinically relevant cells with accurate cell identity can produce meaningful phenotypes and tangible therapeutic solutions. Our main strategy is inspired by the notion that distinct cell types in culture can be distinguished by a myriad of features. To build tools to identify these features, we first devise paradigms for separation and isolation of distinct cell types. We then use genetic tools such as CRIPSPR genome editing, reporter gene expression and lentiviral-based gene introduction, in order to get access to, isolate and manipulate these unique cell types. Finally, we use multiple fluorescent tagging, barcode labeling and single cell sequencing methods to label, detect, track and analyze origin of brain regions in health and disease models.

Scientific highlights from former group

- Consecutive isolation of long-term cultured NSCs from pluripotent stem cells (PSCs) reveals distinct developmental stages that recapitulate human corticogenesis [Edri et al. Nat Commun 2015]
- Epigenetic footprinting and functional analyses of PSC-derived cortical NSC stages reveals transcriptional networks driving cortical NSC ontogeny [Ziller et al. Nature 2015]
- Quantitative live imaging of neural rosettes reveals structure-function dynamics of NSCs within rosettes coupled to corticogenesis [Ziv et al. PLoS Comput Biol 2015]

Scientific honours & selected invited talks

- Invited speaker (selected abstract), International Society for Stem Cell Research (ISSCR), Melbourne, Australia, 2018
- Invited Speaker, Keystone Symposia on Regenerative Biology and Applications, Hong Kong, 2017
- Invited Speaker and member of the organizing Committee: The International Conference on Stem Cells, Greece, 2017
- Invited Speaker, International Meeting for Cellular and Molecular Neuroscience, Israel, 2017

Results of completed projects during reporting period

Combinatorial pathway inhibition in brain organoids dictates distinct regional, stem cell and cytoarchitectural signatures in health and disease

[Naresh Mutukula, Sneha Arora, in collaboration with Elzbieta Gralinska, Martin Vingron, Björn Brändl, Franz-Josef Müller, Stefan Börno, Bernd Timmermann, MPIMG; Ramon Oliveira Vidal, Sascha Sauer, Max Delbrück Center for Molecular Medicine (MDC), Berlin; Rotem Volkman, former lab student at Tel Aviv University, Israel]

Most methods for deriving brain organoids from pluripotent stem cells are highly diverse and lack specific readouts that instruct landmarks for successful differentiation, hence resulting in heterogeneous populations. This reinvigorated us to revisit basic derivation paradigms side-by-side at the transcriptional, cellular, and cytoarchitectural levels. We generated multiple brain organoids by different methods and compared their transcriptional profile with those of different human brain regions at various developmental stages using bulk RNA-Seq and single cell RNA-Seq. We revealed that different derivation methods lead to highly distinct brain identities, and that only combined inhibition of BMP, TGF and WNT pathways exclusively reproduces cortical fates. We also found that cerebral organoids derived by combined inhibition display enhanced stem cell activity (Notch) that overlaps with efficient radial organization capacity (rosettes), and therefore assigned these features as readout for successful cortical specification. Finally, we have provided a proof of concept to this paradigm by showing that microcephaly organoids display massive cell death and significant reduction in cortex-specific gene expression only when derived under combined inhibition, thus establishing this method indispensable for demonstrating cortex-specific cell death.

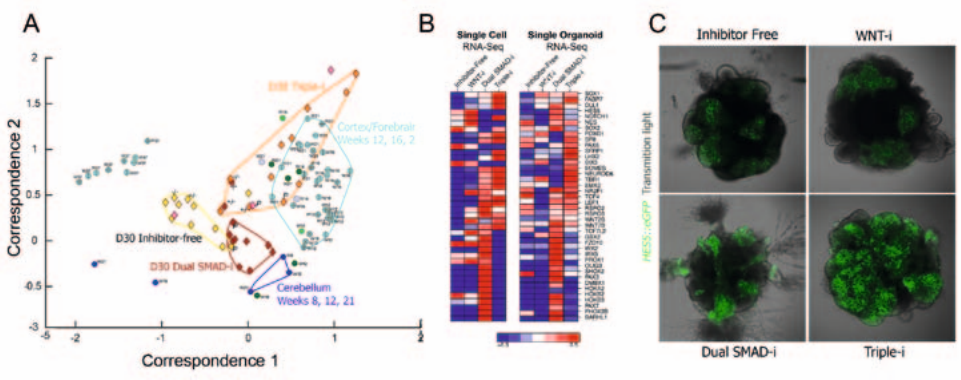


Figure 1.
Integrated analysis of brain organoids
A. RNA-Seq obtained from singly processed organoids derived by different methods and from human brain samples reveals that combined dual SMAD/WNT inhibition best reproduces cortical fates.
B. Remarkable similarity in anterior-posterior marker expression patterns between those obtained by single organoid bulk RNA-Seq or deconvoluted single cell RNA-Seq. Also here, combined dual SMAD/WNT inhibition best reproduces cortical fates.
C. Enhanced stemness (Notch activation; GFP) that also overlaps efficient radial organization (vesicle formation), serve together as readout for cortical specification.

Ongoing and future projects

Human neural stem cell ontogeny: Programming and reprogramming towards neural stem cell (NSC) fates

[Sneha Arora, Naresh Mutukula, Amèlia Aragonés Hernández]

Cellular reprogramming towards distinct pluripotency states and directed reprogramming within and across lineages reinvigorated the idea of developing novel strategies for inducing or maintaining homogeneous, self-renewing NSCs. These can act as "pluripotency" equivalents of the nervous system and therefore are expected to lead an exponential growth in the use of NSCs for various regenerative medicine applications. Taking advantage of our extensively characterized distinct NSC types established in our lab, we are building a platform to readout, evaluate and manipulate the regenerative potential of distinct NSC types *in vitro*. We combine reporter transgenics in pluripotent stem cells, cell sorting, bulk and single cell transcriptomic approaches, gain and loss of function studies and a variety of stem cell assays, to gain powerful readouts and determinants for reprogramming somatic cell populations towards specific NSC stages.

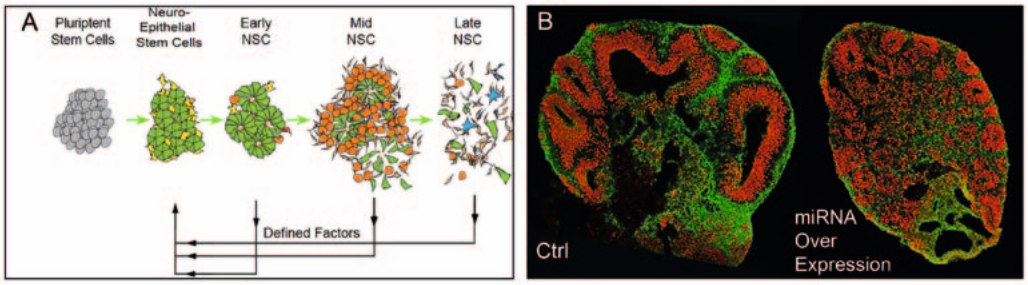


Figure 2.
Reprogramming NSC fates
A: Under "normal" non-reprograming conditions, progenitors progress through distinct stages *in vitro* (left to right). The image illustrates an example for reverting later progressing neural progenitors into early type neural progenitors by defined factors. Reprogramming is performed by biased and unbiased screens for molecules, TFs, shRNAs and micro-RNAs (miRNAs).
B: Impairment of cerebral organoid development induced by miRNA overexpression. Shown are day 30 organoids following miRNA overexpression, and their matched day 30 untreated organoids (Ctrl), stained for the neural stem/progenitor cell marker SOX2 (red) and the neuronal differentiation marker DCX (green).

Lineage fate mapping of cortical neural progenitors – towards comprehensive understanding of lineage relations of cortical NSCs

[Sneha Arora, Naresh Mutukula, Eveline Zeeuw van der Laan, Yakey Yaffe]

Lineage tracing provides a powerful tool for understanding tissue development and disease, especially when combined with manipulation of cell-fate decisions (see previous ongoing project). We are building a lineage tracing platform containing inducible recombinases, multicolor reporter constructs, single cell molecular barcode labeling and sequencing and live-cell imaging, to allow unbiased (ubiquitous) or targeted (promoter-driven) progenitor cell type-inducible genetic labeling. Such a system should enable genetic marking, monitoring, isolation and characterization of unique stem/progenitor cell subpopulations and their progeny during development.

Monitoring and analyzing centrosome dynamics – towards understanding the "clock" that drives changes in cell fate during cell division in pluripotency and differentiation

[Naresh Mutukula, Yakey Yaffe, Amèlia Aragonés Hernández]

While vast changes in cell fate and molecular identity in cortical NSCs occur over weeks or months, it is still not completely understood how these changes are translated at the single cell division level. Similarly, mechanisms of cell fate decisions taken at cell division during transition from the pluripotent stage towards lineage fates are also elusive. We built a photoconvertible centrosome reporter system that allows monitoring and characterizing changes in cell fate following cell division based on their centrosome age. While centrosome duplication and segregation are hallmarks of cell cycle progression, the inheritance of aged and young centrosomes to stem cells and differentiated cells can serve as a hallmark of asymmetrical cell division that leads to changes in cell fate. We can now

photoconvert centrioles in pluripotent stem cells and cortical NSCs and then track them till they are aged. We aim to identify and characterize aged cells from newly born cells, and discover novel determinants during stages of cortical differentiation as well as in pluripotency and lineage commitment.

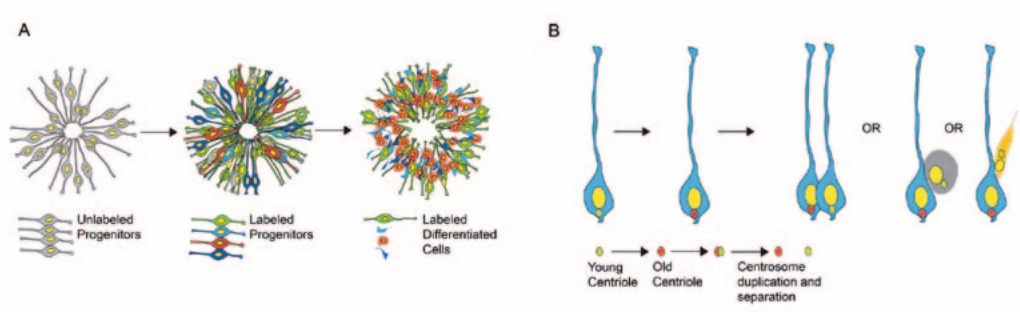


Figure 3.
Tracking NSCs in proliferation and differentiation
A: A paradigm for labeling cell fates in rosettes. Tracking progenitor cells will be allowed based on unbiased or targeted labeling (green, red, blue) according to specific progenitor markers.
B: Labeling based on centrosome age. Cells are photoconverted, following which new centrosomes (non-photoconverted) are duplicated and inherited to daughter cells (a progenitor, an intermediate progenitor, a neuron).

Origin of brain tumorigenesis: Molecular and cellular landmarks of medulloblastoma

[Eveline Zeeuw van der Laan, in collaboration with Thomas Risch, Marie-Laure Yaspo, Daniel Rosebrock, Peter Arndt, MPIMG]

Medulloblastomas (MB) are highly invasive malignant pediatric brain tumors with poor overall survival. MB subgroups exhibit remarkably distinct genomic and transcriptomic features, implying distinctive underlying biology. Yet, MBs are speculated to share common mechanisms with normal neural development. We found significant transcript enrichment of human MB samples in developmentally and molecularly distinct neural stem cell stages we established from human pluripotent cells, suggesting that cell populations within MBs are hierarchically ordered. The goal of this study is to identify, isolate and characterize tumor driving cancer stem cell populations during tumorigenesis. We are generating an *in vitro* pluripotent stem cell system that combines inducible expression of known deriver genes alongside with inducible labeling of key developmental genes suspected to act as cancer stem cells. We aim to model genomic, transcriptional and cellular events during initiation and progression of MB and reveal the interplay of these events to drive MB tumorigenesis.

Department of Computational Molecular Biology

The department has been established in October 2000.

Prof. Dr. Martin Vingron
vingron@molgen.mpg.de
+49 30 8413-1150

Secretary:
Martina Lorse
vinoffic@molgen.mpg.de
+49 30 8413 1151

Structure and organization of the department

The Department of Computational Molecular Biology comprises scientists coming from various backgrounds, ranging from mathematics, statistics and computer science via physics to biology and genetics. They are organized in several project groups, the largest of which is the Transcriptional Regulation Group headed by Martin Vingron. The work of this group focuses on theoretical concepts in regulation and epigenetics, pursues collaboration projects with experimental groups, and develops methods for analysis of sequencing data in the context of genetics and cancer genomics. Experimental aspects of regulation and epigenetics are pursued in the Mechanisms of Transcriptional Regulation Project Group of Sebastiaan Meijsing, who studies glucocorticoid receptor as a model transcription factor. Peter Arndt heads the Evolutionary Genomics Group, a research group within the department, which works on developing models how the DNA in humans and other non-human primates has evolved. Ralf Herwig's project group Bioinformatics concentrates on disease bioinformatics and biological networks.

Research concept of the whole department

With genome research in the focus of the Max Planck Institute for Molecular Genetics (MPIMG), computational biology has become an essential discipline to be represented at the Institute. The Department of Computational Biology combines theory and methods development with collaborative data analysis projects, and pursues biological questions through a combination of computational and experimental approaches. Methods from mathematics, statistics, and computer science form the basis for the bioinformatics analyses performed. Projects may either approach biological questions through theoretical analysis, or may be collaborative projects, where frequent interaction and feedback between experimentalists and theoreticians drive the work forward.

Material resources, equipment, and spatial arrangements

The theoretical work of the Department of Computational Molecular Biology relies heavily on powerful computers. The computer equipment of the department comprises nine compute servers with 48 or 64 processors and containing up to 1 TB of RAM. This architecture serves the classical sequence analysis, the numerical calculations, and the analysis of next generation sequencing data. Many computers are not accessed by the researchers directly, but through a queueing system. This allows for very efficient utilization of the entire infrastructure. Storage space on hard disks in the Institute comprises approximately 5 PB and the department participates in this. The department profits from the IT group's efforts in maintaining this computer system.

The department has its office space on the 3rd floor of tower 3. Sebastiaan Meijsing, who does experimental work, has laboratory space in tower 4.

Ongoing work and planned developments

While regulatory genomics and epigenetics have been a focus of the department for many years, current developments comprise a much closer integration between computation and experiment, both within the department and across groups in the MPIMG and beyond. Close ties have been built to the research group of Stefan Mundlos concerning the analysis of HiC data and with a plan to intensify the joint efforts on analysis of patient genomes. Other collaborative efforts involve the experimental group within the department (Sebastiaan Meijsing), the group of Yechiel Elkabetz (A. Meissner, Dept. of Genome Regulation), the group of Hans Schöler at the Max Planck Institute for Molecular Biomedicine in Münster, Germany, and the group of Marino Zerial at the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, Germany. Another reflection of this tighter integration between computation and experiment is the pending proposal for a new International Max Planck Research School for Biology and Computation, which shall connect the MPIMG with the Departments of Computer Science, Biology, and Biochemistry of the Freie Universität Berlin.

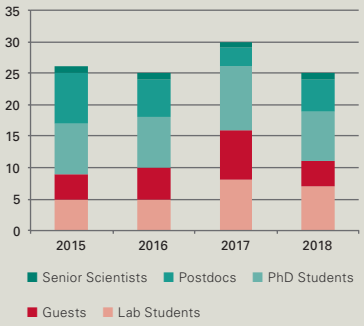
For more information, please see https://www.molgen.mpg.de/en/bioinformatics

Vingron Lab – Transcriptional Regulation Group

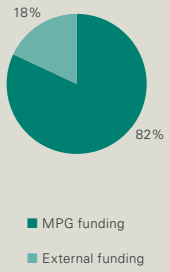
The group has been established in October 2000.

Prof. Dr. Martin Vingron
vingron@molgen.mpg.de
+49 30 8413-1150

Team



Funding



Focus areas

- Computational regulatory genomics and epigenetics
- Analysis of sequencing data in genetics and cancer genomics
- Transcriptome analysis and gene network reconstruction
- Collaborative data analysis with experimental groups

Research concept

The transcriptional regulation group pursues methods development and collaborative data analysis in the fields of transcription factor-based and epigenetic gene regulation, and works on analysis of genomic sequences in search for regulatory effects of mutations.

We try to closely integrate theoretical efforts with applications in a tight feedback loop, where concrete data analysis projects suggest methodical developments and new methods get tested in our collaborations. We develop methods for construction of biological networks and apply statistical and machine learning methods to a variety of data analysis and data integration tasks.

Scientific highlights

- Sequencing of the hexaploid sweet potato (*Ipomoea batatas*) with phasing of a large part of the genome based on Illumina sequencing [Yang et al. Nature Plants 2017].
- In a series of papers led by Stefan Mundlos, we contributed analyses of HiC data with respect to the effect of genomic changes and regulatory interactions [Andrey et al. Genome Res 2017; Franke et al. Nature 2016]

Scientific honours & selected invited talks

- Invited researcher at Simons Institute UC Berkeley, Program on "Algorithmic Challenges in Genomics", spring 2016
- RECOMB Test of Time Award 2018 for the most cited paper from RECOMB 2006
- Invited Speaker at Imperial College, London, April 2018
- Invited Speaker at Southern University of Science and Technology, Shenzhen, China, May 2017
- Invited Speaker for the Open Lecture at Simons Center, UC Berkeley, April 2016
- Invited Speaker at Ecole Normale Superieur, Paris, March 2015

Results of completed projects during reporting period

Haplotyping

[Hossein Moeinzadeh, Stefan Haas, Jun Yang]

In collaboration with a Chinese plant biology group we have sequenced a hexaploid plant, the sweet potato *Ipomoea batatas*, with most of the DNA sequencing done at MPIMG. While it was generally assumed that phasing such a complicated genome would be impossible, it turned out that this particular genome is very heterozygous and even short Illumina sequencing reads contain sufficient polymorphisms to attempt phasing. In the group, we developed the necessary algorithms for phasing and succeeded in assigning almost half of the genome to haplotypes.

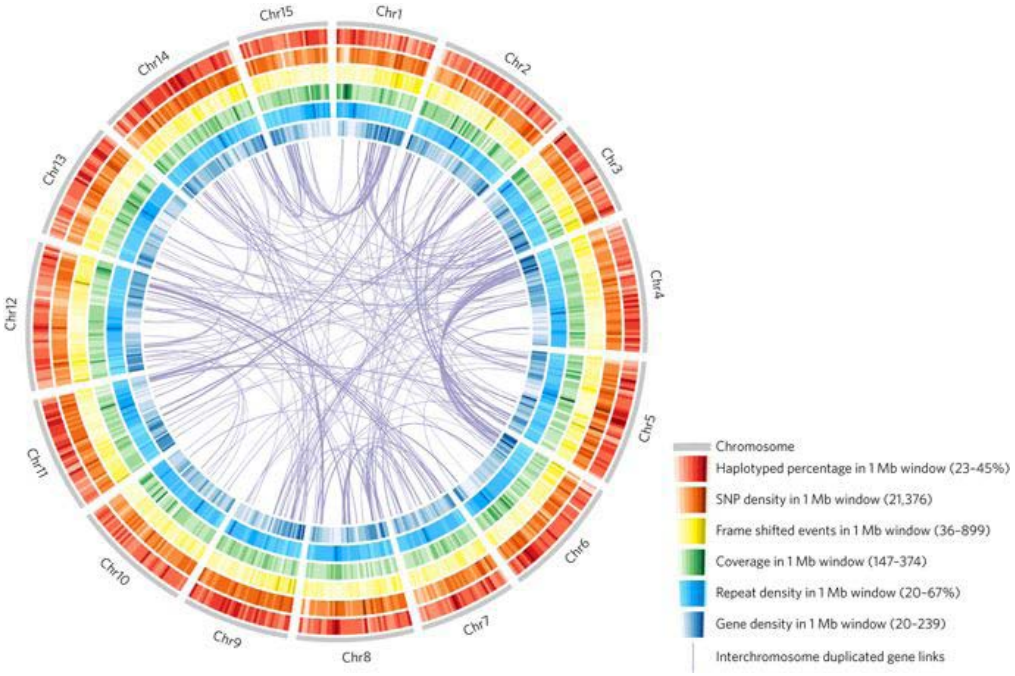


Figure 1. Summary of sweet potato genome. The outer circle quantifies the success of the phasing.

Yang J, Moeinzadeh M H, Kuhl H, Helmuth J, Xiao P, Haas S, Liu G, Zheng J, Sun Z, Fan W, Deng G, Wang H, Hu F, Zhao S, Fernie A R, Boerno S, Timmermann B, Zhang P & Vingron M (2017). Haplotype-resolved sweet potato genome traces back its hexaploidization history. Nat Plants (9):696-703.

3D Chromatin structure

[Robert Schöpflin, Verena Heinrich, Stefan Haas]

We have established a close collaboration with the group of Stefan Mundlos focusing on 3D chromatin structure. Schöpflin, Heinrich and Haas developed tools and provide data analysis support for analyzing 3D chromatin structure data with respect to implications for gene regulation. The emphasis lies on enhancer-promoter interactions, which are affected due to genomic rearrangements with subsequent changes in chromatin interactions.

Franke et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature 538: 265-269 (2016)

Sequence determinants of CpG-islands

[Matthew Huska]

We designed a classification method that predicts non-methylated islands (NMI) in different vertebrates. While in human and mouse, the information comes from CpG islands, in other organisms CpG dinucleotides are not sufficient to predict non-methylated islands. Rather, the NMIs in other organisms are defined by more complex patterns, which nevertheless allow for a good prediction of NMIs.

Huska M & Vingron M. Improved prediction of non-methylated islands in vertebrates highlights different characteristic sequence patterns. PloS Comput Biol 12(12):e1005249 (2016)

PWM enrichment statistics

[Wolfgang Koop]

Motif enrichment analysis (MAE) is a frequently occurring task, where one searches for sequence motifs which are enriched, e.g., in a set of jointly regulated promoters. The corresponding statistical problem lies in determining the distribution of the number of hits of a motif in a long sequence. Based on work of the group done earlier [Pape et al., J Comp Biol 2008], we have now further developed this statistics and can compute extremely accurate p-values even under higher-order Markovian background models. This development establishes a sound basis for a hitherto very heuristic method.

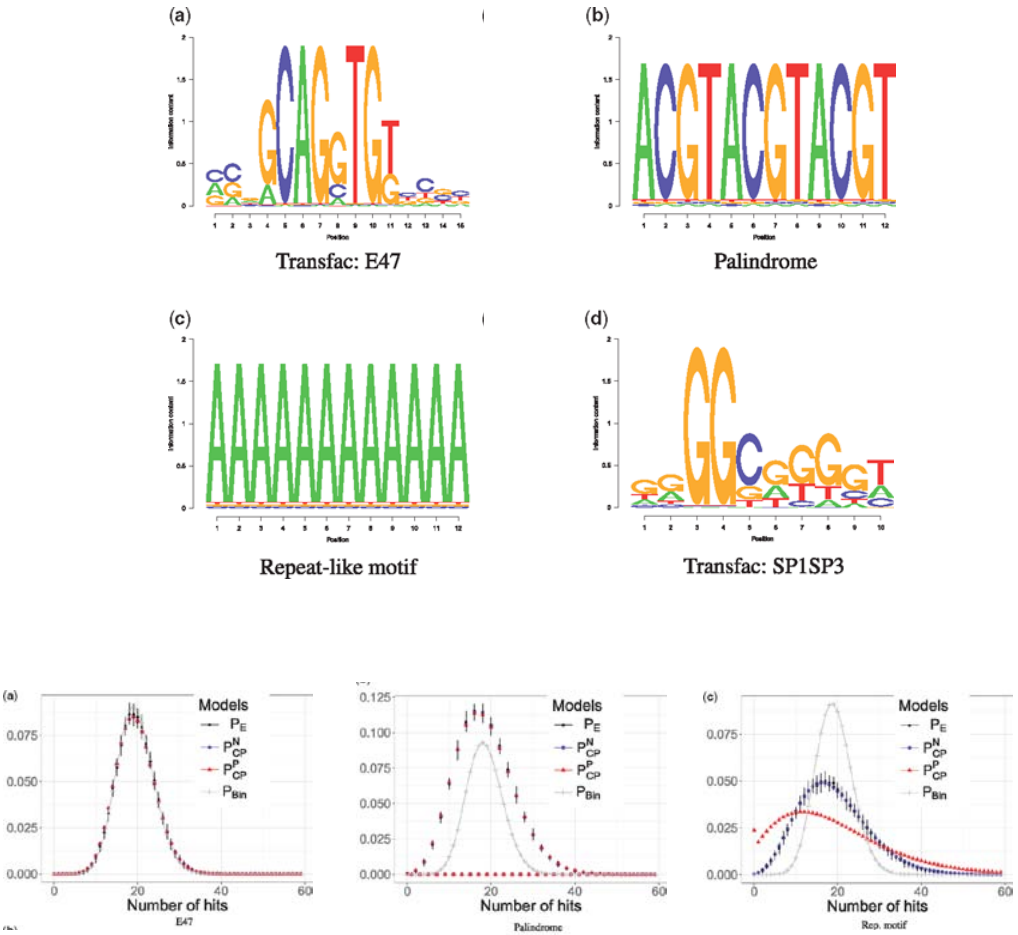


Figure 2.
Determining the statistics of motif hits is particularly difficult when a motif is palindromic or low complexity. The densities show the quality of the approximation to the empirical distribution.

Kopp W & Vingron M. An improved compound Poisson model for the number of motif hits in DNA sequences. Bioinformatics 33(24):3929-3937 (2017)

Reconstruction of biological networks

[Mahsa Ghanbari, Julia Lasserre, Ercan Kuruoglu, Alena van Bömmel, Edgar Steiger]

On the theoretical side of the biological networks, several new algorithms were developed. One algorithm is devoted to the question of integrating prior knowledge about gene interactions into the network reconstruction algorithm. Another method aims at delineating changes in network connectivity, when studying time-course expression data. Lately, we have worked on developing a new approach to network reconstruction that is based on "distance correlation" rather than Pearson correlation, which has the advantage that no linearity assumptions need to be made. We developed the DPM (Distance Precision Matrix) Method, which allows for gene network reconstruction even for non-linear relationships among gene. Networks of interactions among transcription factors in enhancer regions are the focus of another new method, which determines in a tissue-specific manner, which pairs of transcription factor binding motifs characterize the enhancers of that particular tissue. This was applied to provide a comprehensive analysis of regulatory interaction in mES cells.

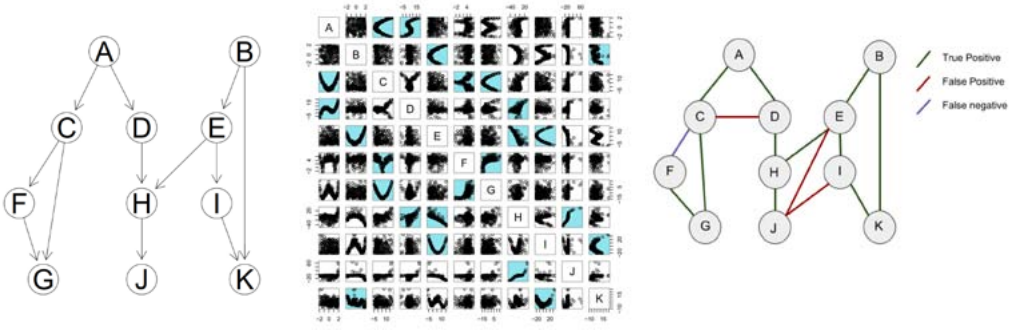


Figure 3.
Reconstruction of the given network on the left when provided with data shown in the middle. The coloured subplots show the simulated non-linear data for the network edges. Reconstructed network on the right.

Ghanbari M, Lasserre J & Vingron M. The Distance Precision Matrix: computing networks from non-linear relationships. Bioinformatics https://doi.org/10.1093/bioinformatics/bty724 (2018)

Transcriptome analysis of lung tumors

[Stefan Haas, Ruping Sun]

In a collaborative project led by Roman Thomas, University Cologne, we analyzed RNAseq samples of five lung tumor types with special emphasis on small-cell lung cancer (SCLC) and large-cell neuroendocrine carcinoma (LCNEC). Despite being histologically very similar the clustering of expression data of SCLC and LCNEC not only showed major differences in gene expression between this tumor types but also revealed two subgroups of LCNEC samples. We also found several tumor-type specific splice isoforms, including recurrent splice variants of transcription factors TP73 and E2F7, which are involved in cell-cycle regulation. In an additional project on SCLC patient-derived xenografts derived from circulating tumor cells we predicted candidate fusion transcripts along multiple passages of xenografts. This analysis showed that key fusion transcripts are robustly expressed even after several xenograft passages.

George J, ... [12 authors] ... , Sun R, ... [36 authors] ... Vingron M, ... [3 authors] ...Haas S A, Olivier M, Foll M, Buttner R, Hayes D N, Brambilla E, Fernandez-Cuesta L & Thomas R K. Integrative genomic profiling of large-cell neuroendocrine carcinomas reveals distinct subtypes of high-grade neuroendocrine lung tumors. Nat Commun 9:1048 (2018)

Regulatory networks in stem cells and reprogramming

[Wolfgang Kopp]

Earlier work on epigenetic regulation and epigenetic networks has continued. In collaboration with the group of A. Valencia we analysed a comprehensive epigenetic data-set for mouse ES cells and determined the network of interactions using the methods that had been developed in our group. Another collaboration with the lab of Hans Schöler focused on the role of Esrrb in reprogramming.

Adachi K, Kopp W, Wu G, Heising S, Greber B, Stehling M, Arauzo M J, Boerno S T, Timmermann B, Vingron M & Schöler H R (2018). Esrrb unlocks silenced enhancers for reprogramming to naive pluripotency. Cell Stem Cell 23(2):266-275.e6 (2018)

Ongoing and future projects

Regulatory sequence motifs and epigenetic regulation

[Xinyi Yang, Anna Ramisch, Tobias Zehnder, Verena Heinrich, Philipp Benner]

Work is continuing on regulatory sequence motifs and epigenetic regulation. One project provides an integrated analysis of promoter patterns together with ChIP-seq data for transcription factors and histone modifications. We find distinct groups of promoters characterized by certain TF combinations as well as promoters to which CTCF binds and apparently mediates promoter-enhancer looping. At the same time, we are developing machine learning methods to predict enhancers from epigenetic marks. Based on correlation to genes within the same topologically associated domain, these predicted enhancers are subsequently linked to potential target promoters/genes. In the future, hidden Markov models shall serve to integrate information about chromatin into a coherent picture about the regulatory state of a cell type or tissue.

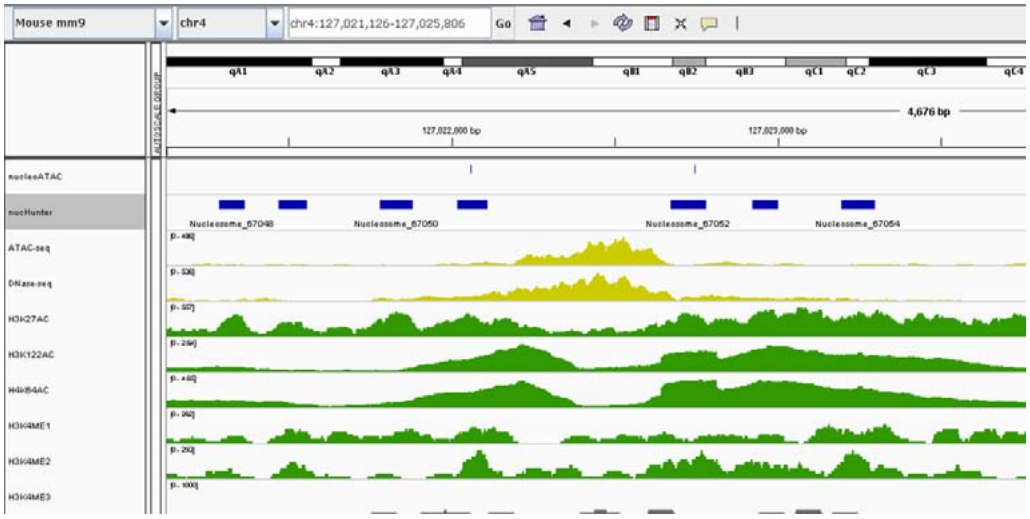


Figure 4.
Enhancers are characterized by a nucleosome free region that is delineated on either side by a nucleosome carrying the enhancer-typical histone modifications.

Sequence analysis

[Hossein Moeinzadeh, Stefan Haas, David Heller, Giuseppe Gallone]

Following our earlier work on sequencing and phasing, and mutation identification, the Mundlos and Vingron groups have joined their efforts in genome analysis. One project deals with the mole *Talpa occidentalis* which is a peculiar model organism for studying skeletal and gonadal development because the female moles develop ovaries as well as a testis-like tissue. We constructed an annotated draft genome assembly covering >17000 known genes. Genomic comparison with human and mouse revealed a triplication of a key enzyme involved in testosterone synthesis and we discovered a small group of genes involved in early development of kidney specifically expressed in the female testis. On the human side, we are currently focusing on the exploitation of long-read sequencing technologies for detection of structural variations in human individuals. In the long run, we aim at combining diverse technologies for the purpose of extensive characterization of individual genomes.

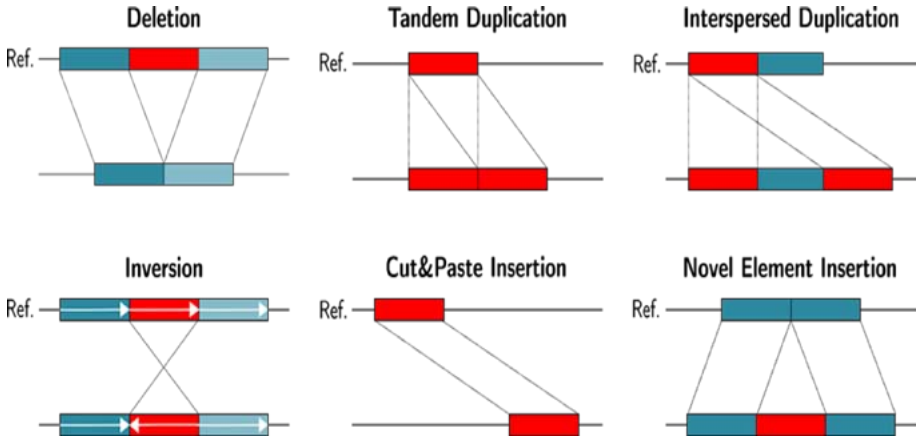


Figure 5.
The different kinds of genomic structural variation to be identified from sequencing data.

Organoids and single cell transcriptomics

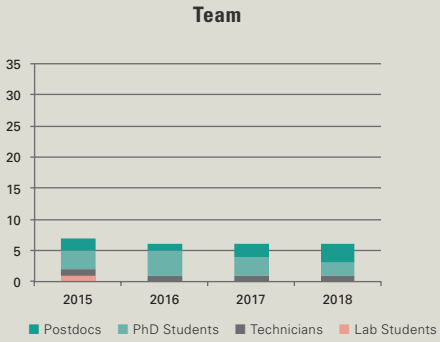
[Ela Gralinska, Lam-Ha Ly, Virginie Stanislas]

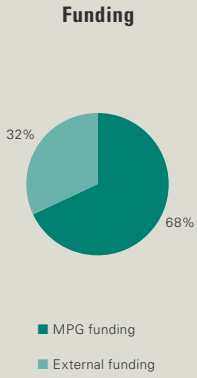
In collaboration with Y. Elkabetz from the Meissner department, we are analysing transcriptomes of brain organoids. New methods are being developed to compare publicly available tissue transcriptomes to the ones generated from organoids as well as their individual single-cell transcriptomes. For this purpose, we are currently exploring normalization and visualization methods (mostly correspondence analysis) that allow us to bridge from single cells to organoids to tissues.

Meijsing Lab – Mechanisms of Transcriptional Regulation Group

The group has been established in September 2009.

Dr. Sebastiaan H. Meijsing
meijsing@molgen.mpg.de
++49 30 8413-1176





Focus areas

- *Functional genomics* – principles of productive TF:DNA interactions
- *Structural biology* – DNA as an allosteric ligand
- *Epigenetics* – transgenerational inheritance of induced chromatin accessibility changes

Research concept

Using the glucocorticoid receptor (GR) as a model transcription factor whose activity is strictly controlled by the presence of hormone, my group uses integrative experimental, structural and computational modeling approaches, with the ultimate goal to qualitatively and quantitatively explain the genome-wide transcriptional consequences of GR signaling. Specific aims include: (i) Understanding how the integration of signals, especially the sequence of DNA binding site, fine-tune the activity of GR towards individual target genes; (ii) Linking genomic GR binding to the regulation of endogenous target genes; (iii) Identifying mechanisms that specify where and when GR target genes are expressed and eventually how they contribute to GR's physiological and therapeutical roles.

Scientific highlight

- Discovery that cell-type specific wiring of promoter-enhancer contacts can direct the activity of an enhancer towards distinct promoters [Thormann et al. Nucleic Acids Res 2018]

Scientific honours & selected invited talks

- Sebastiaan Meijsing, Invited speaker at the New Roles for Nuclear Receptors in Development, Health and Disease meeting, Cancun, Mexico, 2018
- Stefanie Schöne, Elisabeth-Gateff-Prize (PhD prize of the German Genetics Society), 2017
- Sebastiaan Meijsing, Invited speaker at the Nuclear Receptor Network meeting, Amsterdam, 2016
- Sebastiaan Meijsing, Associate member Epigenesys, 2013-2016
- Stefanie Schöne, L'Oreal prize for Woman in Science, 2015

Results of completed projects during reported period

DNA as an allosteric modulator of GR structure and activity

[Stefanie Schöne, Marcel Jurk]

Genes are not simply turned on or off, but their expression is fine-tuned. We study the role of one signal that modulates activity: the exact sequence of the DNA to which GR binds. Computational analysis showed that the nucleotides directly flanking the core-binding site differ depending on the strength of GR-dependent gene activation. Follow-up experiments showed that these flanking nucleotides change the three-dimensional structure of the DNA and of the associated GR protein and that the sequence-induced changes in GR activity cannot be explained by differences in GR occupancy. Rather, mutations that mitigate DNA-induced changes in structure also blunt changes in activity, arguing for a role of DNA-induced structural changes in modulating GR activity downstream of binding.

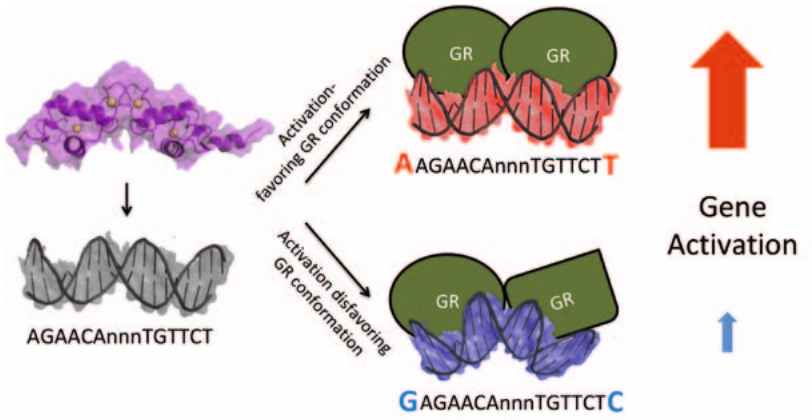


Figure 1.
DNA as an allosteric ligand.

Schöne S, Jurk M, Helabad M B, Dror I, Lebars I, Kieffer B, Imhof P, Rohs R, Vingron M, Thomas-Chollier M & Meijsing S H. Sequences flanking the core-binding site modulate glucocorticoid receptor structure and activity. Nat Commun 7: 12621 (2016)

The role of the genetic and "epigenetic" landscape in guiding GR to specific genomic loci

[Stephan Starick, Jonas Telorac]

Binding-site motifs of TFs are ubiquitously found in the genome whereas binding only occurs at a subset of putative sites. Whereas most studies focus on the identification of sequences that recruit TFs to specific genomic loci, we identified and characterized negative regulatory sequences that interfere with DNA binding of GR. Contrary to our expectation, we found that these sequences exert their effect by mechanisms other than chromatin accessibility, possibly involving anchoring to sub-nuclear regions that are less permissive to TF binding [Telorac

et al. Nucleic Acids Res 2016]. In addition to sequence, chromatin plays a key role in specifying the genomic binding pattern of TFs. One interesting and surprising finding we made was that GR binding is depleted at promoter regions. Sequence composition can explain part, but not all, of this depletion indicating that active mechanisms exist that prevent GR from binding at promoter-proximal regions.

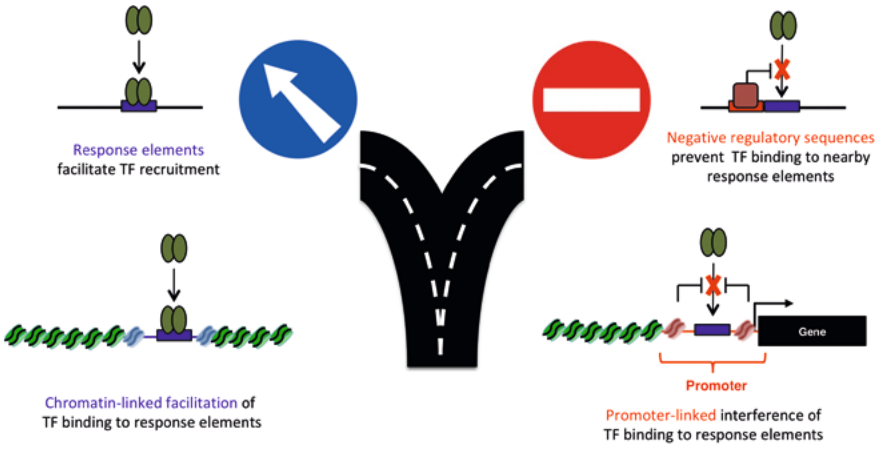


Figure 2.
DNA sequences and chromatin dictate the genomic binding pattern of GR.

Love M I, Huska M R, Jurk M, Schöpflin R, Starick S R, Schwahn K, Cooper S B, Yamamoto K R, Thomas-Chollier M, Vingron M & Meijsing S H. Role of the chromatin landscape and sequence in determining cell type-specific genomic glucocorticoid receptor binding and gene regulation. Nucleic Acids Res 45: 1805-1819 (2017)

Linking GR binding to the transcriptional regulation of genes

[Verena Thormann]

Using genome-editing and by integrating information regarding chromatin state and the 3D organization of the genome, we identified features that distinguish "productive" (resulting in the regulation of a gene) from "non-productive" GR binding events. For example, the simultaneous presence of a cluster of GR binding sites is required for the activity of certain enhancers. Further, by deleting GR binding sites that are shared between different cell types; we found that cell type-specific genome organization and enhancer-blocking can result in cell type-specific wiring of promoter-enhancer contacts. This rewiring allows an individual GR binding site to direct the expression of distinct transcripts and thereby contributes to the cell type-specific consequences of glucocorticoid signaling.

Thormann V, Rothkegel M C, Schöpflin R, Glaser L V, Djuric P, Li N, Chung H R, Schwahn K, Vingron M & Meijsing S H. Genomic dissection of enhancers uncovers principles of combinatorial regulation and cell type-specific wiring of enhancer-promoter contacts. Nucleic Acids Res 46: 2868-2882 (2018)

Ongoing and future projects

Understanding the principles of productive TF:DNA interactions

[Verena Thormann, Laura Glaser]

TFs can bind to tens of thousands of genomic binding sites, yet they seem to regulate a much smaller number of genes. We combine genome-wide and focused studies with individual enhancers to unravel molecular mechanisms that facilitate "productive" TF interactions (Fig. 3). For example, we developed a massively parallelized reporter assay to quantify the intrinsic activity of tens of thousands of enhancers. Furthermore, using computational approaches, we identified candidate features (e.g. sequence motifs) associated with active enhancers. To test the relevance of identified features, we will study the

effect of their deletion on gene regulation ("reverse engineering"). In parallel, we generate synthetic genomic enhancers ("forward engineering") to determine if we really understand the operating principles of productive TF binding.

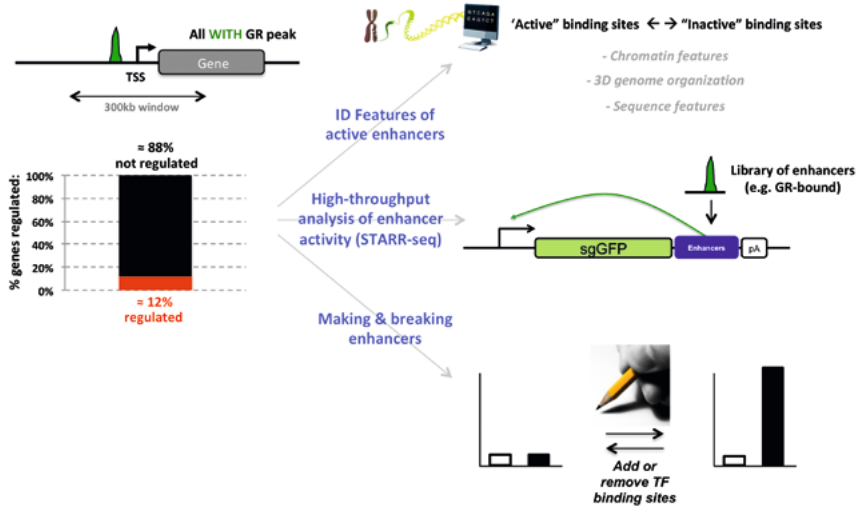


Figure 3.
What defines active enhancers?

Specificity of related TFs

[Marina Borschiwer, Melissa Bothe]

Paralogous TFs often share similar DNA sequence preferences, yet regulate distinct genes and execute unique functions *in vivo*. For example, GR and the androgen receptor (AR) recognize essentially the same sequences, yet whereas AR stimulates muscle gain, GR promotes muscle wasting. To address this paradox, we mapped genome-wide binding and changes in gene expression and chromatin landscape in response to activation of either receptor. Furthermore, we mapped the intrinsic enhancer activity of the majority of regions bound by both receptors. Consistent with previous studies, we find that receptor-specific binding can explain some of the differences in gene regulation. Interestingly, we also identified cases where the specificity appears to arise downstream of binding, given that the genomic loci responsible for the regulation of a gene are occupied by both receptors, yet only one is capable of activating the gene. Ongoing efforts aim at understanding what directs this specificity downstream of binding.

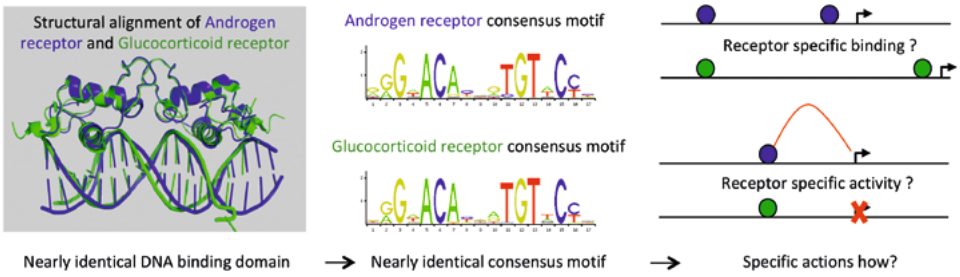
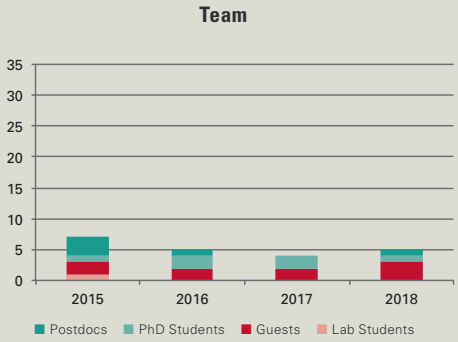


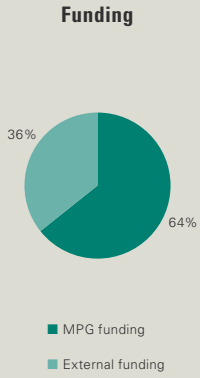
Figure 4.
Understanding what drives specificity of paralogous TFs

Herwig Lab – Bioinformatics Group

The group has been established in February 2001 at the former Dept. of Vertebrate Genomics and moved to the Dept. of Computational Molecular Biology in December 2014.

Dr. Ralf Herwig
herwig@molgen.mpg.de
++49 30 8413-1587





Focus areas

- *Data integration* – analysis and functional interpretation of methylome, transcriptome and proteome data
- *Network propagation* – inferring genotype-phenotype relations in diabetes mellitus and cancer
- *Machine learning* – drug sensitivity prediction in precision medicine

Research concept

Research covers i) the development of computational methods for the analysis of molecular data often in collaborative research projects and ii) the integration and interpretation of these data in the context of biological networks. The group has developed computational methods for RNA-seq and MeDIP-seq and works on the integrative analysis of these data in order to elucidate the interplay of methylation, gene expression and genome structure that are operative in human (disease) processes related to cancer, diabetes and drug toxicity. Furthermore, the group has developed and maintains the ConsensusPathDB molecular interaction resource, which is a widely used research resource (~910 citations). We have published 19 scientific publications in the reporting period.

Scientific highlights

- Identification of a predictive marker for carboplatin resistance in lung cancer therapy [Grasse et al. Genome Medicine 2018]
- QSEA – a novel software for quantifying methylation enrichment experiments [Lienhard et al. Nucleic Acids Res 2017]
- ConsensusPathDB - a resource for molecular interaction networks [Herwig et al. Nature Protocols 2016]

Scientific honours & selected invited talks

- Invited seminar talk at Cologne Center for Genomics, Cologne, 2018
- Invited talk at the German Biotechnology Days, Hannover, 2017
- Invited talk at the 55th Annual Meeting of the Society of Toxicology USA, New Orleans, 2016
- Invited seminar talk, Stuttgart Research Center Systems Biology, Stuttgart, 2015
- Invited talk at the Tumor Liquid Biopsy Symposium, Freiburg, 2015

Results of completed projects during reporting period (selected)

Epigenomics of pre-clinical models of non-small cell lung cancer

[Matthias Lienhard, Ralf Herwig in collaboration with Anna Ramisch, Martin Vingron, Vingron lab, and Michal Schweiger, University Cologne]

Non-Small Cell Lung Cancer (NSCLC) is primarily treated with radiation, surgery and platinum-based drugs like carboplatin and the major challenge is intrinsic or acquired resistance to chemotherapy. Molecular markers predicting the outcome for the patients are urgently needed. In the BMBF-funded project EPITREAT we employed patient-derived xenografts (PDX) to detect predictive methylation biomarkers for platin-based therapies. Using MeDIP-seq we generated genome-wide DNA methylation profiles of PDXs and identified a set of candidate regions with methylation correlated to carboplatin response and corresponding inverse gene expression pattern even before therapy. This analysis led to the identification of a promoter CpG island methylation of LDL Receptor Related Protein 12 (LRP12) associated with increased resistance to carboplatin. Validation in an independent patient cohort confirmed that LRP12 methylation status is predictive for therapeutic response of NSCLC patients to platin therapy with a sensitivity of 80% and specificity of 84% ($p < 0.01$). Additionally, we find a shorter survival time for patients with LRP12 hypermethylation in the validation cohort as well as in the TCGA cohort for NSCLC (LUAD) confirming the results.

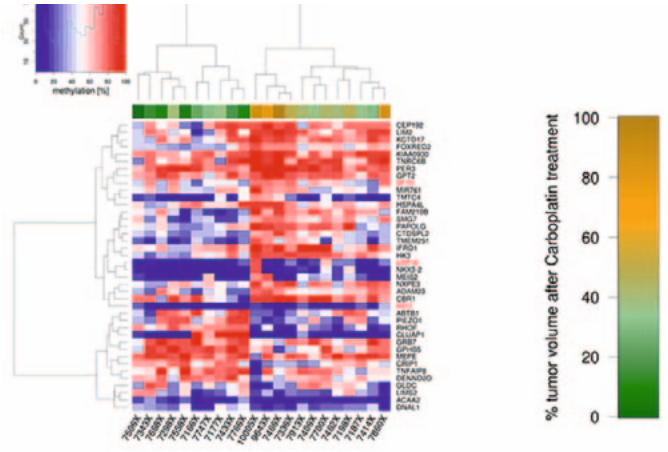


Figure 1.
Response DMRs in gene promoters that correlate with tumor volume of PDXs after therapy.

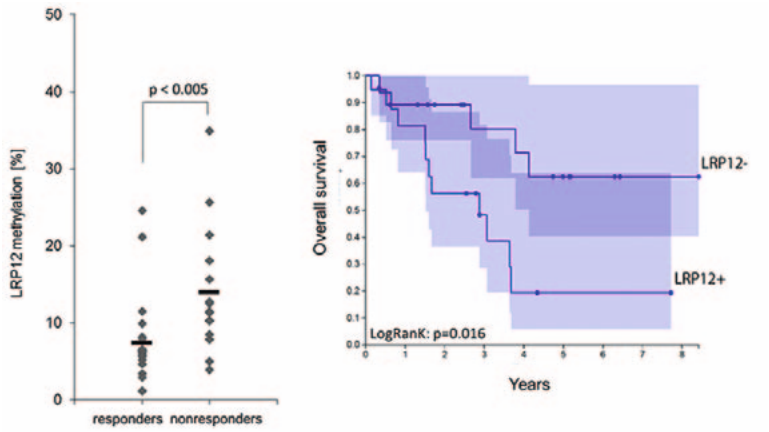


Figure 2.
LRP12 marker validation in independent patient tumors (left); Kaplan-Meier survival curve (right)

Grasse S*, Lienhard M*, Freese S, Steinbach A, Grimm C, Hussong M, Rolff J, Becker M, Dreher F, Schirmer U, Boerno S, Ramisch A, Leschber G, Timmermann B, Odenthal M, Büttner R, Grohé C, Lüders H, Vingron M, Fichtner I, Kein S, Sültmann H, Lehrach H, Herwig R* & Schweiger M*. Epigenomic profiling of Non-Small Cell Lung Cancer (NSCLC) xenografts uncover LRP12 DNA methylation as predictive biomarker for carboplatin resistance. Genome Medicine 10: 55 (2018)
*equal contribution as first and last authors

Statistical modelling of whole-genome methylation enrichment experiments

[Matthias Lienhard, Ralf Herwig, in collaboration with Stefan Börno, Bernd Timmermann, Sequencing Core Facility, and Michal Schweiger, University of Cologne]

Genome-wide enrichment of methylated DNA followed by sequencing (MeDIP-seq) offers a reasonable compromise between experimental costs and genomic coverage. However, the corresponding read-out of the experiments is qualitative only and quantification of the enrichment signals in terms of absolute levels of methylation requires specific transformation. We have developed a Bayesian statistical model that transforms the enrichment read counts to absolute levels of methylation and, thus, enhances interpretability and clinical application. We compared the method with competing models and show that the QSEA workflow outperforms other approaches and retrieves well-known lung tumor methylation markers that are causative for gene expression changes, demonstrating the applicability of QSEA for clinical studies. QSEA is implemented in R and available from the Bioconductor repository 3.4.

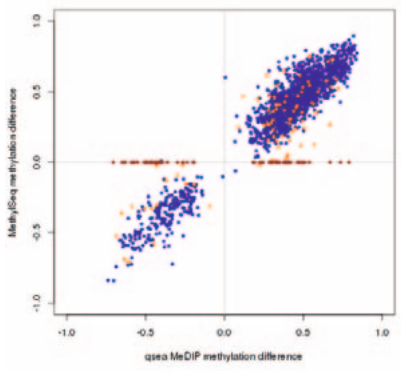


Figure 3.
Agreement of QSEA-estimated % methylation differences between PDX and normal tissues (X-axis) derived from MeDIP-seq data and % methylation differences measured with bisulfite experiments (Y-axis).

Lienhard M*, Grasse S*, Rolff J, Frese S, Schirmer U, Becker M, Boerno S, Timmermann B, Chavez L, Sueltmann H, Leschber G, Fichtner I, Schweiger M* & Herwig R*.QSEA-modelling of genome-wide DNA methylation from sequencing enrichment experiments. Nucleic Acids Res 45: e44 (2017)
*equal contribution as first and last authors

Ongoing and future projects (selected)

Dynamic adverse anti-cancer drug response at multi-omics levels

[Gal Barel, Matthias Lienhard, Ralf Herwig, in collaboration with Jos Kleinjans, University Maastricht, The Netherlands, Ralph Schlappbach, ETH Zurich, Switzerland]

Anthracyclines are widely used as anti-cancer therapeutics despite of the fact that they lead to severe cardiotoxic effects in many patients. The mechanisms of anthracycline-induced cardiotoxicity are most likely multifactorial but remain largely unclear. In a European Consortium funded by Framework 7 we have analyzed human cardiac microtissues, mimicking essential physiological functions of the heart, and performed time-resolved measurements of most-widely used anthracyclines at the proteome, transcriptome and methylome levels. We identified essential molecular players of adverse drug response affecting sarcomere and mitochondrial functions. Furthermore, we agglomerated a large molecular interaction network from the ConsensusPathDB and used network propagation as concept for data integration and show that the overlay of multi-omics genome data onto molecular interaction networks increases the functional information compared to each single approach.

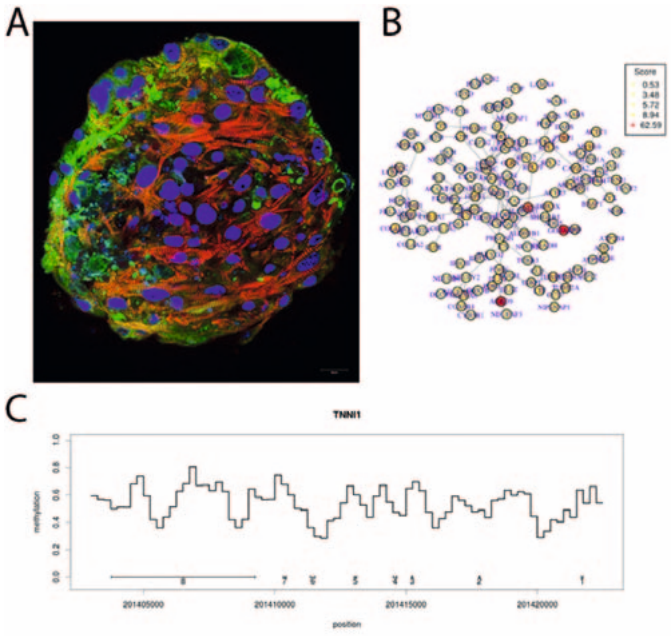


Figure 4.
A) iPS-derived cardiac microtissues (InSphero GmbH). B) Network propagation of temporal quantitative proteomics data is used to identify response network modules of proteins. C) Gene body methylation of essential sarcomere proteins (here TNNI1) lead to changes in gene and protein expression.

Herwig R, Hardt C, Lienhard M & Kamburov A. Analyzing and interpreting genome data at the network level with ConsensusPathDB. Nat Protoc 11(10):1889-1907 (2016)

Genoytpe-phenotype correlations in type-2 diabetes mellitus

[Gal Barel, Ralf Herwig, in collaboration with Hadi Al-Hasani, German Diabetes Center Düsseldorf]

Previous genetic studies in patients have shown that multi-factorial disease phenotypes, such as diabetes mellitus (DM), cannot adequately be explained by a single or few variants but rather agglomerate contributions of multiple variants of genes organized in molecular networks. In this project we analyze patient variant profiles along with detailed phenotypic characterization of patients provided by the German Diabetes Cohort (GDC) Consortium at the level of biological networks in order to allow the differentiation of patients into subgroups according to secondary complications. These subgroups and comorbidities can be associated with characteristic genetic variant profiles and molecular interaction networks that determine the genotype-phenotype network of DM.

Machine learning methods for precision medicine

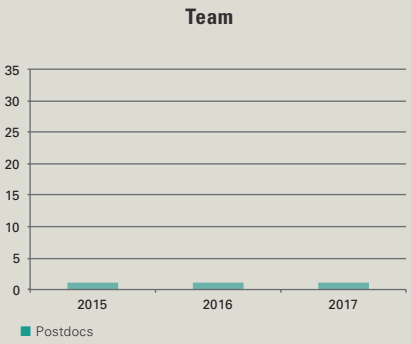
[Ralf Herwig, in collaboration with Tobias Scheffer, University Potsdam]

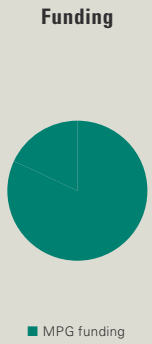
Machine learning approaches have emerged as the state-of-the-art methodology to infer predictions from large-scale data with numerous applications in science and economy. The group recently has successfully applied to the BMBF call on "Machine learning" with a research project that aims to develop novel machine learning methods and to apply these to the prediction of cancer therapy success in precision medicine. The goal of the project is to learn the sensitivity of the drug response of a biological system (cell line, patient tumors) from its molecular features and their relationships. Novel methods will develop constraints that allow better inclusion of background knowledge on biological pathways and gene-gene relationships.

Hoehe Lab – Diploid Genomics Group

Margret Hoehe, MD, PhD
hoehe@molgen.mpg.de

The group has been established in February 2002 at the former Dept. of Vertebrate Genomics and moved to the Dept. of Computational Molecular Biology in December 2014. Due to the retirement of M. Hoehe, the group was closed in December 2017.





Focus areas

- Human genetic variation, haplotypes, *cis*- versus *trans*-configurations
- Large scale analysis of experimentally and statistically haplotype-resolved genomes
- Diploid genome, exome, proteome analysis
- Functional interpretation/significance of phase

Research concept

To fully understand human genetic variation and its functional consequences, the specific distribution of variants between the two chromosomal homologues of (diploid) genes and any functional units of importance must be known. The phase of variants can significantly impact gene function and phenotype. Focusing initially on the development and application of novel molecular genetics and bioinformatics approaches to haplotype-resolve whole genomes, we have aimed during our final years to address key questions concerning the diploid nature of human genomes and its functional implications computationally on a larger scale. We have expanded on preliminary key results reported by Hoehe, Church, Lehrach et al., Nature Communications 2014, performing analyses of 1000 Genomes data and other world-wide unique data resources (PGP).

Scientific highlight

- Establishment of significant abundance of *cis* configurations of mutations as a universal characteristic of diploid human genomes and distinction of two functionally different categories of autosomal protein-coding genes, "*cis*-" and "*trans*-abundant genes" [Hoehe et al. bioRxiv, 2017]

Scientific honours

- Invited speaker, 6th Annual Next Generation Sequencing Asia Congress, Singapore, 2016
- Invited speaker at the MPI for Experimental Medicine, Göttingen, 2015
- Invited talk at GET Conference, Vienna, 2015
- Invited speaker, 5th Annual Next Generation Sequencing Asia Congress, Singapore, 2015

Results of completed projects during reporting period

Establishment of significant abundance of *cis* configurations of mutations in diploid human genomes and distinction of "*cis*-" and "*trans*-abundant genes"

[Margret Hoehe, Thomas Hübsch, in collaboration with Ralf Herwig, MPIMG; George M. Church, Harvard Medical School & Wyss Institute; Radoje Drmanac, Brock Peters, Complete Genomics, Inc.; and Qing Mao, BGI-Shenzen]

This project expands on preliminary key results reported by Hoehe et al. Nat Commun 2014. Results are based on systematic large scale analysis of 18,121 autosomal protein-coding genes in 1,092 statistically phased genomes from the 1000 Genomes Project and an unprecedented number of 184 experimentally phased genomes from the Personal Genome Project (PGP). Mutations predicted to functionally alter the protein and coding variants as a whole were found to exist significantly more frequently in *cis*- than *trans*-configurations, with global *cis/trans* ratios of ~60:40. Significant *cis*-abundance was observed in virtually all individual genomes in all populations. This phenomenon was largely constituted by a shared, global set of phase-sensitive genes, which was significantly enriched with gene sets indicating its involvement in adaptation and evolution. At the bottom of *cis*-abundance were two major categories of autosomal genes, a large group of "*cis*-abundant" genes, which exhibited *cis* configurations of functionally protein-altering mutations in significant excess, and a smaller group of "*trans*-abundant" genes (figure 1). These two gene categories were found functionally distinguishable, and exhibited strikingly different distributional patterns of protein-altering mutations (figure 2). First evidence for potential (physical) interaction of closely spaced mutations in *cis* within tertiary protein structure was obtained. Thus, the distinction of *cis* and *trans* configurations ultimately led to the classification of variable autosomal genes into two major, potentially functionally divergent, categories. Our work exposes the importance of phase for the interpretation of protein-coding genetic variation and the description and analysis of diploid genes, as well as its functional significance.

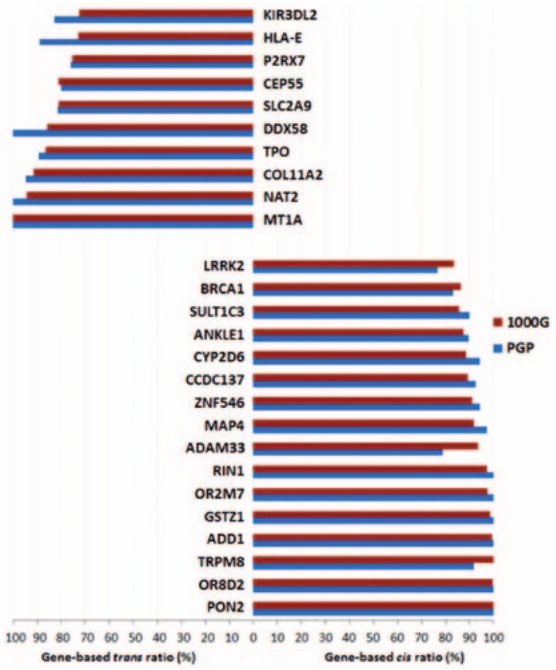


Figure 1.
Examples of *cis*- (right) and *trans*-abundant genes (left) shared by 1000 Genomes and PGP. Gene-based *cis* ratio defined as number of *cis* configurations counted in a gene in a defined sample of genomes, divided by total configuration count of this gene in the sample; gene-based *trans* ratios defined accordingly.

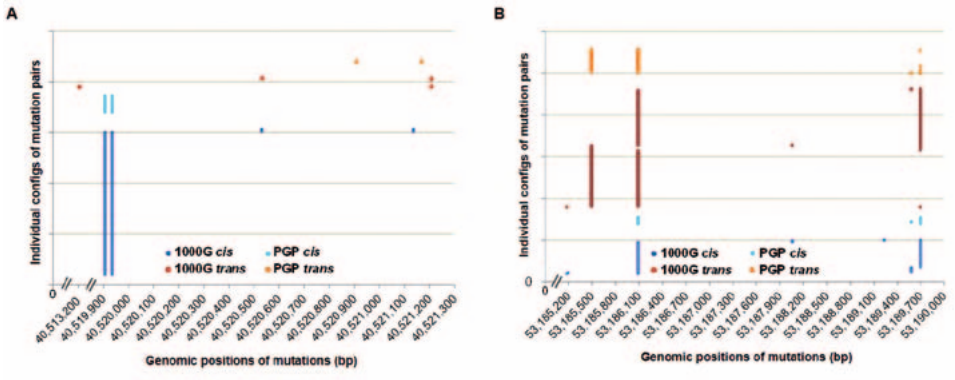


Figure 2.
Pairs of functionally protein-altering mutations in *cis*-abundant genes such as ZNF546 (left) are much more closely spaced than mutations in *trans*-abundant genes (right), such as KRT3. *Cis*-abundant genes are characterized by one highly frequent, "major", pair of mutations.

Hoehe MR, Herwig R, Mao Q, Peters BA, Drmanac R, Church GM & Huebsch T. Significant abundance of *cis* configurations of mutations in diploid human genomes. bioRxiv 221085; doi: https://doi.org/10.1101/221085 (2017)

Genetics, genomics and pharmacogenomics of neuropsychiatric disorders

[Margret Hoehe, in collaboration with Deborah Morris-Rosendahl, Royal Brompton Hospital and University College London, UK]

Review of results/state-of-the-art to assess potential clinical applications toward precision medicine, such as genetic testing for diagnostic purposes as well as the prediction of disease risk and individual response to drug treatment.

Hoehe MR & Morris-Rosendahl DJ. The role of genetics and genomics in clinical psychiatry. Dialogues Clin Neurosci 20(3): 169-177 (2018)

Clinical genetic testing in psychiatric disorders

[Margret Hoehe]

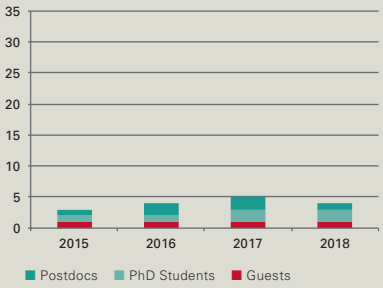
Margret Hoehe is a member of the International Society of Psychiatric Genetics (ISPG) Taskforce on Clinical Genetic Testing and its Writing Committee, an expert panel aiming to review the available evidence from existing literature to provide guidance for the medical community and the general public on the use of genetic testing in diagnostic and treatment decisions.

Arndt Lab – Research Group Evolutionary Genomics

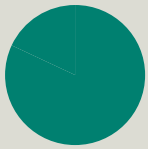
The group has been established in October 2003.

Dr. Peter Arndt
arndt@molgen.mpg.de
++49 30 8413-1162

Team



Funding



Focus areas

- Molecular biology and evolution
- Context-dependent models of sequence evolution
- Dynamical models
- Parameter and model inference
- Cancer genomics

Research concept

Evolutionary processes responsible for the diversification and adaptation of species (or cell populations) leave distinctive marks in their genomes. These marks encompass a multitude of DNA alterations: single nucleotide exchanges, insertions and deletions of short segments, insertions of repeats, segmental duplications, and whole genome duplications. Such genomic changes accumulate not only in the genomes of species over millions of years but also in cancerous cell populations within an individual on a much shorter timescale. The rapidly increasing availability of genomic sequence data now allows describing and understanding the complex phenomena of genomic evolution in a much higher resolution by applying novel and innovative computational methods and mathematical concepts.

Scientific highlights

- Elucidating large scale variations of mutation rates and other features in the human genome [Smith et al. PLoS Genet 2018].

Scientific honours

- Associate Scientist at the Broad Institute of MIT and Harvard University, Boston, since 2015
- Member of the Society of Molecular Biology and Evolution and the German Physical Society
- Invited speaker at conference "Mathematical and Computational Evolutionary Biology", Porquerolles, France, 2015

Results of completed projects during reporting period (selected)

Evolutionary dynamics of selfish DNA explains the abundance distribution of genomic subsequences

[Michael Sheinman, Florian Massip, Peter Arndt, in collaboration with Anna Ramisch, MPIMG]

One intriguing feature of large genomes is that certain subsequences are much more abundant than others. In fact, abundances of subsequences of a given length are distributed with a scale-free power-law tail that resembles a Zipf's law observed in multiple different data types, including human literary texts. We found that selfish DNA elements, such as those belonging to the Alu family of repeats, dominate the power-law tail. Interestingly, the power-law exponent for the Alu elements increases with the length of the considered subsequences. Motivated by these observations, we developed a model of selfish DNA expansion. The predictions of this model qualitatively and quantitatively agree with the empirical observations. The obtained results of our model shed light on how evolution of selfish DNA elements shapes the non-trivial statistical properties of genomes.

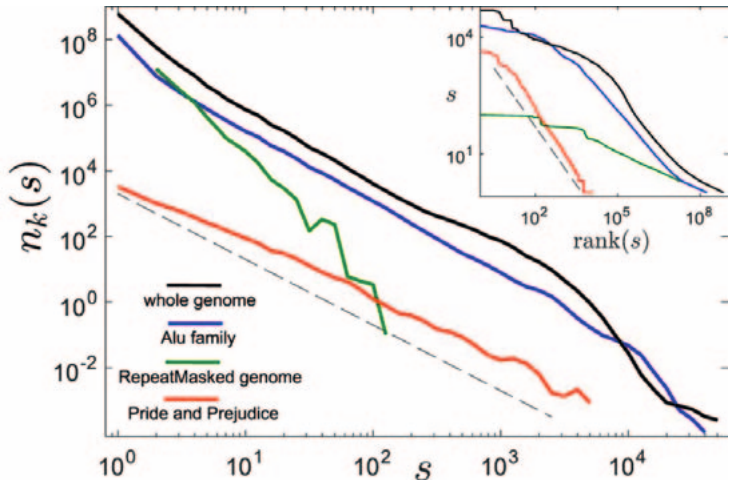


Figure 1.
Distributions of abundances of k-mers for k = 40 in the human genome and in the novel Pride and Prejudice by Jane Austen. s is the number of copies of a certain k-mer and $n_k(s)$ is the number of different k-mers with abundance s.

Sheinman M, Ramisch A, Massip F & Arndt P F. Evolutionary dynamics of selfish DNA explains the abundance distribution of genomic subsequences. Sci Rep 6:30851 (2016)

Large-scale variation in the rate of germline *de novo* mutations, base composition, divergence and diversity in humans

[Peter Arndt, in collaboration with Adam Eyre Walker, University of Sussex]

Using a unique dataset of more than 130,000 *de novo* mutations we found a large-scale variation in the mutation rate at 100KB and 1MB scales. We observed that different types of mutations vary in concert and in a manner that is not expected to generate variation in base composition. Hence, mutation bias is not responsible for the large-scale variation in base composition observed across human chromosomes. As expected, large-scale variation in the rate of divergence between species and the variation within species across the genome are correlated to the rate of mutation. However, the correlation between divergence and the mutation rate is not as strong as it could be. We showed that biased gene conversion is responsible for weakening the correlation. In contrast, we find that most of the variation across the genome in diversity can be explained by variation in the mutation rate. Most interestingly, we found that the correlation between the rate of mutation in humans and the divergence between humans and other species weakens as the species become more divergent.

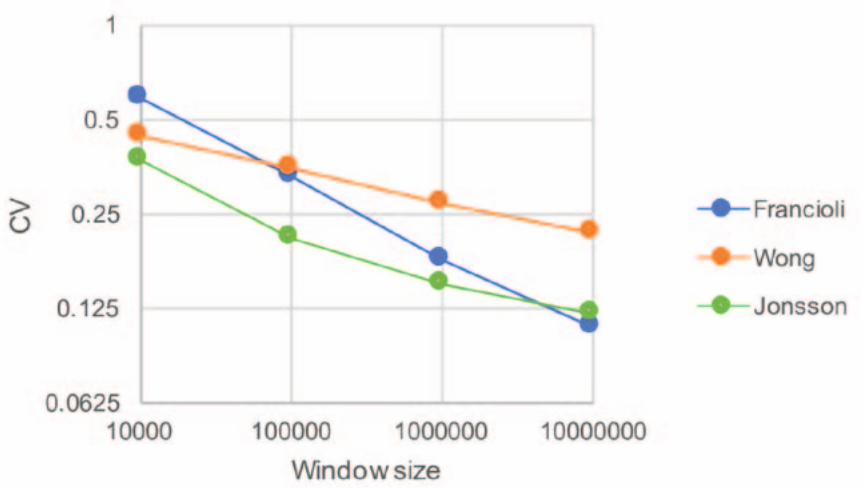


Figure 2.
The coefficient variation of mutation rates for different window sizes in 3 analysed data sets. All curves are shallower than -1/2. Therefore, variation at a larger scale cannot be explained by variation at a smaller scale.

Smith T C A, Arndt P F & Eyre-Walker A. Large scale variation in the rate of germ-line *de novo* mutation, base composition, divergence and diversity in humans. PLoS Genet 14(3):e1007254 (2018)

Ongoing and future projects (selected)

Quantitative modeling and functional analysis reveal continuous horizontal gene transfer across the bacterial domain

[Florian Massip, Peter Arndt, in collaboration with Misha Sheinman, Utrecht University]

Horizontal gene transfer (HGT) is an essential force that shapes microbial populations and drives their evolution. This process occurs via different mechanisms and has been studied in detail for a variety of biological systems. However, the coarse-grained global picture of HGT in the microbial world is still unclear. We are developing a method to detect recent HGT events, are modeling the evolution with HGT and are validating our model using empirical data. Our results are consistent with a model of continuous HGT on evolutionary time scales. Analyzing recently transferred genomic regions we assessed involvement of different taxa in HGT, transferred gene repertoire per taxa, function diversity and enrichment of certain functional categories. Our study thereby provides a first-time global overview over the evolution of microbial populations with HGT.

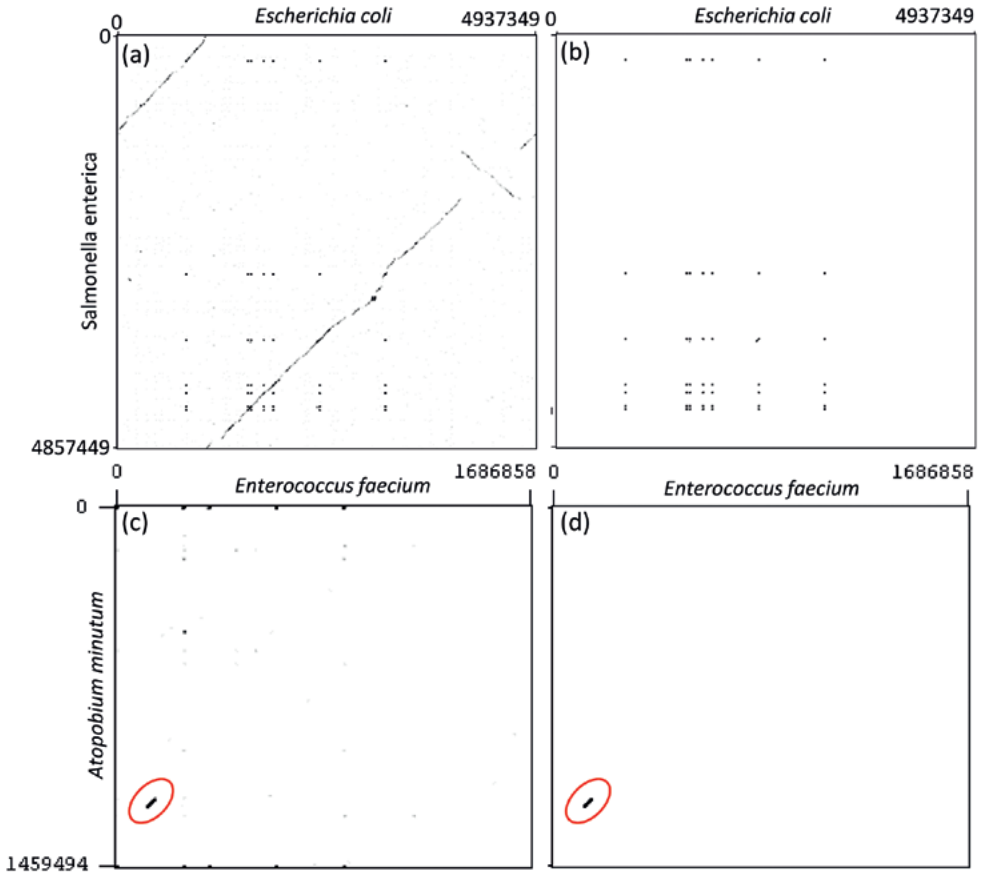


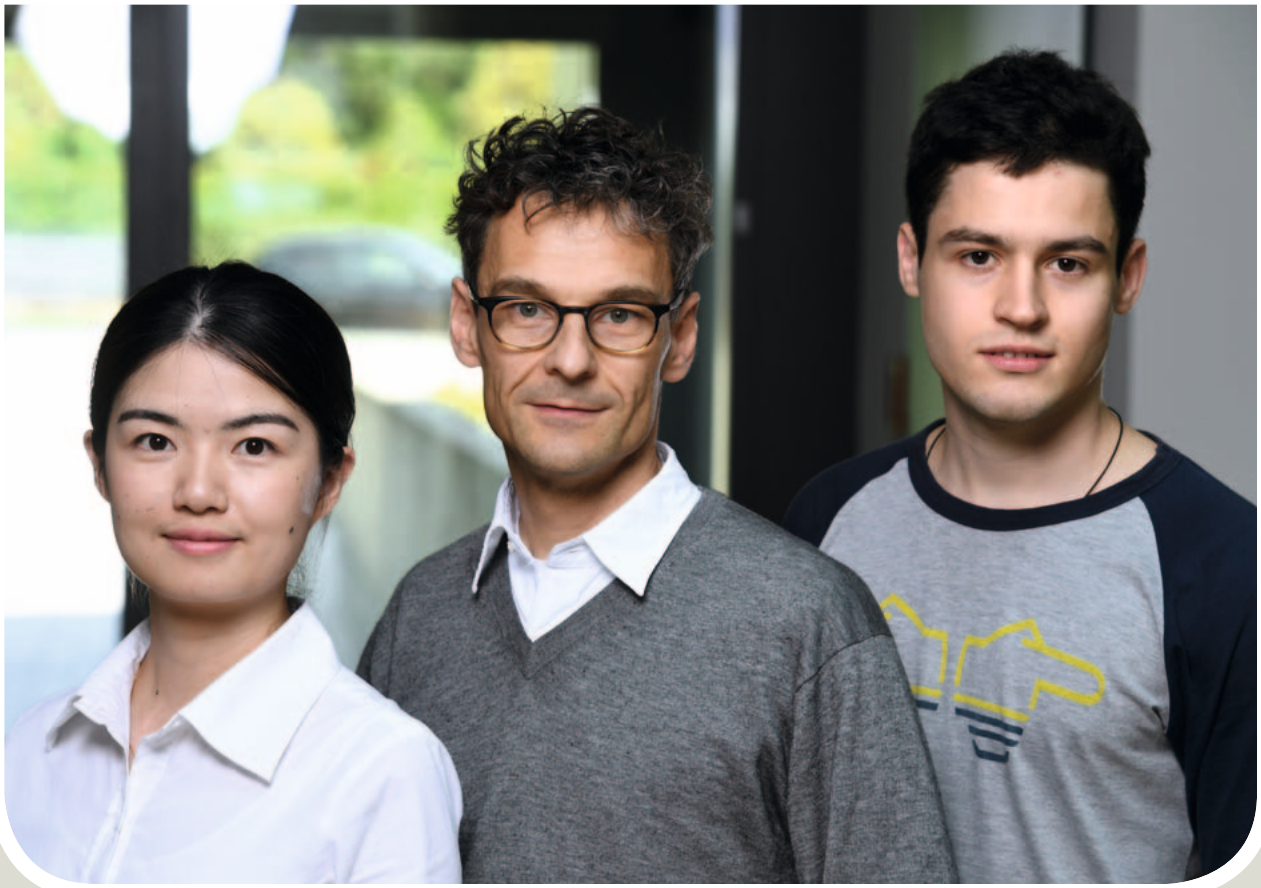
Figure 3.
Dot plot of two pairs of long contigs from genomes of different bacteria. (a-b) contigs from *E.coli* and *Salmonella enterica* (both belong to the Enterobacteriaceae family). Only matches longer than 20bp in (a) and 300bp in (b) are shown. (c-d) contigs from *Enterococcus faecium* (phylum Firmicutes) and *Atopobium minutum* (phylum Actinobacteria). Only matches longer than 20bp in (c) and 300bp in (d) are shown.

Modeling genome evolution

[Eldar Abdullaev, Ren Ren, Peter Arndt]

In one ongoing research project we study segmental duplications, i.e. long DNA segmental duplications (SDs) are long DNA sequences (typically > 1kb) with high sequence homology (> 90% identity) that exist in multiple copies in a genome. Several biological mechanisms are responsible for the emergence of SDs, including non-homologous recombination, non-homologous end joining and DNA polymerase slippage during replication. Segmental duplications are involved in multiple biological processes. For example, the duplication of genes is an important mechanism that reduces evolutionary constraints and leads to the creation of new genes. In addition, changes in the copy number of different genomic regions (CNVs) are associated with cancer. Even though we know some biological mechanisms that lead to segmental duplications it is currently not clear which forces shape SDs distribution in the genome. Therefore, we here evaluate for the first time a model that will explain the pattern of SDs distribution. This project will give us an insight on fundamental mechanisms of SD propagation in the genome.

In another project we model the evolution of genomic CpG islands using a dynamical model that allows for the protection of regions with high G+C and CpG content, i.e. CpG islands. Such a model can elegantly describe the dichotomy of the sequence composition along eukaryotic genomes.

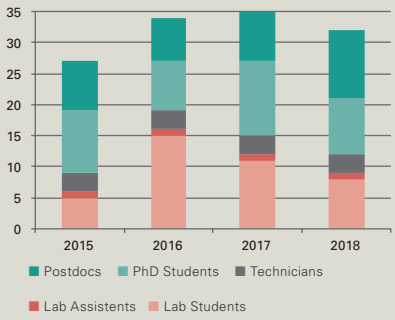


Mundlos Lab – Research Group Development & Disease

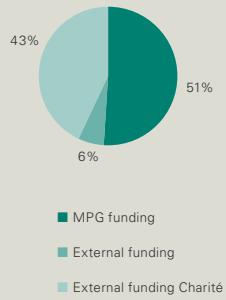
The group has been established in May 2000. In January 2018, Stefan Mundlos has been appointed as External Scientific Member of the Max Planck Society.

Prof. Dr. Stefan Mundlos
mundlos@molgen.mpg.de
++49 30 8413-1449

Team



Funding



Focus areas

- *Human Genetics* – molecular pathogenesis of congenital malformations
- *Human Genetics* – structural variations, detection and functional consequences
- *Developmental Biology* – morphogenesis of the skeleton and limb development
- *3D Genomics* – chromatin architecture in development and disease
- *Long range gene regulation* – enhancer function in a development and disease context

Structure and organization of the research group

The research group Development & Disease focuses on fundamental questions regarding normal and abnormal development. The research group is part of and works in close collaboration with the Institute for Medical and Human Genetics (IMHG) at the Charité-Universitätsmedizin Berlin. The IMHG provides clinical and diagnostic genetic service for the Charité and beyond. Medical doctors in training get scientific education at the Development & Disease group and scientists from the MPIMG have the opportunity to specialize in Medical Genetics at the IMHG. Thus, the IMHG covers the clinical aspects, genome analysis and diagnostics of human genetic disease, whereas the research group is more focused on functional analysis and basic biology of genomic alterations. Shared infrastructures, exchange of technical achievements and expertise, as well as common research goals ensure a successful interdisciplinary approach to study the mechanisms of genetic disease.

Research concept

By combining developmental biology, genetics, and clinical medicine we aim at generating in-depth knowledge of genetic disease and molecular disease pathology. Our phenotype-driven approach aims at dissecting molecular mechanisms of normal and abnormal development with a focus on defects of the skeleton/limb. Through clinical and diagnostic services as well as collaborations, patient cohorts are generated that are analyzed for genetic defects. A particular focus of our current work lies on variants in the non-coding genome and their role in gene regulation in the context of higher order chromatin structure. With the establishment of CRISPR/Cas9 methodology, we are now able to re-engineer basically any mutation in the genome and test their effect *in vivo* in mice. We use this technology to investigate the effects of variants on gene regulation, 3D chromatin folding and disease pathology. The effects of rearrangements on gene expression are not only studied during development and in a disease context, but also as a mechanism in evolution. It is our aim to translate the knowledge gained to develop novel test systems and better data analysis in order to improve diagnostics for genetic diseases.

Scientific highlights

- The hindlimb specific expression of transcription factor Pitx1 is determined by 3D chromatin architecture thereby conferring morphogenetic identity of forelimb vs. hindlimb [Kragesteen et al. Nature Genetics 2018]
- In collaboration with M. Nicodemi, Naples, Italy, we use capture HiC and a polymer physics based model to predict the effects of structural variants on chromatin architecture, enhancer promoter contacts, gene expression and disease [Bianco et al. Nature Genetics 2018]
- Extensive dissection of a regulatory region at the *Ihh* locus demonstrates gene regulation by a multipartite enhancer cluster with high redundancy. Doubling the enhancer quantity results in Ihh over- and misexpression showing dosage sensitivity of enhancers and thereby explaining malformation phenotypes [Will et al. Nature Genetics 2017]
- Large genomic duplications that contain TAD boundaries result in the formation of new chromatin domains with own regulatory potential that are insulated from the rest of the genome. This new pathomechanism is important for the interpretation of copy number variations, but also in an evolutionary setting where copy number differences are thought to play a crucial role in the evolution of genome complexity [Franke et al. Nature 2016]
- Genomic rearrangements such as deletions, duplications or inversions can result in re-wiring of enhancer promoter contacts, gene misexpression and disease through the disruption of 3D genome architecture [Lupianez et al. Cell 2015]

Scientific honours & selected invited talks

- Stefan Mundlos – Koselleck grant by the Deutsche Forschungsgemeinschaft, 2018
- Anja Will – Human Genetics Academic Award, 2018
- Stefan Mundlos – Appointment as External Scientific Member of the Max Planck Society, 2017
- Stefan Mundlos – Election as a member of EMBO, 2017
- Magdalena Socha – ESHG Young Investigator Award, 2017
- Anja Will – ESHG Young Investigator Award, Anja Will, 2017
- Stefan Mundlos – ESHG Science Award, 2016
- Stefan Mundlos – Greenberg Lecture, Baylor College, Houston, USA, 2016
- Martin Mensah – ESHG Young Investigator Award, 2016
- Peter N. Robinson – Eurordis Scientific Award, 2016

Results of completed projects during reporting period (selected)

Over the last years we have been successful in investigating the effects of large scale rearrangements, so called structural variants (SVs), on 3D genome folding, gene regulation and gene expression. Based on our studies, we developed a framework for the interpretation of SVs with regard to the pathogenic potential taken into account their effect on 3D genome architecture and gene regulation. Our studies are based on the seminal discovery of topologically associated domains (TADs) that were shown to be regions of high contact, separated by boundaries with low contact, thereby confining the region an enhancer can interact with. We interpret SVs in regard to their effect on TAD configuration. Deletions and inversions, for example, can lead to the fusion of TADs thereby connecting two regulatory regions that were previously separated. Large duplications that include boundary elements, in contrast, can result in the formation of new chromatin domains with own regulatory domains. Increasing the dosage of enhancers in TADs can lead to gene mis- and overexpression. With these studies we describe a new mutational pathomechanism that is highly relevant for the interpretation of SVs in congenital disease and cancer.

Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, ... [9 authors]...Borschiwer M, Haas SA, Osterwalder M, Franke M, Timmermann B, Hecht J, Spielmann M, Visel A & Mundlos S. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 161(5):1012-1025 (2015)

Spielmann M, Lupiáñez DG & Mundlos S. Structural variation in the 3D genome. Nat Rev Genet 19(7):453-467 (2018)

Dissecting the pathomechanisms of genomic duplications

[Martin Franke, Daniel Ibrahim, Guillaume Andrey, Katerina Kraft, Ivana Jerkovic]

In previous studies we and others described overlapping duplications at the *SOX9* locus that give rise to very different phenotypes: Cooks syndrome, a limb malformation condition with aplasia of nails, female to male sex reversal, or no abnormality. To understand how these rearrangements can result in such disparate outcomes, we studied the 3D genome architecture at the extended *Sox9* locus in mouse limb buds using chromosome conformation capture technologies (4C and capture HiC). These studies showed that the *Sox9* TAD is defined by strong boundaries, one of which is located with the large gene desert between *Sox9* and the neighboring *Kcnj1/16* genes. We re-engineered three different types of human duplications in mice and were able to show that those duplications, that span the TAD boundary, effectively create a new TAD (neo-TAD) that is insulated from the rest of the genome and retains its own regulatory potential. Duplicated genes caught in this neo-TAD (in this case *Kcnj2*) are now regulated by the duplicated region of the neighbor-TAD, resulting in gene misexpression and disease. This mechanism is important for the interpretations of SVs, has been shown to be relevant in cancer, and is likely to be a mechanism for the acquisition of evolutionary traits. Bioinformatic analyses were done in close collaboration with the Vingron Department.

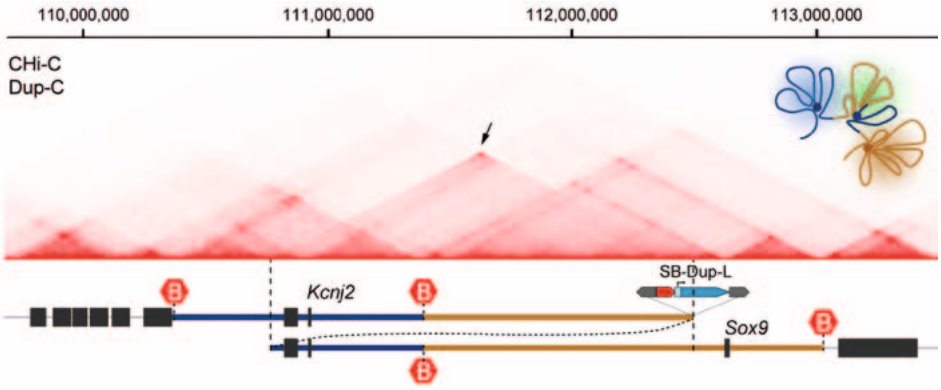


Figure 1.
Sox9/Kcnj2 locus with duplication creating neo-TAD (arrow).

Franke M*, Ibrahim DM*, Andrey G, Schwarzer W, Heinrich V, Schöpflin R, Kraft K, Kempfer R, Jerković I, Chan WL, Spielmann M, Timmermann B, Wittler L, .. [7 authors] ... Vingron M, Spitz F & Mundlos S. Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature 538(7624):265-269 (2016)
* shared first authorship

Predicting the effects of structural variants on genome architecture

[Darío Lupiáñez, Katerina Kraft, Robert Schöpflin, Guillaume Andrey]

SV can change chromatin organization and lead to enhancer-promoter rewiring, gene misexpression and disease. However, it is currently difficult to predict the effect of SVs on TAD organization and gene expression. Together with M. Nicodemi, University of Naples, Italy, we used HiC and polymer physics to predict the effect of SVs on higher order chromatin organization and gene expression at the *EPHA4* locus. The model disentangles the folding complexity of the locus and identifies SV-induced ectopic contacts and alterations of 3D genome organization in homozygous or heterozygous states. The *in silico* prediction provides a tool for analyzing the disease-causing potential of SVs.

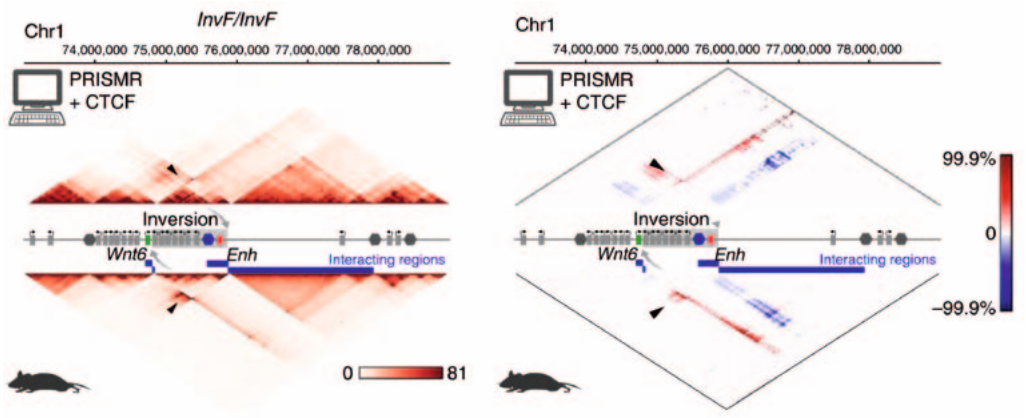


Figure 2.
Prediction and mouse HiC data showing aberrant contacts induced by inversion.

Bianco S*, Lupiáñez DG*, Chiariello AM, Annunziatella C, Kraft K, Schöpflin R, Wittler L, Andrey G, Vingron M, Pombo A, Mundlos S* & Nicodemi M*. Polymer physics predicts the effects of structural variants on chromatin architecture. Nat Genet 50(5):662-667 (2018)
* shared first and last authorship

Enhancer redundancy and dosage sensitivity

[Anja Will, Giulia Cova, Darío Lupiáñez]

The sum of individual enhancers is generally thought to constitute a gene's expression pattern. In this study we investigated the activity and specificity of an enhancer cluster that regulates the expression of Indian hedgehog (*Ihh*), a gene important for skeletal development. We first identified the regulatory domain of *Ihh*, deleted the domain, which resulted in an almost complete loss of *Ihh* expression. We identified a total of eight enhancers that all showed overlapping but individually different regulatory activity in four tissues (finger tips, condensations, growth plates, skull sutures) indicating a complex functional redundancy. We tested the effect of enhancer dosage by duplicating parts and the entire region. This results in Ihh over- and misexpression and in limb phenotypes (syndactyly, polydactyly), thereby recapitulating human malformations associated with similar duplications. This study illustrates that enhancers function with high redundancy and that they are sensitive not only to loss, but also to gain of copies. Developmental enhancer complexes are a fine-tuned regulatory system that works as a whole in a dosage-sensitive manner.

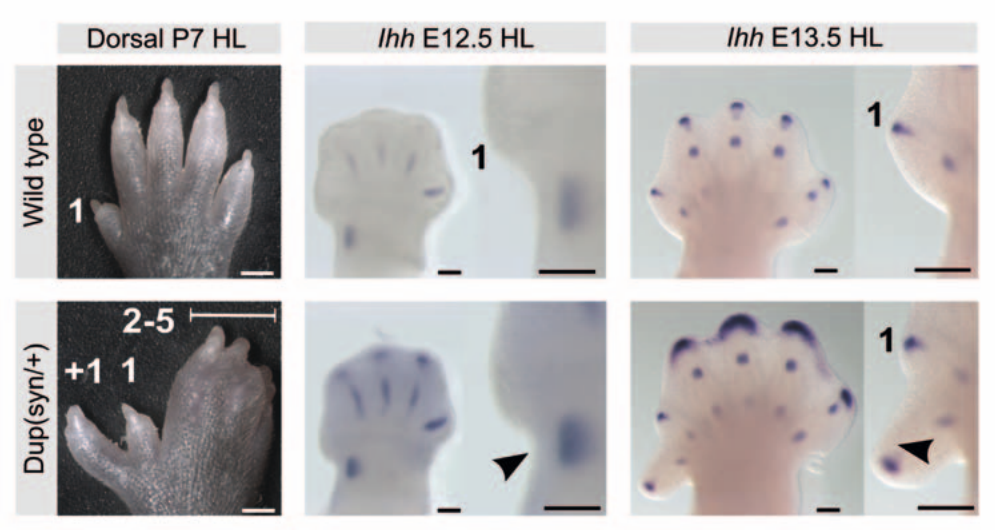


Figure 3.
Duplication of *Ihh* enhancer cluster induces *Ihh* misexpression and polydactyly.

Will AJ, Cova G, Osterwalder M, Chan WL, Wittler L, Brieske N, Heinrich V, de Villartay JP, Vingron M, Klopocki E, Visel A, Lupiáñez DG* & Mundlos S*. Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). Nat Genet 49(10):1539-1545 (2017)
* shared last authorship

Conferring enhancer specificity in 3D

[Bjørt Kragesteen, Malte Spielmann, Christina Paliou, Ivana Jerković, Martin Franke, Darío Lupiáñez, Katerina Kraft, Robert Schöpflin, Guillaume Andrey]

The development of hindlimbs (legs) vs. forelimbs (arms) is a paradigm of morphogenesis and gene regulation. Early during development, the fore- and hindlimb buds are indistinguishable, only later do they develop their characteristic morphology. We investigated this phenomenon by dissecting the regulation of *Pitx1*, the gene that determines hindlimb identity and that is expressed in hindlimbs only. We dissected the *Pitx1* regulatory landscape and find that *Pitx1* is regulated by an enhancer *PEN*, which is active in both, fore- and hindlimbs. cHiC and computational modeling of the locus show that *PEN* is tugged away in the forelimb behind the polycomb repressed neighbor of *Pitx1*, *Neurog1*, in the 3D space of folded chromatin. In contrast, in the active state, i.e. in the hindlimb, *PEN* is next to *Pitx1* and thus able to contact it. Rearrangements of the locus (inversion and deletion) disrupt the inactive forelimb state resulting in a 3D conformation that resembles the active hindlimb. Furthermore, we identified the *HoxC* cluster as upstream regulator of the active configuration. This mechanism explains the molecular pathology of Liebenberg syndrome, a condition with arm-to-leg transformation, that we call "regulatory endoactivation". Bioinformatic analyses for HiC were done in close collaboration with the Vingron Department.

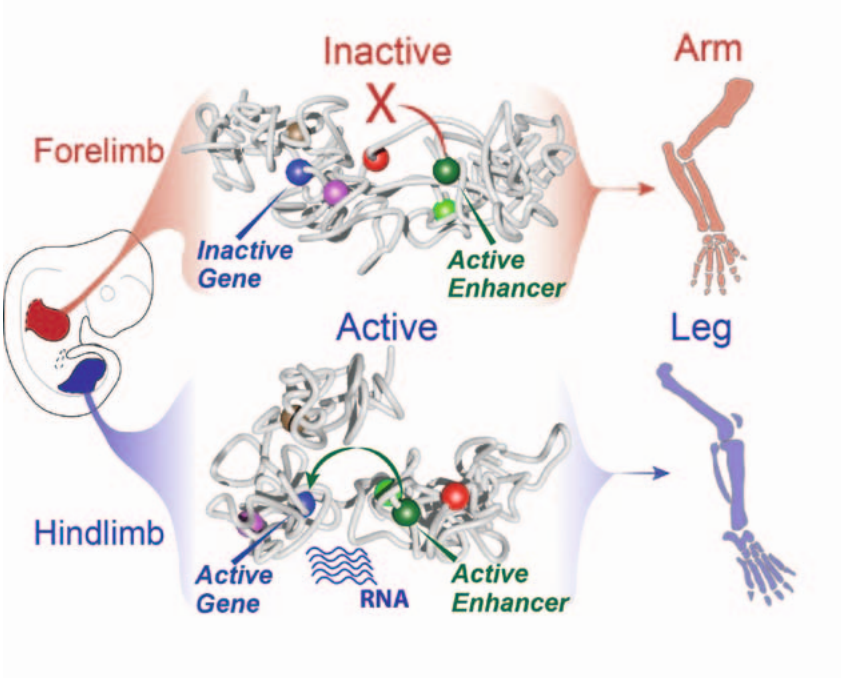


Figure 4.
Different 3D chromatin configurations determine arm vs. leg identity.

Kragesteen BK*, Spielmann M*, Paliou C, Heinrich V, Schoepflin R, Esposito A, Annunziatella C, Bianco S, Chiariello AM, Jerković I, Harabula I, Guckelberger P, Pechstein M, Wittler L, Chan WL, Franke M, Lupiáñez DJ, Kraft K, Timmermann B, Vingron M, Visel A, Nicodemi M, Mundlos S* & Andrey G*. Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. Nat Genet (2018), doi: 10.1038/s41588-018-0221-x
* shared first and last authorship

Genome analysis for disease variant identification and interpretation

[Rocio Acuna Hidalgo, Malte Spielmann, Robert Schöpflin]

With whole genome sequencing the identification of disease causing variants has entered a new stage. Nevertheless, the interpretation of the results remains challenging and only few disease-causing mutations have been identified so far, mainly because non-coding variants are difficult to interpret. In this project, we use whole genome sequencing to identify variants. This project is interdisciplinary and involves several departments at the Charité and elsewhere, clinicians for sampling, diagnosing, and phenotyping as well as bioinformaticians for the analysis of phenotypic and sequence data, and HiC. The continuous supply of patient material provides us with a constant flow of novel genes and mutations.

Flöttmann R, Kragesteen BK, Geuer S, Socha M, Allou L, ... [20 authors] Mundlos S* & Spielmann M.* Noncoding copy-number variations are associated with congenital limb malformation. Genet Med 20(6):599-607 (2018)
*shared last authorship

Spielmann M, Kakar N, Tayebi N, Leettola C, Nürnberg G, Sowada N, Lupiáñez DG, Harabula I, Flöttmann R, Horn D, Chan WL, ... [10 authors] Mundlos S* & Borck G*. Exome sequencing and CRISPR/Cas genome editing identify mutations of ZAK as a cause of limb defects in humans and mice. Genome Res 26(2):183-91 (2016)
*shared last authorship

Ongoing and future projects

Structural organization of TADs

[Daniel Ibrahim, Alexandra Despang, Robert Schöpflin]

TADs are organizational units of the genome that define regulatory domains. However, little is known what keeps a TAD together, what type of rearrangements lead to TAD disruption/fusion and when this results in gene misexpression. We investigated the basic organization of TADs and their effect on gene regulation at the *Sox9* locus, a region of the genome that is prone to rearrangements with often detrimental outcome. The deletion of the boundary and each CTCF site of two flanking TADs resulted in the fusion of the *Sox9* and *Kcnj2* TADs, but no gene misexpression. Only the inversion of the region with its CTCF sites resulted in activation of *Kcnj2* by *Sox9* enhancers. We show that TADs consists of multiple, partially redundant CTCF sites and that boundaries are part of this overall scheme. Enhancers are, against current concepts, not promiscuous but have high specificity towards their target gene. This can be overcome by forced contacts in newly formed domains.

Regulatory genomics of evolution

[Francisca Martinez Real, Dario Lupianez, Robert Schöpflin]

Evolutionary traits are thought to be acquired through multiple alterations in genomic information/sequence. However, few species-specific traits can be explained on a molecular level. In this project we investigated a particularly striking example of an evolutionary trait, true hermaphroditism (presence of ovaries and testis) in female moles. Together with the Vingron Department, we created a high-quality genome with dense information on regulatory activity during gonadal development, based on whole genome sequencing and HiC, as well as RNA-seq, epigenetic histone modification ChIP-seq from different stages of gonadal development in mole embryos. We identified three loci that contribute to the trait that all work through gene regulatory mechanisms. a) enhancer acquisition: gain of expression in a tissue where the gene is not expressed in other species via the acquisition of new active enhancer elements; b) TAD extension: break of synteny that moves a regulatory sequence in the original TAD thereby adding new enhancers to the TAD; c) enhancer duplication: duplication of an enhancer element resulting in gene upregulation. The here identified mechanisms/rearrangements follow the same principle as those described for human mutations.

The super resolution genome

[Rocio Acuna Hidalgo, Robert Schöpflin]

In spite of increasing efforts to diagnose patients with suspected genetic disease, a large percentage remains unknown in its origin. For example, in our cohort of individuals with limb malformations, >70% remain undiagnosed in spite of whole genome sequencing (WGS). We aim at tackling this problem by filling in the gaps of information that are left using current sequencing technology. In addition to the analysis of trio-WGS, we produce high resolution HiC from patient cells and compare these data with controls and the variants identified. In addition, we use long read sequencing technology (PacBio and Nanopore) to produce long reads for better SV detection.

Repetitive elements and their role in 3D genome organization

[Giulia Cova, Guillaume Andrey]

Our genome is full of repetitive elements, but in regard to their role in gene regulation and disease pathology, they are largely ignored. We study the role of repetitive elements in a model system, the Dactylaplasia mouse. This mouse mutant carries an insertion of a MusD element in the *Fgf8* locus, which is associated with a split-hand-foot limb malformation. We investigate the effect of this insertion on nearby gene expression and the molecular pathology of this condition. We will use this model to study the effect of repetitive elements genome-wide and will combine this knowledge with the study in human samples using long read technology. This sequencing technology is an essential part to locate repeats and to investigate their effect on genome integrity and regulation.

Cooperation within the Institute

We have intensive cooperation with the Vingron Department, mainly regarding the analysis of chromosome conformation capture data (4C, HiC), as well as on ChIP-seq and ATAC-seq analysis. This collaboration has been very fruitful and without it, these results could have not been achieved. Cooperation over the past years have been with the Herrman Dept. on mouse transgenic technology, CRISPR/Cas based genome editing in ES cells, and the analysis of mutant mice. The collaboration with the mouse facility is prerequisite for the success of our work and only the excellent expertise in transgenic technologies at the Institute makes these experiments possible. Intensive collaborations exist with the sequencing facility. Collaborations with the Meissner Dept. focus on methylome analysis in mouse mutants.

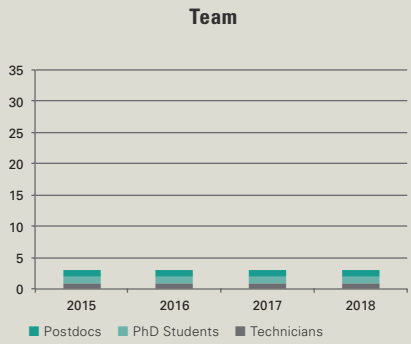
Special facilities / equipment

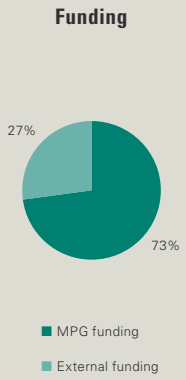
The research group as well as the IMG is equipped with the standard facilities for research into genetics, developmental biology, cell biology, and molecular biology.

Kalscheuer Lab – Chromosome Rearrangements and Disease Group

Dr. Vera Kalscheuer
kalscheu@molgen.mpg.de
++49 30 8413-1293

The group has been established in July 1995 at the former Dept. of Human Molecular Genetics and moved to the Research Group Development & Disease in November 2014.





Focus areas

- *Human genetics* – identification of novel genes for syndromic and non-syndromic intellectual disability, genotype-phenotype analysis
- *Neurobiology of disease* – understanding the pathophysiology of selected monogenic genetic disorders of the nervous system by dissecting the underlying molecular mechanisms

Research concept

Intellectual disability (ID) is characterized by significant limitations both in intellectual functioning and in adaptive behavior, which covers many everyday social and practical skills. It affects 1-2% of the general population and can be caused by a diverse spectrum of factors including gene mutations. During the past years we focused on the identification of novel ID genes and to better understand their functions and the underlying pathophysiologic mechanisms using *in vitro* and *in vivo* model systems. A clear knowledge of cellular processes and molecules that are involved in ID will hopefully pave the way towards the identification of novel targets that can be used for the development of drug therapies for genetically caused ID.

Scientific highlights

- Pathogenic variants in E3 ubiquitin ligase RLIM/RNF12 lead to a syndromic X-linked intellectual disability and behavior disorder [Frints et al. Mol Psychiatry 2018]
- Inherited and *de novo* pathogenic variants in the chloride/hydrogen ion exchanger CIC-4 cause X-linked intellectual disability in males and females [Palmer et al. Mol Psychiatry 2018]

- X-linked *GABRA3* variants are associated with epileptic seizures, encephalopathy and dysmorphic features [Niturad et al. Brain 2017]
- Pathogenic variants in the eukaryotic initiation factor *EIF2S3* cause a syndromic form of X-linked intellectual disability [Skopkova et al. Hum Mutat 2017]

Scientific honours & selected invited talks

- Invited speaker CDKL5 Forum, Boston, MA, USA, 2017
- Invited speaker 11th International Meeting on Genetics of Neurodevelopmental Disorders, Troina, Sicily, 2016
- Invited speaker Focus on CDKL5, Turin, Italy, 2016

Results of completed projects during reporting period

EIF2S3 mutations associated with severe X-linked intellectual disability syndrome MEHMO

[Friederike Hennig, Ute Fischer, in collaboration with Stefan Haas, MPIMG]

We reported that disruption of the X-linked *EIF2S3* gene, which codes for the gamma subunit of eukaryotic translation initiation factor 2 (eIF2), causes MEHMO syndrome. This rare disorder is characterized by profound intellectual disability, epilepsy, hypogonadism, microcephaly and obesity. eIF2 gamma is crucial for initiation of protein synthesis and regulation of the integrated stress response. Interestingly, studies in patient-derived fibroblasts suggested increased integrated stress response activation due to the frameshift mutation identified in the families. Functional assays performed in yeast demonstrated that this mutation impaired eIF2 gamma function. Our results added MEHMO syndrome to a group of disorders that are associated with eIF2, causing dysregulation of translation.

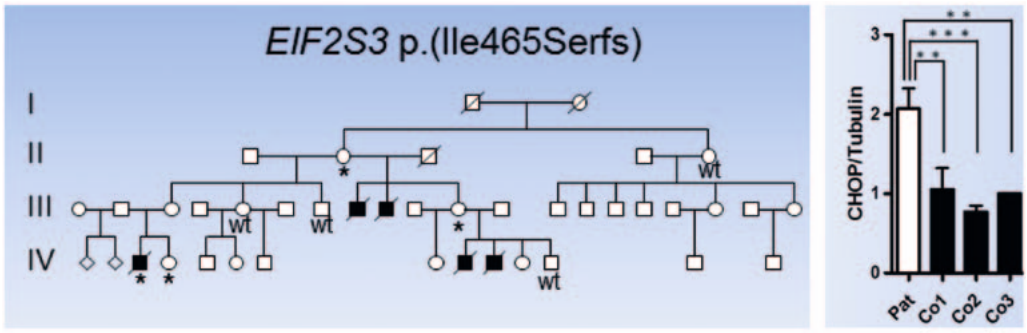


Figure 1.
Co-segregating truncating mutation in *EIF2S3* in a large family with MEHMO syndrome and associated elevated protein level of the stress response protein CHOP in patient fibroblasts compared to controls.

Skopkova M*, Hennig F*, Shin B S*, Turner C E, Stanikova D, Brennerova K, Stanik J, Fischer U, ... [16 authors] ... Dever T E & Kalscheuer V M. *EIF2S3* Mutations associated with severe X-linked Intellectual Disability Syndrome MEHMO. Hum Mutat 38: 409-425 (2017)
*equal contribution

Rare *GABRA3* variants are associated with epileptic seizures, encephalopathy and dysmorphic features

[Ute Fischer, Melanie Bienek, Corinna Jensen, Hao Hu, in collaboration with Stefan Haas, MPIMG]

In this study we for the first time linked the X-linked *GABRA3* gene, which encodes the alpha3-subunit of $GABA_A$ receptors, to a human phenotype. We identified pathogenic variants in families and sporadic cases with a range of epileptic seizure types, a varying degree of intellectual disability and developmental delay, sometimes with dysmorphic features. The variants co-segregated mostly but not completely with the disorder in the families, indicating in some

cases incomplete penetrance, involvement of other genes, or presence of phenocopies. Functional studies in Xenopus laevis oocytes revealed a variable but significant reduction of GABA-evoked anion currents for all mutants compared to wild-type receptors.

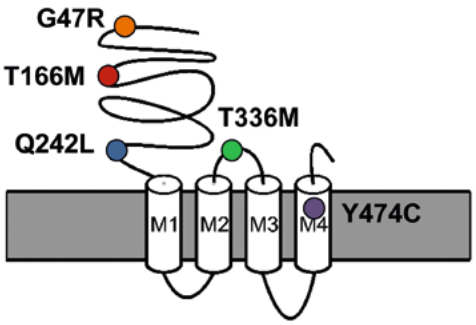


Figure 2.
Schematic representation of the $GABA_A$ receptor alpha3-subunit and the predicted positions of the mutated amino acids.

Niturad C E*, Lev D*, Kalscheuer V M*, ... [18 authors] Fischer U, Bienek M, Jensen C, Hu H et al. [15 authors]. Rare *GABRA3* variants are associated with epileptic seizures, encephalopathy and dysmorphic features. Brain 140: 2879-2894 (2017)
*equal contribution

De novo and inherited mutations in the X-linked gene *CLCN4* are associated with syndromic intellectual disability and behavior and seizure disorders in males and females

[Friederike Hennig, Ute Fischer, Astrid Grimme, Vanessa Suckow, Luciana Musante]

CLCN4 encodes the chloride/hydrogen ion exchanger CIC-4 prominently expressed in brain. We previously identified pathogenic *CLCN4* variants in a small number of families with X-linked intellectual disability and seizure disorder (Hu et al., Mol Psychiatry 2016). In this study we reported additional families with novel *de novo* or inherited *CLCN4* variants, extended the molecular spectrum and suggested phenotype-genotype correlation. Intellectual disability ranged from borderline to profound. Behavioral and psychiatric disorders were common in both child- and adulthood. Epilepsy was common, severity ranging from epileptic encephalopathy to well-controlled seizures. Of note, heterozygous females can be as severely affected as males. Pathogenicity of missense variants was further supported by electrophysiological studies in *Xenopus laevis* oocytes.

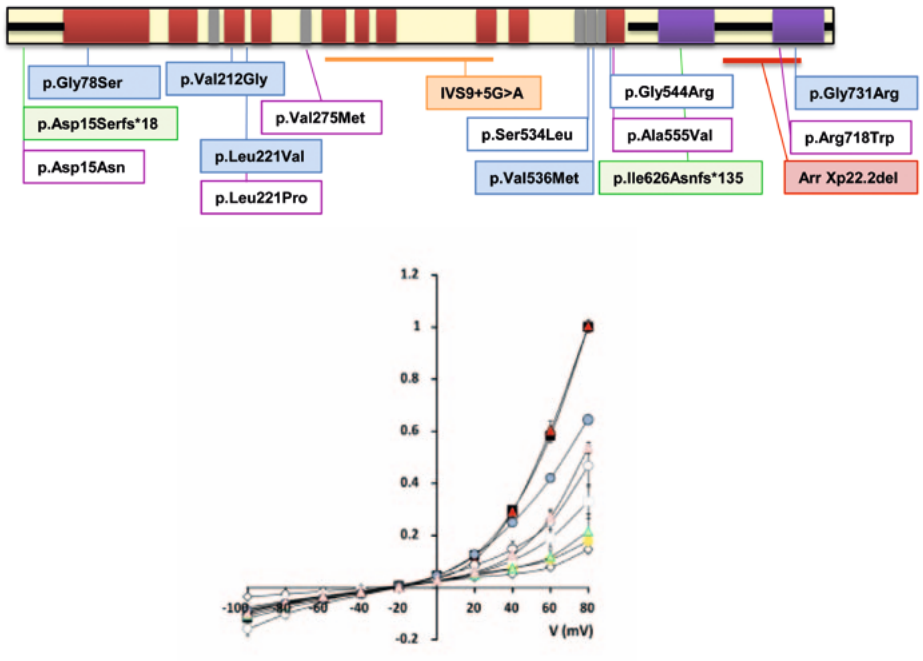


Figure 3.
Newly identified pathogenic variants in CIC-4 protein and functional results in *Xenopus* oocytes.

Palmer E E, ...[34 authors]...Hennig F, Fischer U, Grimme A, Suckow V, Musante L, ... [8 authors] ...Field M & Kalscheuer V M. *De novo* and inherited mutations in the X-linked gene *CLCN4* are associated with syndromic intellectual disability and behavior and seizure disorders in males and females. Mol Psychiatry 23: 222-230 (2018)

Ongoing and future projects

Functional significance of missense variants identified in affected probands and families

[Friederike Hennig, Vanessa Suckow, in collaboration with Lars Wittler, Judith Fiedler, Stefan Mundlos and Malte Spielmann, MPIMG]

Interpreting the pathogenicity of germline missense variants identified in affected probands and families can be challenging, especially when they are associated with disease variability. We use the CRISPR/Cas9 approach for introducing defined point mutations into mouse embryonic stem cells. The mutant cells are subsequently used for performing follow-up studies *in vitro* and *in vivo*. One of the variants we currently investigate in more detail is a single nucleotide exchange identified in *ROR2*. *ROR2* encodes the receptor tyrosine kinase like orphan receptor 2 protein, which belongs to a family of receptor tyrosine kinases with an essential role in skeletal development and early formation of chondrocytes.

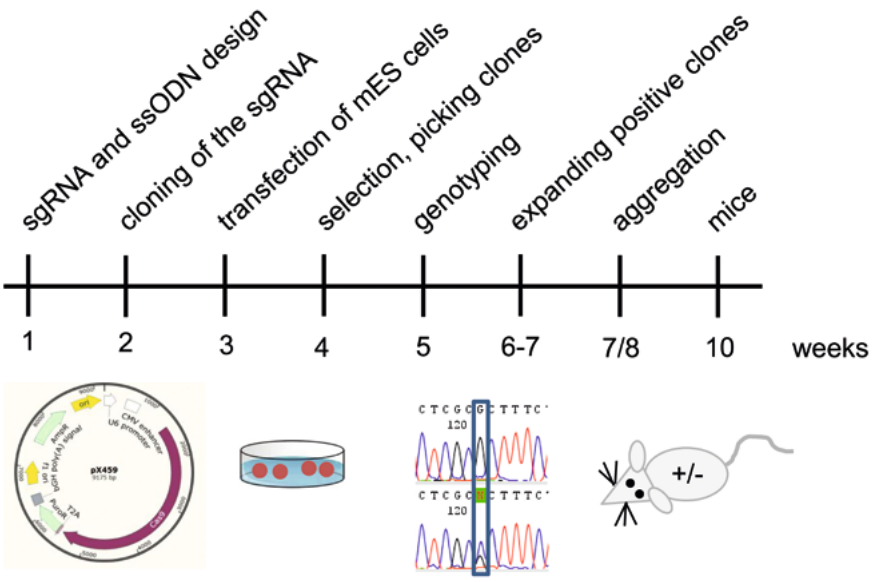


Figure 4.
Approach for generating mouse ES cells and transgenic mice with defined point mutations.

Pathomechanisms underlying the CDKL5 deficiency disorder

[Luciana Musante, Friederike Hennig, Melanie Hambrock, Vanessa Suckow, Aswin Pyakurel, in collaboration with David Meierhofer, MPIMG]

We previously linked *CDKL5* (cyclin-dependent kinase 5), which codes for a serine/threonine kinase, with a human phenotype (Kalscheuer et al., Am J Human Genet 2003, Tao et al., Am J Hum Genet 2004) and showed that CDKL5 ensures excitatory synapse stability by reinforcing NGL-1-PSD95 interaction in the postsynaptic compartment (Ricciardi et al., Nat Cell Biol 2012). Deleterious variants of *CDKL5* cause a severe neurodevelopmental disorder with early onset, mostly difficult to control, epileptic seizures, which usually occur within the first three to six months after birth. Other prominent clinical features include global developmental delay, cognitive dysfunction, general hypotonia, and severely impaired gross motor function, but there is clinical variability. The disorder primarily affects girls and most of the pathogenic variants are *de novo*. To gain more insight into the cellular functions of CDKL5 and the signaling pathways it is involved in, a prerequisite to

better understanding the pathophysiology of the CDKL5 deficiency disorder, we have established a Cdkl5 mouse model and investigate newly identified CDKL5 protein complex partners. One of them is implicated in mechanistic target of rapamycin (mTOR) signaling, suggesting that CDKL5 may play a role in regulating mTOR activity.

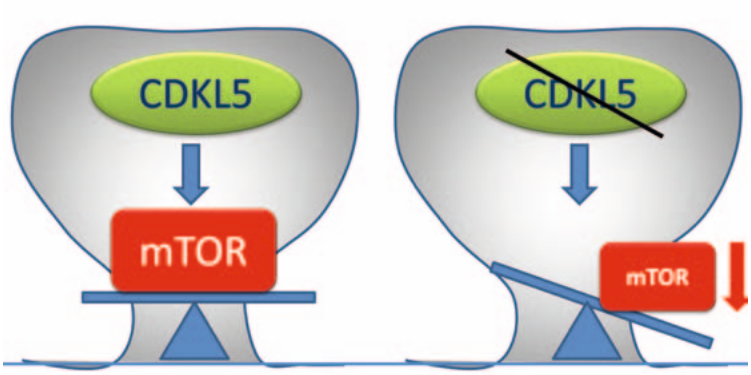


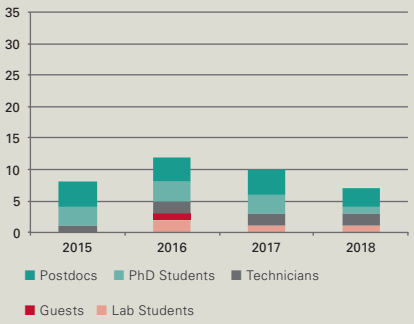
Figure 5.
Schematic showing CDKL5 and mTOR at the synapse under normal condition and in disease.

Chung Lab – Max Planck Research Group Epigenomics

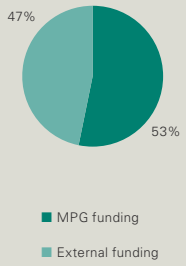
Dr. Ho-Ryun Chung
chung@molgen.mpg.de
++49 30 8413-1122

The group has been established in December 2011. Due to the appointment of Ho-Ryun Chung to a W3 professorship at the University of Marburg, the lab will be closed in November 2018.

Team



Funding



Focus areas

- Transcriptional regulation
- Chromatin and heterochromatin
- Epigenetics
- Nucleosome positioning
- Machine learning and data analysis

Research concept

The group focuses on chromatin-mediated transcriptional regulation. We study *cis*- and *trans* epigenetic mechanisms that lead to stable cell-type-specific gene expression. These mechanisms are pivotal for normal embryonic development and ensure normal cellular function during the life span of an organism.

In particular, we are interested in the mechanisms of heterochromatin nucleation, spreading, and maintenance in *cis*, and gene regulatory networks in *trans*. We develop and use both experimental and computational approaches that allow for generating and analyzing established and novel data types to answer eminent questions in the field. Our research should yield a better understanding of epigenetic mechanisms acting during development but also during the etiology of human diseases.

Scientific highlight

- reChIP-seq reveals widespread bivalency of H3K4me3 and H3K27me3 in CD4(+) memory T cells [Kinkley et al. Nat Commun 2016].

Scientific honours & selected invited talks

- Ho-Ryun Chung – W3 Professor for Medical Bioinformatics and Director of the Institute for Medical Bioinformatics at the University of Marburg, 2018
- F1000 Faculty Member for Bioinformatics, Biomedical Informatics & Computational Biology
- Invited speaker at seminar series of TRR81 Chromatin Changes in Differentiation and Malignancies, University of Marburg, 2017
- Invited speaker at seminar series Progress in Musculoskeletal Medicine, University of Münster, 2017

Results of completed projects during reporting period

reChIP-seq reveals widespread bivalency of H3K4me3 and H3K27me3 in CD4(+) memory T cells

[Sarah Kinkley, Johannes Helmuth, Ilona Dunkel]

The combinatorial action of co-localizing chromatin modifications and regulators determines chromatin structure and function. However, identifying co-localizing chromatin features in a high-throughput manner remains a technical challenge. Here we describe a novel reChIP-seq approach and tailored bioinformatic analysis tool, normR that allows for the sequential enrichment and detection of co-localizing DNA-associated proteins in an unbiased and genome-wide manner. We illustrate the utility of the reChIP-seq method and normR by identifying H3K4me3 or H3K27me3 bivalently modified nucleosomes in primary human CD4(+) memory T cells. We unravel widespread bivalency at hypomethylated CpG-islands coinciding with inactive promoters of developmental regulators. reChIP-seq additionally uncovered heterogeneous bivalency in the population, which was undetectable by intersecting H3K4me3 and H3K27me3 ChIP-seq tracks. Finally, we provide evidence that bivalency is established and stabilized by interplay between the genome and epigenome. Our reChIP-seq approach augments conventional ChIP-seq and is broadly applicable to unravel combinatorial modes of action.

Kinkley S, Helmuth J, Polansky J K, Dunkel I, Gasparoni G, Fröhler S, Chen W, Walter J, Hamann A & Chung H-R. reChIP-seq reveals widespread bivalency of H3K4me3 and H3K27me3 in CD4(+) memory T cells. Nat Commun 7: 12514 (2016)

PHF13 is a molecular reader and transcriptional co-regulator of H3K4me2/3

[Alisa Fuchs, Sarah Kinkley, Ilona Dunkel]

PHF13 is a chromatin affiliated protein with a functional role in differentiation, cell division, DNA damage response and higher chromatin order. To gain insight into PHF13's ability to modulate these processes, we elucidate the mechanisms targeting PHF13 to chromatin, its genome wide localization and its molecular chromatin context. Size exclusion chromatography, mass spectrometry, X-ray crystallography and ChIP sequencing demonstrate that PHF13 binds chromatin in a multivalent fashion via direct interactions with H3K4me2/3 and DNA, and indirectly via interactions with PRC2 and RNA Pol II. Furthermore, PHF13 depletion disrupted the interactions between PRC2, RNA Pol II S5P, H3K4me3 and H3K27me3 and resulted in the up and down regulation of genes functionally enriched in transcriptional regulation, DNA binding, cell cycle, differentiation and chromatin organization. Together our findings argue that PHF13 is an H3K4me2/3 molecular reader and transcriptional co-regulator, affording it the ability to impact different chromatin processes.

Chung H-R, Xu C, Fuchs A, Mund A, Lange M, Staege H, Schubert T, Bian C, Dunkel I, Eberharter A, Regnard C, Klinker H, Meierhofer D, Cozzuto L, Winterpacht A, Di Croce L, Min J, Will H & Kinkley S. PHF13 is a molecular reader and transcriptional co-regulator of H3K4me2/3. Elife 5. pii: e10607 (2016)

The Non-Specific Lethal complex creates a transcription-competent nucleosome landscape at promoters

[Ho-Ryun Chung]

Nucleosomal organization at gene promoters is critical for transcription, with a nucleosome-depleted region (NDR) at transcription start sites (TSSs) being required for transcription initiation. Here, we report that the *Drosophila* Non-Specific Lethal (NSL) complex is necessary to maintain this stereotypical nucleosomal organization at promoters. Upon NSL1 depletion, nucleosomes invade the NDRs at TSSs of NSL-bound genes. NSL complex member NSL3 binds to TATA-less promoters in a sequence-dependent manner. The NSL complex biochemically and genetically interacts with the NURF chromatin remodeling complex and is necessary and sufficient to recruit NURF to target promoters. The NSL complex is not only essential for transcription but is required for accurate TSS selection for genes with multiple TSSs. Further, loss of NSL complex leads to an increase in transcriptional noise. Thus, the NSL complex establishes a canonical nucleosomal organization that enables transcription and determines TSS fidelity.

Ongoing and future projects

My tenure as Max Planck Research Group leader ends in November 2018. As of the 1st of July 2018, I will start my new position as W3 Professor for Medical Bioinformatics and Biostatics and Director of the Institute for Medical Bioinformatics and Biostatistics at the Philipps-University Marburg. In the remaining time at the Institute, we hope to finish two projects:

STARR-FAIRE

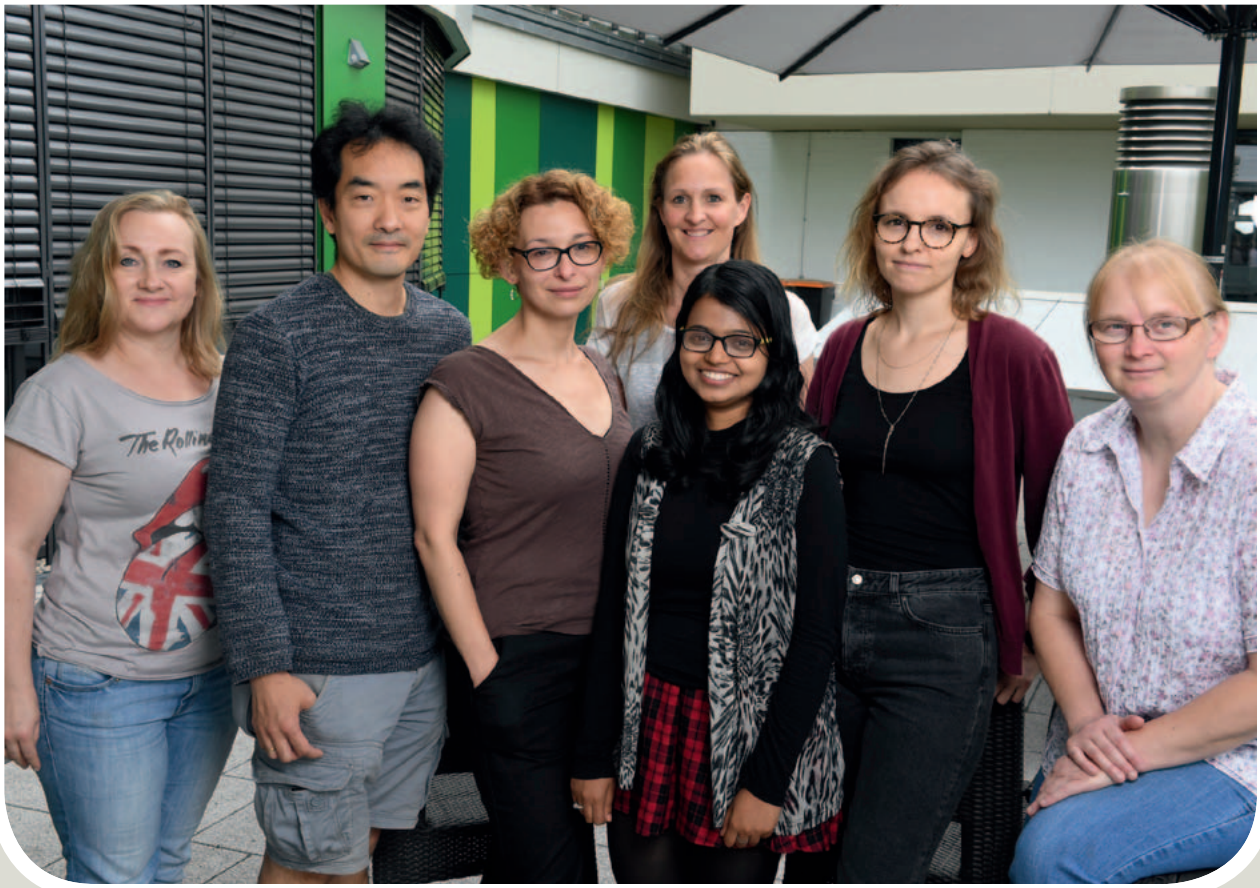
[Alisa Fuchs]

We successfully developed a method that uses the formaldehyde-assisted isolation of regulatory elements (FAIRE) assay to enrich for open chromatin fragments that serve as input for the self-transcribing active regulatory regions sequencing (STARR-seq) method. In addition we also introduced unique molecular identifiers (UMIs) in the first strand synthesis during library preparation for Illumina sequencing. With the help of the FAIRE assay we can measure enhancer activity for accessible chromatin regions in species with large genomes. Moreover, by the introduction of UMIs, we can measure the enhancer activity of individual DNA fragments.

DNase I HS HiC, referred to as Enhancer-Promoter-Interaction sequencing (EPI-seq)

[Alisa Fuchs, Astrid Grimme]

We successfully developed a method that uses a partial DNase I digest in the nucleus to cut chromatin at accessible chromatin regions. After digestions, the DNA ends are repaired using biotinylated nucleotides and ligated. DNase I hypersensitive regions are enriched by selecting DNA fragments between 500 and 2,000 base pairs. Illumina sequencing libraries are prepared using the Nextera system on streptavidin beads. Preliminary analysis reveals that this approach allows for high spatial resolution (500 – 1,000 base pairs) at moderate sequencing costs (~ 2 x 180,000,000 paired reads) uncovering regulatory-regions interactions.

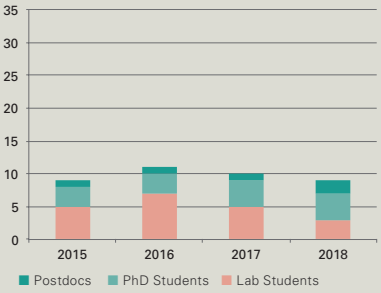


Marsico Lab –Research Group RNA Bioinformatics

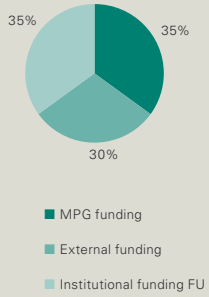
The group has been established in June 2014.

Prof. Dr. Annalisa Marsico
marsico@molgen.mpg.de
++49 30 8413-1843

Team



Funding



Focus areas

- *Applied Machine Learning* – deep learning models for RNA biology and system medicine
- *RNA bioinformatics* – tools for CLIP-seq data analysis and RBP motif finding
- *Long non-coding RNAs* – functional mechanisms and *in silico* classification
- *microRNA regulation* – transcriptional regulation of microRNAs and functions in lung infections

Research concept

Gene regulation occurs at several levels and is a highly controlled process, whose alterations contribute to many diseases. Besides messenger RNAs (mRNAs), most of the detected transcripts in complex organisms are non-coding RNAs (ncRNAs). The functional consequences of some ncRNAs are still poorly understood. Many newly discovered ncRNAs (e.g. microRNAs, long non-coding RNAs) as well as RNA Binding Proteins (RBPs) have been suggested to constitute an additional layer of gene regulation. Our group focuses on the development of machine learning methods to integrate high-dimensional genomic data, such as CLIP-seq, RNA-seq data and RNA sequences into models to elucidate ncRNA and RBP functions. In close collaboration with several experimental groups, we aim at characterizing protein-RNA interactions, their functional consequences, as well as bridge the gap between thousands of annotated lncRNAs and their (lack of) functional classification, mainly in disease settings.

Scientific highlights

- A python package for DNA/RNA sequence classification using Convolutional Neural Networks (CNNs) [Budach & Marsico Bioinformatics 2018]
- Identification of a group of nuclear lncRNAs which function as transcriptional regulators in a *quasi-cis* mechanism [Ntini et al. Nat Commun 2018]
- A *de-novo* motif finder to learn sequence/structure binding motifs of RNA Binding Proteins (RBPs) [Heller et al. Nucleic Acids Res 2017]
- A method to accurately identify RBP-RNA interaction sites from single nucleotide resolution CLIP-Seq data [Krakau et al. Genome Biol 2017]
- A first model to understand the contribution of genetic variation to microRNA regulation [Budach et al. Genetics 2016]

Scientific honours & selected invited talks

- Invited speaker at the London RNA club, Francis Crick Institute, London UK (2018)
- Invited speaker at the Symposium "Computational Challenges in RNA-based Gene Regulation: Protein-RNA Recognition, Regulation and Prediction", Dagstuhl Castle, Germany (2017)
- Successfully completed middle-evaluation Junior Professorship at Freie Universität Berlin, equivalent to German Habilitation towards a Full Professorship (2017)
- Invited speaker at the IRI Symposium "From RNA Pools to single RNA Molecules" Berlin, Germany (2015)

Results of completed projects during reporting period

PureCLIP: capturing target-specific protein–RNA interaction footprints from single-nucleotide CLIP-seq data

[Sabrina Krakau in collaboration with Hugues Richard, UPMC Paris, France]

The iCLIP and eCLIP techniques facilitate the detection of protein–RNA interaction sites at high resolution, based on diagnostic events at crosslink sites. We have developed *PureCLIP* (https://github.com/skrakau/PureCLIP), a non-homogeneous hidden Markov model-based approach, which simultaneously performs peak-calling and individual crosslink site detection. Unlike previous methods, it explicitly incorporates non-specific background and transcript abundance, as well as cross-linking sequence biases into the model. On both simulated and real data, PureCLIP is more accurate than other state-of-the-art methods and has a higher agreement across replicates, making it a reliable tool for the identification of genuine RNA-RBP interaction sites.

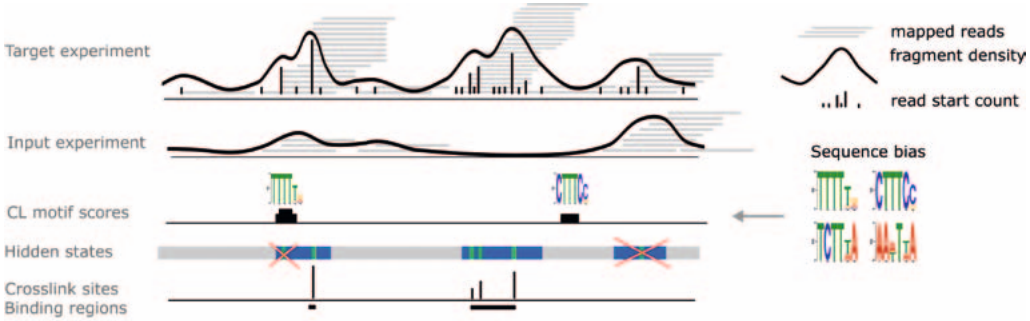


Figure 1. Overview of the PureCLIP approach. Based on read coverage and read start counts from iCLIP/eCLIP data it infers for each position the most likely hidden state. The goal is to identify sites corresponding to true RNA-RBP interactions. Information from input controls, as well as sequence bias, such as cross-linking (CL) motifs, is incorporated to reduce false calls.

Krakau S, Richards H & Marsico A. PureCLIP: Capturing target-specific protein-RNA interaction footprints from single-nucleotide CLIP-seq data. Genome Biol 18(1):240 (2017)

Learning sequence-structure motifs from high-throughput RNA-binding protein data

[David Heller, Stefan Budach in collaboration with Martin Vingron, MPIMG, and Ralf Krestel, Hasso Plattner Institute Potsdam]

RNA-binding proteins (RBPs) play an important role in RNA post-transcriptional regulation and recognize target RNAs via sequence-structure motifs. We have developed *ssHMM*, a *de novo* RNA motif finder based on a hidden Markov model (HMM) and Gibbs sampling to capture the relationship between RNA sequence and secondary structure preference of an RBP. ssHMM has been applied to find novel sequence-structure motifs of uncharacterized RBPs. It is freely available at https://github.molgen.mpg.de/heller/ssHMM and as a Docker image.

In a follow-up project, we have extended the framework of ssHMM to learn more complicated RBP-RNA interaction patterns from CLIP-seq data with Convolutional neural networks (CNNs), and designed *pysster*, a Python package for training CNNs on biological sequence data, mainly RNA sequences. *Pysster* can visualize learned motifs along with information about their positional and class enrichment, making it a powerful instrument in high-throughput genomics applications. The package is freely available at https://github.com/budach/pysster.

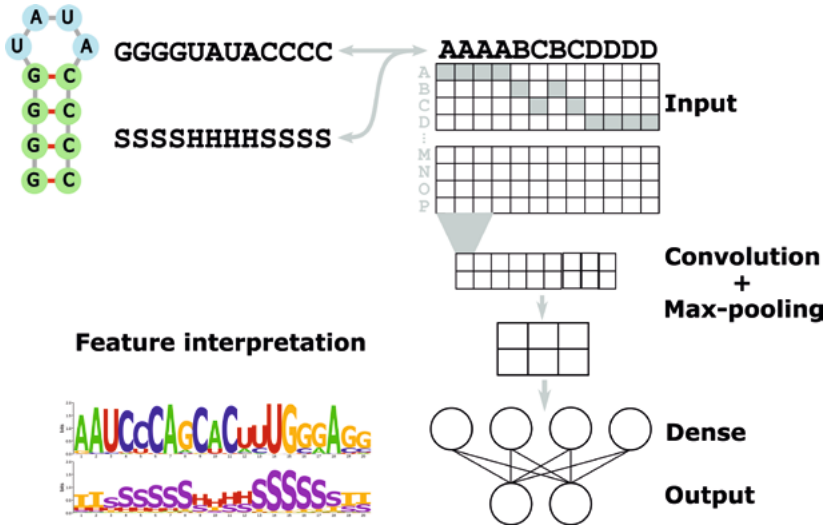


Figure 2.
Basic CNN architecture. RNA sequence and structure input strings are encoded via a combined alphabet, representing primary sequence and RNA secondary structure. Learned features can be interpreted as sequence and structure motifs.

Heller D, Krestel R, Ohler U, Vingron M & Marsico A. ssHMM: Extracting intuitive sequence-structure motifs from high-throughput RNA-binding protein data. Nucleic Acids Res 45(19):11004-11018 (2017)

Budach S & Marsico A. pysster: Learning sequence and structure motifs in DNA and RNA sequences using convolutional neural networks. Bioinformatics doi: 10.1093/bioinformatics/bty222 (2018) [Epub ahead of print]

Ongoing and future projects

Combinatorial prediction of *Xist*-induced gene silencing dynamics from epigenetic and genomic features

[Lisa Barros de Andrade e Sousa, in collaboration with Edda Schulz, MPIMG; Edith Heard, Institute Curie, Paris, France; and John Lis, Cornell University NYC, USA]

X-chromosome inactivation (XCI) in female mammals is initiated by the long non-coding RNA *Xist*, which mediates gene silencing of one X chromosome. The dynamics of gene silencing are highly variable across genes, with some genes even escaping XCI. The rules and molecular mechanisms underlying *Xist*-mediated gene silencing are still not fully understood. We have quantified chromosome-wide gene silencing dynamics upon *Xist* induction by measuring the nascent transcriptome in an allele-specific manner using Precision nuclear Run-On sequencing (PRO-Seq). We have then applied a random forest machine learning model to predict the measured silencing dynamics based on a large set of epigenetic and genomic features and tested few predictions experimentally. Overall, we have identified different feature sets that appear to influence whether a gene is subject to XCI at all and whether silencing occurs with slow or fast dynamics.

Experimental and *in silico* approaches for comprehensive functional classification of lncRNAs with focus on enhancer-like lncRNAs

[Evgenia Ntini, Stefan Budach, in collaboration with Jörg Winkler, Knut Reinert, Freie Universität Berlin, and Ulf Ørom, Aarhus University, Denmark]

Although few classifications of lncRNAs, based either on epigenetic features or kinetic measures of RNAs exist, there is the need of a functional classification of lncRNAs which elucidates their molecular functions. In collaboration with Prof. Knut Reinert (FU Berlin, joint DFG grant MA4454/3-1), we are investigating whether smaller local structure motifs mediate, in some cases, common lncRNA functions, e.g. RBP binding. We will use the tools developed in our lab, *PureCLIP* and *pysster*, to cluster lncRNAs based on RBP binding sites, local sequence-structure motifs, as well as genomic, epigenetic and metabolic data. By performing a transient transcriptome sequencing (TT-seq) experiment to sequence nascent RNAs at different time points, we will define, for the first time, a quantitative measure of „rate of release from chromatin", which we believe to be crucial for the function of a subgroup of enhancer lncRNAs, as hypothesized in our previous paper (Ntini et al. Nature Commun 2018). Ultimately, we hope to link this sub-class of lncRNAs to common local sequence/structure motifs which would help us to elucidate their molecular function.

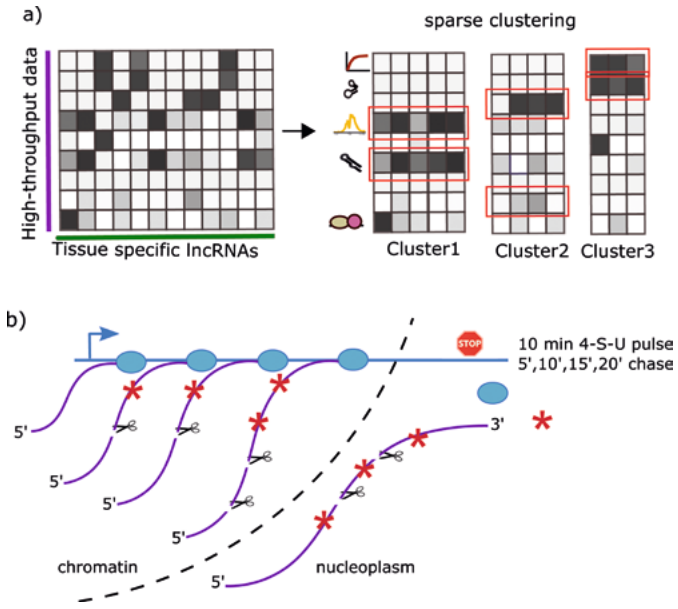


Figure 3.
a) Schematic view of lncRNA clustering based on genomic, epigenetic and metabolic measures. The most important features for the classification will be identified. b) Schematic overview of the planned TT-seq experiment.

Non-coding RNA-mediated regulatory networks in diseases

Dual RNA-seq host-pathogen in *Legionella pneumophila* lung infection

[Brian Caffrey, Roman Schulte-Sasse in collaboration with Martin Vingron, MPIMG; Wigo Bertrams, Christina Herkt, Bernd Schmeck, Institute for Lung Research, Marburg; and Jörg Vogel (Würzburg, Institute for Molecular Infection Biology, Germany)- SFB TR84 grant]

In this project we want to identify the direct molecular genome-wide effects on both the pathogen and the host cell transcriptome upon bacterial infection, in a dual RNA-seq experiment. It refers to the application of high-throughput RNA sequencing to a dual system composed of host cells (in this case human blood-derived macrophages) infected with a bacterium (in this case *Legionella pneumophila*), allowing both host and pathogen transcriptome to be analyzed in parallel at different time points during the course of infection (figure 4 left panel). We will test the hypothesis that an inter-species regulatory RNA network shapes the common battlefields of the pathogen and the host immune defense to the extent that it can be addressed therapeutically. We have identified, among others, an infection-specific long non-coding RNA, linc00158 which enhances expression of miR-155, and three differentially expressed miRNAs (miR-125b, miR-221, miR-579) which have a joint effect on intracellular replication of *L. pneumophila* during the infection via targeting of MX1 and LGALS8 genes. More in general, we are addressing the task of prioritizing coding and non-coding disease genes from multi-omics data using deep learning methods such as Graph Convolutional Network (GCNs).

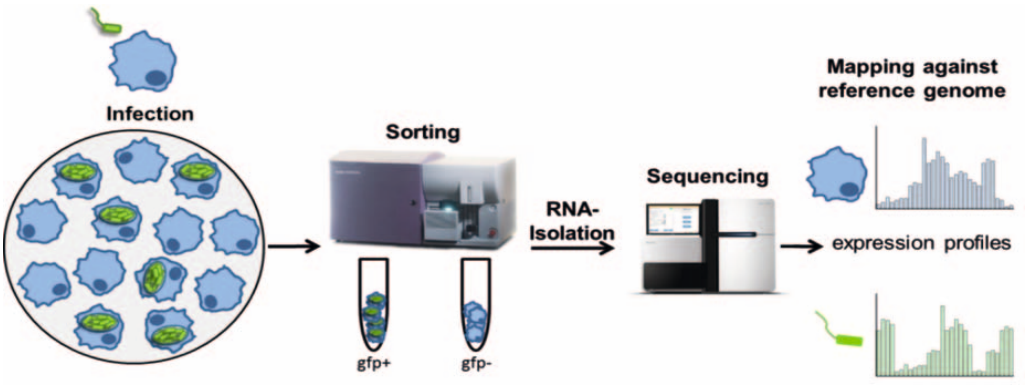
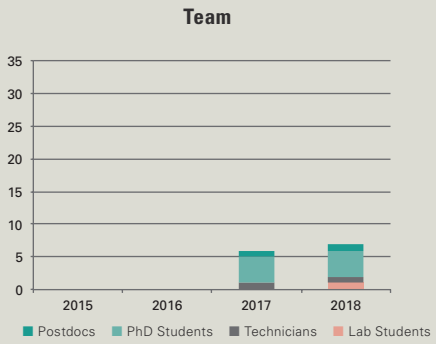


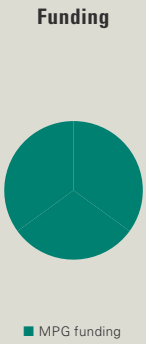
Figure 4.
Left: Schematic representation of the dual RNA-seq procedure; right: Simultaneous overexpression of miR-125b, miR-221 and miR-579 leads to an increased Legionella replication through a miRNA-mediated regulatory network.

Mayer Lab – Max Planck Research Group Nascent Transcription & Cell Differentiation

The group has been established in January 2017.

Dr. Andreas Mayer
mayer@molgen.mpg.de
++49 30 8413-1264





Focus areas

- *Genome transcription* – regulatory mechanisms that control RNA polymerase II transcription
- *Cell differentiation* – transcriptional dynamics and chromatin regulation during neurogenesis
- *Methods development* – high-resolution transcriptional profiling approaches

Research concept

The primary goal of our research is to reveal the key mechanisms that underlie the regulation of chromatin-mediated nascent RNA polymerase II transcription, including non-coding transcription, in differentiated mammalian cells and during cell differentiation. Specifically, we aim to understand how post-initiation regulatory events are established in a dynamic chromatin environment *in vivo* and how these regulatory mechanisms control and coordinate cellular differentiation. In order to achieve these goals, we are also developing new high-resolution genome-wide methods to study the full spectrum of genome transcription with nucleotide precision and at high kinetic resolution.

Scientific highlight

- Development of a new RNA-seq approach to analyze transcript abundance in subcellular compartments of human cells [Mayer et al. Current Protoc Mol Biol 2017]

Scientific honours & selected invited talks

- Invited speaker at the Institute of Genetics and Molecular and Cellular Biology (IGBMC), Illkirch, France, 2018
- Invited speaker at the Max F. Perutz Laboratories (MFPL) and IMBA, Vienna, Austria, 2018
- FEBS long-term postdoctoral fellowship for Olga Jasnovidova (postdoctoral fellow), 2018
- Invited speaker at EMBO Conference: From Single – to Multiomics, EMBL, Heidelberg, Germany, 2017
- Invited speaker at the Center for Regenerative Therapies, Technical University of Dresden, Germany, 2017

Results of completed projects during reporting period

subRNA-seq: a new genomic method for studying transcript abundance in subcellular compartments of mammalian cells

[Andreas Mayer & L. Stirling Churchman]

In eukaryotic cells, RNAs at various stages of maturation and processing are distributed across different cellular compartments. The standard approach to globally measure transcript abundance *in vivo* is RNA sequencing (RNA-seq). RNA-seq usually relies on RNA isolation from whole-cell lysates and thus mainly captures fully processed cytoplasmic RNAs. To overcome this major limitation, we here developed a protocol for subcellular RNA-seq (subRNA-seq). subRNA-seq allows the quantitative measurement of RNA polymerase II–generated RNAs from the chromatin, nucleoplasm, and cytoplasm of mammalian cells. This approach relies on an optimized cell fractionation prior to RNA isolation and sequencing library preparation (figure 1). High-throughput sequencing of the RNAs from the different cellular compartments reveals the identity, abundance, and subcellular distribution of transcripts, thus providing insights into RNA processing and maturation. Deep sequencing of the chromatin-associated RNAs further offers the opportunity to study nascent RNAs.

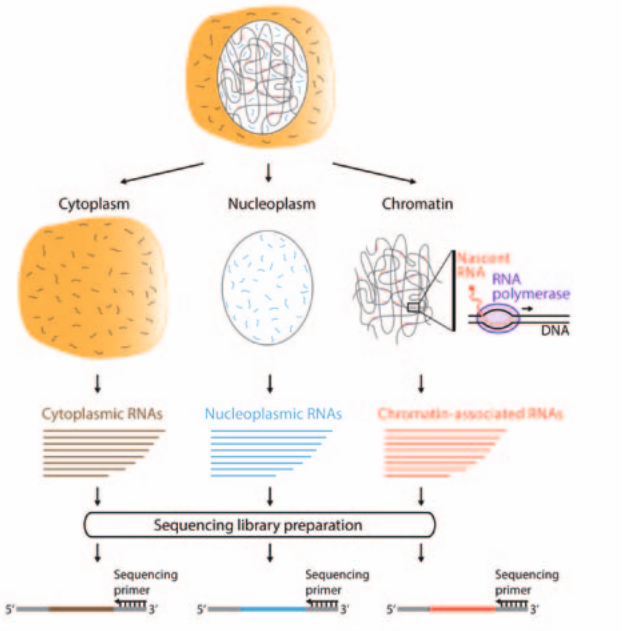


Figure 1.
Schematic overview of the subRNA-seq approach (adapted from Mayer and Churchman Curr Protoc Mol Biol 2017).

*Mayer A & *Churchman LS. A detailed protocol for subcellular RNA sequencing (subRNA-seq). Curr Protoc Mol Biol 120:4.29.1-4.29.18 (2017)
*co-corresponding authorship

Ongoing and future projects

Regulation of RNA polymerase II transcription by BET proteins

[Mirjam Arnold, Annkatrin Bressin]

Recently, evidence has accumulated that BET bromodomain proteins BRD4, BRD3 and BRD2 are involved in RNA polymerase II (Pol II) transcription. However, their functions in genome transcription remain unclear. In collaboration with the James Bradner laboratory (Dana-Farber Cancer Institute, Boston, USA), we have previously applied induced BET protein degradation to specifically and rapidly deplete BET proteins from human cells (figure 2A). BET degradation resulted in a global collapse of productive transcription elongation (figure 2B) suggesting a regulatory role in Pol II transcription elongation (Winter, Mayer et. al. Mol Cell 2017). Since induced BET degradation results in the simultaneous depletion of all BET proteins, BET protein-specific functions remain unknown. In the present study, we have generated a BRD4 degron-tagged human cell line to investigate BRD4-specific roles in Pol II genome transcription.

Since BRD4 predominantly localizes to promoter and enhancer regions, future studies will focus on elucidating the role of BRD4 in enhancer transcription and enhancer-promoter communication. In future, we will also investigate other BET-specific functions in the regulation of Pol II transcription.

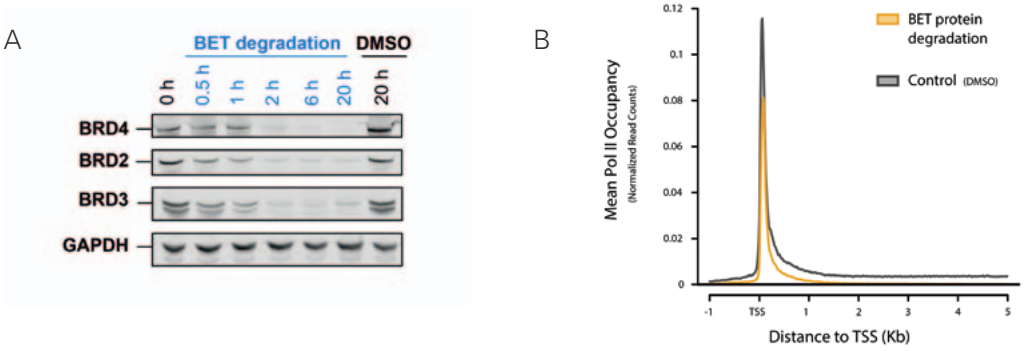


Figure 2.
Induced BET degradation (A) disrupts productive transcription elongation genome-wide (B)

Role and regulation of RNA polymerase II pausing

[Yelizaveta Mochalova, Susanne Freier]

We and others have recently developed a native elongating transcript sequencing (NET-seq) approach for human cells which provides a quantitative measure of RNA polymerase II (Pol II) density with single-nucleotide resolution across the genome (Mayer & Churchman Nat Protoc 2016). Application of the human NET-seq approach revealed pervasive Pol II pausing at the majority of active genes. Surprisingly, pausing of Pol II is not restricted to promoter-proximal regions but also occurs throughout gene-bodies of expressed genes (figure 3) (Mayer et al. Cell 2015). This observation emphasizes the existence of major rate limiting steps and thus regulatory potential downstream of transcription initiation (Mayer et al. Curr Opin Cell Biol 2017). In the present study, we aim to reveal the determinants of transcriptional pausing as well as its role for controlling the transcriptional output of cells.

In future, we will investigate the role and dynamics of Pol II pausing during cell differentiation.

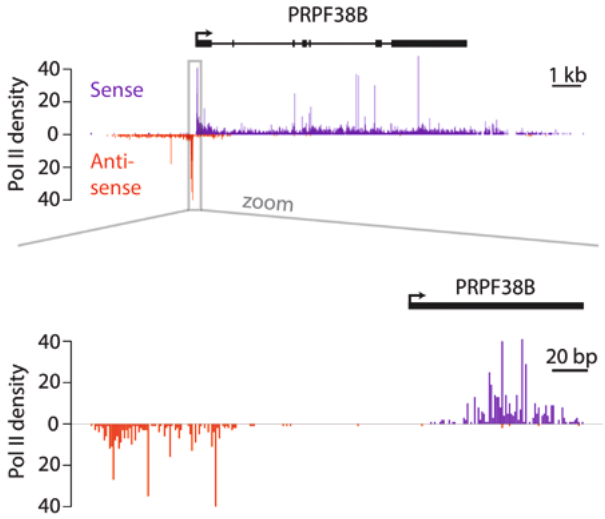


Figure 3.
Widespread Pol II pausing along a representative human gene (adapted from Mayer et al. Cell 2015).

Regulatory interplay between chromatin organization and transcription during cell differentiation

[Olga Jasnovidova, Martyna Gajos]

The regulatory links between the chromatin organization and RNA polymerase II (Pol II) transcription that underlie cell fate determination and cell identity are poorly defined. In order to identify the regulatory principles that govern cell fate decisions and that control cell identity, we employ an *in vitro* neuronal cell differentiation system in collaboration with Volker Busskamp's group at the TU Dresden (figure 4). Using this neurogenesis model, induced human pluripotent stem cells differentiate into bipolar retinal neurons with high purity and yield recapitulating neuronal lineage determination *in vitro*. We determine transcriptional switches and chromatin-state-changes throughout all stages of the neuronal differentiation process. To expose causal connections between chromatin organization and Pol II transcription, we disrupt specific Pol II transcription factors and chromatin regulators at distinct developmental stages.

In future, we will extend our analysis to other differentiation systems such as to a human retinal disease model.

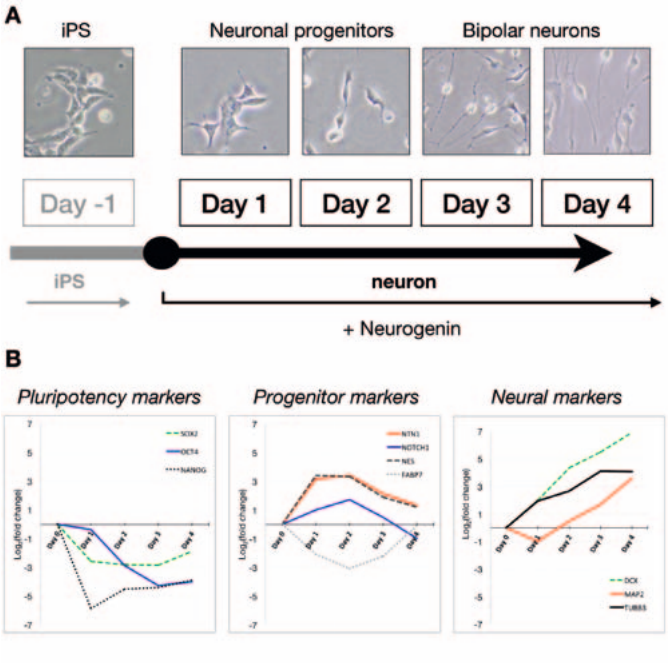
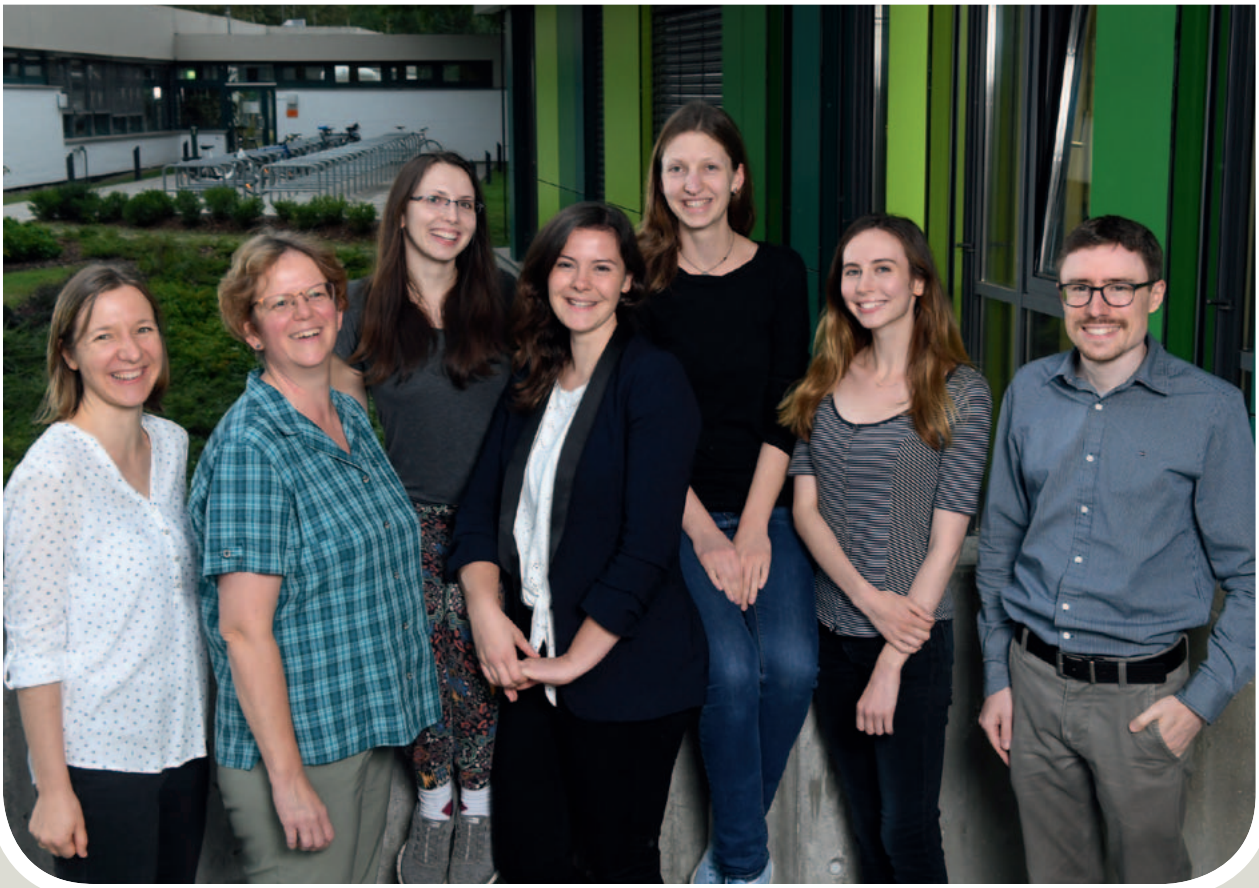


Figure 4.
In vitro neuronal differentiation as a model for studying cell lineage determination

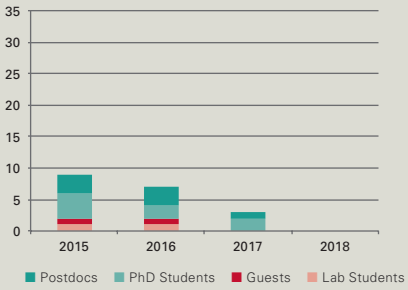


Ørom Lab – Research Group Non-coding RNA

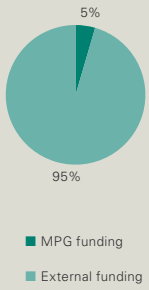
Ulf Andersson Vang Ørom, PhD
ulf.orom@mbg.au.dk

The group has been established in January 2012. Due to the appointment of Ulf Ørom as associate professor at the University of Aarhus, Denmark, the lab has been closed in November 2017.

Team



Funding



Focus areas

- Long non-coding RNA
- miRNA biogenesis
- Nascent RNA

Research concept

In my lab we have worked towards the functionality of non-coding RNAs by interrogating where and how functional interactions with proteins take place. In this direction we have described the nuclear RNA-protein interactome (Conrad et al. Nat Commun 2016) and the nascent m6A RNA methylation pattern (Louloupi et al. Cell Rep 2018). We show for the long ncRNA A-ROD as well as for primary miRNA processing an important involvement of chromatin-release dynamics in non-coding RNA processing and function (Ntini et al. Nat Commun 2018; Louloupi et al. RNA 2017).

Scientific highlights

- Chromatin retention is responsible for long ncRNA function [Ntini et al. Nat Commun 2018]
- m6A Modification on nascent RNA controls splicing efficiency [Louloupi et al. Cell Rep 2018]

Scientific honours

- Annita Louloupi – Invitation to the 68th Lindau Nobel Laureate Meeting, 2018
- Ulf Ørom – Appointed Associate Professor at Aarhus University, 2017
- Ulf Ørom – Hallas-Møller Investigator Grant, Novo Nordisk Foundation, 2017

Results of completed projects during reporting period

Transient N-6-methyladenosine transcriptome sequencing reveals a regulatory role of m6A in splicing efficiency

Splicing efficiency varies among transcripts, and tight control of splicing kinetics is crucial for coordinated gene expression. N-6-methyladenosine (m6A) is the most abundant RNA modification and is involved in regulation of RNA biogenesis and function. The impact of m6A on regulation of RNA splicing kinetics is unknown. Here, we provide a time-resolved high-resolution assessment of m6A on nascent RNA transcripts and unveil its importance for the control of RNA splicing kinetics. We find that early co-transcriptional m6A deposition near splice junctions promotes fast splicing, while m6A modifications in introns are associated with long, slowly processed introns and alternative splicing events. In conclusion, we show that early m6A deposition specifies the fate of transcripts regarding splicing kinetics and alternative splicing.

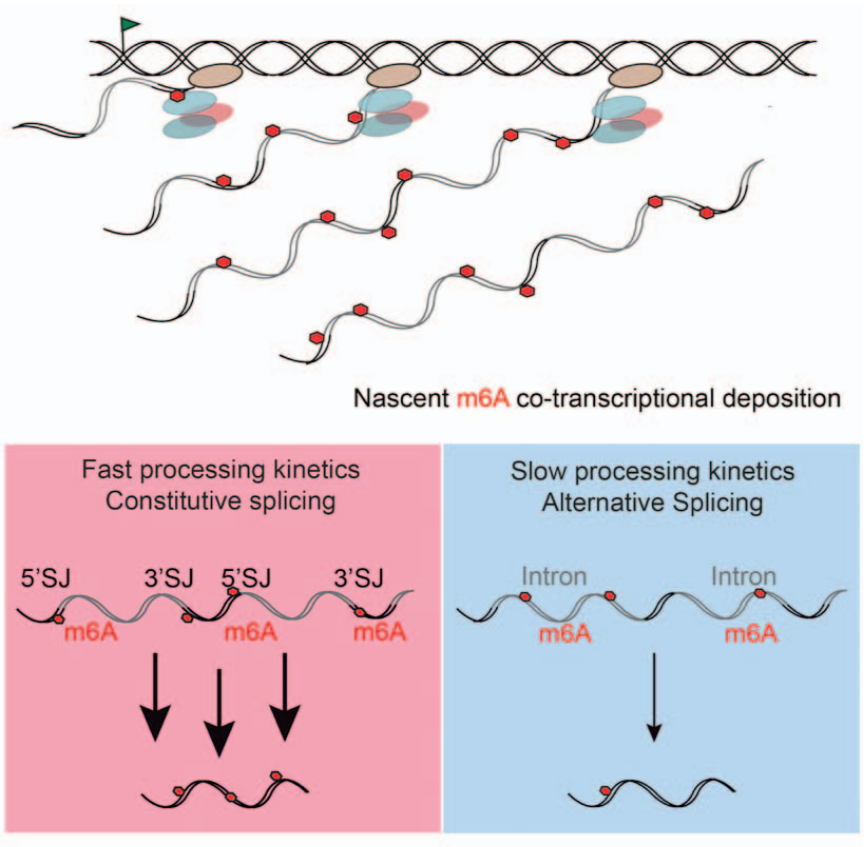


Figure 1.
Overview of how m6A at splice junctions and in introns affect splicing efficiency.

Louloupi A, Ntini E, Conrad T & Ørom UAV. Transient N-6-methyladenosine transcriptome sequencing reveals a regulatory role of m6A in splicing efficiency. Cell Rep 23(12):3429-3437 (2018)

Long ncRNA A-ROD activates its target gene *DKK1* at its release from chromatin

Long ncRNAs are often enriched in the nucleus and at chromatin, but whether their dissociation from chromatin is important for their role in transcription regulation is unclear. Here, we group long ncRNAs using epigenetic marks, expression and strength of chromosomal interactions; we find that long ncRNAs transcribed from loci engaged in strong long-range chromosomal interactions are less abundant at chromatin, suggesting the release from chromatin as a crucial functional aspect of long ncRNAs in transcription regulation of their target genes. To gain mechanistic insight into this, we functionally validate the long ncRNA A-ROD, which enhances *DKK1* transcription via its nascent spliced released form. Our data provide evidence that the regulatory interaction requires dissociation of A-ROD from chromatin, with target specificity ensured within the pre-established chromosomal proximity. We propose that the post-transcriptional release of a subset of long ncRNAs from the chromatin-associated template plays an important role in their function as transcription regulators.

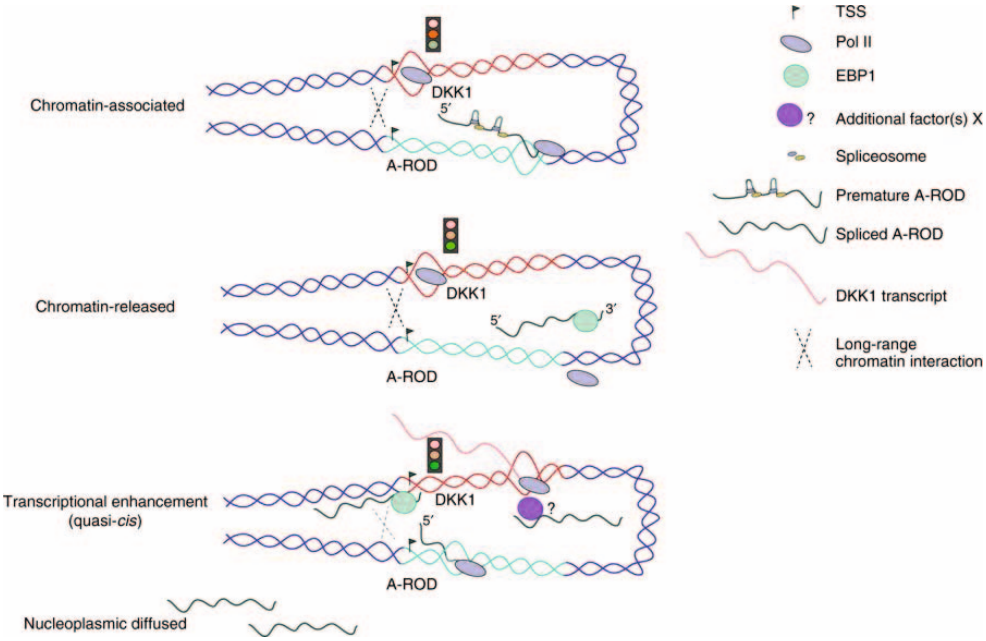


Figure 2.
Long ncRNA A-ROD is functional as an enhancer-like long ncRNA at its release from chromatin where it recruits transcription factors to the promoter of its target gene *DKK1*.

Ntini E, Louloupi A, Liz J, Muino JM, Marsico A & Ørom UAV. Chromatin-release of the long ncRNA A-ROD is required for transcriptional activation of its target gene DKK1. Nat Commun 9(1):1636 (2018)

Microprocessor dynamics shows co- and post-transcriptional processing of pri-miRNAs

miRNAs are small regulatory RNAs involved in the regulation of translation of target transcripts. miRNA biogenesis is a multistep process starting with the cleavage of the primary miRNA transcript in the nucleus by the Microprocessor complex. Endogenous processing of pri-miRNAs is challenging to study and the *in vivo* kinetics of this process are not known. Here, we present a method for determining the processing kinetics of pri-miRNAs within intact cells over time, using a pulse-chase approach to label transcribed RNA during 15 min, and follow the processing within a 1-hour window after labeling with bromouridine. We show that pri-miRNAs exhibit different processing kinetics ranging from fast over intermediate to slow processing, and we provide evidence that pri-miRNA processing can occur both cotranscriptionally and post-transcriptionally.

Louloupi A, Ntini E, Liz J & Ørom UAV. Microprocessor dynamics shows co- and post-transcriptional processing of pri-miRNAs. RNA 23(6):892-898 (2017)

Serial interactome capture of the human cell nucleus

Novel RNA-guided cellular functions are paralleled by an increasing number of RNA-binding proteins (RBPs). Here we present 'serial RNA interactome capture' (serIC), a multiple purification procedure of ultraviolet-crosslinked poly(A)-RNA-protein complexes that enables global RBP detection with high specificity. We apply serIC to the nuclei of proliferating K562 cells to obtain the first human nuclear RNA interactome. The domain composition of the 382 identified nuclear RBPs markedly differs from previous IC experiments, including few factors without known RNA-binding domains that are in good agreement with computationally predicted RNA binding. serIC extends the number of DNA-RNA-binding proteins (DRBPs), and reveals a network of RBPs involved in p53 signalling and double-strand break repair. serIC is an effective tool to couple global RBP capture with additional selection or labeling steps for specific detection of highly purified RBPs.

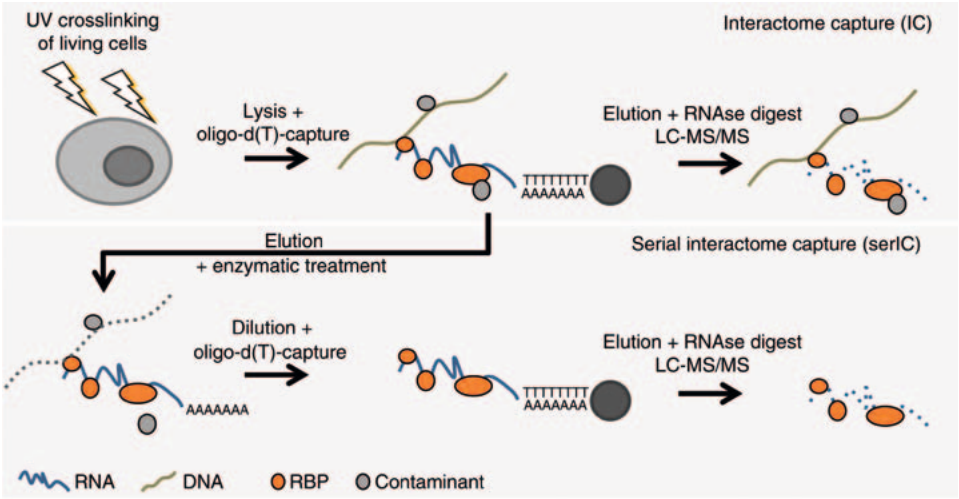


Figure 3.
Graphical overview of tandem purification of RNA-binding proteins used for chromatin interactome capture.

Conrad T, Albrecht AS, de Melo Costa VR, Sauer S, Meierhofer D & Ørom UAV. Serial interactome capture of the human cell nucleus. Nat Commun 7:11212 (2016)

Ongoing and future projects

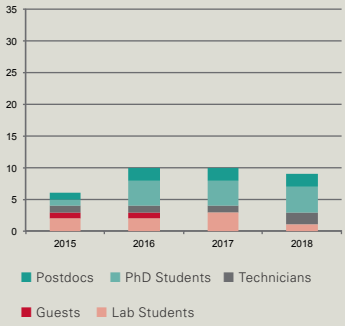
All projects have been successfully completed. In December 2017, the lab has moved to Aarhus University, Denmark.

Schulz Lab – Max Planck Research Group Regulatory Networks in Stem Cells

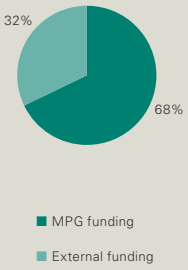
The group has been established in January 2015.

Dr. Edda G. Schulz
edda.schulz@molgen.mpg.de
++49 30 8413-1226

Team



Funding



Focus areas

- *Systems Biology* – gene-regulatory networks during development
- *Epigenetics* – regulatory principles governing the onset of X-chromosome inactivation
- *Stem Cells* – X-chromosomal dosage effects on pluripotency and differentiation

Research concept

Complex gene regulatory networks ensure the tight coordination of different molecular programs during embryonic development. To identify the underlying control principles we study the interplay of two essential developmental processes, namely exit from the pluripotent state and X-chromosome inactivation. Both processes integrate information on X-chromosomal dosage, which promotes up-regulation of *Xist*, the master regulator of X inactivation, and attenuates the exit from the pluripotent state. We use an interdisciplinary approach to understand how the underlying regulatory network reads out information on X-chromosomal dosage and ensures that each female cell up-regulates *Xist* from exactly one out of two chromosomes during exit from the pluripotent state.

Scientific highlight

- Developed the first model of the *Xist* regulatory network that can explain *Xist* patterns across evolution [Mutzel et al, BioRxiv, 2017]

Scientific honours & selected invited talks

- HFSP Career development award (2018)
- Invited speaker at the IRI Symposium "Advances and challenges in single cell biology", Berlin, Germany (2018)
- Invited speaker at the symposium "The X chromosome - Genetics and Epigenetics", College de France, Paris (2018)
- Invited speaker at the Conference "X-chromosome inactivation: a tribute to Mary Lyon", London, UK (2016)
- Invited speaker at the IRI Symposium "From RNA Pools to single RNA Molecules", Berlin, Germany (2015)

Results of completed projects during reporting period

Identifying the regulatory principles that control the onset of X inactivation

[Verena Mutzel, Ilona Dunkel]

X-chromosome inactivation (XCI) is initiated by mono-allelic up-regulation of *Xist* from either the paternal or the maternal X chromosome. Through mathematical modeling and simulations, combined with experimental testing of model predictions, we have performed a systematic comparison of alternative network architectures and identified the core regulatory network required to ensure female-specific and mono-allelic *Xist* expression (figure 1). A global toggle switch, mediated by a *trans*-acting *Xist* activator must be coupled to a local toggle switch mediated by a *cis*-acting *Xist* repressor. This network can recapitulate the correct *Xist* expression patterns in male, female and several aneuploid cell lines, but also the diverse *Xist* patterns observed in different mammalian species.

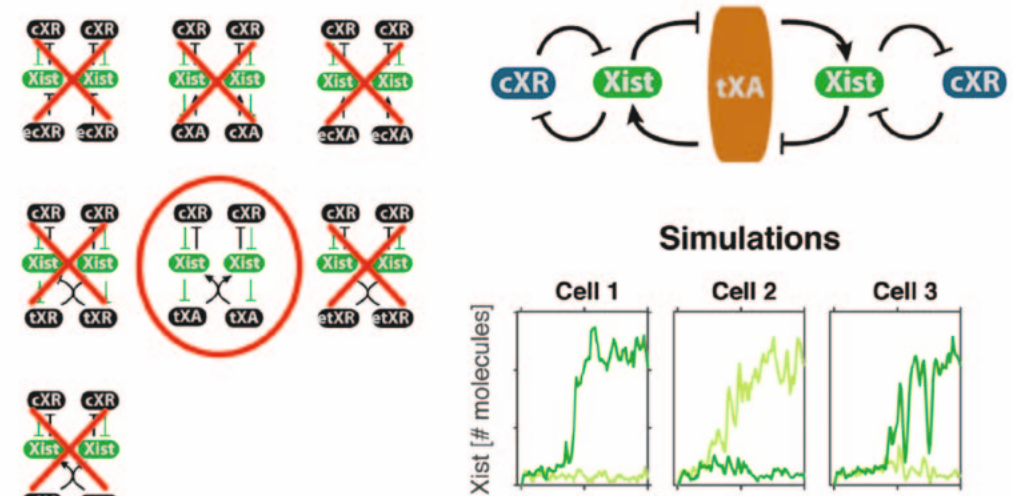


Figure 1.
A systematic model comparison to identify the network that can ensure mono-allelic *Xist* up-regulation in females.

Mutzel V, Okamoto I, Dunkel I, Saitou M, Giorgetti L, Heard E & Schulz EG. A symmetric toggle switch explains the onset of random X inactivation in different mammals. BioRxiv, 2017, doi: https://doi.org/10.1101/204909

Ongoing and future projects

How does X-chromosomal dosage stabilize the pluripotent state?

[Oriana Genolet, Zeba Sultana, Anna Monaco]

Double X-chromosomal dosage shifts female ES cells towards the naive pluripotent state and delays their differentiation through inhibiting the MAPK signaling pathway (figure 2). Through a pooled lentiviral CRISPR screen we have identified X-linked MAPK inhibitors. These were then further characterized in a series of smaller screens with respect to their

ability to modulate Mek phosphorylation, to increase pluripotency factor expression and to delay differentiation. In this way we have identified a small number of genes that mediate X-dosage effects in ES cells, which are currently being characterized in more detail. Moreover, we have profiled the signaling state of cells with one and two X chromosomes through multiplexed phosphorylation measurements in response to systematic perturbations with small molecules. In collaboration with Nils Blüthgen (Charite, Berlin) we are currently reconstructing the signaling network for each cell type based on this data set, through modular response analysis. Through this approach we expect to identify the differences in the signaling networks between male and female ES cells and to gain a better understanding of signaling cross-talk in ES cells (figure 2).

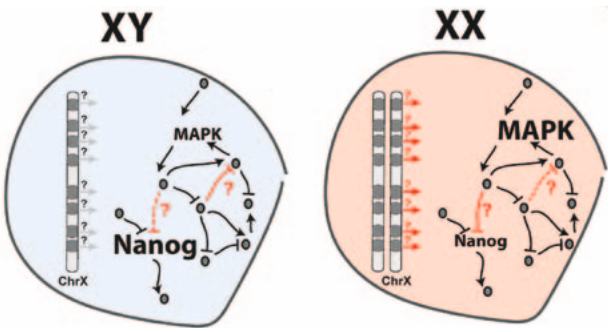


Figure 2.
Identification of sex differences in the ESC signalling network and of the underlying X-linked signaling modulators.

Identifying *Xist* regulators that transmit information on X dosage and differentiation

[Liat Ravid-Lustig, Vera Schmiedel, Guido Pacini, Ilona Dunkel]

Xist up-regulation is restricted to female cells and to a precise developmental window. To identify the underlying regulators we perform pooled CRISPR screens, based on a protocol we have developed to sort cells based on *Xist* expression. Through combining this approach with a CRISPR activation system (Suntag), we have screened the entire X chromosome for putative *Xist* activators (figure 3). Moreover, we will use the same approach to screen a library of putative developmental regulators of *Xist* identified by single-cell RNA-sequencing and to identify *cis*-regulatory elements in the *Xist* locus in collaboration with Charles Gersbach, Duke University, Durham. To further characterize each regulator, we will use CRISPR to perturb all *Xist* regulators in a multiplexed manner and analyze the effects through single-cell transcriptomics. In the long term we aim to use this data to functionally dissect the *Xist* regulatory network through a data-driven modeling approach.

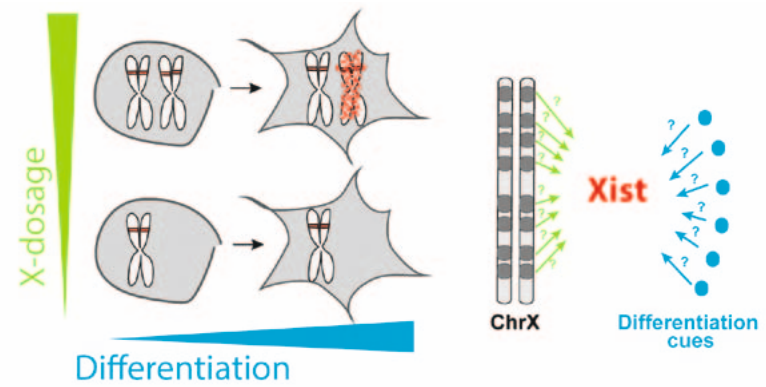


Figure 3.
Screening for regulators that control female-specificity and developmental timing of *Xist*.

Dissecting X-chromosome inactivation with single cell transcriptomics

[Ilona Dunkel, Guido Pacini, in collaboration with Annalisa Marsico, MPIMG]

Since each individual cell randomly silences either the paternal or the maternal X chromosome, the genome-wide analysis of the X inactivation process is challenging. In collaboration with Annalisa Marsico we use single-cell RNA-sequencing to investigate the determinants and effects of X-dosage compensation. We have performed a time-course experiment of differentiating ES cells with one or two X chromosomes and analyzed it in an allele-specific fashion (figure 4). This data set allows us to investigate (1) silencing dynamics in a chromosome-wide manner, (2) effects of X-dosage changes during X inactivation and (3) the transient bi-allelic expression state that we have recently described (Mutzel et al., BioRxiv, 2017).

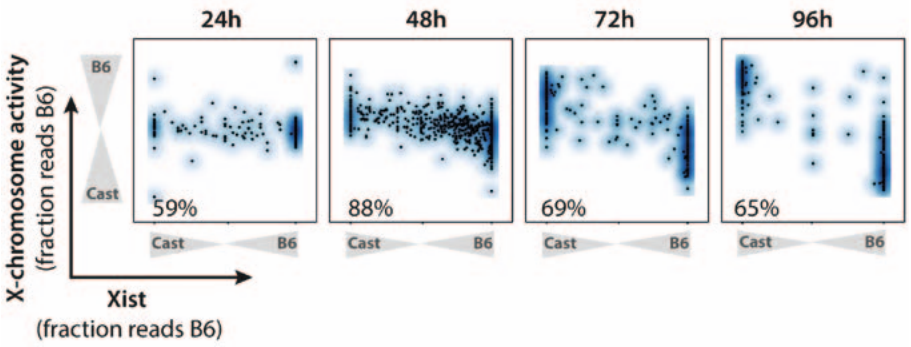


Figure 4.
Relative expression from the two alleles (B6/Cast) of *Xist* (x-axis) and the whole X chromosome (y-axis) in differentiating female ES cells.

Regulatory effects of antisense transcription

[Verena Mutzel, Rutger Gjaltema]

Our analysis of the *Xist* regulatory network predicted a *cis*-acting positive feedback, which we proposed to be mediated by *Xist's* repressive antisense transcript *Tsix*. Through mathematical modeling we could show that mutual repression of *Xist* and *Tsix* can generate such a local bistable switch. Through theoretical and experimental approaches we are currently investigating, whether antisense transcription might generate similar switch-like behavior at other loci. We use mathematical modeling to investigate the constraints on locus architecture (length of overlap, antisense transcription through promoters) and model parameters for switch-like behavior to occur. Moreover, we measure nascent transcription in a genome-wide manner through TT-Seq to identify sense-antisense pairs that switch their transcriptional states during ES cell differentiation. Through inducible CRISPR activation and repression we will subsequently modulate antisense transcription to test model predictions and understand whether switching between states can be induced ectopically.

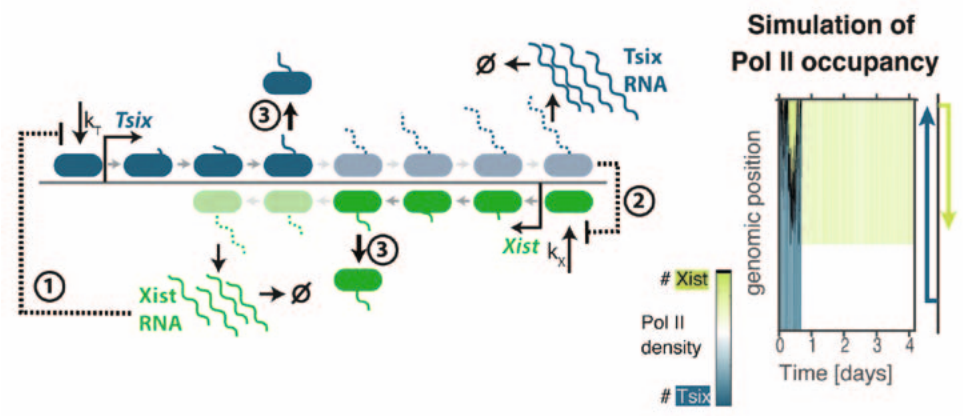
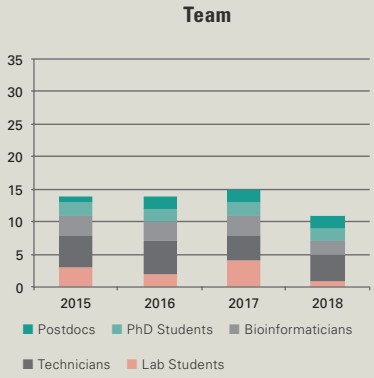


Figure 5.
Antisense transcription at the *Xist/Tsix* locus (left) and simulation of the *Tsix*-to-*Xist* transition at the onset of X inactivation (right).

Yaspo Lab – Research Group Gene Regulation & Systems Biology of Cancer

The group has been established in 1996 at the former Dept. of Vertebrate Genomics and moved to the Otto Warburg Laboratory in December 2014.

Dr. Marie-Laure Yaspo
yaspo@molgen.mpg.de
++49 30 8413-1356



Focus areas

- *Cancer omics* – deep profiling of tumors by NGS and imaging mass cytometry
- *Personalized medicine* – translational omics, metastatic melanoma
- *Immune status* – deciphering human B and T cell repertoires in cancer and autoimmune conditions

Research concept

We are interested in deciphering molecular mechanisms underlying tumorigenesis and drug sensitivity in several cancer types. Sequencing genomes and transcriptomes of tumor cohorts, we developed integrative bioinformatics pipelines for identifying tumor molecular features, as well as biomarkers of treatment response. In a translational trial, we transform deep NGS tumor/patient data into personalized reports that we have implemented in clinical setting, supporting treatment choice for metastatic melanoma. We are also investigating functional relationships between cancer cells and microenvironment, combining deconvolution of bulk transcriptomics with CyTOF-based imaging mass cytometry allowing spatially resolved single cell proteomics. Further, we developed methods for identifying B and T receptor clonotypes for monitoring adaptive immune response in disease contexts.

Scientific highlights

We contributed deep molecular profiling, in particular bulk RNAseq data, for several adult and pediatric tumor types in the framework of European and National cancer genomics projects, generating high-quality standardized datasets for the scientific community that led to > 20 high impact publications led by consortium partners or by us. In Treat20plus,

we coordinate a project for transforming tumor molecular analysis into clinical reports implemented at hospital tumor boards.

- Molecular evolution of early onset prostate cancer identifies novel molecular risk markers and clinical trajectories [Gerhauser et al. Cancer Cell 2018]
- Predicting sensitivity to EGFR inhibitors in colon cancer [Schütte et al. Nat Commun 2017]
- NGS profiling of fatal TCF3-HLF acute lymphoblastic leukemia identifies therapeutic options [Fischer et al. Nat Genet 2015]

Scientific honours & selected invited talks

- Invited speaker at 33rd German Cancer Congress, Berlin, 2018
- Keynote speaker at the Annual Meeting of CESAR (Central European Society for Anticancer Drug Research), Berlin, 2018
- Invited speaker at Roche Symposium "Molecular tumor profiling" Vienna, 2017
- Invited speaker at DKTK annual retreat, Heidelberg, 2017
- VIB Conference series „Revolutionizing NGS", Antwerp, Belgium, 2017

Results of completed projects during reporting period

Treat20Plus: a personalized medicine pilot trial for metastatic melanoma

[Thomas Risch, Rene Buschow, Praneeth Devulapally, Marie-Laure Yaspo serving as coordinator, project in collaboration with Charite Comprehensive Cancer Center and Alacris Theranostics, Berlin]

Great strides have been made for treating cutaneous metastatic melanoma with *BRAF* V600E and/or a high mutation burden eligible for immune checkpoint inhibitors (ICIs) boosting the immune response against tumor cells. But for BRAF-wildtype cases, sun-shielded mucosal, acral, or uveal melanomas (UM) presenting different etiologies with low mutation burden, the options are limited. We profiled the genomes, exomes and transcriptomes of 65 "hard cases" to identify somatic landscapes and key oncogenic events. In this translational pilot, we designed NGS-based comprehensive molecular tumor analysis (CMTA) reports implemented in the clinic, which guided treatment for >1/3 patients. We established patient-derived spheroids for pre-clinical screening platform. The deconvolution of bulk RNAseq allows us to infer the stromal and tumor microenvironment (TME). Figure 1A shows two main UM groups with different immune cells infiltrates. We combine RNAseq data with spatially resolved imaging mass cytometry (IMC) at single cell resolution on histological sections (figure 1B) for deciphering the TME and its role in the response to ICIs.

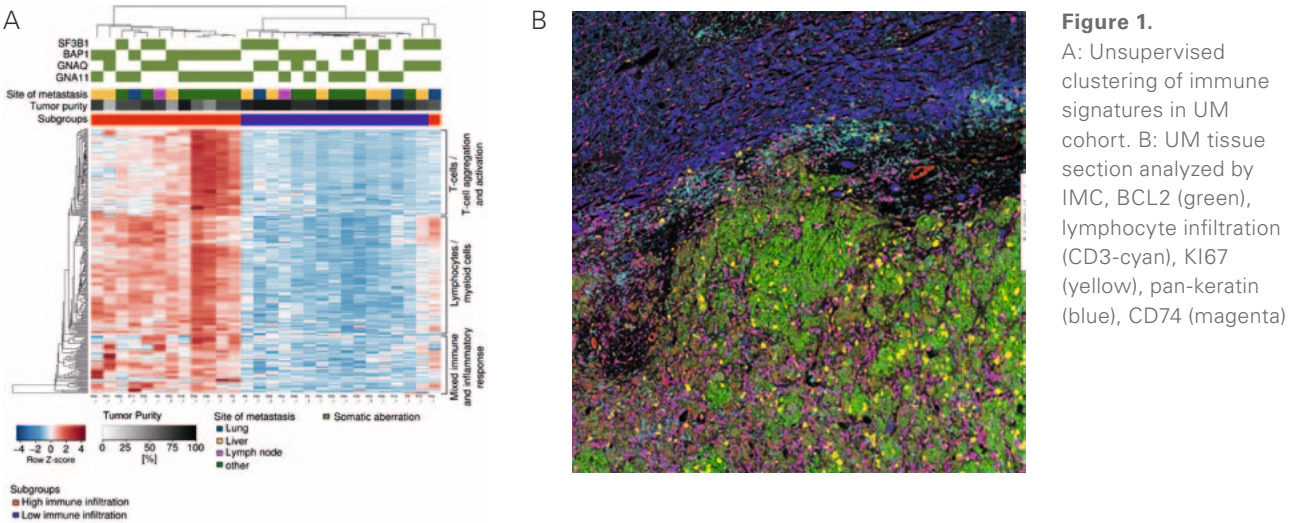


Figure 1.
A: Unsupervised clustering of immune signatures in UM cohort. B: UM tissue section analyzed by IMC, BCL2 (green), lymphocyte infiltration (CD3-cyan), KI67 (yellow), pan-keratin (blue), CD74 (magenta)

Schütte M, Ogilvie LA, Rieke DT, Lange BMH, Yaspo ML & Lehrach H. Cancer precision medicine: Why more is more and DNA is not enough. Public Health Genomics 20(2):70-80 (2017) (review)

Oncotrack: Innovative Medicines Initiative (IMI, EU) "Identification of novel biomarkers of drug response in colorectal cancer"

[Thomas Risch, Nilofar Abdavi-Azar, Christine Jandrasits, Vyacheslav Amstislavskiy]

Colorectal cancer (CRC) is a complex disease associated with deregulated WNT, MAPK, and EGFR pathways. Tumors without mutations in *BRAF*, *KRAS* or *NRAS* are eligible to receive antibodies against EGFR, however there is an unmet need for identifying novel positive biomarkers predicting drug sensitivity. We performed a multi-omics analysis of a prospective cohort of > 100 CRCs (stages I-IV) with their derived xenograft and organoid models. We identified specific CRC subgroups, and found that the models recapitulated tumor types featuring stemness, Wnt signalling, or colonic specialized processes, but lost the TME signature characteristic of one tumor group. This feature was instrumental for analysing pre-clinical drug response data generated in the consortium. Using support vector machines, we identified novel gene expression signatures predictive of 5-FU, Avastin, and EGFR inhibition treatment outcomes.

Schütte M, Risch T, Abdavi-Azar N, ...[4 authors]... Jandrasits C, Borodina T, Amstislavskiy V, Worth CL, Schweiger C, Liebs S, Lange M, Warnatz HJ, Butcher LM, Barrett JE, Sultan M, ... [26 authors] ... Lehrach H & Yaspo ML. Molecular dissection of colorectal cancer in pre-clinical models identifies biomarkers predicting sensitivity to EGFR inhibitors. Nat Commun 8:14262 (2017)

Identifying molecular pathways and therapy options in pediatric acute lymphoblastic leukemia

[Thomas Risch, Hans-Jörg Warnatz, Vyacheslav Amstislavskiy, together with Meryem Ralser, Marc Sultan, two former team members, as part of the UFOplan consortium]

Acute lymphoblastic leukemia (ALL), the most common type of cancer in children, can occur in various forms differing by their etiology, severity and response to therapies. Despite large improvements achieved for ALL overall survival, the aggressive subtype associated with t(17;19) chromosomal translocation fusing *TCF3* and *HLF* genes, remains as yet incurable. To decode the molecular basis of this dismal leukemia, we compared the molecular pattern of t(17;19) versus the t(1,19) good prognosis subtype associated with *TCF3-PBX1* fusion. The cardinal differences characterizing t(17;19) ALL were activating mutations in the RAS pathway, loss of *PAX5*, and a global expression profile reflecting a de-differentiated stem-like phenotype with elevated BLC2 levels. This work led to design new therapeutic approaches in phase I clinical trials targeting BCL2.

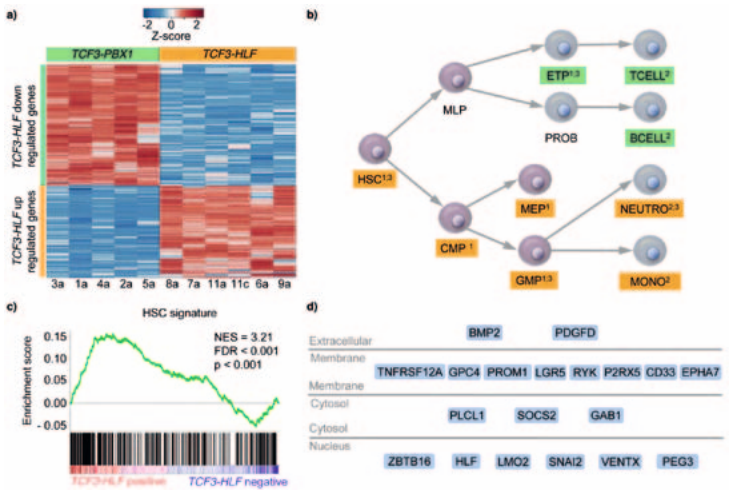


Figure 2.
Heatmap of the 401 differentially expressed genes between the two ALL subtypes.
(b) Enriched hematopoietic stages in *TCF3-HLF* (orange) and *TCF3-PBX1* (green) ALL: hematopoietic stem cells (HSC), common myeloid progenitors (CMP), lymphoid-specified progenitors (GMP and MEP), neutrophils (NEUTRO), monocytes (MONO), multilymphoid progenitor (MLP), early T cell precursors (ETP), pro-B cells (PROB), T cells (TCELL), and B cells (BCELL).
(c) Enrichment plot for the HSC signature
(d) Components of the *TCF3-HLF*- ALL signature reveal stem cells features

Fischer U, Forster M, Rinaldi A, Risch T, Sungalee S, Warnatz HJ, ...[4 authors] ... Sultan M, Tchinda J, Worth CL, Amstislavskiy V, ...[32 authors]... Ralser M, ...[18 authors]... Stanulla M & Yaspo ML. Genomics and drug profiling of fatal *TCF3-HLF*-positive acute lymphoblastic leukemia identifies recurrent mutation patterns and therapeutic options. Nat Genet 47(9):1020-1029 (2015)

Large-scale transcriptomics

[Thomas Risch, Hans-Jörg Warnatz, Vyacheslav Amstislavskiy, Christine Jandrasits; Sören Matzk as part of the projects EU-Blueprint, ICGC Pedbrain and ICGC prostate cancer]

We generated high quality, large-scale transcriptomic data and developed bioinformatics analytic procedures in the course of those projects, altogether contributing an essential resource to the community and a series of publications led by consortium partners, reporting the integration of epigenetic and transcriptomic landscapes of human blood cells and leukemias (e.g. Chen et al. Cell 2016), targetable *MET* gene fusions in glioblastoma (Int. Cancer Genome Consortium Pedbrain Nat Med 2016), active enhancers in medulloblastoma subgroups (Lin et al. Nature 2016) and gene expression signatures predictive of disease severity in early onset prostate cancer.

Gerhauser C*, Favero F*, Risch,T*, ... [19 authors]...Matzk S, Warnatz H-J, Amstislavskiy V, ...[25 authors]...Yaspo M-L (co-senior author), Korbel J O, Schlomm T & Weischenfeldt J. Molecular evolution of early onset prostate cancer identifies novel molecular risk markers and clinical trajectories. Cancer Cell 2018, in press.
*shared first author

Deciphering immune cell repertoires: H2020 TREGeneration: Monitoring T cell repertoires by next generation sequencing following Treg infusion

[Hans-Jörg Warnatz, Praneeth Devulapally]

We have developed a simple method for sequencing paired heavy- and light-chain antibody repertoires. We are monitoring TCR repertoire modifications induced by regulatory T cells (Treg) therapy in phase I/II clinical trials designed to treat patients suffering from Graft-versus-Host Disease (GvHD) after bone marrow transplantation. We are using our quantitative DeepImmune analysis for monitoring TCRβ receptors in longitudinal samples from donor bone marrow and infused GvHD patient samples (figure 3A). We detected up to 150,000 clonotypes per sample with specific repertoires in different T cell subsets (figure 3B), tracking the fate of infused donor Treg cells (figure 4C) and clonotype patterns associated with therapy response.

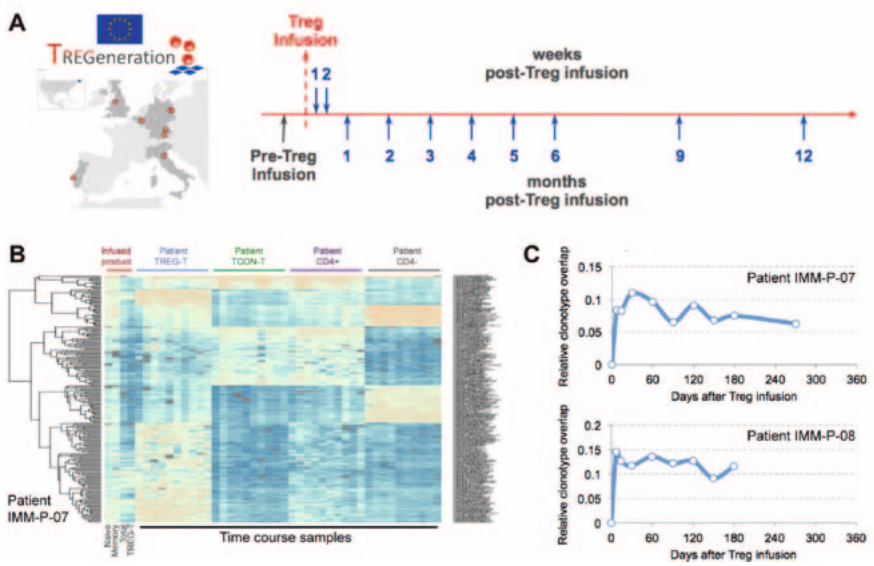


Figure 3.
A: Clinical trials.
B: Analysis of TCRβ sequences in T cell subsets. C: TCRβ clonotypes specific to infused Treg cells.

Devulapally PR, Bürger J, Mielke T, Konthur Z, Lehrach H, Yaspo ML, Glökler J & Warnatz HJ. Simple paired heavy- and light-chain antibody repertoire sequencing using endoplasmic reticulum microsomes. Genome Med 10(1):34 (2018)

Ongoing and future projects

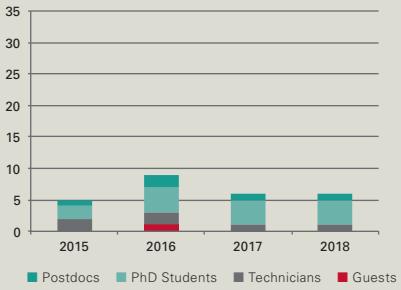
We are currently evaluating clinical outcomes and data generated in Treat20Plus, focusing on spatial architecture, TME and immune infiltration of the different melanoma types; we plan a multi-centric follow-up European project. In IMI CancerID, our ongoing contribution is to sequence mutations in liquid biopsies. For pediatric leukemia, we now focus on identifying the targets of TCF3-HLF fused protein (cooperation with J-P Bourquin, Zurich, Switzerland), analyzing Down syndrome-associated myeloid leukemias (cooperation with J- H Klusmann, Halle, Germany), and start a project addressing relapse in pediatric ALL (cooperation with C. Eckert, Charite, Berlin, Germany). We are partner in a DFG-funded pediatric leukemia network, coordinated by A. Bergmann, Hannover, and in the EU-iPC-Pediatric Cure initiative, performing functional screens of cancer-relevant genes by CRISPR technologies and exploring tumor architecture by imaging mass cytometry.

Zi Lab – Research Group Cell Signaling Dynamics

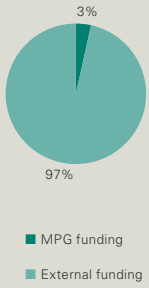
The group has been established in June 2014.

Dr. Zhike Zi
zhike.zi@molgen.mpg.de
++49 30 8413-1660

Team



Funding



Focus areas

- *Systems biology* – cell signaling dynamics
- *Signaling heterogeneity* – molecular mechanisms governing cell signaling heterogeneity
- *Synthetic biology* – spatiotemporal control of cell signaling using optogenetics

Research concept

Cell signaling networks are complex and dynamic systems that enable living cells to sense and respond to changes in their immediate environment. A challenge in cell signaling research is to understand how multiple signaling proteins act together and determine cell fate decisions. Our group uses a systems biology approach that combines mathematical modeling and experimental analyses to study the molecular mechanisms by which heterogeneous cell signaling dynamics control cell proliferation and apoptosis in different types of cells and in individual cells.

Scientific highlights

- Developed an optogenetic system (optoTGFBRs) that enables the precise control of TGF-β signaling in time and space [Li et al. ACS Synth Biol 2018]
- Mathematical modeling and quantitative experiments showed that tissue-specific Chk1 activation determines apoptosis by regulating the balance of p53 and p21

Scientific honours & selected invited talks

- Selected talk at the 3rd International SystemsX.ch Conference, Zurich, Switzerland, 2017
- Invited talk at Korea Advanced Institute of Science and Technology, South Korea, 2016

Results of completed projects during reporting period

Spatiotemporal control of TGF-β signaling with light

[Yuchao Li, Guoyu Wu, Difan Deng]

Transforming growth factor beta (TGF-β) is an important growth factor that regulates many cellular functions in development and disease. Although the molecular mechanisms of TGF-β signaling have been well studied, our understanding of this pathway is limited by the lack of tools that allow the control of TGF-β signaling at high spatiotemporal resolution. In this work, we developed an optogenetic system (optoTGFBRs) that enables the spatiotemporal control of TGF-β signaling at the single cell level (figure 1). This tool can be used to study the dynamics of canonical TGF-β signaling in response to complex inputs and could be generalized to other members of the TGF-β superfamily.

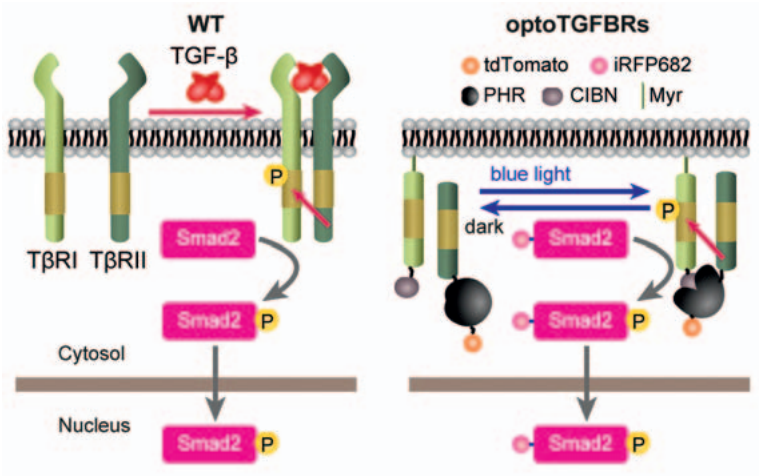


Figure 1.
The design of optoTGFBRs systems

Li Y, Lee M, Kim N, Wu G, Deng D, Kim JM, Liu X, Heo WD & Zi Z. Spatiotemporal control of TGF-β signaling with light. ACS Synth Biol 7(2):443-451 (2018)

Tissue-specific Chk1 activation determines apoptosis by regulating the balance of p53 and p21

[Marijn TM van Jaarsveld, Difan Deng]

The DNA damage response (DDR) protects cells against genomic instability. We found that the DDR is regulated differently in human breast and lung primary cells. Equal levels of cisplatin-DNA lesions caused stronger Chk1 activation in lung cells, leading to resistance. In contrast, breast cells were more resistant and showed more Chk2 activation in response to doxorubicin. Further analyses indicate that Chk1 activity played a regulatory role in p53 phosphorylation, whereas Chk2 activity was essential for p53 activation and p21 expression. We proposed a novel "friction model", in which the balance of p53 and p21 levels determines the apoptotic response in different tissues (figure 2). Our results help us to understand why distinct tissues respond differently to chemotherapy and suggest that modulating the balance of p53 and p21 dynamics could optimize the response to chemotherapy.

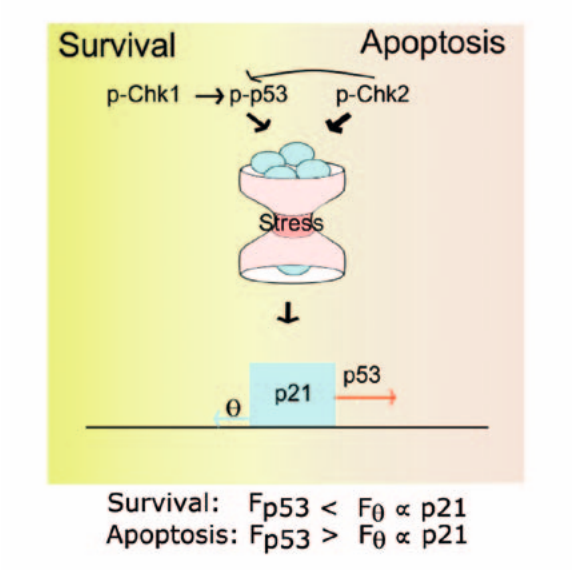


Figure 2.
The "friction model" of apoptosis response to DNA damage in different tissues

Ongoing and future projects

Unraveling the molecular mechanism for TGF-β signaling resistance in cancer cells

[Yuchao Li, Guoyu Wu, Difan Deng]

The duration of the TGF-β/Smad signaling response appears to be cell type specific. We hypothesize that the TGF-β signaling network is rewired at a few critical points in cancer cells, which leads to the adaptation and resistance in cell proliferation regulation. To test this hypothesis, we have developed mathematical models for all the possible feedback regulations (figure 3). The models are refined and optimized by fitting to quantitative experimental data sets. We are analyzing these models to predict where and when the relevant feedback regulators might act in the TGF-β signaling network. In addition, we are performing perturbation screen experiments to find the regulators and study their corresponding molecular mechanisms. This project will improve our understanding of the mechanisms that regulate TGF-β signaling dynamics in different cellular contexts.

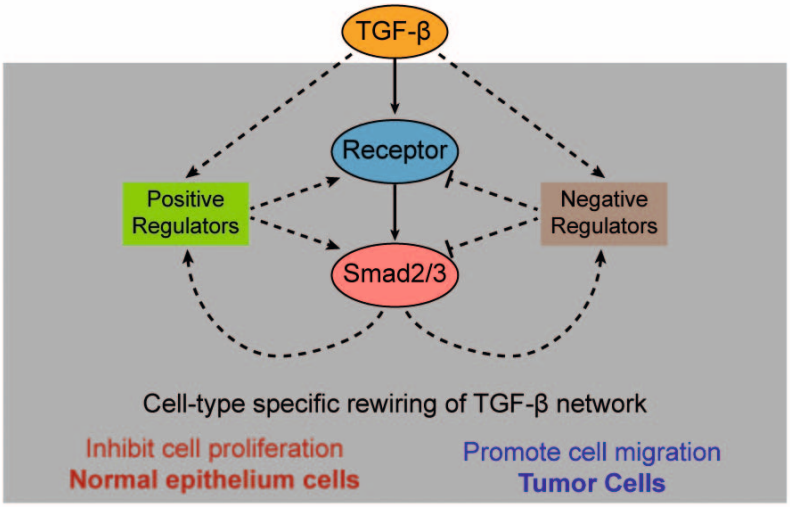


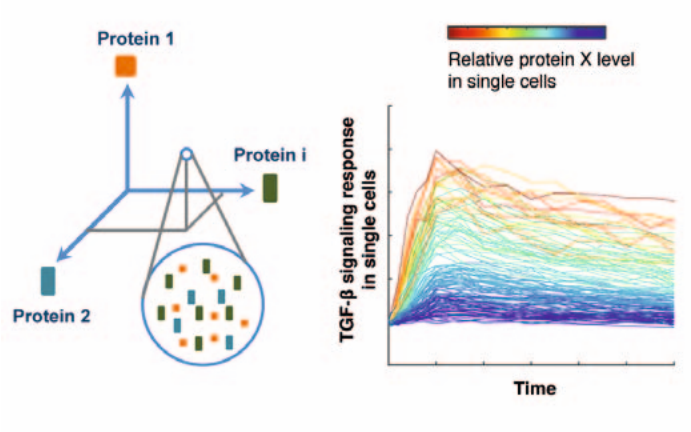
Figure 3.
Identify cell-type specific rewiring of TGF-β network for differential cell proliferation responses in normal and tumor cells

Studying the heterogeneous TGF-β signaling response in single cells

[Yuchao Li, Guoyu Wu, Hongqing Han]

Most components of biological networks have unavoidable stochastic fluctuations in their expression levels and activities. As a result, even genetically identical cells can react differently to the same input signal. Our recent data and others show that protein abundance variation is likely the main source of cell-to-cell variability in TGF-β signaling (figure 4). However, it is still not clear which protein is crucial in determining the heterogeneity of TGF-β signaling. In this project, we are using mathematical modeling and single cell assays to study how the dynamics of Smad signaling and downstream gene expression are controlled by the variation of TGF-β signaling proteins at the single cell level.

Figure 4.
Protein abundance variations determine cell-to-cell variability of TGF-β signaling

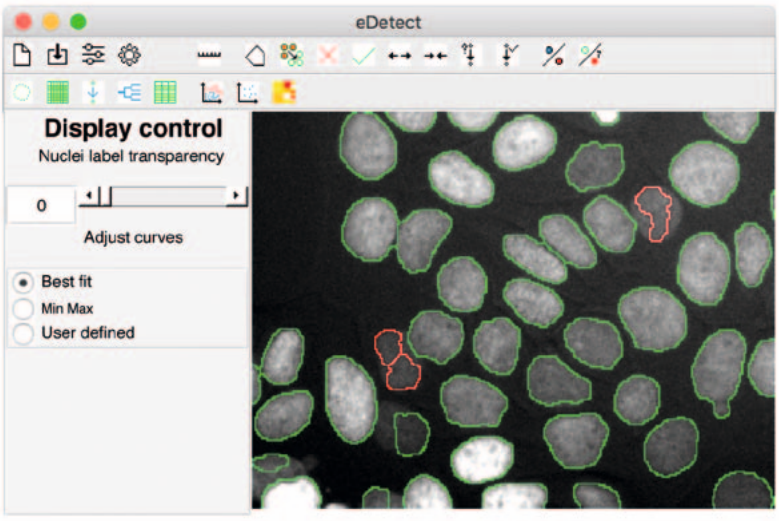


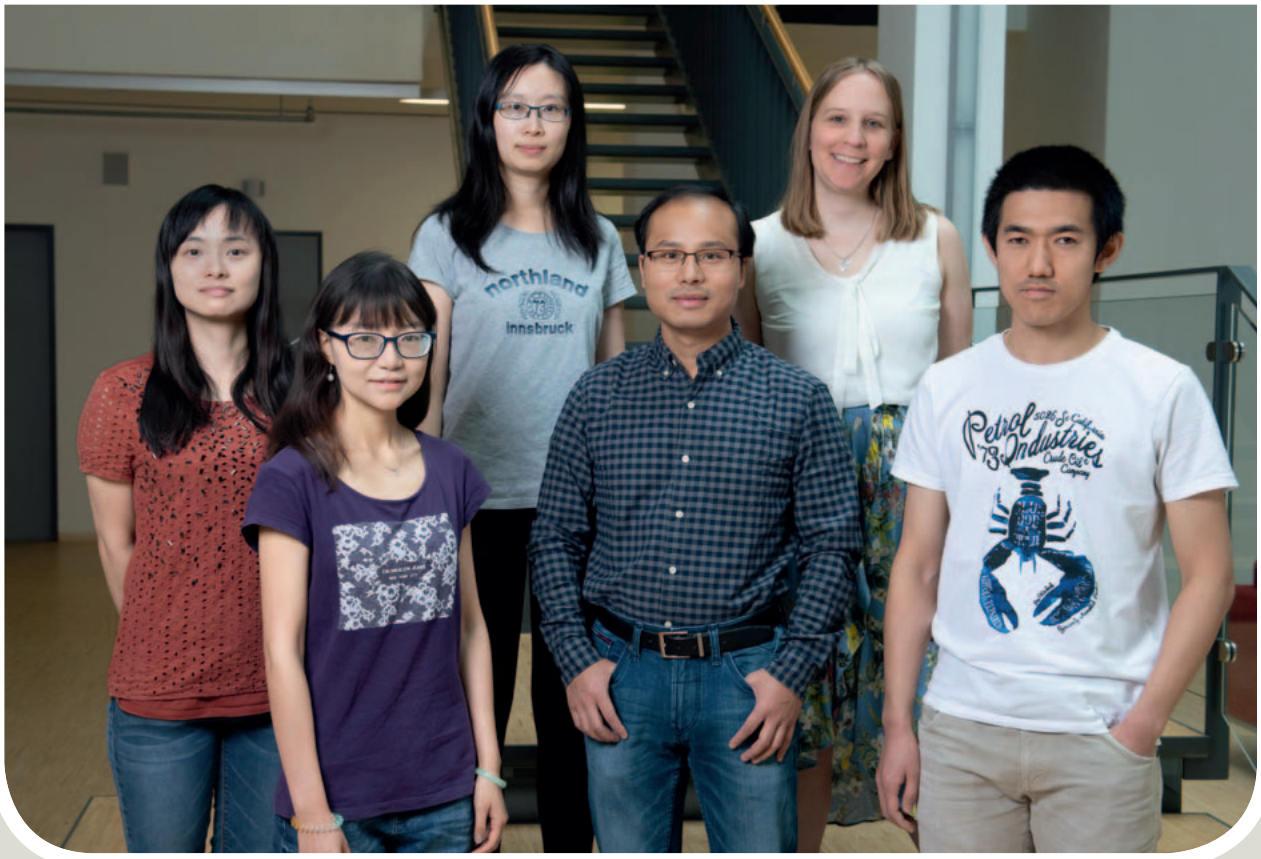
Developing a software tool for live-cell imaging data analysis

[Hongqing Han]

Live cell imaging plays an important role in quantitative understanding of cellular processes and dynamics. However, due to the challenges in cell segmentation and tracking, the automatic analysis of live cell imaging data often is error prone and thus requires tedious manual data correction. We are developing an error detection and correction tool, eDetect, which provides a powerful and user-friendly solution for the analysis of long-term live cell imaging data (figure 5). The eDetect tool visualizes cell image features based on principal component analysis, which provides an efficient way to detect and correct cell segmentation and tracking errors. This is a useful tool to support the quantification of single cell dynamics data.

Figure 5.
Interface of the eDetect tool for live cell imaging data analysis




northland
innsbruck
Petrol
Industries

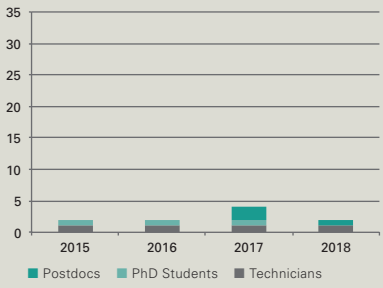
Lehrach Lab – Emeritus Group Vertebrate Genomics

The Department of Vertebrate Genomics has been active from September 1994 until November 2014. The Emeritus Group has continued work from December 2014 until November 2018.

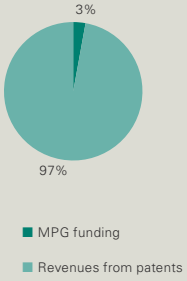
Prof. Dr. Hans Lehrach
lehrach@molgen.mpg.de
++49 30 8413-1220

Secretary:
Ingrid Stark
stark@molgen.mpg.de
++49 30 8413-1221

Team



Funding



Focus areas

- Systems medicine
- Digital twins
- -omics
- Cancer
- Autoimmune diseases
- Aging

Research concept

We are all different, with different genomes, environments and behaviour. It is therefore not surprising that we react very differently to the drugs we receive. For close to twenty years, my group has focussed on developing the basis of a new, data- and model-driven truly personalised medicine and prevention. This is also the core concept behind "Digital Twins for Better Health" (www.digitwins.org), a FET-Flagship application I am coordinating. Concepts and tools from this work will also impact biomedical research, since the tools available tend to be limited to "simple" biological mechanisms with strong causality; however, this restricted range is likely to make up only a small fraction of the mechanisms, evolution has used in developing complex organisms.

Scientific highlight

- Molecular dissection of colorectal cancer in pre-clinical models identifies predicting sensitivity to EGFR inhibitors [Schütte et al. Nat Commun 2017]

Scientific honours & selected invited talks

- Invited speaker at the 5th China-Canada Systems Biology Symposium, China, 2018
- Invited speaker at the 3rd European Conference on Translational Bioinformatics (ECTB), Barcelona, Spain, 2018
- Invited speaker at Labvolution mit Biotechnica 2017, Hannover, Germany, 2017
- Invited speaker at the CLINAM Conference 2017 – 10th European and Global Conference and Exhibition for Clinical Nanomedicine & Targeted Medicine, Basel, Switzerland, 2017
- Invited speaker at Advances in Cellular, Genomic and Epigenomic Insights on Environmental Mutagenesis and Health and 41st Annual Meeting of Environmental Mutagen Society of India (EMSI), Manipal, India, 2017

Results of completed projects during reporting period

In collaboration with the group of Marie-Laure Yaspo, a number of cancer analysis projects have been completed (described in more detail in the report of M.-L. Yaspo, see page 106 ff.).

See e.g. Schütte M, Ogilvie LA, Rieke DT, Lange BMH, Yaspo ML & Lehrach H. Cancer precision medicine: Why more is more and DNA is not enough. Public Health Genomics 20(2):70-80 (2017) (review)

Schütte M, Risch T, Abdavi-Azar N, ...[4 authors]... Jandrasits C, Borodina T, Amstislavskiy V, Worth CL, Schweiger C, Liebs S, Lange M, Warnatz HJ, Butcher LM, Barrett JE, Sultan M, ... [26 authors] ... Lehrach H & Yaspo ML. Molecular dissection of colorectal cancer in pre-clinical models identifies biomarkers predicting sensitivity to EGFR inhibitors. Nat Commun 8:14262 (2017)

Ongoing and future projects

Precision medicine of cancer

[Hans Lehrach, in collaboration with M.-L. Yaspo, MPIMG, and Alacris Theranostics, Berlin]

As a prerequisite to develop a truly personalised therapy choice for oncology patients, I continue my involvement in the deep molecular analyses of tumor and patient for a range of cancer types e.g. colon cancer (OncoTrack), metastatic melanoma (Treat20Plus) and triple-negative breast cancer (ITFoC) etc.).

Modelling the models: numerical experiments to develop and validate strategies for model validation and parameter estimation

[Hans Lehrach, in collaboration with Costin Ciudel, Lucian Itu, Transilvania University of Brașov (CTBAV), Aleksandra Kovachev, Christoph Wierling, Alacris Theranostics, Berlin]

To be able to move from precision medicine to a true personalisation of therapy, prevention and well being, we will have to rely heavily on mechanistic models of the molecular networks in the different cell types of every individual. These models are based on the wealth of information on molecular mechanisms of the interactions within cells and tissues, generated over decades of basic research. These processes are represented in PyBios, an object oriented modelling environment, and can then be individualised with data from the detailed molecular analyses on every patient. This object representation is used to generate large systems of ordinary differential equations, which can be solved numerically. However, this process critically depends on a large number of (typically unknown) parameters, which could, in theory, be determined by parameter optimisation strategies, if sufficient preclinical and/or clinical data and sufficient computing power are available.

To estimate the type, quality and amount of data required for models of different complexity, and to optimise its computational implementation, we take advantage of simulation results, providing "a ground truth" in an admittedly idealised situation, with potentially

unlimited (simulated) data of effectively unlimited accuracy. Once the overall strategy works on realistic models, we can explore the effect of e.g. reducing the amount of input data (e.g. only proteome data), explore the effect of data transformations (e.g. ratios versus absolute values), and determine the effect of different types of noise on the results of the analysis.

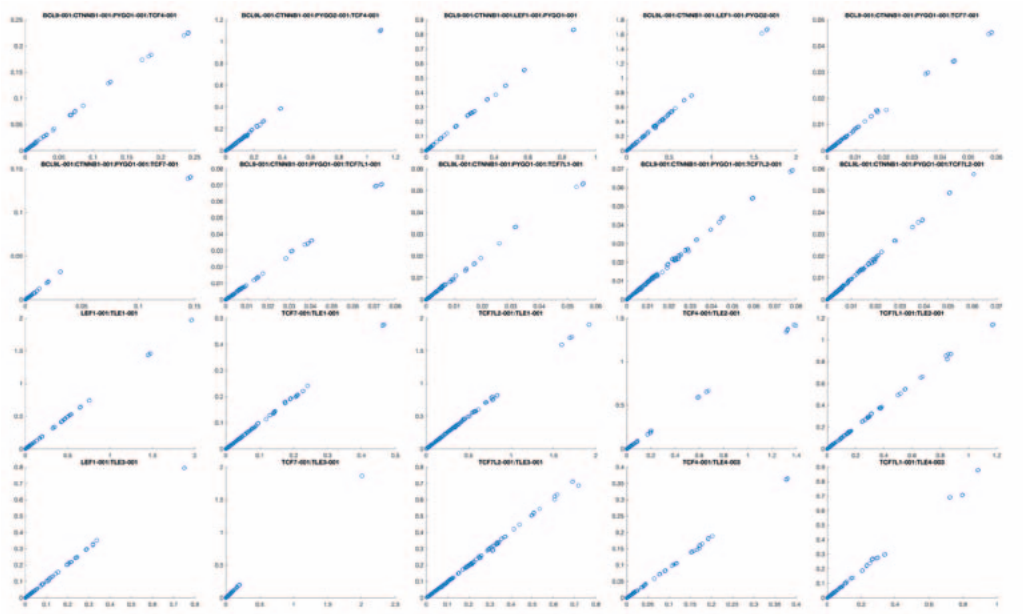


Figure 1.
Scatter plots of "real" (x coordinate) versus "predicted" (y coordinate) concentrations of 20 components (out of 40 used as "observables" in the cost function) in a Wnt pathway model. "Real" values are derived through simulations with an assumed parameter vector. "Predicted" values are generated from simulations using parameter vectors derived through parameter optimisation on a (simulated) training set assuming noise given by a flat distribution up to +/- 20% around the "ground truth" value.

See e.g. Wierling C, Kessler T, Ogilvie LA, Lange BMH, Yaspo ML & Lehrach H. Network and systems biology: essential steps in virtualising drug discovery and development. Drug Discov Today Technol 15: 33-40 (2015)

Reverse engineering of biological networks by single cell genetics

[Michael Boettcher, Hans Lehrach]

Determining the parameter values in large *in silico* models with many relevant parameters, and improving and validating the structure of these models will, almost certainly, require very large amounts of data. We are therefore exploring a modification of the CROPseq strategy for determining the effects of CRISPR-based knockouts/overexpression constructs, potentially facilitating the high-throughput functional dissection of large and complex biological networks. Each construct is sequenced in parallel, using a modified single cell transcriptome sequencing (scRNAseq) strategy; a strategy similar in concept to Next Generation Sequencing.

As proof of principle, we are using this approach to analyse twenty pluripotency factors in mouse embryonic stem cells (mESCs) at single cell transcriptome resolution, allowing us to determine changes in gene expression following perturbation of each factor and to cluster different factors based on the similarity of their perturbation profile. This pilot screen will also lay the foundations for future experiments. We are planning to delete all 1,500 transcription factors expressed in mESCs in order to organize them into transcriptional networks that orchestrate cell fate decisions. To directly address the structure and parametrisation of our large cancer model by this approach, we plan to randomly generate knockouts and deletions, as well as overexpression constructs in all genes relevant to the model; this will initially be conducted within a set of colon and breast cancer cell lines and organoids. The effect of these alterations on the transcriptome (plus the gene targeted in every cell) will again be analysed by a modification of CROP-seq/scRNAseq.

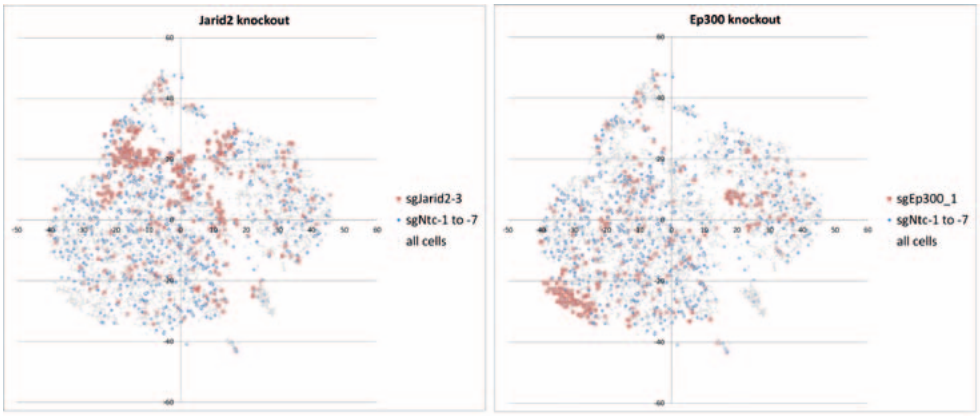


Figure 2.
tSNE analysis for the knockout of two genes (Ep300 und Jarid2) in a mouse ES cell line with one sgRNA each. All cells analysed are indicated in grey, in blue all cells with a non-target control knockout, and in red the cells with knockout of the candidate genes – Ep300 and Jarid2.

Analysis of monozygotic twin pairs using immune repertoire sequencing

[Hans-Jörg Warnatz, Hans Lehrach]

Many diseases and/or their therapies involve the immune system, making the immune system one of the key components of a potential ‚Digital Twin'. To improve our understanding of disease mechanisms in autoimmunity diseases in an isogenic situation, we are analysing monozygotic twin pairs discordant for autoimmune-related diseases, using the deep immune status analysis strategy. We have established collaborative projects for the investigation of Type I Diabetes (TwinsUK project with David Leslie at the Queen Mary University of London) and of Systemic Lupus erythematodes (collaboration with Thomas Dörner, Rheumatology and Clinical Immunology at the Charité University Medicine in Berlin).

Using T cells and B cells isolated from the twins' peripheral blood, we extracted total RNA and sequenced the T cell receptor beta chains and antibody heavy chains using our robust and quantitative DeepImmune sequencing pipeline. The resulting data allows us to cluster twin pairs according to their individual repertoires and to look for differences of affected twins versus their healthy counterparts. Thus, we are looking to identify features within whole repertoires and on the level of individual clonotype signatures that may influence disease states and disease progression.

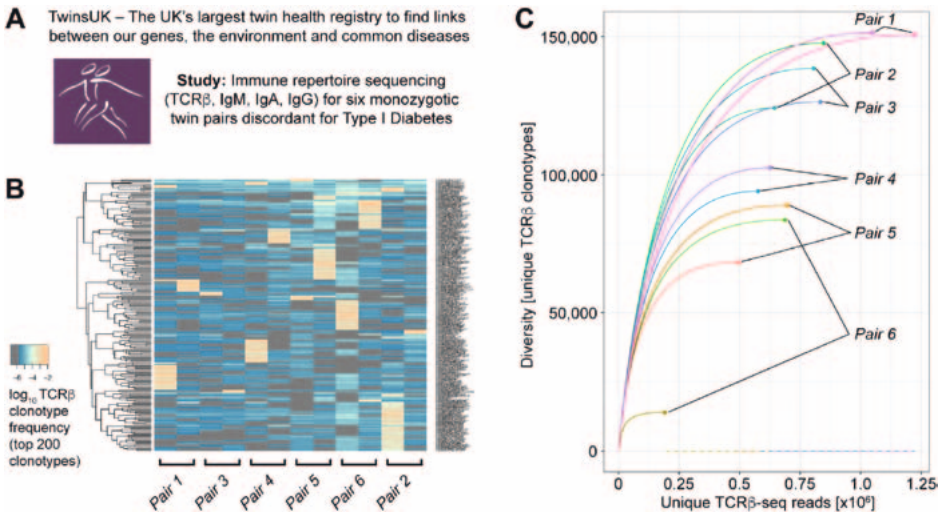


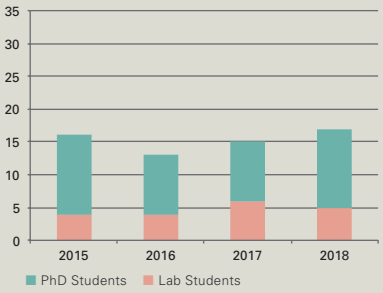
Figure 3.
Exemplary TCRβ sequencing analysis. (A) We have performed T cell and B cell receptor repertoire analysis for six twin pairs from the TwinsUK cohort that are discordant for Type I Diabetes. (B) Heatplot showing the clonotype frequencies of the top 200 most common TCRβ clonotypes among the 12 analyzed samples. These most frequent clonotypes are specific for each individual and do not overlap significantly. (C) Rarefraction plot showing the number of detected TCRβ clonotypes in each sample depending on the number of identified TCRβ-seq reads. This kind of saturation analysis shows that the six different twin pairs vary strongly in their repertoire diversities. Interestingly, the individual twin pairs tend to cluster together in the plot, showing that the twins exhibit similar T cell diversities in their blood. We also observed a tendency for reduced repertoires in the twins affected by Type I Diabetes versus their healthy counterparts.

Reinert Lab – Max Planck Fellow Group Efficient Algorithms for Omics Data

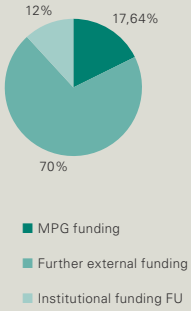
The Max Planck Fellow group has been established in July 2014.

Prof. Dr.-Ing. Knut Reinert
++49 30 838 75 222/218
knut.reinert@fu-berlin.de

Team



Funding



Focus areas

- *Data structures and parallelization* – analysis and storing large –omics data sets
- *Read mapping and variant calling* – finding genomic variations
- *(Pan)-Genome comparison and metagenomics* – qualitative and quantitative analysis of many genomes
- *Genomic RNA analysis* – finding classes of lncRNAs

Research concept

The goal of the Reinert lab is to bridge the existing gap between advanced results in algorithmic research and their practical application as bioinformatics tools for real world data. The group is achieving this by working on solving well posed, novel as well as existing computational problems arising in the context of –omics data analysis. The concept followed is that we always consider the complete analysis cycle, namely the inception of algorithms, their efficient implementation using modern multi-threading and vectorization and their application to relevant problems in collaboration with experimental collaborators at the MPIMG or the FU Berlin.

Scientific highlights

- Devised a comprehensive search method for bidirectional compressed suffix arrays in optimal time [Kianfar et al. RECOMB-Seq 2018]
- Developed currently leading parallelized and vectorised pairwise alignment implementation [Rahn et al. Bioinformatics 2018]

- Developed a state of the art structural variant caller for different kinds of variations [Kim & Reinert WABI 2017 (Lipics)]
- Establishing the software libraries SeqAn and OpenMS as a central German resource [Röst et al. Nat Methods 2016]

Scientific honours & selected invited talks

- Invited speaker at the "Lille Workshop on Stringology", Lille, France, 2017
- Invited speaker at "German-Russian Week of the Young Researcher on Computational Biology and Biomedicine", Skoltech/Moscow, 2017
- Invited speaker at the "12th Workshop on Compression, Text and Algorithms", Palermo, Italy, 2017
- Invited speaker at "MatBio 2017", London, 2017

Results of completed projects during reporting period

Generic accelerated sequence alignment in SeqAn using vectorization and multi-threading

[René Rahn, Marcel Ehrhardt]

Pairwise sequence alignment is a central tool in many bioinformatics analyses. In our group, we implemented a generically accelerated module for pairwise sequence alignment applicable for a broad range of applications. We unified the standard dynamic programming kernel used for pairwise sequence alignment and extended it with a generalized inter-sequence vectorization layout, such that many alignments can be computed simultaneously by exploiting SIMD (Single Instruction Multiple Data) instructions of modern processors. We then extended the module by combining the vector kernel with multithreading on the newest Intel® Xeon® (Skylake) and Intel® Xeon Phi™ (KNL) processors. This resulted in computations for single alignments about 1,600 times faster on the Xeon Phi™ and 1,400 times faster on the Xeon® than executing them with a standard sequential alignment module on a single CPU.

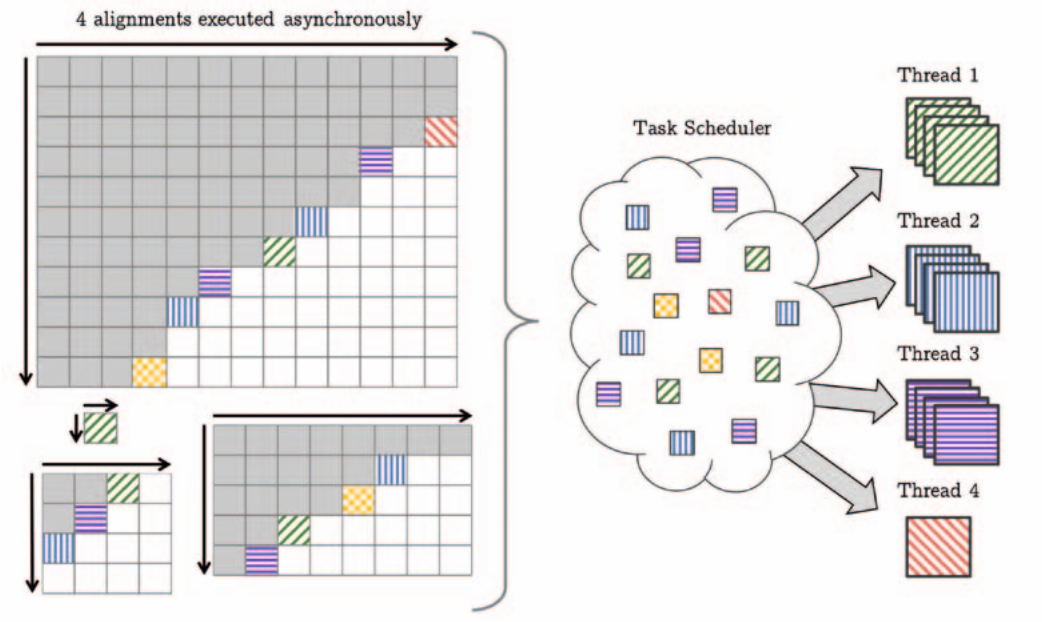


Figure 1.
Generalized wavefront model to compute multiple pairwise sequence alignments split in many sub-alignments. The grey tiles are already computed. The colored (filled) tiles are ready for execution and wait in the task scheduler for the next available thread. If a thread becomes available it tries to pick l = 4 many sub-alignments from the scheduler and computes them vectorized.

Rahn R, Budach S, Costanza P, Ehrhardt M, Hancox J & Reinert K. Generic accelerated sequence alignment in SeqAn using vectorization and multi-threading. Bioinformatics, doi: 10.1093/bioinformatics/bty380 (2018)

Optimum search schemes for approximate string matching using bidirectional FM-index

[Christopher Pockrandt]

Finding approximate occurrences of a pattern in a text using a full-text index is a central problem in bioinformatics and has been extensively researched. Bidirectional indices have opened new possibilities in this regard allowing the search to start from anywhere within the pattern and extend in both directions. In particular, the use of search schemes (partitioning the pattern and searching the pieces in a certain order with given bounds on errors) can yield significant speed-ups. We introduced the first general method to compute optimal search schemes. Our experiments show that the optimal search schemes found by us significantly improve the performance of search in bidirectional FM-index upon previous ad-hoc solutions. For example, approximate matching of 101 bp Illumina reads (with two errors) becomes 35 times faster than standard backtracking.

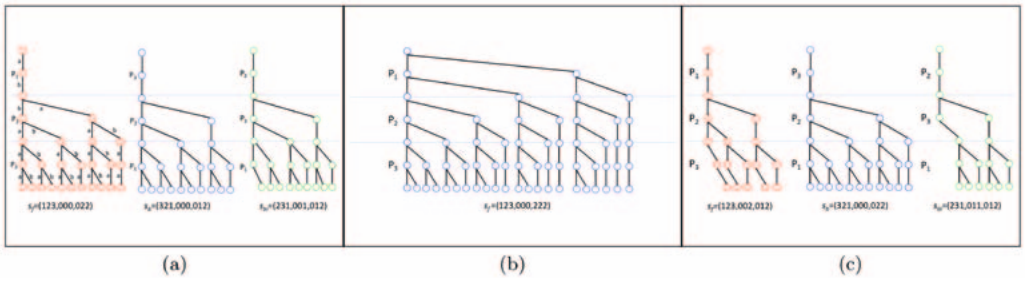


Figure 2.
(a) The search of Lam et al. as described by Kucherov for K = 2 and P = 3, i.e., S_{Lam} = {s_f = (123, 000, 022), s_b = (321, 000, 012), s_{bi} = (231, 001, 012)}, shown for the read "abbaaa" from the alphabet {a,b}, i.e., R = 6 and σ = 2. The read is partitioned into P_1 =ab, P_2 =ba, and P_3 =aa. Partition borders are shown by horizontal lines. A vertical and a diagonal edge represent a match and a mismatch, respectively. Edge labels are only shown for s_f for a cleaner picture. The search corresponding to each trie is designated underneath it by its (π,L,U). The number of edges in S_{Lam} tries is 71. (b) The unidirectional search scheme S_{Uni} = {s_f = (123, 000, 222)} for the same problem. The number of edges in S_{Uni} is 62, (c) The optimal search scheme S_{Opt} = {s_f = (123, 002, 012), s_b = (321, 000, 022), s_{bi} = (231, 011, 012)} for the same problem, found by our method.

Reinert K, Pockrandt C, Kianfar K, Torkamandi B & Luo H. Optimum search schemes for approximate string matching using bidirectional FM-index. Presented at RECOMB-Seq: 1–13 (2018)

Ongoing and future projects

The DREAM framework

[Temesgen Dadi, Svenja Mehringer, Enrico Seiler]

Modern sequencing technologies have brought a super-exponential growth of sequencing capacities. This has enabled the cheap sequencing of the genomic content of pangenomes, metagenomes, or many individuals of the same species (e.g. the 100,000 Genomes Project) that differ only slightly from each other. Yet, the small individual differences are of interest (i.e. SNPs, or small structural polymorphisms) to elucidate the cause of diseases or reconstruct evolutionary events.

These datasets expose interesting characteristics. They are large, while some large fractions are highly redundant (e.g. 100,000 Genomes Project, or storing different strains of bacteria) and hence amenable to compression techniques. On the other hand, compression usually makes it costly to implement the main operations on the data, namely finding approximate matches of (many) queries (approximate in the sense of edit distance). In this project, we hence focus on the development of parallelised and vectorised algorithms for the analysis of highly redundant (i.e. genomic) data. In the work of Dadi we introduced a distributed read mapper in the context of a larger framework, that we call the DREAM (Dynamic seaRchablE pArallel coMpressed) framework. While a lot of research has focused on indexing genomic datasets, the resulting solutions lack, in general, the ability to easily change the underlying datasets. That means it is costly to add or delete complete sequences while maintaining the ability to support fast approximate string searches. For example, in metagenomics, this problem becomes more and more recurrent. Many metagenomics search tools and read mappers use compressed suffix arrays (CSA) which have to index about 50 to 200 gigabases. Due to constant database updates, changes

occur on a daily or weekly basis and thus require a newly constructed index. Recomputing a single index of this size is quite costly in terms of space and time. For example, it takes about one day to compute the index for the dataset used in our experiments. On the other hand, the ability for fast approximate searches in such an index is crucial. In the proposed DREAM framework we will offer various solutions for the above areas depending on some key parameters of the input set (size of the input, amount of redundancy, the importance of rebuilding time vs. search time). The work of several PhDs will go into further developing this framework to allow analyses of terabytes of sequencing databases.

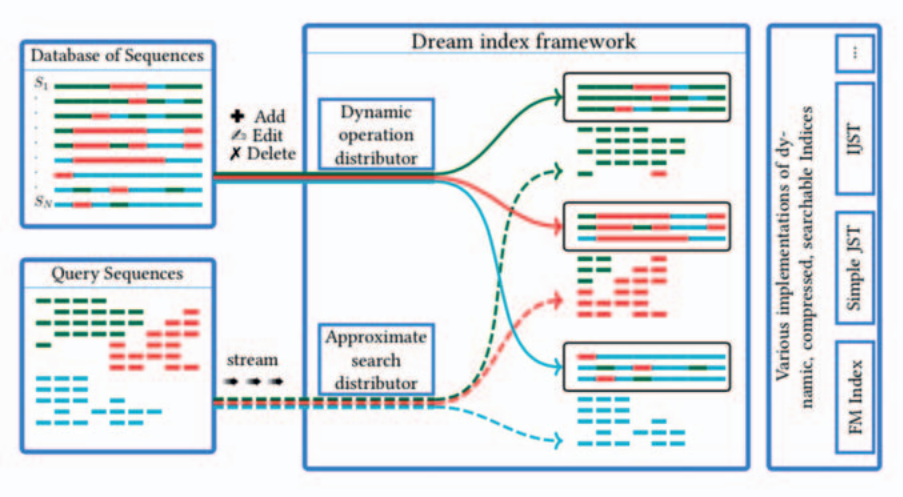
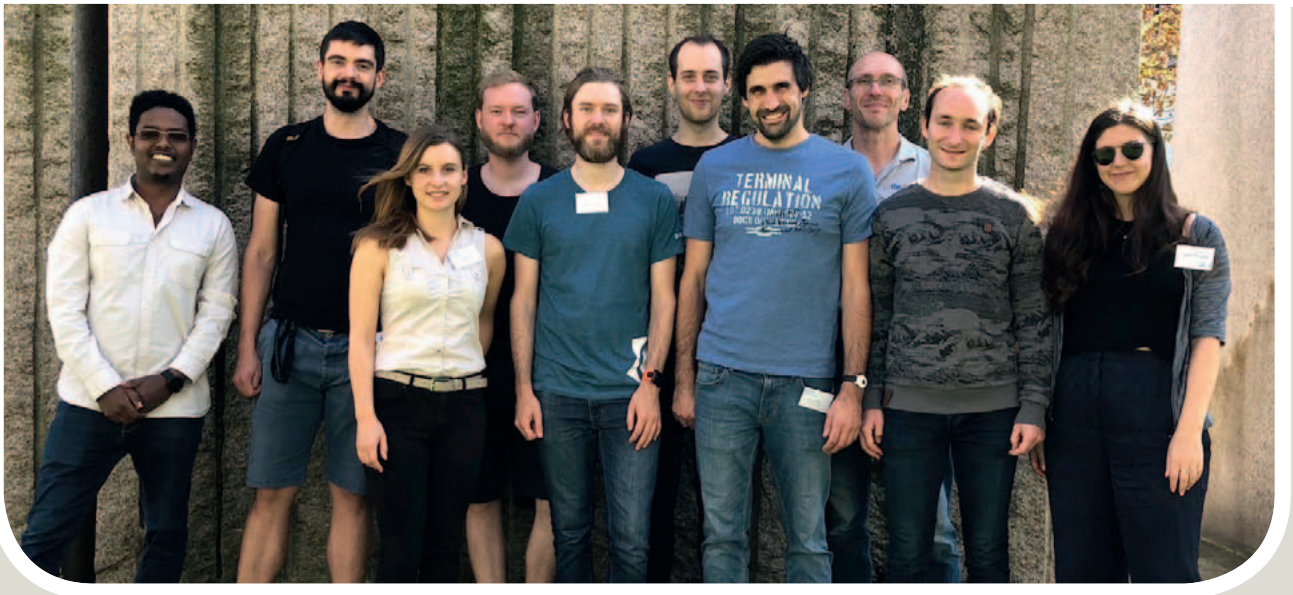


Figure 2.
Sketch of the DREAM index framework. The red sequence piece among the green ones symbolizes that we do not require a perfect partitioning allowing us to use fast methods. The boxes on the right symbolize the potential use of different index implementations.

Motif-based algorithms for detection of functional classes of long non-coding RNAs

[Jörg Winkler, in collaboration with Annalisa Marsico, MPIMG]

In this joint project we want to devise computational methods for aligning (pseudoknotted) RNA families. Then, we derive probabilistic global and local motifs from the alignments, while allowing incorporating experimental evidence and searching the motifs fast and sensitive in large genomic sequences. Subsequently we want to extend and apply this methodology to retrieve sequence structure motifs in human lncRNAs, which will enable the classification of human lncRNAs into different classes based on sequence structure motifs and other genomic features. In a final step we will find and functionally annotate new classes of non-coding RNAs.

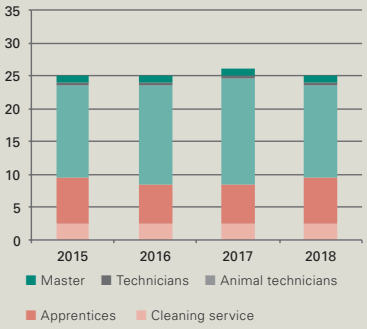


Animal Facility

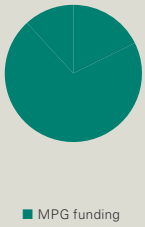
Dr. Ludger Hartmann
hartmann@molgen.mpg.de
++49 30 8413-1189

The animal facility has been established in 2003.

Team



Funding



Service offered by the group

- Animal husbandry for mice and zebrafish
- Computerized colony management
- Assistance with animal experiments (e.g. embryotransfers)
- Health monitoring
- Maintaining of the SPF – hygienic level
- Providing animals, e.g., for the transgenic unit, national and international mouse shipments, quarantine, rederivation

Service concept, methods, and projects

Full service is currently our standard. Operation principle is that all technical work is done by animal facility staff. Mouse users can have access to mouse rooms only with a valid scientific justification and a comprehensive introduction.

Animal welfare highlights

In order to improve animal welfare, two new techniques have been established recently:

- Before embryo-collection of early stages in mice, we perform ultrasonic testing to check, whether the mice are really pregnant.
- Instead of health monitoring of our mouse colonies by using sentinel animals (which therefore have to be sacrificed), we established a sentinel-free system by using Exhaust-Air-Dust Probes, which are tested by PRIA (= PCR Rodent Infectious Agent).

Animal welfare services

The three animal welfare officers of the Institute look at all animal welfare issues in compliance with the German and European animal welfare guidelines and regulations. They consult and assist scientists in terms of methodological refinement and experimental design, severity classification of procedures, ethical considerations and counseling when writing application documents. Also, they control the execution of the animal experiments and are available not only for scientists, but also for all animal technicians in case of any problems.

Above that, the Animal Welfare Committee of the Institute meets several times per year to discuss internal animal care and use. This committee consists of veterinarians (animal welfare officers), animal technicians and two scientists of each department working with mice.

Animal welfare considerations

In 2016, the Max Planck Society published a White Paper – Animal Research in the Max Planck Society, summarizing the results of an intensive discussion about animal research and the responsibility of each individual working with laboratory animals. The White Paper has been intensively discussed at the MPIMG and confirmed us in our ongoing engagement to continuously improve the well-being of our animals in terms of the 3R (= replace-reduce-refine animal experiments). In addition, the Max Planck Society introduced a fourth R for responsibility and committed itself to measures such as improving social environment of experimental animals or engaging proactively in professionalizing the public discourse on animal ethics.

Knowing this, all staff members and other personnel are committed to maintain not only the best hygienic quality standards, but really "to care" for the animals they are responsible for. Everybody dealing with animals has to ensure that all animal care and used procedures are conducted within the highest scientific and ethical principles. Regular advanced training of animal procedures is obligatory for the animal facility staff and scientists working with mice. Basic courses and training, e.g. mouse handling, on-site-training by animal technicians, and FELASA courses are required before starting to work with mice.

To improve the public discussion on animal experiments, we communicate openly and transparently about our work. For years, we publish information about animal experiments on the website of the Institute. In 2018, in a joint effort with the press office of Max Planck Society, the MPIMG realized a short film about our laboratory mouse work which was shot in our facility and is now available on youtube ("Gene, die in die Knochen fahren") as well as on our own website.

Material resources, equipment and spatial arrangements

The mouse facility comprises more than 300 genetically modified and 30 wild type mouse strains kept under specified pathogen free (SPF) conditions in a clean area with ten animal rooms and restricted access. All strains are housed in individually ventilated cages (overall about 6,500 cages). All cages are handled under sterile conditions.

The zebrafish facility (about 150 single tanks) is located in the basement of the animal house. It is used for breeding and maintenance of zebrafish lines, as well as for providing eggs, embryos and larvae to the researchers of the Institute. The facility offers a set-up for zebrafish egg-microinjection.

Current and future developments

The Institute will go on to increase the cage capacity by installing a more compact cage system. Also a pilot rack with "digital ventilated cages" is preferable. This system enables to collect several data out of the cage without the necessity to open the cage (e.g. to collect behavioral information of mice for phenotype characterization).

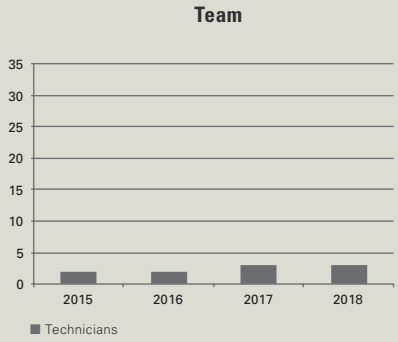
The PyRAT mouse software can be extended by a transgen module ("PyRAT transgenic"). Main features of this module are e.g. a cryopreservation database, embryotransfer reports including offspring, ordering of donors, plug-check. A seamless integration with PyRAT is possible and this should make the teamwork between Animal Facility and Transgenic Unit easier.

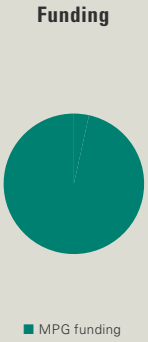
The facility will keep a focus on extension and continuous improvement of animal well-being activities (e.g. further enrichment, where effective).

Transgenic Unit

The group has been established in October 2010.

Dr. Lars Wittler
wittler@molgen.mpg.de
++49 30 8413-1453





Service offered by the group

For the generation of genetically modified mouse models, we employ embryonic stem (ES)-cell based technologies, including tetraploid complementation and ES-cell derivation, as well as zygote manipulation by electroporation and microinjection. Furthermore we provide assisted reproductive technologies (ARTs) as *In-Vitro*-Fertilization (IVF) and cryopreservation of sperm and embryos to retain and distribute mouse strains. We give legal and administrative support for animal experimentation and animal welfare, practical training and assistance in the generation of genetically modified ES cell lines and general education of students and scientists from the Institute in transgenic technologies.

Service concept, methods and projects

The transgenic unit was established to provide a centralized resource and standardized platform to utilize technologies for the generation of genetically modified mouse models for the Institute. To achieve this, we built up a robust routine employing the morula aggregation technique that can be used for both, the conventional generation of chimeras and for tetraploid complementation experiments, a method for generating embryos that are almost exclusively made up by ES cells.

Morula aggregation facilitates to efficiently scale-up the generation of genetically modified mice even with a comparatively small team (four staff members plus support by animal technicians), enabling us to conduct even projects that require a mid- to high-throughput production of mouse models or embryonic material. We process up to 13 different ES cell lines per week, performing up to 375 experiments per year, about two third of them as tetraploid complementation assays. For the majority of these experi-

ments, embryos are taken from the foster mother to directly analyze the resulting phenotype of a genetic alteration during embryonic development. Additionally 213 novel genetically modified mouse lines were established between 2015 and 2018.

This high output demands also a significant administrative effort in the organization and supervision of all steps from the establishment to the phenotype assessment of the generated mouse lines. To cope with the increased legal requirements for animal experimentation, in particular after the revision of the animal welfare act in 2013, we streamlined and optimized the administration of the procedures. At present, we handle approx. 100 independent projects from 36 scientists of the Institute, mastering highest standards not only methodically but also according to legal and animal welfare obligations.

Material resources, equipment and spatial arrangements

The transgenic technology lab provides two microinjection workstations, one of them equipped with piezo drill and laser ablation objective, several microdissection microscopes, including side by side teaching stereomicroscopes and a zygote electroporation device. For cell and embryo culture we are equipped with a small cell culture lab and a variety of incubators that allow also unconventional gassing. Two ramp cooling devices facilitate the cryopreservation of mouse embryos and vitrification of oocytes. Embryo transfers and other surgical interventions are performed on microsurgery workstations that are equipped with inhalation anesthesia units. We are using approx. 600 cages animal space.

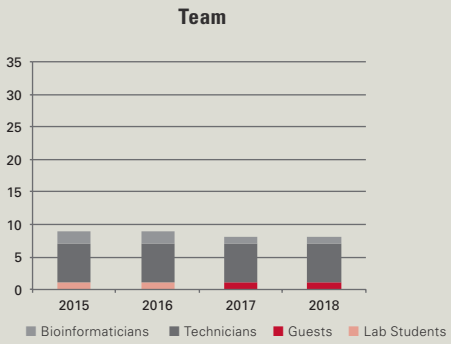
Current and future developments

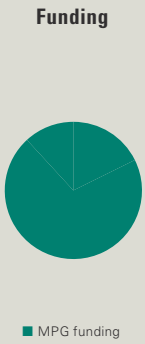
Although the main emphasis of the transgenic lab relies on ES cell based technologies, with the rise of genome editing tools like the CRISPR/Cas9 technology, we also strengthened our capabilities to perform state-of-the-art direct manipulation of zygotes. We acquired a microinjection workstation in 2014 that was supplemented with an ablation laser device in 2017. To expand these methods, we recently obtained an electroporation device that is specifically optimized for zygote electroporation. Currently we are working out the various possibilities for CRISPR/Cas9 based applications by direct manipulation of zygotes through microinjection and electroporation in combination with assisted reproductive technologies. Our goal is to establish similar efficient and productive pipelines for these approaches as we have for the ES cell based genetically modified mouse model generation.

Sequencing Core Facility

Dr. Bernd Timmermann
timmerma@molgen.mpg.de
++49 30 8413-1121

The group has been established in December 2007. Since June 2010, it has been acting as Max Planck Sequencing Core Facility open to all institutes of the BMS.





Service offered by the group

- *Human genome research* – genetic variation in context of common diseases
- *Model organism* – construction and analysis of new genomes of functional relevant organisms

Material resources, equipment and spatial arrangements

The Sequencing Core Facility operates several next generation sequencers and maintains a fully equipped lab and staff able to perform a variety of sequencing applications – from sample preparation to data analysis.

Currently we are providing expertise for two different technologies: The single molecule, real-time (SMRT) sequencing technology from Pacific Biosciences and the sequencing-by-synthesis technology from Illumina. At a read length of up to 40,000 bases the PacBio technology offers a great benefit especially for *de novo* genome sequencing, analysis of structural variations, metagenome analysis and full length transcriptome analysis. The high throughput of our Illumina systems completes our sequencing service and offers a real advantage for many applications. Single cell analysis, expression profiling (RNA-Seq), methylation analysis (MeDIP-Seq, Bisulphite-Seq, RRBSseq), copy number analysis as well as the identification of protein binding sides (ChIP-Seq) and the analysis of whole genomes benefit from the high output of this system to a great extent.

For a detailed list of services and equipment, please have a look at our website https://www.molgen.mpg.de/SeqCoreTechnologies.

Figure 1.
Illumina NovaSeq 6000 system. High-throughput system with an output of up to 6,000 gigabases and 20 billion reads per run in two days.

Scientific highlights

- Protein interaction perturbation profiling at amino-acid resolution [Woodsmith et al. Nature Methods 2017]
- Wild tobacco genomes reveal the evolution of nicotine biosynthesis [Xu et al. PNAS 2017]
- The metabolic background is a global player in *Saccharomyces* gene expression epistasis [Alam et al. Nature Microbiology 2016]
- Completion of the 1000 Genomes Project. Reconstruction of 2,504 individual genomes from 26 populations [1000 Genomes Project Consortium, Nature 2015]

Results of completed projects during reporting period

Wild tobacco genomes reveal the evolution of nicotine biosynthesis

[Heiner Kuhl, Bernd Timmermann in collaboration with Ian T. Baldwin, MPI for Chemical Ecology]

We sequenced and assembled genomes of two wild tobaccos, *Nicotiana attenuata* (2.5 Gb) and *Nicotiana obtusifolia* (1.5 Gb), two ecological models for investigating adaptive traits in nature. We show that after the *Solanaceae* whole-genome triplication event, a repertoire of rapidly expanding transposable elements (TEs) bloated these *Nicotiana* genomes, promoted expression divergences among duplicated genes, and contributed to the evolution of herbivory-induced signaling and defenses, including nicotine biosynthesis. These results provide evidence that TEs and gene duplications facilitated the emergence of a key metabolic innovation relevant to plant fitness.

Xu S, Brockmöller T, Navarro-Quezada A, Kuhl H, ... [9 authors]... Timmermann B, Gaquerel E & Baldwin IT. Wild tobacco genomes reveal the evolution of nicotine biosynthesis. Proc Natl Acad Sci USA 114(23):6133-6138 (2017)

The metabolic background is a global player in *Saccharomyces* gene expression epistasis

[Stefan Börno, Bernd Timmermann in collaboration with Markus Ralser, University of Cambridge, UK]

Here, we address the importance of the metabolic–genetic background by monitoring transcriptome, proteome and metabolome in a repertoire of 16 *Saccharomyces cerevisiae* laboratory backgrounds, combinatorially perturbed in histidine, leucine, methionine and uracil biosynthesis. The metabolic background affected up to 85% of the coding genome. Suggesting widespread confounding, these transcriptional changes show, on average, 83% overlap between unrelated auxotrophs and 35% with previously published transcriptomes generated for non-metabolic gene knockouts. Background-dependent gene expression correlated with metabolic flux and acted, predominantly through masking or suppression, on 88% of transcriptional interactions epistatically. As a consequence, the deletion of the same metabolic gene in a different background could provoke an entirely different transcriptional response. Propagating to the proteome and scaling up at the metabolome, metabolic background dependencies reveal the prevalence of metabolism-dependent epistasis at all regulatory levels.

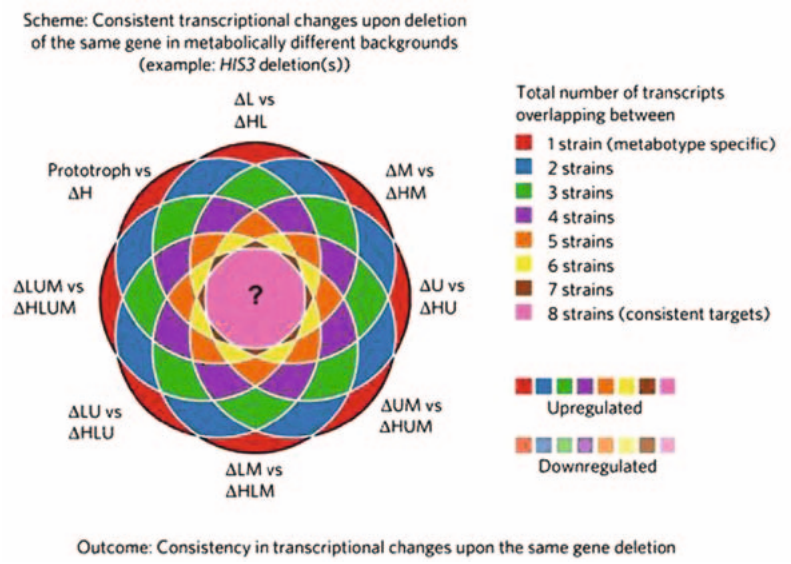


Figure 2. Transcriptional response to a metabolic gene deletion is sensitive to the metabolic-genetic background.Metabolism affects gene function.

Alam MT ...[10 authors]...Börno S, Christen S, Patil KR, Timmermann B, Lilley KS & Ralser M. The metabolic background is a global player in *Saccharomyces* gene expression epistasis. Nat Microbiol 1:15030 (2016)

A global reference for human genetic variation

[Bernd Timmermann as part of the 1,000 Genomes Project Consortium]

The 1,000 Genomes Project set out to provide a comprehensive description of common human genetic variation by applying whole-genome sequencing to a diverse set of individuals from multiple populations. At the end of 2015 the 1,000 Genomes Consortium reported the reconstruction of 2,504 individual genomes from 26 populations using a combination of low-coverage whole-genome sequencing, deep exome sequencing, and dense microarray genotyping. We characterized a broad spectrum of genetic variation, in total over 88 million variants. This resource includes >99% of SNP variants with a frequency of >1% for a variety of ancestries. We describe the distribution of genetic variation across the global sample, and discuss the implications for common disease studies.

The 1,000 Genomes Project Consortium*. A global reference for human genetic variation.Nature 526(7571):68-74 (2015)
*Contributor from the group: Timmermann B

Ongoing and future projects

Sequencing and analysis of the alpine marmot genome

[Stefan Börno, Sven Klages, Bernd Timmermann in collaboration with Markus Ralser, University of Cambridge, UK, and Toni Gossmann, University of Sheffield, UK]

The alpine marmot is a rodent remnant of the 'ice-age' climate of the Pleistocene steppe, and since the disappearance of this habitat, persists in the high altitude alpine meadow. Sequencing its genome, we reveal that the long-term cold-climate adaptation has altered its metabolism, in particular biosynthesis and storage of fatty acids. Paradoxically, despite successful adaptation and a large population size, we detected levels of genetic variation that are among the lowest for mammals, and that purifying selection is ineffective. Reconstructing its demographic past revealed that the two apparently contradictory observations are, in fact, intimately interlinked. The climate-adaptive life history prevented genetic diversity to recover from Pleistocene perturbations. The case of the alpine marmot demonstrates that populations of extremely low genetic diversity can be very successful and persist over thousands of years, but that a climate-adapted life history can deprive a species from the possibility to recover genetic diversity.

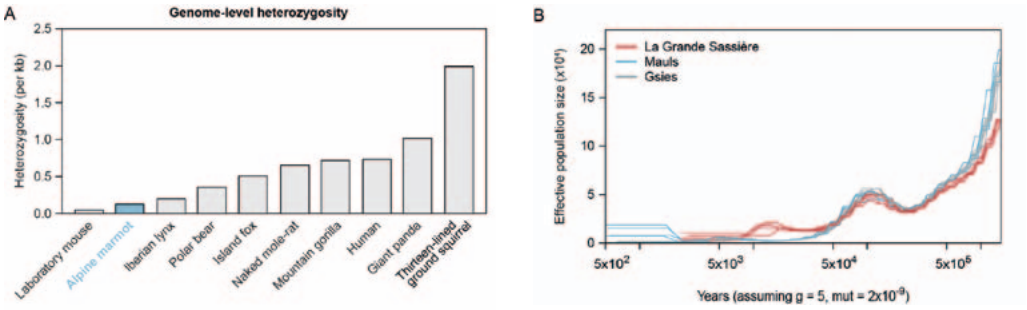
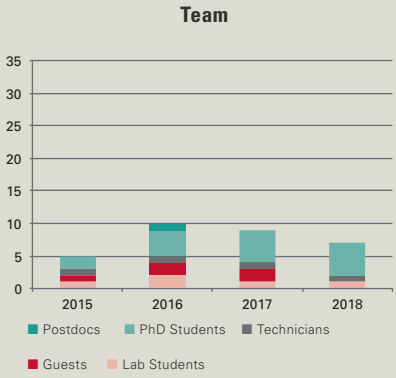


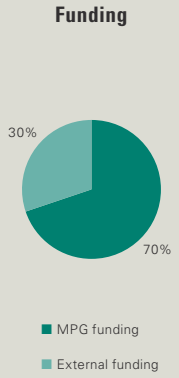
Figure 3.
Genomic diversity of the alpine marmot.
A. The alpine marmot genome is characterized by low heterozygosity at the genome level.
B. Pairwise sequential Markovian coalescent (PSMC) analysis reveals details of the genetic past of the alpine marmot. Evident is a decline in the LGS population after the last glacial maximum.

Mass Spectrometry Facility

Dr. David Meierhofer
meierhof@molgen.mpg.de
++49 30 8413-1567

The Mass Spectrometry Facility has been established in March 2012.





Service offered by the group

The Mass Spectrometry Facility of the MPIMG provides analytical services to the researchers in-house, and assists in the development and analysis of mass spectrometry experiments. The service includes

- Proteome profiling, analysis of IP's, single proteins, PTM's such as methylation, phosphorylation, ubiquitination, acetylation …, and label- or label-free quantification
- Metabolome profiling, analysis of selected metabolites ….
- The Q Exactive Plus Orbitrap mass spectrometer was upgraded to the newest available model HF at the beginning of 2016. This allowed doubling the sequencing speed. Currently, we are setting up and establishing methods to analyze lipids. Lipidomics is thus complementing our current methods in metabolome- and proteome profiling.

Material resources, equipment and spatial arrangements

We are equipped with:

- *QTrap 6500 mass spectrometer*
  The triple quadrupole hybrid ion trap mass spectrometer QTrap 6500 (Sciex) is a low-resolution instrument and can be coupled online to either an LC instrument (1290 series UHPLC; Agilent) or a nanoLC Dionex Ultimate 3000 (Thermo Scientific) device.
- *Q Exactive HF mass spectrometer*
  The Q-Exactive HF Orbitrap mass spectrometer is a high-resolution instrument and can be coupled online to the nanoLC Dionex Ultimate 3000 (both, Thermo Scientific).

Focus area of own research

- *Mitochondrial pathologies* – disease mechanisms
- *Renal cell tumors* – metabolic regulations and dynamics

Research concept

Besides providing analytical mass spectrometry services for researchers of the MPIMG, our facility has its own research group and scientific interests. By applying proteome- and metabolome profiling in renal carcinomas and rare mitochondrial diseases, we try to elucidate metabolic regulations in these diseases. Of special interest is thereby the integration of multiple "omics" data sets to gain a more holistic view of metabolic dysfunctions in diseases.

Scientific highlights

- Establishment of a mass spectrometry-based method for ROS detection [Xiao et al. Antioxidants & Redox Signaling 2018]
- The ROS scavenger glutathione was characterized as a new hallmark in benign renal oncocytomas [Kuerschner et al. Oncotarget 2017]
- An impaired respiratory electron chain triggered down-regulation of the energy metabolism and de-ubiquitination of solute carrier amino acid transporters [Aretz et al. Molecular Cellular Proteomics 2016]

Scientific honors & selected invited talks

- David Meierhofer – Invited speaker at Metabolomic Network Initiative, Munich, 2018
- Yang Ni – Poster award, 12th European Summer School Advanced Proteomics, Brixen, Italy, 2018
- Ayham Alahmad – Invited speaker at the 9. Symposium Urologische Forschung der Deutschen Gesellschaft für Urologie, Freiburg, 2017
- Ayham Alahmad – JKD-Forschungspreis für die Experimentelle Nierenkarzinomforschung (award for experimental research on renal carcinoma), 2017

Results of completed projects during reporting period

Proteomic and metabolomic characterization of cells lacking mitochondrial DNA

[Ina Aretz (former Gielisch)]

ρ0 cells, which entirely lacking mitochondrial DNA, were characterized by metabolome and proteome profiles to elucidate molecular insights of the oxidative phosphorylation. A significant down-regulation of the pyruvate metabolism, the TCA cycle and the respiratory electron transport was observed in ρ0 cells. Pathway and enrichment analysis of regulated proteins revealed the retrograde response, a pathway of communication from mitochondria to the nucleus. The majority of ubiquitination sites in ρ0 cells were decreased, which could indicate an increased activity or re-localization of amino acid transporter and regulation of cytosolic ribosomal proteins.

Aretz I, Hardt C, Wittig I & Meierhofer D. An impaired respiratory electron chain triggers down-regulation of the energy metabolism and de-ubiquitination of solute carrier amino acid transporters. Mol Cell Proteomics 15(5):1526-38 (2016)

LC-MS-based detection of reactive oxygen species (ROS) with hydroethidine- probes in cells

[Yi Xiao]

Detection and quantification of the highly reactive and short-lived superoxide (O^{2-}) can be challenging. A new mass spectrometry (MS)-based method was developed to detect and quantify O^{2-} using three fluorogenic hydroethidine probes: HE, mito-HE, and HPr^{+}, which measure cytosolic, mitochondrial and extracellular O^{2-}, respectively (figure 1). Our method overcomes limitations from spectral overlap of O^{2-} specific and non-specific products in fluorescence microscopy. However, our experiments showed that these hydroethidine probes can be prone to autoxidation during incubation at 37 °C (figure 1). Thus, subtle changes in ROS levels in cell culture experiments might not be quantifiable. Our findings raise the question of whether hydroethidine-based probes can be used for the reliable detection of O^{2-} radicals in cell culture.

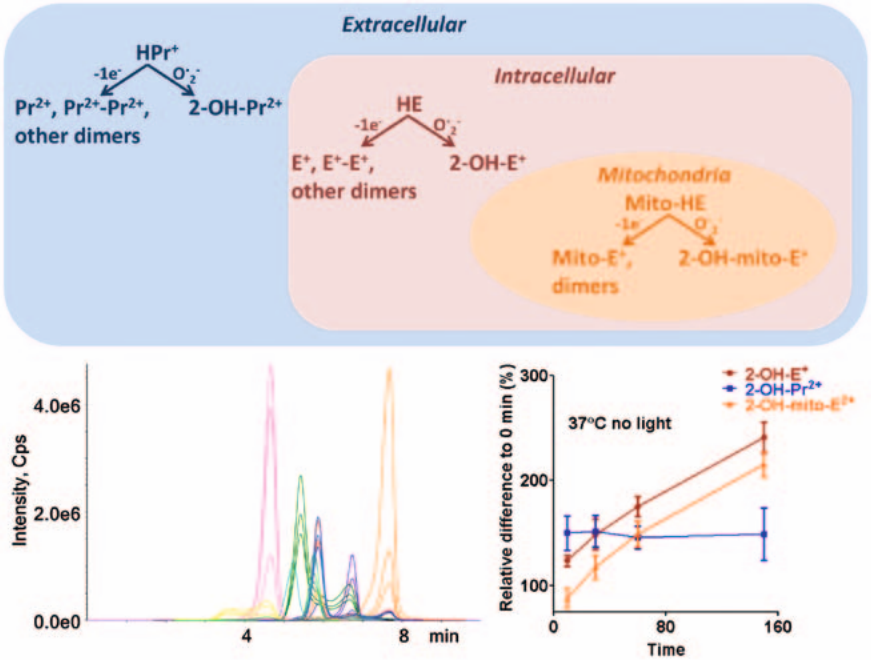


Figure 1.
Three hydroethidine-based probes for O^{2-} detection (top); All probes and their oxidative products were monitored in single LC-MS/MS method (bottom left); O^{2-} specific products increased over time at 37 °C (bottom right).

Xiao Y & Meierhofer D. Are hydroethidine-based probes reliable for ROS detection? Antioxid Redox Signal. 2018, doi: 10.1089/ars.2018.7535. [Epub ahead of print]

Ongoing and future projects

Mutation in NDUFS1 cause global metabolic and proteomic changes and result in massive ROS production in human cells

[Yang Ni]

Mutations in complex I (CI), the first enzyme of the mitochondrial electron transport chain, can cause inborn metabolic errors. The aim of this work is to elucidate the molecular consequences of a pathological mutation in the two core subunits *NDUFS1* and *ND5* of CI. Proteome profiling indicated that the subunits of the N-module were largely absent, resulting in the disassembly of the entire mitochondrial respirasome. Hence, enzyme activity assays revealed an isolated CI deficiency. Furthermore, the interruption of the electron flow leads to electron leakage and in turn to significantly elevated (25-fold) levels of the ROS scavenger oxidized glutathione (figure 2). Consequently, NADH can no longer be oxidized by CI, resulting in an inhibitory feedback on the TCA cycle, thus, malic- and fumaric acid were found to be significantly increased.

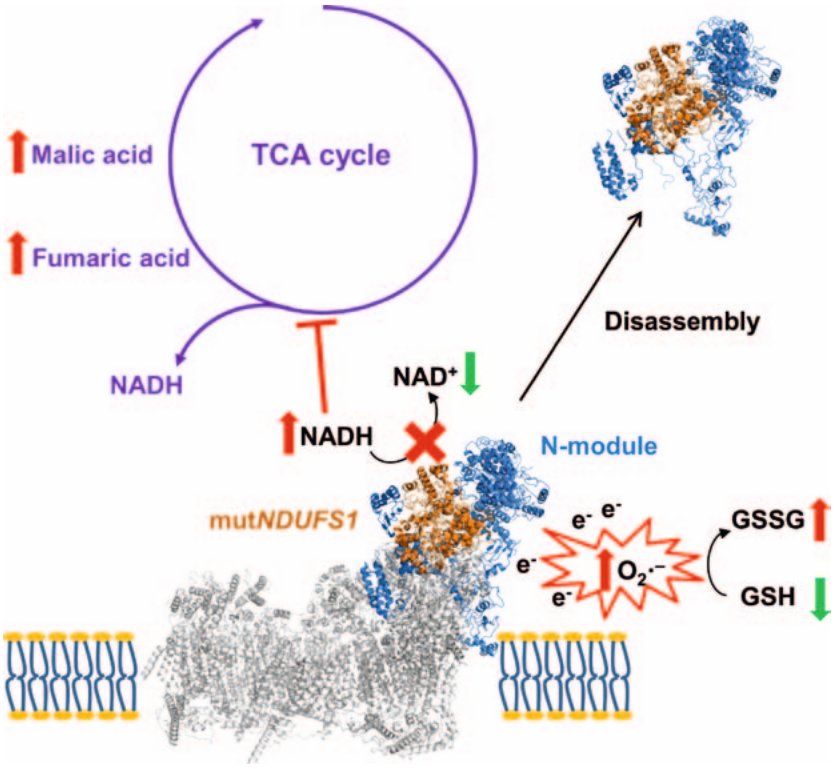


Figure 2.
Schematic representation of the proposed pathomechanism as consequence of a mutation in *NDUFS1* and *ND5*.

Food safety and resilience of food supply chains in crisis situations

[Natalie Witt]

Animal diseases and zoonotic pathogens that can be transmitted via food to humans have a high potential for large-scale emergencies. This work aims to develop a mass spectrometry method to rapidly detect highly pathogenic zoonotic bacteria in food supply chains, which allow simultaneous screening of different bacterial species. Sample preparation methods were established, compared and optimized for three relevant and highly pathogenic bacteria, namely *Brucella* spp., *Francisella tularensis*, and *Bacillus anthracis*. Combining mass spectrometric results with the corresponding next-generation sequencing data, we are able to identify unique species-specific marker peptides, which can be used to establish targeted rapid screening methods for potentially contaminated food samples (figure 3).

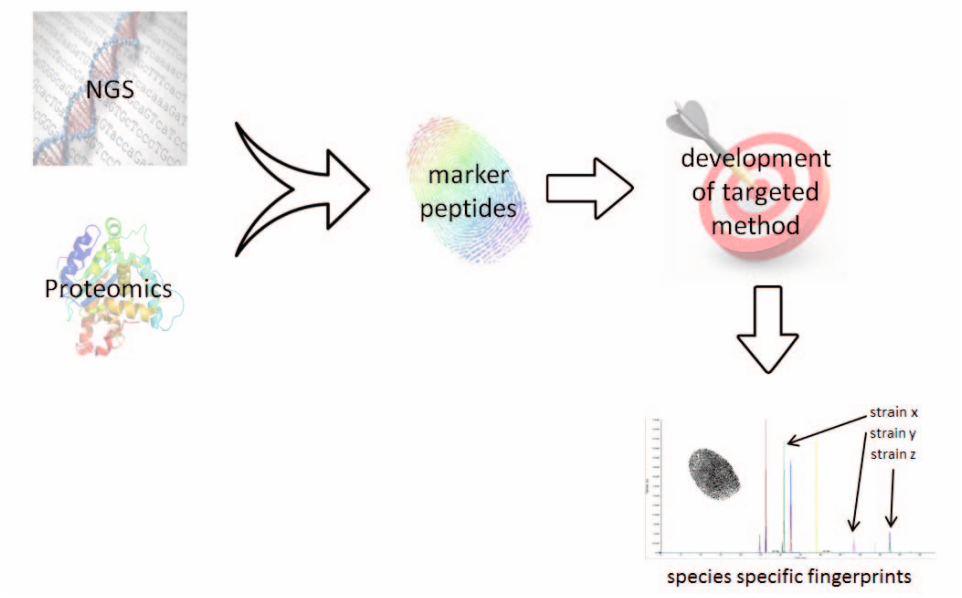


Figure 3.
Workflow for developing a targeted screening method for highly pathogenic bacteria.

Proteomic characterization of papillary renal cell carcinomas

[Ayham Alahmad]

Papillary renal cell carcinomas (pRCC) represent about 15% of renal cell carcinomas. The aim of this work is to understand the regulation on the protein level and underlying molecular mechanisms and alterations of metabolic pathways in this tumor. Our proteome survey identified significant down-regulation of proteins involved in the oxidative phosphorylation (OXPHOS) and significantly up-regulated proteins involved in the spliceosome and ribosome. Our proteome data were in contrast to transcriptome data, which showed no regulation of the OXPHOS system and the spliceosome (figure 4). Interestingly, we found a down-regulation of V-ATPase, which are commonly increased in cancers to excrete lactate.

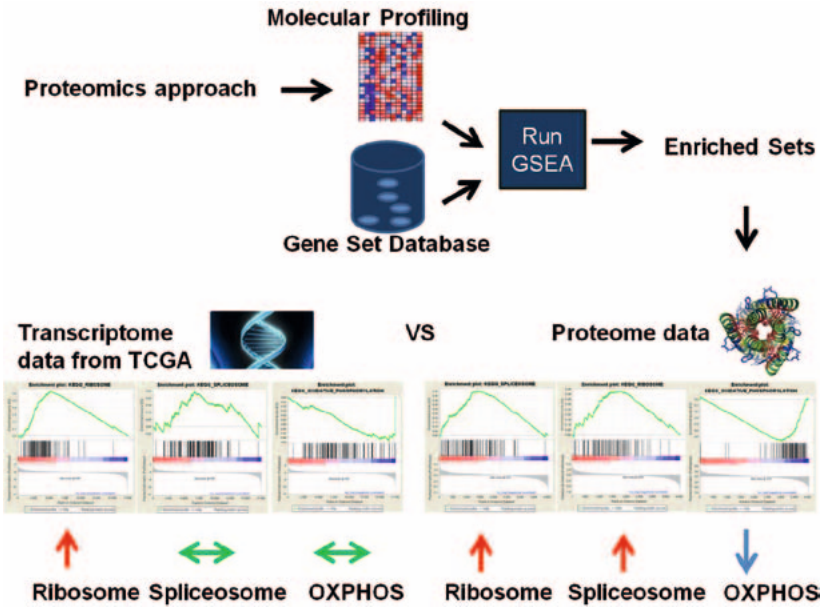


Figure 4.
Comparison of differentially regulated pathways between the transcriptome and proteome in pRCC.

Development of a mass spectrometry-based method to assess the metabolic functionality of the liver

[Robert Gajowski]

We contribute to improving the global liver function assessment by developing an LC-MS-based method, which measures the metabolic turnover of isotopically labeled methacetine to paracetamol in the liver (figure 5). Therefore, plasma samples are collected at different time points after ^{13}C-methacetine injection. Resulting time series show the speed of conversion of ^{13}C-methacetine into paracetamol, which is the main parameter used to investigate the functionality of the liver.

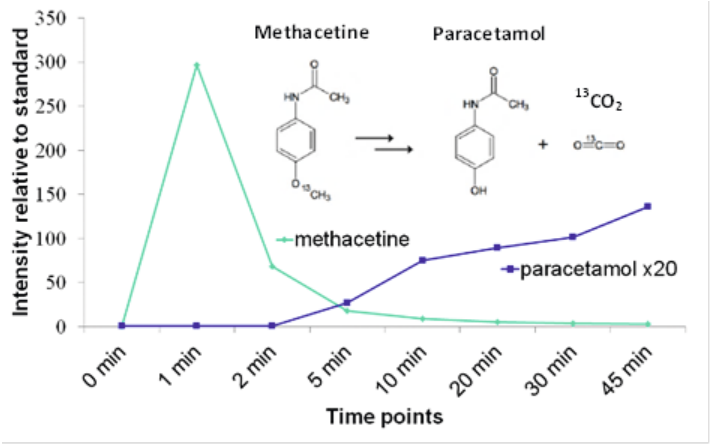
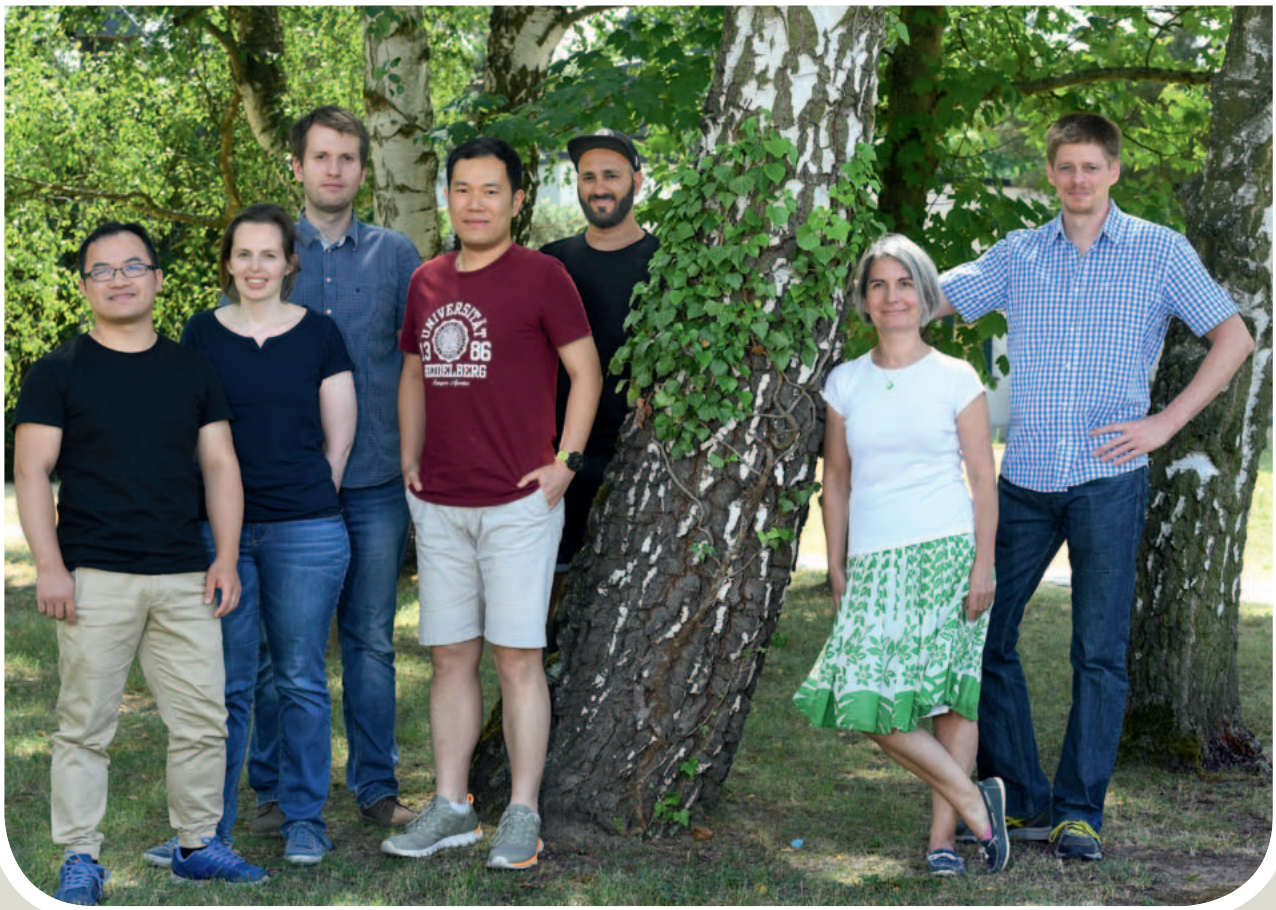


Figure 5.
The conversion of methacetine to paracetamol indicates the metabolic functionality of the liver.

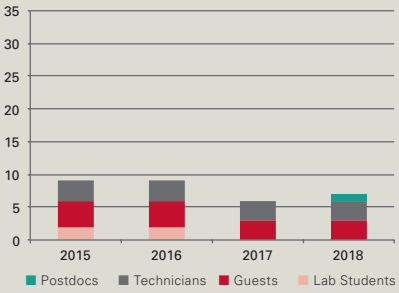

UNIVERSITÄT
13 86
HAMBURG

Service Group Microscopy and Cryo-Electron Microscopy

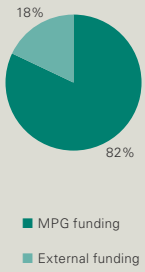
Dr. Thorsten Mielke
mielke@molgen.mpg.de
++49 30 8413-1644

The microscopy part of the group has been established in 1978; the cryo-electron microscopy started in January 2004. In January 2013, both groups were fused to one scientific service group.

Team



Funding



Service offered by the group

The Microscopy and Cryo-Electron Microscopy group provides a broad range of imaging techniques combining both, light microscopy and transmission electron microscopy (TEM). From July 2014 until July 2018, we also supported flow-cytometry. Our service includes:

- Technical support and maintenance of all light-microscopes
- User training on all light microscopes operated as shared equipment
- Application support for advanced light microscopy (e.g. automated imaging, 3D-imaging, screening)
- Implementation of new applications, software and hardware tools
- Conventional TEM sample preparation (plastic embedding, ultrathin-sectioning, immuno-TEM, metal-shadowing)
- Sample preparation for single particle cryo-EM (negative-stain screening, vitrification, cryo-screening)
- TEM imaging including fully-automated high-throughput data collection for single particle cryo-EM and cellular electron microscopy (tiling, stitching, tomography, 3D imaging)
- 2014-2018: Technical and basic application support for flow-cytometry (cell analysis and sorting)

Service concept

In light microscopy, we assist an increasing number of active users (currently about 50) utilizing shared light microscopes of the Institute at all levels ranging from basic user training to implementation of advanced imaging methods.

Due the technical complexity of in particular TEM sample preparation and cryo-EM, TEM is provided as full service and is performed in close collaboration with individual scientists and groups, respectively.

In flow-cytometry, our group provided basic technical and application support for both, cell analysis and sorting. Giving consideration to its crucial role for nearly all departments, the Institute decided to establish an additional service group dedicated to flow-cytometry. Initially, the FACS group consists of one scientist and one technician and started its operation in August 2018.

Material resources, equipment and spatial arrangements

Our group currently hosts the following shared equipment distributed over four publically accessible facility rooms:

- Wide field systems: Zeiss Z1 Axio-Imager, Zeiss Z1 Axio-Observer equipped for life-cell imaging), Zeiss Z1-Observer equipped with an apotome (Dept. Herrmann)
- Confocal laser scanning microscopes: Zeiss LSM700, Zeiss LSM710NLO with 2-photon laser, Zeiss LSM880 with Airyscan unit (the latter both equipped for life-cell imaging)
- Screening systems: Thermo Fischer Scientific ArrayScan VTI (with HCS-Studio software), Zeiss Celldiscoverer 7 (with Perkin-Elmer Columbus software)
- 3D-Imaging: Zeiss Z1 Lightsheet microscope
- Flow-cytometry: Cell analyser (BD Accuri) and two cell sorters (BD FACS Aria II and BD FACS ARIA Fusion)

Furthermore, our group operates three transmission electron microscopes and equipment for TEM sample preparation situated in separate facility rooms including:

- 100 kV Philips CM100 TEM with TVIPS F114 CCD camera
- 120 kV FEI Tecnai T12 Spirit TEM with TVIPS F416 CMOS camera
- 300 kV FEI Tecnai G2 Polara cryo-TEM with Gatan k2summit DED
- Edwards E306A and Denton Bench Top Turbo IV coating system
- Leica UC7 ultramicrotome with FC7 cryo-chamber
- 2 FEI Vitrobot cryo-plungers

Focus area of own research

- *Structural biology* – single particle cryo-EM structure determination of macromolecular protein complexes
- *Cell biology* – TEM analysis of cellular ultrastructure
- *Cell biology* – correlative light and electron microscopy

Scientific highlights

- Cryo-EM analysis of actively translating ribosomes obtained from *ex-vivo* derived human polysomes [Behrmann et al. Cell 2015]
- Cryo-EM analysis of ribosomal assembly intermediates [Nikolay et al. Mol Cell 2018]
- Automated TEM reveals ultrastructural changes involved in age induced memory impairment [Gupta et al. PLoS Biol 2016]

Results of completed projects during reporting period

Overcoming sample heterogeneity - Cryo-EM analysis of natively active protein complexes at near-atomic resolution

[Jörg Bürger, Thorsten Mielke]

Direct Electron Detector systems (DEDs) provoked a resolution revolution in single particle cryo-EM, now enabling near-atomic resolution. Within the collaborative research centre SFB740 (project Z1), we established a pipeline for automated high-throughput data collection, on-the-fly drift correction and CFT-based quality control of DED movie stacks. We further implemented multiparticle refinement strategies developed in Christian Spahn's lab, Charité, Berlin, which are essential to overcome sample heterogeneity exhibited by most protein complexes due to e.g. variable complex assembly and conformational flexibility. These tools enabled us to study protein complexes which are heterogeneous by nature such as actively translating human polysomes *ex-vivo* derived from HEK cells. From a single sample, we could identify eleven distinct functional states, ten of which representing intermediates of the ribosomal elongation cycle (figure 1, Behrmann et al. Cell 2015). Together with density gradient centrifugation, multiparticle refinement was also used to analyze assembly intermediates of *E. coli* 50S ribosomal subunits (Nikolay et al. Mol Cell 2018).

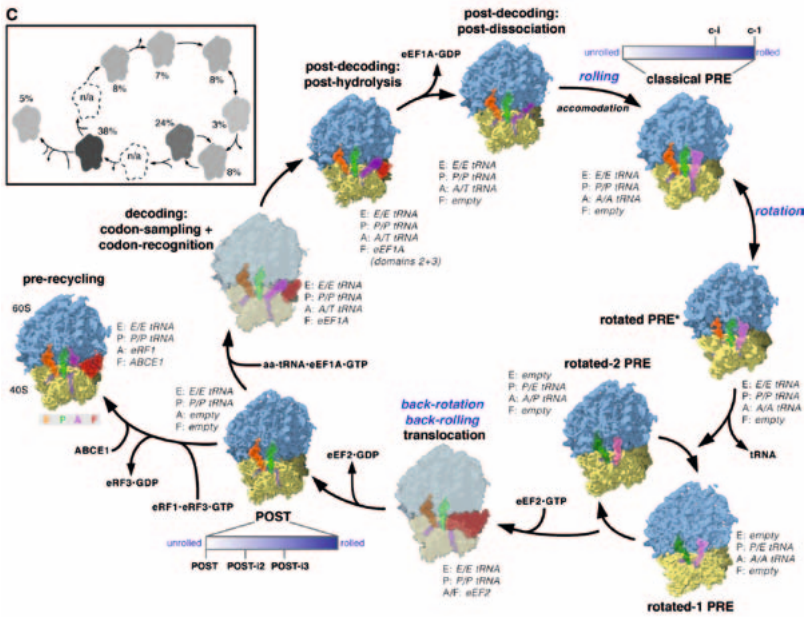


Figure 1. Experimentally observed intermediates of the human ribosomal elongation cycle (from Behrmann et al. Cell 2015)

Behrmann E, Loerke J, Budkevich TV, Yamamoto K, Schmidt A, Penczek PA, Vos MR, Bürger J, Mielke T, Scheerer P & Spahn CMT. Structural snapshots of actively translating human ribosomes. Cell 161: 845-857 (2015)

Automated cellular TEM: Analyzing the ultrastructure of cells and tissues

[Beatrix Fauler, Jörg Bürger, Thorsten Mielke]

We established automated TEM using Leginon and TrakEM2 to routinely image large areas of ultrathin-sections (up to 500 µm^2) at nm resolution. The visualization of entire cells and excerpts from tissues allows for statistical analysis of ultrastructural details such as e.g. T-bar structures in the calyx region of *Drosophila melanogaster* brains. We could show that the number of active zones per bouton area as well as the density of synaptic vesicles decrease with age whereas the average size of T-bar structures increases, indicating that age-induced memory impairment (AMI) in flies is indeed associated with structural re-organization at the presynaptic active zone (figure 2, Gupta et al. PLoS Biol 2016). Automated

TEM is also used to reveal unacquainted events accompanying cellular differentiation or disease processes. Following differentiation of hepatocyte-like cells derived from iPSCs, we could e.g. identify bile canaliculi representing a unique ultrastructural detail of mature liver cells (Matz et al. Sci Reports 2017). Furthermore, we applied automated TEM to characterize neural progenitor cells derived from human iPSCs (Lorenz et al. Cell Stem Cell 2017) and mutations involved in segmental progeroid disorders (Ehmke et al. Am J Hum Genet 2017).

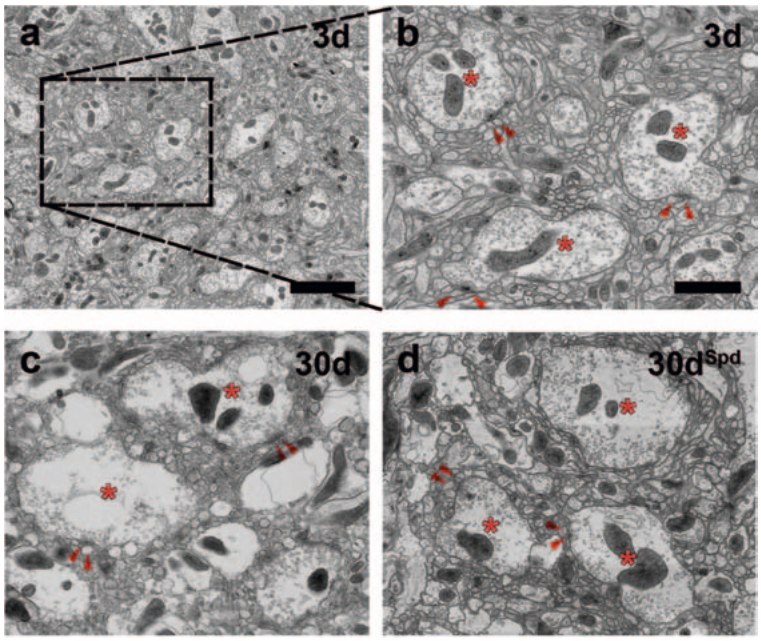


Figure 2.
(a) Montage of TEM images showing a whole calyx cross-section of a 3d w1118 fly (scale bar: 10 µm). (b–d) Higher magnification of PN boutons (asterisks) and dendritic claws of KCs (arrowheads) within the calyx of 3d, 30d, and 30dSpd w1118 flies (scale bar: 2 µm, from Gupta et al., 2016, PLoS Biol.).

Gupta VK, ... [11 authors]...Bushow R, Schwärzel M, Mielke T, Madeo F, Dengjel J, Fiala A & Sigrist SJ. Spermidine suppresses age-associated memory impairment by preventing adverse increase of presynaptic active zone size and release. PLoS Biol 14(9):e1002563 (2016)

Ongoing and future projects

"Multiparticle" cryo-EM – towards exploring the energy landscape of natively active protein complexes

[Jörg Bürger, Thorsten Mielke]

Cryo-EM and multiparticle refinement emerged as key technologies to decipher conformational flexibility of functionally active protein complexes. In fact, conformational states often represent intermediate states within a metastable energy landscape; and tuning this landscape e.g. by ligand binding is essential for the biological function of these complexes. In close collaboration with C. Spahn, Charité, Berlin, we are now analyzing actively translating polysomes *ex-vivo* derived from bacteria (*E. coli*), fungi (yeast) and higher eukaryotes (rabbit, human) – and hence the energy landscape of the translation elongation cycle and its tuning by e.g. antibiotics in different kingdoms of life. In a long-standing collaboration with P. Tavares, CNRS, Gif-sur-Yvette, France, and E. Orlova, Birkbeck, London, UK, we use a combination of conventional TEM and cryo-EM techniques to study the assembly of SPP1 bacteriophages (figure 3). In an ongoing project with M.L. Kraushar, Spahn lab, Charité, Berlin, we study native actively-translating ribosomes derived *ex-vivo* from prenatal mammalian brain tissue, aiming to identify essential translation regulators driving neuronal development with spatiotemporal specificity.

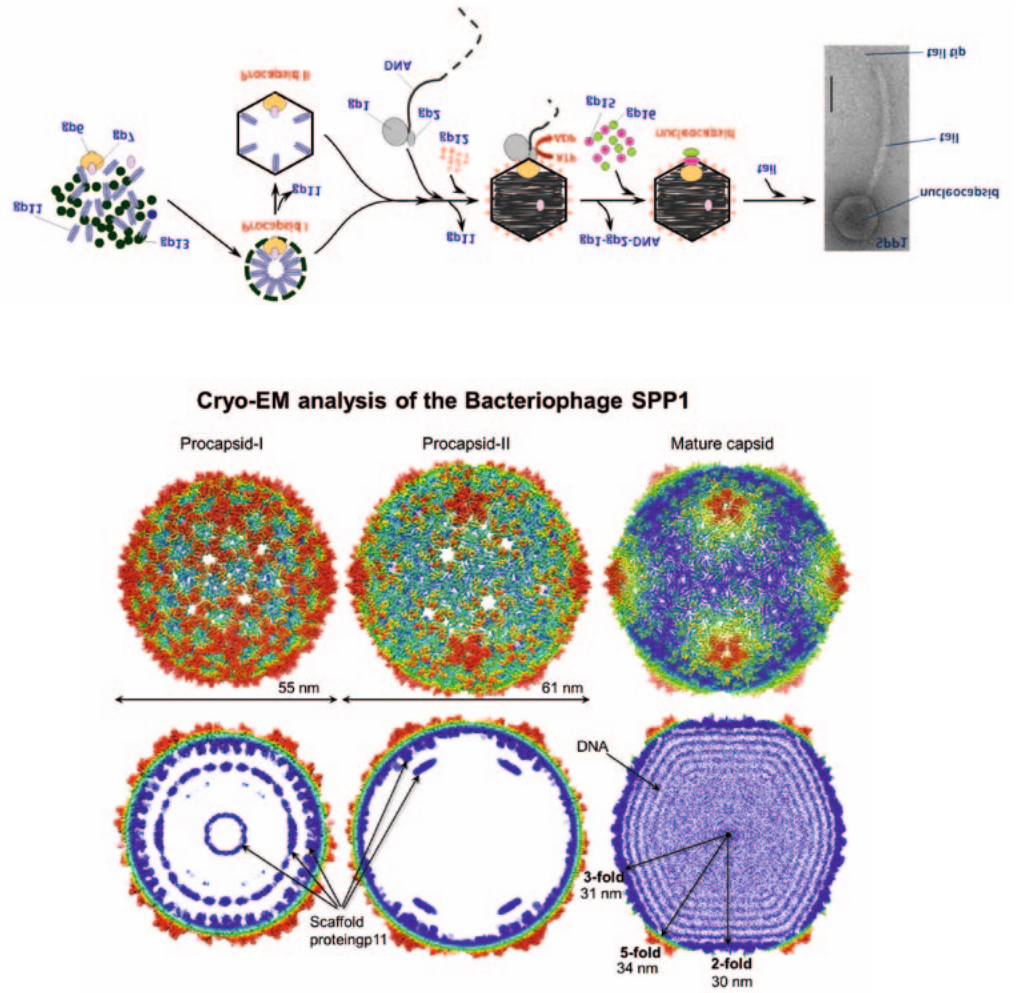


Figure 3.
Top: Assembly pathway of SPP1 bacteriophages. Bottom: Preliminary cryo-EM structures of Sus70 procapsids indicate structural rearrangements of the protein scaffold accompanying phage assembly.

Imaging organoids and other 3D model systems – new challenges for light and electron microscopy service

[Beatrix Fauler, René Buschow, Thorsten Mielke]

3D model systems such as organoids and gastruloids emerged as essential tools for many groups inside the Institute to study cell differentiation, developmental processes as well as related diseases including cancer. Hence, methods for imaging and analyzing these complex biological structures are at the strategic focus of our service group. This involves e.g. implementation of life-cell imaging and screening techniques to follow formation, growth and development of organoids, but also methods to gather high-resolution 2D and 3D information (figure 4). Accordingly, the Institute acquired new shared instruments including a highly-flexible research and screening microscope (Zeiss Celldiscoverer 7) and a confocal LSM880 system with Airyscan unit. In a joined project with Zeiss, we aim to implement protocols for screening and analyzing organoids in various multiwell formats with a strong focus on patient-derived cancer organoids studied in the Yaspo group. We provide application support for 3D imaging using wide field techniques (e.g. structured illumination), confocal Airyscan imaging, 2 photon and light-sheet microscopy. In another project, we will assist the newly established Hnisz lab to develop protocols for life-cell imaging of transcriptional condensates inside the nucleus of mammalian cells. Furthermore, we aim at strengthening support for image processing providing a platform for complex analysis tools (e.g. Columbus, HCS-studio, CellProfiler and deconvolution software).

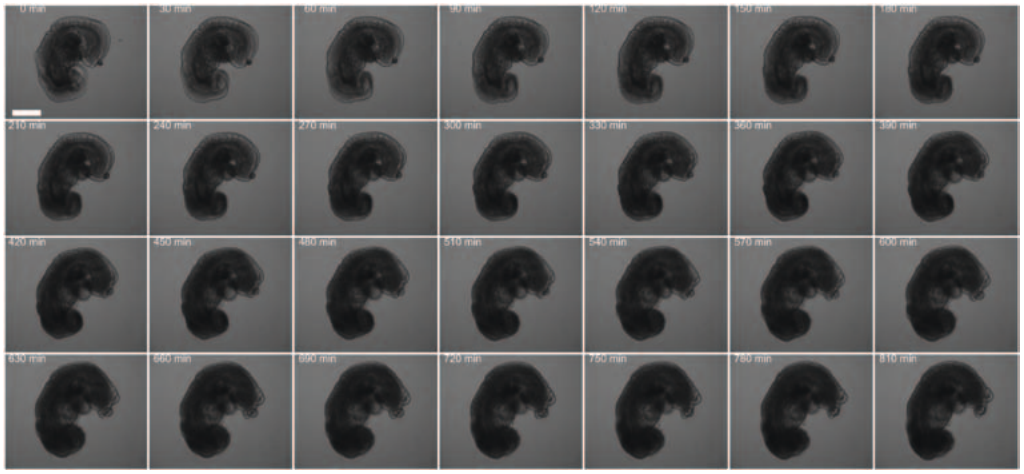
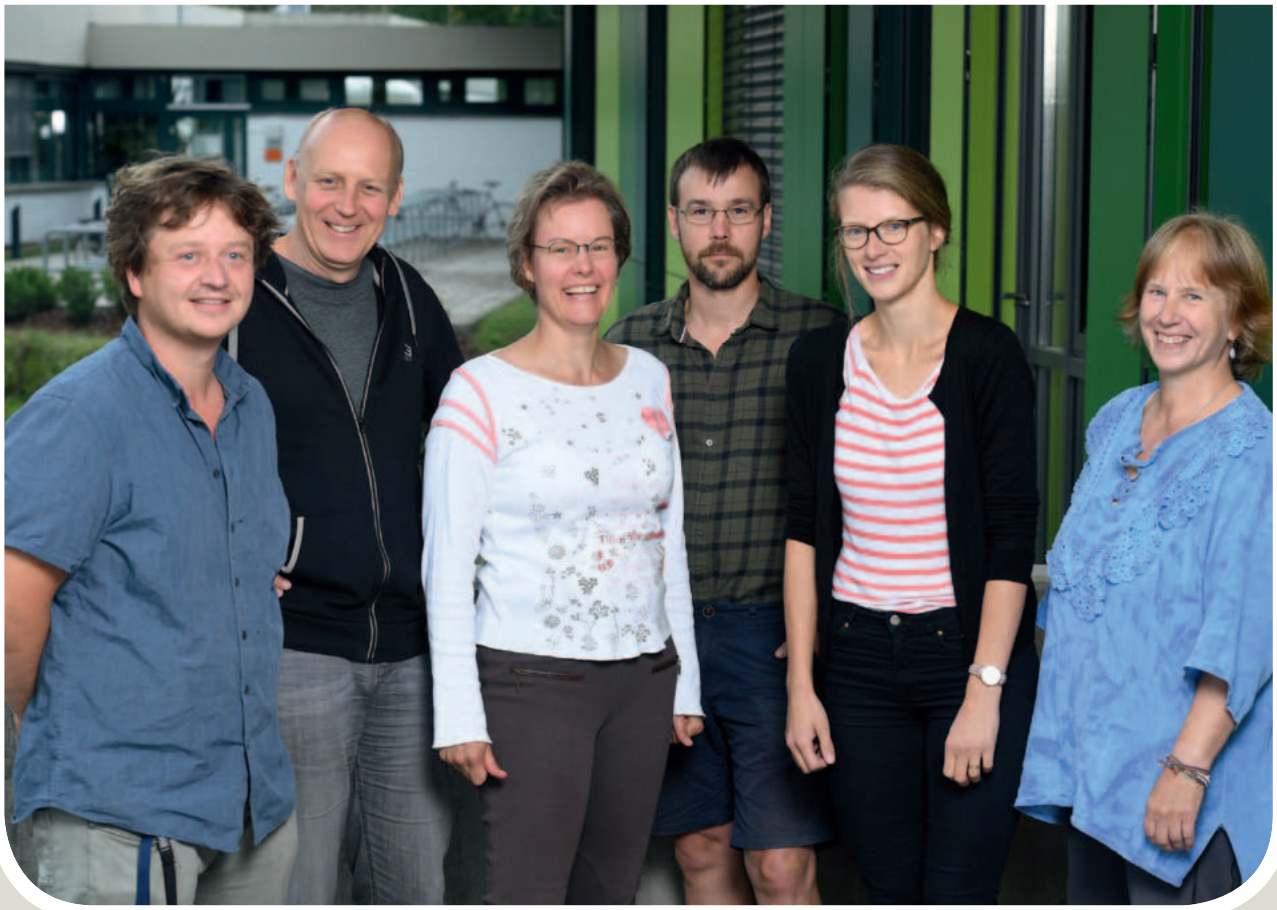


Figure 4.
Life-cell imaging of a mouse embryo with 30 min time interval (scale bar: 500 µm, project of J. Veenvliet, Dept. Herrmann)

Our own research focus hereby lies on ultrastructural analysis of cells, tissues and organoids. We recently used automated TEM to study the effect of BRAF and KRAS oncogenes in intestinal organoids (Brandt et al. BioRxiv 2018). Moreover, we establish correlative approaches required to relocate regions of interest identified by e.g. fluorescence microscopy at the much higher resolution (nm range) provided by TEM. In an ongoing project with W. Azab, FU Berlin, we exploit correlative techniques to image rare events e.g. the transfer of EHV 1 virus particles between PBMCs and endothelial cells. Our future goal here is establishing a workflow combining organoid screening, 3D fluorescence imaging and TEM ultrastructure analysis.

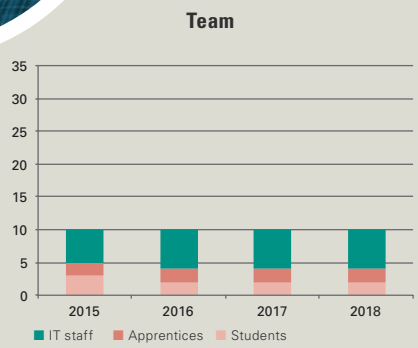


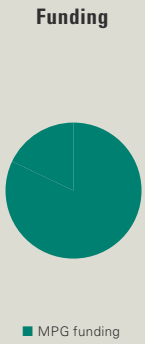
IT group

The IT group of the MPIMG has been established in 1995. It is headed by Donald Buczek and Peter Marquardt since January 2011.

Dipl.-Ing. Donald Buczek
buczek@molgen.mpg.de
++49 30 8413-1433

Dipl.-Ing. Peter Marquardt
marquardt_p@molgen.mpg.de
++49 30 8413-1430





Service offered by the group

The IT group is in charge of the operation and development of the IT-infrastructure of the Institute. This includes

- Global and internal internet connectivity
- Email servers, webmail
- Desktop workstation installation and deployment (hard and software)
- Running and maintaining interactive compute servers and a compute cluster
- Fileservers for scientific data (public data and confidential data)
- Daily data backup and longterm scientific data archives (>15 years)
- Providing a huge range of open source scientific software
- Software development and bug hunting
- Database service installation and maintenance
- Integration of scientific hardware in our network
- Scientific data management, data analysis pipelines
- Support for audio and videoconferencing
- Providing collaborative services
- Github enterprise
- Providing collaboration tools like wikis, etherpads, Nextcloud, syncthing and others
- Optimizing internal and external scientific data exchange
- Printing and plotting
- Helpdesk
- Apprenticeships for qualified IT specialists [Fachinformatiker] and Software Engineers

Service highlights

One of the highlights of the development of our IT infrastructure in the past years has been the establishment of our github enterprise, now running as a local instance. It offers private software repositories based on the github.com platform. The service is offered to all members of the Max Planck Society including their collaborators from anywhere around the world. Currently, we are hosting about 2,500 repositories for more than 700 accounts in 70 institutions.

Our compute cluster, consisting of about 25 multiprocessor nodes with up to 128 threads and 1 TB RAM each, is mainly used for genome analysis, alignments and simulations. Our main goal is to optimize data flows and computations to provide reliable results. The underlying hardware is composed of standard components, which could be exchanged or upgraded anytime. There are neither license fees nor is any proprietary software used to provide this service. This allowed the cluster to scale extremely well the last years, as it will in the future.

Material resources, equipment and spatial arrangements

The online storage capacity of the MPIMG file servers exceeds 9 PB spread over approximately 2,900 hard disks. The active disk-based backup capacity sums up to about 420 TB, whereas the tape and disk-based offline archived data currently comprises about 3 PB.

Presently the group serves about 250 Windows-based PCs and 240 Linux/Unix systems with a variety of hard- and software components and about 100 OSX systems. A variety of web servers are protected by a linux-based firewall, about 50 web servers are active and maintained. The active hard- and software development of the group serves the scientific departments as well as the service and administration groups.

The IT group installed a storage capacity of more than 9,000 TB. Currently, we are running about 6,700 CPU cores with 50,000 GB RAM spread over about 240 Linux systems. They range from single core systems with 256 MB RAM up to multicore servers with 160 threads and 2 TB RAM including our dedicated compute cluster.

Our internal network backbone is based on 10 GbE technology and is currently fed by approx. 50 interconnected network interfaces, from Isilon storage systems via multicore compute servers up to huge file- and archive servers. The in-house LAN is segmented by about 150 manageable switches giving us the flexibility to control each segment and, if necessary, to configure each switch port individually.

The servers of the MPIMG including storage and archive as well as the compute cluster are located in two physically separated server rooms. One of them in tower 4 is capable of supplying 180 kW cooling capacity and houses 20 server racks in a closed cold aisle containment system. The second server room located in tower 3 contains 30 racks in a warm aisle containment system, capable of cooling down 450 kW with full redundancy.

Current and future developments

The IT department is actively involved in planning networking infrastructure for major renovations or new buildings. We are trying to optimize dataflow from lab machines into the computing facilty using 10 GbE as standard. Also we are segmenting external managed lab hardware into VLANs to separate them from our scientific network.

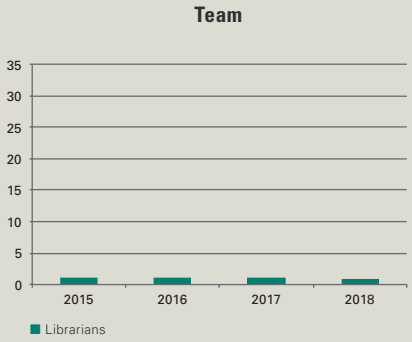
Our data management storage and archive concept providing up to more than 9 Petabyte is currently undergoing a migration from proprietary hardware RAID-based storage to open source software RAID. This migration implies copying and refreshing old archives as

well as establishing new long term concepts for disk-based storage. The main idea is to keep data storage formats as simple as possible to prevent running into manufacturers dependencies again in the future. Our goal is to keep all data accessible in the next ten years without running in a vendor lock-in, regardless of size and type of data.

Also, we are currently setting up a reliable data sharing infrastructure to fulfill the need for flexible scientific data exchange with our collaborators. This includes small document folders, preprocessed results or raw data from machines running in our labs. The key technology we are currently establishing is encrypted P2P data sync protocol based on Syncthing, a client software, which runs on almost all platforms without the need of centralized cloud services. First successful setups involved data exchange of several terabytes with some collaborating labs in Boston, MA.

Library & Scientific Information

Sylvia Elliger
elliger@molgen.mpg.de
++49 30 8413-1314



Service offered by the library

- Reference Service
- Bibliographic Service
- Selective Dissemination of Information Services
- Library Online Catalogue (OPAC)
- Library Website
- Access to Databases
- Institutional Repository (PuRe)
- Access to Open Access Resources
- Bibliometric Analysis
- Interlibrary Loan

Service concept, methods, and projects

The primary goal of the library is to provide the Max Planck Institute for Molecular Genetics community with access to information in every possible form and format covering the respective research areas.

The library of the MPIMG is a special scientific library and offers various methods for searching and acquiring scientific information and gives access to important electronic resources like e journals, e books and databases.

An own classification system is used for the arrangement of books and an online catalogue provides access to all literature held in the library as well as in the departments of

the Institute. Computing resources, including public computers, digital scanning and editing equipment are provided for the staff and guests of the Institute.

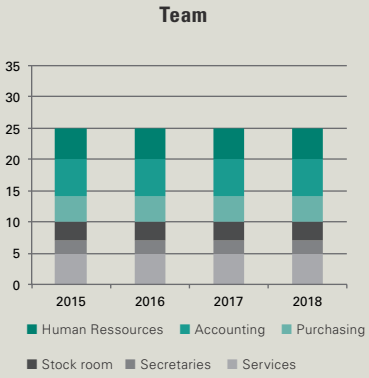
The Library/Scientific Information is developing its services proactively in order to respond to the changing information requirements of the Institute's staff.

Memberships

- Member of the "Arbeitsgemeinschaft der Spezialbibliotheken"
- Member of the Open Access Initiative within the Max Planck Society

Administration, Research Support

Dr. Christoph Krukenkamp
krukenkamp@molgen.mpg.de
++49 30 8413-1360



Service offered by the administration

The administration supports the further development of the institute by providing a dedicated and efficient administrative support in the following areas:

- Human resources
- Accounting and external project funding
- Purchasing
- Stock room
- Common services including reception, guest house, driver, and post office

Administrative concept

The administration of the Max Planck Institute for Molecular Genetics (MPIMG) secures smooth operations and stable infrastructures for the Institute. Besides the core administrative tasks, personnel administration and accounting, the administration takes care of purchasing and financial aspects of national and international grants. Scientists receive support in legal questions pertaining to technology transfer and patenting. This, like many other issues, is dealt with in close cooperation with the respective departments of Max Planck Headquarters in Munich.

Examples of completed and ongoing projects during reporting period

Establishment of new department Meissner

The appointment of Alexander Meissner, former at Harvard University and the Broad Institute, Cambridge, MA, has been successfully completed in August 2016. In July 2017, Alexander Meissner started as director and head of the Department of Genome Regulation

at the MPIMG in principal employment. Meissner moved to the second floor of tower 2, which has been completely renovated and equipped according to his needs. The new department started operation very quickly and meanwhile, Meissner has already successfully recruited four new group leaders, one of them a winner of a prestigious Sofja Kovalevskaja Award. In addition, various third-party funded projects of Alexander Meissner have been transferred from the US to Berlin, the new labs for genetic research according to German safety levels S1 and S2 have been registered with the legal authorities, and an authorization for planned experiments with laboratory animals has been obtained.

New rules of procedure of the MPIMG

In 2018, the MPIMG revised its rules of procedure, thus regulating the responsibilities and procedures within the Institute more precisely. The new rules of procedures govern the tasks of scientific and administrative management, the distribution of tasks and competencies and the interaction of all participants within the MPIMG. In consideration with the statutes of the Max Planck Society (MPG) and the Institute's statutes and based on a template from the Max Planck Headquarters, they also take into account the requirements of the commercial annual accounts.

Funding structures for junior scientists

In 2015, the MPG revised its funding structures for PhD students and postdocs. Starting from July 1st, 2015, all PhD students and postdocs starting at the MPIMG - except those funded by third parties like Humboldt Foundation, Volkswagen Foundation or others - are supported by funding contracts of the Max Planck Society (MPG). The program was implemented very successfully, so that meanwhile all PhD students in the Institute have funding contracts instead of fellowships.

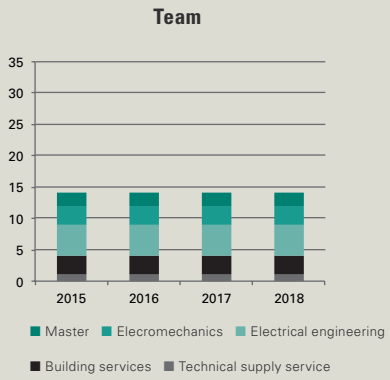
Implementation of commercial accounting system

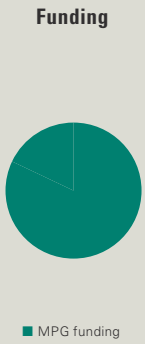
Also in 2015, the Max Planck Society has converted the formerly accounting-based accounting system to a new, commercial accounting system. This means that the MPG as well as each institute has to prepare annual financial statements and management reports in accordance with the provisions of the Third Book of the German Commercial Code for large corporations, taking into account the rules and peculiarities of the company, including an annual auditor's certificate issued by independent auditing companies.

In addition, a range of new regulations have been implemented by the Max Planck Society in recent years. For example, all institutes had to appoint tax and customs officers, who are familiar with the Institute's specific features and legal conditions. Also, a new policy for risk management has been established, including risk officers at the institutes and a central comprehensive risk database. Local versions of the risk database have to be adapted and updated by the institutes annually, and will then be consolidated in a central MPG-wide database, resulting in an annually adjusted risk map for the entire MPG. Furthermore, a Code of Conduct for the protection against sexualized discrimination, harassment and violence has been adopted by the Max Planck Society.

Technical Management & Workshops, Research Support

Dipl.-Ing. Ulf Bornemann
bornemann@molgen.mpg.de
++49 30 8413-1424





Service offered by the technical management & workshops

The technical management of the MPIMG is responsible for the operation and maintenance of the whole institute, including

- Condition of the buildings
- Operation and maintenance of electric energy supply, cooling and heating systems
- Operation and maintenance of steam generators, water conditioning etc.
- Supply with media like gases, liquid nitrogen etc.
- Condition of outdoor facilities
- Emergency service 24 hours/day, seven days/week
- Assistance of scientific departments and groups with lab moves or lab reorganizations
- Assistance with installation of new technical equipment

Highlights of work during reporting period

The most important challenge for the technical management during the reporting period has been the completion and bringing into service of tower 2. The structural and technical refurbishment of tower 2 took place from October 2013 until October 2016. It included a complete renewal of the building shell as well as the (structural and technical) realization of a new fire protection concept. All floor plans have been adapted to new laboratory utilization concepts and the complete technical media for electrical energy, heating, cooling, water, ultra-pure water, steam, gas, and compressed air supplies, ventilation and air conditioning, fire detection technology, communication and building automation have been newly installed. New ventilation plants and smoke extractors have been set up on the roof of tower 2 and new technical equipment for heating, water supply ultrapure water, and pure steam has been installed in the cellar of towers 2 and 1. In addition, a new emer-

gency generator has been put into operation. This technical equipment supplies three laboratory floors with labs of different size, equipment rooms, lab kitchens and offices for the departments of Bernhard Herrmann and Alexander Meissner and the research group of Stefan Mundlos. The fourth laboratory floor contains infrastructural lab space for all scientific groups, including two labs for experiments according to biological safety level S2 (German Genetic Engineering Law), an area for working with radioactive nucleotides, a teaching lab for student courses, and a seminar room for up to 24 people.

In winter 2016/2017, the department of Bernhard Herrmann as well as the research group of Stefan Mundlos moved to their new labs on the third and first floor of tower 2, respectively. This took place very smoothly and all groups could get on with their experimental work shortly after the move. Subsequently, tower 1 has been taken out of operation.

Figure 1.
View into a newly built large-capacity laboratory before start of the experimental work.

Ongoing and future projects of the technical management

The next step in the renovation of the entire property will be the renovation of tower 1. Unfortunately, this, initially, had to be postponed after the handing over of tower 2. Due to a significant increase of costs of the whole renovations measure as well as changes in the planned use, Max Planck Society (MPG) needed a new approval from the Joint Science Conference, the legal authority for major building measures of MPG, at first. By now, the approval is available and dismantling of tower 1 will start in late autumn 2018. Tower 1 will contain three lab floors for the department Meissner as well as for two other scientific departments, and a specialized lab floor for the microscopy and sequencing service groups. In addition, the library as well as a staff canteen will be located in tower 1.

The new technical equipment implicates numerous new challenges for the workshop staff. The employees are trained on an ongoing basis to keep pace with technological developments. In addition, we are working on a personnel development plan to make sure that the necessary know-how will be kept in the case of the age-related departure of our current staff.

Please notice:
At the publications cited in this Research Report, only the members of the respective groups are underlined. The funding charts in the individual group reports show funding in 2017. Only for the Meissner and Elkabetz labs, funding in 2018 is shown.

Imprint

Published by the Max Planck Institute for Molecular Genetics (MPIMG),
Berlin, Germany, October 2018

Editorial Board: Bernhard G. Herrmann, Alexander Meissner, Martin Vingron
Coordination: Patricia Marquardt
Graphic design: pigurdesign
Photos taken by: David Ausserhofer (pp. 20, 27, 28, 29, 30, 35, 36, 39, 40, 42, 49, 50, 54, 63, 67, 68, 76, 81, 82, 85, 86, 91, 92, 96 101 105, 110, 111, 115, 116, 126, 127, 130, 133, 139, 140, 145, 146, 148, 151, 156); Sascha Bachmann (p. 7); Gesine Born (p. 19 bottom); Lisa Bornemann (p. 154), Vera Kalscheuer (p. 75); Kirsten Kelleher (p. 15 top); Andreas Muhs (pp. 17, 155); Helen Nicolai (p. 2); Katrin Ullrich (pp. 19 top, 55, 60, 62, 100, 153).
Scientific Illustrations were provided by the respective research groups.

Contact:
Max Planck Institute for Molecular Genetics
Ihnestr. 63 – 73
14195 Berlin
Germany

Phone: +49 (0)30 8413-0
Email: info@molgen.mpg.de

For further information about the MPIMG, see
http://www.molgen.mpg.de