

RESEARCH REPORT **2022**

MAX PLANCK INSTITUTE FOR MOLECULAR GENETICS













RESEARCH REPORT 2022 MAX PLANCK INSTITUTE FOR MOLECULAR GENETICS

IMPRINT

Published by the Max Planck Institute for Molecular Genetics (MPIMG) Berlin, Germany, October 2022

EDITORIAL BOARD: Bernhard G. Herrmann, Alexander Meissner, Martin Vingron

COORDINATION: Friederike Krupp

GRAPHIC DESIGN: Nicola Graf

CONTACT:

Max Planck Institute for Molecular Genetics Ihnestraße 63 – 73 14195 Berlin Germany

Phone: +49 30 8413-0 Email: info@molgen.mpg.de Website: www.molgen.mpg.de

COVER PAGE:

The image shows mouse embryos at the blastocyst stage whose development has been paused *in vitro*. The inner cell mass is shown in green. Copyright: Vera van der Weijden, Bulut-Karslioglu Lab

IMAGE COPYRIGHT:

Photos were taken by Martin Ballaschk (p. 9, 11, 21, 23 top and 157), Katrin Ullrich (p.10), Anne-Dominique Gindrat (p. 22) and David Ausserhofer (all others).

Scientific illustrations were provided by the respective research groups. Copyright for all photographs and scientific illustrations is held by the Max Planck Institute for Molecular Genetics.



EDITORIAL

We are pleased to present the latest Research Report of the Max Planck Institute for Molecular Genetics (MPIMG), which summarizes both the scientific and organizational progress at our institute from 2018 – 2022. The report looks back at an unprecedently challenging and yet extremely productive period, it reflects on past achievements, and defines a vision for the future. All of these developments would not be possible without bright minds – graduate students, postdoctoral fellows and established researchers – driving innovative science.

We would like to take this opportunity to thank foremost the Max Planck Society but also the many other funding agencies who support our work by granting scientific freedom and independence to do the most exciting research and to tackle important scientific questions. We believe that the results presented in this report have made a significant contribution to the larger scientific community.

We continuously attempt to evolve the report to ensure we present a brief but comprehensive overview of our institute. For the Research Report 2022, we focused on our identity, culture and cutting-edge, collaborative work. As usual, more detailed information about our research can be found in the reports of the individual groups and further information in the Addendum.

Berlin, October 2022 Alexander Meissner on behalf of the MPIMG

Table of Contents

THE MAX PLANCK INSTITUTE FOR MOLECULAR GENETICS (MPIMG)

The MPIMG	9	
Organization of the MPIMG	.10	
Developments at the MPIMG	. 11	
Appointments of New Group Leaders	. 11	
Scientific Success at the MPIMG	.12	
Scientific Highlights	.12	
Awards to MPIMG Scientists	.16	
Careers of Young MPIMG Talents	.18	
Training Young Talents at the MPIMG	.18	
Committees and Associations at the MPIMG	.19	
Local Networks and Scientific Exchange	20	
Working Environment at the MPIMG	20	
Equal Opportunities	20	
Outreach Activities of the MPIMG	.21	
Animal Welfare at the MPIMG	23	
Impact of the COVID-19 Pandemic	24	
Outlook	24	

DEPARTMENT OF DEVELOPMENTAL GENETICS

DEPARTMENT OF GENOME REGULATION

Meissner Lab	36
Kraushar Lab	42
Müller Lab	48
Bulut-Karslioglu Lab	52
Elkabetz Lab	.57
Hnisz Lab	62

DEPARTMENT OF COMPUTATIONAL MOLECULAR BIOLOGY

Vingron Lab)	68
Herwig Lab		74
Kinkley Lab		80
Arndt Lab		85

RESEARCH GROUP DEVELOPMENT & DISEASE

Mundlos Lab	90
Kalscheuer Lab	98
Metzger Lab	03

OTTO WARBURG LABORATORY

 	107
 	113
 	119
 	124

MAX PLANCK FELLOW GROUPS

Reinert Lab	 30
Vallier Lab	 34

SCIENTIFIC SERVICES

Animal Facility	140
Transgenic Unit	143
Mass Spectrometry	145
Flow Cytometry	147
Microscopy & Cryo-electron Microscopy	149
Sequencing	
IT Service	155
Library and Scientific Information	

ADMINISTRATION

TECHNICAL MANAGEMENT & WORKSHOPS



As of September 2022.



The Max Planck Institute for Molecular Genetics: Genes to Genomes to Function

NA is the blueprint for all living organisms on Earth. The four-letter molecule has been at the center of attention for decades, but how primary DNA sequence translates to the range of phenotypes observed in nature is still mostly enigmatic. Major efforts are underway to map the genetic alterations that cause differences between individuals as well as between health and disease. Part of the complexity arises from the interplay of the various genetic, epigenetic, transcriptional and posttranscriptional layers of regulation that orchestrate metazoan development.

Our mission at the Max Planck Institute for Molecular Genetics (MPIMG) is to discover how genes and their regulatory landscapes give rise to organismal function. We are a diverse, collaborative and creative team of scientists unlocking new frontiers at the interface of gene and genome regulation, using various models of developmental and disease biology. Our core values include shifting scientific paradigms, embracing curiosity-driven discoveries, being exceptionally collaborative across scientific disciplines, and investing in the next generation of young talent.

THE MPIMG

The institute is home to over 200 scientists from over 40 different countries, who collaborate in 20 research groups, and are supported by our administrative team and facility management. Together, the groups combine different disciplines and approaches



Figure 2: The building of the MPIMG.

to gain fundamental insights in normal development and disease.

ORGANIZATION OF THE MPIMG

The institute currently has three departments, headed by Bernhard Herrmann (Department of Developmental Genetics), Alexander Meissner (Department of Genome Regulation) and Martin Vingron (Department of Computational Molecular Biology), that are complemented by the research group Development & Disease headed by Stefan Mundlos. Stefan is an External Scientific Member of the MPIMG and chair of the Institute of Medical Genetics and Human Genetics at the Charité – Universitätsmedizin Berlin.

The departmental labs at the institute are complemented by independent research groups organized within the Otto Warburg Laboratory (OWL). The OWL currently include the Max Planck Research Groups of Tugce Aktas and Andreas Mayer and the Lise Meitner Group of Edda Schulz. In addition, Yechiel Elkabetz, Peter Arndt and Marie-Laure Yaspo lead a permanent group at the institute (equivalent to tenured associate professor). All group leaders receive financial support from the Max Planck Society and manage their group independently.

Likewise, aspiring group leaders who have raised their own external funding are assigned to a specific department and – like the OWL group leaders bear full responsibility for their research (Aydan Bulut-Karslioglu, Denes Hnisz and Julia Metzger).
According to the Max Planck Society regulations, the position of these principal investigators is called Research Group Leaders.

To close the gap between postdocs and principal investigators, an additional level of organization was introduced at our institute. Group leaders conduct their research independently and receive funding from the MPIMG and/or other sources. Group leaders in the Meissner department are Matthew Kraushar and Franz-Josef Müller, in the Vingron department Ralf Herwig and Sarah Kinkley, and in the Mundlos group Vera Kalscheuer.

To strengthen ties with local universities and retain excellent researchers at our institute, Knut Reinert (professor at Freie Universität Berlin) and Ludovic Vallier (professor at the Berlin Institute of Health at Charité Universitätsmedizin Berlin) have Max Planck Fellow Groups at the MPIMG.

All directors, (research) group leaders and Max Planck fellows are equally responsible for advancing the institute's research mission, and are supported by our state-of-the-art service groups. Currently, one department at the MPIMG is vacant, and both the Max Planck Society and the institute are constantly screening outstanding scientist to recruit as director or group leader to the MPIMG.

DEVELOPMENTS AT THE MPIMG

Over the past years, we have begun to update the organizational structures at the institute and established the MPIMG Faculty that includes all directors, the research group and group leaders as well as the Max Planck fellows. The faculty meets regularly to discuss organizational matters and topics relevant to the institute's scientific future success. The faculty is informed at regularly scheduled meetings by the Managing Director, advises the Board of Directors in general institutional matters, and may suggest topics to be discussed by the Board of Directors.

As a main component of evolving the institute, we are constantly improving channels of scientific exchange and communication:

In 2022 we organized our first MPIMG Institute Retreat, which was a major highlight this year and will be held every other year in the future. The year in between will be used for individual group retreats and the institute-wide local one-day retreat called the Day of Science. The goal is to provide a comprehensive overview of the current scientific projects at the institute and to encourage in-house communication and collaboration.

We have introduced Faculty Chalk Talks in 2019 to provide a platform for sharing ideas and discussing novel concepts. The use of the chalk board and the small group size make this set-up an interactive highlight of scientific exchange at our institute.

We added User Seminars, where the heads of our service groups present news from their group, advertise new technologies and collect user feedback.

We have increased social networking at the MPIMG, which includes a Welcome Week for new PhD students, a monthly institute-wide Pizza&Beer event and ice-cream on the patio.

We also started a monthly newsletter to share important news and event details with the entire staff. This has been met with great response among all members of the community.

In 2019, a new intranet system was set up to facilitate the retrieval of administrative and organizational information. In addition, it features a news section, a calendar of internal and external events, and the latest publications by the institute's scientists. Article series present key positions in the organization and introduce new staff members.

At the end of 2018, we completely updated the MPIMG website and continue to optimize various aspects. A major step to improve external communication and outreach are our publication news stories.

The Dahlem Colloquium remains our premier seminar series and hosts internationally renowned scientists from around the world to present their current work. Due to the pandemic, we had to run this temporarily online, but now take advantage of the added flexibility and continue to host a wider range of speakers both in person and online. This is a crucial event to inspire the next generation of scientists as well as connect with scientific thought leaders around the world. Likewise, the internal Institute Seminar Series remains a key pillar for the MPIMG community and gives all researchers the opportunity to present their latest work in front of the entire institute.

APPOINTMENTS OF NEW GROUP LEADERS

The MPIMG is constantly looking for outstanding scientists who meet the institute's standards of scientific excellence and high-level leadership. We are currently actively looking for investigators at the director and group leader level. Our efforts to recruit two additional directors have been partially delayed through the COVID-19 pandemic and ongoing constructions. The renovation of tower 1 is expected to be completed in early 2023 (originally scheduled for 2019) and will not only provide crucial laboratory space but



Figure 3: Presentations at the 1st MPIMG Retreat at Lake Scharmützel, Brandenburg.

also essential features for meetings, including a canteen and a library.

Nonetheless, the institute has been able to attract highly motivated researchers at the group leader level that are contributing to the institute's scientific success with innovative ideas and expertise in various fields.

In December 2018, Tugce Aktas started as research group leader in the Otto Warburg Laboratory. She established the Max Planck Research Group Quantitative RNA Biology and works very successfully with her team on aspects of transposon-host interactions and their impact on evolution and wiring of post-transcriptional RNA processing networks since then.

In 2019, Sarah Kinkley established her group in the department of Martin Vingron. She studies higher order organization of DNA and chromatin in the nucleus.

Julia Metzger established the Heisenberg Research Group Veterinary Functional Genomics at the MPIMG in 2020 and, together with her team, is investigating the genetic factors that control length growth in the growth plate of bones.

Matthew Kraushar joined the MPIMG as International Guest Fellow in 2020. One year later he founded his group in the Meissner department. His lab is investigating the landscape of gene expression during brain development.

In summer 2022, Ludovic Vallier, Einstein Professor for Stem Cells in Regenerative Therapies at the Berlin Institute of Health at Charité joined the MPIMG as Max Planck Fellow to pursue fundamental research on the mechanisms of liver development and disease. This collaboration will open new ways to explore stem cells and organoids, and will provide exciting new insights.

An overview of all research groups at the institute can be found in the organization chart (p. 8).

Taking advantage of our institute's funding flexibility, we have very recently promoted additional outstanding postdoctoral researchers, similar to Matthew Kraushar, to establish their own groups. They include Helene Kretzmer, Francisca Martinez-Real and Daniel Ibrahim.

SCIENTIFIC SUCCESS AT THE MPIMG

MPIMG researchers are among the top tier of their peers in genomics, epigenomics and related fields, and in recent years the MPIMG has continued to excel in research work. For details on the accomplishments of individual groups, please refer to the respective chapters.

Scientific Highlights

Overall, MPIMG researchers have authored 444 publications between January 2018 and July 2022. 71 of them had an impact factor of 25 and higher (approximately 17% of all publications).



The strong collaborative links between research are illustrated by the fact that many of the institute's high

Figure 4: Number of peerreviewed publications (from January 2018 until July 2022).



Figure 5: Visualization of the inter-institutional networks for selected high impact factor publications (IF \ge 25).

impact papers are authored by two or more groups of the MPIMG.

Here, we present selected research highlights of the last years:

How Cell Diversity Emerges

Epigenetic regulatory factors act to modify the "packaging" of DNA without altering the underlying genetic information. Specifically, they bookmark the DNA and control which parts of the genome can be accessed in each cell. Most of these regulators are essential, and embryos lacking them tend to die during the time of development when organs begin to emerge. Researchers in the Meissner department investigated the role of epigenetic regulators for embryonic development with an unprecedented, single-cell precision in a comprehensive study published in Nature. Using the CRISPR-Cas9 system, they first specifically removed the genes coding for ten of the most important epigenetic regulators in fertilized oocytes. After the embryos had developed for six to nine days, the team examined the anatomical and molecular changes that resulted from the absence of a specific regulator. They found that the cellular composition

of many of the embryos was substantially altered in unexpected ways. Certain cell types were present in excessive numbers, while others were not produced at all. To understand these changes at a molecular level, the researchers examined hundreds to thousands of individual cells from embryos from which indi-



Figure 6: After 8 days, the developing mouse embryo resembles a seahorse (left). Without the epigenetic regulator PRC2, it is less complex (right).

vidual epigenetic regulators had been systematically removed. Then, the researchers sequenced the RNA of nearly 280,000 individual cells to investigate the consequences of the loss of function. When they compared the data of altered and unaltered embryos, they identified genes that were dysregulated, and cell types that were abnormally over- or underproduced. The study made it possible for the first time to observe very early stages of development in a level of detail that was previously unthinkable. The study paves the way for even more detailed studies, examining for example other factors such as transcription or growth factors or even a combination of these factors.

Grosswendt S.; Kretzmer H.; Smith Z.D.; Sampath Kumar A.; Hetzel S.; Wittler L.; Klages S.; Timmermann B.; Mukherji S.; Meissner A.; Epigenetic regulator function through mouse gastrulation; Nature (584), 2020.

Reading Between the Lines of the Genome

Non-coding DNA sequences control which genes become active when and in which cell. In a 2021 Nature study, an international collaboration with clinical researchers from Brazil and India, human geneticists at the University of Lausanne and the Mundlos group discovered a small piece of DNA in the noncoding genome. This segment appears to be the underlying cause for a rare genetic disease characterized by severe malformations of the limbs, where



Figure 7: Cartilage (blue) and bone (red) staining of the hand-paw isolated from a normal mouse (left) and a diseased mouse (right). The animal model displays malformations of the limbs such as the fused digits in the center, the polydactyly on the left, and the ventral fused digit on the right.

the distinction between the ventral and dorsal, i.e., the palm and back side, of the extremities has been lost. The non-coding DNA segment is located near the developmental gene engrailed-1 (En1) a gene that plays important roles during the development of the limbs, the brain, the sternum, and ribs. In addition, the scientists discovered that an RNA molecule, which they named Maenli (for Master Activator of Engrailed-1 in the Limb), was transcribed from the region that was missing in the patients. Thus, non-coding sequences can have a crucial role in regulating developmental processes and might be involved in the cause of various congenital diseases.

Allou L.; Balzano S.; Magg A.; Quinodoz M.; Royer-Bertrand B.; Schöpflin R.; Chan W.-L.; Speck-Martins C. E.; Rocha Carvalho D. R.; Farage L.; Marques Lourenco C.; Albuquerque R.; Rajagopal, S.; Nampoothiri S.; Campos-Xavier B.; Chiesa C.; Niel-Bütschi F.; Wittler L.; Timmermann B.; Spielmann M.; Robson M. I.; Ringel A.; Heinrich V.; Cova G.; Andrey G.; Prada-Medina C. A.; Pescini-Gobert, R.; Unger S., Bonafé L.; Grote P.; Rivolta C.; Mundlos S.; Superti-Furga A.; Non-coding deletions identify Maenli IncRNA as a limb-specific En1 regulator; Nature (592), 2021.

Unexpectedly Potent Protein Droplets

Repeats of individual building blocks within proteins are the cause of many hereditary diseases, but how such repeats actually cause disease is still largely unknown. MPIMG scientists investigated how repeat elongations lead to synpolydactyly, a developmental disease, caused by a defect in the HOXD13 gene that leads to a surplus and fused finger. The HOXD13 protein forms tiny droplets called condensates together with other molecules. The phenomenon occurs for practical reasons: the cell bundles key molecules it needs for specific tasks, like transcription, into little, droplet-like packages. In a Cell study, MPIMG scientists have shown that the mutated HOXD13 gene produces protein with altered chemical properties, causing a change in the composition and properties of HOXD13 droplets cells and, as a result, the protein's regulatory functions were impaired. This genetically driven change in condensate composition could be a new mechanism underlying many hereditary diseases.

Basu S.; Mackowiak S.D; Niskanen H.; Knezevic D.; Asimi V.; Grosswendt S.; Geertsema H.; Ali S.; Jerković I.; Ewers H.; Mundlos S.; Meissner A.; Ibrahim D.M.; Hnisz D.; Unblending of Transcriptional Condensates in Human Repeat Expansion Disease; Cell (181/5), 2020.



Figure 8: Microscopic image of cell nuclei of cultured cells. HOXD13 condensates are labelled in red. The DNA is stained in blue.

Moles: Intersexual and Genetically Doped

Female moles are intersexual and simultaneously develop functional ovarian and testicular tissue - the ovotestis, something that is unique among mammals. In a 2020 Science study, several MPIMG research groups presented the genetic peculiarities that lead to this characteristic sexual development in female moles. According to the study, changes in the 3D structure of the genome lead to altered control of genetic activity. In addition to the genetic program for testicular development, this also stimulates enzymes for high level testosterone production - a "natural doping" for female moles for their life underground. The researchers found that these evolutionary changes appeared not only in the genes, but, more importantly, in the regulatory regions belonging to these genes. When comparing the genome to that of other animals and humans, the team discovered an inversion - i.e., an inverted genomic segment in a region known to be involved in testicular development. The inversion causes additional DNA segments to get included in the regulatory domain of the gene FGF9, which reorganizes the control and regulation of the gene. The intersexual phenotype of female moles is thus a good example of how important the three-dimensional organization of the genome is for evolution.

Real, F. M.; Haas, S. A.; Franchini, P.; Xiong, P.; Simakov, O.; Kuhl, H.; Schöpflin, R.; Heller, D.; Moeinzadeh, M. H.; Heinrich, V. ; Krannich , T.; Bressin, A.; Hartmann, M. F.; Wudy, S. A.; Dechmann, D. K. N.; Hurtado, A.; Barrionuevo, F. J.; Schindler, M.; Harabula, I.; Osterwalder, M.; Hiller, M.; Wittler, L.; Visel, A.; Timmermann, B.; Meyer, A.; Vingron, M.; Jiménez, R.; Mundlos, S.; Lupiáñez, D. G.: The mole genome reveals regulatory rearrangements associated with adaptive intersexuality. Science 370 (6513), 2020.

Embryonic Development in a Petri Dish

During embryonic development, the emergence of a complex organism from a single cell requires a high degree of self-organization of stem cells and their progeny. Current 2-D models of cell differentiation have not been able to form any embryo-like structures in vitro. Using a new 3-D method, researchers from the Herrmann and Meissner labs have shown for the first time that mouse embryonic stem cells can form trunk-like structures in the Petri dish with anlagen for spinal cord as well as bone, cartilage and muscle tissue. This recapitulates the early shape-generating processes of embryonic trunk development. A key component triggering the morphogenetic processes is a special gel that mimics the properties of the extracellular matrix. The gel provides support for the cultured cells and allows them to orient in space, for example, by distinguishing inside from outside, and to arrange in an embryo-like manner. This new method opens a new area of research, synthetic embryology, and could replace animal experiments in various approaches on a scale impossible in mouse embryos, e.g., in mutant analyses or pharmacological studies.

Veenvliet, J. V.; Bolondi, A.; Kretzmer, H.; Haut, L.; Scholze-Wittler, M.; Schifferl, D.; Koch, F.; Guignard, L.; Sampath Kumar, A.; Pustet, M. ; Heimann, S.; Buschow, R.; Wittler, L.; Timmermann, B.; Meissner, A.; Herrmann, B. G.: Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites. Science 370 (6522), 2020.

How Genes Share Their Workspace

Genes and their genetic switches are organized into functional units to turn genes on or off and disrupting these units can lead to disease. In a Cell study, MPIMG researchers investigated in a collaborative effort what happens when these units get disturbed. In the ancestor of placental mammals, including humans, a new gene *Zfp42* was inserted right between the important developmental gene Fat1, a gene involved in growth and migration of cells, and its enhancer. When comparing cells from different



Figure 9: Collage of fluorescent trunk-like structures generated from stem cells at several stages of development (stained for expression of different key developmental genes).

mouse tissues – the developing embryonic limb and embryonic stem cells - the researchers found that the cells cope with the insertion of the new gene with two separate mechanisms, depending on its developmental state. From these studies, the researchers conclude that in embryonic limbs the enhancer contacts both genes but Zfp42 remains inactive. The gene had been silenced by DNA methylation, a chemical modification that locks genes in an off-state. Surprisingly, in mouse embryonic stem cells, the DNA around the two genes is organized completely differently, and Zfp42 and Fat1 now build their own physically separate "workspaces" with their own respective enhancers. These two novel mechanisms reveal how a single DNA "workspace" can be readily modified to host entirely different gene activities, indicating that functional units in the genome are more robust and flexible than previously thought.

Ringel A.R.; Szabo Q.; Chiariello A.M.; Chudzik K.; Schöpflin R.; Rothe P.; Mattei A.L.; Zehnder T.; Harnett D.; Laupert V.; Bianco A.; Hetzel S.; Glaser J.; Phan M.H.Q.; Schindler M.; Ibrahim D.; Paliou C.; Esposito A.; Prada-Medina C.A.; Haas S.A.; Giere P.; Vingron M.; Wittler L.; Meissner A.; Nicodemi M.; Cavalli G.; Bantignies F.; Mundlos S.; Robson M.I.; Repression and 3D-restructuring resolves regulatory conflicts in evolutionarily rearranged genomes. Cell 185 (20), 2022.

Awards to MPIMG Scientists

In recent years, the scientific work of the MPIMG has been recognized with numerous awards and competitive funding for MPIMG scientists. Here we present a selection of awards for MPIMG scientists.

Principal Investigators

Ludovic Vallier, Einstein Professor for Stem Cells in Regenerative Therapies at the Berlin Institute of Health at Charité, has been appointed as Max Planck Fellow. He will receive funding to establish a Max Planck Fellow group at the MPIMG. (2022)

Stefan Mundlos was honored with the Mendel Medal of the National Academy of Sciences Leopoldina. (2021)

Edda Schulz was honored with the EMBO Young Investigator Award. She has been accepted into the EMBO Young Investigators Program. (2021)

Martin Vingron was elected Chairman of the Biology & Medicine Section of the Scientific Council of the Max Planck Society (2021) (The sections are important bodies within the Max Planck Society in which the appointment of new directors and the establishment or closure of institutes or departments are discussed).

Franz-Josef Müller was appointed Deputy Director of the Department of Psychiatry and Psychotherapy at the University Hospital Schleswig-Holstein Campus, Kiel Germany. (2021)

Igor Ulitsky (Vingron lab) received the Meitner-Humboldt Research Award by the Alexander von Humboldt Foundation and the Israeli Ministry of Science, Technology and Space. (2021)

Edda Schulz was selected by the Max Planck Society for the Lise Meitner Excellence Program. Thus, her prospects as an MPIMG researcher have been extended by another five years. After that, there is the chance of a permanent position at the Max Planck Society. (2020)

Julia Metzger was granted a Heisenberg professorship with the aim of establishing a new research area within the genetic research of body size. (2020)

Alexander Meissner has been appointed honorary professor by the Freie Universität (FU) Berlin. He is thus a corporate member of the FU and he will participate in teaching and may give exams. (2020)

Martin Vingron has been elected as a member of the DFG review board "Foundations of Biology and Medicine: Bioinformatics and Theoretical Biology". (2020)

2019-2021: Alex Meissner und Denes Hnisz are listed as Highly Cited Researchers by the Web of Science. This recognizes the world's most influential researchers of the past decade, who have authored multiple highly cited papers that rank in the top 1% of citations by field and year.

Edda Schulz received the Career Development Award of the Human Frontier Science Program. (2019)

Alexander Meissner was elected Member of the European Molecular Biology Organization (EMBO, 2018).

Aydan Bulut-Karslioglu received the Sofja Kovaleskaja Award by the Humboldt Foundation. (2018)

Postdoctoral Researchers

César Prada (Spielmann lab) was awarded a Young Investigator Award for outstanding science by the European Society of Human Genetics. (2020)

Julia Batki (Meissner lab) joined the MPIMG with an EMBO Postdoctoral Fellowship. (2020)

Lila Allou (Mundlos lab) received the Young Investigator Award for Outstanding Science of the European Society of Human Genetics. (2019)

Uira Melo (Mundlos lab) received the Vienna Medical Academy Award of the European Society of Human Genetics. (2019)

Henri Niskanen (Hnisz lab) was awarded postdoctoral fellowships from the Emil Aaltonen Foundation, the Orion Research Foundation and Instrumentarium Science Foundation (2019).

Doctoral Students

Jana Henck (Spielmann lab) received the Young Investigator Award for Outstanding Science of the European Society of Human Genetics. (2022)

Philine Guckelberger (Meissner Lab) was awarded a Boehringer Ingelheim PhD Fellowship. (2021)

Jelena Ulicevic (Mayer Lab) has received a coveted doctoral scholarship from the German National Academic Foundation. (2021)

Annita Louloupi (Ørom lab) received the Marthe Vogt Award. (2020)

Leah Haut (Meissner Lab) was awarded a Boehringer Ingelheim PhD Fellowship. (2020)

Gemma Noviello (Schulz lab) is funded by the European Marie Sklodowska Curie Network for Innovative Education PEP-NET. (2019)

Nina Bailly (Meissner lab) received a grant from the German National Academic Foundation for her dissertation project. (2019)

Trainees

Jessica Müßigbrodt received the Trainee Award of the Max Planck Society. (2021)

CAREERS OF YOUNG MPIMG TALENTS

Part of our mission is to train and promote the next generation of scientists. Here, we present a selection of the career paths of scientists who have left the institute recently.

In 2019, Annalisa Marsico, former research group leader in the Otto Warburg Laboratory and assistant professor at the Freie Universität Berlin, became head of the research group Computational RNA Biology at the Computational Biology (Helmholtz Zentrum München).

Malte Spielmann (former group leader in the Mundlos research group) has accepted an appointment as Professor of Human Genetics at the Institute of Human Genetics of the University Hospital Schleswig-Holstein in Lübeck and Kiel, University of Lübeck. (2020)

Stefanie Grosswendt (former postdoc in the Meissner lab) established the research group From Cell States to Function at the BIMSB, Max Delbrück Center, Berlin. (2020)

Jesse Veenvliet (former postdoc in the Herrmann lab) is head of the Max Planck Research Group Stembryogenesis at the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden. (2021)

Olga Jasnovidova (former postdoc in the Mayer and Aktas labs) established a Max Planck Partner Group at Tallinn University of Technology, School of Science, Department of Chemistry and Biotechnology (Tallinn, Estland, 2022).

Zhike Zi, a former research group leader in the Otto Warburg Laboratory became professor at the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences. (2022)

Mike Robson, a postdoc in the Mundlos lab, will start in April 2023 as a group leader at the Berlin Institute for Medical Systems Biology at the BIMSB, Max Delbrück Center, Berlin, to further develop multi-omics techniques to dissect how the non-coding genome and its 3D structure interact to control gene expression and drive disease.

Lila Allou, a postdoc in the Mundlos group will start at the University College in London as Research Associate/Junior Research Group Leader in the Department of Neuromuscular Diseases in early 2023.

TRAINING YOUNG TALENTS AT THE MPIMG

In 2019, the new International Max Planck Research School for Biology And Computation (IMPRS-BAC) replaced the former bioinformatics-focused International Max Planck Research School for Computational Biology and Scientific Computing (IMPRS-CBSC). As the name suggests, the new graduate program combines computational and experimental research and provides broad training for our PhD students in



Figure 10: Young MPIMG talents.

close collaboration with the Department of Mathematics and Computer Science and the Department of Biology, Chemistry, and Pharmacy at Freie Universität Berlin. The IMPRS-BAC faculty includes bioinformaticians, computer scientists, biochemists, biologists, and physicists from the MPIMG, Freie Universität Berlin and associated institutes.

Every year, between 8 and 16 IMPRS PhD students are recruited in a structured process. As of May 2022, 117 PhD students are affiliated with our institute. Most of them are part of IMPRS, exceptions are possible for students funded by other sources. IMPRS is very international, with students coming from 24 different countries, including 33% from Germany and 62% from European countries. Currently, 58% female and 42% male students are enrolled in the program.

IMPRS students have been very successful, with 173 scientific articles published with an IMPRS PhD student as first or last author since 2005. In addition to training in research, the IMPRS curriculum offers career support through transferable skills, tutoring opportunities and the annual Otto Warburg Summer School. Despite the COVID-19 pandemic, the IMPRS has been very active, in organizing workshops on transferable skills (e.g., scientific writing, statistical literacy, science communication, project management) and nanocourses (e.g., computational protein design, growth factor signaling, compact data structures, bioethics).

Excellent supervision of the PhD students is ensured by a Thesis Advisory Committee (TAC) and by a supervision agreement that defines the cornerstones of the training structure, funding and supervision quality. These measures have been strongly recommended by the faculty of the MPIMG and have been well received by the students. The TAC consists of the primary supervisor and two or three other scientists from the MPIMG or other institutions. The student and the TAC meet annually and discuss scientific, practical, and career-oriented questions. The TAC also gives the student the opportunity to discuss supervision problems confidentially.

COMMITTEES AND ASSOCIATIONS AT THE MPIMG

Traditionally, Max Planck Institutes are governed by the Board of Directors, which is composed of the directors of the departments who are entitled to vote and permanent guests, including External Member Stefan Mundlos, the speaker of the Otto Warburg Laboratory, the Head of Administration and Scientific Coordination. The Board of Directors delineates the scientific orientation of the institute and develops the future roadmap. For the day-to-day business, the directors elect the Managing Director among them.

MPIMG committees serve as an invaluable additional layer of supportive organization. These structures encourage employees from different levels and of different professional groups to participate in leading the institute into the future. There are several established committees at the MPIMG, including the Animal Welfare Committee, the Housing Committee and the Dahlem Colloquium Committee. In addition, we have recently established new committees. For instance, our Recruitment Committee streamlines the review of applications for research group leader and other key scientific positions, maintains contact with applicants and proposes candidates to the faculty. The Service Group Committee has been organizing an external review of our service groups in spring 2022. One of the most recent additions is the Sustainability Committee which oversees the MPIMG's efforts to transform the institute into a more sustainable research institution. Thus, the involvement of scientific and non-scientific staff in the MPIMG's committees promotes employee participation and drives the institute's development.

The Student Association (STA) is a representative body that has been advocating for the interests of all MPIMG students since 2001. Each department and the OWL groups are represented by a student on the STA Steering Committee. Since 2017, a STA representative has also been invited as a regular guest to the extended meetings of the Board of Directors. Moreover, the STA provides a platform for organizing social and networking events to promote scientific discussion and the exchange of technical and scientific knowledge between students. An example of this is the monthly "bioinformal" meeting organized by PhD students, where short talks and interactive discussions on open questions and new methods take place in a relaxed atmosphere. The regular seminars at the MPIMG and at the Freie Universität Berlin also help to strengthen collaborations and networking among all students.

The Postdoc Association (PDA) was reinstated in January 2019 to promote networking among MPIMG postdocs and to connect with postdocs of other research institutions in Berlin. The PDA organizes seminars given by late-stage postdocs and young group leaders, workshops on career development, and various other meetings to boost scientific careers. The PDA also regularly meets with the managing director to discuss relevant topics.

LOCAL NETWORKS AND SCIENTIFIC EXCHANGE

Over the past years, we have strengthened our ties at all levels with the local universities as well as the various sites of the Charité Universitätsmedizin Berlin. The coordination among the leadership for recruitments and other development questions has been frequent and successful. MPIMG group leaders are typically often associated with local universities or hospitals (e.g., Charité and Berlin Institute of Health, Freie Universität Berlin) and work closely with renowned research institutions around the world. All of these research partnerships are actively pursued and will continue to contribute to the scientific success of the MPIMG in the future.

The MPIMG is a member of Berlin Research 50 (BR50), an alliance of Berlin's non-university research institutions that have joined forces to strengthen Berlin as an international science hub. The initiative intends to develop research strategies, enhance the dialogue between science, politics and society and foster cooperation with Berlin's universities.

The institute is also part of the German Stem Cell Network (GSCN) that brings together expertise in stem cell research in Germany and strengthens synergies between basic research, regenerative medicine and pharmacology by fostering innovative research.

In fall 2019, the Berlin Institute for Molecular Systems Biology (BIMSB) and the MPIMG were spearheading an initiative to improve the connectivity between them and to promote joint research programs with the first BIMSB – MPIMG symposium.

In summer 2022, members of the MPIMG faculty co-organized the Max Planck Frontiers Symposia @ BMS Science Week in Berlin.

Several other seminar set-ups, such as the Berlin Stem Cell Club and the RNA Club Berlin have been established in collaboration with the MPIMG and other research institutions to connect with scientists from various Berlin research institutions and universities.

WORKING ENVIRONMENT AT THE MPIMG

At the MPIMG, we aim to create an inspiring and safe working environment for all. We believe that our research can only thrive when scientists, core staff, and administrative and technical personnel work together in a stimulating environment where everyone is equally valued. We are committed to the well-being of our employees and value all who work at the MPIMG, regardless of gender, race, or disability. We are convinced that diversity at the workplace increases creativity and, thus, improves the quality of research.

Equal Opportunities

The MPIMG pays particular attention to gender equality and the compatibility of career and family. In December 2017, the MPIMG became one of the first institutes in the Max Planck Society to adopt a gender equality policy, especially in respect to career development for female scientists, reconciling work and family, and promoting the visibility of gender issues. The MPIMG Gender Equality Plan has been updated in 2021 and, like the report before, has been awarded the "Gold" label by the Max Planck Society.

At the MPIMG, women are strongly encouraged and supported in applying for career advancement, mentoring programs, awards and prizes. For instance, the Max Planck Society has launched the Boost! Program in 2019 to counteract the underrepresentation of female researchers in scientific leadership positions. Two of our scientists, Sarah Kinkley and Helene Kretzmer, received this funding and have since both founded their own groups. Ongoing efforts are also made to increase the number of women in fields where they are underrepresented. To this end, the Equal Opportunity Officer is involved in recruitment procedures, and statistics for gender-sensitive personnel development have been introduced. In addition, the MPIMG promotes the role model function of female scientists by e.g., inviting more female speakers and compiling statistics on the various career paths of MPIMG alumni.

The Max Planck Society is strongly committed to the compatibility of work and family life and in 2018 was awarded the Work and Family Certificate for the fifth time by the non-profit berufundfamilie Service GmbH for its family-friendly human resources policy. The MPIMG actively supports employees with children



Figure 11: MPIMG female scientists in the lab.

by offering family-friendly and flexible working hours, home office and childcare at nearby daycare centers. Childcare is also provided at institute seminars and events.

Finally, the MPIMG sensitizes its employees to the importance of equality issues through various measures: the Equal Opportunities Officer is a regular guest at the Extended Board of Directors meetings, she reports regularly at the works council meetings, and her work has been made more visible on the MPIMG's website and on the intranet. Care is also taken to ensure that the MPIMG's official documents and forms are worded in a gender-appropriate manner.

OUTREACH ACTIVITIES OF THE MPIMG

MPIMG Institute Retreat 2022

In fall 2022, more than 160 scientists, students and technicians spent three days at a conference hotel on Lake Scharmützel near Berlin. The goal was to connect people across groups and to exchange and

discuss new ideas - especially after the corona pandemic had hindered much of the normal social interaction over the past 2.5 years. The first day started with team activities, including a GPS scavenger hunt through the surrounding woods and parks, an open-air poster session, a dinner with randomly premixed seats, and a pub guiz. On the second day, the participants discussed current scientific projects at the MPIMG and the retreat concluded on the third day with a look into the future. Young group leaders presented their vision of their lab and issues relevant for the future of the institute had been discussed (e.g., equality issues and environmental and financial costs in biomedical research). The retreat was very well received and the next retreat is already scheduled for fall 2024.

Public relations

In recent years, the MPIMG has pushed its effort to inform the public about the institute's scientific achievements and their impact on society. We actively sought contact with the media and provided expert opinions on science-related topics (e.g., SARS-CoV2, animal testing, genome editing), resulting in various appearances on television, radio and print.

Online and social media activities

In 2018, the institute's website received a modern and clean look and a responsive layout. Now, information about the MPIMG research activities is provided for both scientific peers and lay people on the website.

The online reach and social media activities of the MPIMG have grown significantly in recent years. The institute's twitter channel @MPI_MolGen was revived in summer 2019 and has proven to be an effective tool for both science-to-science and science-to-public communication. Not only are the voices of the MPIMG's many Twitter users amplified by re-tweeting, quoting, and commenting, but also talks, press releases, new publications, and media appearances are announced on the account. Since the channel's revival, the audience has grown by 50-60 followers each month, reaching a following of almost 2.800 in September 2022.

Since 2021, we have been using our LinkedIn page to post job openings. As a result, the number of followers has more than doubled to 8.200 in September 2022. In the future, we will expand our presence on this platform to connect with former, current and future employees. To reach YouTube's broader and younger audience, we collaborated with the Max Planck Society communications department for a video commenting on the scientific accuracy of the movie "Gattaca". We plan to pursue similar collaborations in the future.

Networking and spreading the word through participation in conferences and visits was partially interrupted by the COVID-19 pandemic. As direct interactions have been difficult over the past two years, digital networking tools have become even more important (e.g., Twitter, LinkedIn and YouTube). To actively strengthen these skills, science communication experts recently held seminars and workshops to enable MPIMG scientists to improve their science communication skills

Corporate Identity

We have further developed the Institute's brand and corporate identity. We adopted the new corporate design of the Max Planck Society for digital and print media, refreshed our distinctive "Helix" logo, and integrated the Max Planck Society's Minerva emblem.



Figure 12: Impressions of the MPIMG building at the event "Long Night of the Sciences".



Figure 13: High-school students visit the Imaging and the IT service group at the "Girl's Day" and "Boy's Day".

Public outreach

The MPIMG considers educational outreach an integral part of public relations. Despite the COVID-19 pandemic, we have tried to maintain this part of outreach activities. The institute participated in the annual Berlin event "Long Night of the Sciences", where 70-80 universities and research institutions throughout Berlin and Potsdam open their doors for one night and invite the public to visit their laboratories and learn about their work. Many MPIMG employees have volunteered in the past to give talks, demonstrate experiments, and tour the labs with visitors of all ages.

We also continue our efforts to communicate animal research in collaboration with the Charité -Universitätsmedizin Berlin, the Max Delbrück Center for Molecular Medicine, and other research institutions by organizing public events on animal experimentation ("Tierversuche im Gespräch").

Educational outreach

MPIMG research groups regularly organize practical courses, internships for high school students and PhD students and participate in "Girls' Day" and "Boy's Day"– an initiative which introduces young female high school students to male-dominated professions and vice versa.

ANIMAL WELFARE AT THE MPIMG

For the foreseeable future, animal research will remain an essential component of life science research. The MPIMG monitors and, if necessary, improves its efforts in animal welfare, with a special emphasis on the 3R principles (Replacement, Reduction and Refinement). For example, the Animal Facility has implemented a health monitoring system, thereby eliminating the use of sentinel animals, improved the housing of laboratory mice by introducing enrichment materials, and facilitated euthanasia by introducing a safe and stress-free method for euthanization.

In accordance with the new legal requirements, the MPIMG has established an Animal Welfare Board consisting of two animal technicians, two scientists of each department or research group, the head of the Animal Facility and all animal welfare officers. The Board meets regularly to discuss animal welfare issues at the institute (e.g., ethical considerations, classification of severity of procedures, methodolog-



Figure 14: View into the Animal Facility at the MPIMG.

ical refinement). In addition, education and training activities are provided for staff of the Animal Facility and scientists (e.g., advanced and basic courses for experimenters). The MPIMG also proactively engages in internal communication (e.g., through tours of the Animal House and seminars on Animal Science) and public outreach (e.g., video shoots about the work with laboratory animals at the MPIMG).

IMPACT OF THE COVID-19 PANDEMIC

The COVID-19 pandemic had a significant impact on the institute with approximately half of our staff becoming infected with SARS-CoV2 in 2022 alone. But thanks to a high vaccination coverage, a tight (voluntary) testing scheme and the consideration of the staff, the MPIMG has managed the COVID-19 pandemic very well.

In March 2020, following a decision by the Berlin Senate, the institute went into emergency operations for several weeks, and only essential parts kept operating. Most experimental work was suspended at this time, forcing many researchers to focus on work outside the laboratory. All seminars, in-person meetings, the MPIMG institute retreat 2020 and guest visits were cancelled and shifted to online formats. To make the institute a safe place to work during the pandemic, a crisis team met weekly to decide on immediate and long-term measures to manage the pandemic, taking into account national and local legislation and Max Planck Society's central administration regulations. In July 2020, the institute cautiously and gradually returned to normal operations - with several hygiene restrictions in place: access control, minimum distance, obligatory masks, and regulations for the number of persons per office, laboratory and meeting room.

Already in January 2021, testing with RT-LAMP twice a week was offered for the staff. Later, testing with rapid antigen tests was also possible for MPIMG employees. A vaccination initiative at the institute started in June 2021 and MPIMG staff as well as employees of neighboring institutes offered doses of RNA vaccine.

OUTLOOK

Despite some local and global challenges that we will face in the coming years, the MPIMG is well positioned and overall very optimistic about the future. In terms of infrastructure, we look forward to moving into the newly renovated tower 1 in early 2023, where scientific groups and the Sequencing and Imaging service groups will occupy new space that has been specifically designed to meet their needs (e.g., special ventilation systems). The renovation of tower 4 remains a high priority and will hopefully be initiated before critical infrastructure fails. Particularly in view of the climate crisis and rising energy costs, we are currently developing measures to make the institute a more environmentally friendly and sustainable place to conduct research.

Recruitments at all levels remain a high priority and we are confident that the progress and current culture of the institute, as outlined in this report, will be of interest to many investigators around the world and make us competitive in a global market for talent.

Research at the Max Planck Institute for Molecular Genetics



Department of Developmental Genetics

ESTABLISHED IN NOVEMBER 2003, UNTIL FEBRUARY 2024

PROF. DR. BERNHARD G. HERRMANN herrmann@molgen.mpg.de +49 30 8413 / 1344

SCIENTIFIC ASSISTANT: Dr. Marta Caparros Rodriguez caparros@molgen.mpg.de +49 30 8413 / 1344

SECRETARY: Christel Siegers siegers@molgen.mpg.de +49 30 8413 / 1344



Structure and organization of the department

In 2022, the department consists of several project groups (Amaral, Bauer, Koch, Tsaytler and Schifferl) and one research group (Schrewe). Project groups receive technical support, and some comprise a Master and/or a Ph.D. student. All project groups report to and are supervised by the department head. The PhD students are co-supervised by the heads of the project groups and the department head.

Focus areas

- Developmental genetics Gene regulation networks controlling embryonic body formation in the mouse
- Non-Mendelian inheritance mechanisms causing transmission ratio distortion in the mouse

Research concept

The department of Developmental Genetics investigates two major questions:

a) how are trunk and tail formation genetically controlled during embryogenesis, and

b) what are the basic mechanisms of non-Mendelian inheritance in mammals?

Understanding mammalian development requires deciphering the mechanisms by which stem cells give rise to various cell lineages, and how they organize into functional three-dimensional structures and organs. This is an immensely complex process, requiring thousands of protein-coding as well as non-coding genes, which change activity 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 early lineage decisions during trunk and tail development, a major focus of our work is the role of T in shaping the embryonic trunk and tail in relation to these factors.

Bodies are formed from billions of cells. Therefore, understanding development comprises unraveling gene regulation mechanisms at the single cell level and deciphering 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 ratio of alleles from the parents to their offspring. The classical model of non-Mendelian inheritance in mouse (and mammals as such) is the phenomenon of transmission ratio distortion (TRD) caused by the mouse *t*-haplotype. Transmission of the latter from heterozygous (*t*/+) males is almost exclusive; up to 99% of their offspring inherit the *t*-haplotype. We have worked out the molecular mechanisms promoting this selfish behavior of the *t*-haplotype in 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

PWD/Ph-encoded genetic variants modulate the cellular Wnt/β-Catenin response to suppress Apc^{Min-}triggered intestinal tumor formation. *Farrall et al., Cancer Research, 2021*

RAC1 controls progressive movement and competitiveness of mammalian spermatozoa. Amaral and Herrmann, PLoS Genetics, 2021

A 37 kb region upstream of brachyury comprising a notochord enhancer is essential for notochord and tail development. *Schifferl et al., Development, 2021*

Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites. *Veenvliet et al., Science, 2020*

Two isoforms of the RAC-Specific guanine nucleotide exchange factor TIAM2 act oppositely on Transmission Ratio Distortion by the mouse *t*-haplotype. *Charron et al., PLoS Genetics, 2019*

Research projects

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

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). This is 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 events. Naturally, the dynamics of these processes in space and time are difficult to study *in vivo* in mammals. 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.

Recently, we succeeded in generating 3D structures that recapitulate trunk development (we call them trunk-like structures, TLS). TLS induce all three germ layers in a spatially organized manner reminiscent of their *in vivo* counterpart, as shown by extensive characterization of TLS comprising scRNA-seq, in collaboration with the Meissner department. Importantly, TLS contain cells co-expressing Brachyury and Sox2 at their posterior end, resembling neuro-mesodermal progenitors. These cells form paraxial mesoderm and neuroectoderm, developing into somites and a neural tube, respectively, at an embryo-like pace (fig. 1). The TLS system is a powerful platform for studying the morphogenetic changes and lineage decisions, underlying mid-gestational development in the culture dish. We work on further improving TLSs by trying to induce a node and notochord as well as lateral mesoderm.



Figure 1: Generation of trunk-like structures (TLS) with somites and neural tube from mouse embryonic stem cells (mESCs) in a dish. T::H2B-mCherry marks mesodermal, Sox2::H2B-Venus neural cells; Uncx and Tcf15 are somitic marker genes. (From: Veenvliet et al., Science, 2020)

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

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



Figure 2: Identification of a brachyury (T) enhancer essential for notochord formation and axial development. NotomC marks notochordal, α -T and α -Sox2 mesodermal or neural cells, respectively. TNE, T notochord enhancer; T^{cd}, complete deletion of T region; T^{Δ TNE}, deletion of TNE. (From: Schifferl et al., Development, 2021)

We have shown that the neural tube and the mesodermal tissues derive from neuro-mesodermal progenitors (NMP) giving rise to the trunk. We understand in 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, besides single cell transcriptomics, require a toolbox of finely tuned genetic instruments and high-resolution microscopy. We have investigated the control of brachyury in the node, notochord and tail bud NMPs and identified several enhancers that are essential for notochord development and axial elongation. We have published one notochord enhancer and its role in axial elongation recently (fig. 2). Our next goal is to provide a deeper insight into the trunk organizer, lateral mesoderm development, tail NMP formation and tail outgrowth.

Genetic and physiological basis of transmission ratio distortion in the mouse

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 *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. According to our model, 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 progressive movement of sperm carrying the *t*-haplotype and confer 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.



Figure 3: Less progressive motility of +-sperm derived from t/+ males is rescued by RAC1 inhibition with NSC. LIN, linearity of sperm movement. (From: Amaral and Herrmann, PLoS Genetics, 2021)

We have shown that Tiam2, a Rac1-specific GEF, is acting as distorter. We have also shown that its target, the Rho small G-protein Rac1, indeed controls the progressive movement and thus the competitiveness of sperm. Our physiological data now confirm the current model that the distorters control the progressive motility of sperm (fig. 3). Now, we are looking for a direct link between the distorters, SMOK, and components of the axoneme.

Inherited beneficial genetic variants have the power to suppress tumor formation despite strong cancer driver activity

Recently Tomasetti and Vogelstein raised the idea that the majority of human cancers are due to "bad luck", that is, due to cancer driver mutations occurring randomly during stem cell division. However, it has been shown that genetic predisposition affects the penetrance of tumor-initiating mutations, such as APC mutations that stabilize β -catenin and cause intestinal tumors in mice and humans. The mechanisms involved in genetically predisposed penetrance are not well understood yet.



Figure 4: Tumor suppression despite cancer driver activity by APC^{Min} involves multiple cis-controlled gene variants. (From: Farrall et al., Cancer Research, 2021)

We have analyzed tumor multiplicity and gene expression in tumor-prone Apc^{Min/+} mice on highly variant C57BL/6J (B6) and PWD/Ph (PWD) genetic backgrounds. We found that (B6 x PWD)F1-APC^{Min/+} offspring mice were largely free of intestinal adenoma, and several chromosome substitution (consomic) strains carrying single PWD chromosomes on the B6 genetic background displayed reduced adenoma numbers. In collaboration with the Herwig group, we showed that multiple dosage-dependent modifier loci on PWD chromosome 5 contribute to tumor suppression. We showed that the activation of β -catenin-driven and stem cellspecific gene expression in the presence of Apc^{Min} or following APC loss remained moderate in intestines carrying PWD chromosome 5, suggesting that PWD variants restrict adenoma initiation by controlling stem cell homeostasis. Gene expression of modifier candidates and DNA methylation on chromosome 5 were predominantly cis-controlled and largely reflected parental patterns. This data provided a genetic basis for the inheritance of tumor susceptibility (fig. 4). Our analysis highlights the strong impact that multiple genetic variants acting in networks can have on tumor development. Thus, the individual cancer risk relates more strongly to the set of genetic variants inherited from the parents than to "bad luck" in picking up cancer driver mutations during stem cell division.

The Mediator complex and transcriptional regulation

Transcriptional regulation is arguably the most important step controlling the decision of which genes are to be expressed at any given time.

The Mediator complex functions as a co-activator of enhancer activity and serves as a molecular bridge between distal enhancers and the transcriptional machinery at core promoters. It is a general target of transcription factor activation domains. Different transcription factors bind specifically to different Mediator subunits, and Mediator subunit Med12 is the docking site for important developmental transcriptional regulators, e.g., β -catenin, Sox9, and Sox10. By studying conditional Med12 tissue- and cell type-specific knock-out mouse mutants, we have identified key Med12-dependent signaling pathways.

In order to identify novel candidate Med12-interactors, we have targeted the Med12 locus in ES cells that now express fusion proteins of Med12. We used TurboID for proximity- and epitope-biotinylation and mEGFP for Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins (RIME) for analysis of chromatin complexes that involve Med12. Both approaches provided various proteins that directly or indirectly are associated with Med12 in ES cells. We are currently generating mouse lines with these ES cells to expand these analyses to embryos and adult tissues.

Equipment and resources

The Department of Developmental Genetics is housed on the third floor of tower 2. The department provides the developmental biology community with the Molecular Anatomy of the Mouse Embryo Project database (Mamep: http://mamep.molgen.mpg.de). The department provides personnel (1 scientist, 1 engineer) and expertise to the Transgenic Unit of the institute, which produces transgenic embryos and mice from genetically modified ES cells for all groups at the institute.





Department of Genome Regulation

ESTABLISHED IN JULY 2017

PROF. DR. ALEXANDER MEISSNER meissner@molgen.mpg.de +49 30 8413 / 1880 뇌 Meissner Department SECRETARY Cordula Mancini office-meissner@molgen.mpg.de +49 30 8413/1881

Structure and organization of the department

The department of Genome Regulation is headed by Alexander Meissner and contains several outstanding international scientists. All group leaders regularly meet to discuss scientific and organizational questions. We also organize department wide retreats and monthly meetings that provide diverse training opportunities for our students and postdocs. Currently, the department consists of seven independent research groups: Meissner lab (since 2017), Elkabetz lab (since 2017), Bulut-Karslioglu lab (since 2018), Hnisz lab (since 2018), Müller lab (since 2018), Kraushar lab (since 2022) and Kretzmer lab (since 2022).

Research concept of whole department

The department of Genome Regulation is a coalition of international experts that share a passion for innovative and high-impact genomics research. We bring together a strong expertise in mouse and human stem cell models, early mouse development, transcription biology and epigenetics, with a wide range of experimental and computational approaches. The overarching theme of the department is to obtain a comprehensive picture of how our genomes are regulated during development and which parts are derailed in disease states.



Ongoing work and planned developments

Over the past 5 years, we have established a fully operational and lively department that has also successfully integrated into the larger institute. Members of the department have been extremely productive and co-authored over 60 peer-reviewed manuscripts, filed several patents and raised notable thirdparty funding since 2018. Details for the various completed, ongoing and future studies are described in the individual sections.

Material resources, equipment and spatial arrangements

The department currently offers classic molecular biology bench space for around 30 people and has three tissue culture rooms (7 hoods and 12 incubators) with state-of-the-art equipment. All groups share the tissue cultures to maximize interaction within the department and the Meissner, Bulut-Karslioglu and Hnisz 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 the cell biology. Moreover, members of the department have full access to the MPIMG's extensive technological and computational resources.



MEISSNER LAB

Genome Regulation

ESTABLISHED IN JULY 2017, ALEXANDER MEISSNER IS THE MANAGING DIRECTOR SINCE 2019

PROF. DR. ALEXANDER MEISSNER meissner@molgen.mpg.de +49 30 8413 / 1880

凶 Meissner Lab



Focus areas

- Developmental biology genome regulation in early embryogenesis
- Cancer epigenomics mechanism, function and origin of the cancer epigenome
- DNA methylation regulatory dynamics in development, aging and disease
- Pluripotency and reprogramming mechanisms that regulate cell fate 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. The more recent discovery of possible shared regulatory principles between early extraembryonic development and cancer have led to a greater focus on studying this form of non-canonical epigenetic regulation. DNA methylation is part of that and generally remains a central focus in our basic research program to study genome organization, including the annotation of regulatory features and their dynamics during development. We also continue to study the underlying mechanisms that regulate cell fate transitions, in particular exit and entry into pluripotency.
Comprehensive analysis of ALL methylation data reveals an unusual epigenetic landscape. *Hetzel et al., Nature Cancer, 2022*

Discovery of focal methylation turnover at somatic enhancers in pluripotent stem cells.

Charlton et al., Nature Genetics, 2020

Zygotic disruption of epigenetic regulators and single cell profiling of gastrulation embryos. Grosswendt, et al., Nature 2020

Extracellular matrix supports *in vitro* derivation of advanced embryonic structures. *Veenvliet et al., Science, 2020*

Cas9-based molecular barcoding provides new insights into mammalian embryogenesis. Chan et al., Nature, 2019

Scientific honors and invited talks

• Top 1% Highly Cited Researcher by the Web of Science (2019-2021)

Selected invited talks

- Gordon Conference "Chromatin Structure and Function", Barcelona, Spain (2022)
- "ISSCR International Symposium", Jerusalem, Israel (2022)
- Annual meeting of the European Association for Cancer Research (virtual, 2021)
- "9th Epigenomics of Common Diseases" (virtual, 2020)

Research projects

Molecular recording of mammalian embryogenesis

In this study, we used custom technology for high information, multi-channel, and continuous molecular recording and presented a cell fate map underlying mammalian gastrulation. Several key ideas have emerged from this study: the transformative nature of Cas9-directed mutation with a single cell RNA-seq readout, how historical information from such a molecular recorder can complement RNA-seq profiles to characterize cell type, and an early framework for quantitatively understanding stochastic transitions during mammalian development. The modularity of our molecular recorder allows simple substitutions that will increase its range of applications (Chan et al., Nature 2019).



Figure 1: Lineage tracing in mouse experiments. The target site (within the 3' untranslated region of mCherry) and the three-guide array are encoded into a single piggyBac transposon vector. The vector, transposase mRNA and Rosa26::Cas9:eGFP sperm are injected into oocytes to ensure early integration and tracing in all subsequent cells after zygotic genome activation. ITR, inverted terminal repeat.

TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers

Here, we reported a comprehensive analysis of a series of human ESC lines devoid of DNMT3 and/or TET activity. We found that TETs are focally recruited to thousands of somatic enhancers that are highly methylated in the pluripotent state, which results in local 5hmC enrichment and targeted demethylation in the absence of countering DNMT3 activity. Together, our results highlight a novel focal activity of the methylating and demethylating enzymes that adds yet another unique epigenetic mechanism to the regulation of pluripotent cells (Charlton et al., Nature Genetics, 2020).



Figure 2: A representative example of three sites that are highly methylated in undifferentiated pluripotent stem cells and show dramatic loss in DNMT3A/B double knockout (DKO) cells (compare to the background that loses little methylation). The same sites are bound by HAND1 in mesodermal cells upon differentiation and marked by H3K27ac, which indicates their putative enhancer function.

Epigenetic regulator function through mouse gastrulation

In this study, we presented a combined genetic perturbation and scRNA-seq strategy to functionally dissect mammalian embryogenesis. Our platform is designed to understand complex mutant phenotypes, both anatomically and molecularly, and to account for variations across replicates that may be fundamental to a given developmental process. We have investigated a number of key epigenetic regulators that cause lethal post-gastrula-

tion phenotypes and have been difficult to fully characterize because they are presumed to buffer differentiation across many different contexts. Taken together, these strategies may ultimately yield a complete description of the interactions between genetic and epigenetic mechanisms that govern ontogeny (Grosswendt et al., Nature, 2020).



Figure 3: Composition and expression changes in EED-mutant embryos. The mutant state proportions were compared to matched wild-type stages (circle sizes). Gene expression correlation (purple to yellow) was determined by using our defined marker genes. States are organized according to lineage. The earliest states (within conceptus) are less affected compared to later states. Many later states are not observed (grey), but whether some could be produced before lethality remains unclear. FC, fold change.

Acute lymphoblastic leukemia (ALL) displays a distinct, highly methylated genome

Our comprehensive, high-coverage genome-wide methylation data set of ALL patients, healthy control samples and ALL cell lines provided important insights into the particular methylation landscape of ALL compared to other hematopoietic as well as solid malignancies. In the context of a recently proposed model, we believe that the absence of global hypomethylation in ALL and acute myeloid leukemia should be further investigated and will provide potential insights into the underlying deregulation (Hetzel et al., Nature Cancer, 2022)..

Dynamic antagonism between key repressive pathways in the placental epigenome

Here, we utilized trophoblast stem cell (TSC) lines as a model to investigate the unusual placental epigenome. We find that persistent intermediate methylation in extraembryonic cells is maintained through a dynamic, antagonistic relationship between two distinct epigenetic repressive pathways – DNA methylation and the Polycomb repressive complexes (PRC) – that canonically regulate mutually exclusive subsets of the genome of most somatic cells. It is worth noting that similar dramatic genome-wide shifts in methylation away from heterochromatic regions and towards CGIs have been considered characteristic features of many different cancer types since the mid-1980s. It is tempting to speculate that fundamental regulatory principles are shared between our investigated epigenetic regulation in TSCs and primary tumors. If this is indeed the case, it would shine new light on a major cancer hallmark and highlight relevant parallels in normal development and disease (Weigert, Hetzel et al., in revision Nature Cell Biology).



Figure 4: Simplified model of DNA methylation and PRC2 dynamics in somatic cells compared to the cancer-like epigenome found in TSCs: Somatic cells generally regulate genetic loci in a bistable fashion, preserving an overall highly methylated genome with minimal contributions from PRC2, and unmethylated CGIs that are targeted by PRC2 and protected from DNMT3s. In TSCs, the genome shifts to an overall intermediate, seemingly metastable methylation state, which cooccurs with PRC2-deposited H3K27me3. Although this state can be driven to high or low methylation levels by modulating the activity of these two inputs, this form of genome regulation is robust enough to return to the steady intermediate state.

Cancer cell lines converge to distinct non-physiological epigenomic states

Cancer cell lines have long been known to exhibit more extreme methylation levels at selected CGIs and partially methylated domains (PMDs) compared to primary tumors. However, a clear genome-wide assessment of the consistency and genomic nature of these alterations across commonly established cancer cell lines has not been undertaken, including an assessment of how these landscapes may implicate unique forms of epigenetic regulation. Using data from the cancer cell line encyclopedia (CCLE) and the cancer genome atlas (TCGA), we show that most available cancer cell lines exhibit an "inverse bimodal" landscape, characterized by extreme CGI and low PMD methylation levels. Other epigenetic states also exist, including a pattern characterized by genome-wide hypermethylation that includes PMDs as well as CGIs, but very few preserve primary-like features (Hetzel et al., in preparation).



Figure 5: Scatterplots of the relationship between mean methylation entropy and mean CpG methylation at hypermethylated CGIs for normal solid tissues, corresponding cancer samples and cell lines. The increased CGI methylation levels in cancer cell lines are accompanied by reduced methylation entropy.

Spatial transcriptomic maps of whole mouse embryos

We performed embryo-wide spatial transcriptomic profiling using Slide-seq to decipher tightly regulated gene expression patterns of ~27,000 genes in developing embryos during the early stages of organogenesis (E8.5-E9.5). The curation of these datasets led to the creation of highly resolved spatial transcriptomic atlases, which are extremely useful tools for studying gene expression patterns during development. The reconstruction of a digital three-dimensional (3D) E8.5 mouse embryo with our innovative computational tool "sc3D", which enables quantitative exploration of gene expression patterns and gradients on a digital *"in situ"* basis, represents a key advance in the field. (Kumar et al., in revision Nature Biotechnology)



Figure 6: Schematic of the experimental workflow and data analysis. Sagittal sections of mouse embryos at E8.5 (15 sections of 10 μ m thickness from an entire embryo with 30 μ m distance between consecutive sections) and E9.5 (6 sections of 10 μ m thickness each from 3 embryos covering the head and thoracic regions) were obtained and used for Slide-seq V2 spatial transcriptomic mapping. Dotted lines indicate the approximate position of the embryos while sectioning. NMPs, neuromesodermal progenitors; WT, wild type; KO, knockout; A, anterior; P, posterior; D, dorsal; V, ventral.

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.





KRAUSHAR LAB

High-resolution Neurogenetics

ESTABLISHED IN OCTOBER 2021

MATTHEW L. KRAUSHAR, PHD, MD matthew.kraushar@molgen.mpg.de +49 30 8413 / 1660 ↘ Kraushar Lab



Focus areas

- Post-transcriptional regulation in nervous system development
- Protein synthesis in prenatal neocortex neurogenesis

Research concept

The lab's primary research concept is that the cellular transcriptome is dynamically and selectively translated as a regulatory step during the patterning of the evolutionarily advanced nervous system. The mammalian neocortex is the most evolutionarily advanced brain region. Neocortex neural stem cells express broad, regionally defined transcriptomes, which are refined at the protein level to specify precise distinctions in neuronal identity in neocortex layers, circuits, and connectivity. This post-transcriptional regulation converges on the ribosome, the final gatekeeper of gene expression. How ribosome regulation impacts the neocortex proteome during developmental neurogenesis is the specific concept of our lab.

This first paper coming from the Kraushar Lab discovered a transient critical window in brain development when regulation of mRNA translation peaks. With a combined bioinformatic multi-omics approach and *in vivo* analysis, we find spatiotemporal uncoupling of transcription and translation for thousands of mRNAs, in addition to a mechanism controlling large-scale changes in ribosome abundance within distinct neuronal lineages.

Harnett*, Ambrozkiewicz* et al., Nature Structural & Molecular Biology, 2022 (accepted)

Selected invited talks

- Institute of Molecular Pathology (IMP, Vienna, Austria (2022)
- Conference "Neurogenesis from Development to Adulthood in Health and Disease", Ascona, Switzerland (2022)
- "FENS: The Brain Conference RNA Mechanisms and Brain Disease", Rungstedgaard, Denmark (2021)
- EMBL-EMBO conference "Protein Synthesis and Translational Control", Heidelberg, Germany (2021)

Research projects

Building upon the structural biology work recently published (Kraushar et al., Mol Cell, 2021), our lab completed a bioinformatics-focused project analyzing mRNA translation in the mouse brain neocortex across the prenatal and perinatal developmental period during neurogenesis (fig. 1).



Figure 1: Prenatal neocortex development and brain circuit wiring (adapted from DeBoer* & Kraushar* et al., 2013).

Our high-resolution analysis included all the major reactants, synthesis, and products of protein synthesis, with a combination of RNA-seq, tRNA qPCR array, Ribo-seq, and mass spectrometry (fig. 2A-B; Harnett* & Ambrozkiewicz* et al., Nat Struct Mol Biol, accepted ahead of print). In this work, we discovered a transient window of dynamic mRNA translation at mid-gestation (fig. 2C), with timed translation upregulation of chromatin-binding proteins, and translation downregulation of ribosomal proteins, during neurogenesis (fig. 2D-G).



Figure 2: Timed dynamic mRNA translation in neocortex development (adapted from Harnett* & Ambrozkiewicz* et al., Nat Struct Mol Biol, accepted ahead of print).

We find three mechanistic features of timed translational control during neurogenesis: 1) A release of ribosome pausing during the elongation of specific amino acids (fig. 3A); 2) a sharp decrease in ribosome abundance (fig. 3B); and 3) a sharp change in the activity (phosphorylation) of the translation initiation factor eIF4EBP1 (fig. 3C). These data have led us to hypothesize that changes in amino acid supply signal to decrease translation during neurogenesis via the activity of eIF4EBP1.



Figure 3: Mechanisms of mRNA translation control in neocortex neurogenesis (adapted from Harnett* & Ambrozkiewicz* et al., Nat Struct Mol Biol, accepted ahead of print).

mTOR is the cellular sensor for amino acid levels, and signals to eIF4EBP1 to control the translation of ribosomal proteins (fig. 4A). Therefore, we speculate that mTOR signaling is a key mechanism of translational control during neocortex neurogenesis, driving neuronal subtype specification in development. We have begun to test this hypothesis by measuring amino acid levels in mouse neocortex (fig. 4B) and mTOR pathway members (fig. 4C) using mass spectrometry. We find a coordinated shift in amino acid abundance during development. In addition to eIF4EBP1 (fig. 3C), downstream mTOR targets like ribosomal protein S6 are dynamically phosphorylated (fig. 4D). Strikingly, S6 phosphorylation is neuronal subtype-specific in evolutionarily advanced neuronal lineages like upper layers (L2-4) and subplate (SP). We are testing the evolutionary advancement of neuronal S6 phosphorylation in human brain development in collaboration with Dr. Zeljka Krsnik at the University of Zagreb.



Figure 4: Amino acid signaling to mTOR in neocortex neurogenesis (A, adapted from Saba et al, 2021; B-D, unpublished data Kraushar Lab).

Amino acids pass through the blood brain barrier (BBB), which develops in parallel with the neocortex. We are collaborating with Prof. Chenghua Gu at Harvard University, to pursue similar experiments as those above (fig. 2-4) in mice with mutations disrupting BBB integrity - linking amino acid supply, the BBB, and mRNA translation in neocortex development. Such a mechanism is implicated in neurodevelopmental disorders, and we are further collaborating with Dr. James Millonig at Rutgers University to investigate human cellular models of autism spectrum disorders.

Historically, neocortex gene expression has been studied mainly by RNA-seq and mass spectrometry for mRNA and protein steady-state, where the birthdate and lifetimes of these molecules is unknown. We are performing metabolic labeling and subsequent click chemistry to capture snapshots of *timed active* mRNA transcription and translation in developing neocortex organotypic slice cultures, where the developmental birth and maturation of neurons continues *ex utero* (fig. 5). We have been awarded three years of funding from the Minerva Stiftung to utilize this organotypic system in collaboration with Dr. Yonatan Stelzer at the Weizmann Institute, where we will dissociate the neocortex after metabolic labeling of gene expression to push our analyses of bulk tissue to the single-cell level.



Figure 5: Click chemistry labeling of active transcription and translation in neocortex organotypic cultures (unpublished data).

Finally, we have initiated a new direction for the Kraushar lab, focusing on the fundamental physico-chemical forces driving ribosome assembly. The ribosome is composed of ~ 79 core ribosomal proteins on a scaffold of 4 rRNAs. The driving forces for these individual components to concentrate in the nucleolus and to initiate ribosome subunit assembly remains a major open question. We have been awarded three years of funding from the DFG SPP2191 (Molecular mechanisms of functional phase separation) to test the hypothesis that ribosomal proteins form a multicomponent network fluid in the nucleolus to condense around transcribed rRNA (fig. 6A). Our preliminary bioinformatic analyses suggest that ribosomal proteins are composed of highly intrinsically disordered regions (IDRs), which may be drivers of liquid-liquid phase separation (fig. 6B). Indeed, we find that ribosomal protein disordered domains are capable of phase separation *in vitro* (fig. 6C). Future *in vitro* and *in vivo* experiments in this SPP2191 project aim to establish a new paradigm ribosome biogenesis, with a combination of biochemical and biophysical methods, including cryo-electron microscopy, per our experience (Kraushar et al., Mol Cell, 2021).



Figure 6: Condensation of ribosomal proteins in ribosome biogenesis. Ribosomal large subunit (LSU) and small subunit (SSU). Polyethylene glycol (PEG). (Unpublished data, Kraushar Lab)





MÜLLER LAB

Cellular Phenotyping

ESTABLISHED IN JANUARY 2017

PROF. DR. MED. FRANZ-JOSEF MÜLLER fjmuellr@mpg.molgen.de +49 176 8426 / 6729 ↘ Müller Lab



Focus areas

- Pluripotent stem cell generation, differentiation and characterization
- Transcriptomics, epigenetics and genome assembly
- Cancer diagnostics
- Mental health
- Nanopore sequencing
- Basic science & translation

Research concept

The Cellular Phenotyping group's core concept follows Max Planck's famous quote: "Insight must precede application." We are working with the world-leading, basic science experimental toolset established at the Max Planck Institute of Molecular Genetics (MPIMG) to generate novel biological insights that have the potential to create paradigm-shifting clinical translation as the salient goal of our work. This approach allows us to work with and on cutting-edge cellular and molecular technologies in stem cell biology, nanopore sequencing, and classification algorithms to better understand and classify biological phenotypes in normal cellular development and malignant transformation. Only such fundamental insights can lead to truly transformative applications in health care.

Development of the world's most widely used bioinformatic assay for pluripotency in human cells as an open-access tool (www.pluritest.org).

Development of hiPSC models and nanopore epigenome sequencing for characterizing genomic "dark matter regions". *Giesselmann et al., Nature Biotechnology, 2019 Tandon et al., Stem Cell Research, 2018 Vögtle et al., The American Journal of Human Genetics, 2018*

Scientific honors

- Professor of Neuropsychiatry at the Christian-Albrechts Universität zu Kiel (since 2022)
- Deputy Director of the Department of Psychiatry and Psychotherapy at the University Hospital Schleswig-Holstein Campus, Kiel, Germany (since 2021)
- Principal Investigator: DFG Cluster of Excellence "Precision Medicine in Chronic Inflammation" Kiel, Germany (since 2019)
- Deputy Director of the Department of Psychosomatics and Psychotherapy at the University Hospital Schleswig-Holstein Campus Kiel (2019-2021)
- Member of the Steering Committee: DFG Cluster of Excellence "Inflammation at Interfaces", Kiel, Germany (2016-2019)

Research projects

The Cellular Phenotyping Group focuses on identifying and utilizing genomewide patterns in transcriptional profiles and epigenetic DNA modifications for preclinical and clinical applications. This mission has led us to integrate methods ranging from patient-specific induced pluripotent stem cells, molecular analysis with Cas9-guided nanopore sequencing, signal processing algorithms, and disease classification based on epigenetic profiles.

Since 2015, we have been working on discovering novel genomic and epigenetic mechanisms in development, health, and disease by utilizing nanopore sequencing as third-generation, long-read sequencing technology. In our initial project starting in 2017 at MPIMG, we have developed novel signal processing and simulation algorithms for raw nanopore sequence traces towards the characterization of "dark matter" genomic regions, which previously were impossible to study and characterize (Giesselmann et al., Nat Biotechnol, 2019; Giesselmann et al., Bioinformatics, 2019). The use of human induced pluripotent stem cell (hiPSC) models of neuropsychiatric disorders derived from patients with pathogenic repeat-expansion allowed us to observe and determine the exact number of repeat expansions at the level of single DNA molecules, as well as the epigenetic modifications of these hitherto "unsequenceable" DNA polymers.

As a next step, we aimed at pinpointing epigenetic patterns unique to cancer genomes with nanopore sequencing for use in a point-of-care setting. In our



Figure 1: Genomic "dark matter" analyzed in hiPSCs with novel algorithmic molecular methods (adapted from Giesselmann et al., Nat Biotechnol, 2019).

ongoing central project IntraEpiGliom, funded by the BMBF with a total volume of 2.8 million euros, we are working together with neurosurgeons, electrical engineers and a diagnostics company on the intraoperative diagnosis of brain tumors based on epigenetic and genetic patterns measured with nanopore sequencing. The basic science conducted in our group at MPIMG provides the foundational platform for potentially paradigm-shifting clinical translation: Genome-wide patterns of CpG methylation from various CNS malignancies are decoded by a novel pattern recognition algorithm (termed MethyLYZR; fig. 2) that can process data streams online from nanopore sequencers. Molecular workflows are established to ultra-rapidly prepare samples and analyze the resulting data stream from nanopore sequencers in real time at the point of care. As a result, we are enabling the classification and assignment of one of 90 diagnoses with a > 95% accuracy to tumor biopsies intraoperatively within 45 minutes from sample-to-answer (fig.2G). This foundational work could lead to neurosurgical intraoperative decision trees informed by whole-epigenome sequencing and machine learning algorithms.



Figure 2: MethyLYZR intraoperative epigenome sequencing as means to inform neurosurgical decisions.

This project has also led us to develop expertise in the area of genome assembly with a particular focus on those "dark matter regions" of the human genome mentioned above.

With the long-term goal of extracting clinically actionable information from *de novo* assembled medical genomes, we have collaborated with conservation biologists at the San Diego Zoo to complete the first chromosome-level genome assembly of the functionally extinct northern white rhinoceros (NWR). This is an instrumental step in the ongoing pluripotent stem cell-based de-extinction efforts of this second-largest land mammal species (fig. 3). The MPIMG's exceptional computational resources with a combination of nanopore long-read sequencing, linked read, HiC, and optical maps. Using BAC mapping and computational methods, we have compared the resulting high-quality genome (fig. 3A) with the horse genome (fig. 3B). Ongoing work aims to achieve the same results with only one high-accuracy long-read sequencing method and to measure and understand epigenetic roadmaps for targeted differentiation towards *in vitro* derivation of fertile gametes from NWR iPSC.



Figure 3: Genome assembly of a functionally extinct species – the northern white rhinoceros. (Photo: Jack Davidson/NYTimes)

Based on our efforts to assemble highly complex genomic regions, we collaborate with Bernhard Herrmann, director of the Department for Developmental Genetics at MPIMG, on the genome assembly of the t-haplotype. In addition, using nanopore sequencing to record dynamic patterns of CpG methylation and our human induced pluripotent stem cell lines, we have worked with several groups at MPIMG supporting the elucidation of the role of CpG methylation and writing and erasing enzymes on endogenous retroviruses in development and disease. The hiPSC cell lines derived by us have been shared with several groups at MPIMG, and access to relevant human material has been helpful in several completed and ongoing studies at MPIMG (Rosebrock et al., Nat Biotechnol, 2022; Wu et al., Nat Commun, 2021).





BULUT-KARSLIOGLU LAB

Stem Cell Chromatin

ESTABLISHED IN JULY 2018

DR. AYDAN BULUT-KARSLIOGLU aydan.karslioglu@molgen.mpg.de +49 30 8413 / 1126 ↘ Bulut-Karslioglu Lab

Sofja Kovalevskaja Research Group



Focus areas

- Developmental and stem cell biology
- Gene-environment interactions
- Epigenetics
- Metabolism
- Developmental timing
- Gene regulation

Research concept

Our lab focuses on mechanisms regulating stem cell state transitions and fate commitment. Specifically, we study how cells communicate signals from their surroundings to the gene expression machinery, especially in the context of cell fate decisions in development. To delineate principles of stemness, we use *in vitro* and *in vivo* models and probe molecular processes critical for the emergence, maintenance, and resolution of stem states in physiological and adverse conditions. Fundamental questions we aim to answer are:

1. How does the early embryo adjust its developmental timing in response to embryonic and maternal signals? (Mechanisms of embryonic diapause)

- 2. How is pluripotency resolved in favorable conditions? (Early events in pluripotency exit)
- 3. How do uterine conditions affect developmental fitness and gene expression programs? (Gene-environment interactions)

Hypoxia induces an early primitive streak signature, enhancing spontaneous elongation and lineage representation in gastruloids. *López-Anguita et al., Development, 2022*

Scientific honors and invited talks

• Sofja Kovalevskaja Award (Humboldt Foundation), 2018

Selected invited talks

- "2nd Crick Beddington Developmental Biology Symposium", London, UK (2022)
- EMBO Workshop "Timing mechanisms linking development and evolution", Barcelona, Spain (2022)
- "4th International Aegean Stem Cell Conference", Kos, Greece (2021)
- "Queen Mary University London New Horizons in Genomics Conference" (virtual, 2020)

Research projects

Harnessing hypoxia as an environmental factor shaping cell fate choices

The cellular microenvironment together with intrinsic regulators shape stem cell identity and differentiation capacity. Mammalian early embryos are exposed to hypoxia in vivo and appear to benefit from hypoxic culture in vitro. Yet, how hypoxia response influences transcriptional networks and lineage choices remains poorly understood. Here, we investigated the molecular effects of acute and prolonged hypoxia on embryonic and extraembryonic stem cells as well as the functional impact on differentiation potential. We find a temporal and cell type-specific transcriptional response, including an early primitive streak signature in hypoxic embryonic stem cells (ESCs) mediated by HIF1a. Using a 3D gastruloid differentiation model, we show that hypoxia-induced T expression enables symmetry breaking and axial elongation in the absence of exogenous WNT activation. When combined with exogenous WNT activation, hypoxia enhances lineage representation in gastruloids, as demonstrated by highly enriched signatures of gut endoderm, notochord, neuromesodermal progenitors and somites. Our findings directly link the microenvironment to stem cell function and provide a rationale, supportive of applying physiological conditions in models of embryo development (López-Anguita N et al., Development, 2022).



Figure 1: Harnessing hypoxia as a lineage-shaping factor that promotes cell diversity and spontaneous elongation in gastruloids.

Adjusting human developmental timing

Many mammals can control the timing of gestation and birth by pausing embryonic development at the blastocyst stage. It is unknown whether the capacity to pause development is conserved across all mammals and, more specifically, in humans. Here we use the inhibition of mTOR to establish a dormant pluripotent state in human naïve ESCs and blastoids. Paused ESCs and blastoids retain pluripotency and, at the same time, cease proliferation. Release from dormancy restores proliferation with full reversibility of the magnitude and pattern of gene activity at the molecular level. Comparison of the pause response in human and mouse stem cells suggests translational regulation, together with metabolic rewiring as pillars of dormancy in both species, while signaling pathways may be discordantly used. Human early embryos actively utilize the mTOR pathway, with a pattern of activity similar to mouse embryos. Thus, our results suggest that the developmental timing of the human embryo may be controllable, with implications for reproductive therapies (lyer et al., in revision Nature).



Figure 2: Fully reversible developmental pausing of human blastoids and pluripotent stem cells.

Metabolic enhancement of mammalian developmental pausing

The quest to model and modulate embryonic development became a recent cornerstone of stem cell and developmental biology. Inhibition of the growth regulator mTOR (mTORi) pauses mouse development *in vitro*, yet the constraints to pause duration are unrecognized. By comparing the embryonic and extraembryonic stem cells' response to mTORi-induced pausing, we identified lipid usage as a bottleneck to developmental pausing. Enhancing

fatty acid oxidation boosts embryo longevity, while blocking it reduces pausing capacity. Genomic and metabolic analysis of single embryos point to a deeper dormant state in enhanced pausing and reveals a link between lipid metabolism and embryo morphology. Our results lift a constraint in embryo survival and suggest that lipid metabolism may be a conserved metabolic transition required to sustain dormancy (van der Weijden V et al., in revision Nature Cell Biology).



Figure 3: Promoting lipid usage via supplementation of the bottleneck metabolite L-carnitine enhances embryo longevity. R, reactivated.

mTOR as a novel histone kinase, regulating exit from pluripotency

mTOR function is widely studied in the context of the regulation of metabolism and protein translation. These regulatory functions are exclusively cytoplasmic, and mTOR has thus been classified as a cytoplasmic serine/threonine kinase. Recently, we discovered a novel function of this well-known kinase in the nucleus. We show that mTOR binds to chromatin in ESCs, but not in differentiated cells. mTOR localizes to promoters of master developmental regulators in ESCs, such as Hox, Fox and Sox family members. Through *in vitro* kinase assays, we identify mTOR catalytic activity specifically towards histone H2A and particularly at H2AT120p. This is a very interesting finding in terms of regulation of pluripotency due to the fact that developmental genes are repressed in ESCs by concerted action of Polycomb complexes, which catalyze histone H2AK119 ubiquitination and H3K27me3. We are currently investigating the interplay between H2AT120p, H2AK119ub, and H3K27me3 at the onset of pluripotency exit. Our preliminary data suggest that nuclear mTOR function provides a direct connection between the environment and the progress of development.



Figure 4: Identification of mTOR as a novel histone kinase, regulating pluripotency exit by counteracting Polycomb repression.

Other ongoing and future work

Currently, we work on several other projects dissecting the mechanisms of developmental pausing, particularly focusing on miRNA-protein networks, gene regulation via cytoplasmic and nuclear bodies and chromatin. Future work will focus on dissecting transcriptional vs. post-transcriptional gene regulation and how reversible and dynamic gene repression is achieved in pluripotent cells.





ELKABETZ LAB

Human Brain and Neural Stem Cell Studies

ESTABLISHED IN MAY 2017

DR. YECHIEL ELKABETZ elkabetz@molgen.mpg.de +49 30 8413 / 1886 ↘ Elkabetz Lab

Research Group



Focus areas

- Neural stem cell (NSC) patterning, heterogeneity and potential
- Cerebral organogenesis in a dish
- Programming and reprogramming towards NSC fates

Research concept

A major goal of the Elkabetz lab is to develop strategies for generating homogeneous sources of cultured neural stem cell (NSC) types and their progeny in 2D and 3D organoid systems. Our main strategy is the identification, isolation, characterization and manipulation of distinct NSC types in culture. We first devise transgenic and marker-based paradigms for sorting, separation and isolation of distinct cell types. We then use bulk and single cell RNA-Seq, immunostainings and stem cell techniques in 2D and 3D cultures to molecularly characterize and assess the fate potential of these cell types. Finally, we use genome editing, reporters and gain/loss of function studies in order to manipulate these cell types. Ultimately, we aim to identify novel brain stem cells and utilize them for therapeutic approaches.

Developing advanced cortical organoids in the dish. Rosebrock, Arora et al., Nat. Cell. Biol. 2022

Selected invited talks

- Oral presentation: Frontiers In Stem Cell Biology: From Embryogenesis to Tissue Maintenance and Regeneration, Dead Sea, Israel (2022)
- 3rd International Conference on Stem Cells (Aegean Conferences), Crete, Greece (2019, no show due to birth of my twins)
- Oral presentation: 16th Annual Meeting of the International Society for Stem Cell Research (ISSCR), Melbourne, Australia (2018)

Research projects

Enhanced cortical neural stem cell identity through short SMAD and WNT inhibition in human cerebral organoids facilitates emergence of outer radial glial cells

(In collaboration with the labs of Vingron, Arndt and Müller and Tel Aviv University, Israel)

Despite the tremendous upsurge in derivation methods, cerebral organoids exhibit a broad regional heterogeneity accompanied by limited cortical cellular diversity suggesting inadequate patterning of early NSCs. Here we show that a short and early dual SMAD and WNT inhibition course is necessary and sufficient to establish robust and lasting cortical organoid NSC identity, efficiently suppressing non-cortical NSC fates. In contrast, other widely used methods are inconsistent in their cortical NSC-specification capacity. Accordingly, our method selectively enriches outer radial glia NSCs, which cyto-architecturally demarcate well-defined outer sub-ventricular-like regions and propagate from superiorly radially organized, apical cortical rosette NSCs. Finally, our approach culminates in the emergence of molecularly distinct deep and upper cortical layer neurons, and reliably uncovers cortex-specific microcephaly defects. Thus, a short SMAD and WNT inhibition is critical for establishing a rich cortical cell repertoire. This enables us to mirror fundamental molecular and cyto-architectural features of cortical development and establish meaningful disease models (Rosebrock, Arora et al. Nature Cell Biol, 2022).





Figure 1: scRNASeq of Day 50 Organoids iPSC Lines Reveals Consistent Cortical Specification in Triple-i protocol and outer radial glia NSC subpopulation.

Programming and reprogramming NSC fates: specification of hippocampal and choroid plexus fates in cortical organoids by micro-RNAs

(In collaboration with Markus Hafner, NIH)

While NSCs are the founder populations of different brain regions, the information about the early signals that modulate these NSCs to be specified to different brain regions is still limited. In this study, we profiled the expression levels of micro-RNAs across different stages of NSC development during corticogenesis in 2D rosettes and 3D organoid cultures. We hypothesized that early specification of cortical NSCs may involve the concerted action of multiple factors governed through early stage micro-RNAs (miRs). Using a battery of molecular and cellular studies, we show that early expressed miRs are important in the specification of the archicortex. We find hsa-miR-20b-5p (miR-20b) to be one of the highly expressed micro-RNAs in early cortical NSCs in rosette sand organoids. Overexpressing hsa-miR-20b during early cortical organogenesis directs widespread differentiation of organoids towards choroid plexus cells via enhancing WNT and BMP signaling. Strikingly, when overexpressed together with its main target gene *CCND1*, this leads to dramatic cell fate transition towards hippocampal NSCs and neurons within organoids. Thus, we propose that early cortical NSCs transitioning uses a short developmental exposure to miR-20b in order to generate the cortical hem and its derivatives – the hippocampal primordium and the choroid plexus.



Figure 2: Caudalization of cortical organoids towards hippocampal and choroid plexus fates by micro-RNA20b. a) Immunostaining of PAX6/FOXG1 shows loss of cortical vesicles upon miR-20b expression. Immunostaining of SOX2/TTR and SOX2/LMX1A shows specific elevation of choroid plexus and hippocampal markers upon mir-20b expression. UMAP plot for scRNA-Seq data of day 50 wild type (b) and miR-20b expressing (c) organoids shows the expansion of caudal fates, and hippocampal and choroid plexus fates in particular. d) Bar plot-showing percentage of cells in wild type and miR-20b overexpressing organoids for different cell types.

Programming and reprogramming NSC fates: identification and characterization of early cortical NSC markers

One of the major goals in our lab is to develop novel strategies to induce and maintain homogeneous, self-renewing founder cortical NSCs. Such endeavour entails i) the identification of novel markers for early cortical NSCs, ii) using these markers for isolation and characterization of novel NSCs and iii) screening for factors that maintain these cells in culture. By profiling 2D rosettes and 3D organoid cultures at different stages of development and using bulk and single cell RNA-Seq, we identified markers expressed in early stage cortical NSCs. These were validated by immunostaining in rosettes and organoids and by gene expression analysis of FACS-sorted cells. These analyses confirmed that such markers prospectively enrich in early cortical NSC populations in culture. In parallel, we have been developing a screening system to identify transcription factors (TFs) that are involved in the establishment of cortical NSCs. We plan to use the novel markers as readout in this screening system – towards the identification of factors that accelerate, induce or maintain early cortical NSCs.



Figure 3: Identification and characterization of novel early cortical NSC markers. a. Scheme for deriving and harvesting cortical cells. b. scRNA-Seq through 3 developmental time points allows identification of two markers. c. Validation in immunostaining confirms early transient marker expression. d. FACS charts and stack bars confirm transient expression and colocalization of both early markers. Isolated double positive cells display cortical NSC morphology.



THE MAX PLANCK INSTITUTE FOR MOLECULAR GENETICS - RESEARCH REPORT 2022 | 61



HNISZ LAB

Precision Gene Control

ESTABLISHED IN APRIL 2018

DR. DENES HNISZ hnisz@molgen.mpg.de +49 30 8413 1876 ↘ Hnisz Lab

Research Group



Focus areas

- Transcription biology
- Biomolecular condensates
- Transcriptional dysregulation
- Interdisciplinary transcription chemical biology, physics, human genetics, cancer

Research concept

The mission of the Hnisz lab is to discover the principles that underlie the control of transcriptional programs during development and disease. We recently proposed a model that transcriptional regulatory proteins form phase-separated 'transcriptional condensates' that may play important roles in the control of cell identity. The central theme of the lab is to use the transcriptional condensate model to solve major outstanding problems in transcription-, stem cell- and disease biology. Key ongoing and future directions include understanding the molecular basis of condensate formation, investigating the regulatory functions of transcriptional condensates, elucidating their roles in various developmental and disease contexts, and drugging condensates.

Discovered the role of endogenous retroviral RNA as a chemoattractant that controls condensate distribution in embryonic cells. *Asimi*, Sampath Kumar*, Niskanen*, Riemenschnieder* et al, Nature Genetics, 2022*

Predicted the molecular basis of phase separation in human transcription factors. *Basu*, Mackowiak* et al., Cell, 2020*

Provided evidence for changes in condensate composition as a pathomechanism in human genetic disease. Basu*, Mackowiak* et al., Cell, 2020

Scientific honors and invited talks

• Top 1% Highly Cited Researcher by the Web of Science (2019-2021)

Selected invited talks

- "2nd Annual Meeting on Advances on Nuclear Topology and 3D Chromatin Architecture in Cancer", New York, USA (2021)
- "SFB850 International Symposium", Freiburg, Germany (2019)
- "Forbeck Forum on 3D chromosomal architecture and nuclear topology", San Diego, USA (2019)
- "Clinical Translation of Epigenetics in Cancer Therapy", Litchfield Park, USA (2019)

Research projects

Molecular grammar of phase separation in human transcription factors

(In collaboration with the Ibrahim, Mundlos and Meissner labs)

One key focus area in the lab is discovering the molecular features of phase separation (also referred to as "molecular grammar") encoded in human transcription factors and other regulatory proteins. In an early pivotal study, we predicted such molecular features and provided evidence that disease-associated mutations in humans alter those features. In a case study of human synpolydactyly, we demonstrated that mutations in the sequence that drive phase separation of the HOX13 transcription factor led to changes in the composition of biomolecular condensates formed by HOXD13. This work provided a foundation to further investigate the molecular basis of phase separation and new ways of understanding the function of disease-associated variation in humans.

(Basu*, Mackowiak* et al., Cell, 2020)



2x Radius of gyration (nm)

Figure 1: HOXD13 condensates in mouse embryonic limb bud cells. Images were taken with Stochastic Optical Reconstruction Microscopy (STORM). The zoomed-in area on the right is highlighted with a red box on the left.

Biology of transcriptional condensates

(In collaboration with the Meissner lab)

Much evidence suggests that transcriptional regulatory proteins form or partition into nuclear condensates, but the components of these transcriptional condensates are unknown and their biological importance is debated. We have shown that the distribution and function of transcriptional condensates in embryonic cells is partially controlled by RNA species produced by endogenous retroviruses (ERVs). We found that when ERVs are transcribed, transcriptional condensates dissociate from super-enhancers and associate with ERV loci, in a manner dependent on ERV RNAs. This work has opened new avenues for understanding how transposable elements affect cellular states and for understanding the role of numerous non-coding nascent RNAs whose functions have been unknown so far (Asimi*, Sampath Kumar*, Niskanen*, Riemenschneider* et al., Nature Genetics 2022).



Figure 2: Co-localization between the IAP ERV RNA and RNAPII puncta in TRIM28-degraded mESCs that transcribe ERVs. Displayed are separate images of individual z-slices of the RNA-FISH and IF signal, and an image of the merged channels. The nuclear periphery determined by DAPI staining (not shown) is highlighted as a white contour. Scale bar: 2.5µm.

Suboptimization of human transcription factors

(In collaboration with the Vingron lab)

Specificity in transcriptional control is engendered by transcription factors (TFs) binding enhancer DNA elements. Weak enhancers are suboptimal for TF binding motif sequence and spacing, suggesting an evolutionary trade-off between transcriptional activity and specificity. We are pursuing and idea that a similar trade-off is encoded in human transcription factors as submaximal spacing between aromatic residues in their disordered regions. If correct, this idea may be the foundation of a universal strategy to facilitate TF-mediated cellular reprogramming and to enhance or reduce the function of any biomolecule.

Disruption of phase separation as a pathomechanism in human disease

(In collaboration with the Mundlos, Kalscheuer, Kraushar and Spielmann labs)

Thousands of genetic variants in protein-coding genes have been linked to disease. As they mostly occur within intrinsically disordered protein regions that have poorly defined functions, the functional impact of most variants is unknown. Intrinsically disordered regions were recently found to mediate formation of phase-separated biomolecular condensates, suggesting that mutations in disordered proteins may alter condensate properties and function. We are creating and testing models of pathomechanisms to study how disease-associated variants in disordered proteins may contribute to transcriptional dysregulation in human genetics diseases and cancer.

Targeting of oncoprotein condensates with small molecules

Virtually all biopolymers can form condensates, but specific chemical tools to modulate the function of specific biomolecular condensates are lacking. In a collaboration with the Xavier Salvatella lab at the IRB Barcelona, we are developing small molecules that target the phase separation capacity of the androgen receptor, and are investigating the use of such compounds as prostate cancer therapeutics in prostate cancer models.





Department of Computational Molecular Biology

ESTABLISHED IN OCTOBER 2000

PROF. DR. MARTIN VINGRON vingron@molgen.mpg.de +49 30 8413 / 11500 ン Vingron Department SECRETARY Martina Lorse vinoffic@molgen.mpg.de +49 30 8413 / 1151

Structure and organization of the department

The department 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 gene regulation and epigenetics, pursues collaborative projects with experimental groups, and develops methods for the analysis of sequencing data in the context of genetics and cancer genomics. Until 2020, Sebastiaan Meijsing headed an experimental group on regulation and epigenetics, mainly studying glucocorticoid receptor as a model transcription factor. This group has been succeeded by the group headed by Sarah Kinkley, which pursues research on chromatin structure and genome stability. Peter Arndt heads the Evolutionary Genomics Group which works on developing models how the DNA in humans and other organisms has evolved. Ralf Herwig's group concentrates on disease bioinformatics and biological networks.

Research concept of the department

The Department of Computational Molecular 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. The overarching topic of the department is regulatory genomics, with the long-term goal of deciphering the regulatory wiring of cellular processes.



Ongoing work and planned developments

Ongoing work in the department comprises both, the development of methods and tightly integrated experimental collaborations. On the side of the methods development, we focus on the analysis of complex data, AI applications, and evolutionary questions. New technologies like single cell RNA sequencing pose new challenges, for which we import existing tools where possible and develop new methods where we feel it is needed. Today, many collaboration projects also involve single cell data (e.g., Yechiel Elkabetz, Meissner department). Chromatin structure has been studied in the context of gene regulation and is now the focus of the work of Sarah Kinkley. Evolutionary questions like the ones studied by Peter Arndt find applications, e.g., in tumor evolution and cancer genomics and also play a prominent role in the work of Ralf Herwig.

The ties to the research group of Stefan Mundlos had started with the analysis of HiC data but a common interest in the regulatory role of human genetic variants has increasingly taken a central role. As a consequence, we have made an effort to analyze human genomic sequencing data and to search for structural variations.

Equipment and resources

The theoretical work of the Computational Molecular Biology Department relies heavily on powerful computers. The MPI's computer system is shared by all users, including the Computational Molecular Biology Department, and it is maintained by the IT group. The central computer equipment comprises several compute servers with 48 or 64 processors, 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. Some of the processors are meant for interactive use, while others are meant to be accessed through a queuing system. This allows for very efficient utilization of the entire infrastructure. Storage space on hard disks in the institute comprises approximately 7 PB and the department participates in this. Recently we have also acquired a small GPU cluster. 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. Sarah Kinkley has office space on the first floor of Tower 3 and her group's laboratory is located in Tower 4.



VINGRON LAB

Transcriptional Regulation

ESTABLISHED IN OCTOBER 2000

PROF. DR. MARTIN VINGRON vingron@molgen.mpg.de +49 30 8413 / 1150 ン Vingron Lab



Focus areas

- Transcriptional regulation
- Epigenetics
- Genetic variation
- Data analysis methods
- Machine learning

Research concept

Martin Vingron's gene regulation group pursues methods development and collaborative data analysis in the fields of transcription factor-based and epigenetic gene regulation, and works on analyzing 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 data analysis and visualization, and apply statistical and machine learning methods to a variety of data analysis and data integration tasks.

Collaborative analysis of genomic and regulatory evolution of the genome of the Spanish mole, together with Mundlos group. *Real et al., Science, 2020*

Development of the CRUP method for predicting enhancers active in a particular cell type, given ChIP-seq data for only a small number of histone modifications. *Ramisch et al., Genome Biology, 2019*

Development of the SVIM algorithm to detect structural variations in genomic data from PacBio reads. Heller et al., Bioinformatics 2019

Selected invited talk

- "USC Computational Biology Symposium", Los Angeles (2022)
- o "CRG Barcelona", Barcelona, Spain (virtual, 2021)
- "Minerva-Gentner Symposium" (virtual, 2020)
- Keynote lecture at "RECOMB", Washington D.C., USA (2019)
- "Distinguished Speaker Colloquium", Zentrum für Bioinformatik, Universität Saarbrücken, Saarbrücken, Germany (2018)

Research projects

Recognition of enhancers and epigenetic regulation

Active enhancers are characterized by particular histone modification patterns. Therefore, cell-type specific epigenetic data can be used to predict active enhancers. We approached this prediction problem using statistical and machine learning methods, developing a number of different algorithms for different aspects of the problem. Utilizing Hidden Markov Models, we first developed an algorithm, eHMM, that delineates very clear and distinct active enhancer regions in a genome. An extension of this approach, modHMM, produces a comprehensive annotation of a genome with enhancers, promoters, transcribed regions, etc., based on given ChIP-seg and RNA-seg data. In a parallel effort, we used a random forest algorithm to develop the CRUP (cis-regulatory unit prediction) method which takes ChIP-seq data for only three different histone modifications (H3K27ac, H3K4me1, H3K4me3) for a tissue or cell-type as inputs and predicts active enhancers. This method has become the workhorse of our enhancer annotation efforts. It has been applied in several other projects, like the work of Sebastiaan Meijsing and in regulatory annotation of other organisms, e.g., in the mole project. Currently, we are working on predicting which enhancers interact with a given promoter and we are analyzing ATAC-seq data (chromatin accessibility) data to back up the enhancer predictions (Ramisch et al., Genome Biol., 2019).



Figure 1: This figure shows the CRUP-predicted development of the enhancer landscape during mouse midbrain development (experimental data from Gorkin et al, 2017).

Comparing experimentally derived enhancers with computational predictions

Sebastiaan Meijsing and his group have contributed experimental work on the characterization of enhancers. The associated data were analyzed, shedding light on the relationship between experimental enhancer determination (FAIRE-STARR-seq) and the computational predictions by CRUP. The computational predictions from ChIP-seq data can, to a certain degree, make up for the episomal setting (i.e., separate from the chromosomal environment) of the STARR-seq experiment. This data set also allowed to study whether enhancer activity in one particular cell type can be predicted from the sequence. In fact, this is to a certain degree possible (Glaser et al., Nucleic Acids Res, 2021).

Is there a tissue-specificity code in the sequence?

The question that is even harder than predicting the activity of a regulatory element from its sequence, is whether cell type specificity is encoded in the sequence. This implies that one should be able to tell from the sequence that an element would be active in, say, liver, but not in the heart. To tackle this question, Philipp Benner developed specialized machine learning tools for the extraction of patterns from sequence and used modHMM to put together a data set of enhancers that were active in specific tissues. We discovered that even these difficult questions can be solved with some success for particular cell types, while for other cell types it remains a challenge to predict whether an enhancer would be active in particularly that tissue (Benner and Vingron, NAR Genom and Bioinform, 2021).



Figure 2: The figure shows precision-recall curves when trying to predict in which tissue a regulatory element might be active. While for some tissues, e.g. liver, this is doable with reasonable success, the results are near random for hindbrain.

Sequence analysis and genomic variation

Traditionally, human genetics has had a focus on protein coding variants in the genome. With a generally better understanding of gene regulation and epigenetics, also mutations in the non-coding part of the genome have become interesting because they might lead to alterations in the regulatory wiring of the cell. Together with Stefan Mundlos and his group we started an initiative to determine variations, in particular structural variations, in individual genomes. At the same time, we develop methods to predict whether such variations are likely to affect gene regulation (Heller and Vingron, Bioinformatics, 2019; Hertzberg et al., Genome Biol, 2022).

Reference		Locus 1	Locus 2a	Locus 2b
Reference 🐋 🐋	ndividual		**	
	Reference	→→	*	*

Figure 3: How to determine structural variations from the problems when mapping a read? The figure shows read signatures for an interspersed duplication and a novel element insertion.

Single cell data analysis and determination of marker genes

Together with the group of Yechiel Elkabetz, we started to analyze gene expression data for brain organoids, generated by his laboratory. Initially, they produced bulk RNA-seq data and then quickly proceeded to single cell RNA-seq data. The task was to determine the cell types that had developed in an organoid. While the respective paper is meanwhile published (Nature Cell Biology, 2022), the work has spawned new methodological developments in our group. First, we developed a new method, Association Plots, for visualizing the genes that are associated to a particular cluster of cells. The genes are represented by dots, and the further to the right a gene lies, the stronger is its association with a cluster. While an Association Plot focusses on one cell cluster at a time, the underlying ideas can also form the basis for new clustering algorithms that combine clustering and marker gene definition. This is the current focus of development (Gralinska et al., J Mol Biol, 2022).



Figure 4: This example of an Association Plot shows genes associated to a cluster of B-cells and the genes highlighted in red are known B-cell marker genes.



Following our earlier work on the construction of gene regulatory networks from expression data, we also investigated the question how to utilize single cell RNA-seq data for this task. Due to the sparsity of the single cell data, it is common practice to apply imputation methods to make up for missing values. We studied the effect of these methods and uncovered a hitherto unrecognized trap: The imputation methods themselves tend to exploit correlations among the genes, and when applying the network reconstruction after imputation, there is a clear tendency to rediscover the ill-founded assumptions of the imputation method (Ly and Vingron, Patterns, 2022).

Detection of phenotype-related genomic changes

Phenotypic differences across species are often the result of changes in the set of genes available, or in the regulatory landscape affecting phenotype-related genes. Genomic adaptations may range from adjustments of single enhancers to the emergence of alternative TAD structures, driven by genomic rearrangements. Thus, allowing the activation of sets of enhancers in a new regulatory and functional context.



Figure 5: Breaks of synteny predicted via pair-wise genome comparisons. Species from left to right: human, shrew, mole, mouse.

Assuming that a complex, clade-specific feature evolved only once, we screened the genomes of two moles characterized by the development of an ovotestis for shared genomic rearrangements (breaks of synteny) that are not existent in other species without ovotestis (Fig. 5). Although the quality of genome assemblies strongly affects the number of predicted rearrangements in pair-wise genome comparisons, the approach of filtering for shared events for species of the same phenotype greatly reduced the false-positive rate. Further filtering of the 286 candidate breaks of synteny by the function of nearby genes, revealed a major rearrangement of the TAD comprising FGF9 (Fig. 6), a master regulator of gonad development. To pinpoint potential enhancers related to the regulatory network of ovotestis formation, we predicted enhancers from ChIPseq and ATACseq data, now being refined by our new software CRUP (M Real et al., Science, 2020).


Figure 6: Genomic rearrangements specific to ovotestis phenotype at the FGF9 locus. Blue and green boxes connected by lines refer to sequences in human and mouse, respectively. A red arrow indicates the break of synteny.

Recently, we started screening 9 bat genomes for breaks of synteny potentially associated to wing development, yielding 64 candidate regions. Here, with experimental data available for at least two bat species, we aim to identify enhancers conserved clade-specifically. Applying the same strategy of grouping results by phenotype will help to discover distinct enhancers functionally linked to the respective phenotype.







HERWIG LAB

Bioinformatics

• ESTABLISHED IN FEBRUARY 2001

DR. RALF HERWIG herwig@molgen.mpg.de +49 30 8413 1587 ↘ Herwig Lab



Focus areas

- Bioinformatics
- Genome analysis
- Network biology
- Biomedical applications

Research concept

Our research covers the development of computational methods i) for the analysis of molecular data with the focus on human diseases, in particular cancer and type-2 diabetes and ii) for integration of these data in the context of biological networks. The group has developed a novel network propagation framework and applied this to the identification of drug toxicity mechanisms from multiomics data and to the biological explanation of "black-box" machine learning predictions for the survival of cancer patients. Molecular interactions have been substantially updated within our ConsensusPathDB resource (>2,100 citations so far). Additionally, we have developed a software package for long-read transcriptome sequencing and analyzed the effect of aberrant splicing regulation in cancer. We have published 25 scientific publications in the reporting period.

Scientific highlights

ConsensusPathDB – integrative resource for molecular interactions. *Kamburov and Herwig, Nucleic Acids Research, 2022*

Pan-cancer survival prediction model. *Thedinga and Herwig, iScience, 2022*

Modifiers of intestinal tumor formation. Farrall et al., Cancer Research, 2021

NetCore – novel robust method for network propagation. Barel and Herwig, Nucleic Acids Research, 2020 Selevsek et al., Communications Biology, 2020

Selected invited talks

- Tutorial and talk at the "21st European Conference on Computational Biology (ECCB)", Sitges Barcelona, Spain (2022)
- Project presentation at the "All-Hands Meeting of the Al Initiative of the Federal Ministry for Education and Research", Dortmund, Germany (2019)
- "11th World Congress on Alternatives and Animal Use in the Life Sciences", Maastricht, Netherlands (2021)
- "12th annual RECOMB/ISCB Conference on Regulatory & Systems Genomics", New York, USA (2019)
- "PROTEOMIC FORUM 2019 XIII. Annual Congress of the European Proteomics Association", Potsdam, Germany (2019)

Research projects

A novel framework for network genome analysis

The Herwig group has contributed resources and methods for the network-based analysis of genome data. Since 2009 we continuously provide the Consensus-PathDB resource, a meta-database of human (as well as mouse and yeast) molecular interactions integrated from 33 public resources.

This database has recently been substantially updated with content and novel interaction types and currently comprises 859,848 interactions of different types for 200,499 different physical entities such as proteins, drugs, metabolites etc. (Kamburov and Herwig, 2022). In particular, the integrated protein-protein interaction (PPI) network has been used as a scaffold for genome analysis, and we have developed a mathematical framework, called NetCore, based on random walk with restart to analyze experimental data from multiple sources at the network level (fig. 1). Instead of node degree, the algorithm utilizes node core for the re-ranking procedure that is specifically robust against experimental bias of PPI experiments (Barel and Herwig, 2020).



Figure 1: Network propagation framework.

Such network propagation has been applied in the context of multi-omics drug toxicity, exploring toxicity mechanisms and clinically relevant novel biomarkers (fig. 2A-B; Selevsek et al., 2020).

Ongoing projects include the network-based analysis of phosphorylation data to study the temporal response to insulin treatment in human muscle in cooperation with the German Diabetes Center in Düsseldorf as well as the computation of network-derived drug safety indices for the improvement of drug development in a joint BMBF-funded project with the University of Maastricht.



Figure 2: A) Integrated network module identified from 3D cardiac microtissues treated with four anti-cancer drugs. B) Protein expression of network-derived biomarkers in cardiac biopsies of patients (Y-axes) correlates with left-ventricular ejection fraction (LVEF; X-axes).

Enhancing biological plausibility for machine learning methods in biomedicine

The Herwig group participates in the AI initiative of the Federal Ministry of Education and Research (BMBF) and has coordinated a project to improve machine learning for biomedical applications by the incorporation of biological background knowledge and the exploitation of methods that enable *post hoc* interpretability of machine learning methods. A major application domain is the survival risk prediction for cancer patients with molecular data. We have developed an approach, *"Predict and Propagate"*, that is based on ensemble tree learning with XGBoost and subsequent network propagation of the learned features with NetCore that generates plausible network modules from otherwise non-interpretable machine learning methods (fig. 3; Thedinga and Herwig, 2021, 2022). In particular, we found that the tumor microenvironment is highly predictive for pan-cancer patient survival.



Figure 3: Predict and Propagate – workflow for cancer patient survival prediction and identification of biologically plausible features.

A second application domain is the prediction of drug sensitivity from the molecular characterization of the cellular system and chemical descriptions of the drug under study. In cooperation with machine learning experts at the University of Potsdam, we developed a method based on the minimization of ranking loss for drug sensitivity predictions (Prasse et al., 2022a), which outperforms existing deep neural network approaches. Currently, we work on transfer learning protocols and explore deep neural network architectures. This allows us to use the large body of *in vitro* drug sensitivity data for pre-training and subsequently use these pre-trained models for the prediction of drug sensitivity in pre-clinical systems, such as PDXs, PDOs and *ex vivo* patient tissues, where training data is typically of much smaller size (Prasse et al., 2022b). With the developed methods we have contributed to two international machine learning challenges, the BEAT-AML challenge on drug sensitivity predictions for AML patients and the IDG-DREAM challenge on binding properties of kinase inhibitors (Cichonska et al., 2021).

Human disease projects

The group has applied its previously developed workflows for RNA-seq and MeDIP-seq to project collaborations in the field of cancer and type-2 diabetes with partners within and

outside the institute. With the Herrmann department, we have conducted a fundamental cancer genetics study based on consomic mouse strains to identify modifiers of intestinal tumor formation. We have identified candidate gene regions and biomarkers on mouse chromosome 5 that determine tumor multiplicity in the presence of cancerous APC mutations (fig. 4; Farrall et al., 2021).



Figure 4: A) Genotype clusters of mice (1–18; X-axis; only clusters with with > 5 mice) derived from SNP markers on chromosome 5 and tumor multiplicity phenotypes (Y-axis). Color gradient reflects different proportions of B6 (blue) and PWD genotypes (green). B) Pairwise Mann-Whitney tests of the tumor multiplicity across the different clusters.

With collaborators at the University Cologne, we have analyzed differential methylation- and gene-expression patterns in lung cancer patients with resistance and susceptibility, respectively, towards carboplatin therapy. We identified a novel tumor suppressor, LRP12, that is promoter-hypermethylated upon drug resistance. The novel resistance marker has been identified in PDX models of lung tumors and was validated in a retrospective clinical cohort.

With partners at the German Diabetes Center Düsseldorf we analyzed muscle tissues in knock-out mice related to glucose uptake (TBC1D14; submitted) and epigenetic modifications of insulin action (HDAC5; Klymenko et al., 2020).

Long-read transcriptome sequencing (LRTS)

The Herwig group has developed the software package IsoTools, a complete pipeline for mapping, annotation and statistical analysis of PacBio Iso-Seq® experiments. The package comprises data quality control steps, isoform identification and quantification, identification of (coordinated) splicing events, as well as statistical tests for differential splicing (Lienhard et al., 2021).



Figure 5: Splicing regulation in CLL patients. A) Hot-spot mutations in SF3B1 lead to 3' alternative splicing in the MAP3K7 gene (vertical line). B) iCLIP shows double-peak RNA-binding of SF3B1 around -50 and -10nt upstream of the canonical 3' splice site.

In a project funded by the DFG we, in cooperation with groups at the Universities in Cologne and Frankfurt, applied IsoTools to investigate aberrant splicing induced by mutations in the splicing factor SF3B1 in chronic lymphocytic leukemia (CLL) patients and patients with myel-odysplastic syndromes (MDS). We observed that SF3B1 hot-spot mutations specifically affect 3' alternative splicing (fig. 5A) and, using iCLIP data, we identified an RNA binding signature that suggests a double peak at -50 and -10 nt upstream of the 3' splice site (fig. 5B). This might have effects on the binding capabilities of SF3B1 as well as on branchpoint and 3' splice site recognition. Ongoing work includes the experimental validation of these findings.

Additionally, with IsoTools we participated in LRGASP, an international challenge on long-read transcriptome isoform identification and quantification organized by the Gencode consortium. Dissemination was strengthened through joint tutorial sessions at ECCB 2022.

Future projects will cover the extension of the analysis framework to single-cell protocols and the development of refined patient-specific isoform-landscapes. This will be realized by the involvement of the group in the European-wide Marie Sklodowska-Curie action LongTREC on LRTS and its applications in biomedicine, starting in 2023.





KINKLEY LAB

Chromatin Structure and Function

ESTABLISHED IN NOVEMBER 2019

DR. SARAH DOERDRECHTER-KINKLEY kinkley@molgen.mpg.de +49 30 8413 1875 뇌 Kinkley Lab



Focus areas

- Chromatin structure and function: R-loops, higher chromatin order, 3D chromatin architecture, cohesin
- Epigenetics
- o Mitosis

Research concept

We are interested in the mechanisms of higher chromatin order and the impact of chromatin structure on genome stability. Specifically, we are studying the interplay between PHF13 and cohesin, and how these proteins regulate interphase and mitotic chromatin architecture. In addition, we are investigating the role of R-loops as a potential oncogenic precursor state. While R-loops are well known to drive genome instability, high throughput studies exploring their incidence following perturbation of oncogenes/tumor suppressors are missing, due in part to a lack of suitable tools. We are working on developing live cell R-loop sensors to facilitate high throughput studies to understand the incidence of R-loops and if they are a common oncogenic precursor state - targetable across cancer types.

Scientific highlights

Connecting the Dots: PHF13 and cohesin promote polymer-polymer phase separation of chromatin into chromosomes. *Rossi et al., BioRxiv, 2022*

Selected invited talks

- "Forum Experiment", Hannover, Germany (2022)
- "Molecular Mechanisms of Phase Separation", Northeim, Germany (2022)

Research projects

PHF13 and cohesin initiate chromosome condensation

How interphase chromatin is condensed into mitotic chromosomes remains incompletely understood. The condensins drive this process from pro-meta-phase on, however, what initiates this process at prophase remains unclear. Our research shows that PHF13, a tightly regulated H3K4me epigenetic reader, and cohesin, a mediator of chromatin architecture, cooperate to drive higher order compaction of chromatin into chromosomes during early prophase. Mechanistically, PHF13 stabilizes cohesin chromatin interactions and it oligomerizes via direct N- and indirect C- terminal dimerization. This results in a polymer with increased chromatin avidity and the ability to bridge neighboring nucleosomes in a linear and distant manner via its coupling to cohesin (e.g., SMC1 and SMC3). This results in a two-step condensation process of chromatin into chromosomes.



Figure 1: A) IF: Endogenous co-localization. B) IF: Transient PHF13 chromosome condensation. C) Condensation model of PHF13 and cohesion (adapted from Rossi et al., BioRxiv, 2022)

Future Directions: We are generating PHF13 degron cell lines to explore the impact of PHF13 on cohesin chromatin association in a time-resolved manner. To this end, we are optimizing Cut and Tag experiments to look at the genome-wide distribution of cohesin (SMC1, SMC3 and RAD21) following PHF13 depletion or

induction in Dld-1 cells. In addition, Hi-C experiments are planned under the same conditions to look at the impact of PHF13 on 3D chromatin architecture. Finally, we have endogenously tagged PHF13 with Emerald in Dld-1 cells and we have received endogenously tagged EGFP-RAD-21, STAG2, NIPBL and WAPL cell lines from the lab of Jan-Michael Peters (IMP, Vienna, Austria) and endogenously Flag and Streptavidin tagged SMC3, RAD21, PDS5A and NIPBL from the lab of Todd Waldman (Georgetown University, Washington D.C., USA). Using these cell lines, we want to elucidate the molecular context of the PHF13-cohesin interactions. In collaboration with Axel Imhof at the Biomedical Center at LMU (Munich, Germany), we are establishing proximity biotinylation followed by mass spectrometry for PHF13. Additionally, using a tandem affinity purification approach we will enrich PHF13 cohesin complexes and use chemical cross-linking to identify the exact cohesin-PHF13 interaction interface. Finally, our dream experiment would be to perform cryo-EM on these complexes if we can enrich sufficient quantities.

Phase transitions govern higher order chromatin transitions

We have made the observations that PHF13 is capable of undergoing phase separation (PS) and phase transitions from a liquid-like state to a polymer-like state. We have mapped the relevant domains, to either its IDR domains which promotes liquid-liquid phase separation (LLPS) or its ordered domains for polymer-polymere phase separation (PPPS), respectively. Interestingly, the decision between LLPS or PPPS states influences its impact on chromatin structure, suggesting an interplay between these forms of phase separation with a physiological impact on higher chromatin order.



Figure 2: Schematic of PHF13 (A). Deletion of ordered regions (B) promotes LLPS condensates (C) and IDR mediated oligomerization (D).



Figure 3: Deletion of IDRs promotes PPPS and condensation. A) In silico contact map. B) IF: ΔPEST1 and ΔPEST2.

Future Directions: We are interested in understanding the "molecular grammar" underlying these phase transitions at a physiological level. The IDR1 and IDR3 overlap with the PEST domains of the PHF13 protein which are bookmarked by sumoylation sites and contain several CDK1 and GSK3β phosphorylation sites, which may act as molecular cues in the decision making between an LLPS and PPPS state. Furthermore, our preliminary findings suggest that these domains and molecular cues are utilized in other chromatin effector proteins. We aim to explore the relevance of phase transitions in chromatin factors and impact on higher chromatin order. We are collaborating with Martin Vingron and Gözde Kibar on this project.

Incidence of R-loops formation following perturbation of chromatin regulatory proteins

Cancer is believed to be many different diseases, characterized by distinct sets of driver mutations. Much emphasis in cancer research has been given to understanding the direct yet distinct consequences of driver mutations without understanding the commonality between them. Chromosome instability (CIN) and structural and numerical chromosomal abnormalities, are a hallmark of almost all human cancers. Therefore, understanding how distinct driver mutations lead to a state of chromosomal instability does provide a potential therapeutic window that may be applicable across cancer types. Driver mutations in epigenetic regulators, DNA and RNA metabolic enzymes, splicing related proteins, DNA repair proteins and chromatin structural proteins are frequent in different cancers. While each of these mutations impact different cellular processes, several have been shown to also favor aberrant R-loop formation, a stable three stranded nucleic acid structure, which when unresolved leads to CIN.

The impact of R-loops on genome stability is well understood, however to date no highthroughput study looking at the incidence of R-loop formation following perturbation of oncogenes and tumor suppressors has been reported. This is in part due to a lack of suitable tools. We and others have recently identified that the only accepted S9.6 monoclonal antibody against R-loops also detects ssRNA, making it not suitable for IF based screens. Furthermore, there are also no reported live cell sensors of R-loops. Therefore, we are trying to establish optimized tools to allow us to screen the incidence of R-loops following perturbation (siRNA and CRISPRa) of 835 chromatin regulatory proteins (Dharmacon Epigenetic Library screens).



Figure 4: A) IF and statistical analysis of S9.6 antibody. B-C) Tet-inducible EGFP-2xHBD R-loop reporter, IF (B) and WB (C).

To this end, we have recently developed a functional synthetic R-loop reporter cell line, using the Tet-On system.





ARNDT LAB

Evolutionary and Cancer Genomics

o o o o o o o o o o o o o o

ESTABLISHED IN OCTOBER 2003

DR. PETER ARNDT arndt@molgen.mpg.de +49 30 8413 / 1162 ↘ Arndt Lab

Research Group



Focus areas

- Molecular biology and evolution
- Context-dependent models of sequence evolution
- Modelling genome evolution
- Cancer genomics

Research concept

Evolutionary processes are responsible for the diversification and adaptation of species (or cell populations) and leave distinctive marks in their genomes. These marks encompass a multitude of DNA alterations on different scales: 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 somatic and cancerous cell populations within an individual on a much shorter timescale. The rapidly increasing availability of genomic sequence data for such populations 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

Development of a new method to identify and measure rates of Horizontal Gene Transfer (HGT) across the bacterial kingdom. Massip et al., eLife, 2021

Elucidating the dynamics of segmental duplications in the human genome through network analysis.

Abdullaev et al., BMC Genomics, 2021

Scientific honors

- Associate Scientist at the Broad Institute of MIT and Harvard University, Boston, USA, since 2015
- Member of the Society of Molecular Biology and Evolution and the German Physical Society

Research projects

Dynamical aspects of the evolution of segmental duplications

With this project we wanted to investigate a very interesting complex of questions: How are eukaryotic genomes organized? Specifically, what is the role of "segmental duplications", i.e., long stretches of unique genomic DNA which occur in two or more copies in the human genome (and actually make up about 5% of human DNA)? And what type of evolutionary dynamics is needed to build up and sustain this chromosomal organization? At first glance these are purely biological questions. However, to investigate the evolutionary and dynamical aspects of the proliferation of segmental duplications, mathematical models needed to be developed.

In the last years, we made significant progress towards answering the questions above. For the first time, an approach based on network theory was used. This elegantly allows to exclude certain evolutionary/dynamical scenarios for the proliferation of segmental duplications.



Figure 1: a) Example of several duplication events and the constructed network. Alignments appear not only between copied regions, but also when a duplication overlaps with one of the existing duplicated regions. b) The network of segmental duplications of the human genome.

Identical sequences found in distant genomes reveal frequent horizontal transfer across the bacterial domain

(In collaboration with BIMSB Berlin and the University of Utrecht)

Horizontal gene transfer (HGT) is an essential force in microbial evolution. Despite detailed studies on a variety of systems, a global picture of HGT in the microbial world is still missing. Here, we exploit that HGT creates long identical DNA sequences in the genomes of distant species, which can be found efficiently using alignment-free methods. Our pairwise analysis of 93,481 bacterial genomes identified 138,273 HGT events. We developed a model to explain their statistical properties and to estimate the transfer rate between pairs of taxa. This reveals that long-distance HGT is frequent: our study indicates that HGT between species from different phyla has occurred in at least 8% of the species. Thus, our results confirm that the function of sequences strongly impacts their transfer rate, which varies by more than three orders of magnitude between different functional categories. Overall, we provide a comprehensive view of HGT, illuminating a fundamental process driving bacterial evolution.

									10^{7} 10^{8} 10^{9} 10^{10} 10^{11} total length of contigs (bp)
	• •	· · · ·	1 .7 1 12	n i ni	144 J	a aana i Aango i	L 1	2	
			「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」						
Aquificat Plancton Chlamyc Spiroche Balareola Rhoddh Proteoby Acidoba Tenericu Deinoco Cyanobs Actinob Chlorofit Balaneola Rhoddh Proteoby Acidoba Tenericu Deinoco Cyanobs Actinoba Tenericu	e nycetes liae tetes detes detes detes eeota eermaeota acteria cteria cteria cteria cteria ccus-Thermus tcteria ccus-Thermus tcteria ccus-Thermus tcteria ogae	A 10 ⁴ 10 ¹ 10 ¹ 10 ¹							

Figure 2: Effective pairwise HGT rate at bacterial family level. For each pair of families, the prefactor A proportional to the HGT rate is displayed.



Figure 3: Schematic of the mathematical model. Left: The evolutionary fate of a DNA segment following HGT. Right: Exponential MLDs (log-log scale) for many segments originating from different HGT events. The red curve is the sum of all blue, purple and grey curves and follows a power law with exponent -3.

The tumor mutational landscape is a record of the pre-malignant state

(In collaboration with the Broad Institute and the PCAWG consortium)

The chromatin structure has a major influence on the cell-specific density of somatic mutations along the cancer genome. In this project we searched for the putative cancer cell-of-origin (COO) of 2,550 pan-cancer whole genomes, representing 32 cancer types. We matched their mutational landscape to the regional patterns of chromatin modifications ascertained in 104



Figure 4: The chromatin composition of the COO is a major determinant of cancer mutation distribution and provides insights regarding tumor origins. (a) Strategy using normal chromatin profiles and cancer mutation profiles. (b) Mutation counts alongside six histone modification profiles. (c) Random Forest regression. (d) The cell type that fit the model with the highest prediction accuracy was defined as the best match.

normal tissue types and found that in almost all cancer types, the cell-of-origin can be predicted solely from their DNA sequences. We also demonstrate that the technique is equally capable of identifying the cell-of-origin for a series of 2,044 metastatic samples from 22 of the tumor types available as primaries. Our findings highlight that many somatic mutations accumulate while the chromatin structure of the cell-of-origin is maintained and that this historical record, captured in the DNA, can be used to identify the often-elusive cancer cell-of-origin.

Future projects

(In collaboration with BIH Charité, the Broad Institute, the NCI Amsterdam and Mimes ParisTech)

Following up on recently finished projects, the group is now focusing on various research projects. For instance, our recently established method to quantify the amount of horizontal gene transfer between families of bacteria (see above) can be extended to measure rates of horizontal gene transfer and/or recombination even on the species level. Instead of exact matches, one needs to consider locally aligning regions and to analyze the length distribution of exact matches within them. This type of approach can also be used to analyze genomes of archaea and eukaryotes yielding very general results about genome architecture in finite populations. Using these new insights, we will try to significantly improve the reconstruction of ancestral recombination graphs (ARGs).

We further study how pseudo-time ordered single cell expression data from several experiments can be used to identify regulatory networks. For single cell cancer samples such an analysis has the potential to reveal mis-regulated pathways in tumorigenesis.





MUNDLOS LAB

Development & Disease

- • • • •
- • • • •
- ESTABLISHED IN MAY 2000

PROF. DR. STEFAN MUNDLOS mundlos@molgen.mpg.de +49 30 8413 / 1449 ↘ Mundlos Lab

Research Group



Structure and organization of the research group

The research group Development & Disease addresses fundamental questions about the genetic origin of phenotypic variation in the context of disease and evolution. The research group is part of and works in close collaboration with the Institute for Medical Genetics and Human Genetics (IMHG) at the Charité, Universitätsmedizin Berlin. In this setting, the IMHG covers the clinical aspects, genome analysis and diagnostics of human genetic disease, whereas the research group at the Max Planck Institute for Molecular Genetics is focused on the functional analysis and basic biology of genomic function. Shared infrastructures, exchange of technical achievements and expertise, as well as common research goals ensure a successful interdisciplinary approach to study how sequence translates into phenotype and disease.

Focus areas

- Evolutionary genomics genetic basis of phenotypic adaptation
- 3D genomics chromatin architecture and long-range gene regulation
- Human genetics molecular pathogenesis of genetic disease
- Disease-causing variants in the non-coding genome

Research concept

By combining developmental and evolutionary biology, genetics, and clinical medicine, we aim to gain deep insights into how phenotypes are encoded in the genome. We focus on the non-coding regulatory genome and its function in development, disease and evolution – one of the major challenges in current genomic science. Our phenotype-driven approach aims at dissecting the molecular mechanisms of normal and abnormal development, using the skeleton/ limb as a model system. Besides our long-standing interest in deciphering the cause of rare genetic disease, we investigate basic mechanisms of gene regulation including 3D genomics, long-range chromatin interaction and non-coding transcripts such as IncRNAs. Advanced methodologies for genome engineering are used to investigate the effect of genomic manipulations on gene regulation and phenotype in vivo in mice during development. The limb as a model system offers many advantages because of its easy accessibility, different developmental stages with distinct cellular components and gene expression, and wellstudied developmental pathways. Genomic comparisons on multiple levels are used to identify regions in the genome that confer specific phenotypic adaptations, such as the ovotestes in female moles and the development of wings in bats. We use advanced genome engineering and synthetic genomics to manipulate genomic landscapes and to reconstitute parts of a genome of one species in another.

Scientific highlights

Gene regulation conflicts in evolutionarily rearranged genomes are resolved by promoter repression and 3D-restructuring of TADs. *Ringel et al., Cell, 2022*

Small deletions encompassing the IncRNA Maenli result in severe limb malformations. Allou et al., Nature, 2021

A phylogenomic strategy of genome-genome comparisons between Talpa occidentalis and other species has been used to study the genomic origin of hermaphroditism in female moles. *Real et al., Science, 2020*

CTCF loss leads to TAD-fusion but not to major changes in gene expression. The loss of the original target gene and the re-direction of enhancer promoter contacts by rearranged CTCF sites is needed for gene misexpression. Despang et al., Nature Genetics, 2019

Inversions can result in the fusion of genes/promoters with other regulatory regions of the genome. Here we investigate the effect of translocating one strong enhancer towards different target regions in a gene dense region. *Kraft et al., Nature Cell Biology, 2019*

Scientific honors and invited talks

• Mendel Medallie for pioneering achievements in the field of general and molecular biology or genetics, Leopoldina (2021)

Selected invited talks

- "German Society for Endocrinology" (virtual, 2021)
- "NIH, Enhancers, Gene Regulation and Genome Organization", Washington D.C., USA (2020)
- "British Society for Genomic Medicine", London, UK (2019)
- "34th Ernst Klenk Symposium in Molecular Medicine", Cologne, Germany (2018)

Research projects

Over the last years, we have successfully applied our methodologies and expertise to three major aspects of genomics: the non-coding genome in disease, the mechanisms by which regulatory landscapes control gene expression, and evolutionary genomics. The discovery of topologically associated domains (TADs) opened many opportunities for us and helped to answer so far unsolved questions. However, it became clear that many more layers beyond TADs exist, and that the 3D chromatin conformation we observe by HiC is just another part of the complex structure that holds our genome together. The knowledge gained over the past years by our *in vivo* analyses has enabled us to interpret genomic sequence and its variations better. We expanded our interest toward evolutionary genomics based on the hypothesis that phenotypic novelty is generated mainly in the non-coding genome by adapting gene regulation. Advances in comparative genomics in combination with developmental biology enable us to identify such sequences and test them for their function *in vivo* (Ibrahim et al., Curr Opin Genet Dev, 2020, Robson et al., Mol Cell, 2019).

Research focus: evolutionary genomics – genetic basis of phenotypic adaptation

1. Evolutionary adaptation to changes in regulatory landscapes

Genomic rearrangements are thought to be the main drivers of evolution. However, such structural changes can result in TAD disruption and altered gene regulation. It remains unknown how evolution wires novel genes and expression programs into the genome. We addressed this question by studying a locus in detail in which a new gene regulatory program appeared during evolution within a more ancient TAD. The placental mammal-specific gene *Zfp42* emerged in an ancient vertebrate TAD without adopting or disrupting the conserved expression of its gene *Fat1*. Promoter repression and 3D-restructuring enabled the integration of independent Zfp42 regulation and the persistence of Fat1 regulation at one locus. Our study shows that regulatory complexity is common in evolution as most TADs contain multiple independently expressed genes. This demonstrates how evolving genomes modify gene regulation through impressively diverse mechanisms (Ringel et al., Cell, 2022).



Figure 1: Model for the evolution of independent Zfp42R expression within Fat1's ancient TAD.

2. Regulatory genomics of evolution

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. Here, we investigated a particularly striking example of evolutionary adaptation: intersexuality (presence of ovaries and testis) in female moles. We performed a comparative analysis between species based on a high-quality genome and HiC. In addition, expression data and epigenetic modifications from different stages of gonadal development in moles and mice were compared. This phylogenomic strategy revealed a region of enhancer duplication next to a gene involved in androgen synthesis and a break in the TAD configuration next to *Fgf9*, a known sex-determination gene. Furthermore, we produced transgenic mice with a knock-in of the mole enhancer or overexpressing FGF9. These mice showed phenotypes recapitulating mole sexual features, such as high androgen levels and increased muscle strength. This work highlights how integrative genomic approaches can reveal the phenotypic impact of noncoding sequence changes. (Real et al., Science, 2020)



Figure 2: Mole (Talpa occidentalis) in a trap.

Research focus: chromatin architecture and long-range gene regulation

1. Role of TADs and their boundaries in developmental gene regulation

Studies investigating the importance of TADs in gene regulation revealed a putative contradiction: rearrangements of TADs and their boundaries can result in gene misexpression
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N
 N

but the loss of all TADs (CTCF depletion) has no major effect on gene regulation. Our study resolves this question: TADs and their boundaries reinforce enhancer-promoter interaction but they are not essential for this process (fig. 3, top). Intrinsic enhancer-promoter specificity appears to be sufficient to guide enhancers to their cognate promoters even in the absence of CTCF. However, if the target gene promoter is lost and/or enhancers are forced to contact the wrong promoter (e.g., by rearrangements) gene misexpression is possible (Despang et al., Nat Genet, 2019).

2. Mechanisms of inversions

Inversions are common rearrangements in the genome. However, it remains unclear when and how this leads to novel gene activation. Here, we systematically investigated the effect of inversions by placing a known limb enhancer located in the Epha4 TAD into a nearby gene dense region. Our study shows that the activation potential of genes depends on their epigenetic marks. Very sensitive genes like *lhh* can be activated even though several other genes are located between the enhancer and the target gene. Furthermore, we describe architectural stripes as a feature driven by enhancer activity (fig. 3, bottom). We show that architectural stripes are frequent events of developmental three-dimensional genome architecture, often associated with active enhancers. Balanced chromosomal rearrangements can induce ectopic gene expression and the formation of asymmetric chromatin contact patterns that are dependent on CTCF anchors and enhancer activity (Kraft et al., Nat Cell Biol, 2019).



Figure 3: Top: HiC showing the fusion of the Sox9/Kcnj2 TADs after removal of the boundaries and CTCF sites (scissors). 4C tracks below show leakage of contacts into the neighboring TAD, but the effect on gene expression is minimal. Bottom: HiC from limb buds showing the region with the inverted enhancer (purple oval) and the position of the breakpoints.

3. Gene regulation from the nuclear envelope

The nuclear envelope extensively interacts with chromatin to attach it to the nuclear periphery in repressive lamina-associated domains (LADs). However, how LADs are altered during embryonic development or impact gene regulation remains poorly understood. Here, we established a new technology, termed scDam&T, to simultaneously map LADs and tran-

scription in single cells from mouse embryos *in vivo*. With scDam&T, we find that critical developmental genes are frequently released from repressive LAD environments together with the surrounding chromatin that regulates their expression. Contrasting current models, this LAD detachment often occurs later in cell-differentiation in progenitors prior to gene activation. Moreover, such LAD pre-release is linked to the activity of regulatory elements and is constrained by the genome's 3D structure. Thus, we postulate that early LAD re-organisation is controlled by distinct non-coding elements and this primes regulatory domains for later activity. Using this approach, we can now unravel the previously unknown regulatory function of LADs.



Figure 4: HiC and LADs at Sox9 locus at different stages of limb development.

Research focus: molecular pathogenesis of genetic disease

1. Regulation of developmental gene expression by IncRNA Maenli

LncRNAs are important regulators of gene expression but their function often remains elusive or can only be assessed indirectly. This is in part due to the enormous diversity with which lncRNAs regulate genes and also to their overall low degree of sequence conservation. In families with a novel type of limb malformation, small non-coding deletions were identified that contained a novel lncRNA. We characterized this lncRNA called Maenli (for MAster regulator of EN1 in the LImb) extensively *in vivo* and show that it regulates the expression of the En1 gene 300 kb away. En1 is known to regulate the development of the dorsal midbrain and anterior hindbrain and is also essential in the establishment of a dorso-ventral pattern in developing limbs (fig. 5, left). Maenli is expressed in the early ectoderm of the limb bud where it activates En1 specifically in the limb. Here, we provide the first example of a human malformation caused by mutations in a lncRNA. Thus, lncRNAs can activate developmental genes in a time and cell-specific enhancer-like fashion (Allou et al., Nature, 2021).

2. Genome analysis for disease variant identification and interpretation

Whole genome sequencing has taken the identification of disease-causing variants to a new level. Nevertheless, the interpretation of the results remains challenging and only few disease-causing mutations have been identified so far. We have been analyzing large cohorts of exome-negative individuals with a range of sequencing technologies and HiC. In this study, we used HiC to better identify and interpret structural variants (fig. 5, right; Melo et al., Am J Hum Genet, 2020). Other studies that use HiC and other methods to interpret the non-coding genome: Melo et al., Hum Genet, 2021, Elsner et al., Hum Genet, 2021.

 O
 O
 O
 O
 O

 O
 O
 O
 O
 O
 O

 O
 O
 O
 O
 O
 O

 O
 O
 O
 O
 O
 O

 O
 O
 O
 O
 O
 O

 O
 O
 O
 O
 O
 O

 O
 O
 O
 O
 O
 O

 O
 O
 O
 O
 O
 O

 O
 O
 O
 O
 O
 O

 O
 O
 O
 O
 O
 O

 O
 O
 O
 O
 O
 O

 O
 O
 O
 O
 O
 O

 O
 O
 O
 O
 O
 O

 O
 O
 O
 O
 O
 O

 O
 O
 O
 O
 O
 O

 O
 O
 O

In collaboration with the Vingron department, we used long as well as short read sequencing technologies, HiC and allele-specific RNAseq to study very complex congenital rearrangements (chromothripsis, Schöpflin et al., Nature Communications, accepted).



Figure 5: Left: Inactivation of lncRNA Maenli results in an En1-like double-dorsal phenotype with circumferential claws and syndactyly. Right: HiC map of chromothrypsis samples vs. control. Arrows indicate breakpoints.

Future directions

1. Genetic engineering of regulatory evolution



Current concepts in genomics and human genetics rely on the inactivation/interference with genes and/or their pathways to elucidate their function. This is typically achieved by knock outs or other types of genome manipulation aimed at the inactivation of certain functions. However, such studies only reveal what is essential for a certain gene function or regulatory mechanism, but they do not tell us how phenotypes are encoded in the genome. While these studies have helped to unravel gene function, we still lack a basic

knowledge of how regulatory sequences encode complex morphological structures. This is because we do not understand how enhancers, promoters, and other regulatory components work together to control and fine-tune gene expression. In the future, I would like to take a different approach. I plan to study how evolutionary adaptation is encoded and thus investigate the emergence of novelty in genomic information. We will use an extreme case of evolutionary adaptation, the bat wing, to study how sequence encodes for this amazing phenotype. We collected a vast amount of genomic information of three types of bats and the mouse, including full length chromosomes and gene regulatory data during early and late stages of bat and mouse limb development. By comparing fore- and hindlimb development in bats vs. mice, we aim to identify regions that are causally related to bat wing development. Our strategy is based on our successful phylogenomic analysis in moles.

2. Synthetic genomics for artificial sequence design

CRISPR/Cas9-based technologies have been instrumental in the generation of large structural variations, point mutations or smaller insertions. However, large-scale insertions/



replacements are not possible with current technologies. This is a major problem when studying regulatory landscapes which are characteristically several 100 kb large. We will use a technology developed to artificially produce the yeast genome. Yeast can be used to synthesize large fragments of DNA *de novo* (up to 150 kb), by a homologous recombination-based approach. These BAC-size fragments will then be inserted in the mouse genome using a landing pad technology. We will use these technologies to reconstitute parts of the bat genome in mice. These mice will then be used to study the effects on gene regulation and expres-

sion, as well as phenotype. The *de novo* design of sequence gives us the opportunity to generate regulatory landscapes from scratch which can then be used as a testbed for experimental perturbations.

Cooperation within the institute

We have an intensive cooperation with the Vingron department, especially in the analysis of chromosome conformation capture data (4C, HiC), as well as on ChIP-seq, ATAC-seq analysis and comparative genomics. This collaboration has been very fruitful and without it these results could have not been achieved. 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. We collaborate with the EM-facility in studying the structure of nucleosomes. Intensive collaborations exist with the sequencing facility. Collaborations with the Meissner department focus on methylome analysis in mouse mutants. We have been collaborating with the Schulz group on the *Xist* locus and the Hnisz group on the function of Hox mutants and, more recently, on mutations in *HMGB1* that alter phase separation.

Special facilities / equipment

Both the research group and the IMG are equipped with standard facilities for research in genetics, developmental biology, cell biology, and molecular biology.





KALSCHEUER LAB

Chromosome Rearrangements and Disease

ESTABLISHED IN JULY 1995 IN THE FORMER DEPARTMENT OF HUMAN MOLECULAR GENETICS, TRANSITIONED TO THE RESEARCH GROUP DEVELOPMENT & DISEASE IN NOVEMBER 2014

DR. VERA M. KALSCHEUER kalscheu@molgen.mpg.de +49 30 8413 / 1293 ↘ Kalscheuer Lab



Focus areas

- Human genetics
- Neurodevelopmental disorders and intellectual disability
- Genotype-phenotype relationship
- Pathomechanism and pathophysiology of selected monogenic disorders of the nervous system

Research concept

The research group focuses on the identification of novel genes for monogenic neurodevelopmental disorders (NDD) and intellectual disability (ID), particularly on human chromosome X. We aim to better understand the function and the underlying molecular pathology of selected, previously discovered NDD and ID genes. Research activities are carried out at different levels, using *in vitro* and *in vivo* model systems within the group and through international collaborations. A clear knowledge of the cellular processes and molecules that are involved in monogenic disorders will hopefully pave the way towards the identification of novel targets that can be used for the development of therapies for genetically caused disorders.

Scientific highlights

In recent years, we discovered and characterized novel variants of X-chromosome-linked (candidate) disease genes within the group and through collaborations, e.g., *CSTF2*, *TAF1*, *EIF2S3*, *ZC4H2*, *RLIM*, *CLCN4*, and *MSL3*, and studied the associated phenotypes. Kotzaeridou et al., Clinical Genetics, 2020 Frints et al., Human Mutation, 2019 Frints et al., Molecular Psychiatry, 2019 Palmer et al., Molecular Psychiatry, 2018

Scientific honors and invited talks

• Orphan Disease Center on behalf of ZC4H2 Research Foundation award

Selected invited talks

- Speaker and co-organizer of the "1st International Scientific Conference on CLCN4 Condition", London, UK (2022)
- "CDKL5 Workshop in Asia", Tokyo, Japan (2019)
- "1st International scientific meeting on ZC4H2 deficiency", Dubai, UAE (2018)

Research projects

Novel pathogenic *EIF2S3* missense variants causing clinically variable MEHMO syndrome with impaired $eIF2\gamma$ translational function

The *EIF2S3* gene encodes the gamma subunit of eukaryotic translation initiation factor 2 (eIF2) which is critical for initiating protein synthesis and regulating



Figure 1: Novel variants of **EIF2S3** causing clinically variable MEHMO syndrome and presence of mutant eIF2 gamma protein in patient-derived cells.

the integrated stress response. We have previously linked rare variants of *EIF2S3* to MEHMO syndrome. This rare syndrome is characterized by severe intellectual disability, epilepsy, hypogonadism, microcephaly, and obesity. Our findings added MEHMO to a group of disorders associated with eIF2 that cause translation dysregulation (Skopkova et al., 2017). Follow-up studies on novel rare *EIF2S3* missense variants expanded the limited genetic spectrum and underscored the clinically variable expressivity of MEHMO, caused by impaired eIF2 gamma translation function (Kotzaeridou et al., 2020).

Pathogenic variants in the E3 ubiquitin ligase RLIM/RNF12 lead to a syndromic X-linked intellectual disability and behavior disorder.

Pathogenic variants of the X-linked gene *RLIM*, also known as *RNF12*, cause Tönne-Kalscheuer syndrome, an ID syndrome associated with variable clinical features, including craniofacial abnormalities, hypogenitalism and diaphragmatic hernia. RLIM (RING finger LIM domain-interacting protein) is an E3 ubiquitin ligase acting as a negative regulator of LIM-domain containing transcription factors. We first reported *RLIM* variants in three large families (Hu et al., 2016) and recently published the genetic and clinical findings of additional families. Of these nine families 40 affected males presented with ID and variable behavioral anomalies with or without congenital malformations. All identified *RLIM* variants are missense changes that co-segregated with the phenotype. *In vitro* functional studies revealed that the variants located in the RING-H2 zinc finger domain of RLIM decreased ubiquitin ligase activity. *In vivo* functional studies in zebrafish showed that the absence of *rlim* caused microcephaly. None of the patient-specific variants rescued this phenotype, thus, likely representing severe loss-of-function variants. Taken together, we suggest that enzyme activity of RLIM is required for normal development, cognition and behavior (Frints et al., 2019). Ongoing studies on novel *RLIM* variants should expand our understanding of this rare syndrome.



Figure 2: Location of the **RLIM** gene on the X chromosome, schematic of the gene and protein, along with multiple alignments showing high conservation of the amino acids mutated in the families A-I.

Functional and clinical studies of variants in the *CLCN4* gene revealed pathophysiological complexity of the *CLCN4*-related neurodevelopmental condition.

CLCN4 encodes the intracellularly located chloride/proton ion-exchanger CIC-4. We first identified pathogenic *CLCN4* variants in a small number of families with X-linked intellectual disability and seizure disorder (Hu et al., 2016) and reported additional families with novel *de novo* or inherited *CLCN4* variants. Intellectual disability ranged from borderline to profound. Behavioral and psychiatric disorders were common in both childhood 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. Reduced or complete loss-of-function was demonstrated by electrophysiological studies in *Xenopus laevis* oocytes (Palmer et al., 2018). Continued molecular, functional and clinical studies on novel *CLCN4* variants from 43 families revealed the pathophysiological complexity of the *CLCN4*-related neurodevelopmental condition. The analysis of missense variants with respect to electrophysiological properties in *Xenopus* oocytes using extended voltage and pH ranges revealed novel pathophysiological mechanisms, including a toxic gain-of-function associated with a disrupted gate that allows inward transport at negative voltages (Palmer et al., manuscript under review).

Figure 3: A) CLCN4 variants from newly identified families. B) Quantitative analysis of pH dependence of gainof-function variants.

Deleterious variants of X-linked ZC4H2 cause a variable phenotype in males and females with neurogenic arthrogryposis multiplex congenita.

Pathogenic variants of *ZC4H2*, which encodes a zinc-finger protein, cause an infrequently described syndromic form of arthrogryposis multiplex congenita (AMC) with central and





peripheral nervous system involvement (Hirata et al., 2013). Following up on our previous work, we reported the molecular genetic and clinical spectrum of 23 newly identified families with ZC4H2-Associated Rare Disorders (ZARD). *ZC4H2* gene defects comprised novel and recurrent, mostly inherited missense variants in affected males, and *de novo* splicing, frame-shift, nonsense and partial *ZC4H2* deletions in affected females. Of note, 15 females with deleterious *de novo ZC4H2* variants presented with phenotypes ranging from mild to severe, and their clinical features overlapped with those seen in affected males. We propose *ZC4H2* as a good candidate for early genetic testing of males and females with a clinical suspicion of fetal hypo-/akinesia and/or (neurogenic) AMC (Frints et al., 2019). Our ongoing and future studies should improve the understanding of ZC4H2's function and ZARD and pave the way for the development of strategies to benefit patients in the short and long term.



Figure 4: Overview of likely pathogenic variants of ZC4H2. The horizontal bars show **de novo** deletions of ZC4H2 in affected females, and the distribution of variants along the protein is shown below, with **de novo** variants in females underlined.





METZGER LAB

ESTABLISHED IN JULY 2020

Veterinary Functional Genomics

PROF. DR. JULIA METZGER jmetzger@molgen.mpg.de +49 30 8413 / 1598

☑ Metzger Lab

Heisenberg Research Group



Focus areas

- Functional genomics genome architecture in mammalian model organisms
- Bone growth biology targeting complex regulatory effects
- Computational genomics for development and evolution

Research concept

The research group aims at understanding how sequence and structural variation are encoded in the genome and investigates its functional consequences. Latest omics technologies are used to link sequence patterns and variants to the 3D genome and the underlying complex regulatory networks essential for the development and growth of a body. We follow an integrated wet-lab and computational approach to investigate phenotype-genotype interrelations from a developmental and evolutionary genetics point of view. In particular, our research focuses on the consequences of genomic selection on body size to sophisticated processes in limb bone development at single cell resolution in mammalian model organisms.

Scientific highlights

The lab developed an improved workflow for Runs of Homozygosity (ROH) detection in whole genome sequencing data to identify signatures of selection. *Berghöfer et al., BMC Genomics, 2022*

Scientific honors

• Appointment as Heisenberg Professor by the German Research Foundation (DFG), 2020

Research projects

Development of an improved workflow for Runs of Homozygosity detection in pigs

Since the introduction of genomics into research work, we encounter the challenge of scanning millions of variants to find those involved in biological processes, specific phenotypes, or disease traits. Homozygosity mapping has been shown to be a successful method to narrow down regions of interest associated with specific phenotypes or disease traits. Long stretches of homozygous regions, designated as Runs of Homozygosity (ROHs), represent the inheritance of two copies of an ancestral allele and can display past selection events leaving footprints in the genome. In our study, we evaluated different methods for ROH calling, testing customization with regard to sample size, heterozygosity and datasets as well as different settings specifically necessary for high-throughput data. We performed systematic parameter testing for rule- and model-based approaches, using an exemplary dataset of different pigs. With this work, we offer a workflow for ROH detection adjusted for whole genome sequencing (WGS) data and demonstrate its value for the identification of potential footprints of selection in pigs, displaying their breed-specific characteristics or favorable phenotypes.



Figure 1: Workflow of ROH detection pipeline.

Dissecting gene regulatory networks behind the determination of body size ("MEASURE"-project)

The process of growth is a highly complex developmental mechanism affecting the architecture and function of a body. Several regulatory variants are expected to play a major role in cell division rates in key areas of longitudinal bone growth, the so-called growth plates. The current study particularly focuses on the pig, a model animal, which underwent strong artificial selection forces for extremely small and large skeletal size. We are performing comprehensive analyses of whole genome sequencing data to narrow down variants of interest using our improved workflow for Runs of Homozygosity (ROH) detection. As a next step, potential candidate variants in these regions are functionally annotated in order to predict the effect on skeletal growth. Samples from growth plates of miniature and large pig breeds were obtained and underwent sequencing, using multiple omics-techniques to measure differentially expressed genes (DEGs) and enhancer elements. These high-throughput data are currently used for variant enrichment analyses to identify alleles that potentially affect the tight control mechanisms of cis-regulatory elements, like enhancers or promotors on gene expression patterns, prioritized in chondrocyte development. The physical interactions of the identified elements will then be further assigned by topologically associating domains (TADs) and validated in 3D cell culture, modelling the proliferative rates of chondrocytes. These findings will contribute to the understanding of longitudinal growth of the skeleton and provide deep insights into the molecular processes involved in the development of a body.



Figure 2: Schematic of experimental and data analysis workflow. Large and miniature pigs are produced and undergo sampling for multi-omics analyses of growth plates at the age of 80 days. ROHs are detected, common regions (ROHR) are identified and merged. This allows to narrow down the number of variants of interest significantly, which are subsequently annotated using genetic and epigenetic tissue-profiles.



Regulatory networks in limb development at single-cell resolution

The developmental evolution of mammalian limbs is based on a complex interplay of selective pressures for energetic benefits and adaption to the environment. While the earliest development of the limb indicates strong conservation among mammals, the diversity in later stages is rather high. In this project, we aim to investigate this distal progression of the limb buds in an even-toed ungulate, the pig. Limb buds from two developmental stages, gestational day (D) 24 and D31, were sampled and underwent single-cell RNA sequencing. Subsequently, these data will allow comparative identification and annotation of cell types to elucidate the topology and dynamics of gene regulatory networks that determine the patterning of these limbs. Furthermore, an interspecies integration of single-cell expression data from mouse and bat limb buds at equivalent embryonic stages will be performed to learn more about the underlying morphological diversification of the number and shape of toes. This thorough dissection of cell types and their trajectories will contribute to the understanding of how morphology is encoded in the genome.





AKTAS LAB

Quantitative RNA Biology

• • • • • •

ESTABLISHED IN DECEMBER 2018

TUGCE AKTAS, PHD aktas@molgen.mpg.de +49 30 8413 / 1550 ↘ Aktas Lab

Max Planck Research Group



Focus areas

- RNA processing: function of nuclear speckles
- Nuclear architecture: role of nuclear speckles in the 3D genome from flies to humans
- Innate immunity and transposable elements: dsRNA regulation by DHX9 & ADAR1
- Gene expression regulation: mechanistic cross-talk of splicing and transcription

Research concept

The Aktas group combines experimental and computational approaches to investigate fundamental aspects of co-transcriptional RNA processing with a specific focus on transposable elements. Analyses of our genome-wide datasets obtained from multiple model organisms serve as tools to instruct the experimental approaches. These functional assays guide our evolutionary analyses and enable us to extrapolate our findings into non-model organisms. We use a wide range of methodologies in the fields of biochemistry, cell biology and genomics: *in vivo* proteomics, *in vitro* kinetic measurements of protein-RNA interactions, genetically engineered cell lines for functional assessments and genome wide identifications of a large list of RNA binding proteins under perturbed transcriptional states. The combination of alternative methods and functional characterizations are crucial to answer our research questions from a mechanistic perspective.

Scientific highlights

Identification of the essential proteins for nuclear speckle (NS) formation which enable us to address the function of NS from an evolutionary perspective for the first time.

llik et al., eLife, 2020

Scientific honors and invited talks

 Selected as "rising star" scientist and invited for the "Sabri Ülker Symposium" in 2023, organized by Prof. Gokhan Hotamisligil, Director at the Sabri Ülker Research Center at Harvard University

Selected invited talks

- Institute of Molecular Biology (IMB), Mainz, Germany (2022)
- "International meeting of the SPP1935 & the RTG2355" (virtual, 2022)
- Center for Molecular and Cellular Bioengineering at TU Dresden, Germany (2021)
- Interdisciplinary Life Sciences Graduate Program at Molecular Biosciences, University of Texas, Austin, USA (2020)

Research projects

What are nuclear speckles?

An abundant source of disinformation in the gene expression process is transposable elements (TEs). TEs clutter our genome with sequences that recruit transcription factors and RNA binding proteins (RBPs), wasting cellular resources occasionally with fatal results. In our first project we focused on nuclear speckles (NS), which are biomolecular condensates which contain ~360 proteins and spliced RNA. Most of these proteins are RBPs that are associated with pre-mRNA splicing and have a large overlap with spliceosome-associated proteins and other RNA-processing factors involved in 3'-end cleavage/polyadenylation and nuclear export – processes that are tightly connected with each other. Investigating the essential core of NS would therefore enable us to disrupt NS, decipher the role of NS in exon and intron definition and study if NS are involved in keeping transposon derived RNA in the nucleus.

We started by characterizing the well-established NS marker antibody known as SC-35. We found that the first characterization of this antibody in 1992 was incorrect and that the main target of SC35 mAb is SRRM2, a spliceosome-associated protein that sharply localizes to NS. Following this unexpected discovery, our group identified the core of nuclear speckles as two largely disordered and rapidly evolving proteins called SON and SRRM2 (Ilik et al., eLife, 2020). The disordered fractions of the proteins are controlled by expanding a single, unusually long exon that almost perfectly correlates with the length of the protein in a given species.


Taxon:
Invertebrates
Fish
Amphibians
Reptiles
Birds
Mammals
Other Vertebrates

Figure 1: A) Respective localization of the two NS core proteins in human cells. B) Evolution of protein length and disordered regions is shown for SON and SRRM2 in metazoa. C) The protein length of SON and SRRM2 is determined by the length of the longest exon in different organisms.

Evolution of nuclear speckle formation and its function

Now that we have an operational definition of what makes NS, we started working on identifying the role of NS in splicing. We induced an NS-absent state in human cells and performed bulk RNA sequencing. Our preliminary results indicate intron retention as the major splicing defect. Short introns with higher GC-content are especially affected. Our phylogenetic analysis shows that both SON and SRRM2 are conserved in animals. Therefore, we are currently investigating the protein composition and function of NS in *Drosophila melanogaster*. To this end, we expressed a fosmid, that contains the genomic locus of the fly *Son* gene fused to GFP, and observed a speckle-like GFP signal in the nucleus of transfected cells. Actively transcribed gene loci are proposed to get in closer proximity of NS, suggesting a role for NS in nuclear architecture. Therefore, we have started to work on an integrative model of NS in gene expression, RNA-splicing, 3D genome architecture from flies to humans.



*transfection marker

Figure 2: A) Side-by-side alignment (aligned regions: black boxes) of the human and fly SON with annotated C-terminal domains. B) D. melanogaster SON (CG8273) fused to GFP shows a speckle-like nuclear GFP signal.

Nuclear retention and intronization of TE-derived RNA

In order to identify RBPs that mitigate the adverse effects of TE insertions, we carried out screens in human cells on SR- or SR-like proteins, which define exons, and on hnRNPs, which define introns. We identified a family of RBPs that suppresses L1 elements, which are the source of all current retrotransposition activity in humans. This family of RBPs prevents L1-exonization by competing with SR proteins for binding to an exonic splicing enhancer (ESE) motif that seems to be enriched on several autonomous TE families. Interestingly,



Figure 3: A) List of RBPs used in the screen. B) UMAP representation of regions bound by RBPs split into unsupervised clusters (labeled according to target distribution). Known Alu-binders DHX9 and hnRNP-C form separate clusters. C) Co-depletion of L1 binding RBPs lead to L1-exonization and early termination.

this shielding mechanism is lifted in germ cells, possibly allowing TE mobility and copy number increase, while keeping them dormant in somatic cells. We expanded our work into *Mus musculus* and *Drosophila melanogaster* and found that the function of this RBP family in preventing TE-exonization is conserved. These findings propose an evolutionary conserved, RNA-based restriction of TE activity in the soma, where piRNA pathways are mostly non-existing.

Functional study of nuclear dsRNA binding proteins DHX9 and ADAR1

Our previous work before joining the MPI for Molecular Genetics showed that double stranded-RNA (dsRNA) binding RNA helicase DHX9 resolves secondary structures. These structures are formed by the transcription of the inverted repeats and is mainly caused by Alu elements in humans. We found that the interferon induced p150 isoform of RNA editing enzyme ADAR1 directly interacts with DHX9. Both the resolvase function of DHX9 and its interaction with ADAR1^{p150} is conserved in rodents. This intriguing finding suggests that DHX9-ADAR1 interaction precedes the divergence of SINE repeats and might have originally evolved for innate immunity. We generated genetically engineered cell lines which allow us to rapidly degrade DHX9 or ADAR1 independently or simultaneously. These cell lines will help us answer what is happening to the nuclear dsRNA network using proteomics and RNA-RNA interaction mapping approaches. Furthermore, we initially screen RNA target re-distribution of DHX9 upon ADAR1 depletion (or vice versa) to determine dependencies between the two enzyme activities. Currently, we are expanding our research into non-model organisms such as cephalopods that are reported to have extensive re-coding RNA editing. We will use these systems to investigate the power of TE-induced innovations in overcoming acute and chronic stressors, from heat shock to the impending climate crisis.



Figure 4: A) The helicase activity of DHX9 helps genomes tolerate repeat expansion. B) DHX9 directly interacts with the p150 isoform of ADAR1. C) DHX9 and ADAR1 are two dsRNA binding enzymes.

AKTAS LAB



Figure 5: A) Generation of genetically engineered rapid depletion cell lines. B) Induction of protein degradation in double degron cell line (Shield-1 is used to stabilize the tagged ADAR1 protein, whereas Auxin and dTAG13 are the degraders). U2AF2 serves as loading control.



112



DR. ANDREAS MAYER mayer@molgen.mpg.de

+ 49 30 8413 / 1264

☑ Mayer Lab

MAYER LAB

High-resolution Functional Genomics

ESTABLISHED IN FEBRUARY 2017



Max Planck Research Group

Focus areas

- RNA polymerase II transcription
- RNA processing (splicing, 3'-RNA processing)
- BET bromodomain proteins
- Chromatin regulation (chromatin remodeler)
- Cell differentiation
- Methods development

Research concept

Our research aims at understanding the molecular mechanisms that control RNA polymerase II transcription and RNA processing in mature and differentiating mammalian cells. We are particularly interested in post-initiation regulatory control mechanisms acting at the level of transcription elongation, and how elongation is coordinated with RNA processing. To identify regulatory factors and disentangle their individual roles in these fundamental processes, we develop and apply a high-resolution functional multi-omics approach that includes genome-wide experimental methods and computational tools. Ultimate goals are to elucidate how these mechanisms control cell lineage determination and how misregulation contributes to the development of disease.

Scientific highlights

Using a functional multi-omics approach, we revealed a general BRD4-mediated transcriptional checkpoint that couples RNA polymerase II transcription elongation with 3'-RNA processing in human cells *Arnold et al., Molecular Cell, 2021*

Identification of the RNA polymerase II pausing landscape at single-nucleotide resolution, pointing to a universally conserved pausing mechanism *Gajos et al., Nucleic Acids Research, 2021*

Selected invited talks

- Invited speaker and session chair at the Keystone Meeting "Gene Regulation: From Emerging Technologies to New Models", Santa Fe, USA (2022)
- Speaker at the Cold Spring Harbor Meeting "Eukaryotic mRNA Processing" (virtual, 2021)
- Speaker at the EMBO Symposium "Multiomics to Mechanism" (virtual, 2021)
- Speaker at the 45th FEBS Congress, FEBS2021 (virtual, 2021)
- Invited speaker at the meeting "Non-coding Genome", Institut Curie, Paris, France (2019)

Research projects

Determine direct functions of the BET protein BRD4 in transcription and RNA processing

In this project, we elucidated direct BRD4-specific functions in RNA polymerase II (Pol II) transcription and co-transcriptional RNA processing. We established a human cell line that allows rapid and selective degradation of endogenous BRD4 (fig. 1A, B). As an imme-



Figure 1: Direct roles of BRD4 in post-initiation transcription control. For a detailed description of the figure panels, see Arnold et al., Mol Cell, 2021.

diate consequence, acute BRD4 loss impaired the release of Pol II into productive elongation (fig. 1C) and unexpectedly led to massive readthrough transcription at the 3'-end of genes, as revealed by an improved NET-seq method (fig. 1D). An integrated proteome-wide analysis showed that both observations are linked by BRD4, uncovering 5'-elongation and 3'-RNA processing factors as core BRD4 interactors (fig. 1E). Further mechanistic analyses revealed that the loss of BRD4 disrupted the recruitment of 3'-RNA processing factors at the 5'-control region (fig. 1F), leading to RNA cleavage and transcription termination defects. These studies indicate a BRD4-mediated, general elongation checkpoint and establish a molecular link between 5'-elongation control and 3'-RNA processing (Arnold et al., Mol Cell, 2021; Eischer et al., RNA WIREs, 2022; Altendorfer et al., Transcription, 2022).

Identifying the landscape of putative post-initiation transcription regulatory sites in human cells

In this combined computational and experimental study, we developed a transcriptional pause site detection algorithm (PDA) for nucleotide-resolution occupancy data and a new NET-seq approach, termed nested NET-seq, to reveal widespread genomic Pol II pausing at single-nucleotide resolution in human cells (fig. 2A). Unexpectedly, we discovered that most Pol II pauses occur outside of promoter-proximal gene regions, primarily along the gene-body of the transcribed genes (fig. 2B, C). Sequence analysis combined with machine learning modeling reveals the DNA sequence properties that underly widespread transcriptional pausing, including new pause motifs (fig. 2D). Interestingly, the key sequence determinants of RNA polymerase pausing are conserved between human cells and bacteria. This study indicates pervasive, sequence-induced transcriptional pausing in human cells and suggests a new post-initiation regulatory layer to control gene expression (Gajos et al., NAR, 2021).



Figure 2: Define potential post-initiation transcription regulatory sites with single-nucleotide precision. For a detailed description of the figure panels, see Gajos et al., NAR, 2021.

Development of a computational pipeline for processing and visualizing functional features of transcript isoforms

Here, we developed a versatile pipeline, called isoTV, to process, predict, and visualize the functional features of potentially translated transcript isoforms. Attributes such as gene and isoform expression, transcript composition, and functional features are summarized in an easy-to-interpret visualization (fig. 3, Annaldasula et al., Bioinformatics, 2021).





Understanding human disease mechanisms

In this collaborative project with Dr. Bruno Reversade's lab at the Genome Institute, A*STAR (Singapore), we determined the genetic basis and molecular mechanisms of the human disease Complex Lethal Osteochondrodysplasia. Using an interdisciplinary human genetics and functional genomics approach, we could identify a deep intronic mutation in the *TAPT1* locus causal for this disease. The mutated allele is recessively inherited (fig. 4A), leads to mis-splicing and results in an unstable *TAPT1* RNA (fig. 4B). Transcriptome analyses could identify the affected molecular pathways that underlie the disease phenotype (manuscript in revision).



Figure 4: Molecular basis of Complex Lethal Osteochondrodysplasia. A) Pedigree tree of a family with the TAPT1 mutation. Filled symbols, affected. The results of variant detection are indicated below the symbols (- mutated allele; + normal allele). B) TAPT1 mutation provokes exon skipping (red). RNA levels are shown for the region of interest of the TAPT1 locus. The Sashimi plot indicating exon usage is given below. Data is shown for a healthy individual (WT), the healthy mother (IV.2) and the patient (V.5).

Mechanisms of enhancer-target gene communication

In this ongoing study, we investigate how enhancer transcription is regulated and coordinated with transcription at their target genes to control transcript levels in human cells. A particular focus is on BET proteins including BRD4 which we found to bind to enhancers (fig. 5A) and is required for genome-wide enhancer transcription (fig. 5B).



Figure 5: Role of BRD4 in enhancer transcription. A) BRD4 occupancy at promoters and enhancers. B) Change in the occupancy of Pol II at enhancers upon acute BRD4 ablation.

Determine the role and regulatory dynamics of transcription and RNA processing in neuronal cell fate determination

In this project, we address the following questions: are neuron-specific gene expression patterns established gradually or in the form of transcription and splicing switches? What are the determinants of these dynamic transcriptional and splicing changes as cells transition from a pluripotent to a differentiated state? In order to accomplish these goals, we have adapted an *in vitro* human neuronal differentiation system which is amenable for genomeand proteome-wide studies (fig. 6A, B). Currently, we are collecting multi-omics data.



Figure 6: Optimized human neurogenesis model for the analysis of dynamical transcription and RNA processing control mechanisms.

Future directions

In the future, we will further improve and diversify our functional multi-omics approach to continue uncovering new general mechanisms of transcription and RNA processing regulation in human cells. We will extend this approach to other Pol II transcription factors to determine their direct roles, and to identify new factors that link transcription and RNA processing. Finally, we will elucidate their potential functions in neuronal stem cell differentiation and disease development.





SCHULZ LAB

Systems Epigenetics

- • • • •
- • • • •

• ESTABLISHED IN JANUARY 2015



DR. EDDA G. SCHULZ edda.schulz@molgen.mpg.de +49 30 8413 / 1226 ↘ Schulz Lab

Lise Meitner Research Group



Focus areas

- Epigenetics regulatory principles governing the onset of X-chromosome inactivation
- Quantitative biology understanding the complexities of quantitative gene regulation
- Synthetic biology rebuilding and dissecting gene-regulatory modules in isolation

Research concept

The main goal of our research is to understand how multiple quantitative signals are decoded by gene-regulatory networks and cis-regulatory landscapes to precisely control gene expression in mammals. We integrate theoretical approaches, such as mathematical modelling, with experimental techniques, such as (epi) genomic profiling, CRISPR screens and (epi)genetic editing. We use the Xist gene as a model, which is upregulated during early development from one X chromosome in female mammals to initiate X-chromosome inactivation. The Xist regulatory network senses a two-fold expression difference of X-linked genes between the sexes, integrates information on X-dosage and differentiation and maintains two distinct Xist expression states within the same nucleus. This model thus allows us to investigate (1) how quantitative information is decoded, (2) how multiple signals are integrated, and (3) how distinct transcriptional states are maintained through epigenetic memory.

Scientific highlights

New IncRNA-associated enhancer cluster that governs developmental regulation of Xist. *Gjaltema*, Schwämmle* et al., Molecular Cell, 2022*

Identification of X-linked genes that underlie sex differences in embryonic stem cells. *Genolet et al., Genome Biology, 2021*

Model of the Xist regulatory network to explain Xist patterns across evolution. *Mutzel et al., Nature Structural & Molecular Biology, 2019*

Applied machine learning to dissect the rules governing Xist-mediated silencing. *Barros et al., Genome Research, 2019*

Scientific honors and invited talks

- EMBO Young Investigator (2021)
- ERC Starting Grant "CisTune" (2020)
- HFSP Career development award (2018)

Selected invited talks

- "Annual Meeting of the International Society of Stem Cell Research (ISSCR)" (virtual, 2021)
- Symposium "X-chromosome Genetics and Epigenetics", College de France, Paris, France (2018)

Research projects

During embryonic development, genes exhibit complex expression patterns and perform a variety of signal-processing tasks, such as logical integration of two signals, establishing response thresholds and maintaining a memory of past signals (fig. 1). How such advanced signal-processing tasks are performed in higher organisms remains largely unknown. Our research aims at understanding how transcription factors and regulatory elements work together across different layers of regulation to process distinct developmental signals and to fine-tune gene activity.



Figure 1: Using the Xist gene as a model, we investigate how gene-regulatory landscapes process information.

Experimental and theoretical approaches

To dissect the complexities of gene regulation in mammals, our team combines state-ofthe-art high-throughput (epi)genome editing approaches and (single-cell) genomics with mechanistic mathematical modelling. We build on recent advances in CRISPR/Cas9-mediated (epi)genome editing approaches (Genolet et al. 2021; Gjaltema et al. 2022) and develop new tools to profile functional interactions across regulatory layers to investigate quantitative aspects of gene regulation. We have for example developed a new tool, called CasTuner (fig. 2), which allows us to quantitatively tune endogenous gene expression (Noviello et al., submitted). Moreover, we are currently establishing a new research direction, where we aim to build synthetic loci in a controlled genomic environment to test regulatory principles that we have identified at an endogenous model locus.



1. The Xist regulatory network

Through mathematical modelling and experimental testing of model predictions, we have identified two network modules required for female-specific and mono-allelic Xist upregulation (fig. 3): a dose-dependent X-linked Xist activator and a cis-acting positive feedback loop that establishes an activation threshold and maintains epigenetic memory of the allelic expression state (Mutzel et al., 2019). To dissect the molecular implementations of the predicted regulatory modules, we have started to comprehensively profile the regulatory network and the genomic locus that control Xist: We have characterized the onset of XCI at the transcriptional and chromatin levels (Pacini et al, 2021; Gjaltema*, Schwämmle* et al., 2022) and established an assay to investigate Xist regulation through pooled CRISPR screens. In a first screen to find new regulators, we have identified the GATA transcription factor family as potent Xist activators (Ravid Lustig et al., 2022). In a collaboration with the Meissner lab, we could show that GATA factors are essential for Xist upregulation after fertilization. Currently, we are testing all transcription factors in the mouse genome for a role in Xist regulation (Schwämmle et al., unpublished). We already have identified X-linked factors that affect pluripotency and differentiation in an X-dosage dependent manner, which might play a role in coupling X-inactivation and developmental progression (Genolet et al., 2021).



Figure 3: A systematic model comparison to identify the network that can ensure mono-allelic Xist up-regulation in females.

2. Signal-decoding at the Xist locus

To dissect the cis-regulatory landscape that controls Xist upregulation, we have used the same assay in a non-coding CRISPR screen together with extensive chromatin profiling to identify and characterize regulatory elements of Xist in collaboration with the Mundlos lab (Gjaltema*, Schwämmle* et al., 2022). We discovered that X-dosage and differentiation cues control proximal and distal regulatory elements, respectively (fig. 4), which we propose as a regulatory concept that allows integration of two signals with an AND-logic (coincidence detection). We discovered a new distal enhancer cluster associated with a lncRNA we named Xert, that activates Xist in cis in response to differentiation cues. Next, we aim to map the functional interactions between transcription factors and regulatory elements identified in our CRISPR screens in a high-throughput fashion.



Figure 4: Distinct genetic elements underlie X-dosage sensing and developmental regulation at the Xist locus.

3. Synthetic biology

We have started to develop (high-throughput) synthetic biology approaches to build synthetic loci at a neutral location in the mouse genome to understand the regulatory principles underlying gene regulation. Through such an approach we could show that antisense transcription can induce epigenetic memory under certain conditions (Mutzel et al., in preparation). To profile synthetic loci in a high-throughput fashion, we are establishing a new assay for highly parallel chromatin profiling (Kanata et al., unpublished). In the future, we plan to construct and analyze complex genomic loci and perform high-throughput screens in such synthetic genomic contexts to gain a detailed quantitative understanding of mammalian gene regulation and to validate the regulatory principles discovered at the Xist locus. Our ultimate goal will be the ability to predict and control expression patterns, levels and single-cell distributions from newly designed synthetic loci.





YASPO LAB Gene Regulation and Systems Biology of Cancer

DR. MARIE-LAURE YASPO yaspo@molgen.mpg.de +49 30 8413 1356 뇌 Yaspo Lab

Research Group



Focus areas

• Cancer molecular genetics

ESTABLISHED IN APRIL 2002

- Tumor microenvironment
- Relapse pediatric leukemia, rare melanomas
- Gene regulation networks, integrative omics and eRNAs
- Gene expression profiling
- Imaging mass cytometry

Research concept

Our central research goal is to understand complex molecular processes and gene regulation networks involved in the pathogenesis of adult and pediatric refractory cancers associated with dismal prognosis. We analyze cancer molecular landscapes with a systems biology approach, in which we integrate data generated from NGS technologies, gene perturbations, RNAseq deconvolution and imaging mass cytometry. We have developed a bench-to-bedside concept in collaborative projects with university hospitals, thus, translating basic research findings to improve tumor diagnostics and personalized therapies. The studied patient groups include rare melanoma subtypes and pediatric cancers of diverse origins, considered as diseases of dysregulated development, with a focus on acute lymphoblastic leukemia.

Scientific highlights

Molecular and clinical trajectories of early-onset prostate cancer reveal an aggressive group overexpressing ERSP1. *Gerhauser et al., Cancer Cell, 2018*

Selected invited talks

- "3rd Annual Imaging Mass Cytometry User Group Meeting", Zurich, Switzerland (2019)
- "Melanoma Workshop", Organizer, VIB-KU Leuven Center for Cancer Biology, Otranto, Italy (2019)
- Big Data Institute / Ludwig Institute for Cancer Research, Oxford, UK (2019)
- "Horizons in Molecular Biology, Career Fair", Max Planck Institute for Biophysical Chemistry, Göttingen, Germany (2018)
- Symposium "Visualize a New Path Forward Mass Cytometry and Microfluidics for Translational Cancer Research", Amsterdam, Netherlands (2018)

Research projects

Treat20Plus study

(Coordinator, in cooperation with Charité and Alacris GmbH, Berlin)

The molecular profiling of "hard-to-treat" melanomas served as a basis for the identification of associated oncogenic pathways and therapeutic targets supporting treatment decisions. In contrast to cutaneous melanoma, which is associated with UV light exposure and BRAF V600E mutations, rare sun-shielded melanomas (mucosal, acral, uveal) are genetically different and lack effective therapeutic options. We profiled the genome, exome and transcriptome of 100 refractory metastatic melanoma patients by NGS, and established personalized reports presented at the Charité Tumor Board, guiding treatment for > 1/3 of the enrolled patients (Lamping et al., 2020; Leyvraz et al., 2022). We also investigated the gene regulation networks governing the pathogenesis of uveal metastatic melanoma (mUM). mUMs carry recurrent mutations in the G protein subunit GNAQ/GNA11 and belong to one of two subgroups: disomy 3 (D3, mutated in the splicing factor SF3B1) or monosomy 3 (M3) with loss of BAP1, (Polycomb repressive de-ubiquitinase (PR-DUB) complex), showing the worst prognosis.

We investigated the epigenetic patterns in mUM by integrating transcriptome and whole genome bisulfite sequencing. CpG demethylation rewired the activity of enhancer driving oncogene expression, e.g., MET, BCL2, and RAS activators mediating GNAQ mutations (RASGRP3, SOS1; fig. 1). Additionally, mUM group-specific, differentially methylated regions (DMRs) targeting enhancers (fig. 2) provided a rationale behind the biological and clinical differences in disease sub-groups (Risch et al., to be submitted; in cooperation with S. Hetzel, Meissner department).



Figure 2: Altered methylation in mUM subgroups showing key down- and upregulated genes (blue & red arrows).

Currently, we are validating mUM regulatory networks by investigating the impact of BAP1 loss in cell lines using gene perturbations, ATACseq, and ChIPseq/Cut&Run for key transcription factors and histone marks. Furthermore, we investigate patterns of enhancer-associated transcription (eRNAs, in cooperation with the Mayer group).

Taking advantage of spatially-resolved imaging mass cytometry (IMC; Hyperion, Fluidigm), we explore the melanoma tumor architecture and immune microenvironment. With our newly developed image analysis tool, we link deconvolved bulk RNAseq data with pathology (MICCRA).





Molecular pathways in pediatric cancers

Pediatric cancers originate from poorly differentiated progenitor cells (e.g., mesenchymal progenitors in Ewing Sarcoma or lymphoid progenitors in acute lymphoblastic leukemia) and display specific genomic landscapes and frequent chromosome rearrangements, which trigger oncogenic gene fusions (e.g., EWSR1FLI1 fusion in Ewing sarcoma). Although advanced generic therapy protocols have significantly improved survival over the last decades, a fraction of tumors remains refractory with poor prognosis. We are working on several projects with the aim to better understand the molecular basis of relapse and drug resistance.

1. Relapse acute lymphoblastic leukemia (ALL)

ALL is the most common type of cancer in children and occurs in various forms with different etiology and severity. Our earlier work identified BCL2 as clinical target in the rare, incurable ALL with t(17;19) translocation fusing TCF3 and z genes (Fischer et al., Nature Genet, 2015). We have shown that the TCF3-HLF fusion rewires enhancers driving self-renewal properties, and confer EP 300 sensitivity (Huang et al. 2019). Now, we initiated two new studies on B-cell precursor ALLs to identify pathway signatures associated with relapse and metastasis niche by harnessing whole-exome, bulk and single seq RNA-seq (in cooperation with C. Eckert, Charité and J.-P. Bourquin, Children Hospital, Zurich, Switzerland).

- a. ALL with t(1;19) / TCF3-PBX1 fusion with good outcome, except for a fraction of patients with dismal prognosis.
- b. Relapse ALL metastasizing to the testis
- 2. CROPseq in pediatric tumors: perturbation biology at single cell resolution

We harness genetic perturbation methods using CRISPR to better understand the topology and regulation mechanisms of signaling networks in cancer cells, and to identify therapies targeting non-druggable mutations by exploiting synthetic lethality. The CROPseq method uses a plasmid designed to make the sgRNA readable in single cell seq data. This technique addresses the shortcomings of classical gene perturbation experiments and enables us to measure the effect of many perturbations in parallel (fig.4, https://ipc-project.eu).



Figure 4: CROPseq (Datlinger et al., 2017) with guide spacer transcribed into polyA RNA captured with mRNAseq.

CROPseq on cancer cell lines

We established and optimized experimental and bioinformatic CROPseq workflows to analyze CAS-9 expressing cell lines derived from Ewing sarcoma, hepatoblastoma and ALLs. We designed dedicated gRNAs libraries with 200 guides targeting 60 genes (selected from expert knowledge and insights and mined from large dropout screens).

CROPseq results on Ewing's sarcoma cell lines

The analyses are ongoing for all three pediatric tumor types. Result examples for a transduced CROPseq library in A673 Ewing's sarcoma cell line are shown in figure 5. Functional dependencies between different targeted genes are revealed by comparing transcription profiles of different CRISPR-guided targets (fig. 6). For instance, STAG2 knockout affects cell-cell interactions components.



Figure 5: Guide coverage in library (top). Guides detected in 7171/15408 cells (47%) in single cell RNAseq (bottom).



Figure 6: Functional modules identified in CROPseqmediated gene knock-downs.

In the future, we plan to perform gene-drug interaction screens for generating phenotypic
readouts by feeding in silico models of drug response which are established for selected
pediatric cancer types.

3TR Innovative Medicine Initiatives Project

We started a project that addresses the spatial architecture of immune cells which correlate with treatment response in autoimmune diseases (e.g., Lupus Erythematosus) using Imaging Mass Spectrometry.





REINERT LAB

Efficient Algorithms for Omics Data

ESTABLISHED IN JULY 2014

PROF. DR.-ING. KNUT REINERT knut.reinert@fu-berlin.de +49 30 838 75 / 222 ↘ Reinert Lab



Max Planck Fellow Group

Focus areas

- Data structures and parallelization
- Read mapping and variant calling
- (Pan-) Genome comparison and metagenomics
- Genomic RNA analysis

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 achieves this by working to solve well-posed, new as well as existing computational problems arising in the context of omics data analysis. The concept is that we always consider the complete analysis cycle. The inception of algorithms, their efficient implementation, using modern multi-threading and vectorization, and their application to relevant problems is done in collaboration with experimental scientists at the MPI for Molecular Genetics and the Freie Universität Berlin.

Scientific highlights

The resulting metagenomics tool was evaluated very favorably in the CAMI challenge. Meyer et al., Nature Methods, 2022

We acquired funding and devised new data structure to distribute approximate string queries. Seiler et al., IScience, 2021

We used a bidirectional index to compute genome mappability very quickly. *Pockrandt et al., Bioinformatics (Oxford, England), 2020*

Scientific honors

• Member of the "International Society for Computational Biology"

Research projects

The DREAM framework

"Raptor: A fast and space-efficient pre-filter for querying very large collections of nucleotide sequences" and "Hierarchical interleaved Bloom filter: Enabling ultrafast, approximate sequence queries"

The recent improvements of sequencing technologies, commonly subsumed under the term NGS (Next-Generation Sequencing) or 3rd (and 4th) generation sequencing, have triggered incredibly innovative diagnoses and treatments in biomedicine, but also tremendously increased the sequencing throughput. The costs for producing one million base pairs could be reduced from hundreds of thousands of dollars to a few cents. As a result of this development, the number of new data submissions has grown dramatically and is expected to continue to increase, faster than the cost per capacity of storage devices will decrease. This poses challenges for existing sequence analysis pipelines. They are usually designed to run on a few recent samples but cannot be used in reasonable time for *many* samples, let alone for reanalyzing existing data.

The most basic task of such pipelines is to (approximately) search sequencing reads or short sequence patterns like genes in large reference data sets. Our group introduced a data structure in 2018, the Interleaved Bloom Filter (IBF). Based on this, the application Raptor extends the data structure with winnowing minimizers and probabilistic thresholding. Raptor's performance is currently state-of-the-art but the method has two limitations: Firstly, the runtime deteriorates when the number of samples (called colors or user bins) becomes large. This is the case for metagenomics data, large collections of RNA-seq files, or indexing k-mers in contigs of a deBruijn graph, among others. Secondly, the size of the index depends on the maximum sample size in the collection. As a result, the method uses more space than necessary when the samples are unevenly sized.

Based on the IBF we introduced a new data structure, the Hierarchical Interleaved Bloom Filter (HIBF) that overcomes the main limitations of the IBF data structure. The HIBF successfully decouples the user input from the internal representation and can handle unbalanced size distributions and millions of bins. This is achieved by computing a hierarchical layout that considers the sequence similarity between bins and automatically sets internal parameters (fig. 1).



Figure 1: An HIBF Layout computed on 20 user bins (or colors) with tmax= 5. Shown on the top is the input of 20 user bins (UB) sorted in decreasing order by their size (k-mer content). Below is the HIBF layout with 4 levels.

The resulting data structure can index for example 98.8 GiB of RefSeq data divided in about 26000 user bins in less than 12 minutes and is orders of magnitudes faster than the competitors Mantis and Bifrost, both in building and querying (up to 960 and 230 times, respectively). In addition, the index can be built in only 45 GiB of space using minimizers, which is 10 less than the size used by Mantis. The approximate search is up to 330 and 75 times faster than Bifrost and Mantis. Several PhD students will work on further developing this framework to allow analyses of Terabytes of sequencing databases.

Quantitative filters

"Needle: A fast and space-efficient prefilter for estimating the quantification of very large collections of expression experiments"

As a non-obvious extension of the research on the IBF and HIBF described above, we also investigated, how we can filter quantitative information. Tools like Raptor, Mantis or Bifrost can only answer simple membership queries, with no ability to quantify found transcripts. Traditionally, quantification involves a computationally expensive alignment step, where the reads of one RNA-sequencing (RNA-seq) experiment are aligned to a transcriptome to measure the expression of e.g., STAR. Lately, alternative approaches have been proposed, where this alignment step was replaced by faster methods such as analyzing the k-mers of a transcript (e.g., Sailfish) or using pseudoalignments (e.g., the tools kallisto and Salmon). All these approaches use an EM algorithm to handle ambiguity of reads or k-mers and a transcriptome to determine relevant k-mers beforehand. Although these developments have been a considerable improvement with respect to exact alignments, analyzing a data set of thousands of experiments is still too costly.

REINDEER was the first tool to not only store representative k-mers for all given experiments, but also how often they occur. Like REINDEER, the recently published Gazelle stores k-mers and their counts, but unlike REINDEER, it always performs a log-transform of the count values to save space.

We developed Needle, a tool for the semi-quantitative analysis of large collections of expression experiments. Needle is based on two ideas: Firstly, it uses the Interleaved Bloom Filter (IBF) with minimizers instead of contiguously overlapping k-mers to efficiently index and query these experiments. Secondly, rather than storing the exact raw count value of every minimizer, Needle splits the count values of one experiment into meaningful buckets and stores each bucket as one IBF (Fig. 2). We show how this discretization can efficiently and accurately approximate the expression of given transcripts for all given files at once. Due to the efficiency of the IBF, Needle can build the index 3-54 times faster than REIN-DEER and the count values can be obtained 16-100 times faster while only using 3-39% of the space required by REINDEER, where the speed advantage depends on the chosen minimizer window size. A direct comparison to Gazelle was not possible as the tool is not publicly available yet. But based on the provided analysis of Gazelle, we are quite confident that Needle outperforms Gazelle as well.

Construction







VALLIER LAB

Liver organogenesis

ESTABLISHED IN SEPTEMBER 2022

PROF. DR. LUDOVIC VALLIER vallier@molgen.mpg.de +49 172 3273244

Max Planck Fellow Group

Focus areas

- In vitro models: hiPSCs and organoids
- Liver development: hepatocyte maturation and biliary tree formation
- Liver disease: NAFLD/NASH
- Regenerative medicine: cell therapy and tissue repair

Research concept

The overarching goal of the Vallier group based at the MPIMG, is to understand liver organogenesis and to use the resulting knowledge for the development of new therapies. For that, we combine hiPSCs, primary organoids, gain and loss of function approaches and single cell analyses to study human liver development *in vitro*. Of particular interest, we aim to uncover (i) the molecular mechanisms controlling the specification of hepatocytes and cholangiocytes and (ii) the factors driving the functional maturation of the same cell types. We will then exploit this knowledge to generate cells for disease modelling and for cell-based therapy. This translational activity is currently located in the Vallier lab based at the Berlin Institute of Health Centre for Regenerative Therapy.

Scientific highlights

Single cell map of human liver development reveals factors driving functional maturation of hepatocytes. Wesley at al., Nature Cell Biology, 2022

Cholangiocytes from different regions of the biliary tree are interchangeable for cell therapy. Sampaziotis et al., Science, 2021

134

Disease modelling using hiPSCs reveal the function of PNPLA3 in NAFLD. *Tilson et al., Hepatology, 2021*

Genetic background leads to variability between pluripotent stem cells. *Ortmann et al., Cell Stem Cell, 2020*

Scientific honors and selected invited talks

- W3 Einstein strategic Professorship 2022 (Berlin)
- Max Planck Fellow 2022 (Berlin)
- Fellow of the Academy of Medical Science 2020 (UK)
- Fellow of St. Edmund's College 2018 (Cambridge, UK)
- Keynote speaker at the "FASEB liver meeting" (virtual, 2021)

Research projects

A new platform to study human liver development

The sequential events of differentiation that ensure the emergence of endoderm, foregut, liver bud and hepatic organogenesis have been extensively studied in animal models. The resulting knowledge has been exploited to develop protocols for differentiating human induced Pluripotent Stem Cells (hiPSCs) into a diversity of hepatic cell types. However, this knowledge-based approach has been more challenging to apply to produce fully functional cells, especially hepatocytes. This limitation is due in part to a lack of understanding the mechanisms driving organ maturation in human. To address this limitation, we have generated a single cell-map of the developing human liver (Wesley and Ross, Nature Cell Biology, 2022; fig.1). This analysis identified factors controlling liver development while allowing the establishment of culture conditions for the derivation of human hepatoblast organoids (HBOs; fig.1). Hepatoblasts represent the "natural" stem cell of the developing liver since they can differentiate into hepatocytes and cholangiocytes. Thus, HBOs provide a novel platform to study liver organogenesis *in vitro*.



Figure 1: Single cell map of human liver development and derivation of primary HBO (Hepatoblast Organoids). Right: UMAP analyses of the developing human liver. Left: Bright field image of human HBOs and immunostaining showing expression of hepatoblast markers in HBOs. (Carola Morell; Wesley et al., Nature Cell Biology, 2022).

Biliary tree development

The biliary tree is a network of conduits which transports the bile generated in the liver. Cholangiocytes represent the main cell type of the biliary epithelium and they have two different embryonic origins. Cholangiocytes located in the extrahepatic part of the biliary tree originate from a pancreaticobiliary progenitor while intrahepatic cholangiocytes originate from hepatoblasts. The impact of this divergence on cholangiocytes function remains to be fully defined as well as the mechanisms by which the intra- and extrahepatic biliary system connect during early development. To address these questions, we recently performed single cell RNA-Seq on primary human intra- and extrahepatic cholangiocytes (Sampaziotis et al., 2021). Interestingly, cholangiocytes display a strong regional identity in vivo. However, such difference is lost when the same cells are grown as organoids in vitro (fig.2). Furthermore, transplantation experiments in mouse models have shown that intra- and extrahepatic cholangiocytes are interchangeable. Thus, the transcriptomic identity of cholangiocytes is mainly imposed by their environment and not programmed by their embryonic origin. However, functional differences remain between cholangiocytes. Indeed, cholangiocytes isolated from the intrahepatic regions have the unique capacity to transdifferentiate into cells expressing hepatocytes markers and to form tube networks in vitro (Rimland et al., 2021; Roos et al., 2022). Thus, these results show that the developmental origin might not be important for cholangiocytes function but could have a major impact in their role in liver regeneration.



Figure 2: Single cell comparison of primary cholangiocytes located in different regions of the biliary tree and grown as organoids in vitro. (PRI = Primary; ORG = Cholangiocytes organoids, BTO = Cholangiocytes organoids grown in the presence of bile acid; IHD=Intrahepatic duct, CBD=Common bile duct; GB=Gallbladder; from Sampaziotis et al., Science 2021)

Improving single cell map of human liver development

In collaboration with the Human Cell Atlas consortium, we are currently improving the resolution of our single cell map of human liver development. For that, we are performing detailed time course analyses on human liver development using single nuclei multi-omics (RNA-Seq and ATAC-Seq; fig. 3). Ongoing analyses have identified factors and signalling pathways that are potentially involved in differentiation of hepatocytes/cholangiocytes, and in functional maturation of hepatocytes. Future work will focus on validating the importance of these factors using hiPSCs derived hepatocytes and hiPSCs derived HBOs (see below). These experiments will be combined with functional validations *in vivo* in mouse models, taking advantage of the unique knowledge available at the MPIMG. We also plan detailed analyses of the molecular mechanisms driving cellular maturation, especially in the interplay of transcription, epigenome and cell cycle progression.



Figure 3: Single map of human liver development. Single nuclei were collected from foetal liver aged 5 pcw to 18 pcw. The resulting nuclei were then processed for RNA-Seq and ATAC-Seq using the 10X multi-omics platform (Carola Morel, Arash Shahsavari, Irina Mohorianu)

Derivation of hepatoblasts from hiPSCs

The comparison of the developmental trajectory of primary hepatocytes and of hepatocytes generated *in vitro* have revealed specific divergences. This difference has major practical consequences since the derivation of HBOs from hiPSCs remains challenging. Consequently, we are currently developing culture conditions to derive HBOs *in vitro* (fig. 4). Several projects are ongoing (i) to further characterize the resulting cells and compare them with primary HBOs, (ii) to use these cells for investigating the function of potential maturation factors *in vitro*, and (iii) to define the importance of proliferation in the functional maturation of hepatocytes. Future work will focus on the function of WNT signalling in hepatoblast self-renewal. We hypothesize that WNT plays a key role in maintaining the competence of hepatoblasts to differentiate into hepatocytes and cholangiocytes by coordinating key epigenetic marks and cell cycle. Such mechanisms could also be relevant for regenerative processes in the adult liver and ultimately for tumorigenesis.



Figure 4: Hepatoblast organoids generated from hiPSCs. We refined our protocol to generate foregut and hepatic endoderm from hiPSCs. We then identified the best stage of differentiation for growing hiPSCs derived hepatoblast organoids expressing Albumin and HNF4 (Charlotte Grey-Wilson).

Developmental differences between intra- and extrahepatic cholangiocytes

To further understand the development of the biliary tree, we are performing detailed single cell mapping of the human foetal biliary tree in combination with 3D imaging. Our goal is to understand the mechanisms by which the intra- and extrahepatic biliary trees are established and connected. We hypothesize that these mechanisms ultimately establish the functionality of cholangiocytes and their role in regenerative processes during chronic liver disease. Future work will involve detailed epigenetic analyses of cholangiocytes organoids derived from different regions of the biliary tree to define the mechanisms by which their functions is established during development and in adult tissues. These analyses will be done in collaboration with the Meissner lab.

Scientific Services

2



DR. LUDGER HARTMANN hartmann@molgen.mpg.de +49 30 8413 / 1189

Focus areas

- Animal husbandry and breeding management of laboratory mice in compliance with animal welfare standards (in- and outbred strains, genetically modified mice)
- Technical and experimental work with laboratory mice with respect to animal welfare aspects
- Training of personnel in laboratory animal science

Service concept and team

The Animal Facility provides state of the art laboratory animal care and services for the scientific needs of the MPIMG in compliance with the German and European Animal Welfare Guidelines and Regulations. The team currently consists of a veterinarian (head of the facility), an animal care manager and 12 animal technicians (two shared with the Transgenic Unit). All staff members are committed to maintain the highest standards of hygienic quality and ensure that all animal care and operating procedures are conducted in accordance with the highest scientific, humane and ethical principles.

Services of the Animal Facility

- Animal husbandry including regular animal welfare checks
- Breeding of in- and out-bred strains (breeding management for scientists)
- Computerized colony management to ensure transparency, especially of genetically modified mouse strains using PyRAT database software
- Providing animals for the Transgenic Unit (e.g., embryo donors, pseudo-pregnant foster animals, vasectomized male animals)
- Technical and experimental work (e.g., timed matings, biopsies, injections of e.g., hormones, blood and organ collection, ultrasonic testing to check for pregnancy)
- Import and export of mice (national and international)
- Operation of a quarantine station (transfer to the animal house by embryo transfers)



Figure 1: The Animal Facility of the MPIMG

- Ordering animals from vendors
- Maintaining SPF (specific-pathogen-free) hygienic level
- Documentation: daily work reports, genotyping lists, update of PyRAT mouse software, detecting and observing harmful phenotypes, final evaluations, score sheets, protocols within the scope of approvals and progress reports
- Operation of a veterinary home pharmacy
- Operation of a quality management system with TÜV-certified quality managers, based on standard operating procedures

Zebrafish facility: No experiments with fish were conducted in 2018-2019. Therefore, the institute has decided to close this facility at the end of 2019.

The Animal Facility offers a complete service for breeding and keeping mice exclusively for users of the MPIMG. All standard technical work is done by the Animal Facility staff (not a must, but strongly encouraged). There is no free access to the Animal Facility or mouse rooms for users. However, access is possible with a valid scientific justification (e.g., to look at phenotypes in person) and a comprehensive introduction. All work done by a scientist is performed together with experienced Animal Facility staff members. The budget is provided in full by the institute and the research groups are not charged additional fees for the services of the Animal Facility.

Equipment and resources

The animal house is a free-standing, two-level building with a total area of approximately 1.900 m². The area for the laboratory animals (only mice) is 300 m², plus laboratories for the Transgenic Unit, social rooms and areas for cleaning, storage and technical supplies. A quarantine station is located outside the building. The mice are kept under SPF conditions in areas with restricted access. All mouse strains are housed in individually ventilated cages and are handled under sterile conditions (a total of approximately 7.000 cages can be used).

ANIMAL FACILITY





DR. LARS WITTLER wittler@molgen.mpg.de +49 30 8413 / 1453

Focus areas

- Genetically engineered mouse models and animal transgenesis
- Assisted reproductive technologies: microinjection and zygote electroporation
- Morula aggregation and tetraploid complementation
- Mouse strain cryopreservation and rederivation

Service concept and team

Genetically engineered mouse models (GEMMs) are invaluable tools for biomedical research and the molecular life sciences. A high level of expertise, training and routine is required to efficiently execute the methods and technologies to create a GEMM while adhering to best animal welfare practice. The Transgenic Unit is a scientific service group of the MPIMG that provides a centralized platform for these technologies. Currently, the group consists of a staff scientist (head of the facility) and six technicians (tow shared with the Animal Facility). The team of the Transgenic Unit has more than ten years of experience and expertise in mouse embryology, transgenesis, embryonic stem (ES) cell culture, assisted reproductive technologies, mouse husbandry and general animal experimentation.

Services of the Transgenic Unit

- Morula Aggregation Platform (including tetraploid complementation): Morula aggregation is employed as the primary method to generate GEMMs for the MPIMG and is a robust way for using tetraploid complementation to generate embryos derived almost entirely from ES cells
- Microinjection: Standard pronucleus and blastocyst injections, laser assisted morula injection, 2C injection, cell ablation and separation and intracytoplasmic sperm injection
- Zygote electroporation: as an efficient way for the generation of Cas9 based site-directed mutagenesis. Based on this high efficiency, the Transgenic Unit, in cooperation with the Department of Genome Regulation, has established a platform that facilitates the production of knock-out embryos for direct use in phenotypic and multi-omics analyses

- Assisted reproductive technologies (ARTs): central methods for the conservation and long-term storage of mouse strains, e.g., cryopreservation. The use of ARTs is an important component to implement the 3R principles in contemporary mouse husbandry
- Practical training and assistance for MPIMG scientists in transgenic and assisted reproductive technologies
- Legal and administrative support for planning and implementation of projects involving the creation of GEMMs

All services are tailored to the needs of the research groups and are offered exclusively for internal projects at the MPIMG. All tasks related to embryo manipulation and animal work required for GEMM creation are performed by well-trained staff of the Transgenic Unit in close cooperation with the Animal Facility. Each step follows standard operating procedures and is quality controlled to ensure high efficiency and optimal application of animal welfare and 3R principles. The budget is fully provided by the institute and no additional fees are charged to the research groups for the services of the Transgenic Unit

Equipment and resources

The Transgenic Unit includes two microinjection workstations, a zygote electroporation/cell fusion setup and equipment for mouse surgery. In addition, the facility has freezers for ARTs and cryopreservation, incubators, several stereomicroscopes and equipment for the fabrication of injection needles.

The Transgenic Unit is housed in three laboratory rooms in the Animal Facility. Embryo transfers are performed in a laboratory inside the SPF barrier of the Animal Facility. Two animal rooms within the SPF barrier foster pseudo-pregnant females, stud males and vasectomized males. The Transgenic Unit is the main recipient of animals from the wildtype mouse breeding colonies that are maintained in the Animal Facility.




Mass Spectrometry

ESTABLISHED IN 2012

DR. DAVID MEIERHOFER meierhofer@molgen.mpg.de +49 30 8413 / 1567

Focus areas

- (Single cell) Proteomics
- Metabolomics

Service concept and team

The Mass Spectrometry (MS) group of the MPIMG provides sophisticated support in deciphering the molecular mechanisms of various biological samples by integrating multi-omics techniques, including the proteome, metabolome, transcriptome, and epigenome thereby gaining more detailed insights from their experiments. The MS team currently comprises one staff scientist (head of the group), a bioinformatician, a PhD student and one technician.

Services of the Mass Spectrometry group

- Proteomics
- Single Cell Proteomics to tackle proteomes of single cells or posttranslational modifications in a few cells that are morphologically or functionally similar
- Metabolomics for more than 600 metabolites
- Fluxomics and stable isotope tracing to allow a more dynamic view of cell metabolism or to verify the functionality of a metabolic enzyme
- Joint project planning (e.g., feasibility, limitations, number of samples)
- Data processing and visualization
- Support and training for data analysis and statistics
- Support in describing the LC-MS/MS methods for e.g., publications, grant applications and theses
- Establishment and modification of existing and new LC-MS methods based on user requirements

All services can be used with a wide range of sample types, like tissues, cell-culture samples, body fluids (e.g., blood, plasma) or gel bands. To maintain high standards, the final steps of sample preparation (e.g., protein digestion) and the liquid chromatography MS (LC-MS) instruments are operated exclusively by the service staff. The LC-MS equipment is maintained and serviced by the staff, and there are no service contracts. The group routinely calibrates the mass spectrometers and verifies good operating conditions by performing quality control measurements prior to each LC-MS/MS project. The services of the MS facility are offered to primarily all groups of the MPIMG and the budget is provided in full by the institute. The research groups are not charged additional fees for the services of the MS facility.

In a dedicated research project, the MS group aims to decipher molecular mechanisms in renal cell carcinomas (RCC), specifically in malignant chromophobe and papillary RCC and benign renal oncocytomas. The research strategy integrates proteome and metabolome profiling to study dysregulations between the tumors and healthy control tissues on the molecular level.

Equipment and resources

The following LC-MS/MS systems are available in-house and are operated by the MS team:

- QTrap 6500 (Sciex): simultaneous quantitation and confirmation of known peptides or metabolites
- Q Exactive HF (Thermo Scientific): Proteomics and the measurements of sub-proteomes by isolating specific organelles or cell fractions or by the enrichment of specific post-translational modifications
- UltiMate 3000 RSLCnano: liquid chromatography system, online coupled with the MS system; proteome profiling and relative quantification (label-free, SILAC)
- TimsTOF SCP (Bruker Daltonics): designed for quantitative deep single-cell 4D-Proteomics, parallel accumulation serial fragmentation acquisition methods with extremely high speed (120 Hz) and sensitivity

The premises of the MS group are located on the ground floor of tower 4.





DR. CLAUDIA GIESECKE-THIEL giesecke@molgen.mpg.de +49 30 8413 1544

Focus areas

- Multidimensional flow cytometry and high-end cell sorting
- Method development to improve measurement reliability

Service concept and team

Sorting cells according to specific characteristics is often prerequisite for research questions pursued at the MPIMG. The team of the Flow Cytometry group supports researchers in the operation of high-end cell sorting systems. The group currently consists of three people, a facility head (part-time position) and two technical assistants (one part-time and one full-time position).

Services of the Flow Cytometry group

- Full-service cell sorting of rare and/or sensitive populations from heterogeneous samples using FACS
- Review and discussion of project plans, staining panels, and support in selection of fluorescent protein construct for target cell properties
- Discussion of analysis strategies and data interpretation
- Provision of protocols
- Adjust instrument settings to the needs of the researchers
- Development of novel flow cytometric techniques to meet the diverse scientific needs of the institute and continuously improve the quality of measurements
- Maintenance of cell sorters and analyzer

To improve sorting results and reduce the number of experimental animals, the staff is constantly seeking ways to improve the specificity and efficiency of live-dead cell staining in combination with cell lines, cell types, and fluorescent proteins and to enhance sorting efficiency through target cell enrichment and debris depletion.

The Flow Cytometry group offers operator-assisted cell sorting by FACS and flow cytometric analyses to researchers of the MPIMG. Self-use of our cell



sorters is also possible for qualified and trained in-house users. The facility is fully funded by the institute budget. No booking fees are charged for the use of the flow cytometer and cell sorters, for training or other services.

Equipment and resources

The Flow Cytometry operates

- a BD FACSAriall Special Order System (SORP): in operation since December 2008, upgraded with a yellow-green laser in August 2009
- a BD FACSAria Fusion: in operation since March 2018 and re-configured in August 2018 (the original BD standard configuration was unsuitable for in-house needs for simultaneous analysis and sorting of CFP and GFP or YFP (double) positive cells)
- a BD Accuri RUO SORP CSampler

Both the BD FACSAriaTMII SORP and the BD FACSAriaTM Fusion are used very frequently by researchers from all departments and research groups.

The facility is located in a room on the ground floor of tower 4, which houses all the equipment.





Microscopy & Cryo-electron Microscopy

ESTABLISHED IN JANUARY 2013 BY FUSING THE FORMER MICROSCOPY SERVICE GROUP, ESTABLISHED IN 1978, AND THE CRYO-ELECTRON MICROSCOPY GROUP, ESTABLISHED IN 2004

DR. THORSTEN MIELKE mielke@molgen.mpg.de +49 30 8413 / 1644

Focus areas

- State-of-the art light microscopy imaging techniques
- Room temperature and cryo-TEM
- Technical support and user training
- Image analysis, statistical analysis and data interpretation

Service concept and team

The Microscopy and Cryo-Electron Microscopy Service Group provides a broad range of imaging techniques combining both, light microscopy (LM) and transmission electron microscopy (TEM). The facility supports the research groups in planning, performing and analyzing their imaging tasks. The service offered is a highly interactive workflow development process that includes experiment design, sample preparation, data acquisition as well as image analysis solutions, elaborate processing pipelines, automation, and data visualization and interpretation. All these steps are critical for extracting and quantifying meaningful biological information from complex 2D, 3D and higher dimensional data sets. Currently, the team consists of one staff scientist (head of the group), one post-doctoral fellow and two technicians.

Services of the Microscopy and Cryo-Electron Microscopy Service Group

- Light microscopy: wide-field, high content, confocal, STED, FLIM, lightsheet and holotomography imaging
- Automated 2D, 3D and higher-dimensional imaging
- Conventional room temperature TEM (plastic embedding, ultrathin-sectioning, Immune-EM)
- Single particle cryo-TEM (negative-stain screening, vitrification, cryo-screening, automated data collection)
- Technical support and maintenance of the institute's microscopy infrastructure
- Basic and advanced user training



- Application support and method implementation
- Image analysis, statistical analysis, data interpretation and workflow development

Users are trained to operate the light microscopes themselves. In contrast, TEM is offered as a full service due to the technical complexity and lack of standardization of TEM sample preparation and imaging. Samples are processed and imaged by members of the service group. However, all steps of TEM analysis are performed in close collaboration with the scientists.

As a service group, the team is constantly monitoring new developments, both in-house and in the field, to identify new needs and implement new imaging methods and. The budget of the microscopy group is largely covered by the institute and the research groups are not charged booking fees for the use of the microscopes.

Equipment and resources

The facility has 13 advanced light microscope systems (ten operated by the service group and three by other departments). These include

- five wide-field epifluorescence microscopes (Z1 Imager, 2x Z1 Observer, 1x Observer 7, V16 Stereo Zoom)
- four confocal systems (LSM700, 2 LSM880 (one with 2-photon excitation), Leica Stellaris Tau-STED)
- two automated screening platforms (CellDiscoverer 7, Cellomics Arrayscan VTI), one light sheet microscope (Z1 Lightsheet) and one holotomography system (Nanolive CX-A)

All microscopes function as shared instruments. The staff also operates three transmission electron microscopes (Philips CM100, 120 kV Tecnai G2 Spirit, 300 kV FEG Tecnai G2 Polara cryo-TEM with a Gatan k2 summit DED) and a wide range of sample preparation equipment

(2x Vitrobot plunge freezers, UC7 ultramicrotome with FC7 cryo-stage, VT1200S vibratome, coating systems and plasma cleaners) which are available upon training.

The group has four rooms for light microscopy, dedicated TEM rooms, two sample preparation labs and offices for staff and processing workstations. Once the renovation of tower 1 is complete, the group will centralize all light microscopes together with offices and sample preparation labs into newly built and specially designed group rooms on the ground floor of tower 1. This area will have approximately the same surface area, but will partially be equipped with special high precision air conditioning for confocal and STED microscopes.



• • • • • • • • • • • • • • • • • • • •	
	· · · · · · · · · · · · · · · · · · ·
Sequencing	
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 	· · · · · · · · · · · · · · · · · · ·
••••••••••••••••••••••••••••••••••••••	

PROF. DR. ALEXANDER MEISSNER meissner@molgen.mpg.de

DR. HELENE KRETZMER kretzmer@moglen.mpg.de

Dr. Bernd Timmermann, the previous head of the sequencing group, has moved to Roche Diagnostics and Drs. Helene Kretzmer and Alexander Meissner act as interim co-directors until the search for a new head is completed.

Focus areas

- Provision of state-of-the-art NGS technologies
- o Development of new sequencing applications

Service concept

The Sequencing group is prerequisite for a research institution like the MPIMG. The group provides excellent expertise in all next generation sequencing (NGS) technologies and significantly supports MPIMG researchers in their scientific endeavors. The team currently consists of a project leader, a data manager and two technicians and provides help for researchers to process DNA and RNA samples in an efficient and economical manner – from sample preparation to data analysis. The group operates several next generation sequencers and maintains a fully equipped lab. Automation solutions for sample preparation have been established to maximize throughput and minimize variation. To handle the increasing number of samples, a laboratory information management system has been established. Thus, all samples are tracked from delivery to data processing.

Services of the Sequencing team

- Large variety of NGS based applications, e.g., RNA sequencing, genomic sequencing, ChIP sequencing, targeted re-sequencing and methylation analysis
- Project advice and development: discussion of sequencing strategies, protocols, choice of technologies and data analysis
- Preparation of biological material: DNA and RNA preparation for short and long read protocols, preparation of cell solutions for single cell experiments
- Library preparation: large number of different protocols for short and long read sequencing, development of new protocols



Figure 1: A PacBio long read sequencer is loaded.

- Data analysis: e.g., pipelines for transcriptome profiling, whole exome and methylation analysis
- Support with the preparation of manuscripts: data upload to short read archives, providing information for material and methods
- Training for different library preparation protocols

Biological samples (e.g., tissue, DNA/RNA, prepped libraires) are processed by the staff of the sequencing unit. After data processing or additional data analysis, the user will be informed via email. The basic financing principle of the Sequencing group is to refinance sequencing kits and general consumables. To achieve this goal, we have developed a transparent pricing system for all our sequencing services.

Due to our expertise and short turn-around time, we received an official request in 2020 from the Charité / Berlin Institute of Health to collaborate in COVID-19-related projects.

Equipment and resources

We currently provide expertise in two different sequencing technologies: the single molecule, real-time sequencing technology (Pacific Biosciences) and the sequencing-by-synthesis technology (Illumina). At a read length of up to 400,000 bases, the PacBio technology offers great benefit especially for *de novo* genome sequencing, analysis of structural variations, metagenome analysis and full-length transcriptome analysis.

• Sequel II system (PacBio): produce reads with a sequencing length of several hundred kilobases with the continuous long read mode or high-quality single-molecule reads with an accuracy of >99.9 % (HiFi reads) with the circular consensus sequencing method

 O
 O
 O
 O
 O

 O
 O
 O
 O
 O

 O
 O
 O
 O
 O

 O
 O
 O
 O
 O

 O
 O
 O
 O
 O

 O
 O
 O
 O
 O

 O
 O
 O
 O
 O

 O
 O
 O
 O
 O

 O
 O
 O
 O
 O

 O
 O
 O
 O
 O

 O
 O
 O
 O
 O

 O
 O
 O
 O
 O

 O
 O
 O
 O
 O

 O
 O
 O
 O
 O

 O
 O
 O
 O
 O

 O
 O
 O
 O
 O

 O
 O
 O
 O
 O

 O
 O
 O
 O
 O

The high throughput of our Illumina systems completes our sequencing service and provides a real advantage for many applications. Single cell analysis, expression profiling, methylation analysis, copy number analysis as well as protein binding site identification and the analysis of whole genomes benefit greatly from the high throughput of this technology.

- ISeq: system for library testing
- MiSeq and NextSeq 500 system: fast track lane
- NovaSeq 6000: high throughput system

For long read sequencing, we use a Sequel II from Pacific Biosciences and two Minlon devices from Oxford Nanopore.

Currently, the premises are located on the ground floor of tower 4. Once the renovation of tower 1 is complete, the group will move to its new space on the ground floor of tower 1 (expected in spring 2023).





IT Service

ESTABLISHED IN JANUARY 2013 BY FUSING THE FORMER MICROSCOPY SERVICE GROUP, ESTABLISHED IN 1978, AND THE CRYO-ELECTRON MICROSCOPY GROUP, ESTABLISHED IN 2004

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

Focus areas

- Operation and maintenance of the IT infrastructure of the institute
- Hardware and software support
- Data archiving and security

Service concept and team

The IT Service Group is essential for all IT and computational needs of the research groups, the service facilities and the administration of the MPIMG. To enable scientists to concentrate on their research, two engineers (heads of the group), six technicians and a trainee ensure the smooth operation of the IT infrastructure, hardware and software. The group supports about 1,000 computer systems, ranging from server to laptop. Additionally, it hosts about 450 active user accounts. The IT group runs a sizable Linux cluster with extensive storage capacity for scientific computing. It develops procedures and rules covering the usage of the IT infrastructure. Moreover, it helps the bioinformaticians of the institute with developing and integrating applications and data processing pipelines.

Services of the IT Service Group

- Helpdesk for all users
- Hardware management (e.g., computers, printers, presentation technology)
- Data management, data backup and archive
- Data and IT security
- Software management
- Maintenance of internet, intranet, WLAN and email

The use of the hard disk storage is chargeable for all research groups and amounts to 0,67 EUR / GB / year for data that is backed up and 0,07 EUR / GB / year if the data is mirrored.

Equipment and resources

Storage: All scientific data is stored on Linux storage systems. Currently, there are about 5.6 PB of data on about 2,500 disks. All data is kept on RAID6 storage and additionally mirrored onto a secondary storage in another server room. There is a system for daily backup of data where appropriate. The maintenance of software (e.g., for mirror, archive, backup and raid maintenance) has all been developed in-house.

Compute Cluster: The group provides a compute cluster with approximately 32 machines. The standard compute node is set up with 128 cores $(2 \times 32 + HT)$ and has 1 TB RAM. All nodes add up to 2,800 CPU cores and 31 TB RAM. Job scheduling is established with a job queuing software (mxqd), which has been developed in-house. It integrates closely with the overall IT structure, e.g., the global file system name space. In 2022, support for GPU queues was added. Four Nvidia A100 GPU nodes with 378 GB RAM and 32 cores and a few smaller GPU cards are available for the cluster users.

The server systems, including the storage and archive systems and the compute cluster, are located in two spatially separated server rooms. One of these, located 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, in tower 3, contains 30 racks in a warm-aisle containment system, capable of cooling down 450 kW with full redundancy.





Library & Scientific Information Service

SYLVIA ELLIGER library@molgen.mpg.de +49 30 8413 / 1314

Focus areas

- Support in all matters concerning publishing scientific articles
- Preservation and expansion of the library collection (analog and digital media stock)

Service concept and team

The primary goal of the Library & Scientific Information Service is to provide the staff of the MPIMG with access to information in all possible forms and formats related to the respective research areas. The library is a specialized scientific library and offers various methods for researching and obtaining scientific information and gives access to a wide range of electronic resources. The library is developing its services proactively in order to respond to the changing information requirements of the institute's staff. The library is run by one librarian.

Services of the Library & Scientific Information Service

- Reference and bibliographic services
- Maintenance of the publication database (PuRe)
- Maintenance of the library online catalogue (OPAC) and the MPIMG library website
- Interlibrary loan
- Access to electronic information worldwide (e.g., e-books, e-journals, databases)
- Support in the process of publication: Open Access, Creative Commons Licenses, Copyright, Law, APCs (Article Processing Charges)
- Bibliometric analyses (e.g., h-index and impact factor)
- New library/resource discovery system according to Aleph (evaluation in 2023)

A proprietary classification system is used to list books, and an online catalog provides access to all literature held in the library and in the MPIMG departments. The library as well as computing resources (i.e., public computers, digital scanning and editing equipment) are provided free of charge to the staff and guests of the MPIMG.

Equipment and resources

The stock of the library includes 4.500 monographs and periodicals, 15.000 journal volumes (library and magazine) and numerous electronic resources like e-journals, e-books and databases.

Currently, the library is closed to visitors due to construction work. Once the renovation of tower 4 is complete, the library will move to its new premises on the fourth floor of tower 4. The total area of the reopened library will then be 300 m² and will include 18 bookshelves, 14 computer workstations, a study room, the librarian's office and a copy room.



DR. CLAUDIA THUROW +49 30 8413 / 1360

Claudia Thurow, a lawyer and commercial manager of an art foundation and previously head of Legal Department & Technology Transfer at the Max Delbrück Center, Berlin, will take over the administrative lead on November 1, 2022. The former head of administration, Christoph Krukenkamp, moved as administrative managing director to DBFZ (German Biomass Research Center).

Administrative concept

The administration supports the institute in the following areas:

- Human resources
- Accounting and external project funding
- Procurement
- Warehousing
- Common services: welcome office, guest house and postal services

The administration of the MPIMG ensures smooth operations and a stable infrastructure for the institute. In addition to the core administrative tasks, the administration takes care of the financial processing of national and international third-party funding. Furthermore, the scientists are supported in legal questions of technology transfer and patenting. These and many other issues are handled in close cooperation with the Administrative Headquarters of the Max Planck Society.

Selected highlights

Introduction of the virtual workplace (2020)

The COVID-19 pandemic posed major challenges for the MPIMG administration. To ensure continued operation and smooth workflows, the Max Planck Society provided the administrative staff with a modernized "virtual workplace". This enables the administrative staff to remotely access files and process essential administrative tasks.

Introduction of elnvoice (2021)

With this project, the Max Planck Society is implementing the federal e-invoice regulation resulting from the EU Directive 2014/55/EU. In the course of this regulation, all federal authorities and downstream institutions are obliged to receive, process and store invoices in digital form without media discontinuity. All employees of the MPIMG administration have participated in training courses and the conversion of invoice processing from paper to a digital workflow has led to substantial improvements, particularly for employees in accounting and purchasing.





Technical Management & Workshops

DIPL.-ING. ULF BORNEMANN bornemann@molgen.mpg.de +49 30 8413 / 1424

Service concept

The technical management of the MPIMG is responsible for all areas related to building management and technical infrastructure:

- Maintenance and development of the MPIMG's building stock and the technical infrastructure
- Operation and maintenance of the electrical power supply, building automation, cooling and heating systems
- Operation and maintenance of steam generators and water purification
- Supply of media (e.g., gases, liquid nitrogen)
- Support of the scientific departments and research groups during relocations or laboratory reorganizations
- Assistance with the installation of new technical equipment
- Organization of the emergency service (24/7)
- Maintenance of the outdoor facilities

Selected highlights

The greatest challenges for the technical management during the last years were the coordination of the construction measures for the renovation of tower 1 and the maintenance of the institute's critical technical infrastructure during the COVID-19 pandemic.

The refurbishment of tower 1 began in early 2019, and the reopening of the building is planned for the beginning of 2023 after completion of the installation work and commissioning of the complex technical systems engineering. As with tower 2, the construction work on tower 1 includes the complete renewal of the building shell and the structural and technical implementation of a new fire protection concept. All floor plans were adapted to new laboratory utilization concepts and the complete technical media for electrical energy, heating/ cooling, (ultrapure) water, steam, gas and compressed air supply, ventilation and air-conditioning technology, fire alarm technology, communication and building automation were newly installed. New ventilation and smoke extraction systems were installed on the roof of tower 1. Now, the technical equipment provides

four floors with laboratories of various sizes, equipment rooms and offices for the Meissner department and for two departments, whose directors have yet to be appointed. The Imaging and Sequencing service groups will move from tower 4 to the ground floor of tower 1. After the renovation, the library and a canteen reopen on the fourth floor of tower 1.

The renovation of tower 1 had to be carried out under the difficult circumstances of the COVID-19 pandemic and the resulting shortages of personnel at the construction and installation companies. Due to the lack of personnel and the considerable supply problems with materials and equipment, the completion of the building had to be postponed several times.

