2016

Research at a Glance
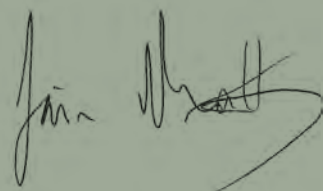
EMBL

Research at a Glance 2016

Contents

EMBL is Europe's flagship laboratory for the life sciences. It was founded in 1974 by its member states as an intergovernmental organisation to promote the molecular life sciences in Europe and beyond.

EMBL pursues cutting-edge research across its five sites in Heidelberg, Grenoble, Hamburg, Hinxton and Monterotondo. The Laboratory's contribution to the European life sciences, however, extends beyond its research mission. EMBL is a provider of world-class research infrastructure and services for the life sciences and a centre of excellence for advanced training, which has over the course of its history helped launch the careers of several thousand life scientists. EMBL is broadly engaged in technology development, and drives innovation through a successful technology transfer programme, allowing scientists and society at large to benefit from its inventions and discoveries. Finally, as Europe's only intergovernmental laboratory in the life sciences, EMBL plays a leading role in the integration of science initiatives and helps shape European science policy and strategy.

The synergy and integration of these diverse activities at EMBL is complemented by a variety of successful organisational principles, including international recruitment of the most talented scientists, regular staff turnover, and rigorous peer review of the Laboratory's activities. The resulting scientific excellence, along with a culture that promotes intellectual freedom and flexibility, create a vibrant environment that offers unmatched opportunities to young creative scientists.

Research at a Glance provides a concise overview of the work of EMBL's research groups and core facilities, which address some of the most challenging questions in the molecular life sciences. The overarching goal of the Laboratory's research is to comprehensively understand the underlying principles and mechanisms of living systems by navigating across scales of biological organisation – from single molecules, to cells and tissues, to entire organisms. Research is pursued in a truly interdisciplinary and collaborative fashion – a particularly distinctive feature of EMBL – whereby researchers with complementary expertise from different disciplines work together to tackle specific biological problems. Importantly, the skills that scientists develop in the dynamic, interdisciplinary and international EMBL environment are exported to our member states when they leave to assume key positions in other institutes.

The critical mass of expertise and resources concentrated at EMBL has produced many important achievements. EMBL is continuously ranked as one of the top research institutions worldwide based on the quality and impact of its scientific publications. EMBL's scientific excellence, its attractiveness to talented young scientists, and the continued support from our member states are a testimony to the Laboratory's success, and drive us to continue pushing the forefront of life science research in the future.

Iain Mattaj
EMBL Director General

European Molecular Biology Laboratory

Research topics

Columns (left to right):

Directors' Research: Matthias Hentze, Maria Leptin

Cell Biology and Biophysics: Alba Diz-Muñoz, Jan Ellenberg, Darren Gilmour, Christian Häring, Lars Hufnagel, Martin Jechlinger, Péter Lénárt, François Nédélec, Pierre Neveu, Rainer Pepperkok, Robert Prevedel, Jonas Ries, Carsten Schultz, Yannick Schwab

Developmental Biology: Detlev Arendt, Alexander Aulehla, Stefano de Renzis, Anne Ephrussi, Marcus Heisler, Takashi Hiiragi, Francesca Peri

Genome Biology: Eileen Furlong, Wolfgang Huber, Maja Köhn, Jan Korbel, Christoph Merten, Kyung-Min Noh, Mikhail Savitski, Lars Steinmetz, Athanasios Typas

Structural and Computational Biology: Theodore Alexandrov, Orsolya Barabas, Martin Beck, Peer Bork, John Briggs, Anne-Claude Gavin, Toby Gibson, Janosch Hennig, Edward Lemke

Research topic rows (left labels):

- Cell signalling and cell differentiation
- Cellular organisation and dynamics, and cell division
- Chemistry and chemical biology
- Computational genomics and metagenomics
- Computational modelling of biological systems and processes
- Disease mechanisms, pathogens, molecular medicine, stem cells
- Evolution
- Functional genomics, genetics and gene networks
- Gene regulation, transcription, chromatin and epigenetics
- Imaging and image analysis
- Macromolecular complexes, interaction networks
- Neurobiology
- Physics and biophysics
- Plant biology
- Proteomics
- RNA metabolism, transport and processing, ncRNAs and miRNAs
- Robotics and automation, engineering
- Software development and bioinformatics
- Tissue morphogenesis, cell polarity and migration
- X-ray crystallography, NMR, electron microscopy

Research group leaders grouped by unit and the research topics they work on:

- European Bioinformatics Institute — pages 78–93
- Structural Biology — pages 102–111
- Structural Biology — pages 112–121
- Mouse Biology — pages 122–129

Research topic	Christoph Müller	Kiran Patil	Carsten Sachse	Judith Zaugg	Georg Zeller	Alex Bateman	Pedro Beltrão	Ewan Birney	Alvis Brazma	Anton Enright	Paul Flicek	Moritz Gerstung	Nick Goldman	Gerard Kleywegt	John Marioni	Oliver Stegle	Christoph Steinbeck	Janet Thornton	Imre Berger	Florent Cipriani	Stephen Cusack	Marco Marcia	José Márquez	Andrew McCarthy	Daniel Panne	Christiane Schaffitzel	Stefan Fiedler	Victor Lamzin	Christian Löw	Rob Meijers	Thomas Schneider	Dmitri Svergun	Matthias Wilmanns	Hiroki Asari	Philip Avner	Cornelius Gross	Paul Heppenstall	Christoph Lancrin
Cell signalling and cell differentiation							■																		■								■		■			■
Cellular organisation and dynamics, and cell division	■		■																		■					■			■	■			■					
Chemistry and chemical biology																	■	■						■				■	■	■								
Computational genomics and metagenomics				■	■		■	■	■		■	■	■		■	■																						
Computational modelling of biological systems and processes		■		■	■				■		■	■			■	■		■										■						■				
Disease mechanisms, pathogens, molecular medicine, stem cells			■		■							■				■				■	■				■					■			■	■	■	■	■	
Evolution						■	■				■	■	■		■			■															■			■		
Functional genomics, genetics and gene networks		■				■	■	■	■	■		■				■																						■
Gene regulation, transcription, chromatin and epigenetics	■			■					■		■				■	■									■								■					
Imaging and image analysis	■		■																							■	■						■					■
Macromolecular complexes, interaction networks	■					■	■			■	■							■	■		■	■			■	■	■	■	■	■	■	■						
Neurobiology																							■						■					■		■	■	
Physics and biophysics	■		■																					■				■	■	■	■	■	■					
Plant biology																								■														
Proteomics							■		■	■								■	■						■	■	■					■						
RNA metabolism, transport and processing, ncRNAs and miRNAs						■				■											■	■			■													
Robotics and automation, engineering																			■	■				■			■			■								
Software development and bioinformatics	■	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■	■									■			■		■						
Tissue morphogenesis, cell polarity and migration																								■					■									■
X-ray crystallography, NMR, electron microscopy	■		■																■	■	■	■	■		■		■	■	■	■			■					

About EMBL

The European Molecular Biology Laboratory (EMBL) is a world-class international research organisation, with some 85 independent groups covering the spectrum of molecular biology. Scientists represent disciplines including biology, chemistry, physics and computer science, working across the laboratory's five sites.

Europe's flagship laboratory for the life sciences

EMBL was founded in 1974 to create a central European laboratory in the emerging field of molecular biology.

It remains the only intergovernmental research organisation in Europe that performs research in the molecular life sciences, and is directly supported by 22 member states, two associate members outside of Europe, and four prospect member states. EMBL's goals are:

- Undertaking outstanding life science research: setting trends and pushing the limits of technology.

- Providing world-class research infrastructure and services to the member states.

- Training and inspiring the next generation of scientific leaders.

- Driving research, innovation and progress through technology development, interactions with industry and technology transfer.

- Taking a leading role in the integration of life science research in Europe.

The European Bioinformatics Institute (EMBL-EBI) is located on the Wellcome Trust Genome Campus in Hinxton, near Cambridge. As a European hub for biomolecular data, EMBL-EBI offers the scientific community access to a variety of bioinformatics services, alongside which a number of active research groups work in areas that complement and extend these services.

EMBL Hamburg develops novel, innovative technologies in structural biology, such as high-throughput crystallisation and data interpretation software, as well as operating cutting-edge synchrotron radiation beamlines and offering world-leading facilities and expertise to the research community. It also has an ambitious research programme for structures of multifunctional proteins and protein complexes of biomedical relevance.

EMBL Heidelberg is home to five research units, central scientific services, the administration, and the laboratory's technology transfer arm, EMBL Enterprise Management (EMBLEM). Heidelberg is the largest centre for biomedical research in Germany and there are many bilateral links between EMBL scientists and local research institutions.

Hinxton

Hamburg

Heidelberg

Grenoble

Monterotondo

EMBL Monterotondo, near Rome, focuses on mouse genetics and functional genomics, and offers expertise in mammalian physiology and production of mouse models of human diseases. Researchers form dynamic partnerships with other international research and clinical centres. The outstation shares a campus with Italian national research groups (IBC-CNR) and the headquarters of the the European Mouse Mutant Archive.

EMBL Grenoble builds and operates beamlines for macromolecular crystallography, develops instrumentation and techniques, and provides facilities and expertise to visitors in collaboration with its campus partners, the European Synchrotron Radiation Facility (ERSF) and the Institut Laue-Langevin (ILL). The outstation is also part of the Unit of Virus Host Cell Interactions (UVHCI).

European Molecular Biology Laboratory

Career Opportunities

Across EMBL's five sites there are opportunites spanning the spectrum of life science research for PhD students, postdoctoral fellows, group leaders, and many other professionals, from software developers to chemists and engineers.

PhD programme

Training is one of EMBL's core missions and our International PhD Programme is renowned for offering excellent education to prospective scientists.

Research independence, dedicated mentoring and an international environment are the cornerstones of the programme, in which around 230 students from all over the world are currently enrolled.

Students have the opportunity to obtain joint PhD degrees between EMBL and one of its partner universities or from a recognised university of their choice.

EMBL recruits PhD students twice a year.
For more details please contact predocs@embl.de.

www.embl.de/training/eipp

Postdoctoral fellows

Postdoctoral fellows at EMBL benefit from the expertise of world class scientists, state-of-the-art scientific equipment, training in career development and an excellent seminar programme.

Our research groups encourage a balance between senior and young scientists, creating the ideal environment to share and discuss research endeavours while supporting junior colleagues to develop and grow into new positions.

The EMBL Interdisciplinary Postdocs (EIPOD) programme builds on highly interactive research between units and is aimed at candidates whose research crosses scientific boundaries.

Please contact group leaders directly to find out if a position is available, or visit www.embl.de/jobs.

An international and inclusive workplace offering equal opportunities

EMBL is committed to creating and maintaining an environment that values and supports diversity and gender equality. EMBL's staff comprises more than 1700 people from more than 60 different countries. In 2015 43% of staff members and 51% of the new PhD students starting at EMBL were women.

EMBL's primary mission is to promote excellence in the life sciences: to achieve this goal EMBL depends on hiring and training outstanding research, service, and administrative personnel. EMBL aims to have gender balance among its personnel in order to maximize the diversity of experiences, perspectives, and skills that are required to conduct world-class science, and ensure that EMBL serves as a model for research organisations across Europe.
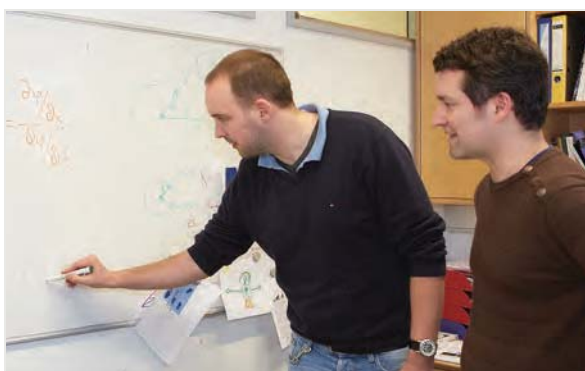
Group and team leaders

EMBL fosters the pursuit of ambitious and long-term research projects at the highest level. Group and team leaders have the freedom to set their own scientific directions and are encouraged to explore the most challenging research areas.

Support for team and group leaders includes funding for a number of staff, and laboratory space with equipment. Research collaborations between groups are an integral part of EMBL's scientific culture.

In addition to advanced scientific development, EMBL offers vocational training to improve skills in areas such as coaching, team management and communication. Establishing a good work-life balance is emphasised at every career stage.

Other careers

EMBL has ongoing opportunities for physicists, computer scientists and electronic engineers, especially early in their careers. Ever-more sophisticated analysis of very large data sets at the European Bioinformatics Institute (EMBL-EBI) draws on a skilled workforce from many disciplines: from scientific expertise in the life sciences to technical know-how in software development. Similarly, qualified technical staff are highly sought after to operate beamlines at EMBL's outstations in Hamburg and Grenoble.

Other positions include interface development, communications, user support, industry liason and training. We offer advanced scientific development and vocational training to improve skills in areas such as coaching and communication.
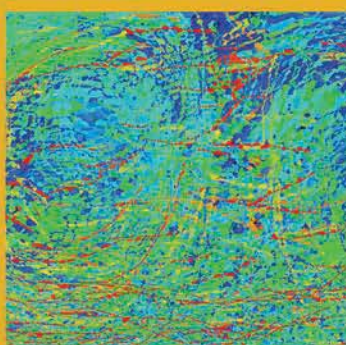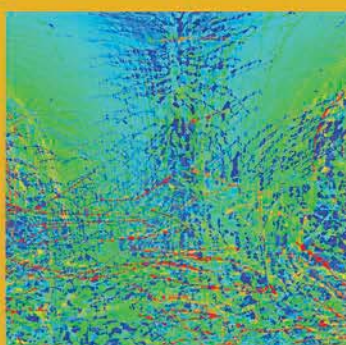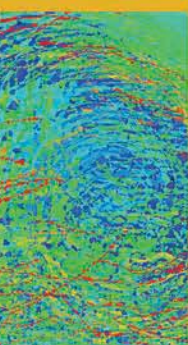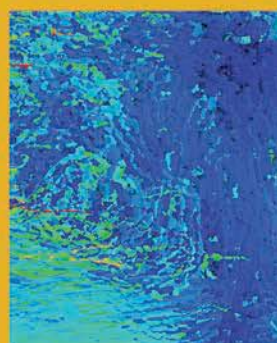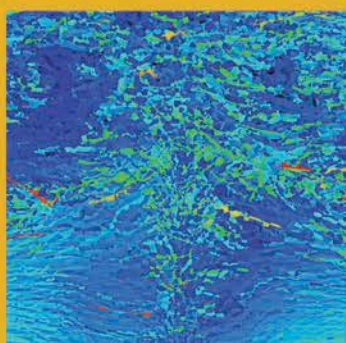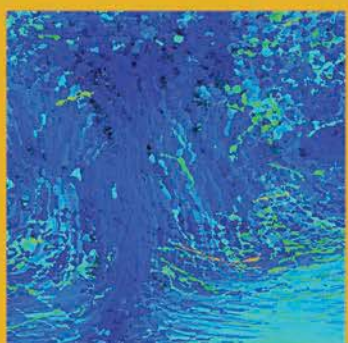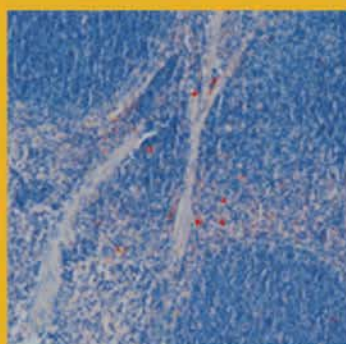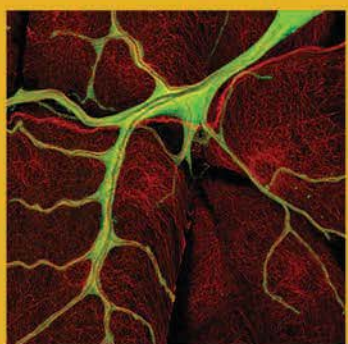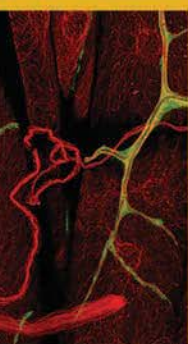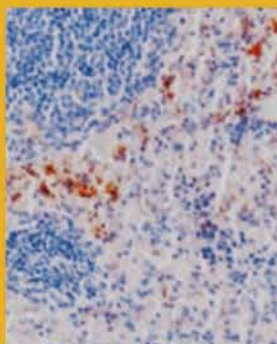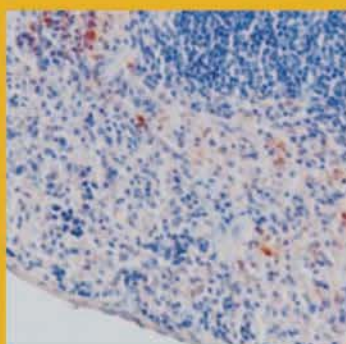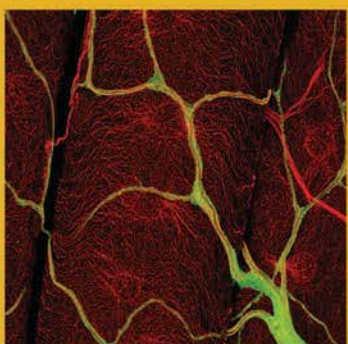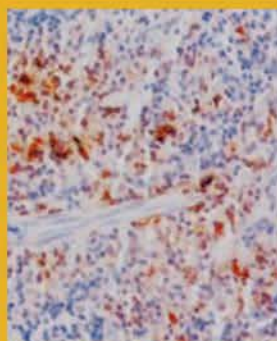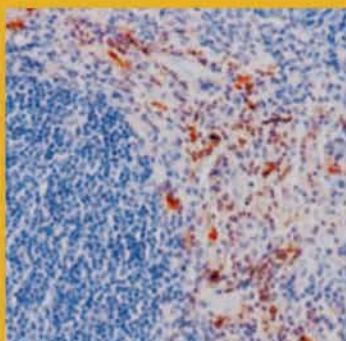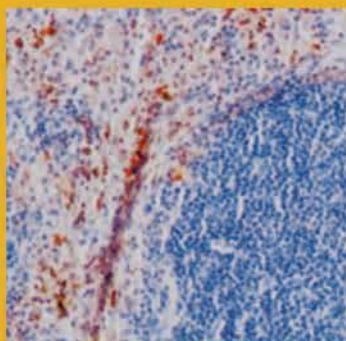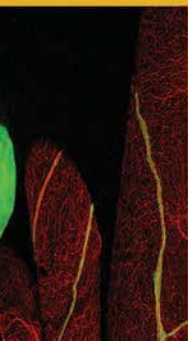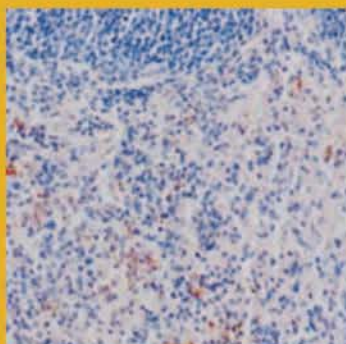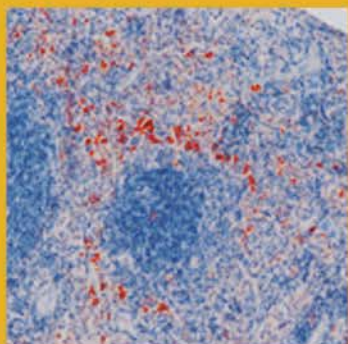
Directors' Research

Directors' Research covers two thematically distinct research groups, headed by the Director of EMBL and the Director of EMBO, an organisation of more than 1500 leading researchers that promotes excellence in the life sciences.

The Hentze group combines biochemical and systems-level approaches to investigate the connections between gene expression, cell metabolism, and their role in human disease. Key goals of the group include collaborative efforts to: uncover the basic mechanisms underlying protein synthesis and its regulation by miRNAs and RNA-binding proteins in cell metabolism, differentiation, and development; explore, define, and understand enigmRBPs and REM networks; help elucidate the role of RNA metabolism in disease, and to develop novel diagnostic and therapeutic strategies based on this knowledge; and to understand the molecular mechanisms and regulatory circuits underlying physiological iron homeostasis.

In investigating the mechanisms and forces that determine cell shape in *Drosophila*, **the Leptin group** studies two cell types. They look at how the cells at the tips of the fruit fly's tracheal system rearrange their components as they grow rapidly and branch out to carry air to the animal's tissues. And at the tissue level, the group investigates how forces generated by single cells give the embryo's ventral furrow its final shape. The group also studies medaka and zebrafish to understand how signals from damaged cells are recognised by the innate immune system. They are developing methods to assay immune and stress responses in real time as the fish's cells encounter pathogens and stress signals.

RNA Biology, metabolism and molecular medicine

Matthias Hentze

MD 1984, University of Münster.

Postdoctoral training at the NIH, Bethesda.

Group Leader at EMBL since 1989. Senior Scientist since 1998. Associate Director 2005-2013. Director since 2013.

Co-Director of the EMBL/University of Heidelberg Molecular Medicine Partnership Unit since 2002.

ERC Advanced Investigator since 2011.

SELECTED REFERENCES

Beckmann BM, *et al.* (2015) The RNA-binding proteomes from yeast to man harbor conserved enigmRBPs. *Nature Comm.* 6, 10127

Castello A, *et al.* (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 149, 1393-1406

Hentze MW, *et al.* (2010) Two to tango: regulation of mammalian iron metabolism. *Cell* 142, 24-38

Hentze MW & Preiss T. (2010) The REM phase of gene regulation. *Trends Biochem. Sci.* 35, 423-6

The Hentze group combines biochemical and systems level approaches to investigate the connections between gene expression and cell metabolism, and their role in human disease.

Important steps in the control of gene expression are executed in the cytoplasm by regulation of mRNAs via RNA-binding proteins (RBPs) and non-coding regulatory RNAs. We are elucidating these regulatory mechanisms, combining 'reductionist' biochemical and systems level approaches in mammalian, yeast and *Drosophila* model systems.

We developed the techniques of 'mRNA interactome capture' – to define 'all' RBPs associated with mRNAs *in vivo* (Castello *et al.*, 2012) – and 'RBDmap' – to identify the RNA-binding domains of previously unknown RBPs. This work led to the discovery that hundreds of seemingly well characterised cellular proteins also bind RNA (enigmRBPs). These discoveries offer an ideal starting point for exploration of 'enigmRBPs' and 'REM networks' (Hentze & Preiss, 2010), which we expect to shed new light on the relationships between cell metabolism and gene expression (figure 1). Within the Molecular Medicine Partnership Unit (MMPU), we are investigating the post-transcriptional processes of nonsense-mediated decay (NMD) and 3' end processing and their importance in genetic diseases, together with Andreas Kulozik, University of Heidelberg. Our second major interest is the biology of mammalian iron metabolism (figure 2). This work includes the definition of the functions of the IRE/IRP regulatory network and its crosstalk with the iron hormone hepcidin. Within the MMPU, together with Martina Muckenthaler, University of Heidelberg, we study the molecular basis of genetic and non-genetic diseases of human iron metabolism. Our work employs conditional knockout mouse strains for IRP1 and IRP2 and mouse models of iron metabolism diseases.

Future plans

- To uncover the basic mechanisms underlying translational regulation by RNA-binding proteins in cell metabolism, differentiation, and development.

- To explore, define, and understand enigmRBPs and REM networks.

- To help elucidate the role of RNA metabolism in disease, and to develop novel diagnostic and therapeutic strategies based on this knowledge.

- To understand the molecular mechanisms and regulatory circuits underlying physiological iron homeostasis.

For research themes and projects of the teams in the MMPU, see:
The Molecular Medicine Partnership Unit (MMPU): www.embl.de/research/partnerships/local/mmpu/index.html
The University Hospital Heidelberg: www.klinikum.uni-heidelberg.de/Molecular-Medicine-Partnership-Unit.114597.0.html
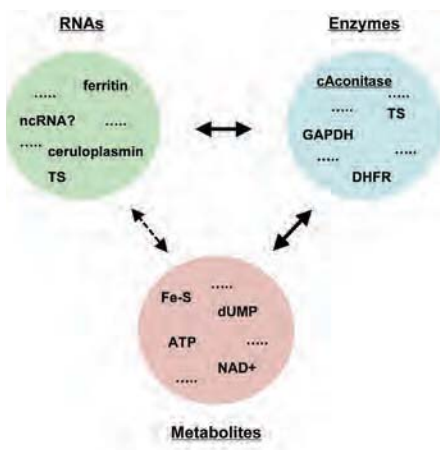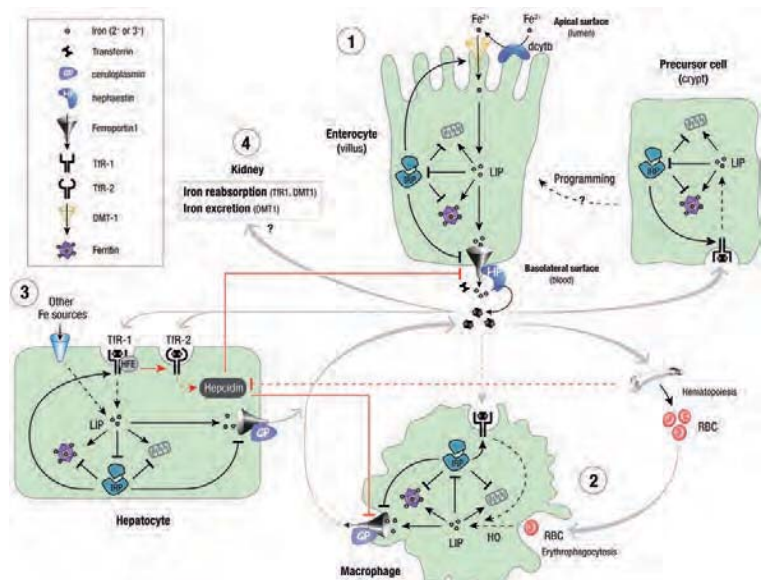

Figure 1: Exploring REM networks.

Figure 2: Systems biology of mammalian iron metabolism.

Cell shape and morphogenesis: subcellular and supracellular mechanisms

Maria Leptin

PhD 1983, Basel Institute for Immunology.

Postdoctoral research then Staff Scientist at the MRC Laboratory for Molecular Biology, Cambridge, UK.

Group Leader at the Max Planck Institute for Developmental Biology, Tübingen.

Professor, Institute for Genetics, University of Cologne.

Director of EMBO and Group Leader at EMBL since 2010.

SELECTED REFERENCES

Rauzi M, *et al.* (2015) Embryo-scale tissue mechanics during *Drosophila* gastrulation movements. *Nat Commun.* 6, 8677

Bajoghli B, *et al.* (2015) Noninvasive in toto imaging of the thymus reveals heterogeneous migratory behavior of developing T cells. *J Immunol.* 195, 2177-86

Sigurbjörnsdóttir S, Mathew R, & Leptin M. (2014) Molecular mechanisms of de novo lumen formation. *Nat Rev Mol Cell Biol.* 15, 665-76

Rembold M, *et al.* (2014) A conserved role for Snail as a potentiator of active transcription. *Genes Dev.* 28, 167-81

Cell shape determination during development

The shape of a developing organism is generated by the activities of its constituent cells: growth and proliferation, movements and shape changes. We are particularly interested in shape changes.

One study concerns an extremely complex single cell, the terminal cell of the *Drosophila* tracheal system. It is highly branched and carries air to target tissues through an intracellular tube bounded by plasma membrane (see figure 1). During its rapid growth, the cell faces the task of synthesising large amounts of membrane and sorting it correctly to defined membrane domains. Extensive re-organisation of the secretory organelles precedes membrane growth. We are investigating how the cytoskeleton, small GTPases and polarity determinants direct the process, and how membrane trafficking processes contribute to building the tube.

In another project, we are aiming to understand how the forces generated by individual cells are integrated within the supracellular organisation of the whole organism to give the tissue its final shape (see figure 2). We study the formation of the ventral furrow in the early *Drosophila* embryo. The cells that form the furrow are the major force generators driving invagination, but to allow furrow formation, neighbouring cells must respond and they may contribute to the process. To understand force integration across many cell populations, we use simultaneous time-lapse imaging of multiple-angle views of the gastrulating embryo. We measure the specific shape changes in all the cells of the embryo, as well as the speed and direction of their movements. Genetic and mechanical manipulations reveal the underlying control circuits.

In vivo *imaging of innate immune responses*

The innate immune system provides rapid defence against pathogens and also deals with non-pathogenic stresses. Macrophages and dendritic cells, two key players in this system, patrol the body and respond to stimuli from damaged cells via extra and intracellular sensors. We aim to understand how such signals are recognised and how the appropriate subcellular and intercellular responses are triggered. We have discovered that one family of sensors – the cytoplasmic NOD-like receptors (NLRs) – are particularly abundant in fish.

Fish model systems allow *in vivo* observation of physiological processes. Specifically, we watch pathogens and the cells that attack them. We use genetically and chemically engineered *in vivo* fluorescent reporters to assay immune and stress responses in real time and at high spatial and temporal resolution as the cells of the fish encounter pathogens and stress signals.

The Leptin group studies the mechanisms and forces that determine cell shape in Drosophila *and uses the zebrafish to analyse innate immune signalling.*

Figure 1: Tracheal cell (green) ramifying on a set of muscles (microtubules stained in red) in a *Drosophila* larva.
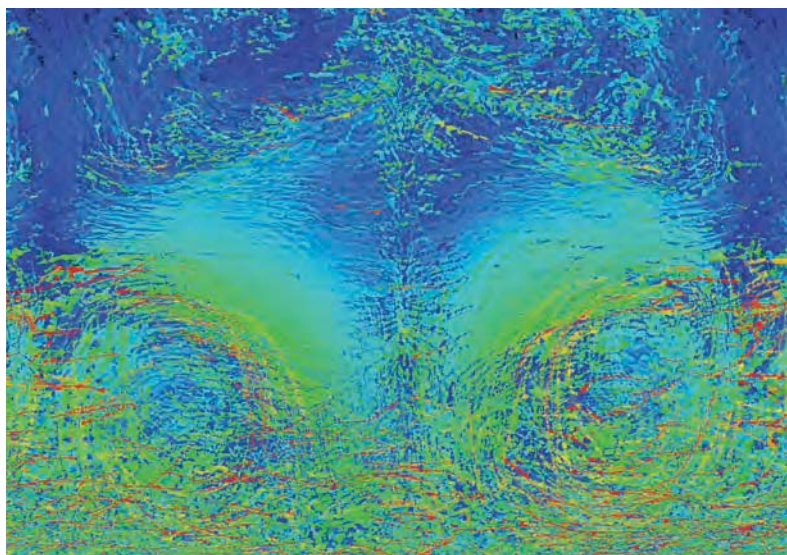
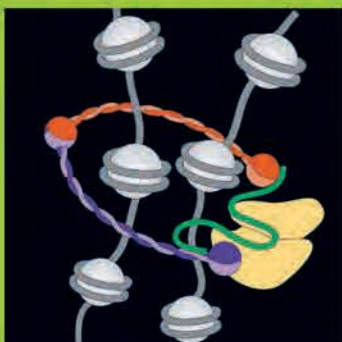Figure 2: A flat projection of the entire surface of a *Drosophila* embryo in which the position and speed of 6000 cells is followed over a 40 minute period. The head of the embryo is at the top, the center of the image is the ventral midline towards which the lateral cells are moving. Image by Matteo Rauzi.

EMBL Heidelberg
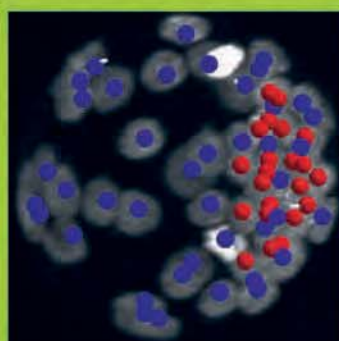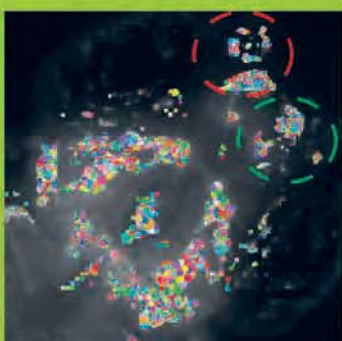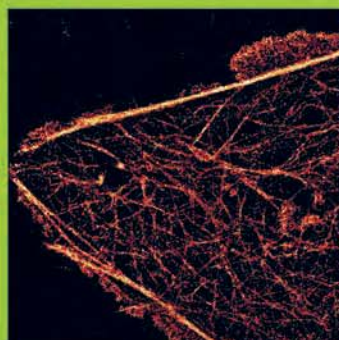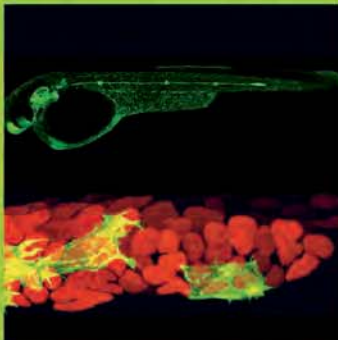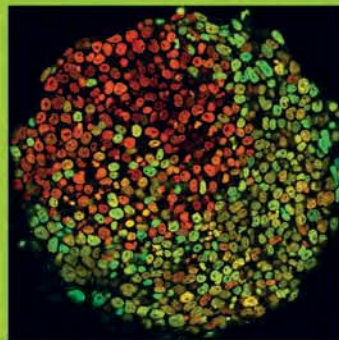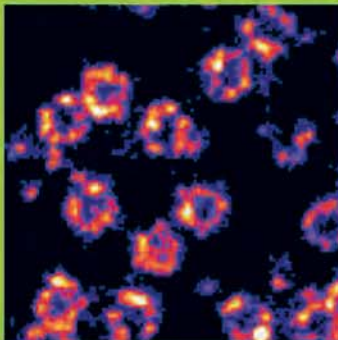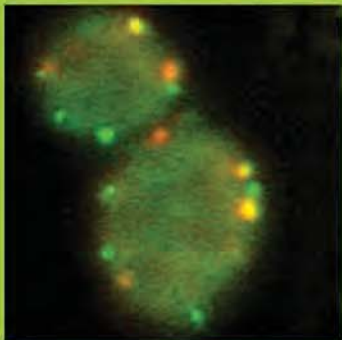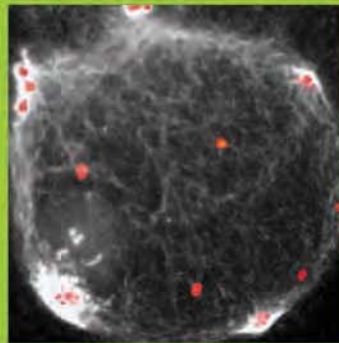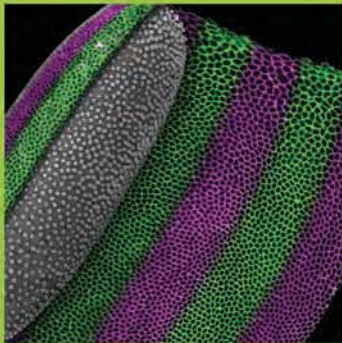
Cell Biology and Biophysics

In this Unit, physicists and chemists work closely together with biologists to elucidate the fundamental rules that govern dynamic cell organisation and function. At the same time, groups are developing new instruments and technologies in order to reach this ambitious goal.

Cells are the smallest autonomous units of life and occupy the midpoint between the molecular and macroscopic scales. In order to understand how living systems are built and function, we need to understand the physical principles that underlie cellular organisation and function.

It is in the cell where we will first understand the basic processes of life at the molecular level in a physiological context. The cell provides the natural coordinate system in space and time onto which we have to map and integrate genomic, transcriptomic, proteomic, structural and biophysical information about the molecules that make up living systems. In short, cell biology has become an integrative hub of much of modern biological research.

This is a time of tremendous opportunity for cell biology, but realising it also represents a formidable challenge and requires new concepts and approaches. Individual cellular processes – such as signalling, membrane trafficking, cytoskeletal dynamics and cell migration, gene expression or cell division – can no longer be studied in isolation but need to be considered as integrated events. The default situation is that the molecular machinery that performs these functions is complex and combinatorial at the single protein, protein complex, and pathway level. This requires new ways of thinking about cellular functions that use network biology and employing quantitative theoretical methods to generate mechanistic and predictive models that rely on realistic physical principles at the cellular, subcellular and molecular scale. Therefore, cell biology needs to integrate traditionally separate disciplines to realise its potential.

Novel developments in microscopy, computer simulations and chemical biology-based probes are a particular strength of the Unit. We constantly explore new directions and integrate new approaches and disciplines to answer cell biological questions. New correlative light/electron and super-resolution imaging methods, as well as mechanistic biochemistry, allow us to directly interface between cell and structural biology to understand molecular mechanisms. Furthermore, advances in live and deep tissue imaging methods now allow us to carry out cell biology in developing organisms to understand how collective cell behaviour leads to organ formation, and how cells interact with their physiological microenvironment.

Mechanisms of cellular functions are often best understood when the organisation of the cell changes dramatically to carry out new functions. This is the case when cells divide, when they change their fate. Both opportunities are exploited in the Unit. As a cell prepares to divide, all the microtubules suddenly depolymerise to reassemble into the mitotic spindle. At the same time, the nucleus is disassembled, mitotic chromosomes are formed, the Golgi complex fragments and membrane traffic ceases. After segregation of the genome is achieved, cellular organisation is re-established. Thus every cell cycle provides the opportunity to study the principles of the biogenesis of cellular compartments. Similarly, the genetic programme is changed and a reorganisation of cellular architecture takes place, guided by rules that we begin to unravel when progenitor cells differentiate into new cell types or start to migrate. Understanding these rules and principles is our challenge in the years to come.

Jan Ellenberg
Head of the Cell Biology and Biophysics Unit

Mechanics of cellular signalling

Alba Diz-Muñoz

PhD 2011, MPI of Cell Biology and Genetics, Germany (Paluch and Heisenberg laboratories).

Postdoctoral research between the Fletcher and Weiner laboratories at UC Berkeley and UCSF, Bay Area, USA.

Group leader at EMBL from April 2016.

SELECTED REFERENCES

Diz-Muñoz A, *et al.* (2013) Use the force: membrane tension as an organizer of diverse cellular functions. *Trends Cell Biol.* 23, 47-53

Diz-Muñoz* A, Krieg* M *et al.* (2010) Control of directed cell migration in vivo by membrane to cortex attachment. *PLoS Biology 8*, e1000544

The Diz-Muñoz group studies the crosstalk between membrane tension and signalling for cell polarity and migration of immune cells and zebrafish mesendodermal precursors.

For decades, research aimed at understanding cellular behaviour has largely focused on biochemical reactions. It is now well known that binding of soluble extracellular ligands to receptors on the plasma membrane can trigger lipid modifications, protein phosphorylation, and changes in protein localisation. More recently, it has become clear that physical forces also transmit critical information in cells and tissues, where they can regulate a wide variety of cell behaviours, including differentiation, death, movement and shape. At the cell surface, plasma membrane tension has been shown to integrate a wide variety of cell behaviours, ranging from determining leading edge size to regulating the balance between exocytosis and endocytosis.

We are only beginning to understand the many ways in which physical forces, and in particular plasma membrane tension, modulate behaviour at the molecular, cellular, and tissue length cells. Current approaches to measure and manipulate forces have important limitations and new tools and techniques are needed to address these challenges.

In the past we focused on the roles mechanics have in zebrafish cell migration and morphogenesis. We demonstrated that increasing cellular blebbing in mesendodermal cells (by reducing membrane tension or increasing intracellular pressure) impairs directed migration. Furthermore, we showed that the ratio of cellular protrusions (blebs to lamellipodia) controls the overall directional persistence of migration in a changing environment, such as in the developing zebrafish embryo. More recently, we have identified key components of a membrane tension-sensing pathway that controls leading edge formation in neutrophils (Figure 1). Our work sheds new light on how the plasma membrane integrates physical forces and intracellular signals to organise cell polarity during movement.

Future plans

Our primary interest lies in understanding the reciprocal interactions between physical forces and cellular signalling cascades during the establishment of cell polarity and subsequent cell motility in single cells and intact organisms. We will use a multi-scale and multi-disciplinary approach to develop novel ways to precisely quantify and modulate membrane tension (and membrane tension only) in an acute manner. Combining optogenetics, TIRF microscopy, 2-photon imaging, FRET sensors and atomic force microscopy, we will tackle how membrane tension/curvature affects signalling and vice-versa (Fig. 2) to organise cell polarity during movement.
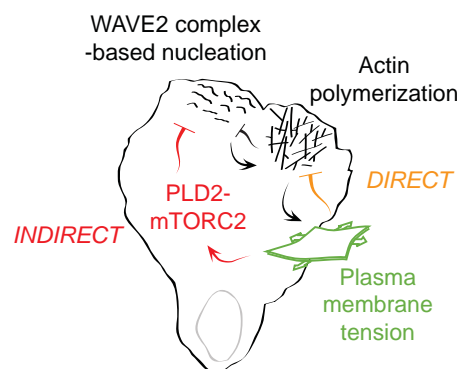

Figure 1: Working model of PLD2–TORC2 membrane tension negative feedback loop. This circuit forms a mechanosensory negative feedback loop that acts as a membrane tension homeostat and controls the spatial organisation of actin assembly during neutrophil polarity and movement.
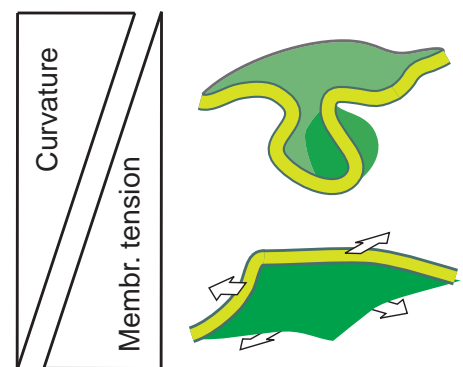
Figure 2: Feedback between curvature and membrane tension.

Systems biology of cell division and nuclear organisation

Jan Ellenberg

PhD 1998, Freie Universität Berlin.

PhD and postdoctoral research at the Cell Biology and Metabolism Branch, NICHD, NIH, Bethesda.

Group leader at EMBL since 1999.

Head of Gene Expression Unit 2006–2010.

Head of Cell Biology and Biophysics Unit since 2010.

SELECTED REFERENCES

Strnad P, *et al.* (2016). Inverted light-sheet microscopy microscope for imaging mouse pre-implantation development. *Nature Methods* (Dec 2015, Epub ahead of print)

Wachsmuth M, *et al.* (2015). High-Throughput fluorescence correlation spectroscopy enables analysis of proteome dynamics in living cells. *Nature Biotechnology* 33, 384-9

Szymborska A, *et al.* (2013). Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging. *Science*, 341, 655-8

Kitajima TS, *et al.* (2011). Complete kinetochore tracking reveals error-prone homologous chromosome biorientation in mammalian oocytes. *Cell* 146, 568-81

The Ellenberg group studies how cells divide and organise in mitosis and meiosis, where errors can lead to problems such as cancer and infertility.

Our overall goal is to systematically elucidate the mechanisms underlying cell division and nuclear organisation. We are developing a broad range of advanced fluorescence-based imaging technologies to assay the functions of the involved molecular machinery non-invasively, automate imaging to address all its molecular components, and computationally process image data to extract biochemical and biophysical parameters. Our research focuses on three areas: systems biology of mitosis, nuclear structure, and molecular mechanisms of meiosis and early embryonic mitosis.

We have previously identified hundreds of new cell division genes by RNAi-based screening of the entire human genome and we are now studying – in live cells and with high-throughput – protein function, interactions and networks during somatic mitosis by automating advanced fluorescence imaging and single molecule techniques, such as fluorescence (cross) correlation spectroscopy.

We also determined the positions of various nuclear pore complex (NPC) components and directly resolved the ring-like structure of the NPC, combining stochastic super-resolution microscopy (SRM) with single particle averaging. Currently, we elucidate the assembly mechanism of the NPC and chromatin dynamics over the cell cycle.

By complete kinetochore tracking we demonstrated that meiotic chromosome biorientation is highly error-prone. We have now developed a gentle light-sheet-based microscope for high-throughput imaging of mouse oocytes and embryos to enable systematic molecular analysis of meiosis and early embryonic mitosis (figure 1).

Future plans

We want to gain comprehensive mechanistic insight into the division of human mitotic cells, provide a biophysical basis to understand nuclear organisation, and establish methods for systems analysis of the meiotic and first mitotic divisions of mammalian oocytes and embryos.

For a systems-level understanding of all crucial protein interactions during cell division, we will combine automated bulk, single molecule imaging and computational data analysis with advanced machine learning and modelling approaches to integrate all interactions into one canonical 4D model of a human dividing cell (figure 2).

To come to a structural understanding of nuclear organisation, we will explore and further improve correlative imaging approaches, combining live cell confocal microscopy, SRM and electron tomography to unravel the mechanism of NPC assembly and disassembly, as well as the human genome architecture, chromatin organisation and compaction.

To be able to apply systems biology tools to oocyte meiosis and early embryonic mitosis, we will push light-sheet-based imaging technology development further to improve its light efficiency and resolution in order to establish a physiological molecular model for early mammalian development and infertility.
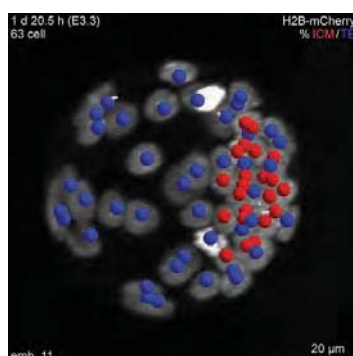
Figure 1: Embryonic pre-implantation development: Tracking of all cells throughout the first mitotic divisions, from zygote to blastocyst stage, using an inverted light-sheet microscope. Nuclei detected by their H2B signal are shown. Colour-coded spheres indicate cells of different fate: trophoectoderm (TE; blue) or inner cell mass (ICM; red) (Strnad P *et al.* 2016).
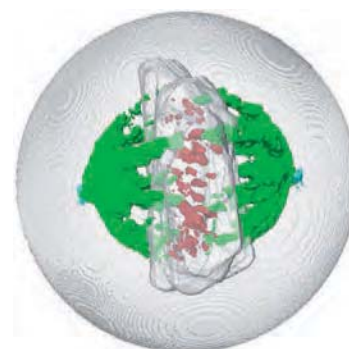
Figure 2: A 3D metaphase model reconstructed from confocal images of HeLa cells. The colours identify different elements/proteins important for mitosis (centrosome, cyan; chromatin, green; kinetochores/centromeres, red).

Multicellular morphogenesis

Darren Gilmour

PhD 1996, University of Cambridge.

Postdoctoral research at the Max Planck Institute for Developmental Biology, Tübingen.

Group Leader at EMBL Heidelberg since 2004.

SELECTED REFERENCES

Durdu S, *et al.* (2014) Luminal signalling links cell communication to tissue architecture during organogenesis. *Nature* 515, 120-4

Revenu C, *et al.* (2014) Quantitative cell polarity imaging defines leader-to-follower transitions during collective migration and the key role of microtubule-dependent adherens junction formation. *Development* 141, 1282-91

Donà E, *et al.* (2013) Directional tissue migration through a self-generated chemokine gradient. *Nature* 503, 285-9

Streichan SJ, *et al.* (2011) Collective cell migration guided by dynamically maintained gradients. *Phys Biol* 8, 045004

Using the zebrafish as a model, the Gilmour group takes an integrative, multiscale approach to study how cells collectively migrate and assemble into functional organs.

Collective behaviour lies at the heart of all biological design. Whether it is the assembly of proteins into complexes or the organisation of animal societies, collective interaction creates something much greater than the sum of the parts. A breathtaking example of such behaviour is seen during embryogenesis, when thousands of collectively migrating cells self-organise to form functional tissues and organs. Given the key role played by collective migration in organ formation, wound repair and cancer, it is surprisingly how little we know about how cells organise each other.

We take an integrative, multi-scale approach to study how cells collectively migrate and assemble into functional organs, using the zebrafish lateral line organ as an experimental model. Here, a migrating epithelial primordium comprising of 100 cells, assembles and deposits a series of rosette-like mechanosensory organs across the surface of the embryo. Its superficial migration route, beneath a single transparent cell layer, makes it the dream *in vivo* sample for quantitative imaging. Moreover, the process can be interrogated using a range of perturbation approaches, such as chemical and optogenetics, and many of the molecular regulators of its migratory behaviour are of general interest due to their role in human disease. For example, the migrating collective is guided by Cxcr4/SDF1 signalling, a chemokine-receptor pair known to control many human cancers.

Future plans

The focus of our group is to use the lateral line to address the general question of how cell behaviours are regulated and coordinated within collectively migrating tissues. We have developed *in vivo* imaging, analysis and perturbation tools that allow the entire morphogenesis process to be addressed at different spatiotemporal scales. By integrating these data, using statistical methods and modelling, we are aiming to understand the interplay between 'opposing' behaviours – namely, cell migration and differentiation. In this way, we hope to move towards a systems-level understanding of how dynamic cell organisation and gene expression are integrated during tissue morphogenesis.
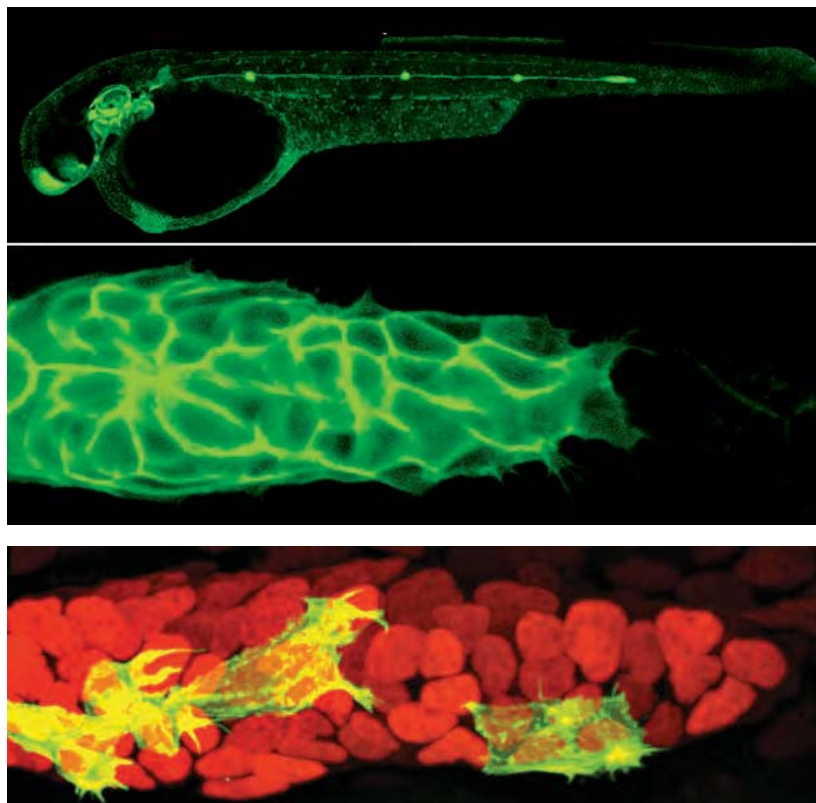
Figure 1: The zebrafish migrating lateral line organ allows collective migration to be easily studied *in vivo*.

Figure 2: Visualising actin dynamics (LifeAct-GFP) within migrating primordium.

Chromosome structure and dynamics

Christian Häring

PhD 2003, Institute of Molecular Pathology, Vienna.

Postdoctoral research at the University of Oxford.

Group Leader at EMBL Heidelberg since 2007.

SELECTED REFERENCES

Piazza I, *et al.* (2014) Association of condensin with chromosomes depends on DNA binding by its HEAT-repeat subunits. *Nat. Struct. Mol. Biol.* 6, 560-8

Cuylen S, *et al.* (2013) Entrapment of chromosomes by condensin rings prevents their breakage during cytokinesis. *Dev. Cell* 4, 469-78

Petrova B, *et al.* (2013) Quantitative analysis of chromosome condensation in fission yeast. *Mol. Cell. Biol.* 5, 984-98

Cuylen S, Metz J, & Haering CH. (2011) Condensin structures chromosomal DNA through topological links. *Nat. Struct. Mol. Biol.* 8, 894-901

The Häring group aims to understand the molecular machinery that organises chromosomes to allow their correct distribution among daughter cells.

Eukaryotic chromosomes undergo enormous changes in structure and organisation over the course of a cell cycle. One of the most fascinating changes is the transformation of interphase chromatin into rod-shaped mitotic chromosomes in preparation for cell division. This process, known as chromosome condensation, is a key step for the successful segregation of chromosomes during mitosis and meiosis. The underlying mechanisms are, however, still poorly understood.

The overall aim of our research is to unravel the action of molecular machines that organise the 3D architecture of eukaryotic genomes. Insights into the general working principles behind these machines will be of great importance to our understanding of how cells inherit a complete set of their chromosomes every time they divide and thereby prevent the emergence of aneuploidies, which are hallmarks of most cancer cells and the leading cause of spontaneous miscarriages in humans.

One of the central players in the formation of mitotic chromosomes is a highly conserved multi-subunit protein complex, known as condensin. We have shown that condensin encircles chromosomal DNA within a large ring structure formed by its structural maintenance of chromosomes (SMC) and kleisin subunits. Our working hypothesis is that condensin uses this topological principle to tie together loops of chromatin (figure 1), which ensures that chromosome arms clear the site of cell cleavage before cytokinesis.

In an independent project, we use a newly developed time-resolved light microscopy assay to quantitatively measure chromosome condensation in live fission yeast cells in high-throughput (figure 2). This has identified, in addition to condensin, new players that direct the formation of mitotic and meiotic chromosomes.

Future plans

We will continue to use a highly interdisciplinary approach to advance our understanding of condensin function in yeast and mammalian cells by combining biochemical, molecular, structural, and cell biology methods. In collaboration with other groups, we are taking further advantage of chemical biological techniques as well as single-molecule approaches to discover how condensin loads onto chromosomes, how it interacts with other chromosomal components, and how its activity is controlled. In addition, we are further investigating the novel candidates identified in the screen for mitotic chromosome condensation proteins to understand the basis of their functions on mitotic chromosomes.

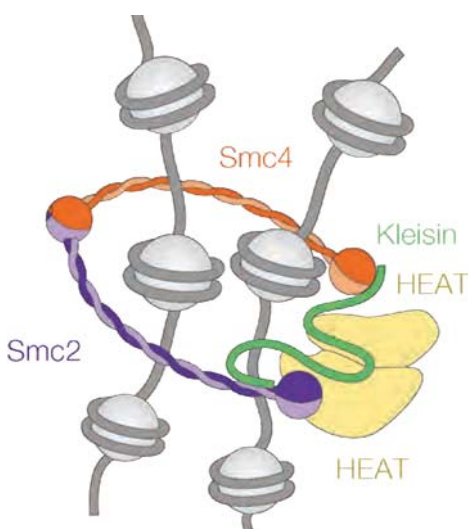Figure 1: Model for the organisation of mitotic chromosomes by condensin rings.
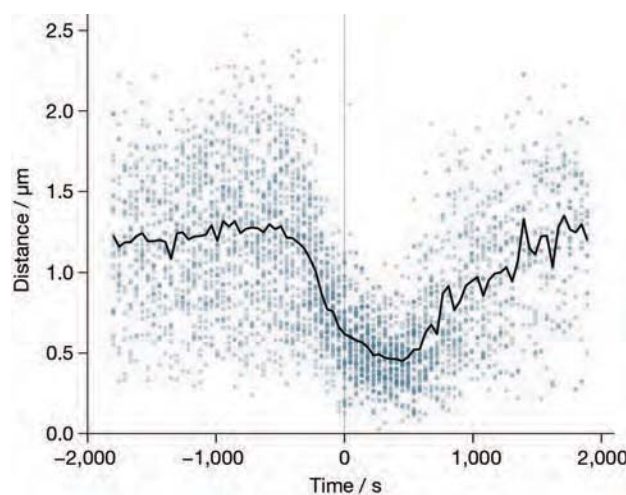

Figure 2: The live-cell chromosome condensation assay tracks the distances between two fluorescently labelled chromosome loci over time. Alignment of a large number of single cell tracks (circles) to the time of anaphase onset (t = 0) generates an average distance plot (line) as a quantitative read-out of condensation dynamics.

Dynamics of cell growth and tissue architecture

Lars Hufnagel

PhD 2001, Max Planck Instutite for Dynamics and Self-Organisation, Göttingen.

Postdoctoral research at the Kavli Institute for Theoretical Physics, Santa Barbara, California.

Group leader at EMBL Heidelberg since 2007.

SELECTED REFERENCES

Krzic U, *et al.* (2012) Multiview light-sheet microscope for rapid *in toto* imaging. *Nat. Methods* 9, 730-3

Capoulade J, *et al.* (2011) Quantitative fluorescence imaging of protein diffusion and interaction in living cells. *Nat. Biotechnol.* 29, 835-9

Streichan SJ, *et al.* (2011) Collective cell migration guided by dynamically maintained gradients. *Phys Biol* 8, 045004

The Hufnagel group studies the role of mechanical cons-traints on processes such as cell growth, programmed cell death, orientation of division, intra-tissue rearrange-ments and cell differentiation.

Biological processes are highly dynamic and span many temporal and spatial scales. During development, cells must integrate and respond to a multitude of biochemical and biophysical signals: for example, changes in intracellular signalling networks, cytoskeleton remodelling, cell shape changes, long-range signalling and tissue remodelling. A whole-embryo view of morphogenesis with subcellular resolution is essential for unravelling the interconnected dynamics at varying scales of development – from interactions within cells to those acting across the whole embryo. Bridging scales from the submicron to the millimeter range with a temporal resolution of several seconds – combined with a total imaging time of several hours – not only poses tremendous challenges for modern microscopy methods but also requires powerful computational approaches for data handling, processing, and image analysis.

The central question that we are interested in is how a complex multi-cellular tissue or organism is formed from individual cells by spatio-temporal regulation of biophysical and intracellular signalling processes. We address all experimental steps, from innovative transgenic lines and microscope development to systematic image processing and biophysical modelling. This requires a multidisciplinary environment of biologists, physicists and computer scientists working closely together.

In order to address these questions we develop novel imaging techniques based on selective plane illumination microscopy (SPIM). SPIM yields optical sectioning by uncoupling the optical path for sample illumination from emitted photon detection. The illumination branch creates a thin light sheet to illuminate a specimen from the side and the emitted light is collected and imaged onto a high speed and high sensitivity camera by a second objective lens. The unprecedented speed of light sheet-based microscopy poses challenges for data handling and image processing, which we address by developing novel image processing tools.

Currently, we investigate cell shape changes and growth patterns in the *Drosophila* embryo with emphasis on the role of mechanical constraints on organ formation and tissue differentiation, complemented by mammalian cell culture studies investigating cell cycle response of an epithelial tissue to external and internal mechanical perturbations. Our group is part of the Centre for Modelling and Simulations in the Biosciences (BIOMS).

Future plans

We are focused on the control and regulation of cell proliferation, apoptosis and cellular rearrangement processes in developing tissues, with a specific emphasis on epithelial tissues and the role of mechanical interactions as a regulator. We seek to characterise and quantify the spatio-temporal effects of mechanical stress, deformations and fluid flow-induced sheer stress on cell growth, gene expression and cellular polarity in two-dimensional epithelial tissues. To address this issue, we pursue an interdisciplinary approach combining classical biological techniques with detailed modelling methods from various fields, ranging from statistical physics to applied mathematics and computer science. We will continue to not only tailor light-sheet microscopes to match specific biological questions, but also push the boundaries of light-sheet microscopy towards high speed intracellular imaging with extremely thin light sheets, super-resolution techniques, and quantitative *in toto* imaging.

Figure 1: MuVi-SPIM image of a *Drosophila* embryo. Eight views were fused to yield an *in toto* reconstruction of the embryo (one side membrane unrolled). The high speed of the microscope enables a detailed reconstruction of cell lineage and shape changes over extended periods of development.

Mechanisms of oncogene dependence and tumour relapse

Martin Jechlinger

PhD 2002, Institute of Molecular Pathology (IMP) Vienna.

Postdoctoral research at Memorial Sloan-Kettering Cancer Center (MSKCC) New York.

Group leader at EMBL Monterotondo 2010-15.
Group leader at EMBL Heidelberg since 2015.

SELECTED REFERENCES

Jechlinger M, *et al.* (2009) Regulation of transgenes in three-dimensional cultures of primary mouse mammary cells demonstrates oncogene dependence and identifies cells that survive deinduction. *Genes Dev.* 23, 1677-88

Podsypanina K, *et al.* (2008) Seeding and propagation of untransformed mouse mammary cells in the lung. *Science* 321, 1841-4

Jechlinger M, *et al.* (2006) Autocrine PDGFR signaling promotes mammary cancer metastasis. *J. Clin. Invest* 116, 1561-70

Jechlinger M, *et al.* (2003) Expression profiling of epithelial plasticity in tumor progression. *Oncogene* 16, 7155-69

In patients and mouse models, interference with the activity of cancer-initiating oncogenes can result in tumour regression. However, novel therapies that target the products of mutant alleles in human cancers are only partly successful, since maintenance of remission requires long-term treatment and relapse often occurs in the presence of therapeutic agents. Hence, a better understanding of drug resistance and tumour recurrence is needed for the design of more successful anti-cancer strategies.
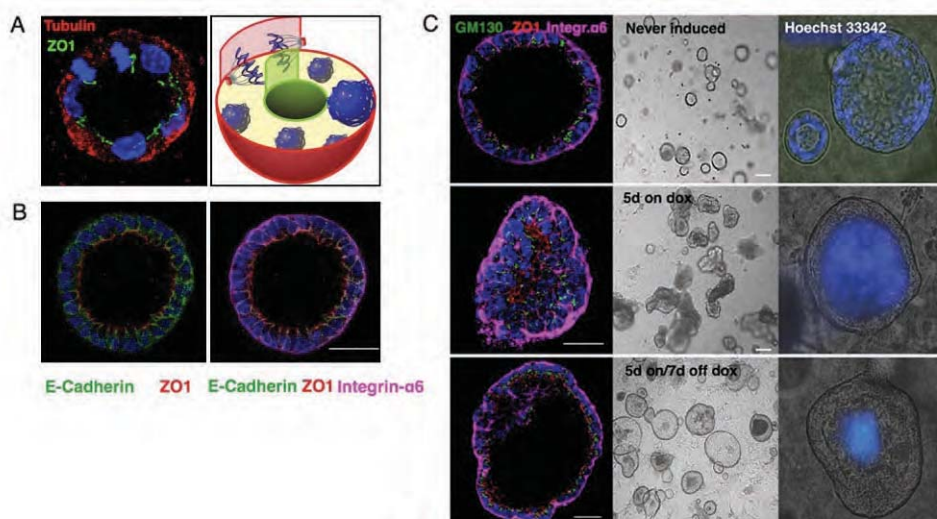
Transgenic mice carrying regulatable transgenes represent tractable systems for studying the mechanisms of oncogene dependence, the response and resistance to targeted drugs and tumour recurrence. In a complementary approach, we have developed a 3D culture system of primary mouse mammary epithelial cells to study detailed responses to the induction and de-induction of oncogenes (mimicking treatment with an ideally targeted drug). This 3D system produced phenotypic changes similar to those observed in the mammary glands of the transgenic mice from which the cultures were derived. In addition, this new approach identified and isolated cells that had survived oncogene withdrawal, which represent a pure population of residual 'dormant' tumour cells. We recently verified that these residual cells progress to form recurrent tumours similar to relapses observed in the animals and reminiscent of the patient situation. Molecular profiling revealed unique hallmarks of surviving residual cells, which we are currently analysing to better understand mechanisms of breast tumour recurrence.

The Jechlinger group uses a 3D culture system of primary mouse mammary epithelial cells to study cancer-initiating oncogenes.

Future plans

We are employing genomic methods, a range of molecular biology techniques, histopathological analysis as well as live cell imaging to investigate our organotypic 3D cell culture systems and the corresponding mouse models. The main aims include:

- Determine at which point during tumourigenesis cells acquire the ability to escape targeted therapy.

- Identify the molecular properties that distinguish surviving-residual cells, from naive cells.

- Interfere with the mechanisms important for survival of residual 'dormant' cells.

(A) One cell in anaphase divides with the sets of chromosomes perpendicular to the apical surface, while keeping ZO1 localised to the luminal membrane.

(B) Confocal microscopy (5mm projection through the middle) shows epithelial cell polarity. E-Cadherin, (adherens junctions, lateral); ZO1, (tight junctions, apical); Integrin a6 (basolateral).

(C) Doxycycline exposure causes loss of epithelial polarity and filling of the lumen; Removal of doxycycline results in survival of a re-polarised cell layer, that acquires the ability to exclude Hoechst 33342. Left panels: Confocal microscopy (5mm projection through the middle) shows Dapi, GM130 (apical), ZO1 (tight junction), Integrina6 (basal) at indicated times. Middle panels: Bright-field pictures show: (top) small, hollow acini; (middle) filled, irregular shaped spheres (bottom) hollow, irregular shaped spheres that show debris of internal cells. Right panels: Exclusion of Hoechst 33342 (1 hour incubation) at indicated times.

Cytoskeletal dynamics and function in oocytes

Péter Lénárt

PhD 2004, EMBL and University of Heidelberg.

Postdoctoral research at the Institute of
Molecular Pathology (IMP), Vienna.

Staff scientist at EMBL Heidelberg since 2008.

Group leader since 2011.

SELECTED REFERENCES

Borrego-Pinto J, *et al.* (2016) Distinct mechanisms eliminate mother and daughter centrioles in meiosis of starfish oocytes. *J Cell Biol* jcb.201510083 (in press)

Mori M, *et al.* (2014) An Arp2/3 nucleated F-actin shell fragments nuclear membranes at nuclear envelope breakdown in starfish oocytes. *Curr Biol.* 24, 1421-8

Mori M, *et al.* (2011) Intracellular transport by an anchored homogeneously contracting F-actin meshwork. *Curr. Biol.* 21, 606-11

Lénárt P, *et al.* (2005) A contractile nuclear actin network drives chromosome congression in oocytes. *Nature* 436, 812-8

Using starfish as a model organism, the Lénárt group combines bio-chemistry with imaging assays to investigate how the fertilisable egg cell develops from the oocyte.

All animal life begins with the fusion of sperm and egg. Our research is focused on the egg cell, specifically investigating how the fertilisable egg develops from the oocyte through meiotic divisions. Oocytes are exceptionally large cells, with diameters up to millimetres in size, because they store large amounts of nutrients to support embryonic development. Therefore, in oocytes and eggs, the cytoskeleton has to transport organelles, separate chromosomes, and organise cellular architecture in a very large cytoplasm. How the cytoskeleton adapts to this unusual size, and how these mechanisms differ from those in small somatic cells, is largely unknown.

We use starfish oocytes as a model system because they are easy to handle, complete meiosis rapidly, develop simply in seawater at room temperature, and are transparent – ideal for high-resolution imaging of cytoskeletal dynamics in live cells. We use confocal microscopy to image live starfish oocytes and employ computational image analysis tools to extract quantitative parameters from these 3D time-lapse datasets. Parameters such as local concentrations or velocities of cellular components provide a quantitative assay for the biological process and, at the same time, serve as inputs for computational models of cytoskeletal dynamics. Model predictions are then tested in perturbation experiments using physical or molecular manipulations. Biochemistry, in combination with the imaging assays, is used to identify the key molecular components in the process.

We have recently shown that meiotic chromosomes scattered in the large oocyte nucleus are collected by an actin meshwork and transported to the spindle, whose short microtubules cannot reach the chromosomes directly, as they do in somatic cells. This novel actin-based chromosome transport system forms as the nuclear envelope breaks down and fills the nuclear space with an actin meshwork, physically entrapping chromosomes. We showed that the actin meshwork contracts homogeneously; however, because it is mechanically anchored to the cell cortex, this contraction is translated into directional transport towards the cortex where the spindle forms. By understanding the mechanism of chromosome transport essential to oocyte division and fertility, our studies revealed a novel design principle for a cytoskeletal 'transport machine' that is very different from previously known mechanisms of actin-driven intracellular transport.

Future plans

Immediate goals include determining the detailed structure of the F-actin meshwork, understanding the molecular mechanisms of meshwork contraction, and identifying the mechanisms by which chromosomes attach to the meshwork. We will employ high-resolution imaging methods to resolve single actin-filaments and to identify, localise and perturb molecules regulating actin filament dynamics to explore the underlying molecular mechanisms. Longer term, we are interested in related cytoskeletal processes that occur in oocytes, eggs and early embryos, with the aim of understanding mechanistically the organisational principles of the actin and microtubule cytoskeleton.
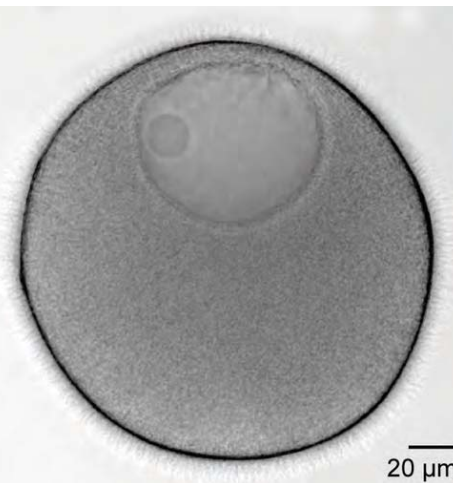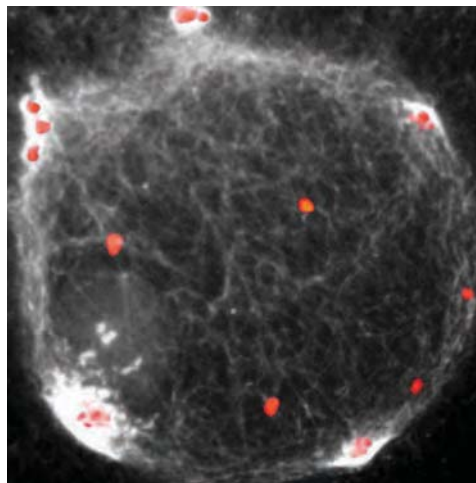

Figure 1 (far left):
The actin filament network (gray) embedding the chromosomes (red).

Figure 2 (near left):
Transparent starfish oocytes are uniquely suited for imaging meiotic divisions.

François Nédélec

PhD 1998, Université Paris-Sud II.

Postdoctoral Research at EMBL.

Group Leader since 2005.

Joint Appointment with the
Stuctural and Computational Biology Unit.

SELECTED REFERENCES

Loughlin R, Heald R, & Nédélec F. (2010) A computational model predicts Xenopus meiotic spindle organization. *J. Cell Biol.* 191, 1239-49

Dinarina A, *et al.* (2009) Chromatin shapes the mitotic spindle. *Cell* 138, 502-13

Jékely G, *et al.* (2008) Mechanism of phototaxis in marine zooplankton. *Nature* 456, 395-9

Kozlowski C, Srayko M & Nedelec F. (2007) Cortical microtubule contacts position the spindle in C. elegans embryos. *Cell* 129, 499-510

The Nédélec group develops in vitro experiments and modelling tools to explore complex intracellular processes, such as mitosis.

Modern microscopy has demonstrated the dynamic nature of biological organisation. The mitotic spindle, for example, is a stable and solid cellular structure: in a given cell type, it has a precise symmetry and very reproducible dimensions. Yet, except for the chromosomes, all the components of a spindle – polar filaments called microtubules and associated proteins – are in rapid turnover. Microtubules grow, shrink and disappear in less than a minute and their associated proteins continuously and stochastically bind and unbind even faster. The resulting assembly, although highly dynamic, is remarkably precise: it can remain steady for hours waiting for the right signal, to eventually apply the balanced forces necessary to position and segregate the chromosomes exactly.

The spindle is thus a fascinating structure that illustrates a central question in biology: how can the uncoordinated and inevitably imperfect actions of proteins and other molecules collectively fulfil the biological needs with the required accuracy? Today, understanding biological phenomena from their multiple biological components seems within our reach, as testified by the rise of systems biology. Yet, collective behaviours in biology require more than statistical averages. Understanding such complex collective behaviours is challenging for many reasons: 1) the diversity of molecular players is enormous; 2) their interactions are often dynamic and out-of-equilibrium; and 3) the properties of the constituents have been selected by natural evolution.

We approach this topic in practical terms by developing *in vitro* experiments and modelling tools, allowing us to reduce the number of components in the system: we can either remove specific proteins, or start from scratch by mixing purified components. Modelling allows us to recapitulate the process of protein organisation in a framework in which all the interactions are known exactly and can even be specified at will. We have developed an advanced simulation engine – called Cytosim – to simulate ensembles of multiple polar fibres and associated proteins, which can simulate problems involving microtubules, actin filaments or both. Simulations are often used to validate or refute existing ideas, but we also try to use them in a more creative way: one can generate systematically various properties for the molecules and automatically test their ability to form stable structures. The analysis of successful scenarios leads to the formulation of new hypotheses.

Future plans

We will study systems in which experiments and theory can be synergistically combined. We currently focus on *Xenopus* egg extracts, an experimental system in which many aspects of mitosis can be recapitulated. We are also generally interested in modelling cellular processes in which the cytoskeleton is a major player, such as the different stages of mitosis, the generation of cell shape in *S. pombe*, or the generation of asymmetry during cell division.
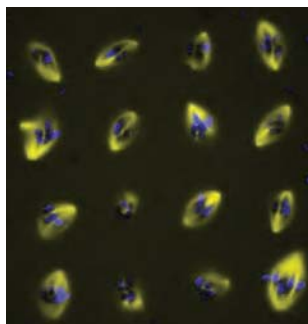
Figure 1: An array of mitotic spindles obtained *in vitro* with Xenopus laevis egg extracts (Dinarina *et al.*, 2009).

Figure 2 (below): The metaphase spindle, a dynamic bipolar structure of filaments called microtubules (white) that are connected by molecular motors (orange). This simulation elucidates how a spindle can remain stable for hours, even though it is made of filaments that individually exist for less than a minute (Loughlin *et al.* 2010).
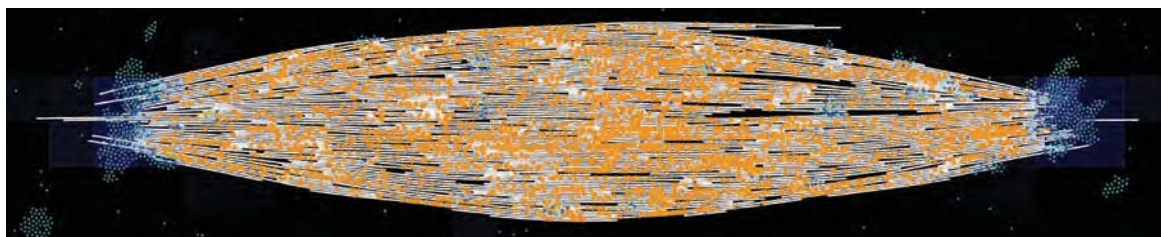
Systems biology of stem cell differentiation

Pierre Neveu

PhD 2007, Ecole Normale Supérieure, Paris.

Postdoctoral research at the Kavli Institute for Theoretical Physics and the Neuroscience Research Institute, Santa Barbara.

Group leader at EMBL Heidelberg since 2011.

SELECTED REFERENCES

Sladitschek HL, Neveu PA (2015). The bimodally expressed microRNA miR-142 gates exit from pluripotency.
Mol Syst Biol 11, 850

Neveu P, *et al.* (2010) MicroRNA profiling reveals two distinct p53-related human pluripotent stem cell states.
Cell Stem Cell 7, 671-81

The Neveu group takes an integrated systems biology approach to investigate the molecular changes that determine what a stem cell becomes.

Pluripotent cells have the dual ability to self-renew and differentiate. Therefore, in pluripotent cells, the expression of hundreds of genes should be stable in the self-renewal case, but gene expression can also be directed in a coordinated manner towards particular states upon external signalling cues (lineage commitment towards terminal differentiation). Deciphering this complex problem has garnered much attention at the systems level.

Tackling this challenge requires good characterisation of the pluripotent state. miRNAs are suitable marker candidates because they are excellent classifiers of tissue types or cellular states and they also play a crucial role in differentiation. By profiling miRNA expression in human cells, we have previously shown that pluripotency surprisingly emerges as a much more diverse state than previously believed: variability in miRNA expression is comparable to that found in differentiated cells and cancer cells. We have also shown that it is possible to dramatically reduce the complexity of miRNA expression patterns to a few meaningful dimensions, suggesting that complex processes of the stem cell system, such as differentiation and reprogramming, can be mapped quantitatively.

Currently, we are employing a dynamic approach at the single cell level to resolve the dynamics of differentiation and the different molecular and cellular processes at play during fate determination. Indeed, differentiation is intrinsically a dynamic process, where individual cells have to change from one state to another. Having developed fluorescent reporters to assess miRNA expression in single cells, we are characterising mouse embryonic stem cell (ESC) self-renewal using single-cell live imaging. We have found that the miRNA miR-142 is bimodally expressed in self-renewing mouse ESCs. miR-142 expression stratifies ESCs into two distinct subpopulations that interconvert stochastically: one subpopulation able to differentiate and a subpopulation deaf to differentiation cues. miR-142 expression creates this phenotypic diversity by switching the activation status of key intracellular signalling pathways.

Future plans

We plan to study the dynamics of differentiation at the single-cell level both *in vitro* in mouse embryonic stem cells and *in vivo*. The ultimate goal is to dissect the transcriptional regulation and gene networks and the associated cellular changes underlying stem cell differentiation. We are taking an integrated systems biology approach that combines single-cell live imaging of miRNA expression, image processing, perturbation approaches, and mathematical modelling.
We wish to address the following questions:

- How dynamic is the pluripotent state?

- What are the *in vitro* dynamics of differentiation of mouse ESCs?

- How do *in vitro* findings compare to *in vivo* differentiation behaviour?

Figure 1: Molecular cartography of stem cells: miRNA expression classifies pluripotent cells, cancer cells and differentiated cells. This map allows us to follow quantitative changes in cell identity such as differentiation and reprogramming. It reveals that reprogramming goes through a cancer-like behaviour.
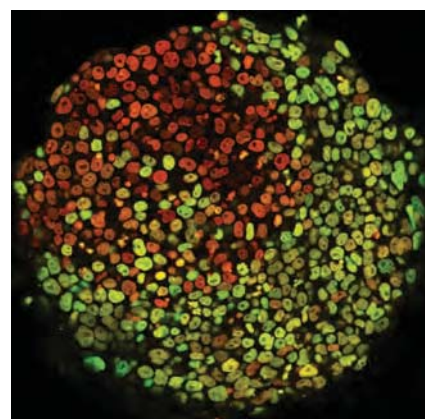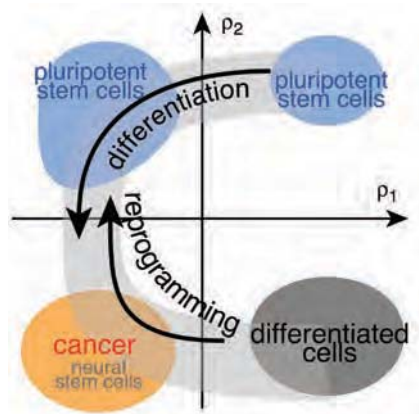

Figure 2: mouse ESC colony expressing a fluorescent reporter of miR-142 activity.

Membrane traffic and organelle biogenesis

Rainer Pepperkok

PhD 1992, University Kaiserslautern.

Postdoctoral research at University of Geneva.

Lab head at the Imperial Cancer Research Fund, London.

At EMBL since 1998.

Senior scientist since 2012.

Head of Core Facilities and Scientific Services since 2014.

SELECTED REFERENCE

Ronchi P, *et al.*, (2014) Positive feedback between Golgi membranes, microtubules and ER exit sites directs de novo biogenesis of the Golgi. *J. Cell Sci.* 127, 4620-33

Blattmann P, *et al.* (2013) RNAi-based functional profiling of loci from blood lipid genome-wide association studies identifies genes with cholesterol regulatory function. *PLoS Genet.* 9, e1003338

Simpson JC, *et al.* (2012) Genome-wide RNAi screening identifies human proteins with a regulatory function in the early secretory pathway. *Nat. Cell Biol.* 14, 764-74

Conrad C, *et al.* (2011) Micropilot: automation of fluorescence microscopy-based imaging for systems biology. *Nat. Methods* 8, 246-9

While many of the core components of the secretory machinery have been identified and characterised to some detail in the past decades, still little is known about how all components function together and how they are regulated in response to extracellular stimuli, stress or differentiation. Transport of material from one organelle to the other involves several steps, which have to occur sequentially and thus require a high degree of control at the molecular level (see figure 1). In order to understand such regulation in the physiological system that contains all possible components involved in the intact cell, we have developed and applied microscopy-based approaches to systematically identify components that regulate the early secretory pathway and the biogenesis and maintenance of the Golgi complex, down to the genome level. We have also developed and applied high-throughput microscopy techniques to quantitatively image genetic or physical interactions of the components we identified.

Network analyses of the components identified in our large-scale screens revealed links between early secretory pathway function, small GTP-binding protein regulation, actin and microtubule cytoskeleton organisation and growth factor mediated signalling. It provides a basis for understanding the global cellular organisation and regulation of the secretory pathway.

In order to investigate the mechanisms of Golgi biogenesis we have developed an approach, using laser nanosurgery, to deplete living cells from their Golgi complex and subsequently analyse the 'Golgi-less' karyoplast by time-lapse light and electron microscopy (figure 2). With this approach we are able to show that Golgi biogenesis in mammalian cells occurs *de novo* from ER derived membranes by a self-organising mechanism that integrates Golgi biogenesis, ER-exit sites biogenesis and the organisation of the microtubule network.

The Pepperkok team develops novel approaches to study the temporal and spatial organisation of membrane traffic and organelle biogenesis in the secretory pathway.

Future plans

We will study the complement of components that our genome-wide screens identified as being involved in the early secretory pathway in further detail. An important question in this context will be if and how they participate in the temporal and spatial organisation of ER-exit sites and their function, and the biogenesis of the Golgi complex.

Ultimately, we hope to be able to define and understand the molecular network(s) underlying trafficking at the ER/Golgi boundary and Golgi function, also considering their relationship to other cellular processes such as transcriptional control, lipid or general metabolism, or signalling and thus contribute towards a global molecular understanding of the living cell.
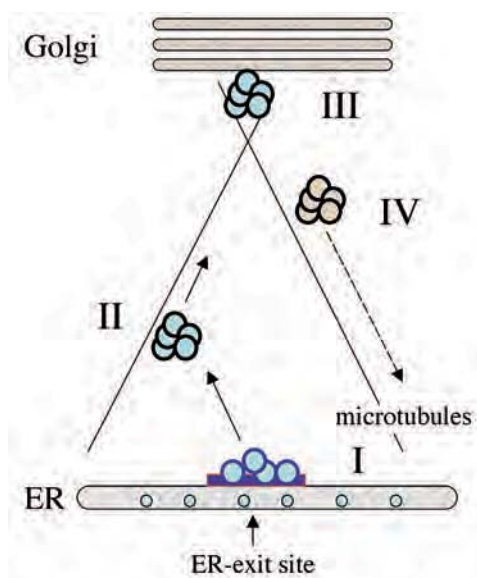

Figure 1: The four steps involved in ER to Golgi transport in mammalian cells. (I): Biogenesis of COPII coated vesicles occurs at specialised ER exit sites of the ER.(II): COPII vesicles homotypically fuse to form larger vesicular tubular transport carriers (VTCs) that are transported to the Golgi complex along microtubules.(III): VTCs arrive at the Golgi complex and fuse to it todeliver their cargo. (IV): Transport machinery and misrouted proteins are return back to the ER by a distinct class of carriers.
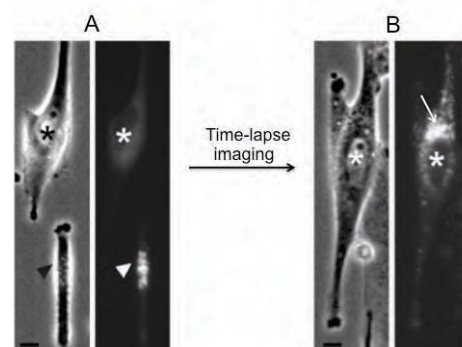

Figure 2: (A) Cells are cut by laser nano-surgery to generate a Golgiless karyoplast and Golgi containing Golgiplasts (arrowhead). Karyoplasts are then followed by time-lapse microscopy to monitor *de novo* Golgi biogenesis in living cells (B). The arrowhead points to the Golgi-like structure reforming after nano-surgery in karyoplasts.

Advanced optical techniques for deep tissue microscopy

Robert Prevedel

PhD 2009, University of Vienna, Austria.

Postdoctoral research at the University of Waterloo, Canada, and the Research Institute of Molecular Pathology and Max F. Perutz Laboratories, University of Vienna, Austria.

Group leader at EMBL from August 2016.

SELECTED REFERENCES

Rupprecht P, *et al.* (2015) Optimizing and extending light-sculpting microscopy for fast functional imaging in neuroscience. *Biomedical Optics Express* 6, 353-68

Prevedel R, *et al.* (2014) Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy. *Nature Methods* 11, 727-30

Schrödel T, *et al.* (2013) Brain-wide 3D imaging of neuronal activity in Caenorhabditis elegans with sculpted light. *Nature Methods* 10, 1013-20

Prevedel R, *et al.* (2007) High-speed linear optics quantum computing using active feed-forward. *Nature* 445, 65-9

The Prevedel group develops new optical techniques for investigating dynamic cellular processes deep inside tissue in vivo.

Light microscopy has revolutionized our understanding in many areas of biology, and over the years tremendous progress has been achieved by imaging cellular and subcellular processes in transparent, fixed, or thin sample preparations. Yet in order to obtain a complete understanding of biological processes, in-vivo studies inside thick, three-dimensional living tissues are often required. However, when light interacts with thick biological tissue, the process of light scattering leads to low-resolution, 'blurry' images and an effective loss of excitation power with increasing imaging depth. This has severely limited the biomedical usefulness of light microscopy to, e.g., cultured cells *in vitro* or superficial layers of tissue *in vivo*. New approaches and tools are thus required to noninvasively image biological function at depth inside living tissue with sufficient resolution, speed and contrast.

In the past, we have developed novel optical techniques for high-speed imaging, with a particular focus on functional imaging in the neurosciences. Amongst others, we have put forward a two-photon microscopy technique based on light-sculpting that has enabled the first whole-brain calcium imaging in *C. elegans*. In other work, we have established light-field deconvolution microscopy, an elegant approach to perform volumetric imaging that achieves unprecedented acquisition speeds while requiring no mechanical scanning. Currently, we extend our imaging methods to the scattering tissue domain and combine our approach with longer-wavelength excitation and red-shifted indicators, which collectively allow for larger imaging depth. Together with our collaborators, we apply our methods to study neuronal activity in a range of model organisms such as *C. elegans*, zebrafish larvae and behaving mice.

Future plans

The future focus of the group is to push the frontiers of deep tissue microscopy in terms of imaging depths and resolution by developing advanced and innovative optical imaging techniques. To do so we will draw from diverse fields such as multi-photon microscopy, active wave-front shaping, photo-acoustics as well as computational imaging. In particular, we aim to bridge the technological gap in spatial resolution and imaging depth that exists between optical microscopy and longer wavelength approaches such as photo-acoustics. We intend to combine these into multi-modal imaging systems that allow studying cellular processes and dynamics at depths inaccessible so far by conventional microscopy. Our multidisciplinary team comprises of physicists, engineers, computer scientists and biologists, and we engage in close collaboration with fellow groups at EMBL in the fields of cell and developmental biology as well as neuroscience.

Whole-brain Ca^{2+}-imaging of larval zebrafish *in vivo* using light-field deconvolution microscopy at 20Hz volume rate. a) Maximum intensity projection of cross sections with individual cells represented by colored overlays (scale bar 100μm). b) Extracted Ca2+ intensity signal ($\Delta F/F_0$) of GCaMP5 fluorescence using spatial filters shown in a. Each row shows a time-series heat map. Arrow denotes the addition of an aversive odor.

Cellular Nanoscopy
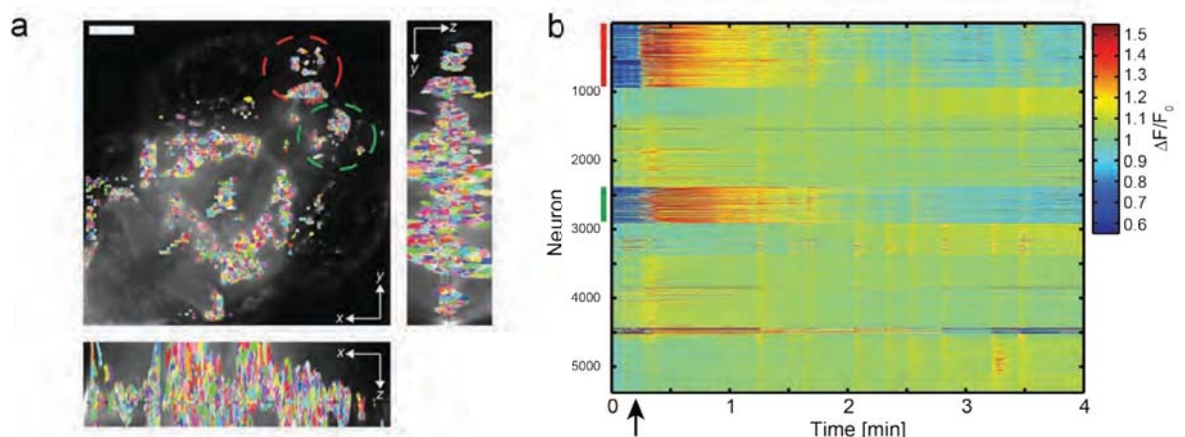
Jonas Ries

PhD 2008, TU Dresden.

Postdoctoral research research at the ETH Zurich.

Group Leader at EMBL Heidelberg since 2012.

SELECTED REFERENCES

Picco, A. *et al.* (2015) Visualizing the functional architecture of the endocytic machinery. *eLife* 4, e04535

Deschamps J, Mund M, & Ries J (2014) 3D superresolution microscopy by supercritical angle detection. *Opt Express* 22, 29081-91

Ries J, *et al.* (2012) A simple, versatile method for GFP-based super-resolution microscopy via nanobodies. *Nat. Methods* 9, 582-4

Schoen I, *et al.* (2011) Binding-activated localization microscopy of DNA structures. *Nano Lett.* 11, 4008-11

The resolution of optical microscopy is limited by diffraction to about 200 nm, which is much larger than the relevant length-scales in cell biology, defined for example by the size of organelles or multi-molecular complexes. Single-molecule localisation-based super-resolution microscopy (localisation microscopy) overcomes this limit by stochastic activation and subsequent localisation of individual fluorophores, reaching a resolution in the 10 nm range.

In the past, we worked on improved labelling schemes for super-resolution microscopy. We established nanobodies as tiny, high-affinity labels, which allow any GFP-tagged protein to be used directly for localisation microscopy. As an alternative to using photo-switchable fluorophores, we introduced binding-activated localisation microscopy (BALM), which employs fluorescence enhancement of fluorogenic dyes upon binding to target structures for superresolution microscopy, to study DNA structures and alpha-synuclein amyloids and demonstrated a superb labelling density combined with a very high resolution.

Currently, one focus of the group is the development of new tools for superresolution microscopy. In one project, we are establishing a robust and simple method for isotropic 3D resolution based on supercritical angle fluorescence detection. Furthermore, we aim at measuring absolute copy numbers of proteins in large complexes by using artificial brightness standards. Combining localisation microscopy with electron microscopy in a correlative approach allows us to add molecular specificity to the ultrastructure. Single-molecule microscopy with light-sheet illumination reduces the background in thick samples.

A second focus of our work is the application of our newly developed tools to address cell biological questions. Here, we are aiming to chart a comprehensive superresolved structural picture of the endocytic machinery as well as of the kinetochore complex in *S. Cerevisiae*. This has been impossible so far with conventional techniques due to their complexity and small size. Furthermore, we are investigating intracellular aggregation of Parkinsons' alpha-synuclein.

We are also developing novel data analysis tools and an open-source software platform for super-resolution microscopy. This will allow us to extract information about protein structures from super-resolution microscopy data.

Future plans

Our vision is to establish super-resolution microscopy as a tool for structural cell biology *in situ* to bridge the methodological gap that currently exists between cell biology and structural biology techniques. We aim to push its limits on all fronts to establish a technique which combines nanometer 3D resolution with maximum labelling efficiencies, absolute measurements of protein copy numbers, precise dual-colour measurements, high-throughput for large scale statistics and novel data analysis approaches, to address exciting biological questions, which were previously inaccessible.

The Ries group develops cutting -edge super-reso- lution microscopy methods to deter- mine structures of multi-protein assemblies in the cellular context.

Figure 1: Dual-colour super-resolution images of Cdc11 (red) and the cell-wall marker ConA (green) show the formation and disassembly of the Cdc11 ring.
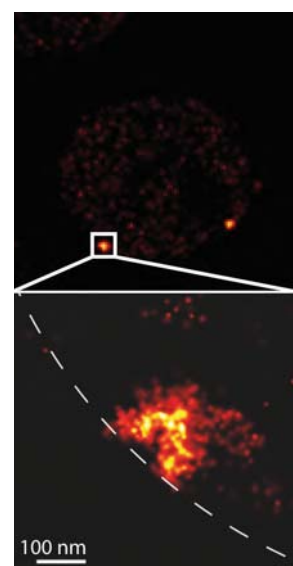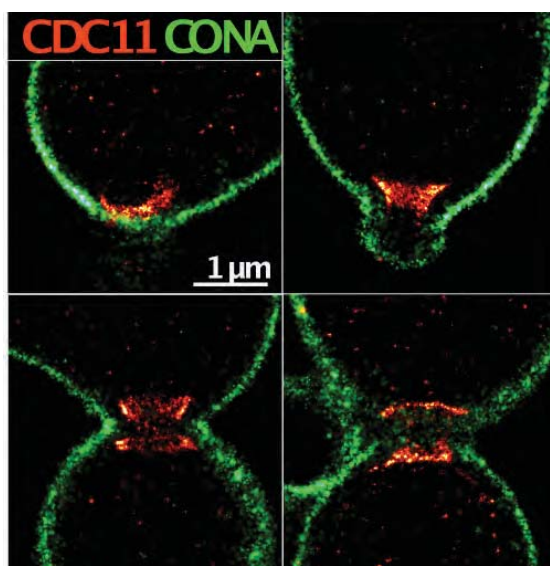

Figure 2: Actin in yeast. Yeast expressing Abp1-mMaple imaged by localisation microscopy. Five fixed example sites at different endocytic time points.

Chemical cell biology

Carsten Schultz

PhD 1989, University of Bremen, Germany.

Postdoctoral research at the University of California, San Diego.

Habilitation 1997, Organic Chemistry, University of Bremen, Germany.

Group leader, MPI for Molecular Physiology, Dortmund, Germany.

Group leader at EMBL since 2001. Senior Scientist since 2008.

Group leader in the Molecular Medicine Partnership Unit (MMPU).

SELECTED REFERENCES

Hoffmann JE, *et al.* (2015) Highly stable trans-cyclooctene amino acids for live cell labeling. *Chem. Eur. J.* 21, 12266-70

Hu H, *et al.* (2015) In vivo visualization of osteoarthritic hypertrophic lesions. *Chem. Sci.* 6, 6256-61

Nadler A, *et al.* (2015) Exclusive photorelease of signaling lipids at the plasma membrane. *Nat. Commun.* 6, 10056

Höglinger D, *et al.* (2015) Sphingosine releases calcium from lysosomes. *eLife* 4

The Schultz group develops tools for imaging and for manipulating cellular enzyme activities, with a particular emphasis on lipid signalling in diabetes and the hereditary disease cystic fibrosis.

Past projects: Our research has previously focused on finding novel ways to stimulate chloride and water secretion of epithelial cells in understanding cystic fibrosis (CF). Our compounds helped to investigate some of the underlying intracellular signalling pathways and provided drug candidates to eventually treat CF patients. Of particular significance was the development of chemical methods to convert highly polar signalling molecules like cyclic nucleotides, inositol phosphates, and phosphoinositides to membrane-permeant, bioactivatable derivatives ('prodrugs') (Schultz 2003; Laketa *et al.* 2009, Laketa *et al.* 2014).

Current projects: Our interest in CF has shifted to the development of lung emphysema – the ultimate cause of death in the patient. In a truly translational collaboration with the Mall group (MMPU), we develop FRET reporters to sense enzyme activities detrimental to lung tissue, such as macrophage and neutrophil elastases. At the cell biology level, our interest focuses on signalling networks regulated by G-protein-coupled and growth factor receptors. We developed a wide range of fluorescent reporter molecules, either genetically encoded (Piljić & Schultz, 2011) or as small molecule fluorescent probes (see figure). We hope to provide a more complete picture of the signalling network and to help find compounds beneficial in unravelling basic principles in signal transduction and, ultimately, in ion and enzyme secretion relevant to CF patients or in insulin secretion of ß-cells. In addition, we prepared a large number of tools to manipulate signalling networks and are able to locally activate the important messenger such as PIP3 and DAG with a light flash in subcellular resolution in living cells (Mentel *et al.* 2011; Nadler *et al.* 2013, Nadler *et al.* submitted). In order to specifically label molecules with fluorophores in intact cells, we prepare highly stable unnatural amino acids that rapidly and irreversibly undergo cycloaddition reactions (click chemistry) with unsurpassed speed and study their application in collaboration with the Lemke group.

Hot projects: Currently, we are very excited about making highly charged dyes pass cell membranes. In collaboration with the Häring group, we are developing a method to visualise protein-protein interactions in cells in real time. By using a novel set of photoactivatable lipid molecules, we are able to modulate the signalling underlying insulin secretion, likely to provide new means of identifying targets important in diabetes.

Future plans

We will continue work aimed at bringing fluorescent reporters for enzyme activities closer to the clinic. We will also focus on lipid signalling and lipid-controlled cell biology, and examine the effect of sphingo- and phospholipids on endocytosis, lipid trafficking, and insulin secretion. In addition, we will improve our possibilities to fluorescently label molecules in intact cells by using faster and more complete bioorthogonal reactions and new fluorophores. Most projects rely on organic chemistry and the group has a significant number of preparative chemists at the graduate student and postdoc level. The symbiosis of chemistry, biochemistry, and cell biology opens new doors and grants novel insights into how cells exhibit their function.

Figure 1: Several reporter and modulator molecules have been developed in our lab, including: small molecule sensors for lipases and proteases; genetically encoded reporters for kinase and phosphatase activities; membrane-permeant and photoactivatable lipid molecules; and lipid derivatives that can be fluorescently labeled in living cells.
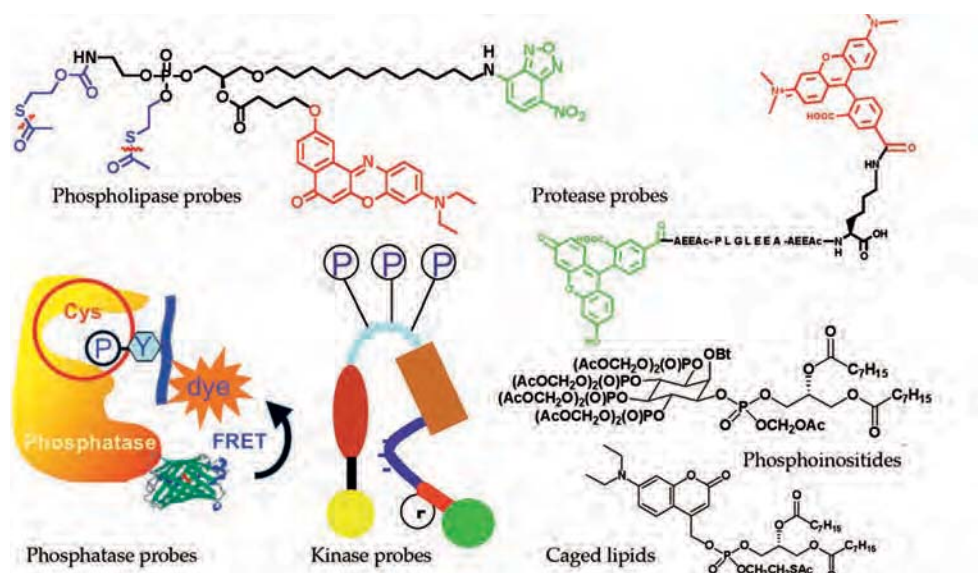
Volume correlative light and electron microscopy

Yannick Schwab

PhD 2001, Louis Pasteur University, Strasbourg.

Postdoctoral research at the University of Calgary, Canada and at the IGBMC, Illkirch, France.

Head of Electron Microscopy at the Imaging Center, IGBMC, Illkirch, France.

Facility head and team leader at EMBL since 2012.

SELECTED REFERENCES

Karreman MA, *et al.* (2016) Fast and precise targeting of single tumor cells in vivo by multimodal correlative microscopy. *J Cell Sci.* 129, 444-56

Karreman MA, *et al.* (2014) Correlating intravital multi-photon microscopy to 3D electron microscopy of invading tumor cells using anatomical reference points. *PLoS ONE* 9, e114448

Goetz JG, *et al.* (2014) Endothelial cilia mediate low flow sensing during zebrafish vascular development. *Cell Rep* 6, 799-808

Durdu S, *et al.* (2014) Luminal signalling links cell communication to tissue architecture during organogenesis. *Nature* 515, 120-4

Correlative light and electron microscopy (CLEM) is a set of techniques that allow data acquisition with both imaging modalities on a single object. It is a growing field that now includes a large variety of strategies, reaching a high degree of precision, even in complex biological models. Before joining EMBL, we were developing tools and protocols to track rare objects or dynamic phenomena on cultured cells and bulk specimen such as nematodes and murine tissues.

One common challenge when trying to combine imaging modalities on the same sample is to identify space cues (external or internal) to track single objects when switching from light microscopy (LM) to electron microscopy (EM). On adherent cultured cells, we have previously developed specific substrates with coordinates to precisely record the position of cells (Spiegelhalter *et al.*, 2009). Currently, we are exploiting these approaches to develop new workflows allowing the study of a higher number of cells.

On more complex specimens, such as multicellular organisms, this targeting is even more critical, as systematic EM acquisition of their entire volume is close to impossible. For this reason, we are developing new methods to map the region of interest (ROI) within large living specimens, taking advantage of structural hallmarks in the sample that are visible with both LM and EM. The position of the ROI is mapped in 3D by confocal or multiphoton microscopy and then tracked at the EM level by targeted ultramicrotomy (Kolotuev *et al.* 2009; 2012; Goetz *et al.* 2014). Relying on structural features of the sample as anchor points, the cell or structure of interest can then be retrieved with sub-micrometric precision (Durdu *et al.* 2014, Karreman *et al.* 2014, 2016).

Future plans

In parallel to the fast evolution of CLEM techniques over the past decade, acquisition methods in electron microscopes have significantly evolved with special breakthroughs in the volume analysis of cells by TEM tomography and automated serial imaging in scanning electron microscopy (ASI-SEM). Our team, in collaboration with other scientists and our industrial partners, combines these advanced techniques to perform CLEM in the 3D space on complex model specimens for cell and developmental biology. We aim to develop new techniques and software to facilitate and automate the correlation and acquisition of large amounts and volumes of sample. By automating these tedious procedures, we intend to improve the throughput of data collection.

The Schwab team is interested in developing tools for the 3D correlation of data generated by fluorescent imaging and by electron microscopy.

Live cell imaging | Coordinates recording
Targeting in the FIBSEM | Volume Electron Microscopy

Intravital imaging | Micro-CT | Volume Electron Microscopy
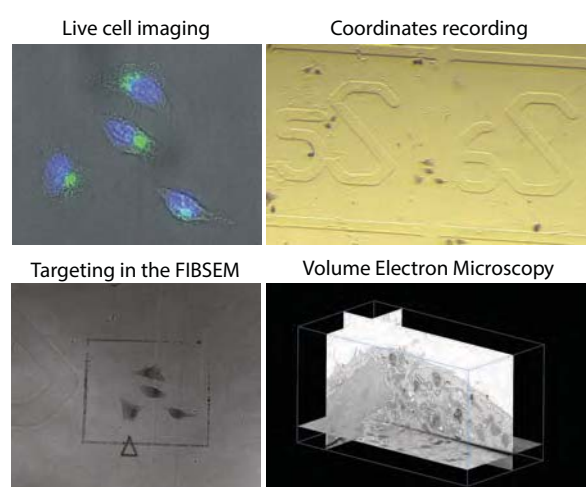Ultrastructural analysis

Figure 1: On adherent cultured cells, CLEM relies on coordinates that are visible by both light and electron microscopy. Precise targeting enables the acquisition of selected subcellular volumes with a FIB-SEM.
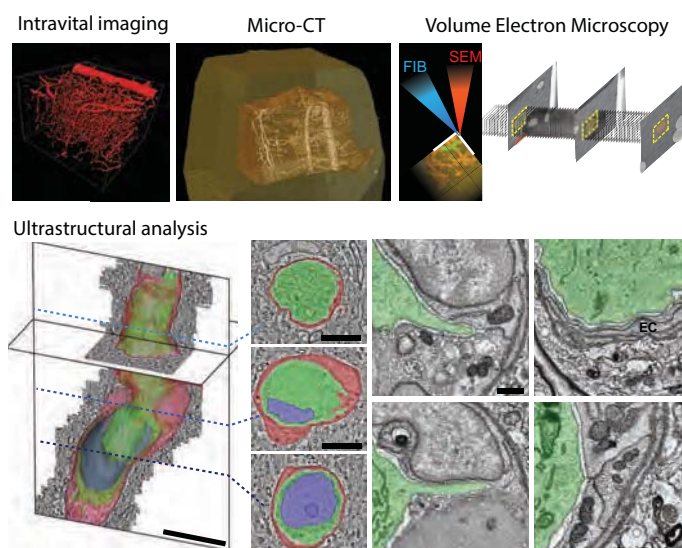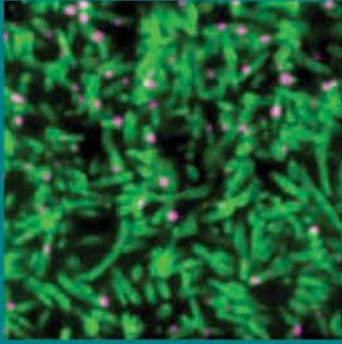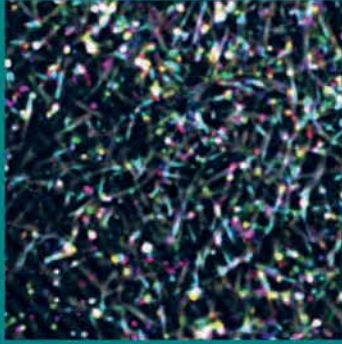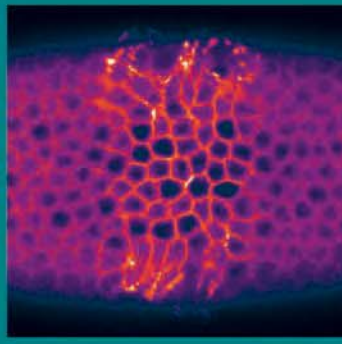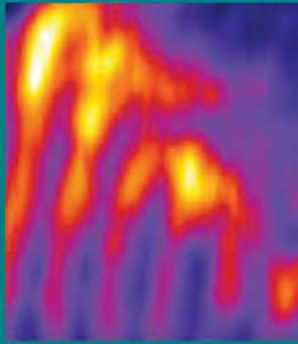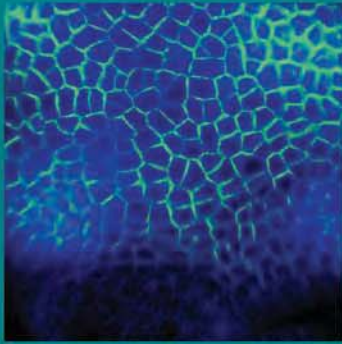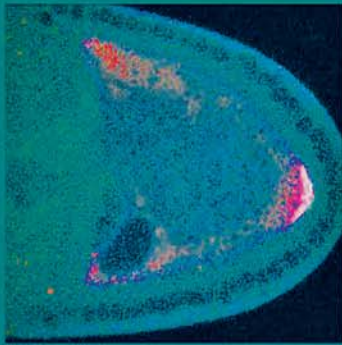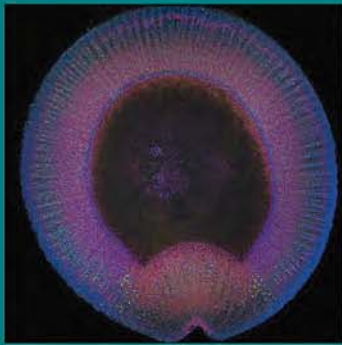
Figure 2: On multicellular specimens (here, mouse brain), our targeting methods enable studying identified cells (tumour cells) by multimodal correlative microscopy (intravital microscopy, microCT and FIB-SEM; from Karreman *et al.* 2016).
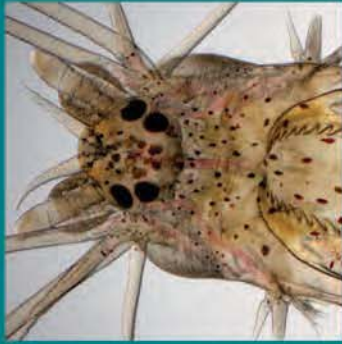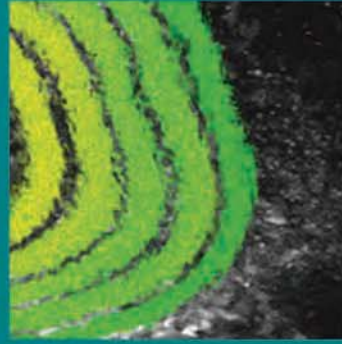
Developmental Biology

The development of living organisms requires precise coordination of all basic cellular processes, in space and time. Groups seek to elucidate the principles, mechanisms and dynamics of fundamental developmental events. Using animal and plant models, research in the Unit integrates numerous complementary approaches to understand how cellular and morphological processes are coordinated and evolve to shape and maintain living organisms in their environment.

A fundamental question in developmental biology is the mechanism by which symmetry is broken and cells with distinct fates are specified. Researchers in the Unit are studying a number of related research areas, including the mechanisms underlying cell polarisation, mRNA transport, and translational control in *Drosophila*; how auxin specifies different cell types in *Arabidopsis*; and a systems-level understanding of the symmetry breaking processes operating in the early mouse embryo.

During development, progenitor cells divide and differentiate into tissues of characteristic shape and function. Another aim is to elucidate how cells in the early *Drosophila* embryo reorganise their content in response to the expression of key developmental transcription factors and, specifically, how tissue-specific gene expression controls protein and membrane trafficking, and how this trafficking regulates cell fate and behaviour.

Elucidating the temporal organisation of embryonic development is a further goal. Using the mouse model, the mechanisms controlling overall developmental rate at an organismal level, as well as the timing of individual patterning processes and the dynamics of underlying signalling pathways, are being investigated. Analysis of novel mouse reporter lines using real-time imaging techniques allows visualisation of the activity and dynamics of signalling pathways in the context of a developing embryo.

The marine annelid *Platynereis* is an ideal model for exploring the evolution of cell types. Large-scale expression profiling at cellular resolution has revealed the evolutionary origin of the vertebrate hypothalamus. Using this model, research in the Unit aims at solving one of the major remaining mysteries in animal evolution: the evolution of the central nervous system.

Several groups seek to understand both normal development and its deviations in disease. During brain development, vast numbers of neurons are targeted for death and are cleared rapidly and efficiently by a resident lineage of phagocytes, the microglia. Combining live imaging and genetic approaches, the dynamic relationship between neurons and microglia in zebrafish is actively investigated.

Re-shuffling of regulatory inputs after chromosomal rearrangements is the likely cause of several human genetic disorders. Focusing on the regulatory architecture of key developmental loci, another goal in the Unit is to understand the molecular mechanisms that control functional interactions between genes and remote *cis*-regulatory elements, and to determine how they contribute to phenotypic variations during vertebrate evolution and in humans.

Anne Ephrussi
Head of the Developmental Biology Unit

Evolution of the nervous system in bilateria

Detlev Arendt

PhD 1999, Albert-Ludwigs-Universität, Freiburg.

Team leader at EMBL since 2002.

Group leader and senior scientist since 2007.

Academic mentor, postdoctoral training since 2007.

ERC Advanced Investigator since 2012.

SELECTED REFERENCES

Tosches MA & Arendt D. (2013) The bilaterian forebrain: an evolutionary chimaera. *Curr. Opin. Neurobiol.* 23, 1080-9

Christodoulou F, *et al.* (2010) Ancient animal microRNAs and the evolution of tissue identity. *Nature* 463, 1084-8

Tomer R, *et al.* (2010) Profiling by image registration reveals common origin of annelid mushroom bodies and vertebrate pallium. *Cell* 142, 800-9

Arendt D. (2008) The evolution of cell types in animals: emerging principles from molecular studies. *Nat. Rev. Genet.* 9, 868-82

By studying and comparing simple marine organisms, the Arendt group looks to understand the origin and evolution of our central nervous system.

We are intrigued by one of the remaining great mysteries in animal evolution: how did our central nervous system (CNS) come into existence? What did it look like at first and how did it function? We are especially interested in the CNS of an extinct animal known as Urbilateria, the last common ancestor of humans, flies and most other 'higher' animals that live today, which lived some 600 million years ago in the ocean.

Our lab has chosen to investigate a new molecular animal model, the marine annelid *Platynereis dumerilii*. As a 'living fossil', *Platynereis* represents an ideal connecting link between vertebrates and the fast evolving protostome models, *Drosophila* and *Caenorhabditis*. *Platynereis* is amenable to high throughput imaging techniques and functional interference approaches – for example the first genetic knockout lines have been generated. With the recent development of the PrImR (Profiling by Image Registration) resource, *Platynereis* has become the first animal model for which gene expression profiling data can be obtained in cellular resolution for the whole organism. We have discovered that their brains harbour sensory-associative parts and a neurosecretory centre that corresponds to the vertebrate pallium and hypothalamus, respectively. A clear picture is emerging that the *Platynereis* brain harbours many cell types so far known only for vertebrates, but in a much more simple and different overall arrangement, revolutionising our current understanding of brain evolution.

To broaden our comparative approach, we have introduced two new model species to the lab, the lancelet amphioxus and the sea anemone *Nematostella*, representing distinct divisions of the animal kingdom: chordates and cnidarians. Amphioxus has a very simple brain uniting invertebrate- and vertebrate-like features. The *Nematostella* nervous system is very simple and is a good proxy for an early stage of nervous system evolution.

Future plans

Our aim is to gain a systems view of the *Platynereis* brain and nervous system and to track the evolutionary history of all constituent cell types by identifying and investigating their evolutionary counterparts in sea anemone and amphioxus. This will involve investigations of cell type-specific gene regulatory networks in all species as well as neurobiological and behavioural approaches.

Our ERC-funded project – BrainEvoDevo – aims at generating a neuron-type atlas of the annelid larval brain by combining neuronal morphologies, axonal projections and cellular expression profiling for an entire bilaterian brain. Working with collaborators, it will be the first cellular resolution expression atlas for a whole animal nervous system involving early developmental and differentiation stages. Building on the Atlas, we will dissect *Platynereis* chemosensory-motor forebrain circuits, by laser ablation of GFP-labelled single neurons, gene knockout studies and behavioural assays based on microfluidics to explore duplication, divergence and expansion of neural circuits in central nervous system development and evolution.

We are also interested in exploring population genetics and the variability of development and differentiation in different habitats and conditions. To this end, we are collecting strains of *Platynereis* and amphioxus as part of the TARA Oceans expedition and as an active member of the EMBL Oceans Team.

As a 'living fossil', *Platynereis* represents an ideal connecting link between vertebrates and the fast evolving protostome models, *Drosophila* and *Caenorhabditis*. Genomic resources and molecular techniques have been generated that make it a model marine invertebrate for ocean biology and for organismal systems biology.

Timing of mammalian embryogenesis

Alexander Aulehla

MD 2002, Albert-Ludwigs-University, Freiburg.

Research at the MD Anderson Cancer Center, Houston, USA and the MPI, Freiburg.

PhD 2008, Paris VI University.

Postdoctoral research at the Stowers Institute, Kansas City, USA.

Group leader at EMBL since 2009.

ERC investigator since 2015.

SELECTED REFERENCES

Tsiairis C D & Aulehla, A. (2016) Self-organization of embryonic genetic oscillators into spatiotemporal wave patterns. *Cell* 164, 656-67

Lauschke VM, *et al.* (2013) Scaling of embryonic patterning based on phase-gradient encoding. *Nature* 493, 101-5

Aulehla A, *et al.* (2008) A beta-catenin gradient links the clock and wavefront systems in mouse embryo segmentation. *Nat. Cell Biol.* 10, 186-93

Aulehla A, *et al.* (2003) Wnt3a plays a major role in the segmentation clock controlling somitogenesis. *Dev. Cell* 4, 395-406

During an embryo's journey from a single cell to a complex organism, countless patterning processes unfold with remarkable precision, spatially but also in respect to temporal sequence, or timing. This temporal aspect of embryonic development is the focus of our research. How is time measured during embryonic development and what extrinsic and intrinsic signals control this timing? How are embryonic oscillators/clocks employed during patterning? What are the dynamics of signalling pathways?

To approach these questions, novel methodologies are required (see video 1, online). We are generating novel real-time reporter mouse lines using knock-in technology that enables visualisation and quantification of temporal dynamics at different levels in the context of mouse embryonic development. Using *in vivo* imaging, we are focusing on the somite segmentation clock, an oscillatory system that is thought to control the formation of the pre-vertebrae that form periodically in a head-to tail sequence within the paraxial mesoderm. In mouse embryos this clock, with a periodicity of around two hours, drives oscillatory activity of several signalling pathways (Wnt, Notch and Fgf signalling) in the developing mesoderm.

We recently developed an *ex vivo* assay that, in combination with real-time imaging reporters, has become instrumental for our approach: the assay recapitulates mesoderm patterning, including segment formation and spatio-temporally controlled oscillatory signalling activities, within the simplified context of a monolayer of primary mesoderm cells put in culture (see figure & video 2, online).

Scaling and phase-shifted oscillators: One fundamental property of vertebrate segment formation is its ability to maintain proportions even when overall embryo size is experimentally altered, a process termed scaling. Intriguingly, scaling behaviour can be observed in the *ex vivo* assay system as well. This enabled us to identify a novel scaling mechanism employing phase-shifted oscillatory activity (Lauschke *et al*, 2013). One major interest is how temporal devices, or oscillators, mechanistically encode spatial information for patterning, particularly at an integrated, higher-order level, so as to reveal emergent properties, incorporating mathematical modelling into our approach.

The Aulehla group studies how the precise timing and sequence of events that unfold as an embryo develops are controlled.

Oscillatory Wnt-signalling: This signalling pathway serves a multitude of evolutionarily conserved functions during development and has been shown to play an essential role during somite formation. Our novel real-time reporter system is designed to reflect oscillatory Wnt-signalling activity both at gene activity and at protein levels. This will enable us to determine how the striking oscillations of Wnt- signalling activity are generated in the first place and, moreover, to functionally test their role in embryonic patterning, particularly identifying the intrinsic and extrinsic factors that are responsible for controlling these oscillations within the segmentation process.

Future plans

- Quantitative (imaging) approach to understand the role of dynamic oscillatory signalling during patterning and scaling.

- Study of emergent properties of coupled oscillator populations.

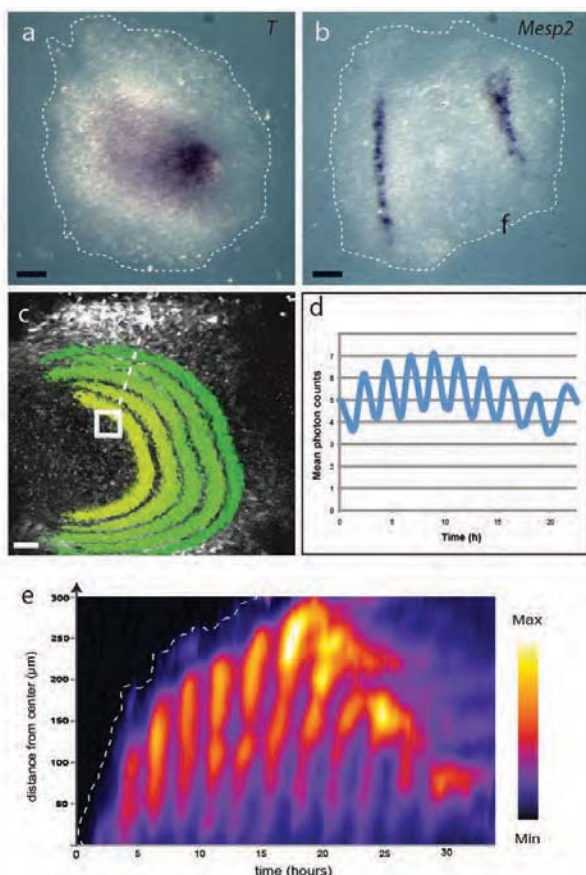- Discovery of oscillatory signalling phenomena during embryogenesis.

Figure 1: *Ex vivo* cell culture model for mesoderm patterning and oscillations. a, b) Primary mesoderm cells retain undifferentiated PSM identity in the centre of the culture, before initiating a differentiation program in the periphery. c) Snapshot of time-series using fluorescent lunatic fringe reporter mouse line (LuVeLu), overlaid with time-projection of activity patterns (in green) seen during the time-lapse recordings (see Video 2). d) Raw photon counts (measured in quadrant shown in c) demonstrating robust oscillatory activity for extended recording times). e) Time-space kymograph along dashed arrow in c. From this quantification, critical oscillation parameters, wave speed and phase-distributions can be calculated.

Cell dynamics and signalling during morphogenesis

Stefano De Renzis

MD 1997, University Federico II, Naples.

PhD 2002, EMBL Heidelberg.

Postdoctoral work at Princeton University.

Group leader at EMBL since 2008.

SELECTED REFERENCES

Guglielmi G, *et al.* (2015) An optogenetic method to modulate cell contractility during tissue morphogenesis. *Dev Cell* 35, 646-60

Reversi A, *et al.* (2014) Plasma membrane phosphoinositide balance regulates cell shape during *Drosophila* embryo morphogenesis. *J. Cell Biol.* 205, 395-408

Fabrowski P, *et al.* (2013). Tubular endocytosis drives remodelling of the apical surface during epithelial morphogenesis in *Drosophila. Nat Commun 4,* 2244

De Renzis S, *et al.* (2007) Unmasking activation of the zygotic genome using chromosomal deletions in the *Drosophila* embryo. *PLoS Biol. 5*, e117

Cell shape changes are of fundamental importance during embryonic development. How cells form and change shape during morphogenesis are the key questions addressed by the De Renzis group.

Tissue morphogenesis is triggered by shape changes in single cells or group of cells. This remodelling depends on a complex interaction between cortical forces exerted by the actin cytoskeleton and membrane homeostasis (i.e. vesicular trafficking and lipid metabolism). Using the early *Drosophila* embryo as model organism, we wish to understand how membrane trafficking and cytoskeletal dynamics are regulated during morphogenesis and how this, in turn, impacts on specific cell and tissue behaviour.

We have recently developed new optogenetic methods to manipulate key components of the membrane transport machinery and cytoskeletal regulators with cellular precision *in situ* during embryonic development (Figure 1). Using these novel tools in combination with two-photon microscopy we can now control protein function with high-spatio temporal precision and thus functionally link subcellular to supracellular processes during morphogenesis. We are also implementing optogenetics to study signalling systems and induce tissue-specific differentiation programs at will.

Future plans

Using a combination of imaging, genome engineering, optogenetic, and biochemical approaches, we wish to elucidate how machineries controlling intracellular trafficking re-organise during differentiation and how this in turn impacts on global changes in tissue morphology. One long-term goal of the group will be to reconstitute morphogenetic processes in naïve tissues using synthetic approaches.
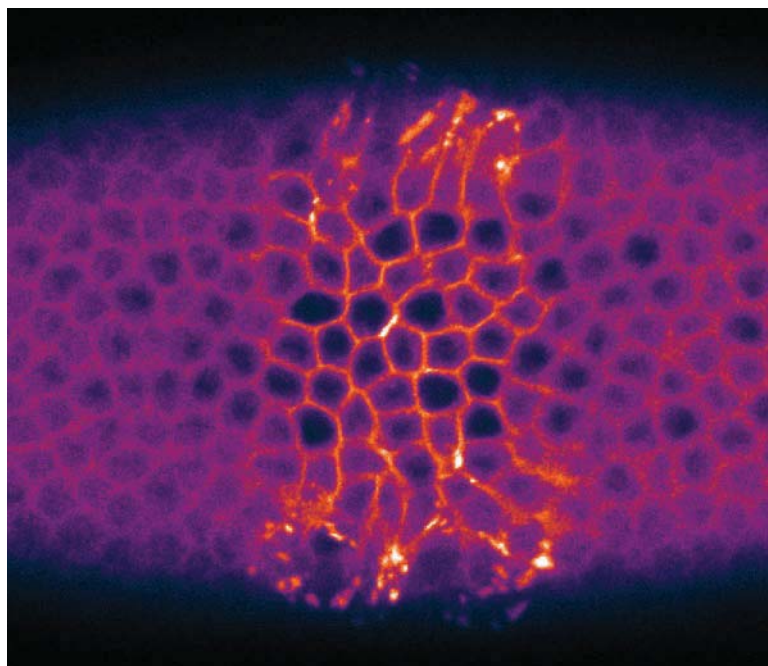

Figure 1: Surface view of a *Drosophila* embryo at the onset of gastrulation. A pulse of laser light has been used to trigger the translocation of an actin inhibitor to the plasma membrane in a central stripe of cells. Optogenetically activated cells lose their regular hexagonal shape and will fail to gastrulate. Image is pseudo-coloured according to fluorescence intensity (blue=low/yellow=high).

RNA localisation and localised translation in development

Anne Ephrussi

PhD 1985, Massachusetts Institute of Technology.

Postdoctoral research at Harvard University and Whitehead Institute, MIT, Cambridge, Massachusetts.

Group leader at EMBL since 1992.

Head of EICAT since 2005.

Head of Unit since 2007.

SELECTED REFERENCES

Jeske M, *et al.* (2015) The crystal structure of the *Drosophila* germline inducer Oskar identifies two domains with distinct vasa helicase- and RNA-binding activities. *Cell Rep* 12, 1-12

Halstead J, *et al.* (2015) An RNA biosensor for imaging the first round of translation from single cells to living organisms. *Science* 347, 1367-71

Ghosh S, *et al.* (2012) Control of RNP motility and localization by a splicing dependent structure in *oskar* mRNA. *Nat. Struct. Mol. Biol.* 19, 441-9

Chekulaeva M, Hentze MW. & Ephrussi A. (2006) Bruno acts as a dual repressor of *oskar* translation, promoting mRNA oligomerization and formation of silencing particles. *Cell* 124, 521-33

Post-transcriptional RNA regulation is central to organismal development and function. The combination of intracellular mRNA localisation and localised translation is a powerful strategy that allows quick and local deployment of protein activities in cells in response to extrinsic signals. mRNA localisation is widespread and conserved from yeast to man. It is involved in the establishment of cell asymmetry, is particularly evident in large cells, such as eggs and neurons, and has key roles in cell fate decisions, cell migration, cell morphology, and polarised cell functions. Asymmetric RNA localisation can be achieved by different mechanisms, such as active transport of RNAs by motor proteins moving on cytoskeletal elements, local protection of RNAs from degradation, facilitated diffusion and trapping.

In the large *Drosophila melanogaster* oocyte, asymmetrically localised cell fate determinants specify the body axes and pattern the future embryo and fly, making it an ideal model for the study of RNA localisation. During oogenesis, the embryonic axis determinant-encoding *oskar, bicoid* and *gurken* mRNAs are transported to specific sites within the oocyte, where they are anchored and locally translated, thus ensuring spatial restriction of their protein products. A polarised cytoskeleton and specific motor proteins mediate mRNA transport and anchoring within the oocyte. Our research combines live-imaging, super-resolution microscopy, genetics and biochemistry to understand how mRNAs are transported, anchored and locally translated.

Of particular interest is *oskar* mRNA, which encodes Oskar protein, which is endowed with the unique ability to induce the formation of germ cells, the cells that ensure perpetuation of the species. *oskar* mRNA is transported to the posterior pole of the oocyte, where it is translated. Oskar protein, which contains LOTUS and SGNH-hydolase-like domains, nucleates formation of germ granules, RNP complexes containing RNAs and proteins essential for germ cell formation and function. Taking a structure-function approach that combines structural biology, genetics, proteomics and transcriptomics, we are addressing how germ granules form and function.

Other classes of RNAs also have important developmental functions. We are investigating the roles of long non-coding RNAs, piRNAs, and of non-canonical RNA binding proteins in *Drosophila* development.

With its exceptional genetic tools, *Drosophila* is ideally suited for investigation of the fundamental mechanisms that govern animal development.

The Ephrussi group dissects the mechanisms that underlie intracellular RNA transport and localised translation – fundamental processes that mediate the functional polarisation of cells during development and in the nervous system.

Future plans

We combine genetics, biochemistry and a broad spectrum of cell biological and imaging approaches to study:

- How RNA targeting signals and proteins associated to form RNPs competent for mRNA transport and translational control.
- Spatial and temporal control of translation.
- Polarisation of the cytoskeleton.
- The roles and regulation of cytoskeletal motors in RNA localisation.
- Developmental roles of non-canonical RNA binding proteins.
- Germ plasm assembly and function.
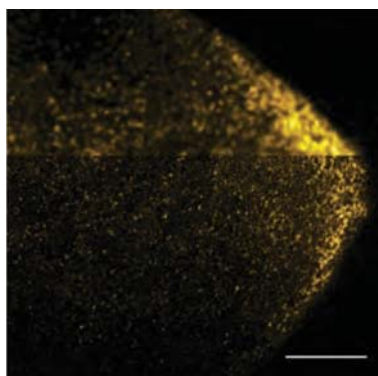
Confocal (upper half) and super-resolution (STED, lower half) micrograph of localising *oskar* mRNPs in close proximity to the posterior pole of a developing *Drosophila* oocyte. *oskar* mRNPs are labeled with a an array of 37 short single stranded DNA oligonucleotides complementary to *oskar* mRNA, each labeled with a single Atto-Rho14 fluorescent dye (images courtesy of Imre Gaspar).
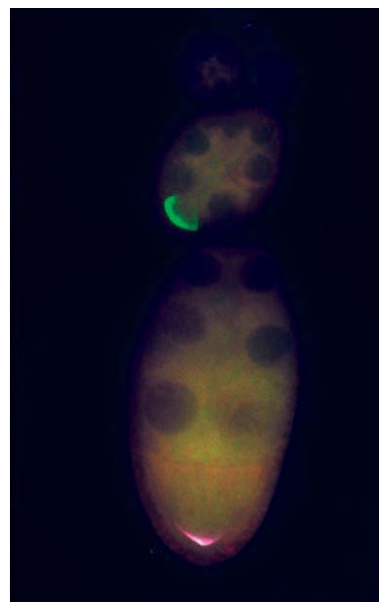
An mRNA biosensor (TRICK) for *oskar* mRNA visualises the first round of *oskar*-TRICK translation in *Drosophila* oocytes. In early-stage oocytes *oskar*-TRICK mRNA is labeled by both NLS-PCP-GFP (green) and NLS-MCP-RFP (red), fluorescent RNA-binding proteins, indicating translational repression. In later stages, the NLS-PCP-GFP fluorescent signal is reduced at the posterior pole and Oskar protein (blue) is detected by immunofluorescence, consistent with translation of a portion of the transcripts (see Halstead*, Lionnet*, Wilbertz*, Wippich* *et al., Science* 2015).

Developmental patterning in plants

Marcus Heisler

PhD 2000, Monash University, Australia.

Postdoctoral research at the California Institute of Technology.

Senior Research Associate at the California Institute of Technology 2007-2009.

Group leader at EMBL since 2009.

ERC Investigator.

SELECTED REFERENCES

Brennecke P, *et al.* (2013) Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* 10, 1093-5

Heisler MG, *et al.* (2010) Alignment between PIN1 polarity and microtubule orientation in the shoot apical meristem reveals a tight coupling between morphogenesis and auxin transport. *PLoS Biol.* 8, e1000516

Hamant O, *et al.* (2008) Developmental patterning by mechanical signals in Arabidopsis. *Science* 322, 1650-5

Jönsson H, *et al.* (2006) An auxin-driven polarized transport model for phyllotaxis. *Proc. Natl. Acad. Sci. U.S.A.* 103, 1633-8

Using A. thaliana *as a model, the Heisler group seeks to understand patterning in plant development and how it is established and regulated.*

In addition to providing us with the air we breathe, the food we eat and much of the energy and materials we use, plants exhibit a unique beauty associated with their strikingly symmetrical patterns of development. Multicellularity also evolved independently in plants giving us an opportunity to compare and contrast the developmental strategies used in different kingdoms.

Lateral organ formation in the model plant species *Arabidopsis thaliana* provides an ideal system for investigating plant development since it involves the coordination of several fundamental processes, including cell polarity, gene expression and morphogenesis. Our previous work reveals that patterns of cell polarity control both morphogenesis at the cellular level as well as at the tissue level. This integration occurs through the co-alignment of microtubule arrays with the polar localisation patterns of the auxin efflux carrier PIN1. The microtubule cytoskeleton regulates growth direction at the cellular level, while PIN1 works to concentrate the hormone auxin at the tissue level to localise growth. Our data so far suggests a role for mechanical stresses in orienting these factors and we are further investigating this possibility.

More recently we have found that organogenesis correlates spatially with a boundary between the expression domains of genes normally associated with the top (dorsal) and bottom (ventral) tissues of leaves. In fact, we have found that juxtaposition of dorsal and ventral gene expression domains is necessary for leaves to initiate properly and the creation of new dorsoventral gene expression boundaries provokes the formation of new leaf tissues along the boundary. Similar juxtaposition dependent development occurs in response to the formation of dorsoventral boundaries during insect wing and vertebrate limb development suggesting a similar logic operates in these otherwise very different contexts.

Future aims

Our current and future research focuses on the establishment and function of dorsoventral boundaries in plants. Our main questions include: How do dorsoventral gene expression boundaries regulate organ morphogenesis and positioning, for example cell polarity patterns? And how are dorsoventral gene expression boundaries established and regulated?
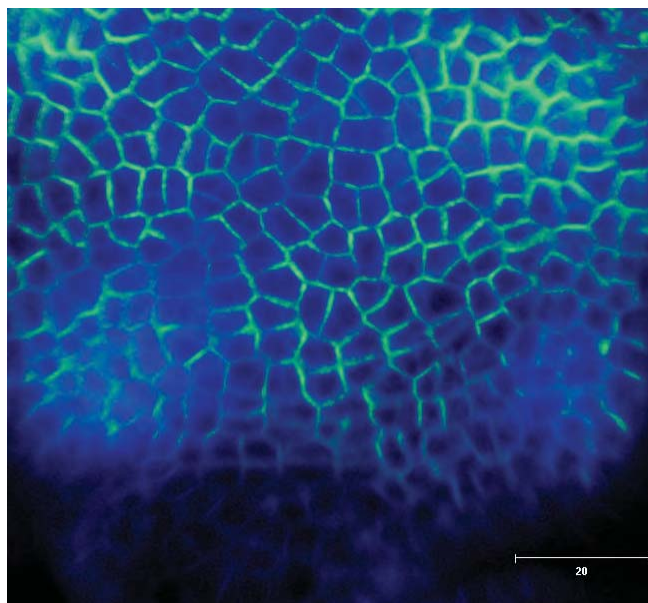
Confocal projection showing polar localisation of the auxin efflux carrier PIN1 fused to GFP. At organ inception PIN1 polarities are directed away from adjacent organ sites and towards the new site.

Symmetry breaking and self-organisation

Takashi Hiiragi

MDPhD 2000, Kyoto University, Japan.

Postdoctoral research at the Max Planck Institute of Immunobiology, Freiburg, Germany.

Group leader at the MPI of Immunobiology 2002-7.

Independent group leader at the MPI for Molecular Biomedicine, Münster, 2007-11.

Group leader at EMBL since 2011.

ERC Investigator.

SELECTED REFERENCES

Maître J-L, *et al.* (2015) Pulsatile cell-autonomous contractility drives compaction in the mouse embryo. *Nat. Cell Biol.* 17, 849–855

Dietrich J-E, *et al.* (2015) Venus trap in the mouse embryo reveals distinct molecular dynamics underlying specification of first embryonic lineages. *EMBO reports 16,* 1005-21

Ohnishi Y, *et al.* (2014) Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. *Nat. Cell Biol. 16,* 27-37

Wennekamp S, *et al.* (2013) A self-organization framework for symmetry breaking in the mammalian embryo. *Nat. Rev. Mol. Cell Biol.* 14, 452-9

Mammalian development begins with cells that are equivalent in their position and developmental potential. This initial symmetry among cells is broken during development to form the blastocyst consisting of two major cell types, the inner cell mass and trophectoderm, which are distinct in their position and gene expression. Recent studies unexpectedly revealed that morphogenesis and gene expression is highly dynamic and stochastic during this process (figure 1). Determining which signal breaks the initial symmetry and how stochastic gene expression leads to the reproducibly patterned blastocyst remain fundamental open questions about the beginning of mammalian life.

We have developed new imaging and experimental systems to monitor early mouse development at unprecedented spatio-temporal resolution. Using genetics, high-resolution microscopy and computational analysis, we established the complete map of mouse pre-implantation development and identified the precise moment of symmetry breaking. This breakthrough now provides the basis to investigate the cellular and molecular mechanism of symmetry breaking.

Upon symmetry-breaking, gene expression varies stochastically between cells before it progressively stabilises into a reproducible pattern segregating the first lineages of the blastocyst. This self-organising process likely relies on feedbacks between gene regulatory networks and cell and tissue mechanics to achieve a coordinated developmental program. To understand how the tissue architecture regulates cell fate specification, we study the mechanical properties of cells that shape the embryo. Using a non-invasive micropipette aspiration method, we map the surface tensions of cells in space and time within the developing mouse embryo (figure 2). An integrative understanding based on the complete maps of cell lineage, gene expression and cell mechanics allows prediction and testing of our models.

Looking at the molecular, cellular and systems levels, the Hiiragi group studies how, early in mammal development, the embryo is shaped from a spherical mass of cells.

Future plans

We adopt a wide variety of experimental strategies including embryology, molecular genetics, live-imaging, biophysics and theoretical modelling in order to address fundamental questions in development and cell biology at a molecular, cellular and systems level. Our goals include:

- identification of the symmetry-breaking cue in mammalian embryos;
- molecular characterisation of the *de novo* formation of epithelial polarity;
- understanding how tissue mechanics, cell polarity and fate are coupled;
- identification of the trigger and mechanism for the *de novo* centriole formation.

Figure 1: Molecular heterogeneity during mouse blastocyst patterning. Cells expressing Nanog (green), Gata6 (red) or Serpinh1 (blue).
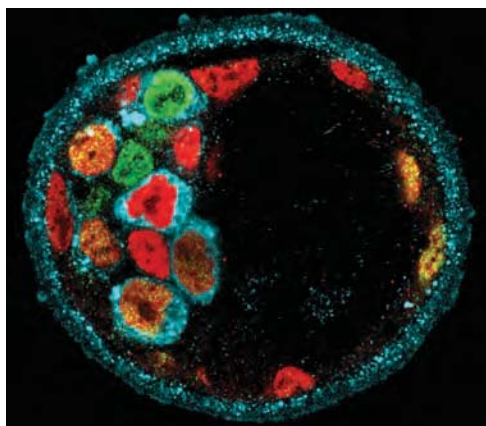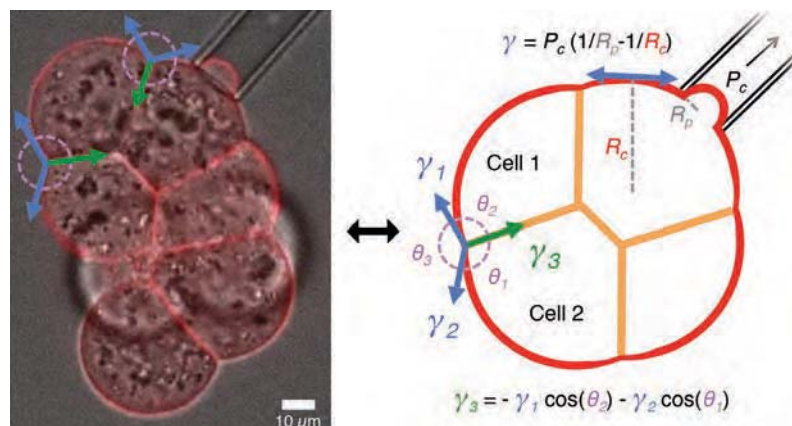
Figure 2: Mapping of surface tensions in a developing mouse embryo.

$$\gamma = P_c \, (1/R_p - 1/R_c)$$

$$\gamma_3 = -\gamma_1 \cos(\theta_2) - \gamma_2 \cos(\theta_1)$$

Microglia: the guardians of the developing brain

Francesca Peri

PhD 2002, University of Cologne.

Postdoctoral research at the Max Planck Institute for Developmental Biology, Tübingen.

Group leader at EMBL since 2008.

ERC Investigator.

SELECTED REFERENCES

Casano AM & Peri F. (2015) Microglia: multitasking specialists of the brain. *Developmental Cell* 32, 469-77

Sieger D & Peri F. (2013) Animal models for studying microglia: the first, the popular, and the new. *Glia*, 61, 3-9

Sieger D, *et al*. (2012) Long-range Ca2+ waves transmit brain-damage signals to microglia. *Dev. Cell*, 22, 1138-48

Peri F & Nüsslein-Volhard C. (2008) Live imaging of neuronal degradation by microglia reveals a role for v0-ATPase a1 in phagosomal fusion *in vivo*. *Cell*, 133, 916-27

The Peri group combines genetic approaches with quantitative imaging techniques to study interactions between neurons and the microglia that eliminate cellular debris in the brain.

During brain development, neurons are generated in great excess and only those that make functional connections survive, while the majority is eliminated via apoptosis. Such huge numbers of dying cells pose a problem to the embryo, as leaking cell contents damages the surrounding environment. Therefore, the clearance of dying cells must be fast and efficient and is performed by a resident lineage of 'professional' phagocytes, the microglia. These cells patrol the entire vertebrate brain and sense the presence of apoptotic and damaged neurons. The coupling between the death of neurons and their phagocytosis by microglia is striking; every time we observe dead neurons we find them already inside the microglia. This remarkable correlation suggests a fast acting communication between the two cell types, such that microglia are forewarned of the coming problem and may even promote the controlled death of neurons during brain development. Despite the importance of microglia in several neuronal pathologies, the mechanism underlying their degradation of neurons remains elusive.

The zebrafish *Danio rerio* is an ideal model system to study complex cell-cell interactions *in vivo*. As the embryo is optically transparent, the role of molecular regulators identified in large-scale forward and reverse genetic screens can be studied *in vivo*. Moreover, a key advantage of the system is that zebrafish microglia are extremely large, dynamic cells that form a non-overlapping network within the small transparent fish brain. Labelling microglia, neurons and organelles of the microglial phagocytotic pathway simultaneously in the living zebrafish embryos allows us to image, for the first time, the entire microglial population in order to study the interaction between neurons and microglia.

Future plans

How microglia collectively ensure that the entire brain is surveyed and how they react to damage with high precision is still entirely unknown. Recent findings suggest that diffusible molecules such as lipids and nucleotides could attract microglia in response to neuronal apoptosis and injury, respectively. While these molecules can trigger dynamic changes in microglia motility *in vitro*, elucidating how their activity is controlled within the intact brain, both in space and time, remains the most important challenge in understanding this fascinating biological problem. We aim to further exploit the massive imaging potential of the transparent zebrafish embryo for studying microglial biology *in vivo*. By combining forward and reverse genetic approaches with quantitative imaging technology, we will directly address the mechanisms underlying the attraction of microglia towards apoptotic, sick and injured neurons. By applying cutting-edge microscopy technology, such as the SPIM/DSLM (Selective Plane Illumination Microscopy), we will image all interactions between neurons and microglia and derive from this time-lapse analysis real quantitative data in a spatiotemporal manner.
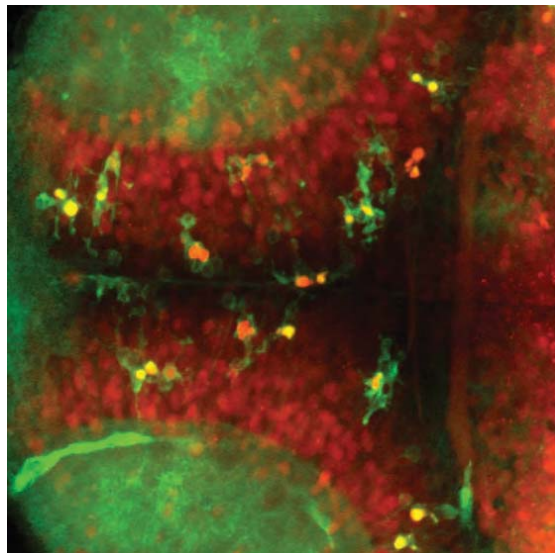
Microglia (green) and neurons (red) in the zebrafish embryonic brain.
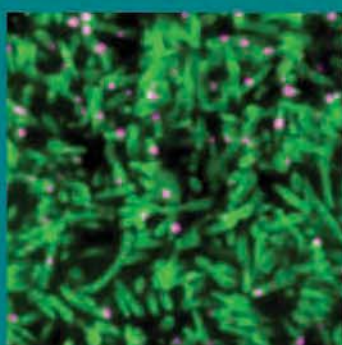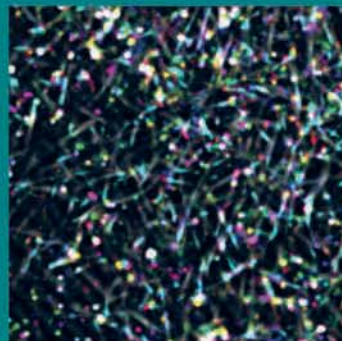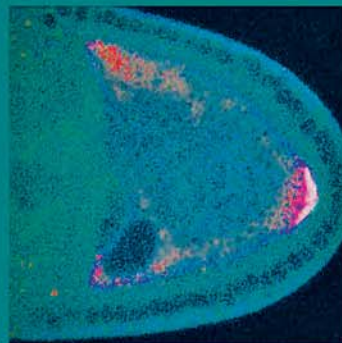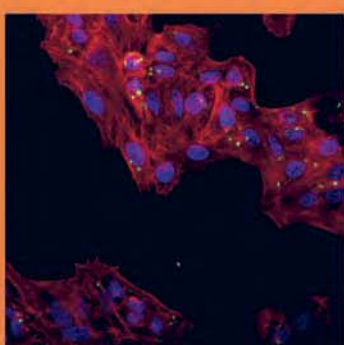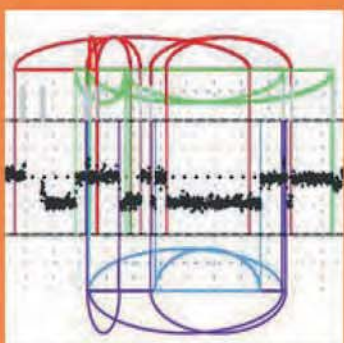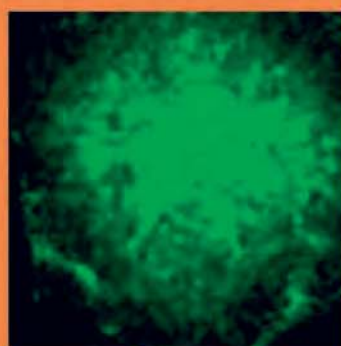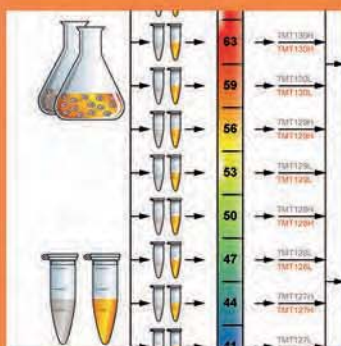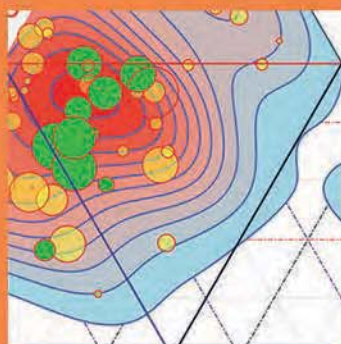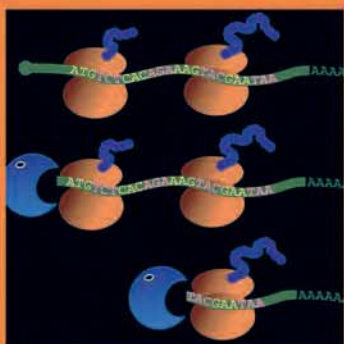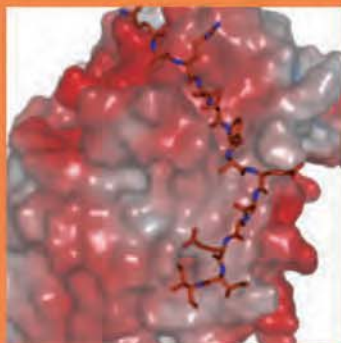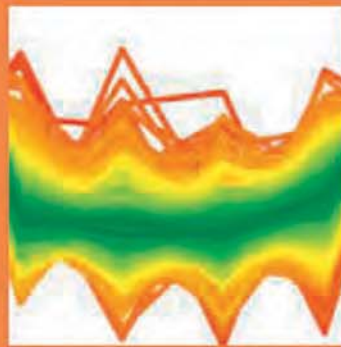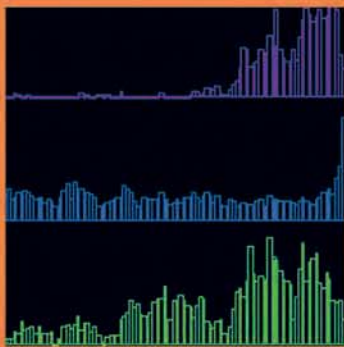
Genome Biology

The genome encodes the genetic blueprint that coordinates all cellular processes, which ultimately give rise to phenotype. The expression of genetic information is tightly regulated in both time and space at multiple steps, including transcriptional, post-transcriptional and post-translational. The Genome Biology Unit takes an integrated systems-level approach to unravel these complex processes at all scales, integrating cutting-edge experimental and computational approaches.

In eukaryotes, many steps of gene expression, such as transcription and RNA processing, take place in the structurally complex environment of the nucleus and often involve remodelling of chromatin into active and inactive states. Messenger RNAs, once exported from the nucleus, undergo additional regulatory steps. Their translation results in the production of proteins, whose functions define the characteristics of different cell types, or cellular phenotypes. Not all RNAs are translated, however. In recent years, multiple types of non-coding RNAs have been discovered that display diverse functionality. Genetic variation in non-coding and protein-coding genes alike, as well as the regulatory elements that govern their expression, can adversely affect the function of these genes, leading to diseases such as cancer. Groups within the Unit are investigating various aspects of genome biology in order to understand these processes leading from genotype to phenotype.

A notable strength of the Unit is its ability to address questions at different scales, ranging from detailed mechanistic studies (using biochemistry, genetics, microfluidics and chemistry) to genome-wide studies (using functional genomic, proteomic and computational approaches), often by developing new enabling technologies. For example, the development and integration of chemistry and microfluidic devices with the recent advances in next-generation sequencing will facilitate major advances in these areas in the coming years. Global, dynamic and quantitative measurements of biological molecules at all levels (DNA, RNA, proteins, cells, organisms, etc) as well as the integration of hypothesis and discovery-driven research characterise the Unit. The synergy between computational and wet-lab groups provides a very interactive and collaborative environment to yield unprecedented insights into how genetic information is 'read' and mediates phenotype through molecular networks.

Eileen Furlong
Head of the Genome Biology Unit

Genome regulation and topology during embryonic development

Eileen Furlong

PhD 1996, University College Dublin.

Postdoctoral research at Stanford University.

Group leader at EMBL since September 2002.
Senior Scientist since 2009.
Joint Head of Genome Biology Unit 2009-13.

Head of Genome Biology Unit since 2013.

ERC Advanced Investigator since 2013.

Elected EMBO Member since 2013.

SELECTED REFERENCES

Ghavi-Helm Y, *et al.* (2014) Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* 512, 96-100

Rembold, M.*, Ciglar, L.*, *et al.* (2014) A conserved role for Snail as a potentiator of active transcription. *Genes Dev.*, 28, 167-81

Erceg J, *et al.* (2014) Subtle changes in motif positioning cause tissue-specific effects on the robustness of enhancer activity. *PLoS Genet.* 10, e1004060

Bonn, S.*, Zinzen, R.P.*, Girardot, C.*, *et al.* (2012) Tissue specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature Genetics,* 44, 148-56

The Furlong group dissects fundamental principles of transcriptional regulation, and how that drives cell fate decisions during development, focusing on organisational and functional properties of the genome.

Precise regulation of gene expression is essential for almost all biological processes, and a key driving force in development, evolution and disease. Expression states are initiated through diverse cues modulating transcription factor activity, which converge on *cis*-regulatory elements such as enhancers. Enhancers thereby act as integration platforms to control specific patterns of expression, telling genes when and where to be expressed. Given their central role, mutations in enhancers often lead to devastating developmental defects and are becoming increasingly linked to human disease.

Our research focuses on dissecting general mechanisms of transcriptional regulation, including how the *cis*-regulatory genome is organised within the nucleus (figure 1), and how chromatin state and transcription factor occupancy influence this process (figure 2). We investigate how natural sequence variation (both within and between species) affects transcription, leading to specific phenotypes. Our work combines genomic, genetic and computational approaches to understand these processes, including the development of new genomic tools to facilitate this analysis within the context of a multi-cellular embryo; *Drosophila* mesoderm specification.

Figure 1: Enhancers interact with genes over very long distances within the *Drosophila* genome, as shown by 4C-Seq (top) and DNA FISH (bottom) during embryogenesis (Ghavi-Helm Y, *et al., Nature* 2014).
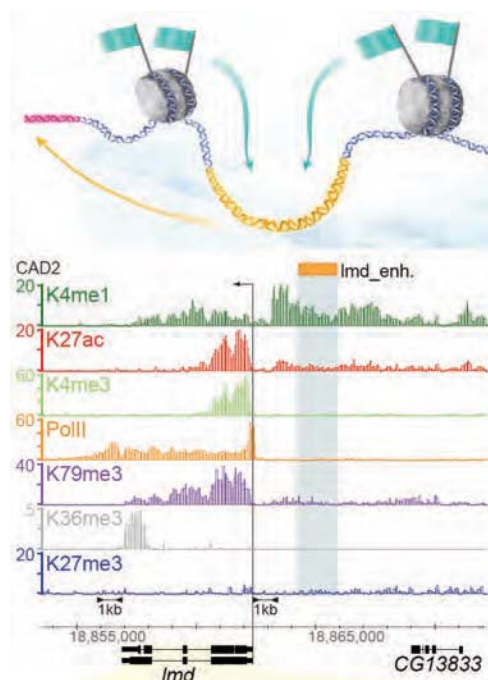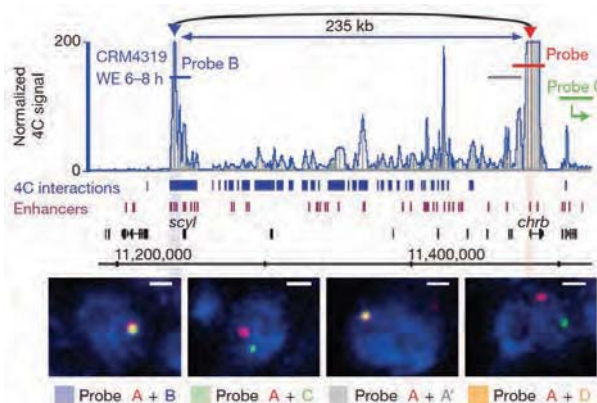

Figure 2: Chromatin state and Pol II occupancy on enhancers (yellow) is highly predictive of enhancers' activity, with Pol II being predictive for the precise timing during development (Bonn*, Zinzen*, Girardot*, *et al., Nature Genetics* 2012).

Future plans

Chromatin topology – the 3D Genome:
How our huge genome is packaged within the confined space of the nucleus to facilitate transcription remains a key challenge in genome regulation. For enhancers to function, they must come in proximity to their target genes. We recently discovered a surprising stability of these 'looping' interactions during two stages of embryonic development (*Nature* 2014). Going forward, we will examine this over an extensive developmental time-course, integrating high-resolution imaging and genetics to dissect the relationship between proximity and transcriptional regulation.

Chromatin remodelling during cell fate decisions:
To uncover general properties of enhancer function we developed a method to investigate cell type-specific changes in chromatin state in the context of a multicellular embryo's development (Figure 2; *Nature Genetics* 2012). Going forward, we will combine this technology with single cell methods and CRISPR/Cas9 technology to understand regulatory properties associated with developmental transitions.

Variation and plasticity in regulatory networks:
Variation in the non-coding genome is widely associated with Quantitative Trait Loci and GWAS loci. Many of these disease-associated variants are in *cis*-regulatory elements, yet it is very difficult to pinpoint the causal SNP within the human genome and therefore dissect the underlying mechanism. We are developing methods to bridge this gap, taking advantage of the extremely high sequence variation among wild *Drosophila* isolates, to dissect functional regulatory variants involved in transcription and genome organisation during embryonic development.

Multi-omics and statistical computing

Wolfgang Huber

PhD 1998, Statistical Physics, University of Freiburg.

Postdoctoral research at IBM Research, San Jose, California and at DKFZ, Heidelberg.

Group leader at EMBL-EBI 2004-2009, EMBL Heidelberg since 2009. Senior Scientist since 2011.

SELECTED REFERENCES

Ignatiadis N, *et al.* (2015) Data-driven hypothesis weighting increases detection power in big data analytics. *bioRxiv*, dx.doi.org/10.1101/034330

Huber W, *et al.* (2015) Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods* 12, 115-21

Love MI, *et al.* (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15, 550

Fischer* B, Sandmann* T, Horn* T, Billmann* M, *et al.* (2015) A map of directional genetic interactions in a metazoan cell. *eLife*, 4

A central challenge of biomedicine is to understand how the biological systems that underlie healthy life and disease react to variations in their make-up (genetic variation, for example) or their environment (drugs, for example). Our group brings together researchers from quantitative disciplines – mathematics, statistics, physics and computer science – and from different fields of biology and medicine.

We employ statistics and machine learning to discover patterns in large datasets, understand mechanisms, and act upon predictive and causal relationships to, ultimately, address questions in personal genomics and molecular medicine. We use large-scale data acquisition and quantitative modelling of phenotypes and molecular profiles, systematic perturbations (such as drugs or high-throughput genetics) and computational analysis of non-linear, epistatic interaction networks.

Genomics and other molecular profiling technologies have resulted in increasingly detailed biology-based understanding of human disease. The next challenge is using this knowledge to engineer treatments and cures. We integrate observational data – such as from large-scale sequencing and molecular profiling – with interventional data – systematic genetic or chemical screens – to reconstruct a fuller picture of the underlying causal relationships and actionable intervention points. A fascinating example is our work on genotype-specific vulnerability and resistance of tumours to targeted drugs in our precision oncology project.

As we engage with new data types, our aim is to develop high-quality computational and statistical methods of wide applicability. We consider the release and maintenance of scientific software an integral part of scientific publishing, and we contribute to the Bioconductor Project, an open source software collaboration to provide tools for the analysis and study of high-throughput genomic data. An example is our DESeq2 package for analysing count data from high-throughput sequencing.

The Huber group develops large-scale statistical models that integrate genomic, molecular and phenotypic data to understand the variations between individuals in health and disease.
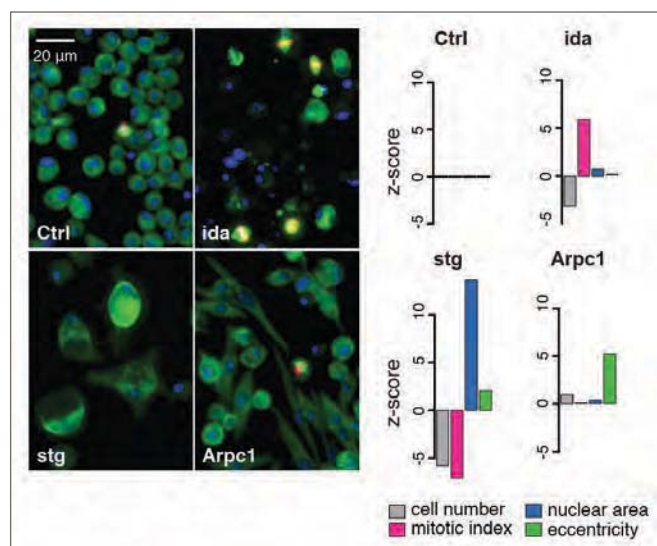

Figure 1: Automated multivariate phenotyping of cells by combinatorial RNAi and automated image analysis.

Future plans

We aim to develop the computational techniques needed to understand novel biological data:

- Clinical multi-omics: we work with clinical researchers to develop predictive assays and algorithms.

- Many powerful mathematical ideas exist but are difficult to access. We translate them into practical methods and software that make a real difference to biomedical researchers, an approach we term 'translational statistics'.

- Quantitative proteomics and *in vivo* drug-target mapping.

- Single-cell and single-molecule -omics.

- High-throughput multidimensional phenotyping: mapping gene-gene and gene-drug interactions through computational image analysis of cell and tissue microscopy, machine learning and mathematical modelling.
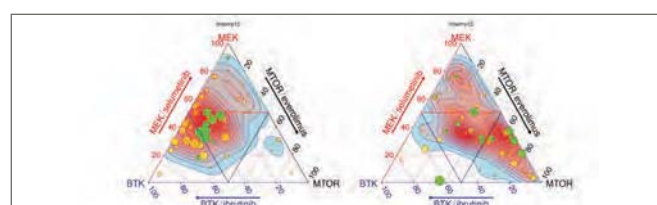

Figure 2: Ternary plots of relative sensitivities to targeted kinase inhibitors for a cohort of primary tumour samples of chronic lymphocytic leukaemia (CLL).

Phosphatase chemistry and biology

Maja Köhn

PhD 2005 MPI for Molecular Physiology, Dortmund, Germany.

Postdoctoral work at Harvard University, Cambridge, USA.

Group Leader at EMBL since 2007.

ERC Investigator.

SELECTED REFERENCES

Hoeger B, *et al.* (2014) Biochemical evaluation of virtual screening methods reveals a cell-active inhibitor of the cancer-promoting phosphatases of regenerating liver. *Eur J Med Chem* 17, 89–100

Pavic K, *et al.* (2014) Unnatural amino acid mutagenesis reveals dimerization as a negative regulatory mechanism of VHR's phosphatase activity. *ACS Chem. Biol.* 9, 1451-9

Li X, *et al.* (2013) Elucidating human phosphatase-substrate networks. *Sci Signal* 6, rs10

Chatterjee J, *et al.* (2012) Development of a peptide that selectively activates protein phosphatase-1 in living cells. *Angew. Chem. Int. Ed. Engl.* 51, 10054-9

The Köhn group combines molecular biology, biochemistry and synthetic chemistry to develop new approaches to study phosphatases, which can play a major role in cancer.

Within intracellular signalling networks, phosphatases are counter players of kinases and play crucial roles in health and disease. Despite their importance, knowledge about their function, regulation and substrate interaction is still limited, and their investigation is challenging also because of the lack of tools to selectively target them. We aim to fill that void using interdisciplinary approaches.

We study the molecular mechanisms of the cancer-promoting PRL (phosphatase of regenerating liver) phosphatases using biochemical and molecular cell biology approaches, and we develop specific inhibitors for them. We observed phosphoinositide-phosphatase activity for PRL-3 (McParland *et al.*, *Biochemistry* 2011), prompting us to use phosphoinositides for substrate-based inhibitor design. Therefore, we developed a solid-phase synthesis strategy (Bru *et al.*, *Chem. Sci.* 2012; figure 1) enabling the parallel synthesis of phosphoinositide analogues. Moreover, through a combined *in silico* and biochemical approach, we discovered a cell-active inhibitor for the PRLs (Hoeger *et al.*, *Eur. J. Med. Chem.* 2014).

Protein phosphatase-1 (PP1) is the ubiquitous phosphatase responsible for a majority of all dephosphorylation reactions on Ser/Thr residues inside cells. We developed the first and only selective chemical PP1-modulator, which activates it inside cells (Chatterjee *et al.*, *Angew. Chem. Int. Ed.* 2012; figure 1). We are extending the PP1 toolkit, and will apply it to study PP1.

We created and maintain the human DEPhOsphorylation Database: DEPOD (figure 2), and have used it to re-classify the human phosphatome and to analyse phosphatase substrate specificities and their relation to kinases (Li *et al.*, *Sci. Signal.* 2013; Duan *et al.*, *Nucleic Acids Res.* 2015).

In the area of chemical tool development, we have established a strategy to design protein tyrosine phosphatase (PTP) inhibitors that can also function as detection tools (Meyer *et al.*, *ACS Chem. Biol.* 2014). Moreover, using unnatural amino acid mutagenesis we established site-directed covalent crosslinking as a principle to detect interacting proteins of PTPs, and to study the effect of the interaction on the biological activity and regulation of the PTP (Pavic *et al.*, *ACS Chem. Biol.* 2014).

The lab combines the expertise of molecular biologists and organic chemists opening up new ways to approach challenges in phosphatase research, and broadening the views and skills of every lab member.

Figure 1:
(A) Solid phase synthesis of phosphoinositides for the preparation of libraries for SAR studies with lipid phosphatases (Bru *et al.*, *Chem. Sci.* 2012).
(B) Selective activators of PP1 in cells enable us to gain new insights into PP1 biology (Chatterjee *et al.*, *Angew. Chem. Int. Ed.* 2012; Reither *et al.*, *Chem. Biol.* 2013).
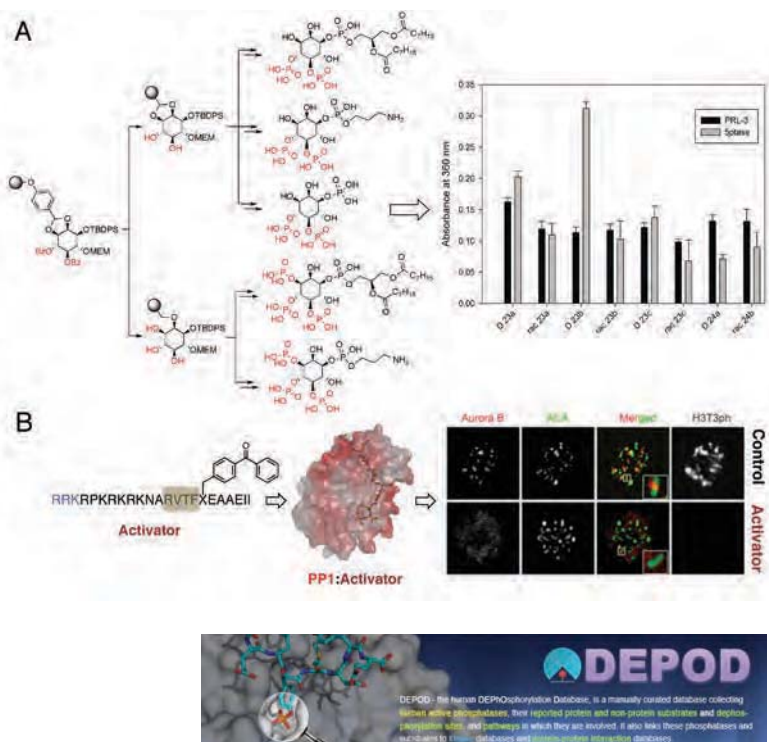
Future plans

- Understand the role of PRLs and inhibit them in oncogenesis.

- Further the development of chemical methods to use peptides and inositides as phosphatase modulators inside cells.

- Design modulators for the highly complex serine/threonine phosphatases.

- Continue to develop and maintain DEPOD.

Figure 2: DEPOD, the human dephosphorylation database - **www.depod.org** (Li *et al.*, *Sci. Signal.* 2013; Duan *et al.*, *Nucleic Acids Res.* 2015).

From genomic variation to molecular mechanism

Jan Korbel

PhD 2005, EMBL Heidelberg/Humboldt University, Berlin.

Postdoctoral research at Yale University, New Haven, Connecticut, USA.

Group leader at EMBL since October 2008.

Joint appointment with EMBL-EBI.

Group leader in the Molecular Medicine Partnership Unit.

ERC Investigator.

SELECTED REFERENCES

Sudmant PH, *et al.* (2015) An integrated map of structural variation in 2,504 human genomes. *Nature* 526, 75-81.

Fischer U, *et al.* (2015) Genomics and drug profiling of fatal TCF3-HLF-positive acute lymphoblastic leukemia identifies recurrent mutation patterns and therapeutic options. *Nat. Genet.* 47, 1020-9.

Mardin BR, *et al.* (2015) A cell-based model system links chromothripsis with hyperploidy. *Mol Syst Biol* 11, 828

Northcott PA, *et al.* (2014) Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* 511, 428-34.

Genetic variation is a fundamental reason why humans differ from one another and why some individuals are more susceptible to diseases than others. Our group is investigating mechanisms and the phenotypic consequences of genetic variation, with a focus on genomic structural variations (SVs). Genome-wide techniques we employ include genome sequencing, single cell and epigenomic assays (including ChIP-Seq, ATAC-Seq, and Hi-C). Our laboratory uses a 'hybrid' approach (integrating laboratory and computational methodology) to combine data generation and analysis with hypothesis generation and experimental testing. Computational methodologies range from analytical integration approaches (with the purpose to uncover biological novelty) to statistically-oriented methods development.

Our group plays leadership roles in research consortia, including currently the largest ongoing global "big data" initiative – the Pan-Cancer Analysis of Whole Genomes (PCAWG) project. Integrating data from somatic and germline whole genomes, DNA methylomes, transcriptomes, and clinical data from more than 2800 cancer patients, we aim to unravel commonalities and discrepancies between the emergence of cancer types and subtypes at the molecular level, to facilitate the molecular classification of malignancies with impact on diagnostics and treatment, and to uncover causalities linking genotype, environment, and phenotype.

Within the 1000 Genomes Project and its follow-up studies, we are conducting further human deep population surveys using short-read and single-molecule long DNA read sequencing. Recent work by us uncovered the most comprehensive catalogue of genetic variation to date, providing insights into the origin of diverse structural variation classes (figure 1).

We are also making progress in understanding the occurrence of complex genetic variants with experimental model systems, especially with regard to a process known as chromothripsis – a cellular catastrophe causing massive simultaneous DNA rearrangements. We recently developed an integrative method termed 'Complex Alterations after Selection and Transformation' (CAST) allowing mechanistic dissection of chromothripsis (figure 2). CAST revealed an increased rate of chromothripsis in hyperploid cells and following telomere shortening (through siRNA induced silencing of a shelterin complex component), findings that we verified by genome sequencing in primary brain tumours.

The Korbel group combines experimental and computational approaches to decipher determinants and consequences of germline and somatic DNA variation.

Future plans

- Completion of human genome variation maps by third-generation sequencing.

- Combining genomic and epigenetic studies to identify molecular determinants for genetic variation in cancer and normal cells.

- Deciphering the basis of genomic instability processes using cell-based models.

- Computational methods development, for example single-cell and single-molecule sequencing, as well as integration of "big data" in omics and health.
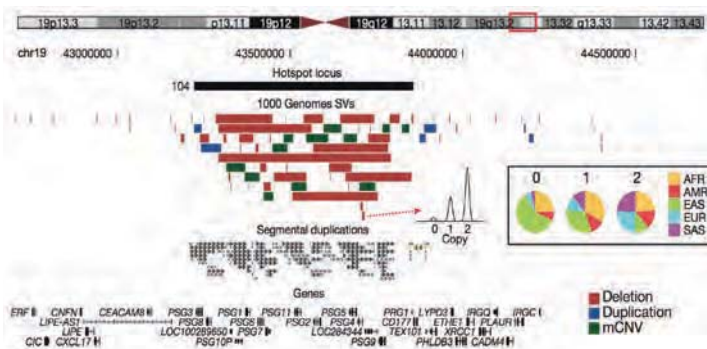

Figure 1: A germline genomic hotspot of structural variation with the PSG locus, a pregnancy-associated gene cluster.
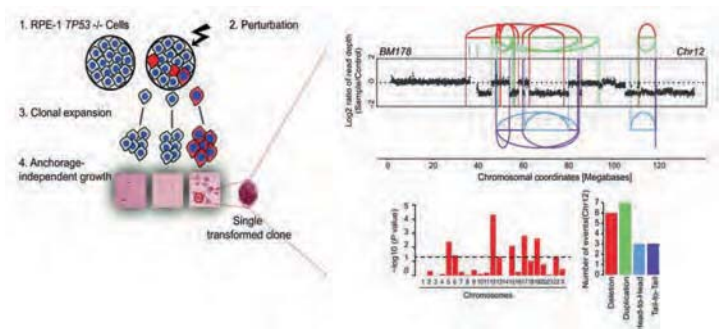
Figure 2: The CAST approach providing insights into determinants of chromothripsis.

Microfluidic approaches in drug discovery and personalised medicine

Christoph A. Merten

PhD 2004, University of Frankfurt.

Postdoctoral research at the MRC Laboratory of Molecular Biology, Cambridge.

Junior group leader at the Institut de Science et d'Ingénierie Supramoléculaire, Strasbourg.

Group leader at EMBL since 2010.

SELECTED REFERENCES

Hu H, *et al.* (2015) Efficient cell pairing in droplets using dual-color sorting. *Lab Chip* 15, 3989-93

Hara H & Merten CA (2015) Dynein-based accumulation of membranes regulates nuclear expansion in *Xenopus laevis* egg extracts. *Dev Cell* 33, 562-75

El Debs B, *et al.* (2012) Functional single cell hybridoma screening using droplet-based microfluidics. *PNAS* 109, 11570-5

Clausell-Tormos J, *et al.* (2008) Droplet-based microfluidic platforms for the encapsulation and screening of mammalian cells and multicellular organisms. *Chem. Biol.* 15, 427-37

The Merten group develops microfluidic technology to address complex questions in biomedical sciences and developmental biology.

Working on the micro-scale offers some unique advantages:

- Drastically increased throughput (processing up to a million samples an hour).
- Possibility of performing assays on the single-cell level.
- Low sample consumption, enabling the use of patient material.

During the past couple of years we have established powerful microfluidic platforms for cell-based and biochemical assays. We continuously develop novel microfluidic chips, detection systems and software (mainly LabVIEW) for the discovery of new drugs and antibodies, partially in collaboration with large pharma industry. Furthermore, we use microfluidic technology to predict optimal (personalised) drug cocktails for cancer therapy. The group is very interdisciplinary and consists of people with various backgrounds, including biologists, engineers, and programmers. However, prior knowledge in microfluidics is not mandatory for joining!

For most applications we use two-phase microfluidics, in which aqueous droplets within an immiscible oil phase serve as miniaturised reaction vessels. As they can be generated at kilohertz frequencies, they are of particular interest for high-throughput screens. Furthermore, the small assay volumes (pico- to nanoliters) facilitate the obtainment of high concentrations of nucleic acids (mRNA, DNA) or proteins (such as secreted antibodies) from individually encapsulated cells, paving the way for single-cell assays. Sometimes we also use continuous-phase microfluidics, for example to generate laminar flow patterns, to apply spatial constrictions, or to provide cells sequentially with different stimuli. This has turned out to be particularly useful for various developmental biology applications.

Future plans

Having a comprehensive microfluidic toolbox at hand (and expanding it continuously), we are now focusing on applications in three different research fields:

Therapeutic antibodies: Droplet based microfluidics enables to screen a large fraction of the murine and human immune repertoire in a single experiment. We are actively exploiting this for novel therapeutic approaches.

Personalised medicine: We use microfluidic devices to test drug combinations directly on human tumor samples and to predict optimal therapies.

Genomics: We are actively developing microfluidic modules for single-cell barcoding and sequencing. This will enable the study of the mechanisms of drug resistance in great detail.

Microfluidic approaches in drug discovery and personalised medicine

Epigenetic mechanisms of neurodevelopment and diseases
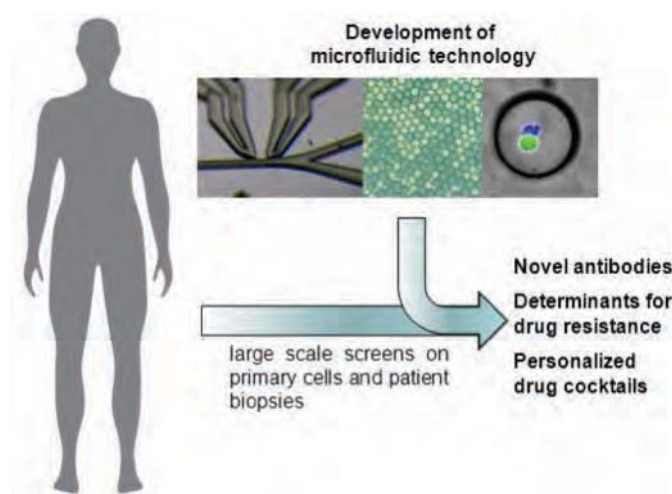
Kyung-Min Noh

PhD in Neuroscience 2008, Albert Einstein College of Medicine, New York.

Postdoctoral research in Epigenetics at The Rockefeller University, New York.

Group leader at EMBL since 2014.

SELECTED REFERENCES

Noh KM, *et al.* (2015) Engineering of a histone-recognition domain in Dnmt3a alters the epigenetic landscape and phenotypic features of mouse ESCs. *Mol Cell* 59, 89-103

Noh KM, *et al.* (2015) ATRX tolerates activity-dependent histone H3 methyl/phos switching to maintain repetitive element silencing in neurons. *Proc Natl Acad Sci USA* 112, 6820-7

Elsasser SJ, *et al.* (2015) Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature* 522, 240-4

Maze I, *et al.* (2015) Critical role of histone turnover in neuronal transcription and plasticity. *Neuron* 87, 77-94

The Noh group studies chromatin links vital for neuro-development and brain disorders.

Chromatin, the faithful association of DNA with histone proteins, exists as the physiological form of our genome and the substrate for processes that regulate cellular gene expression. Numerous diseases including neurodevelopmental disorders are associated with mutations in genes that encode for chromatin-binding proteins or chromatin-modifying enzymes, which together act as epigenetic regulators. A central question grounding our research is how the normal/mutated epigenetic regulators engage in brain development, function and disease.

Histone variants and covalent modifications of both DNA and histones act in concert to define the landscape of our epigenome. Previously, we explored the interconnections between histone and DNA modifications by focusing on a conserved chromatin-binding regulatory domain, the ATRX-DNMT3-DNMT3L (ADD) domain. We showed that the ADD domain is capable of sensing, and therefore integrating, the status of multiple histone modifications. This in turn dictates the in vivo localisation of the full-length ADD-containing protein and its ability to function in downstream chromatin remodelling events. In parallel, we uncovered the localisation and function of a histone H3 variant, H3.3. Guided by distinct chaperone systems, H3.3 marks the genomic regions of histone turnover. We mapped the genome-wide localisation of H3.3 in mouse embryonic stem cells (mESCs) and neuronal precursor cells, and further expanded to terminally differentiated neurons for studying its functional role in promoting neuronal plasticity.
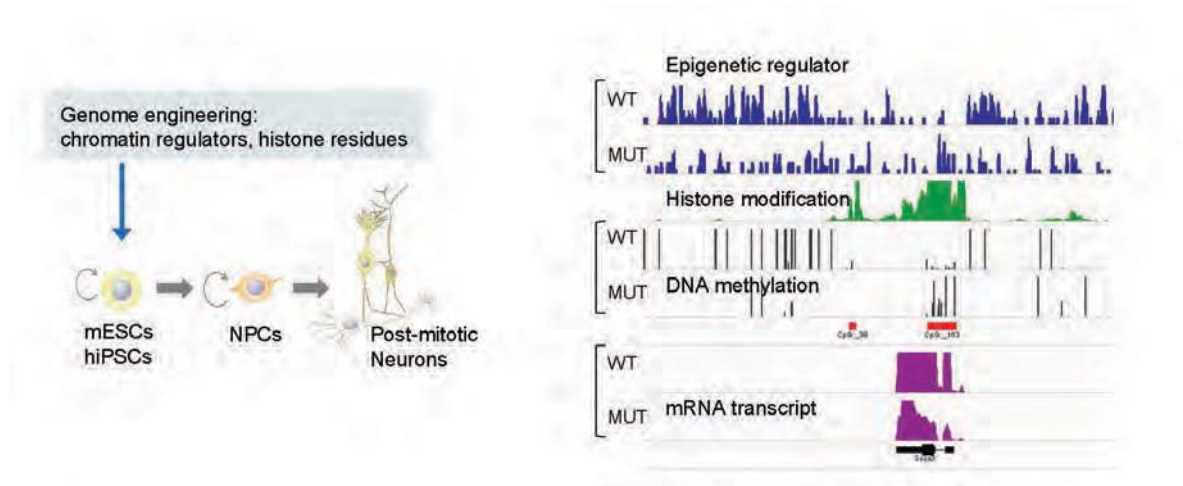
Future plans

We aim to study chromatin regulation, its interpretation during brain development, and its misinterpretation in relation with brain cognitive and developmental disorders. We will focus on the molecular mechanisms that link genetic mutations encoded in epigenetic regulators to the widespread chromatin alterations. To model developmental stages and facilitate the necessary genome editing/engineering, we will use differentiating neurons from mESCs and human induced pluripotent cells (hiPSCs). Ultimately, defining the 'epigenetic landscape' – both in normal and abnormal brain cells – will not only help to provide novel potential targets for therapeutic intervention for cognitive and developmental disorders of the brain, but also advance our basic understanding of the fundamental principles in epigenetic regulation.

Our research projects are to:

- Identify the location/function of mutated epigenetic regulators specific to cognitive deficits, and explore alterations of the epigenetic landscape in developing neurons.

- Investigate the function roles of covalent modifications of DNA and histones in neurodevelopmental transcription programs.

Top: biological system of interest. Bottom: concept of epigenetic mechanisms.

Example of genome-wide approaches with wild-type and mutant epigenetic regulator.

Stability proteomics for assessing the state of the proteome

Mikhail Savitski

PhD 2007, Uppsala University.

Group leader Cellzome, Heidelberg.

Team leader at EMBL from 2016.

SELECTED REFERENCES

Savitski MM, *et al.* (2015) Proteomics. Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* 346,1255784

Reinhard FB, *et al.* (2015) Thermal proteome profiling monitors ligand interactions with cellular membrane proteins. *Nat. Methods* 12, 1129-31

Franken H, *et al.* (2015) Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry. *Nat. Protoc.* 10, 1567-93

Werner T, *et al.* (2014) Ion coalescence of neutron encoded TMT 10-plex reporter ions. *Anal. Chem.* 86, 3594-601

The Savitski team uses and develops stability proteomics for understanding the phenomenon of aggregation and disaggregation, cell phenotyping, and detection of protein interactions with drugs, metabolites, DNA and RNA.

The rapid evolution of high-end mass spectrometers in recent years has enabled an in depth quantitative assessment of the proteome. This development combined with recent advancements in multiplexing technologies (Werner *et al., Analytical Chemistry,* 2012 & 2014), enables new lines of scientific research aimed at proteome wide understanding of the state and function of proteins.

It has been known since 1959 that enzymes become more thermostable when bound to their ligand. In 2013 it was demonstrated for the first time for a handful of proteins that this could enable detection of drug protein interactions in living cells, by heating the cells to different temperatures spinning down the unfolded and subsequently aggregated proteins and measuring the amount of the soluble protein at different temperatures thus inferring its melting curve (Martinez Molina *et al., Science,* 2013). More recently my former colleagues and I extended this methodology to monitoring the thermal stability of thousands of proteins in an unbiased way using multiplexed quantitative mass spectrometry, (figure 1, Savitski *et al., Science,* 2014). This new proteomics method, which we termed Thermal Proteome Profiling (TPP), enables identification of drug targets in situ on a proteome wide scale. We also realised that protein components of pathways affected by the drugs underwent changes in thermal stability. Moreover, comparison of cell lines with different constitutively activated pathways revealed differences in thermal stability of pathway components confirming that TPP can be used as a complementary approach to, for example, phosphoproteomics to study signalling in cells. It was also observed that disruption of protein-protein interactions had an effect on thermal stability. The TPP technology thus holds a big promise for becoming a powerful tool for fundamental biology studies. At the same time, the technology still needs to be significantly further developed with regard to experimental strategies and data analysis. Below the key points for improvement and application of the technology are outlined, which will be the focus of the lab.

Future plans

Developing and applying protein stability proteomics for:

- understanding the phenomenon of disaggregation;
- key pathway components mapping;
- protein-metabolite interaction;
- identification of novel drug targets in living cells;
- detecting protein/DNA and protein/RNA interactions;
- effect of post-translational modifications on protein stability and function;
- further improving mass spectrometry workflows and analysis for stability proteomics;
- creating the human meltome atlas.
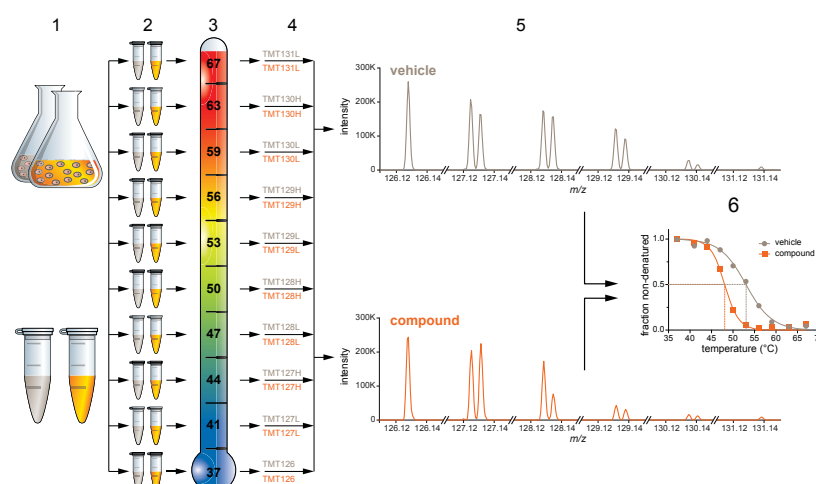
Workflow for performing thermal proteome profiling experiments.

Systems Genetics and Precision Health

Lars Steinmetz

PhD 1997-2001, Stanford University.

Postdoctoral research at Stanford Genome Technology Center.

Visiting group leader at Stanford Genome Technology Center since 2003. Co-Director since 2013.

Group leader at EMBL since 2003. Senior Scientist since 2009. Joint Head of Genome Biology Unit 2009-2012. Associate Head of Unit since 2013.

Professor of Genetics, Stanford University since 2013. ERC Advanced Investigator since 2012.

SELECTED REFERENCES

Brennecke P, *et al.* (2015) Single-cell transcriptome analysis reveals coordinated ectopic gene expression patterns in medullary thymic epithelial cells. *Nat. Immunol.* 16, 933-41

Pelechano V, *et al.* (2015) Widespread co-translational RNA decay reveals ribosome dynamics. *Cell* 161, 1400-12

Landry J, *et al.* (2013) The genomic and transcriptomic landscape of a HeLa cell line. *G3: Genes, Genomes, Genetics* 3, 1213-24

Pelechano V, *et al.* (2013) Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* 497, 127-31

The Steinmetz group develops and applies interdisciplinary, genome-wide technologies to study genome regulation, the genetic basis of complex phenotypes and the genetic and molecular systems underpinning disease.

One of the most daunting obstacles in biomedicine is the complex nature of most phenotypes (including cancer, diabetes, heart disease and some rare diseases) due to interactions between multiple genetic variants and environmental influences. A central challenge is to understand how genetic and environmental perturbations affect health, wellness and disease. Our research is directed at understanding such complex traits. To do so, we develop novel genomic approaches to study the molecular processes that link genotype to phenotype, identify the causal underlying factors, and quantify their contributions. We investigate inter-individual variation at the level of the genome, transcriptome, and proteome, which we integrate with higher-level phenotypes. Our projects are mainly in the following areas:

Quantitative genetics: We have piloted new technologies to dissect the genetic and environmental interactions that underlie complex, multifactorial phenotypes. We are interested in studying the consequences of genetic variation, learning to predict phenotype from genotype, and integrating multiple layers of molecular data to define intervention points that can be targeted to modulate phenotypes of interest (figure 1).

Functions and mechanisms of gene regulation: We have developed several technologies to characterise and quantify transcriptome architecture as well as its functional impact. In particular, we are interested in the function and regulation of non-coding RNAs, antisense transcription, transcriptional heterogeneity, and the molecular phenotypes that arise from pervasive transcription. Recently, we discovered that translation and degradation occur in parallel on the same mRNA allowing ribosome movement to be captured (figure 2).

Disease models: Using multiple model systems, primarily yeast and human cells, we have characterised the genetic and cellular processes affected in certain diseases and assessed potential therapeutic strategies. We apply personalised functional genomics to study diseases in patient-derived cells using systematic and targeted approaches to unravel mechanisms and discover novel treatments. We also develop point-of-care biosensors that monitor an individual's health and facilitate early disease diagnosis and intervention, even before symptoms set in.

Future plans

Using novel algorithms, we aim to identify causal intervention points from multi-omic datasets to modulate phenotypes of interest, including those associated with diseases. We are expanding our studies of transcriptional regulation through targeted investigations of the functional consequences of complex transcriptome architecture and its contributions to single-cell heterogeneity. Ultimately, by integrating genome technologies, biomedical engineering and computational modelling, we aim to uncover the molecular features of health and enable personalised and preventative medicine.

Our lab operates in an integrated manner across sites in Heidelberg, Germany, and at Stanford University in the US.
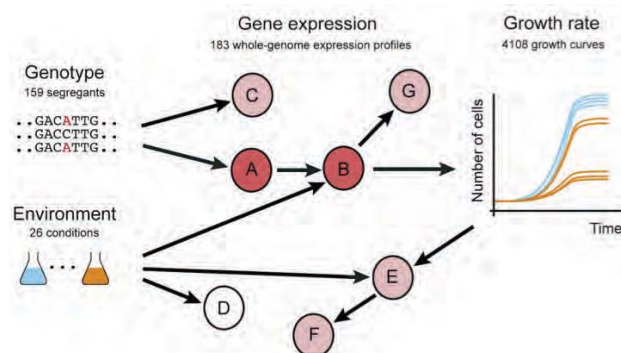
Figure 1: Gene-environment interactions reveal causal pathways (A-B) that mediate genetic effects on phenotype (Gagneur *et al.*, *PLOS Genetics* 2013).
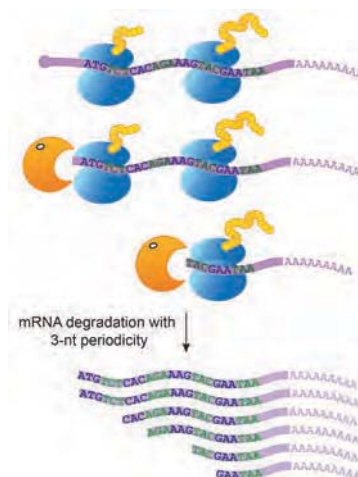
Figure 2: Dying mRNA tells a story of its life: Co-translational degradation revealed by 5P-Seq, a novel method that profiles ribosome dynamics (Pelechano *et al.*, *Cell* 2015).

Dissecting bacterial lifestyle and interspecies interactions with systems approaches

Nassos Typas

PhD 2006, Institute of Microbiology and Plant Physiology, Freie Universität Berlin, Germany.

Postgraduate research, University of California, San Francisco.

Group leader at EMBL since 2011.

Joint appointment with the Structural and Computational Biology Unit.

Humboldt Sofja Kovalevskaja Award Winner 2012.

SELECTED REFERENCES

Gray AN, *et al.* (2015) Coordination of peptidoglycan synthesis and outer membrane constriction during Escherichia coli cell division. *Elife* 4

Choo SH, *et al.* (2014) Detecting envelope stress by monitoring β-barrel assembly. *Cell* 159, 1652-64

Ezraty B, *et al.* (2013) Fe-S cluster biosynthesis controls uptake of aminoglycosides in a ROS-less death pathway. *Science* 6140, 1583-7

Nichols RJ, *et al.* (2011) Phenotypic landscape of a bacterial cell. *Cell* 144, 143-156

The Typas group develops and utilises high-throughput methods to study the cellular networks of different bacterial species, and how these bacteria interact with the environment and with each other.

The recent explosion of genomic sequence information provides a first step towards better understanding diverse bacteria, but also makes it crucial to develop large-scale phenotyping approaches to characterise functions of novel genes and to map them within pathways. We are developing such high-throughput, multi-readout, automated approaches to quantitatively assess gene-gene, gene-drug and drug-drug interactions in many different bacteria and at many different levels (figure 1). We then use the data as starting points for generating new mechanistic insights into targeted cellular processes, and also for uncovering how function, regulation and cross-talk between cellular processes changes across evolution and how this impacts the phenotype.

Our biological focus is on the bacterial envelope – its mode of assembly and growth, and its ability to sense the environment. The bacterial envelope is vital for pathogenesis, cell morphogenesis and cell developmental programs. Although many envelope structural components have been characterised, we often have limited information on how their biosynthesis and transport are interconnected, regulated, or linked to the overall status of the cell, how the cell senses perturbations in these process and how signals are transduced to achieve homeostasis. Working at the intersection between genomics and mechanistic molecular biology, we have discovered key missing players of major envelope components, uncovered niche-specific regulation of conserved envelope processes, identified linking proteins that allow coordination between processes, and mapped network rewiring under different stresses.
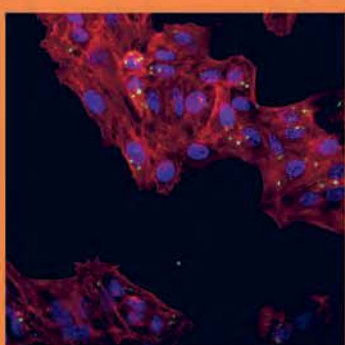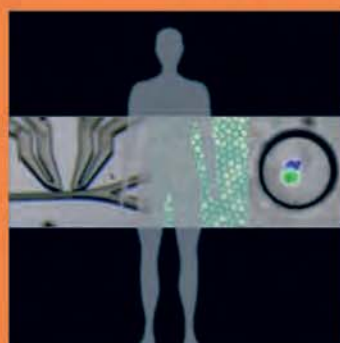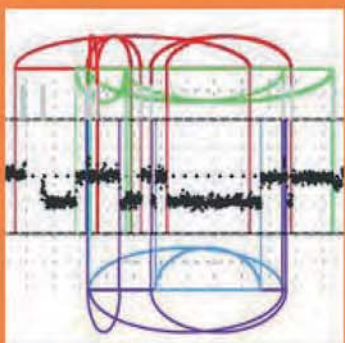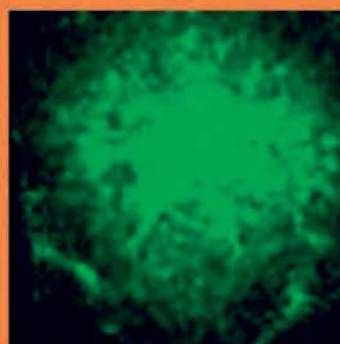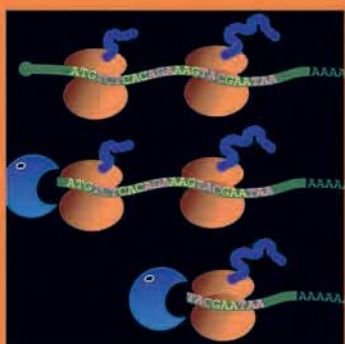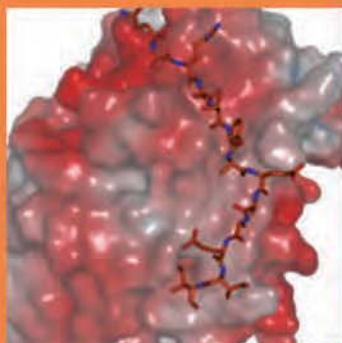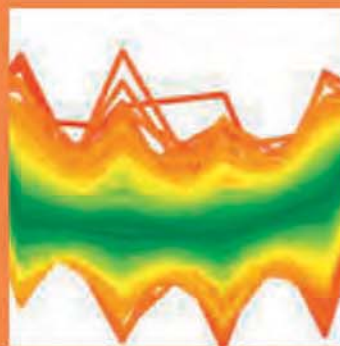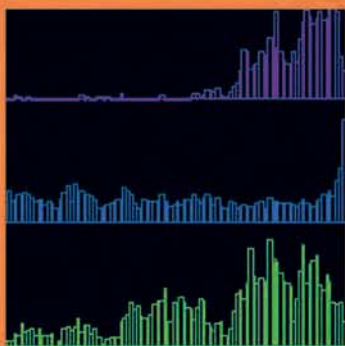
We are also developing large-scale automated platforms for elucidating the mode-of-action of new antibacterials, for large-scale profiling of combinatorial drug therapies and for dissecting the underlying mechanism(s), and for identifying adjuvants that re-sensitise multi-resistant bacterial pathogens or target chronic infections (persisters). Our ultimate goal is to identify rules underlying drug-drug interactions that will allow rational design, and to find solutions for difficult to kill pathogens.
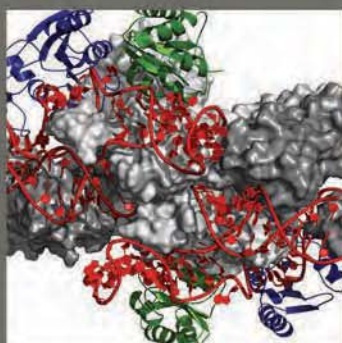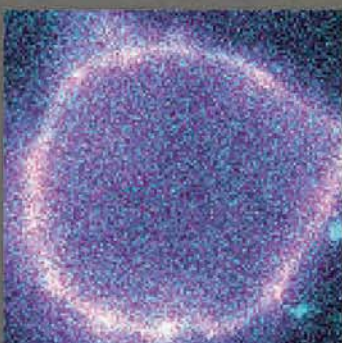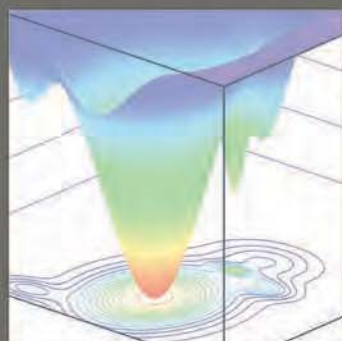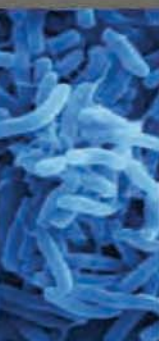
Future plans

We are now expanding our efforts in two directions. First, we are introducing our high-throughput screening approaches into abundant and prevalent species of the human gut microbiome. In collaboration with the Alexandrov, Bork, and Patil groups, and utilising a plethora of complementary technologies – such as imaging mass spectrometry, cutting-edge m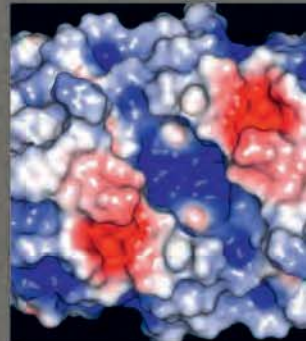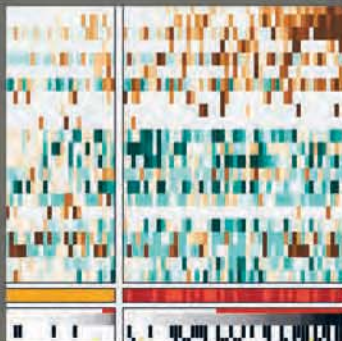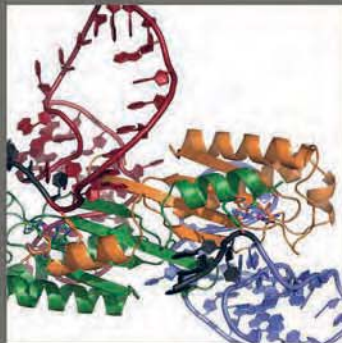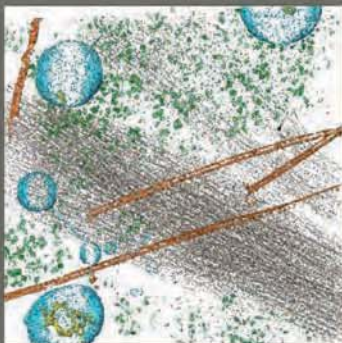icroscopy approaches, meta-omics, modelling –, we aim at understanding the dynamics of such communities, and how their composition is affected by drugs, natural and dietary compounds, physical parameters, and host molecules. Secondly, we are setting up a multi-pronged systematic approach aimed at gaining novel insights into the host-pathogen interface. Here, we combine high-throughput reverse genetics, high-content microscopy and different types of quantitative proteomics to dissect the Salmonella-host interaction.

High-throughput gene-gene, gene-drug and drug-drug interaction profiling provides novel mechanistic insights into the cellular network architecture and into drug mode-of-action.
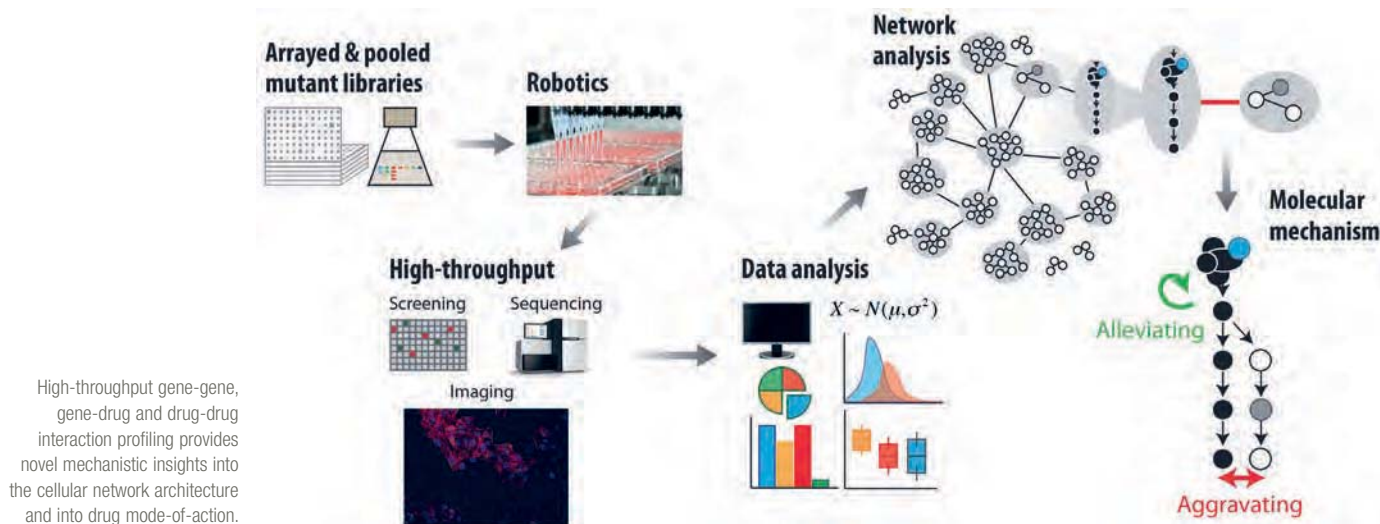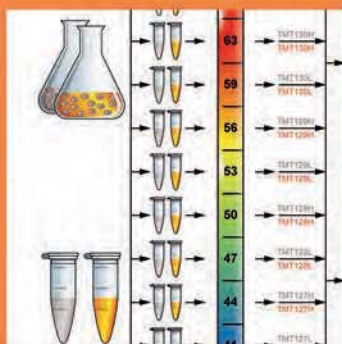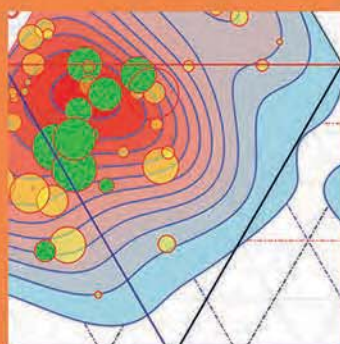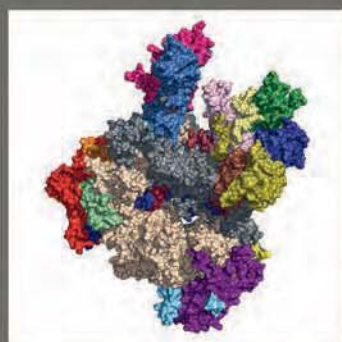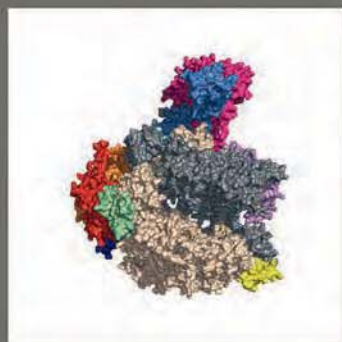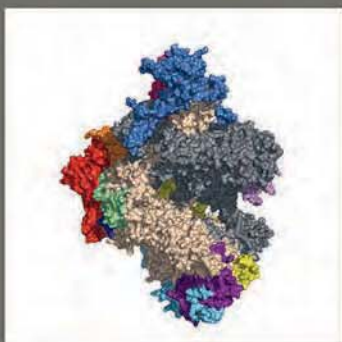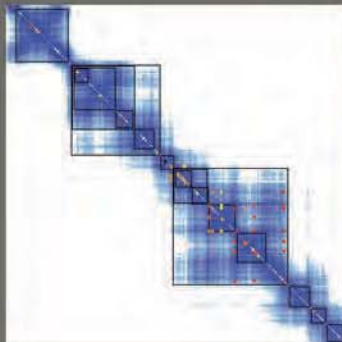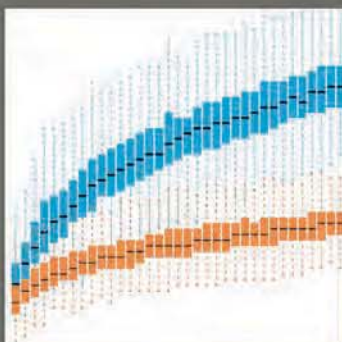
Structural and Computational Biology

The Unit pursues an ambitious research programme in integrated structural and computational systems biology bridging between various spatial and temporal scales.

The Unit currently consists of 14 groups covering a broad methodological spectrum that allows tackling problems at different ranges of spatial and temporal resolution from single amino acid side chain conformations to organismal communities. Molecular structures and dynamic information obtained by X-ray crystallography, NMR and high-resolution single-particle electron microscopy are integrated into the cellular context by electron tomography and correlated light microscopy. Dedicated large scale biochemistry, proteomics, metabolomics, chemical biology, biophysics, and cell biology approaches complement these structural biology activities enabling new research directions. These coordinated experimental activities (8 groups) are synergistic to a large computational biology programme (6 groups), which integrates the different information layers to be able to work towards comprehensive descriptions of biological functions at different spatial scales.

Within the Unit, there is a continuing interplay between groups with expertise in different methodologies. This reflects our belief that a combination of structural and functional studies is the most rewarding route to an understanding of the molecular basis of biological function, and that computational biology is essential to integrate the variety of tools and heterogeneous data into a comprehensive spatial and temporal description of biological processes. Along those lines, groups in the Unit pursue a few common large projects. For example, several groups contribute to the characterisation of the thermophilic fungus *Chaetomium thermophilum*, a model organism for structural biology, to understand its molecular networks and the molecular and cellular principles for eukaryotic thermophily.

Beyond its respective core technologies in structural and computational biology, each group reaches out into different areas, for example, there is considerable expertise in proteomics, metabolomics and next generation sequencing. In addition, several groups based in other Units have shared appointments with the Unit. Several service activities are setup around major technologies, for example the Unit serves as hub for bioinformatics coordination across all EMBL Heidelberg research groups, provides access to high-throughput crystallisation and NMR instrumentation, and engages in the organisation of EMBL-wide facilities and centres.

The SCB Unit is very well equipped for experimental and computational work. Experimental facilities include: crystallisation robot and automated crystal visualisation; rotating anode and X-ray detector; 800 MHz, 700 MHz, 600 MHz and 500 MHz NMR spectrometers; and several transmission electron microscopes, including a high-throughput Titan Krios microscope equipped with a direct detector. The Unit also has facilities for single-molecule light microscopy, metabolic imaging, isothermal calorimetry, circular dichroism, static and dynamic light scattering and analytical ultracentrifugation, as well as for large-scale growth of prokaryotic and eukaryotic cells. The computing environment offers access to around 4000 CPU cores and petabytes of storage, whereby large central and distributed resources are conveniently networked.

Peer Bork and Christoph Müller
Joint Heads of the Structural and Computational Biology Unit

Spatial metabolomics

Theodore Alexandrov

PhD 2007, St. Petersburg State University.

Postdoctoral Researcher, University of Bremen.

Group Leader, University of Bremen.

Assistant Adjunct Professor,
University of California San Diego.

Team leader at EMBL since 2014.

SELECTED REFERENCES

Bouslimani A, *et al.* (2015) Molecular cartography
of the human skin surface in 3D.
PNAS, 112, E2120-9

Watrous J, *et al.* (2013) Microbial metabolic exchange in 3D.
ISME J 7, 770–80

Alexandrov T, Kobarg JH (2011) Efficient spatial segmentation of large
imaging mass spectrometry datasets with spatially-aware clustering.
Bioinformatics 27, i230–8

The Alexandrov team develops novel computational biology tools that aim to reveal the spatial organisation of metabolic processes.

Metabolomics, the study of the chemical fingerprints left by cellular processes, is considered as a crucial research area, promising to advance our understanding of cell biology, physiology, and medicine. In the last years, metabolomics has progressed from cataloguing chemical structures to answering complex biomedical questions. The next frontier now lies in spatial metabolomics, where the challenge is to map the whole metabolome with cellular and sub-cellular spatial resolution and to develop a mechanistic understanding of metabolic processes in space, at the levels of cell populations, organs, and organisms.

Our team contributes to the emerging field of spatial metabolomics by developing computational biology tools that enable imaging and functional interpretation of metabolites in tissue sections, agar plates, and cell cultures. The team is highly interdisciplinary and brings together expertise in mathematics, bioinformatics, and chemistry. We combine dry-lab research with the work in our wet lab equipped with cutting-edge instrumentation for metabolic imaging.

Our tools exploit various analytical techniques based on mass spectrometry – in particular high-resolution imaging mass spectrometry, which can generate 100 gigabytes of information-rich data in a single sample. Recently, we developed techniques for the molecular annotation of this big amount of data and applied it to various biological systems. We were able to visualise hundreds of metabolites with spatial resolution down to 5 um in both 2D and 3D. Our applications include studying metabolic interactions of co-cultured microbial colonies, alterations in metabolic pathways due to therapy response in both cell cultures and model systems, and performing large-scale analysis of the human skin surface.

Future plans

- High-throughput metabolic imaging of biological tissues, agar plates and cell cultures in 2D and 3D.
- Spatial analysis of metabolic pathways and spatial pharmacometabolomics.
- Open bioinformatics engine for spatial metabolomics.

Figure 1: Surface mapping of two metabolites on the skin of female and male individuals. The models are overlaid with a molecular network in background showing the structural relations between these and hundreds of other detected metabolites (Bouslimani A, *et al.*, *PNAS,* 2015). This study was highlighted as an Image of the Year by *Nature* in 2015.
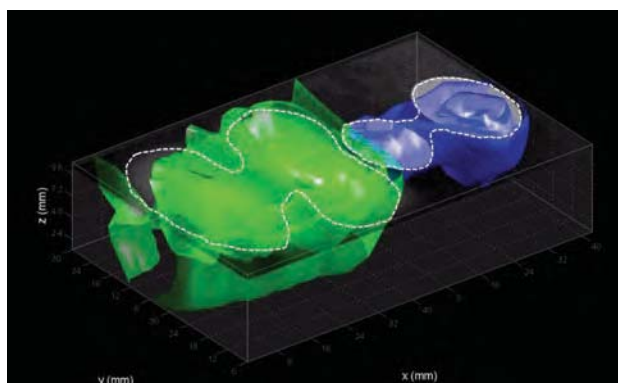

Figure 2: 3D spatial localisation of two metabolites (green, a rhamnolipid with an inhibitory function, and blue) secreted within the agar medium by the interacting colonies of *P. aeruginosa* and *C. albicans* (Watrous *et al., ISME J.,* 2013).

Mechanism, regulation, and exploitation of mobile DNA

Orsolya Barabas

PhD 2005, Eötvös Loránd University, Budapest, Hungary.

Postdoctoral research at the National Institutes of Health, Bethesda, USA.

Group leader at EMBL since 2009.

SELECTED REFERENCES

Smyshlyaev G, *et al.* (2013) Acquisition of an Archaea-like ribonuclease H domain by plant L1 retrotransposons supports modular evolution. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20140-5

Voigt F, *et al.* (2012) Crystal structure of the primary piRNA biogenesis factor Zucchini reveals similarity to the bacterial PLD endonuclease Nuc. *RNA* 18, 2128-34

Guynet C, *et al.* (2009) Resetting the site: redirecting integration of an insertion sequence in a predictable way. *Mol. Cell* 34, 612-9

Barabas O, *et al.* (2008) Mechanism of IS200/IS605 family DNA transposases: activation and transposon-directed target site selection. *Cell* 132, 208-20

Our research focuses on transposons, a class of mobile genetic elements that can autonomously move from one location to another in the genome. They drive genetic diversity and evolution and constitute about half of the human genome. However, the physiological roles of transposons are just starting to be unravelled. Recent studies show that they have key functions in gene regulation, development, immunity, and neurogenesis (Beck *et al.*, 2011). Moreover, these 'jumping' DNA elements offer attractive tools for genetics and human gene therapy.

To better understand transposition and facilitate its applications we investigate the molecular mechanisms of their movement and regulation using structural biology, molecular biology, biochemistry, biophysics, bioinformatics, microbiology, and cell biology approaches. We strive to understand the structure of functional transposition complexes, the chemistry they use to cut and paste DNA, their target-site selection, their abundance, and their regulation in the cell.

Sleeping Beauty: This transposon is a prime tool in vertebrate genetics. We study its structure and mechanisms, investigate how it interacts with other components of human host cells, and develop advanced variants for genetic engineering.

Target site-specific transposons: One of the main obstacles in gene therapy is integration of the therapeutic gene at unwanted locations. Our work revealed that the IS608 transposon uses part of its own sequence to guide integration to a specific site via base pairing (Barabas *et al.*, 2008), and could provide a solution. We are now testing if this target recognition mode can be extended to select unique genomic sites.

Antibiotic resistance carrying elements: The spread of antibiotic resistance is one of today's biggest public health concerns. Conjugative transposons provide a powerful mechanism to transfer resistance between bacteria: we survey their abundance in bacterial genomes and communities and study the mechanisms of two elements from *Helicobacter* and *Enterococcus*.

Transposon regulation: To avoid deleterious outcomes, cells must keep their transposons under control. One major control mechanism is provided by small RNAs and our group investigates these processes in prokaryotes and eukaryotes. Our recent work on the piRNA pathway has revealed the structure and function of a novel factor called Zucchini.

Future plans

- Develop novel genetic engineering tools and explore their applications in transgenesis and synthetic biology.

- Study the mechanism and regulation of a class of 'beneficial' transposons that are involved in the development of ciliated protists.

The Barabas group uses structural and molecular biology approaches to investigate how DNA rearrangements are carried out and regulated, with the ultimate goal of facilitating their applications in research and therapy.

Figure 1: Five crystal structures of the IS608 transpososome together with associated biochemical data elucidate the entire pathway of single-stranded DNA transposition and show how it selects its integration site.
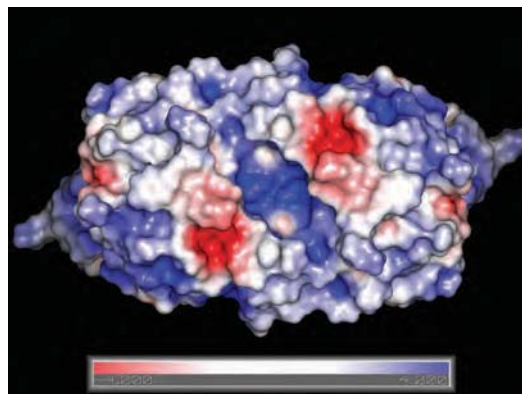

Figure 2: Crystal structure of the primary piRNA biogenesis factor Zucchini reveals its endonuclease function.

Structure and function of large macromolecular assemblies

Martin Beck

PhD 2006, Max Planck Institute of Biochemistry, Martinsried, Germany.

Postdoctoral research at the Institute for Molecular Systems Biology, ETH Zurich, Switzerland.

Group leader at EMBL since 2010.

ERC Investigator.

SELECTED REFERENCES

Ori A, *et al.* (2015). Integrated transcriptome and proteome analyses reveal organ-specific proteome deterioration in old rats. *Cell Systems* 1, 224–37

von Appen, A *et al.* (2015). *In situ* structural analysis of the human nuclear pore complex. *Nature* 526, 140-3

Bui HK, *et al.* (2013) Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* 155, 1233-43

Ori A, *et al.* (2013). Cell type-specific nuclear pores: A case in point for context-dependent stoichiometry of molecular machines. *Molecular Systems Biology* 9, 648

Research in the Beck group combines biochemical approaches, proteomics and cryo-electron microscopy to study large macromolecular assemblies.

Integrated structure determination approaches: Research in our laboratory combines biochemical approaches, proteomics and cryo-electron microscopy to study the structure and function of large macromolecular assemblies. Cryo-electron tomography is the ideal tool to observe molecular machines at work in their native environment (figure 1). Since the attainable resolution of the tomograms is moderate, the challenge ahead is to integrate information provided by complementary approaches in order to bridge the resolution gap towards high-resolution techniques (NMR, X-ray crystallography). Mass spectrometry approaches can provide the auxiliary information that is necessary to tackle this challenge. Targeted mass spectrometry can handle complex protein mixtures and, in combination with heavy labelled reference peptides, provides quantitative information about protein stoichiometries. Using this together with cross-linking techniques can reveal protein interfaces. The spatial information obtained in this way facilitates the fitting of high-resolution structures into cryo-EM maps in order to build pseudo-atomic models of entire molecular machines (figure 2).

Large macromolecular assemblies: Megadalton protein complexes are involved in a number of fundamental cellular processes such as cell division, vesicular trafficking and nucleocytoplasmic exchange. In most cases such molecular machines consist of a multitude of different proteins that occur in several copies within an individual assembly. Their function is often fine-tuned towards context specific needs by compositional remodelling across different cell-types. Structural variations occur through stoichiometric changes, subunit switches or competing protein interfaces. Studying the structure and function of Megadalton protein complexes is a challenging task, not only due to their compositional complexity but also because of their sheer size, which makes them inaccessible to biochemical purification.

Future plans

- To develop integrated workflows for structure determination of large macromolecular assemblies such as the nuclear pore complex (figure 2).

- To reveal the function of cell-type specific variations of macromolecular assemblies.
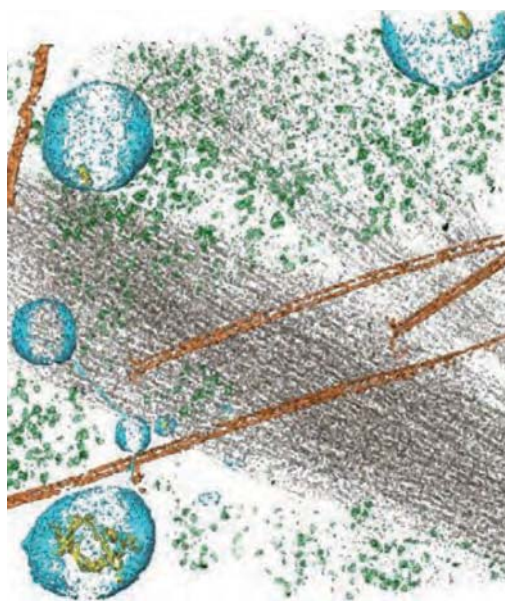

Figure 1. Cryo-electron tomogram of a fraction of the cytoplasm of a human cell. Microtubules are coloured in orange, stress fibres in grey, protein complexes in green, membranes in cyan and vesicular contents in yellow.
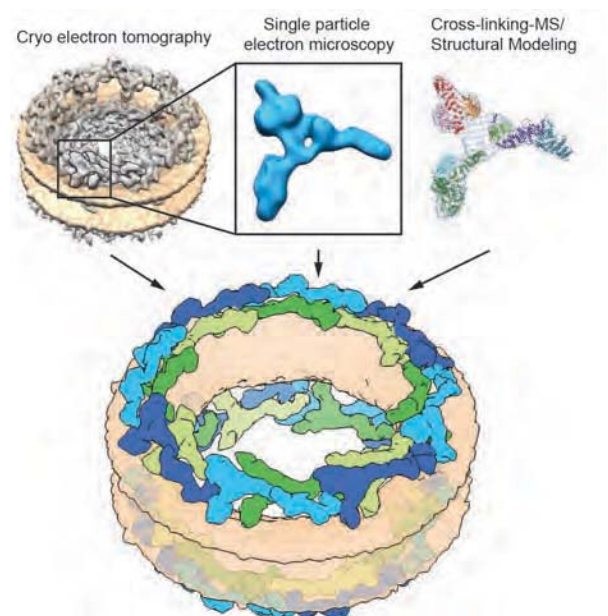

Figure 2. Model of the scaffold arrangement of the human Nuclear Pore Complex revealed by an integrated approach consisting of cryo-electron tomography, single particle EM, cross-linking MS and structural modelling (Bui, von Appen *et al.*, *Cell*, 2013).

Deciphering function and evolution of biological systems

Peer Bork

PhD 1990, University of Leipzig.

Habilitation 1995, Humboldt University, Berlin.

At EMBL since 1991. Joint Head of Unit since 2001.

Strategic head of bioinformatics at EMBL Heidelberg since 2011.

Group leader in the Molecular Medicine Partnership Unit.

ERC Advanced Investigator.

SELECTED REFERENCES

Forslund K, *et al.* (2015) Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* 528, 262-6

Schloissnig S, *et al.* (2013) Genomic variation landscape of the human gut microbiome. *Nature* 493, 45-50

Arumugam M, et al. (2011) Enterotypes of the human gut microbiome. *Nature* 473, 174-80

Bork P, et al. (2015) Tara Oceans studies plankton at planetary scale. *Science* 348, 873 (and associated papers in this special issue)

The group currently works on three different spatial scales, but with common underlying methodological frameworks:

- genes, proteins and small molecules;
- molecular and cellular networks;
- microbial communities.

We usually work in new or emerging research areas and balance methodological work with biological discoveries. Past frontier research projects include the participation in the Human Genome Project (Lander *et al., Nature* 2001), foundational work on the study protein interaction networks (von Mering *et al., Nature* 2002) and comparative metagenomics (Tringe *et al., Science* 2005), and exploration of drug-target interactions using global human "readouts" such as side effects (Campillos *et al., Science* 2008).
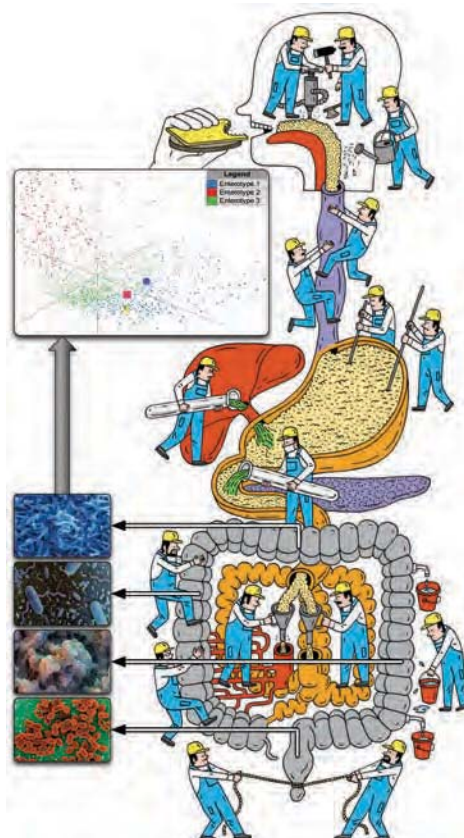
We currently have a number of ocean microbiome projects in the context of the TARA Oceans expedition (Bork *et al., Science* 2015 and references therein), however we mainly focus on the human gut microbiome. We employ metagenomics to uncover the principles of microbial communities in healthy and diseased humans. We identified three main "enterotypes" – or gut microbial community compositions – in developed countries (Arumugam *et al., Nature* 2011), and showed that each human appears to carry individual strains (Schloissnig *et al., Nature* 2013). We found microbial markers for a number of diseases such as obesity (Le Chatelier *et al., Nature* 2013) and colon cancer (Zeller *et al., Mol.Sys.Biol.* 2014) and could demonstrate the impact of drugs on the gut microbiome (Forslund *et al., Nature* 2015).

Furthermore, the environment in the human gut also impacts the efficacy of orally administered drugs: we try to repurpose existing drugs and to understand more about human biology using large-scale integration of various molecular and phenotypic datasets (Kuhn *et al., Mol. Sys. Biol.* 2013, Iskar *et al., Mol. Sys. Biol.* 2013) and also study the impact of drugs on the microbiome.

The main focus of the Bork group is to gain insights into the functioning of biological systems and their evolution by comparative analysis and integration of complex molecular data.

Future plans

We aim to develop the basics for community-based population genetics to understand how microbial communities are transmitted or evolve. This requires studies of communities at the strain level (i.e. of genetic variations). We will monitor strains worldwide and try to use them to understand the principles of successful fecal microbiota transplantations. In the future, we hope to connect microbiomics with diet, host interactions and drug intake. In this regard, we will continue to explore networks between proteins and chemicals such as lipids or carbohydrates and link them to phenotypic data such as disease status, side effects or toxicology. To foster translational research, the group is also partially associated with the Max Delbrück Center for Molecular Medicine in Berlin and with the Molecular Medicine Partnership Unit at the University of Heidelberg.

Multiple roles for different microbial communities in the human gut (modified from the German newspaper Zeit covering the work of the group, original designed by J. Schievink). Metagenomic data from thousands of individuals from all over the world are analysed. For example, three stratifying gut microbial community types (enterotypes) have been discovered in the human population (Arumugam *et al.*, 2011); shown are 1000 individuals clustered by their gut microbial composition. Each individual is a dot, coloured by enterotype.

Viruses and vesicles – cryo-electron microscopy and tomography

John Briggs

PhD 2004, Oxford University.

Postdoctoral research at the University of Munich.

Group Leader at EMBL since 2006.

Senior Scientist since 2013.

Group leader in the Molecular Medicine Partnership Unit.

SELECTED REFERENCES

Avinoam O, *et al.* (2015) Endocytic sites mature by continuous bending and remodeling of the clathrin coat. *Science* 348, 1369-72

Dodonova SO, *et al.* (2015) A structure of the COPI coat and the role of coat proteins in membrane vesicle assembly. *Science* 349, 195-8

Schur FK, *et al.* (2015) Structure of the immature HIV-1 capsid in intact virus particles at 8.8 A resolution. *Nature* 517, 505-8

Kukulski W, *et al.* (2012) Plasma membrane reshaping during endocytosis is revealed by time-resolved electron tomography. *Cell* 150, 508-20

The Briggs group develops and applies cryo-electron microscopy techniques to study the assembly mechanisms of enveloped viruses such as HIV and influenza, as well as coated trafficking vesicles.

We study the structure and molecular assembly mechanisms of important, pathogenic, enveloped viruses (e.g. HIV and Influenza), and of cellular trafficking vesicles (e.g. clathrin and COPI coated vesicles). These extraordinary machines are able to self-assemble, collect cargo and other components, reshape the lipid bilayer to release a vesicle or virus, and then structurally rearrange to identify and fuse with the target membrane. The understanding we aim for could be envisaged as a 3D, functionally-annotated movie, with molecular resolution.

To reach this goal we need detailed structural information at different stages during assembly, ideally under almost native conditions, even within cells. This is difficult with current techniques, so we develop methods for cryo-electron microscopy and tomography, correlated fluorescence and electron microscopy, and image processing. Group members have complementary skills, including biochemistry, cell biology, physics, engineering and computing.

HIV and Influenza viruses
We have a strong interest in the HIV lifecycle, and recently used cryo-tomography methods optimised in the lab to determine the immature capsid structure within heterogeneous HIV particles. We also study the structure and assembly of influenza virus.

Coated vesicles
We study coated vesicles assembly *in vivo* using correlative fluorescence and electron microscopy to find and image intermediate budding steps. Using *in vitro* systems we can get detailed structural information on the arrangement of coat proteins in assembled vesicles. Together these give important insights into how clathrin and COPI mediate vesicle formation.

Innovative methods
Variable, membrane-containing systems such as influenza, HIV, or a COPI coated vesicle cannot be crystallised or averaged using single particle cryo-electron microscopy. We have been developing optimised combinations of cryo-electron tomography and image processing. Using these, we were recently able to resolve individual alpha-helices within intact viruses. We develop correlative fluorescence and electron microscopy methods to find and image rare, transient structures in 3D within cells, interacting with companies to design and apply new technologies.

Figure 1: Correlated fluorescence and electron microscopy can be used to locate a defined intermediate stage in endocytosis and extract quantitative information. This can be applied to multiple stages to understand the whole process (Kukulski *et al.* 2012).
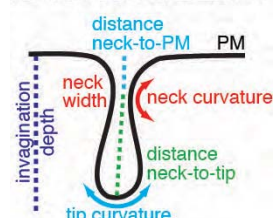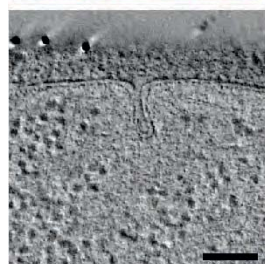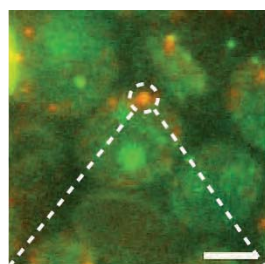
Future plans

Our overarching goal is to understand the interplay between proteins, membrane shape and virus/vesicle structure. What drives virus assembly while maintaining structural flexibility? How do viruses and vesicles that have finished assembly switch to start disassembling? How do proteins reshape cell membranes into vesicles? How do viruses hijack cellular systems for their own use? How does membrane curvature influence protein binding? We also aim to generate detailed mechanistic information on HIV and influenza virus assembly. We develop novel microscopy and image processing approaches to address these questions, and for wide application by other researchers.

Figure 2: An optimised combination of cryo-electron tomography and image processing, developed in the lab, allowed the structure of the immature HIV-1 capsid to be resolved *in situ* – within heterogeneous virus particles. (Schur *et al.* 2015).

Biomolecular networks
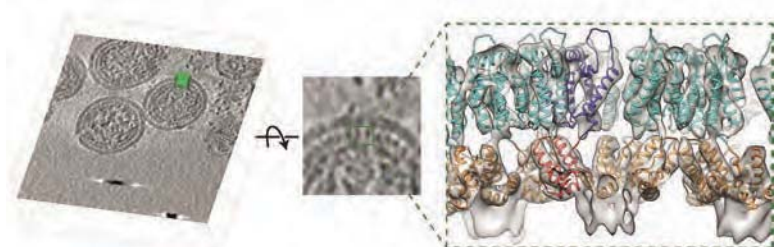
Anne-Claude Gavin

PhD 1992, University of Geneva.

Postdoctoral research at EMBL.

Director, Molecular and Cell Biology, Cellzome AG, Heidelberg.

Group leader at EMBL since 2005.

Senior scientist since 2011.

Elected EMBO Member since 2013.

Group leader in the Molecular Medicine Partnership Unit.

SELECTED REFERENCES

Maeda K, *et al.* (2014) A generic protocol for the purification and characterization of water-soluble complexes of affinity-tagged proteins and lipids. *Nat Protoc* 9, 2256-66

Saliba AE, *et al.* (2014) A quantitative liposome microarray to systematically characterize protein-lipid interactions. *Nat. Methods* 1, 47-50

Maeda K, *et al.* (2013) Interactome map uncovers phosphatidylserine transport by oxysterol-binding proteins. *Nature* 501, 257-61

Kühner S, *et al.* (2009) Proteome organization in a genome-reduced bacterium. *Science* 326, 1235-40

Models of biological systems are expected to be predictive of different healthy and pathological conditions and to provide the general principles for the (re)engineering of biological systems. Our group has pioneered biochemical methods, coupled to quantitative mass-spectrometry, to systematically link dynamic protein interaction networks to various phenotypes in model organisms, human cells and human pathogens. On the long term, we aim to advance network biology and medicine through the integration of quantitative biochemistry, proteomics and structural biology, and define system-wide hypotheses explaining complex phenotypes and human diseases. We will contribute new strategies for the targeting of human pathologies and provide insight into fundamental principles and rules guiding biomolecular recognition.

Charting biological networks: The organisation of biological systems in dynamic, functional assemblies with varying levels of complexity remains largely elusive. One of our main focuses is on deciphering the molecular mechanisms of cell function or dysfunction, which relies to a large extent on tracing the multitude of physical interactions between the cell's many components. We apply a range of biochemical and quantitative mass spectrometry approaches to organisms including yeast, a human pathogen and human somatic stem cells. We aim to identify drug targets and understand the mechanisms and side-effects of therapeutic compounds. Incorporation of structural models, single-particle electron microscopy, and cellular electron tomograms (in collaboration with structural groups at EMBL) provide supporting details for the proteome organisation.

Development of new methods for charting new types of biological networks: While the study of protein-protein and protein–DNA networks currently produce spectacular results, other critically important cellular components – metabolites – have rarely been studied via systematic interaction screens. We currently focus on lipids and have developed new technologies with the capacity to produce systematic datasets measuring protein-lipid interactions. We designed miniaturised arrays of artificial membranes on a small footprint, coupled to microfluidic systems. We have also combined protein fractionation and lipidomics to characterise soluble protein-lipid complexes. We aim to extend the analyses to the entire proteome and lipidome and develop more generic approaches measuring all protein-metabolite interactions.

The Gavin group focuses on detailed and systematic charting of cellular networks and circuitry at molecular levels in time and space.
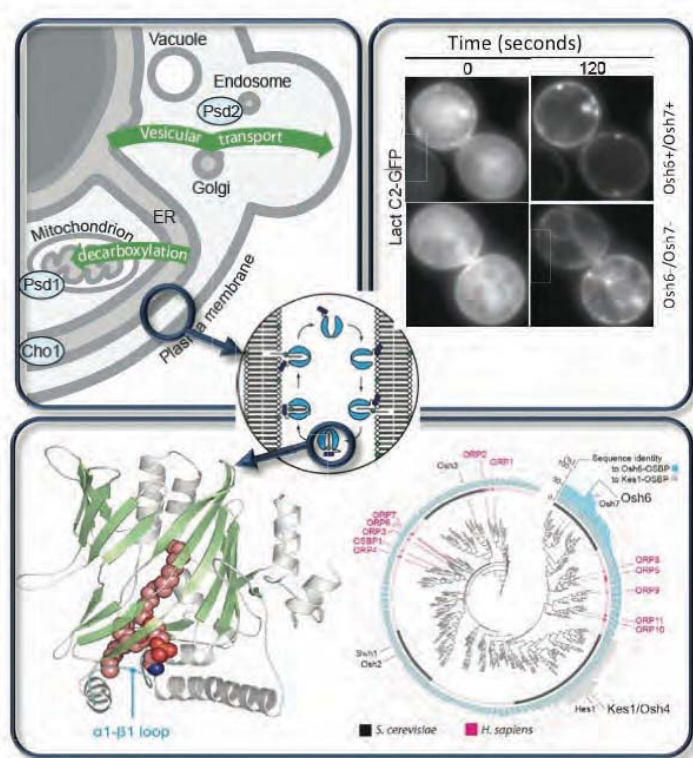
Future plans

- Development of chemical biology methods based on affinity purification to monitor protein-metabolite interactions.

- Global screen aiming at the systematic charting of the interactions taking place between the proteome and the metabolome in the model organism *Saccharomyces cerevisiae* and in human.

- Development of new and existing collaborations to tackle the structural and functional aspects of biomolecular recognition.

The group studies diverse organisms: the yeasts *Saccharomyces cerevisiae*, *Chaetomium thermophilum* (thermophilic eukaryote), the human pathogen *Mycoplasma pneumoniae*, and human somatic stem cells (MMPU group and EU-funded SyStemAge), with datasets contributing detailed cartographies of biological processes relevant to human health or disease. Another major goal is the generation of organism-wide, systematic datasets of protein-metabolite regulatory circuits, and hypotheses or models concerning the consequences of dysfunction in human diseases.

Biological sequence analysis

Toby Gibson

PhD 1984, Cambridge University.

Postdoctoral research at the Laboratory of Molecular Biology, Cambridge.

At EMBL since 1986.

Team leader at EMBL since 1996.

SELECTED REFERENCES

Di Fiore, B. *et al.* (2015). The ABBA motif binds APC/C activators and is shared by APC/C substrates and regulators. *Dev. Cell.*, 32, 358-72

Gibson, T. J. *et al.* (2015). Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad. *Cell Commun. Signal.*, 13, 42

Tompa. P. *et al.* (2014). A million peptide motifs for the molecular biologist. *Mol. Cell.*, 55, 161-9

Van Roey, K. *et al.* (2014). Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem. Rev.*, 114, 6733-78

The Gibson team investigates protein sequences and inter-actions, undertakes computational analyses of macromolecules, and develops tools to enhance sequence analysis research.

Regulatory decisions during eukaryotic cell signalling are made within large dynamic protein complexes by in-complex molecular switching (see Van Roey *et al.*, 2014). Cell regulation is networked, redundant and, above all, cooperative: the proteins involved make remarkable numbers of interactions, and thus have highly modular architectures (Tompa *et al.*, 2014). Tight spatial and temporal regulation goes against the traditional but misleading 'kinase cascade' metaphor.

We host the Eukaryotic Linear Motif (ELM) resource dedicated to short functional site motifs in modular protein sequences, as well as switches.ELM, a compendium of motif-based molecular switches. Linear motifs (LMs or SLiMs) are short functional sites used for the dynamic assembly and regulation of large cellular protein complexes: their characterisation is essential to understand cell signalling. 'Hub' proteins, that make many contacts in interaction networks, have abundant LMs in large Intrinsically Unstructured Protein segments (IUP). Viral proteomes are rich in LMs that are used to hijack cell systems required for viral production. ELM data is now being used by many bioinformatics groups to develop and benchmark LM predictors. We are now actively hunting for new LM candidates and we look to collaborate with groups undertaking validation experiments – see, for example, the recent interdisciplinary collaboration on the ABBA cell cycle regulatory motif (Di Fiore *et al.*, 2015).

We also undertake more general computational analyses of biological macromolecules. Where possible, we contribute to multidisciplinary projects involving structural and experimental groups at EMBL and elsewhere.

Future plans

We will continue to hunt for regulatory motifs and undertake proteome surveys to answer specific questions. Protein interaction networks are anticipated to become increasingly important to our work. Due to the tight integration of protein and RNA molecules in cell regulation, we have a growing transcriptomics focus. We seek to take protein architecture tools, such as 'switches.ELM', to a new level of power and applicability to investigate modular protein function and, in the future, the proteome and protein networks in general. We aim to improve how bioinformatics standards represent cooperative molecular interactions. As part of the EU consortia SyBoSS and SYSCILIA we have been looking at interaction networks and systems in stem cells and in primary cilia.

Linear motifs in T cell signalling, colour-coded by functional class (see Gibson *et al.*, 2015).

Integrated structural biology of translation regulation mechanisms
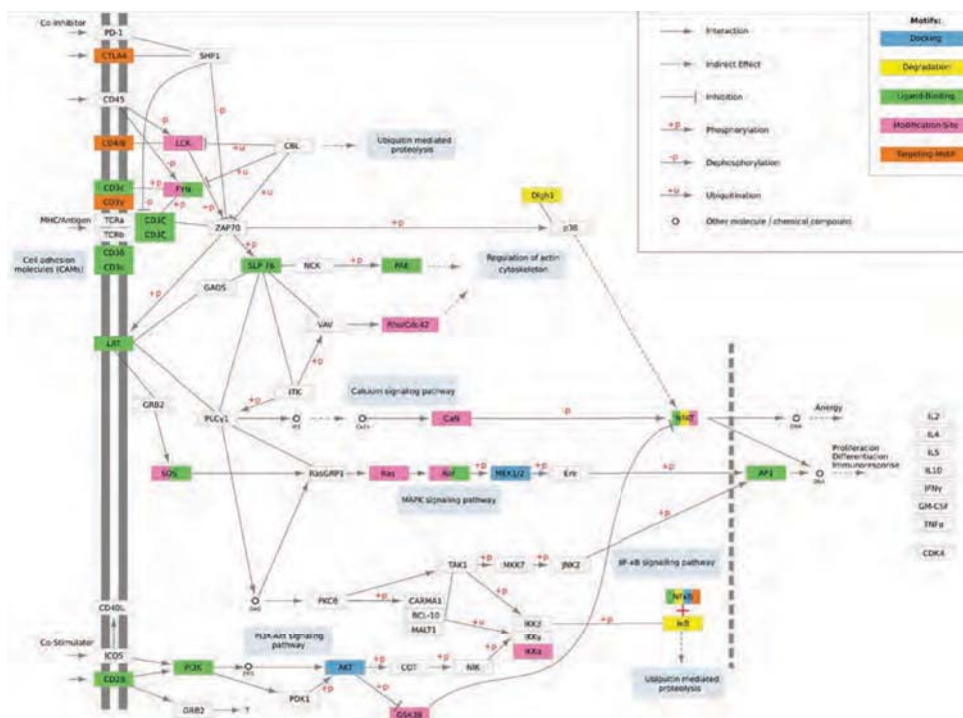
Janosch Hennig

PhD 2009, Linköping University, Sweden.

Postdoctoral research at Helmholtz.

Zentrum München and Technische Universität München.

Group Leader – Emmy Noether Programme since 2015.

SELECTED REFERENCES

Loedige I, *et al.* (2015) The crystal structure of the NHL domain in complex with RNA reveals the molecular basis of Drosophila brain-tumor-mediated gene regulation. *Cell Rep* 13, 1206-20

Hennig J, *et al.* (2015) Structural analysis of protein-RNA complexes in solution using NMR paramagnetic relaxation enhancements. *Meth. Enzymol.* 558, 333-62

Hennig J, *et al.* (2014) Structural basis for the assembly of the Sxl-Unr translation regulatory complex. *Nature* 515, 287-90

Loedige I, *et al.* (2014) The NHL domain of BRAT is an RNA-binding domain that directly contacts the hunchback mRNA for regulation. *Genes Dev.* 28, 749-76

Recognition of mRNA by RNA binding proteins (RBPs) plays essential roles during all stages of gene expression, including splicing, translation, degradation and transport. Also, biogenesis of non-coding RNAs depends on RNA-protein interactions. It is not yet understood in a spatial and temporal manner, how RBPs specify for their cognate RNA sequences on untranslated regions (UTRs). Single RBPs have often a low sequence specificity, especially if they feature only one RNA binding domain. Specificity and affinity increases if RBPs harbour more RNA binding domains. Recently, we have shown that multiple but distinct RBPs act cooperatively to specify for their right target RNA (Figure 1), adding another recognition mechanism to the list of demonstrated protein-RNA interactions (Figure 2).

The atlas of RBPs has been extended to protein domains so far not known to bind RNAs. One of them, the NHL repeat domain, is featured in several members of the eukaryotic tripartite motif (TRIM) protein family. Members of this family featuring this domain are intricately involved in stem and progenitor cell development via interaction with the miRNA-induced silencing complex (miRISC) and are associated to different types of cancer. Biochemically, TRIM proteins are known to act as E3 ligases within ubiquitin signalling. Our research aims to understand the link between ubiquitylation and mRNA translation/degradation mediated by this multi-domain protein family in a structural and functional way.
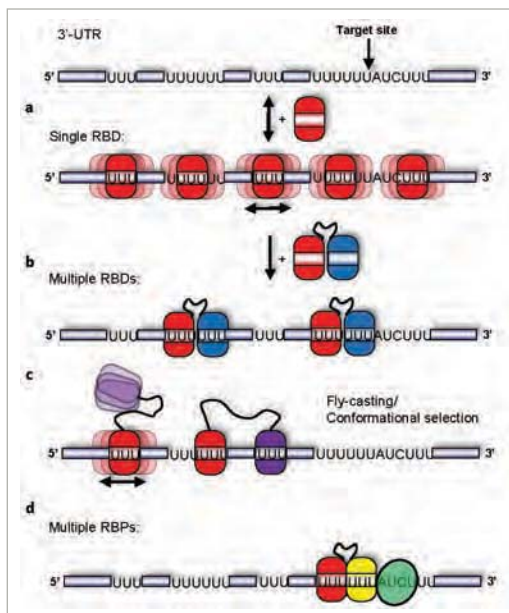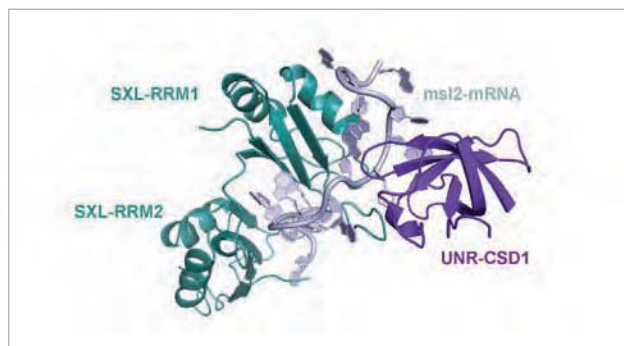
Due to the intrinsically flexible regions in multi-domain proteins and transient protein-protein interactions with other RBPs and proteins of the ubiquitin system, nuclear magnetic resonance spectroscopy (NMR) is well-suited to study TRIM proteins and protein-RNA interactions. However, the size limitations of NMR are disadvantageous, making it necessary to combine sparse NMR data with small-angle scattering and X-ray crystallography.

Future plans

Our future efforts will focus on TRIM proteins and other protein-RNA regulatory networks and will encompass the following main tasks and goals:

- Investigating the structural basis for RNA recognition by the C-terminal NHL domain of different TRIM proteins.

- Investigating the structural basis for E2-E3 recognition within the ubiquitin system, between the N-terminal domains of TRIM proteins and protein interaction partners.

- Linking the complexes formed at the C-terminal and N-terminal domain and get a detailed picture of how TRIM proteins regulate stem cell development.

- Transfer the knowledge to other TRIM proteins involved in immunity.

- Develop and improve methods for integrated structural biology.

The Hennig group employs integrated structural biology, combining nuclear magnetic resonance spectroscopy, X-ray crystallography, and small-angle scattering together with other biophysical techniques to investigate the molecular mechanisms underlying translation regulation with a special focus on multi-protein-RNA assemblies and the tripartite motif protein family.

Figure 1: An example of cooperative, highly specific RNA recognition by two general but distinct RNA binding proteins, Sex-lethal and UNR, regulating the translation of msl-2 mRNA during *Drosophila* dosage compensation (Hennig *et al.,* 2014).

Figure 2: Different modes of RNA binding by RNA binding proteins, where single RBDs, having low specificity can bind to multiple sites on the mRNA transcript (a). Affinity and specificity increases upon cooperative binding of multiple RBDs within a single RBP (b), conformational selection or fly-casting (c), or cooperative interaction between multiple but distinct RBPs.

High resolution studies of protein plasticity

Edward Lemke

PhD, Max Planck Institute for Biophysical Chemistry, Göttingen.

Research Associate, the Scripps Research Institute, USA.

Group leader at EMBL since 2009.

Joint appointment with Cell Biology and Biophysics Unit.

Emmy Noether group leader since 2010.

ERC Investigator since 2015.

SELECTED REFERENCES

Milles S, *et al.* (2015) Plasticity of an ultrafast interaction between nucleoporin and transport receptors. *Cell* 163, 734-45

Nikić I, *et al.* (2014) Minimal tags for rapid dual-color live-cell labeling and super-resolution microscopy. *Angew. Chem. Int. Ed. Engl.* 53, 2245-9

Milles S & Lemke EA. (2014) Mapping multivalency and differential affinities within large intrinsically disordered protein complexes with segmental motion analysis. *Angew. Chem. Int. Ed. Engl.* 53, 7364-7

Tyagi S, *et al.* (2014) Continuous throughput and long-term observation of single-molecule FRET without immobilization. *Nat. Methods* 11, 297-300

The Lemke group uses an interdisciplinary approach to elucidate the nature of intrinsically disordered proteins in biological systems and disease mechanisms.
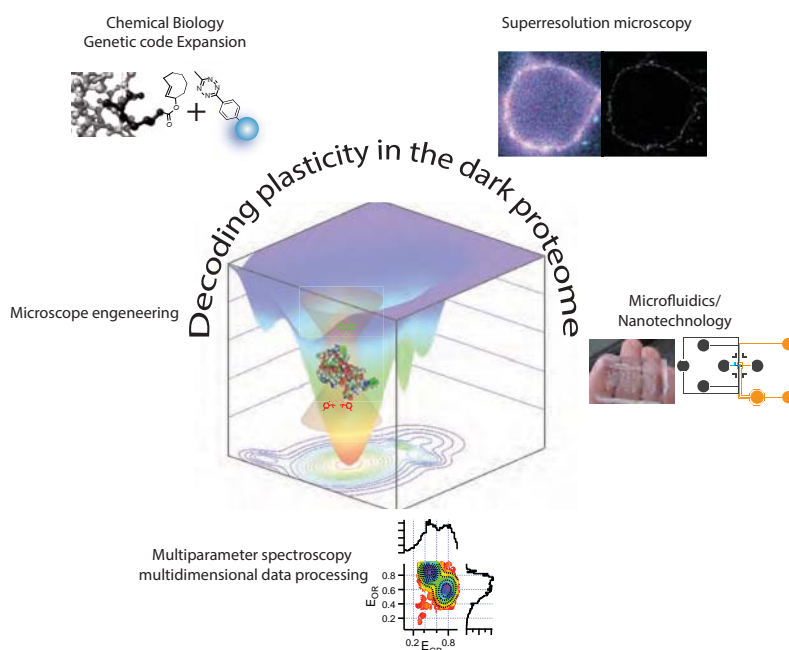
Currently, more than 100 000 protein structures with atomic resolution are available from the protein databank. However, even if all 3D protein structures were available, our view of the molecular building blocks of cellular function would still be incomplete, as we now know that many proteins are intrinsically disordered – unfolded in their native state. Interestingly, the estimated percentage of intrinsically disordered proteins (IDPs) grows with the complexity of the organism (eukaryotes ≈ 50%). Their ability to adopt multiple conformations is considered a major driving force behind their evolution and enrichment in eukaryotes.

Most common strategies for probing protein structure are incompatible with the highly dynamic nature of molecular disorder, so that 50% of the proteome remains "dark", or invisible. In contrast, single molecule and super-resolution techniques, which directly probe the distribution of molecular events, can reveal important mechanisms that otherwise remain obscured. In particular, highly time-resolved advanced fluorescence tools allow probing of molecular structures and dynamics at near atomic scale down to picosecond resolution. While such experiments are now possible in the natural environment of the entire cell, single molecule fluorescence studies *in vitro* and *in vivo* suffer from several limitations such as low throughput and the need for site-specific labelling with special fluorescent dyes. In particular the latter one has become a major bottle neck in state of the art fluorescence.

Besides developing new spectroscopy and microscopy methods, we are utilising a large spectrum of chemical biology and protein engineering tools to overcome these limitations. Our bioengineering efforts allow us to reprogram cells in a way that enables the custom tailoring of proteins with diverse probes, such as dyes and post-translational modifications. This will ultimately enable us to transform living organisms into ideal test beds for molecular, biophysical, and even physiochemical studies of molecular function. Our chemical biology tools also present an ideal interface between the life and material sciences. Furthermore, microfluidics and its potential to miniaturise lab efforts and increase throughput of single molecule science is an area we explore efficiently.

Future plans

Recent studies have shown that even the building blocks with an absolutely critical role in cell survival are largely built from IDPs. For example, many nucleoporins are central to nucleocytoplasmic transport, but also in oncogenesis, chromatin organisation, epigenetic mechanisms, and transcription. Furthermore, viruses extensively use reprogramming of critical IDPs to gain access to, and modify, cellular genomes. How multifunctionality can be encoded into protein disorder is a central question in biology that we aim to answer, as well as integrating our knowledge about such biopolymers towards a better understanding of the life sciences, better drug design, and exploration for bio-inspired material sciences.

Chemical Biology Genetic code Expansion

Superresolution microscopy

Decoding plasticity in the dark proteome

Microscope engeneering

Microfluidics/ Nanotechnology

Multiparameter spectroscopy multidimensional data processing

We innovate new tools, and combine those with our home-built, highly sensitive single molecule and super-resolution equipment to study structure and dynamics of heterogeneous biological systems and pathways, such as viral host pathogen mechanisms and nuclear pore complexes in 4D, and how malfunction of those can lead to disease. We also aim to explore the potential of these IDP biopolymers for novel applications in the life and material sciences.

Molecular mechanisms of transcriptional regulation in eukaryotes

Christoph Müller

PhD 1991, University of Freiburg.

Postdoctoral work at Harvard University, Cambridge, USA.

Group leader at EMBL Grenoble since 1995.

Joint Head of Unit at EMBL Heidelberg since 2007.

Joint appointment with the Genome Biology Unit.

ERC Advanced Investigator.

SELECTED REFERENCES

Hoffmann NA, *et al.* (2015) Molecular structures of unbound and transcribing RNA polymerase III. *Nature* 528, 231-6

Glatt S, *et al.* (2015) Structure of the Kti11/Kti13 heterodimer and its double role in modifications of tRNA and eukaryotic Elongation Factor 2. *Structure* 23, 149-60

Alfieri C, *et al.* (2013) Structural basis for targeting the chromatin repressor Sfmbt to Polycomb response elements. *Genes Dev.* 27, 2367-79

Fernández-Tornero C, *et al.* (2013) Crystal structure of the 14-subunit RNA polymerase I. *Nature* 502, 644-9

We study how sequence-specific transcription factors assemble on DNA and how they interact with co-activators and general transcription factors to recruit RNA polymerases to the transcription start site. We also study the overall structure, architecture and the inner workings of large molecular machines like RNA polymerases or chromatin modifying complexes involved in the transcription process. Finally, we would like to gain insight into how DNA sequence information and epigenetic modifications work together to regulate gene transcription.

To achieve these goals, we use structural information mainly obtained by X-ray crystallography and cryo-electron microscopy combined with other biophysical and biochemical approaches. Systems currently under investigation include multi-protein complexes involved in chromatin targeting, remodelling and histone modifications, yeast RNA polymerase I and III, and the Elongator complex.

Chromatin modifying complexes: The accessibility of chromatin in eukaryotes is regulated by ATP-dependent chromatin remodelling factors and histone modifying enzymes. We study the molecular architecture of chromatin modifying complexes – such as Polycomb group (PcG) protein complexes –, how they are recruited, interact with nucleosome, and are regulated.

RNA polymerase I and III transcription: RNA polymerase I (Pol I) and III (Pol III) consist of 14 and 17 subunits, respectively. Whereas Pol I is responsible for the biosynthesis of ribosomal RNA, Pol III synthesises small RNAs like tRNA and 5S RNA. Misregulation of Pol I and Pol III has been associated with different types of cancer. We are studying the overall architecture of the Pol I and Pol III enzymes and of their pre-initiation machineries using a broad and interdisciplinary approach, combining integrated structural biology with *in vitro* and *in vivo* functional analysis. Ultimately, we would like to understand what features make Pol I and Pol III particularly suitable to fulfil their respective tasks.

Elongator: The 6-subunit Elongator complex is involved in the specific modification of uridines at the wobble base position of tRNAs. Our group recently solved the structure of the Elp456 subcomplex: a ring-like heterohexameric structure resembling hexameric RecA-like ATPases, as well as that of the Kti11/Kti13 heterodimer involved in the regulation of Elongator. We are now pursuing the structural and functional analysis of the entire Elongator complex.

The Müller group uses integrated structural biology, biophysical and biochemical approaches to learn about the molecular mechanisms of transcription regulation in eukaryotes, where DNA is packaged into chromatin.

Future plans

- Molecular insights into the recruitment of transcriptional regulators through the combination of DNA sequence-specific recognition and epigenetic modifications.

- Structural and functional analysis of macromolecular machines involved in transcription regulation, chromatin remodelling and chromatin modification.

- Contributing to a better mechanistic understanding of eukaryotic transcription and epigenetics using integrated structural biology combined with biochemical and cell biology approaches.
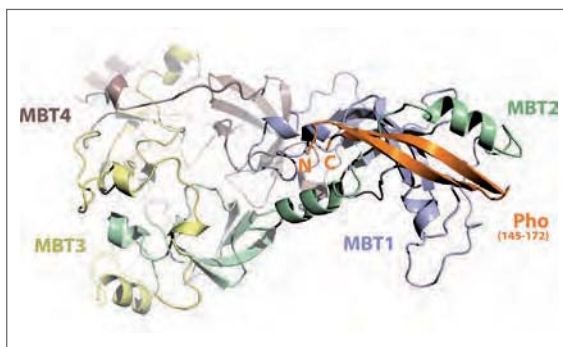

Figure 1: Interaction between the Polycomb group protein Sfmbt 4MBT domain and the Pho spacer region (Alfieri *et al.*, 2013).
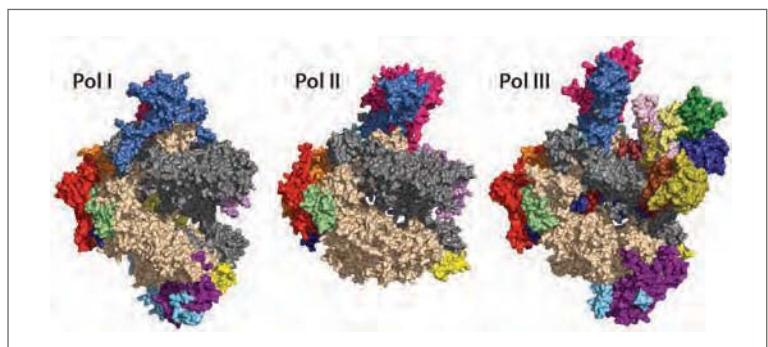

Figure 2: Gallery of eukaryotic RNA polymerases comprising 14-subunit Pol I (Fernández-Tornero *et al.*, 2013), 12-subunit Pol II (PDB 1wcm) and 17-subunit Pol III (Hoffmann *et al.*, 2015).

Architecture and regulation of metabolic networks

Kiran Patil

M. tech. (Chemical engineering) 2002,
Indian Institute of Technology, Bombay.

PhD (Systems biology) 2006, then Assistant Professor,
2006 – 2010, Technical University of Denmark.

Group leader at EMBL since 2010.

SELECTED REFERENCES

Zelezniak A, *et al.* (2015) Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc. Natl. Acad. Sci. U.S.A.* 112, 6449-54

Zelezniak A, Sheridan S & Patil KR (2014) Contribution of network connectivity in determining the relationship between gene expression and metabolite concentration changes. *PLoS Comput Biol.* 10, e1003572

Brochado AR, *et al.* (2010) Improved vanillin production in baker's yeast through *in silico* design. *Microbial Cell factories* 9, 84

Patil KR & Nielsen J. (2005) Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2685-9

The Patil group uses a combination of modelling, bioinformatics, and experimental approaches to study metabolic networks and how they are controlled.

Metabolism is a fundamental cellular process that provides molecular building blocks and energy for growth and maintenance. In order to optimise the use of resources and to maximize fitness, cells respond to environmental or genetic perturbations through a highly coordinated regulation of metabolism. Research in the Patil group focuses on understanding the basic principles of operation and regulation of metabolic networks. We are particularly interested in developing models connecting genotype to the metabolic phenotype (metabolic fluxes and metabolite concentrations) in cell factories and in microbial communities.

With a foundation in genome-scale metabolic modelling, optimisation methods and statistics, we develop novel computational algorithms that are driven by mechanistic insights. For example, we have previously shown that the transcriptional changes in metabolic networks are organised around key metabolites that are crucial for responding to the underlying perturbations (see figure). We complement our computational analyses with experimental activities carried out within our group (microbial physiology and genetics) and in close collaboration with other groups at EMBL and elsewhere (high-throughput phenotyping, metabolomics, proteomics etc.). Such combination of computational and experimental approaches has previously enabled us to improve yeast cell factories producing vanillin – a popular flavouring agent. We are currently developing novel tools, concepts and applications in the following research areas:

i) Metabolic interactions in microbial communities: Microbial communities are ubiquitous in nature and have a large impact on ecological processes and human health. A major focus of our current activities is the development of computational and experimental tools for mapping competitive and cooperative metabolic interactions in natural as well as in synthetic microbial communities. With the help of these tools, we aim at uncovering the role of inter-species interactions in shaping the diversity and stability of complex microbial communities.

ii) Computer-aided design of cell factories: Cell factories, such as yeast and CHO cells, are at the heart of biotechnological processes for sustainable production of various chemicals and pharmaceuticals. We are using modelling and bioinformatics tools to identify genetic re-design strategies towards improving the productivity of such cell factories. These strategies guide our experimental implementation, which in turn help us to further improve the design algorithms in an iterative fashion.

Future plans

We are keenly interested in expanding the scope of our computational and experimental models to gain mechanistic insight into the following biological processes: i) xenobiotic metabolism in microbial communities; ii) crosstalk between metabolism and gene regulatory networks; and iii) metabolic changes during developmental processes. To this end, we are actively seeking collaborative projects within EMBL and elsewhere.

A microbial community can use the biosynthetic capabilities of its members to decrease the collective dependence on nutritional availability from the environment. The figure illustrates this concept using a toy model (left panel) and shows simulation results for communities of different size assembled from a pool of 1200 species (right panel). The metabolic models used here contain around 1000 reactions per species.
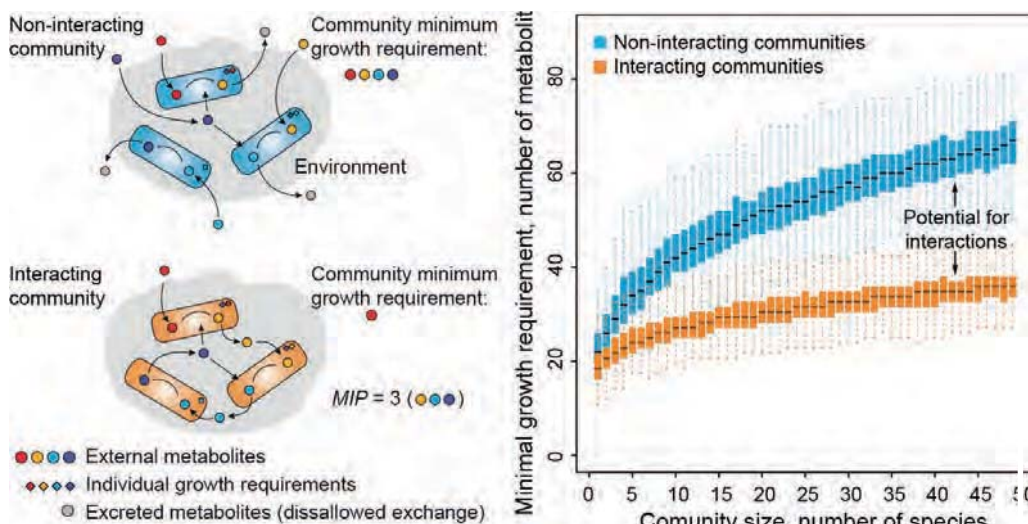
Single-particle electron cryomicroscopy of the autophagy machinery

Carsten Sachse

PhD 2007, University of Jena/FLI-Leibniz-Institute for Age Research and Brandeis University, Waltham, Massachusetts.

Postdoctoral research at Max Planck Research Unit for Enzymology of Protein Folding, Halle, and at MRC Laboratory of Molecular Biology, Cambridge.

Group leader at EMBL since 2010.

SELECTED REFERENCES

Ciuffa R, *et al.* (2015). The selective autophagy receptor p62 forms a flexible filamentous helical scaffold. *Cell Rep* 11, 748-58

Desfosses A, Ciuffa R, Gutsche I, and Sachse C. (2014). SPRING – an image processing package for single-particle based helical reconstruction from electron cryomicrographs. *J Struct Biol* 185, 15-26

Hoffmann NA, *et al.* (2015). Molecular structures of unbound and transcribing RNA polymerase III. *Nature* 528, 231-6

Gutsche I, *et al.* (2015). Structural virology. Near-atomic cryo-EM structure of the helical measles virus nucleocapsid. *Science* 348, 704-7

Autophagy (from the Greek, meaning 'to eat oneself') is the cell's housekeeping mechanism to engulf and degrade long-lived proteins, macromolecular aggregates, damaged organelles and even microbes in double-membrane vesicles called autophagosomes. In our group, we investigate the molecular structures involved in autophagy as they provide fundamental insights for our understanding of aberrant cellular processes like cancer, ageing or infection.

We study the structures of molecular assemblies using biochemical and biophysical techniques, and subsequently visualise them by electron cryomicroscopy (cryo-EM). By this technique, large macromolecular structures and multi-protein complexes can be studied in their near-native environment without the need for crystalisation. Small amounts of material are sufficient to obtain 'snapshots' of single particles in the electron cryomicroscope. The molecular images are combined by computer-aided image processing techniques to compute their 3D structures. As recent advances in hardware and software have led to a wave of atomic-resolution structures, cryo-EM shows great promise in becoming a routine tool for high-resolution structure determination of large macromolecules. To further realise the potential of the technique, the scientific community is still in great need of hardware-based improvements and software enhancements. Therefore, we are also interested in developing techniques, including sample preparation and data processing, to routinely achieve atomic-resolution structures by single-particle cryo-EM. For example, in our group we actively develop the software SPRING for high-resolution cryo-EM structure determination of specimens with helical symmetry.

The Sachse group uses electron cryo-microscopy to study the structures of autophagy complexes to elucidate the mechanisms by which cells eliminate aberrant structures such as large protein aggregates.

Future plans

Multiprotein complexes are essential mediators in the events leading to autophagy. On the structural level however, little is known about their 3D architecture. Fundamental questions on the nature of these complexes need to be addressed:

- How are protein deposits structurally linked to autophagy?

- What are the shapes of these multiprotein assemblies at the membrane?

- How do they give rise to the cellular structure of the autophagosome?

Cryo-EM: High-resolution helical reconstruction of tobacco mosaic virus at 3.3 Å resolution using single-particle cryo-EM from direct electron detectors. Top: helical rod superimposed on cryomicrograph. Center: cross section comprising 17 subunits. Bottom: close-up of α-helix including side-chain density.
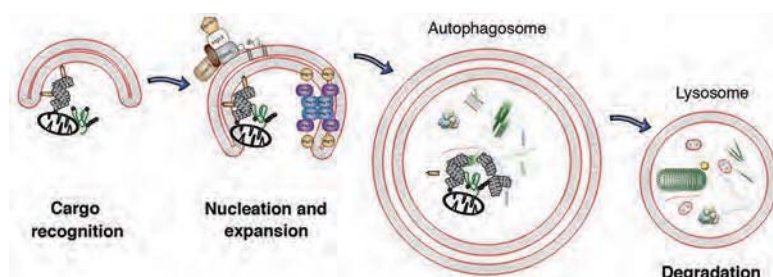
Autophagy: A *de novo* double membrane vesicle entraps large cytosolic cargo such as macromolecules, organelles, protein aggregates and even pathogens destined for degradation in the lysosome.

Personalised genomics to study genetic basis of complex diseases

Judith Zaugg

PhD 2011, EMBL-EBI and Cambridge University

Postdoctoral research fellow, Stanford University

Research group leader at EMBL since 2014

Joint appointment with Genome Biology

Joint appointment with EMBL-EBI

SELECTED REFERENCES

Grubert F*, Zaugg JB*, Kasowski M*, Ursu O*, et al. (2015) Genetic control of chromatin states and gene expression in humans involves local and distal chromosomal interactions. *Cell* 162, 1051-65

Kasowski M*, Kyriazopoulou-Panagiotopoulou S*, Grubert F*, Zaugg JB*, Kundaje A* et al. (2013). Extensive variation in chromatin states across humans. *Science* 342, 750-2

Zaugg JB & Luscombe NM. (2012). A genomic model of condition-specific nucleosome behavior explains transcriptional activity in yeast. *Genome Res* 22, 84-94.

Tan-Wong SM*, Zaugg JB*, Camblong J*, et al. (2012). Gene loops enhance transcriptional directionality. *Science* 338, 671-5

The Zaugg group investigates the variation of molecular phenotypes among individuals along with their genetic variation with the aim of better understanding the molecular basis of complex genetic diseases and inter-individual differences in drug response.

One of the continuing challenges in biomedical research, in particular in translating personalised molecular medicine to the clinic, is to understand the contribution of genetic variation to hereditary traits and diseases. Genome-wide association studies have revealed thousands of associations between genetic variants and complex diseases. However, since most of these variants lie in non-coding parts of the genome, our understanding of the molecular mechanisms underlying these associations is lagging far behind the number of known associations.

To gain a better mechanistic insight into potential causes of known genotype-disease associations, our lab is investigating the variability of molecular phenotypes among individuals and trying to link them to genetic variation. In addition, since many of the disease-associated SNPs are located in regulatory elements, we have a general interest in understanding gene regulatory mechanisms.

Our recent findings indicate that about 15 percent of all regulatory elements, measured through chromatin marks by ChIP-Seq, have a genetic basis, thus challenging the traditional view of chromatin being an epigenetic mark. Many of these so-called histone quantitative trait loci (hQTLs) also have an effect on distal genes or regulatory elements (A) through a mechanism that is likely mediated by transcription factors (B). Importantly, these hQTLs are highly enriched for SNPs that have previously been found to associate with complex traits or diseases, highlighting the functional significance of studying inter-individual variation of molecular phenotypes. We are currently investigating potential mechanisms, such as enhancer compensation models as well as transcript isoform variation, to understand the complex relationship between gene expression and regulatory elements.

Future plans

In the future we will expand our efforts to contributing to the understanding of complex traits and diseases along three lines of research:

- We will apply our models to current genome-wide association studies to increase our power of understanding known associations between genetic variants and complex diseases.

- We will expand our models to include more downstream molecular phenotypes, such as protein levels and complex composition, to estimate the impact of genetic variation on biological pathway activity.

- We will use drug response as a model to investigate the role chromatin in mediating genotype-environment interactions across individuals.
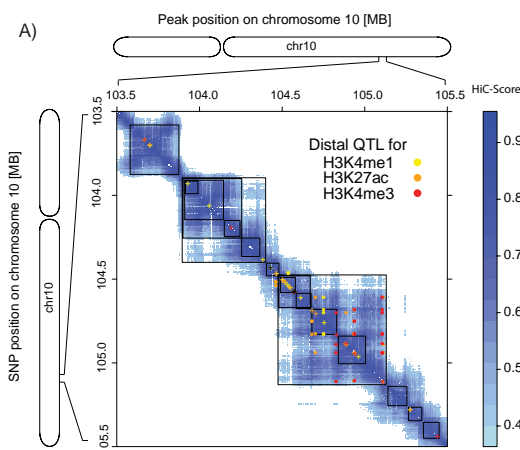

Figure 1: An example region on chromosome 10 displaying the HiC score (blue), local histone QTLs (on the diagonal) and distal hQTLs (off diagonal). Most distal-QTLs lie within the same chromatin domain (indicated in black squares) as their target peak.
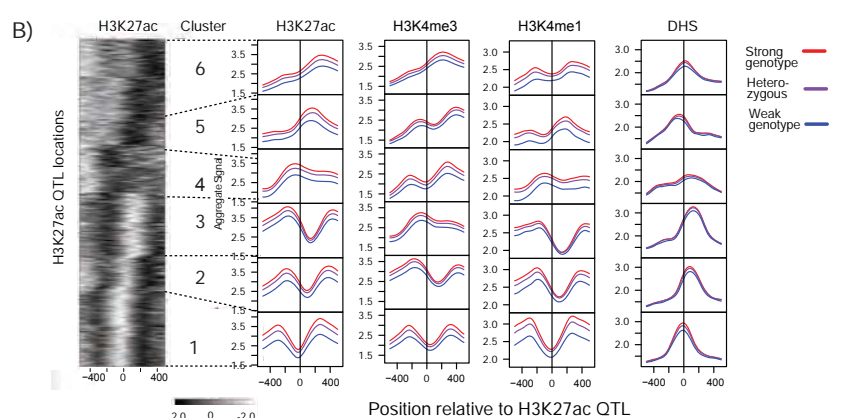
Figure 2: H3K27ac ChIP-seq signal surrounding H3K27ac QTLs was extracted and grouped into six clusters. The aggregate signals for the six clusters are shown for the high-, heterozygous- and low-genotypes (blue, purple, red) for H3K27ac, H3K4me3, H3K4me1, and DNase hypersensitivity sites (DHS) (left to right). The QTL SNPs lie in the nucleosome free regions, indicating that TFs might be driving hQTLs.

Understanding human microbiome perturbations in the context of personalised health and disease

Georg Zeller

PhD 2010, Max Planck Institutes & University of Tübingen.

Postdoctoral research at EMBL Heidelberg.

Team leader at EMBL since 2015.

SELECTED REFERENCES

Sunagawa S, *et al.* (2015) Ocean plankton. Structure and function of the global ocean microbiome. *Science* 348, 1261359

Voigt AY, *et al.* (2015) Temporal and technical variability of human gut metagenomes. *Genome Biol.* 16

Zeller G, *et al.* (2014) Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol. Syst. Biol.* 10

Sunagawa S, *et al.* (2013) Metgenomics species profiling using universal phylogenetic marker genes. *Nat. Methods* 10, 1196-1199

The human microbiome, the complex ecosystem of microorganisms colonising our body, has increasingly been recognised as an important determinant of human physiology. Detailed investigations of microbes in situ (without culturing) have become possible through advances in sequencing technology and computational analysis methodology. These have now started to be applied in large clinical studies to associate changes in microbiome composition and function with human diseases. However, analysis and interpretation of such data remains challenging:

- Quantifying microbial (sub-)species and functions with high accuracy and consistently across various sequencing readouts (16S, shotgun metagenomics and metatranscriptomics) is still difficult for complex communities consisting of many uncultured organisms.

- Microbiome data interpretation is often complicated by many factors that vary in addition to the phenomenon of interest; typical confounders include differences in life-style, co-morbidities or treatments. Comparisons across studies (meta-analyses) are hampered by batch effects arising from technical variation in sample preservation and preparation.

- Perturbations of the microbiome are poorly understood to date. Systematic data and predictive models on the specific effects of environmental exposures (such as host-targeted drugs) on the microbiome are lacking despite this being a key aspect of personalised health and a potential entry point for designing intervention strategies targeted at the microbiome.

To address these challenges, we develop software tools for accurate profiling of both previously sequenced as well as uncharacterised microbial species and the functions encoded in their genomes and transcriptomes. To associate changes in these profiles with host phenotypes, we evaluate various statistics and machine learning tools for their applicability to microbiome sequencing data; we integrate these into software pipelines that make such analyses widely accessible. Using these, we have recently demonstrated that gastro-intestinal diseases can be accurately detected from faecal microbiome readouts. For colorectal cancer in particular this has potential for developing novel non-invasive screening methods (see figure).

The Zeller team develops computational tools for metagenomics data analysis to elucidate the microbiome's role in human health and disease and its responses to chemical perturbations, such as drug treatments.

Colorectal cancer (CRC) can be detected using a classification approach based on microbial markers (top panel) quantified in faecal samples by metagenomic sequencing. Its accuracy was evaluated in cross validation and independent external validation (bottom panels) in comparison to the standard non-invasive screening test (FOBT Hemoccult).

Future plans

- Develop better statistical models specifically for the analysis of microbiome data which are characterised by much larger dispersion than many other types of count data; these also need to deal with confounding in a principled way.

- Develop meta-analysis tools for microbiome research that assess batch effects and try to correct for them where possible.

- Integrative analysis of 16S, metagenomics and metatranscriptomics data with the goal of associating these microbiome readouts to molecular profiles of host health states.

- Contribute analysis methodology to the collaborative efforts involving the Typas, Bork and Patil groups aiming to systematically investigate the effect of chemical (human drugs) and dietary perturbations on the gut microbiome.

EMBL Heidelberg
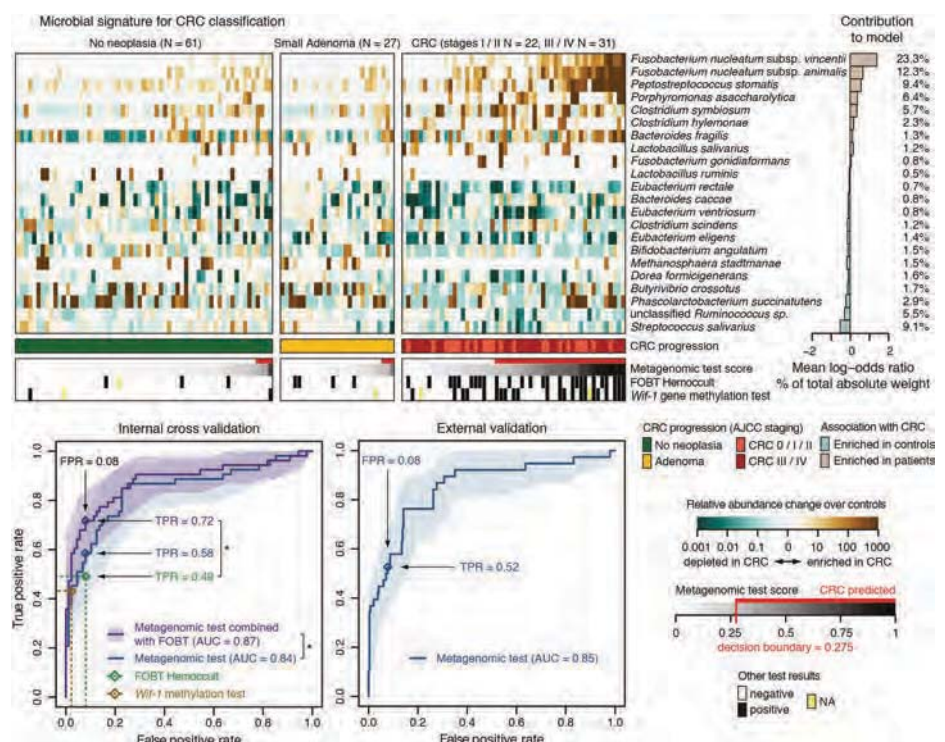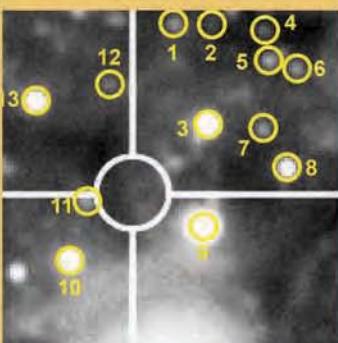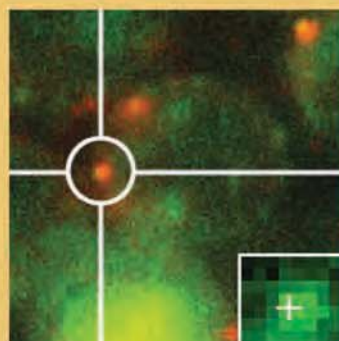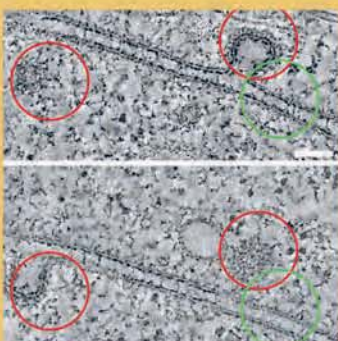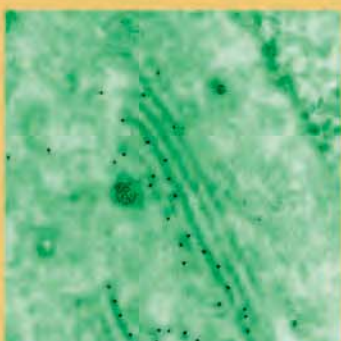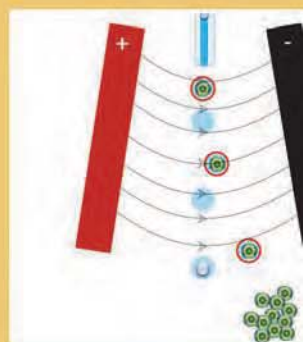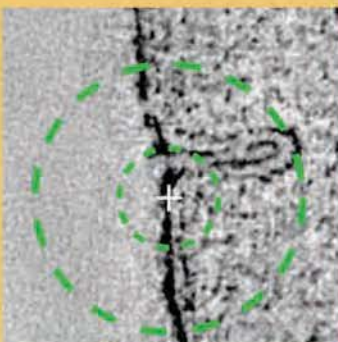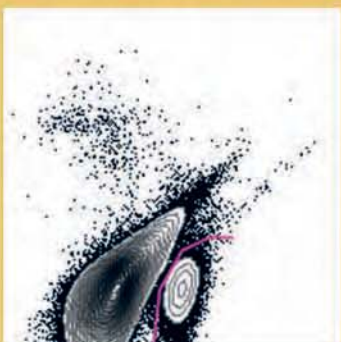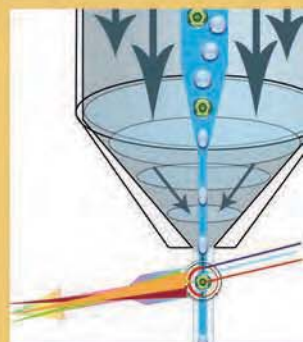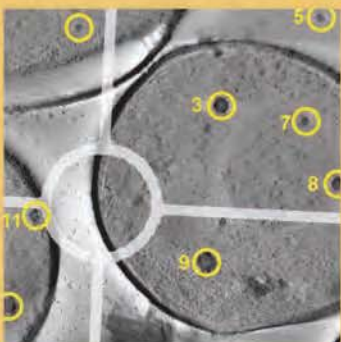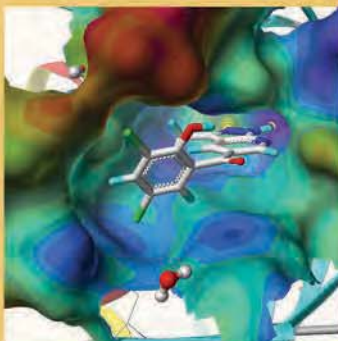
Core Facilities

The EMBL model for Core Facilities has developed a first-rate reputation in the European life sciences community. The Core Facilities contribute significantly to internal and external training courses and workshops, often in collaboration with industrial partners. Moreover, institutions in member states frequently seek our advice and guidance in setting up their own core facilities and services to enhance the efficiency and effectiveness of their scientific research.

EMBL's Core Facilities play a crucial role in enabling scientists to achieve ambitious research goals in a cost effective way. Following the establishment of a small set of facilities in 2001, the support of EMBL Council has enabled significant expansion, with the development of a number of high-level support teams that help focus diverse sets of expertise and multiple cutting edge technologies on specific biological problems. Currently, facilities cover the following areas: Advanced Light Microscopy, Chemical Biology, Electron Microscopy, Flow Cytometry, Genomics, Protein Expression and Purification, Metabolomics and Proteomics. In line with EMBL's mission to provide services to Member States, Core Facilities are open to both internal and external scientists, who benefit significantly from our contributions and advice and are able to conduct research at and beyond normal state-of-the-art.

Core Facilities are staffed by technology experts who focus entirely on service provision, delivering technologies to be used in research projects designed and run by others. Each is run by a Head of Facility who is responsible for daily operations and ensuring high user satisfaction. Close attention is given to the delivery of quality services, fast reaction times to user demands, affordable prices and the complete integration of Core Facilities with the scientific objectives of EMBL.

Such attributes are enhanced by a user committee, which consists of representatives of EMBL's research units. The committee helps to ensure that support activities are tailored to the demands of the research community, supports the introduction of new services, helps to define future strategies and provides valuable feedback on current operations.

Rainer Pepperkok
Head of Core Facilities and Services Unit

Advanced Light Microscopy Facility

Rainer Pepperkok

PhD 1992, University Kaiserslautern.

Postdoctoral research at University of Geneva.

Lab head at the Imperial Cancer Research Fund, London.

At EMBL since 1998.

Senior scientist since 2012.

Head of Core Facilities and Scientific Services since 2014.

SELECTED REFERENCES

Wachsmuth M. *et al.* (2015) High-throughput fluorescence correlation spectroscopy enables analysis of proteome dynamics in living cells. *Nat Biotechnol.* 33, 384-9

Tischer C. *et al.* (2014) Adaptive fluorescence microscopy by online feedback image analysis. *Methods Cell Biol.* 123, 489-503

Ronchi P, Terjung S, & Pepperkok R. (2012) At the cutting edge: applications and perspectives of laser nanosurgery in cell biology. *Biol. Chem.* 393, 235-48

Simpson JC, *et al.* (2012) Genome-wide RNAi screening identifies human proteins with a regulatory function in the early secretory pathway. *Nat. Cell Biol.* 14, 764-74

The Advanced Light Microscopy Facility (ALMF) offers a collection of state-of-the-art light microscopy equipment and image processing tools.

The facility was set up as a cooperation between EMBL and industry to improve communication between users and producers of high-end microscopy technology. The ALMF supports in-house scientists and visitors in using light microscopy methods for their research. The ALMF also regularly organises in-house and international courses to teach basic and advanced light microscopy methods.

Major projects and accomplishments

- The ALMF presently manages about 20 state-of-the-art microscope systems and 10 high-throughput microscopes from leading industrial companies.
- Several workstations for image analysis are provided.
- More than 50 visitors per year come to carry out their own experiments in the ALMF or to evaluate microscopy equipment.
- The ALMF was the seed for the European Light Microscopy Initiative (ELMI) that establishes links between core facilities, users and industry throughout Europe.
- A number of proof-of-concept studies have been hosted in the framework of Eurobioimaging.
- Five genome-wide screens were supported by the ALMF.
- Usage of the facility has exceeded 50,000 hours per year.

Services provided

- Project planning, sample preparation, microscope selection and use, image processing and visualisation.
- Support of advanced microscopy techniques e.g. FRAP, FRET, FCS, laser nanosurgery and super-resolution.
- Developing accessory software and microscopy equipment, co-developments with industrial partners, pre-evaluation of commercial equipment.
- Supporting all aspects of automated microscopy and high-throughput microscopy projects, including RNAi technology.
- Image and data analysis for light microscopy.

Technology partners

The ALMF collaborates with a number of technology partners.
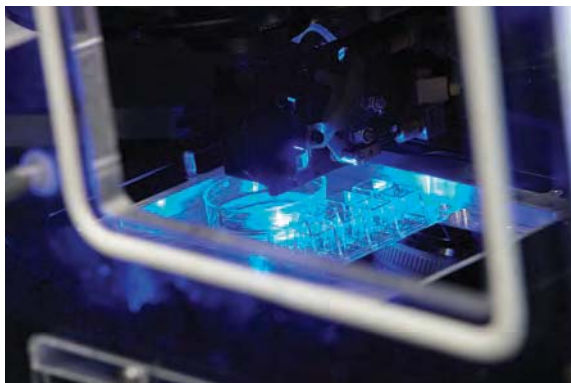
The ALMF currently manages 20 advanced microscope systems and 10 high-content screening microscopes.

Chemical Biology Core Facility

Joe Lewis

PhD 1991, Institute of Molecular Pathology, Vienna.

Postdoctoral research at EMBL.

Group and Global HCV project leader at Anadys Pharmaceuticals, Heidelberg.

MBA 2008, Mannheim Business School.

Facility head at EMBL since 2004.

SELECTED REFERENCES

Kesisova IA, *et al.* (2013) Tripolin A, a novel small-molecule inhibitor of aurora A kinase, reveals new regulation of HURP's distribution on microtubules. *PLoS ONE* 8, e58485

Sehr P, *et al.* (2013) High-throughput pseudovirion-based neutralization assay for analysis of natural and vaccine-induced antibodies against human papillomaviruses. *PLoS ONE* 8, e75677

Bartonova V, *et al.* (2008) Residues in the HIV-1 capsid assembly inhibitor binding site are essential for maintaining the assembly-competent quaternary structure of the capsid protein. *J. Biol. Chem.* 283, 32024-33

Small molecules play essential roles in many areas of basic research and are often used to address important biological questions. Our aim is to enable research groups to address biological questions by identifying and developing 'biotool' compounds against novel targets. We can assist groups in the development of primary and secondary assays for screening against our in-house compound library and guide them through the process of developing tool compounds for their specific target. Chemical optimisation projects can be done in collaboration with our chemistry partners. The facility is a collaboration between EMBL, the German Cancer Research Center (DKFZ), and the University of Heidelberg (since February 2012) to provide the infrastructure and expertise to open up small molecule development to research groups at these institutions.

The facility assists groups in developing primary and secondary assays for screening against the in-house compound library and guide them in developing tool compounds for their specific target.

Major projects and accomplishments

The facility was established at the beginning of 2004. We have a very strong pipeline of projects from all three institutes covering biochemical- and cell-based targets. At the end of 2009 we established computational chemistry as part of the facility offering. Elara Pharmaceuticals GmbH and Savira Pharmaceuticals GmbH have been founded to further develop and commercialise active compounds identified in the facility, targeting specific cancer cell signalling pathways and the influenza virus respectively.

Services provided

Our screening library is composed of around 80 000 compounds. The selection focused on compound catalogues from three leading vendors in the field. Each vendor offers access to significantly larger collections, with low redundancy and highly competitive prices, coupled with attractive options for resupply and follow-up synthesis services. Selected compounds were checked for drug-likeness, structural and shape diversity, novelty, and compliance with medicinal chemistry requirements.

Individual compound selection was done by picking representative compounds around selected scaffolds. A scaffold-based selection offers the advantage of high information screening: as the structural space around each scaffold is covered appropriately, any hit compounds from a high throughput screen can be rapidly followed up by selecting similar compounds to enable initial structure-activity relationships to be discerned. This will help in the prioritisation of the hit compounds for further medicinal chemistry optimisation.

Further services include:

- Selection of appropriate assay technology platforms.
- Developing assays for medium throughput screening.
- Assisting in the design of secondary specificity assays.
- Compound characterisation.
- Managing compound acquisition through our chemistry partners.
- Computational screening using ligand-based and structure-based design strategies.

Ligand docked into target protein.
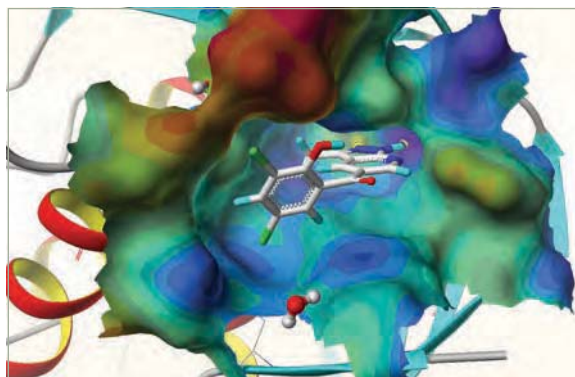
Partners

- **Technology partners:** Perkin Elmer, IDBS, Certara, GE, TTP Labtech.
- **Chemistry partners:** ChemDiv, Chembridge and Enamine.

Electron Microscopy Core Facility

Yannick Schwab

PhD 2001, Louis Pasteur University, Strasbourg.

Postdoctoral research at the University of Calgary, Canada and at the IGBMC, Illkirch, France.

Head of Electron Microscopy at the Imaging Center, IGBMC, Illkirch, France.

Facility head and team leader at EMBL since 2012.

SELECTED REFERENCES

Avinoam O, *et al.* (2015) ENDOCYTOSIS. Endocytic sites mature by continuous bending and remodeling of the clathrin coat. *Science* 348, 1369-72

Francis MK, *et al.* (2015) Endocytic membrane turnover at the leading edge is driven by a transient interaction between Cdc42 and GRAF1. *J Cell Sci.* 128, 4183-95

Mori M, *et al.* (2014) An Arp2/3 nucleated F-actin shell fragments nuclear membranes at nuclear envelope breakdown in starfish oocytes. *Curr Biol.* 24, 1421-8

Foresti O, *et al.* (2014) Quality control of inner nuclear membrane proteins by the Asi complex. *Science* 346, 751-5

The facility provides advanced expertise in electron micros- copy, from sample preparation to image analysis, for a large variety of biological samples.

The EMCF activities cover a large spectrum of EM techniques with a major focus on sample preparation, immuno-localisation of proteins, ultrastructural analysis in 2D and 3D, correlative light and electron microscopy and data processing. Staff in the facility can help you to define optimal experimental conditions for your project – we have experience spanning virtually the full spectrum of biological specimens, with high-level resources for both research and training.

Major projects and accomplishments

Advanced equipment: We offer access to a set of high-pressure freezing machines that are routinely used to vitrify biological samples. Specimens can then be dehydrated, stabilised and embedded in resins in specific freeze-substitution units. Strong expertise has been developed in yeast cells, adherent cultured cells, *Drosophila* embryos, nematodes, zebrafish embryos, and mouse tissues. A microwave-assisted sample processor, used for chemical fixation, dehydration and embedding, greatly reduces time spent preparing the samples (from days to hours). Our electron tomography equipment includes a transmission electron microscope (a FEI 30 300kV microscope with a field emission gun and Eagle FEI 4K camera) and computing set-up with programs for 3D reconstruction and cellular modelling. Specialised EM engineers have expertise in tomography data acquisition and processing.

The Electron microscopist 'savoir faire': We are deeply involved in method development and training. A recent example in correlative light and electron microscopy (CLEM) is the implementation of a technique developed by the Briggs and Kaksonen groups, which tracks the signal of fluorescent proteins in resin sections with high precision.

The future in perspective: Since 2014, the facility has started to provide services in automated serial imaging in scanning electron microscopy (ASI-SEM), such as focused ion beam SEM (FIB-SEM) and serial block face SEM (SBEM). These techniques complement our portfolio of 3D imaging applications (serial section, electron tomography) by providing new opportunities for understanding the cellular fine architecture of multi-cellular specimens. The facility is also offering support for the recently developed cryo-CLEM techniques (see Schorb and Briggs 2014), which will nicely bridge structural and cell biology.

On section CLEM: the signal from fluorescent proteins is used to target the EM analysis to specific subcellular timepoints during the endocytic process (adapted from Avinoam, *et al.* 2015).
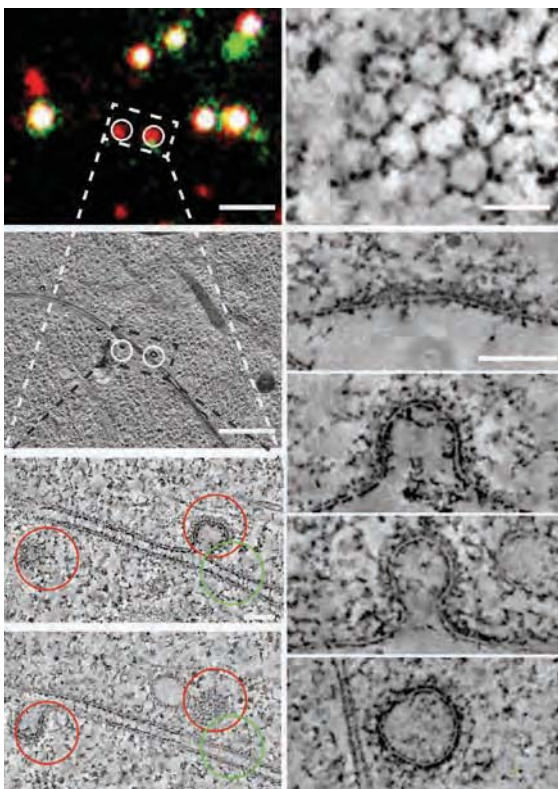
Services provided

Techniques:

- Negative staining.
- Chemical fixation, high pressure freezing of cells and multi-cellular specimens.
- Resin embedding.
- Ultramicrotomy (including serial sectioning).
- Cryo-ultramicrotomy (for the Tokuyasu technique and for CEMOVIS).
- Immuno-labelling.
- 3D EM imaging (TEM tomography, FIB-SEM, SBEM).
- Correlative light and electron microscopy (on section CLEM, cryo-CLEM, 3D CLEM).
- Image analysis and 3D cellular modelling.

Technology partners

FEI Company, Carl Zeiss Microscopy, Leica Microsystems, AbraFluid.

Flow Cytometry Core Facility

Malte Paulsen

PhD 2011, German Cancer Research Center
and the University of Heidelberg

Head of the Flow Cytometry Core Facility at the
Boehringer-Ingelheim Foundation funded Institute for
Molecular Biology (IMB), Mainz.

Head of the Flow Cytometry Core Facility at the National
Heart and Lung Institute, Imperial College London.

Facility Head at EMBL since 2015.

SELECTED REFERENCES

Riddell A, *et al.* (2015) Rmax: A systematic approach to evaluate instrument sort
performance using center stream catch. *Methods* 82, 64-73

Mahen R, *et al.* (2014) Comparative assessment of fluorescent transgene
methods for quantitative imaging in human cells. *Mol. Biol. Cell.* 25, 3610-8

Blake J, *et al.* (2014) Sequencing of a patient with balanced chromosome
abnormalities and neurodevelopmental disease identifies disruption of multiple
high risk loci by structural variation. *PLoS One* 9, e90894

Bonn S, *et al.* (2012) Cell type-specific chromatin immunoprecipitation from
multicellular complex samples using BiTS-ChIP. *Nat Protoc* 7, 978-94

We offer a wide range of flow cytometric techniques, hands on service, data analysis and tailored training sessions for users. Our specialised equipment adds flexibility in the preparation and execution of experiments, allowing tailored approaches to addressing scientific problems. Our facility strives to meet researchers' specific needs by providing dedicated help with experimental design, enabling the highest possible resolution in terms of analysis and product.

We work with equipment from Beckman Coulter, Cytopeia Inc., Becton Dickinson, Union Biometrica, Coherent Inc., and Miltenyi Biotec. We are open to testing new technological developments to best serve the needs of the scientific community.

The goal of the facility is to proactively introduce flow cytometric methods into new research areas while supporting and extending current research.

Major projects and accomplishments

- High-throughput sorting of tissue-specific nuclei from *Drosophila melanogaster* embryo as a preparative step in the analysis of genome regulatory activity during tissue development.

- Establishment of clonal cell lines carrying fluorescent protein-tagged genome-edited genes for 4D life cell imaging, and assessment of protein interactions and concentrations during mitosis.

- Large scale sorting of protoplasts carrying multiple fluorescence reporter constructs for RNAseq.

- Single cell sorting of high efficiencies for single cell quantitative genome, transcriptome, CRISPR/Cas9 genome editing and *in vitro* studies.

- Establishing of in-house training courses for cytometry data analysis and, from 2016 onwards, annual theoretical and practical flow cytometry courses in cooperation with the German Cancer Research Center (DKFZ) and Becton Dickinson.

Services provided

- Complex multi-colour analysis of cell populations based on light scatter, fluorescent probes content and light intensities (including polarisation).

- Full service sorting of rare populations out of a heterogeneous sample. Cell cloning, particle enrichment and high purity bulk sorts.

- Providing EMBL scientific staff with expertise in flow cytometric techniques required in their research projects.

- Providing our researchers with advice and dedicated training sessions in the use of flow cytometry, instrument operation and post-acquisition data analysis.

- Developing novel flow cytometric techniques to meet EMBL's diverse scientific needs.

Single cell sorting of potentially correctly CRISPR-Cas9 edited mES cells

The facility provides key services, such as sorting heterogeneous cell populations into homogeneous populations based on their fluorescence.

Genomics Core Facility

Vladimír Beneš

PhD 1994, Czech Academy of Sciences, Prague.

Postdoctoral research at EMBL.

Facility head since 2001.

SELECTED REFERENCES

Lin Q, *et al.* (2015) Epigenetic program and transcription factor circuitry of dendritic cell development. *Nucleic Acids Res.* 43, 9680-93

Bleckmann M, *et al.* (2015) Genomic analysis and isolation of RNA polymerase II dependent promoters from *Spodoptera frugiperda. PLoS One* 10, e0132898

Benes V, *et al.* (2015) Identification of cytokine-induced modulation of microRNA expression and secretion as measured by a novel microRNA specific qPCR assay. *Sci Rep.* 5, 11590

Agarwal P, *et al.* (2015) CGGBP1 mitigates cytosine methylation at repetitive DNA sequences. *BMC Genomics* 16, 390

GeneCore is the in-house genomics service centre at EMBL equipped with state-of-the-art technologies required for functional genomics analyses and operated by highly qualified staff

The Genomics Core Facility (GeneCore) provides its services to a broad range of users ranging from small research groups to international consortia. Our massively parallel sequencing (MPS) suite boasts HiSeq2000, HiSeq4000 and cBot instruments, as well as MiSeq and NextSeq sequencers. We also work in close collaboration with the University of Heidelberg and the German Cancer Research Center (DKFZ) to further enhance the capabilities of these sequencing technologies. Preparation of MPS libraries for various applications is supported by a robust instrumentation infrastructure (e.g. Covaris, Bioanalyzer, AAT Fragment Analyzer, Qubit, and more). To deal with increasing numbers of incoming samples, we recently reinforced our instrumentation infrastructure through acquisition of Beckman FX liquid handling robots.

Major projects and accomplishments

GeneCore provides the following analyses in a single- or a pair-end sequencing mode, including multiplexing and mate-pair libraries:

- Genome-wide location analysis of nucleic acid-protein interactions – ChIP-Seq, CLIP-Seq.
- Transcriptome sequencing: RNA-Seq.
- Discovery of small non-coding RNAs: ncRNA-Seq.
- Genome-wide DNA methylation analysis: Methyl/BS-Seq.
- *De novo* sequencing & re-sequencing of genomic DNA.
- Targeted enrichment (sequence capture) in solution coupled with MPS.
- Single-cell genomic applications (scRNA-Seq, scDNA-Seq).

GeneCore continues to establish new protocols enabling the processing of challenging samples such as low input or metagenomics samples. For analysis of MPS data, we work intensively with EMBL's bioinformatics community on the development of in-house, freely accessible tools. To date, GeneCore has generated around 100 terabases of MPS sequence data for its users. GeneCore staff also train individual researchers and organise practical courses on corresponding subjects.

Services provided

- MPS sequencing, microarrays (homemade, commercial).
- miRNA qPCR profiling, Bioanalyzer, liquid handling robotics.
- Access to instruments and complete support: qPCR, NanoDrop, PCR cyclers.

We offer processing of samples for a range of microarray applications (mRNA, miRNA and other ncRNA expression profiling, comparative genome hybridisation) available from Affymetrix and Agilent platforms and, upon demand, spotting of customised arrays. In addition to three qPCR instruments managed by GeneCore, our qPCR capacity has been considerably enhanced by a Fluidigm Biomark HD instrument – a device capable of quantitation of transcripts on a single cell level.

Technology partners

MPS continues to be a very dynamic and rapidly evolving technology. We collaborate with several companies involved in developing MPS-related products, for instance testing them in our workflows. GeneCore is a member of the early-access program of Illumina, Agilent, NuGEN and Beckman Coulter. During 2014 we began an extensive collaboration with New England Biolabs and Hamilton aiming at implementation of NEB MPS protocols to automated liquid handling robots.

GeneCore processes users' samples with the help of top-end instruments.

Metabolomics Core Facility

Theodore Alexandrov

PhD 2007, St. Petersburg State University.

Postdoctoral Researcher, University of Bremen.

Group Leader, University of Bremen.

Assistant Adjunct Professor, University of California San Diego.

Team leader at EMBL since 2014.

The Metabolomics Core Facility will open in 2016. It will provide a comprehensive infrastructure for analysis of metabolites and lipids in several biological systems. It aims to provide EMBL scientists with a new and cutting-edge service to facilitate future research into the emerging field of metabolomics and lipidomics.

Metabolomics is associated with detection, quantification, and localisation of small molecules metabolites, which play central roles in biological processes of health and disease within a cell, organ, and organism and in chemical communication between them (Baker, 2011, *Nat Methods*). The facility is equipped with high-resolution mass spectrometry coupled with a chromatographic system (LC-MS) for separation and detection of several classes of metabolites and lipids.

The Metabolomics Core Facility provides a comprehensive infrastructure for analysis of metabolites and lipids.

Major projects and accomplishments

- Identification and detection of several classes of small molecule metabolites in mammalian cell and microbial samples.

- Mass spectral library for over 500 metabolite standards.

- Lipidomic profiling platform across several lipid classes.

- Quantification of polar and non-polar metabolites with a validated quality control system.

- Analysis of drugs and related metabolites.

Services provided

We are providing services for metabolomics mass spectrometry-based analysis. We are performing targeted and untargeted detection and quantification of various classes of molecules ranging from small metabolites to lipids in a variety of biological matrices.

Further services include:

- Global metabolite/metabolomic profiling.

- Targeted metabolomic profiling.

- Lipidomic analysis.

- Metabolite identification.
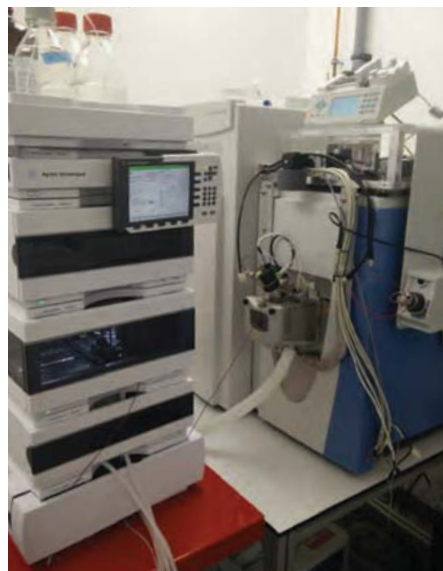
Protein Expression and Purification Core Facility

Kim Remans

PhD 2010, Free University of Brussels

Postdoctoral research, Max Planck Institute of Molecular Physiology, Dortmund

Facility head at EMBL since 2015

SELECTED REFERENCES

Scholz J, *et al.* (2013) A new method to customize protein expression vectors for fast, efficient and background free parallel cloning. *BMC Biotechnol.* 13, 12

Costa SJ, *et al.* (2012) The novel Fh8 and H fusion partners for soluble protein expression in Escherichia coli: a comparison with the traditional gene fusion technology. *Appl. Microbiol. Biotechnol.* 97, 6779-91

Mackereth CD, *et al.* (2011) Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature* 475, 408-11

Gallego O, *et al.* (2010) A systematic screen for protein-lipid interactions in Saccharomyces cerevisiae. *Mol. Syst. Biol.* 6, 430

The Protein Expression and Purification Core Facility produces and purifies proteins both from Escherichia coli *and from insect cells. We implement a large array of chromatographic techniques and also provide support for the biophysical characterisation of the purified proteins.*

The main aim of the facility is to provide assistance and scientific advice concerning all experiments related to protein expression, purification and biophysical characterisation. We can perform quality control assays probing the folding state and stability of the recombinant proteins. We strive to stay up-to-date with the latest scientific developments and trends in the field and test out new products and protocols for our users. Furthermore, we maintain an extensive collection of expression vectors and bacterial strains and we also have some frequently used proteins such as proteases and polymerases readily available. This helps to considerably reduce the expenses of our users.

Services provided

- Expression testing of the protein of interest in various *E. coli* and insect cell strains and in different culturing conditions.
- Purification of recombinant proteins from *E. coli* and insect cells.
- Quality control and biophysical characterisation of purified proteins (for example: analytical gel filtration, analytical ultracentrifugation, ITC, CD, and thermofluor assays).
- Preparation of injection material for immunisations and purifications of antibodies from serum.
- Production of frequently used proteins (for example: proteases, polymerases and fluorophores) for general use within EMBL.
- Providing scientific and technical advice to internal and external researchers.
- Teaching and training of users in all aspects of protein expression and purification techniques.
- Maintaining a large collection of expression vectors and bacterial strains.
- Evaluating new protocols and products for our users.

Technology partners

We are open to collaborations with academic and industrial partners to assess new products or technological developments. Furthermore, we have initiated a network of protein facilities across Europe called Protein Production and Purification Partnership in Europe (P4EU) to improve information exchange and the evaluation of novel technologies.
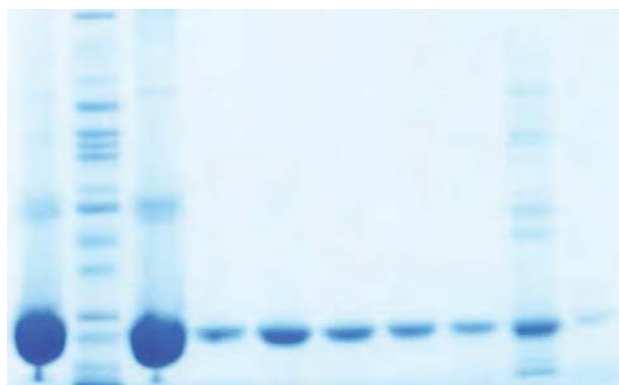
SDS-PAGE analysis of His-TEV purification by ion exchange.

Proteomics Core Facility

Mikhail Savitski

PhD 2007, Uppsala University.

Group leader Cellzome, Heidelberg.

Team leader and Head of Proteomics Core Facility at EMBL since 2016.

SELECTED REFERENCES

Kronja I, *et al.* (2014) Widespread changes in the posttranscriptional landscape at the *Drosophila* oocyte-to-embryo transition. *Cell Rep* 7, 1495-508

Polonio-Vallon T, *et al.* (2014) Src kinase modulates the apoptotic p53 pathway by altering HIPK2 localization. *Cell Cycle* 13, 115-25

Dastidar EG, *et al.* (2013) Comprehensive histone phosphorylation analysis and identification of Pf14-3-3 protein as a histone H3 phosphorylation reader in malaria parasites. *PLoS ONE* 8, e53179

Yokoyama H, *et al.* (2013) CHD4 is a RanGTP-dependent MAP that stabilizes microtubules and regulates bipolar spindle formation. *Curr. Biol.* 23, 2443-51

Infrastructure in the Proteomics Core Facility is centered around state-of-the-art mass spectrometry for MS and LC-MSMS experiments. This is complemented by chromatographic and electrophoretic systems for protein and peptide separation.

Major projects and accomplishments

- Molecular weight determination of intact proteins.
- Identification of proteins from coomassie and silver-stained gels.
- Identification of post-translational modifications.
- Nano flow liquid chromatography coupled to high resolution mass spectrometry: (LC-MSMS) for the identification of proteins in complex mixtures.
- Protein quantification by various stable-isotope labelling strategies (e.g. SILAC).

Services provided

Proteomics:

- Protein identification from gel or in solution.
- High resolution and high mass-accuracy MS, MSMS, and LC-MSMS (Thermo Orbitrap Velos Pro and Q-Exactive) for identification and quantification of proteins in complex mixtures.
- Ion trap (Bruker HCT) LC-MSMS for routine identification of proteins from coomassie and silver-stained gels.
- Triple-quad mass spectrometry (Thermo Vantage) for targeted protein analysis.
- Protein quantification by stable-isotope labeling (SILAC, TMT and dimethyl labelling).
- Identification of post-translational modifications.
- Enrichment of phosphopeptides (TiO2 and IMAC).
- Multi-dimensional peptide separation (isoelectric focusing and liquid chromatography).

Analysis of intact proteins:

- Molecular weight determination of intact proteins by ESI mass spectrometry.
- Determination of N- and C-termini of proteins and products of limited proteolysis.
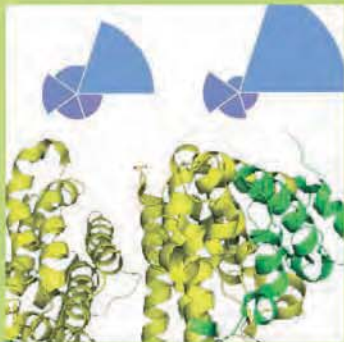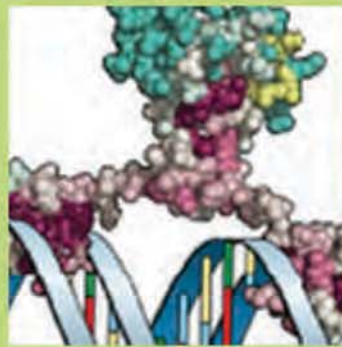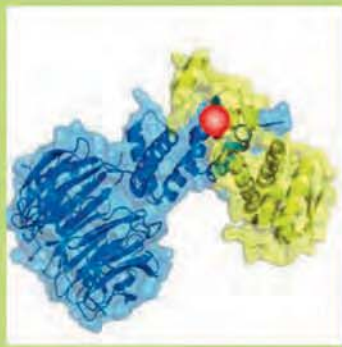- Verification of incorporation of non-natural amino acids.

Technology partner

- BIO-RAD
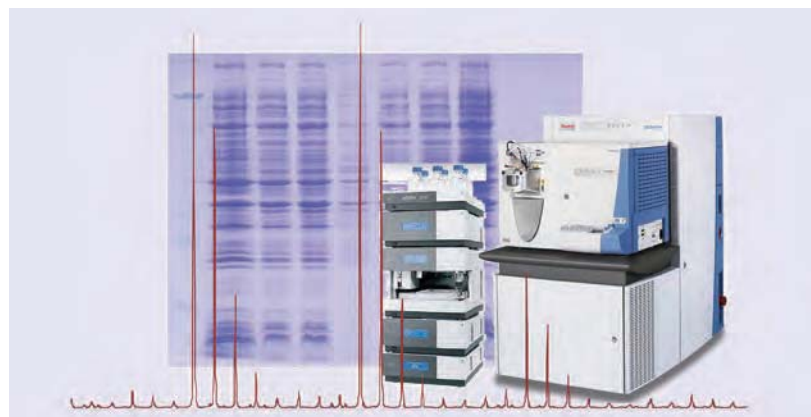
The Proteomics Core Facility provides a full proteomics infrastructure for the identification and characterisation of proteins.

L1   L2   L3

D1   D2   D3   D4

EMBL-EBI Hinxton
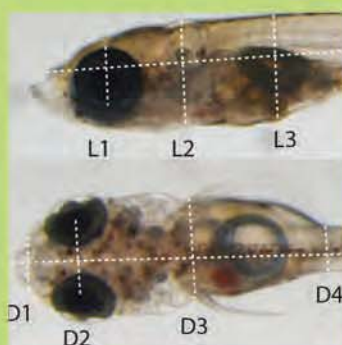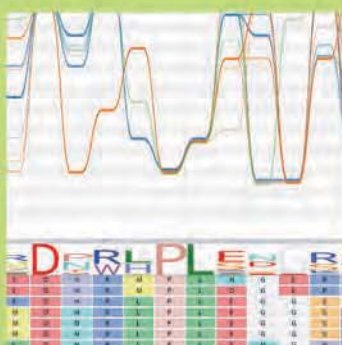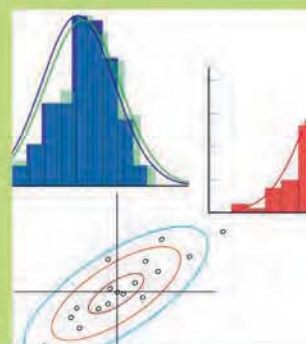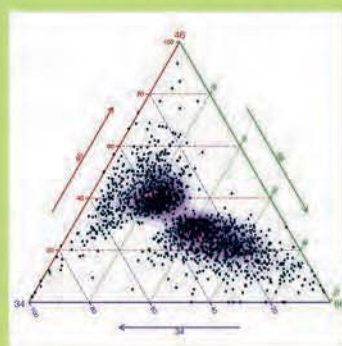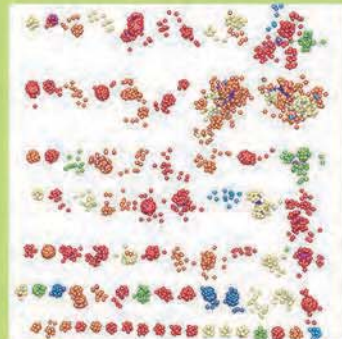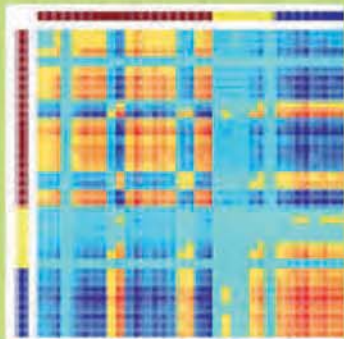
European Bioinformatics Institute

The European Bioinformatics Institute (EMBL-EBI) has been a world leader in computational biology research since its inception in 1994, with work spanning fundamental methods in sequence analysis, multi-dimensional statistical analysis and data-driven biological discovery, from plant biology to mammalian development and disease.

We are highly collaborative and interdisciplinary, regularly publishing high-impact works on sequence and structural alignment, genome analysis, basic biological breakthroughs, algorithms and methods of widespread importance.

Located on the Wellcome Genome Campus just south of Cambridge in the UK, EMBL-EBI is at the centre of one of the world's highest concentrations of technical and scientific expertise in genomics and Big Data in biology, perfectly placed for collaborations of all shapes and sizes. Our researchers collaborate extensively within EMBL and with research groups throughout the world. Notably, a formal partnership between EMBL-EBI and the Sanger Institute and the establishment of the Sanger Institute/EMBL-EBI Single-Cell Genomics Centre have provided a context for a collaboration that has been on going since the Human Genome Project.

Research at EMBL-EBI is carried out both in groups devoted solely to research and by some service team leaders whose research complements their service responsibilities. Most of our research groups collaborate closely with experimentalists, and some generate experimental data themselves. Our research is increasingly related to problems of medical significance, for example developing new ways to understand the molecular drivers of cancer and other diseases. Emerging research themes at EMBL-EBI include computational methods to link genomic information with clinical phenotypes; robust statistical analyses of single cell data; biostatistical models to understand the mechanisms of cancer; the pharmacogenetics of pro-drug activation; exploring regulatory wiring and disease-causing variants with data produced using CRISPR technology; the

effects of proteome variation on cell types, cell states and individuals; and new methods for metabolomics and imaging mass spectrometry.

The future of Big Data analysis in biology is only growing, and we expect to see more opportunities in basic science, data analysis methods development, drug discovery, medicine and other applied fields such as agriculture. We also place a very high value on 'blue skies' research, and can only imagine what discoveries it may unearth.

Ewan Birney
Director of EMBL-EBI and Senior Scientist

Rolf Apweiler
Director of EMBL-EBI and Senior Scientist

Nick Goldman
Research Coordinator and Senior Scientist

Analysis of protein and RNA sequence

Alex Bateman

PhD 1997, University of Cambridge.

Postdoctoral work at the Sanger Centre.

Group leader at Wellcome Trust Sanger Institute 2001-2012.

Head of Protein Sequence Resources at EMBL-EBI since 2012.

SELECTED REFERENCES

Prakash A & Bateman A (2015) Domain atrophy creates rare cases of functional partial protein domains. *Genome Biol.* 16, 88

Petrov A, *et al.* (2015) RNAcentral: an international database of ncRNA sequences. *Nucleic Acids Res* 43, D123-9

Buljan M, *et al.* (2012) Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol Cell* 46, 871-83

Buljan M, Frankish A, & Bateman A. (2010) Quantifying the mechanisms of domain gain in animal proteins. *Genome Biol.* 11, R74

The Bateman group endeavors to classify proteins and certain RNAs into functional families with a view to producing a 'periodic table' of these molecules.

Our work has centred around the idea that there are a finite number of families of protein and RNA genes. We wish to enumerate all of these families to gain an understanding of how complex biological processes have evolved from a relatively small number of components. We have produced a number of widely used biological database resources such as Pfam, Rfam, TreeFam and MEROPS to collect and analyse these families of molecules. Over the years we have published a large number of novel protein domains and families of particularly high interest. For example, we discovered the Paz and Piwi domains which allowed us to identify the Dicer proteins as having an important role in RNAi several months before this was experimentally verified. More recently, we showed that the scramblase genes may act as membrane tethered transcription factors.

Our research interests focus on how proteins and non-coding RNAs interact with each other and how these interaction networks can be rewired due to disease mutations or natural variation. We are interested in how proteins have evolved through the gain and loss of new protein domains. Recently we have been involved in using Wikipedia for collecting community annotation and other biological information for biological databases. Wikipedia provides an enormous opportunity for public engagement in science and we have been encouraging scientists in a number of ways to edit Wikipedia. Current research is looking at identification of non-coding RNAs and understanding the function through computational analysis.

Future plans

We will continue to develop tools and databases to understand the function and evolution of RNA and proteins. Using this data and computational analyses we aim to investigate interaction networks. We will explore the large and growing set of important molecular interactions involving RNA that are currently dispersed among diverse databases and experimental studies. By bringing this data together we wish to uncover the extent and evolution of the RNA interaction network compared to the protein interaction network. In another strand of our research we will develop automated techniques to identify spurious protein predictions that are polluting sequence databases. We have collected thousands of examples of proteins that are unlikely to be translated. These examples will form a good training set for machine learning techniques to identify further suspicious proteins. We are also investigating how we can crowdsource protein families and their annotation to enhance existing databases such as Pfam.
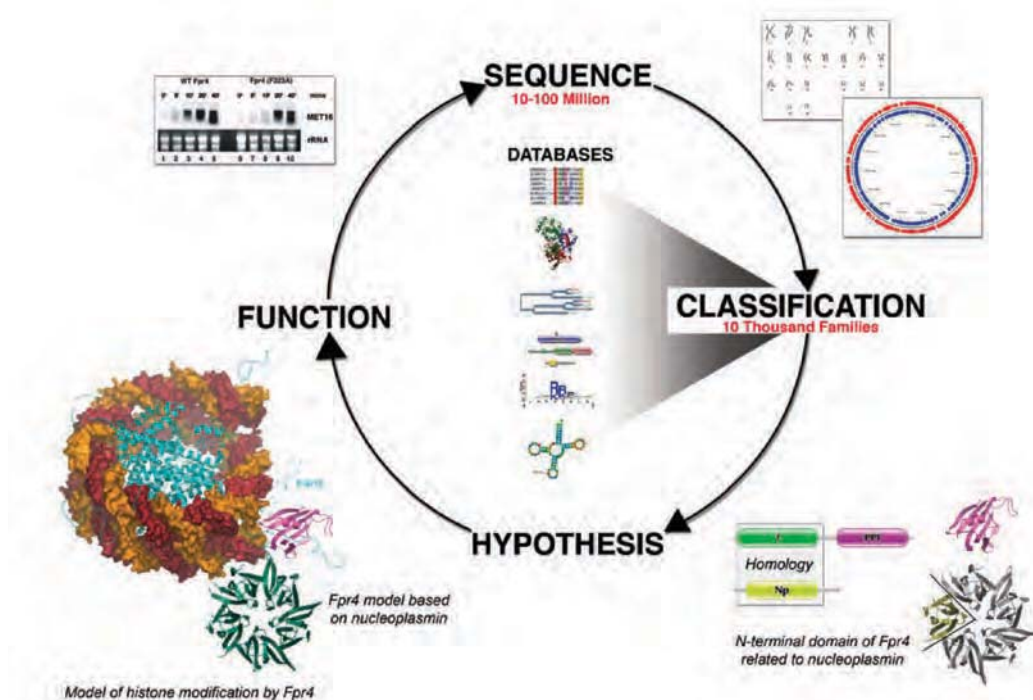
The flow of information from sequence to knowledge of function through classification and hypothesis generation used in the Bateman group.

Evolution of Cellular Networks

Pedro Beltrão

PhD 2007, University of Aveiro (research conducted at EMBL Heidelberg).

Postdoctoral research at the University of California San Francisco.

Group leader at EMBL-EBI since 2013.

SELECTED REFERENCES

Wagih O, *et al.* (2016) Uncovering Phosphorylation-based specificities through functional interaction networks. *Mol Cell Proteomics* 15, 236-45

Johnson JR, *et al.* (2015) Prediction of functionally important phospho-regulatory events in Xenopus laevis oocytes. *PLoS Comput Biol.* 11, e1004362

Strumillo M & Beltrao P. (2015) Towards the computational design of protein post-translational regulation. *Bioorg Med Chem.* 23, 2877-82

Martin H, *et al.* (2015) Differential genetic interactions of yeast stress response MAPK pathways. *Mol. Syst. Biol.* 11, 800

Our group is interested in understanding how novel cellular functions arise and diverge during evolution. We study the molecular sources of phenotypic novelties, exploring how genetic variability that is introduced at the DNA level is propagated through protein structures and interaction networks to give rise to phenotypic variability.

Within the broad scope of this evolutionary problem, we focus on two areas: the function and evolution of post-translational regulatory networks, and the evolution of genetic and chemical-genetic interactions. Looking beyond evolutionary process, we also seek to understand the genomic differences between individuals and improve our capacity to devise therapeutic strategies. In collaboration with mass-spectrometry groups, we develop a resource of experimentally derived, post-translational modifications (PTMs) for different species in order to study the evolutionary dynamics and functional importance of post-translational regulatory networks. We use these data to create novel computational methods to predict PTM function and regulatory interactions. Our goal is to gain insights into the relationship between genetic variation and changes in PTM interactions and function.

The Beltrão group studies the molecular impacts of genetic variability on phenotypic variability in order to understand better the function and evolution of cellular networks.

Future plans

In 2016 we aim to continue our studies of function and evolution of cellular networks with a very strong emphasis on phospho-regulatory networks. Human cells have on the order of 500 kinases that are used by the cell to react to different stimuli and to reach decisions on what changes need to occur to cellular state. We are very interested in understanding how these kinase signalling networks are used, in different environmental conditions, to define specific cellular responses. To study this we have been developing approaches to study cell signalling states using comparative phosphoproteomics. We are also interested in understanding the structural properties that define protein kinase specificity and we continue to develop ways to combine different types of information to predict specificity from sequence. We hope to be able to study the evolution of kinase specificity using these methods. We are also applying some of the tools we have developed to study protein kinases for more applied studies. In collaboration with the Typas group, we are aiming to shed light on how Salmonella cells subvert the normal functions of a human cell to their benefit. In collaboration with Oliver Billker and Jyoti Choudhary at the Sanger Institute, we are studying the phosphoregulation of the malaria parasite in the transition between the human and insect hosts.

Prediction of kinase specificity from interactions networks and protein phosphorylation data. a) Method description 1 – Experimentally identified phosphosites on functional interaction partners of a kinase are collected. 2 – The sites are then subject to motif enrichment to identify over-represented motifs, likely representing the kinase specificity. 3 – Phosphosites matching the top k significant motifs are then retained and used to construct a specificity model. b-e) Examples of predicted specificity models. The top and middle panel of each example shows the specificity of the kinase as constructed from known substrates and as predicted by our method, respectively. The bottom panel shows the top five extracted motifs and the number of phosphosites matching them.

Sequence algorithms and intra-species variation

Ewan Birney

PhD 2000, Sanger Institute.

At EMBL since 2000.

Joint Associate Director, 2012-2015.

Director since 2015.

SELECTED REFERENCES

Ding Z, *et al*. (2014) Quantitative genetics of CTCF binding reveal local sequence effects and different modes of X-chromosome association. *PLoS Genet* 10, e1004798

ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57-74

Goldman N, *et al*. (2013) Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* 494, 77-80

Spivakov M, *et al*. (2012) Analysis of variation at transcription factor binding sites in *Drosophila* and humans. *Genome Biology* 13, R49

The Birney group's research focuses on developing sequence algorithms and using intra-species variation to study basic biology.

Our group has a long-standing interest in developing sequencing algorithms. Over the past four years a considerable focus has been on compression, with theoretical and now practical implementations of compression techniques. Our 'blue skies' research includes collaborating with the Goldman group on a method to store digital data in DNA molecules. Our group continues to be involved in this area as new opportunities arise – including the application of new sequencing technologies, in particular Oxford Nanopore, and the interaction with imaging techniques.

The other of strand of our research focuses around using natural DNA sequence variation to understand basic biology. Over the past five years there has been a tremendous increase in the use of genome-wide association to study human diseases. However, this approach is very general and need not be restricted to the human disease arena. Association analysis can be applied to nearly any measureable phenotype in a cellular or organismal system where an accessible, outbred population is available. We are pursuing association analysis for a number of both molecular (e.g. RNA expression levels and chromatin levels) and basic biology traits in a number of species where favourable populations are available including human and *Drosophila*. In humans we are exploring both molecular traits and physiological traits, such as 4D resolution of human hearts in healthy volunteers. We hope to expand this to a variety of other basic biological phenotypes in other species, including establishing the first vertebrate near-isogenic wild panel in Japanese Rice Paddy fish (Medaka, Oryzias latipes).

Future plans

In 2016 our group will continue to work on sequence algorithms and intra-species variation. Our work with human data will focus on molecular phenotypes in an induced pluripotent stem cell (iPSC) panel generated as part of the HipSci consortium, and on a project based on normal human cardiac data. Our work in *Drosophila* will investigate multi-time-point developmental biology measures. We will also assess the near isogenic panel in Japanese Rice Paddy fish for a number of molecular and whole body phenotypes.

Region-specific changes in left ventricular wall thickness depending on SNP genotype for a locus on chromosome 6. The locus was discovered by a genome-wide association study of 9 Mio SNPs and latent factors of heart morphology in a cohort of 1000 healthy volunteers (Caucasians).
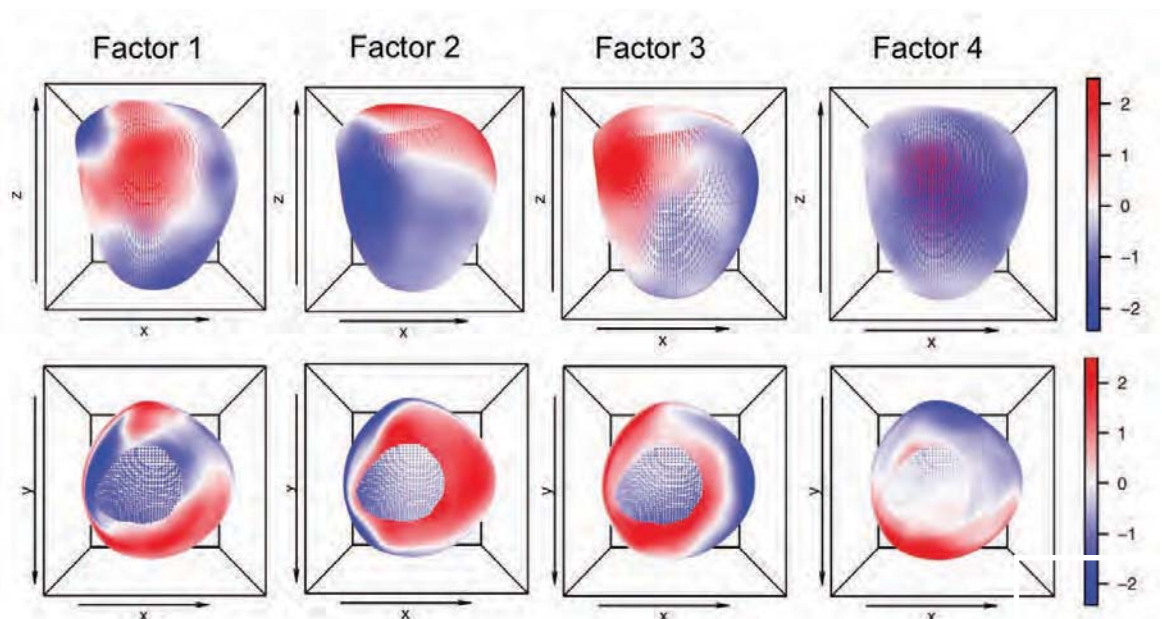
Alvis Brazma

MSc in Mathematics, University of Latvia.

PhD 1987 in Computer Science, Moscow State University.

At EMBL-EBI since 1997.

SELECTED REFERENCES

Frankish A, *et al.* (2015) Comparison of GENCODE and RefSeq gene annotation and the impact of reference geneset on variant effect prediction.
BMC Genomics 16, S2

Scelo G, *et al.* (2014) Variation in genomic landscape of clear cell renal cell carcinoma across Europe.
Nature Communications 5, 5135

Our group focuses on integrative analysis of transcriptomics, proteomics and other types of data across multiple platforms. We are particularly interested in cancer genomics, and in relationships between transcriptomics and proteomics. We collaborate closely with several groups at EMBL-EBI, including the Marioni and Stegle groups.

With our collaborators from Canada, France, UK, Latvia and other countries we co-led a European renal cancer project CAGEKID, a part of the International Cancer Genome Consortium (ICGC). In addition to supporting previous reports on frequent aberrations in the epigenetic machinery and PI3K/mTOR signaling, we uncovered novel pathways and genes affected by recurrent mutations and abnormal transcriptome patterns including focal adhesion, components of extracellular matrix (ECM) and genes encoding FAT cadherins. Jointly with research groups from the Broad Institute, University of California in Santa Cruz and the Sloan Kettering institute, we are now co-leading the Data Integration Working group of the Pan-Cancer Analysis of Whole Genomes project of the ICGC.

We are particularly interested in the usage patterns of different transcripts originating from the same genes and how they are translated in proteome. We have demonstrated that most genes in the human genome express only one transcript to significantly higher levels, that for many genes this dominant transcript does not change between different tissues, and that there is some evidence that the dominant transcript is contributing to the proteome.

Research in the Brazma team focuses on analysis of gene, transcript and protein expression, cancer genomics and proteomics, and integrative analysis of functional genomics data.

Future plans

Large-scale data integration and systems biology will remain in the focus or our research. We will be extending our work on cancer genomics as a part of the pan-cancer project of the International Cancer Genome Consortium, in which we are co-leading the transcriptomics/genomics integration working group that aims to study aberrant transcription patterns across many cancer types. We will extend our research on dominant transcripts to newer, much larger datasets to study how dominant transcripts switch over between different tissues and cell types, and what implication this has on proteome. Jointly with Marioni and Teichmann groups we are starting a new project to define cell types in mammalian species by analysing gene expression data from single cells.

Loss of function through alternative splicing in renal cancer, using the example of the protein phosphatase 2A activator (PPP2R4). Ensembl data shows a switch between two PC transcripts. The APPRIS database, which houses protein structure, function and conservation information for splice variants (developed at CNIO and INB in Spain), shows the principal transcript in N, but not in T. Using the EMBOSS, Needle and UniPDB tools, we see there is less than 35% protein overlap.

Functional Genomics and Analysis of Small RNA Function

Anton Enright

PhD 2003 in Computational Biology, University of Cambridge.

Postdoctoral research at Memorial Sloan-Kettering Cancer Center, New York.

At EMBL-EBI since 2008.

SELECTED REFERENCES

Clark MB, *et al.* (2015) Quantitative gene profiling of long noncoding RNAs with targeted RNA sequencing. *Nature Methods* 12, 339-342

Murray MJ, *et al.* (2015) Solid tumors of childhood display specific serum microRNA profiles. *Cancer Epidemiol. Biomarkers Prev.* 24, 350-360

Cossetti C, *et al.* (2014) Extracellulary via Ifngr1 to activate Stat1 signaling in target cells. *Mol Cell 56,* 193-204

Bussotti G, Notredame C, Enright AJ (2013) Detecting and comparing non-coding RNAs in the high-throughput era. *Int. J. Mol. Sci.* 14, 15423-58

The Enright group focuses on small non-coding RNAs and develops computational tools, systems and algorithms to predict their functions and interactions.

Complete genome sequencing projects are generating enormous amounts of data. Although progress has been rapid, a significant proportion of genes in any given genome are either unannotated or possess a poorly characterised function. Our group aims to predict and describe the functions of genes, proteins and regulatory RNAs as well as their interactions in living organisms. Regulatory RNAs have recently entered the limelight, as the roles of a number of novel classes of non-coding RNAs have been uncovered. Our work involves the development of algorithms, protocols and datasets for functional genomics. We focus on determining the functions of regulatory RNAs including microRNAs, piwiRNAs and long non-coding RNAs. We collaborate extensively with experimental laboratories on commissioning experiments and analysing experimental data. Some laboratory members take advantage of these close collaborations to gain hands-on experience in the wet lab.

Future plans

Our long-term goal is to combine regulatory RNA target prediction, secondary effects and upstream regulation into complex regulatory networks. We are particularly interested in the roles that long non-coding RNAs play in the complex landscape of eukaryotic gene regulation. We are currently working on the largest ever study of the coding and non-coding transcriptome in spermatogenesis with the O'Carroll lab. Recently, epitranscriptomic changes to RNAs have entered the limelight. We have developed a new tool "Chimira" which can accurately assess RNA modifications (e.g. uridylation) in processed microRNAs from NGS data. Additionally, we are working on the exploration of other epitranscriptomic changes (e.g. N-6 Methyladenosine and N-5 Methylcytosine) and how they relate to RNA expression and translation. We will continue to build strong links with experimental laboratories that work on miRNAs and long non-coding RNAs in different systems, as this will allow us to build better datasets with which to train and validate our computational approaches. The use of visualisation techniques to assist with the interpretation and display of complex, multi-dimensional data will continue to be an important parallel aspect of our work.
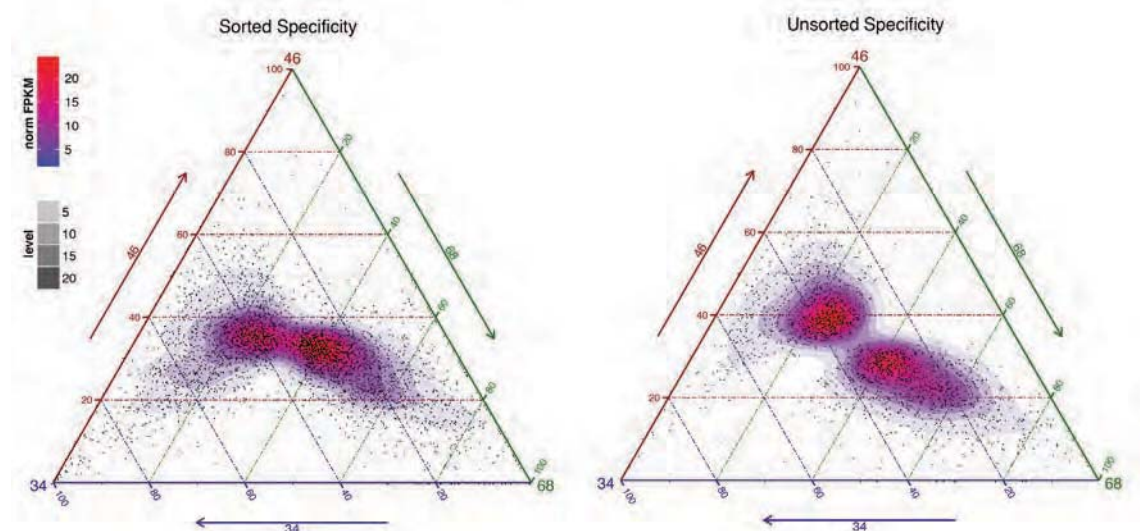
Expression of novel lncRNAs predicted in *D. melanogaster*. Each ternary plot shows expression levels of *de novo* predicted lncRNAs across three time points from both unsorted cells (left) and sorted cells (right).

Evolution of transcriptional regulation

Paul Flicek

DSc 2004, Washington University.

Honorary Faculty Member, Wellcome Trust Sanger Institute since 2008.

At EMBL-EBI since 2005.

Team leader since 2008.

SELECTED REFERENCES

Ing-Simmons E, *et al.* (2015) Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. *Genome Res* 25, 504-13

Villar D, *et al.* (2015) Enhancer evolution across 20 mammalian species. *Cell* 160, 554-566

Zepeda-Mendoza CJ, *et al.* (2015) Quantitative analysis of chromatin interaction changes upon a 4.3 Mb deletion at mouse 4E2. *BMC Genomics* 16, 982

Our team creates and maintains the genomic resources of the Ensembl project and is responsible for data management for a number of large-scale international projects, including the 1000 Genomes Project and, in collaboration with the Brazma team, the International Mouse Phenotyping Consortium. We also collaborate on the development of EMBL-EBI's major variation databases, including the European Genome-phenome Archive (EGA) and the DGVa database of copy number and structural variation. All of these resources are publicly available and are widely used by the scientific community and by the team itself as part of our research into evolution, epigenetics and transcriptional regulation.

Our specific research projects focus on the evolution of transcriptional regulation. Recently we have expanded 'comparative regulatory genomics' techniques including mapping the same DNA–protein interactions in matched tissues in multiple species to understand how gene regulation has evolved while the tissue-level functions are largely conserved. We are also interested in the role of chromatin conformation in tissue-specific gene regulation and have investigated both the CTCF and cohesin complex in this context.

The Flicek team develops large-scale bioinformatics infrastructure, explores the evolution of transcriptional regulation and develops algorithms to study epigenomic data.

Future plans

With the issue of major datasets from the EU-funded Blueprint project and the NIH-funded KOMP2 project, we continue to play an end-to-end role in major genomics projects from raw-data management for the project to summary- data presentation to the wider scientific community.

Our research projects are expanding in number of species, tissues and specific DNA–protein interactions. We will also focus on understanding the differentiation process and components of cell- and tissue-specific regulation. We will address these questions both in the context of our established collaborative projects with the Odom group at the University of Cambridge and as part of other collaborations, including larger EU-funded projects.

1. TF binding site variant
2. Splice donor variant
3. Intron variant
4. Splice acceptor variant
5. 5' UTR variant
6. Initiator codon variant
7. Missense variant
8. Synonymous variant
9. Frameshift variant
10. Stop gained
11. Stop lost
12. 3' UTR variant

A set of annotation terms used to describe the potential effects of sequence variants according to the genic regions they fall in and their allele sequences. The terms are drawn from the Sequence Ontology and are depicted on the molecules they are predicted to affect. Variants categorized as any of the terms 2, 4, 9 and 10 are often collectively referred to as 'loss-of-function' variants, and are typically expected to severely affect gene function.

Computational Cancer Biology

Moritz Gerstung

PhD in Computational Biology, ETH Zurich.

Postdoctoral fellow, Wellcome Trust Sanger Institute.

At EMBL-EBI since August 2015.

SELECTED REFERENCES

Gerstung M, *et al.* (2015) Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. *Nature Comm* 6, 5901

Beerenwinkel N, *et al.* (2015) Cancer evolution: mathematical models and computational inference. *Syst. Biol.* 64, e1-25

Gerstung M, Papaemmanuil E, & Campbell PJ (2014) Subclonal variant calling with multiple samples and prior knowledge. *Bioinform.* 30, 1198-1204

Papaemmanuil E, *et al.* (2013) Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* 122, 3616-27

The Gerstung group investigates the underlying mechanisms of cancer development and seeks to understand the differences in therapy success and outcomes in different individuals.

Cancer is a genetic disease caused by mutations to the genome. When such mutations hit critical genetic elements, they perturb cellular signalling resulting in overly proliferative cells. The availability of cheap sequencing technologies has led to large international efforts such as the International Cancer Genome Consortium for charting the genomic lesions leading to cancer. A revelation of these projects was an even greater genomic complexity of cancer genomes than previously anticipated: Despite having the same disease each patient harbours a unique constellation of mutations. The genetic complexity of cancer is a challenge and an opportunity at the same time. A challenge to understand the underlying mechanisms of cancer development – and an opportunity for finding an explanation for differences in therapy success and outcome.

We have developed statistical models for relating different layers of genomic, molecular and clinical data to extract the precise connections among variables to understand the connection of genotype and phenotype. Moreover we have been working on biostatistical models and informatics tools for predicting outcome based on comprehensive high-dimensional data sets.

Another area of our research is the evolutionary dynamics of cancer. The process of developing cancer is driven by mutation and selection; hence the language to quantify that process is that of evolutionary dynamics. Deep sequencing unmasks the clonal composition of a cancer, which sheds some light on its evolutionary history. Accurate detection of subclonal mutations and reconstruction of phylogenies requires, however, accurate bioinformatics tools that we are actively developing.

Future plans

Our research programme covers different aspects of computational cancer biology. Part of this research will be conducted in local collaboration, and also within national and international initiatives. Future research will involve developing and deploying tools to decipher mutational signatures based on data of comprehensive screens of genotoxins and genetic repair deficiencies. We will continue developing bioinformatical methods for reconstructing the evolutionary history of cancer using NGS data from individual and multiple samples generated as part of international efforts. Lastly, we will work on statistical methods for data-driven personalised outcome predictions.

(a) Predicted risk and observed outcome in 1540 AML patients. (b) Personal constellation of genomic and clinical risk factors for 399 patients. Although there exist some recurrent similarities, each patient has a unique pattern of risk factors.

Evolutionary tools for genomic analysis
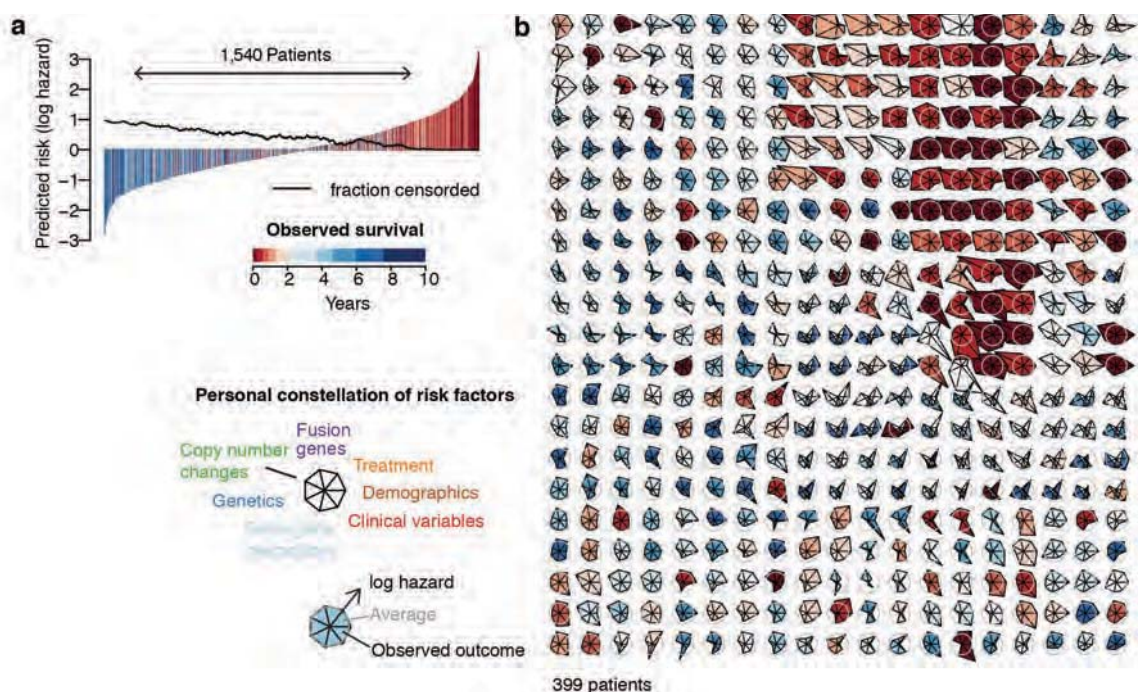
Nick Goldman

PhD University of Cambridge, 1992.

Postdoctoral work at National Institute for Medical Research, London, 1991-1995, and University of Cambridge, 1995-2002.

Wellcome Trust Senior Fellow, 1995-2006.

At EMBL-EBI since 2002.
EMBL Senior Scientist since 2009.

SELECTED REFERENCES

Lowe R, *et al.* (2015) The human blood DNA methylome displays a highly distinctive profile compared with other somatic tissues. *Epigenetics* 10, 274-81

Schwarz RF, *et al.* (2015) Changes in postural syntax characterize sensory modulation and natural variation of *C. elegans* locomotion. *PLoS Comp Biol* 11, e1004322

Schwarz RF, *et al.* (2015) Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. *PLoS Medicine* 12, e1001789

Tan GM, *et al.* (2015) Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Systematic Biology* 64, 778-91

To understand the evolutionary relationships between all organisms, it is necessary to analyse molecular sequences with consideration of their underlying structure. This is usually represented by an evolutionary tree indicating the branching relationships of organisms as they diverge from their common ancestors, and showing degrees of genetic difference between them. We develop mathematical, statistical and computational techniques to reveal information from genome data, draw inferences about the processes that gave rise to these interrelationships and make predictions about the biology of the systems whose components are encoded in those genomes. We develop new evolutionary models and methods, sharing them via stand-alone software and web services, and apply new techniques to interesting biological questions. We participate in comparative genomic studies, both independently and in collaboration with others. Our evolutionary studies involve the analysis of next-generation sequencing (NGS) data, which enables enormous gains in our understanding genomes but poses many new challenges.

Future plans

We are dedicated to using mathematical modelling, statistics and computation to enable biologists to draw as much scientific value as possible from modern molecular sequence data. We will continue to improve and develop new methods for phylogenetic analysis, and new techniques to analyse incomplete datasets. We will apply of our cell lineage tree algorithm to single-cell sequencing data, and further develop the method to identify lineage divergences that are extremely difficult to pinpoint by manual analysis.

Past work in the group was some of the first in phylogenetics to relate protein sequence evolution to features of the entire evolving protein, rather than assuming mutations at different locations had independent effects. We will further develop our methods to model interactions between binding pairs of molecules in signalling networks, instead of focussing on proteins' 3D structures. This will be one the first evolutionary models to account for the fact that proteins do not evolve in isolation but are influenced by the larger network of proteins they interact with, and will enable us to infer how maintenance of cellular information processing has constrained sequence evolution.

Clinicians are looking to genome sequencing to provide diagnostic aids and inform treatment decisions, for example in determining the correct antibiotic based on rapid determination of pathogen species and strain, or detecting mutations known to impact antibiotic resistance. We believe that state-of-the-art genomic analysis methods can assist clinicians and be further optimised to be fast and accurate. In collaboration with clinicians who have expertise in diagnostics and treatment policy, we will work on methods for informing their choices based on bacterial whole-genome sequencing. We will analyse the performance of existing methods for detecting antibiotic resistance using limited data sets, producing knowledge of value for linking the latest NGS technologies with the appropriate software for diagnostic and clinical applications.

The diversity of all life has been shaped by its evolutionary history. Our research focuses on the processes of molecular sequence evolution, developing data analysis methods that allow us to exploit this information and glean powerful insights into genomic function, evolutionary processes and phylogenetic history.

Distribution of yeast trees in tree-space
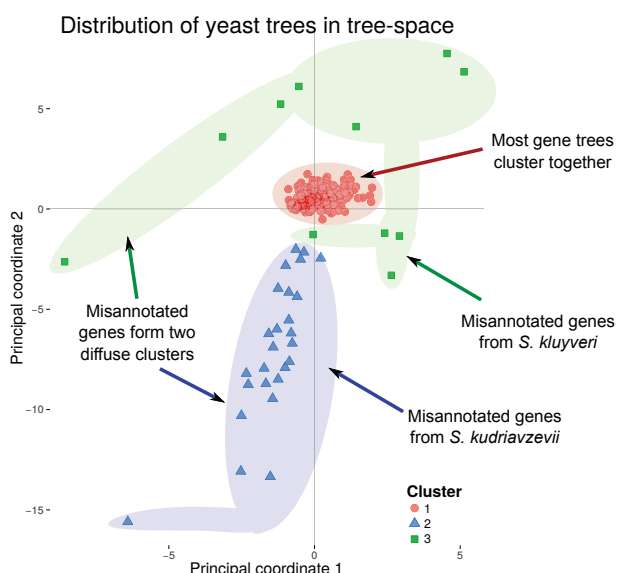
Application of the treeCl clustering method to analyze incongruence in the evolutionary history of 344 yeast genes. The scatterplot shows the embedding in 2D, by multidimensional scaling, of the pairwise geodesic distances between the 344 gene trees. Three clusters were found by spectral clustering: red circles indicate the largest cluster, with 307 members; 37 remaining loci are indicated by blue triangles (cluster 2) and green squares (cluster 3). Loci belonging to the first cluster are tightly grouped and, analysed in combination, yield the correct species phylogeny. In contrast, the second and third clusters are diffuse and the trees belonging to them all have odd and inconsistent phylogenies. This was subsequently determined to be the result of incorrectly called orthology of S. *kudriavzevii* (cluster 2) and S. *kluyveri* (cluster 3) genes.

Protein Data Bank in Europe – Research

Gerard Kleywegt

PhD 1991, University of Utrecht.

Postdoctoral researcher, then independent investigator, then Professor of Structural Molecular Biology, University of Uppsala.

Co-ordinator, then Programme Director of the Swedish Structural Biology Network.

Research Fellow of the Royal Swedish Academy of Sciences.

At EMBL-EBI since 2009.

SELECTED REFERENCES

Velankar S, *et al.* (2012) PDBe: Protein Data Bank in Europe. *Nucleic Acids Res.* 40. D519-24

Gore S, *et al.* (2012) Implementing an X-ray validation pipeline for the Protein Data Bank. *Acta Crystallogr.* D68, 478-83

Patwardhan A, *et al.* (2012) Data management challenges in three-dimensional EM. *Nature Struct. Mol. Biol.* 19, 1203-7

Berman HM, *et al.* (2012) The Protein Data Bank at 40: Reflecting on the past to prepare for the future. *Structure* 20, 391-6

We aim to transform the structural archives into a truly useful resource for biomedical and related disciplines.

The Protein Data Bank in Europe (PDBe) is the European partner in the Worldwide Protein Data Bank organisation (wwPDB), which maintains the single international archive for biomacromolecular structure data. The other wwPDB partners are the Research Collaboratory for Structural Bioinformatics (RCSB) and Biological Magnetic Resonance Bank (BMRB) in the USA and the Protein Data Bank of Japan (PDBj). PDBe is a deposition and annotation site for the Protein Data Bank (PDB) and the Electron Microscopy Data Bank (EMDB).

The major goal of PDBe is to provide integrated structural data resources that evolve with the needs of biologists. To that end, our team endeavours to: handle deposition and annotation of structural data expertly; provide an integrated resource of high-quality macromolecular structures and related data; and maintain in-house expertise in all the major structure-determination techniques (X-ray crystallography, Nuclear Magnetic Resonance spectroscopy and 3D Electron Microscopy). Our specific focus areas are: advanced services, ligands, integration, validation and experimental data.

Future plans

To transform the structural archives into a truly useful resource for biomedical and related disciplines, we focus our developments on five key areas: advanced services (e.g. PDBePISA, PDBeFold, PDBeMotif and the new PDB browsers); annotation, validation and visualisation of ligand data; integration with other biological and chemical data resources; validation and presentation of information about the quality and reliability of structural data; and exposing experimental data in ways that help all users understand the extent to which they support the structural models and inferences.
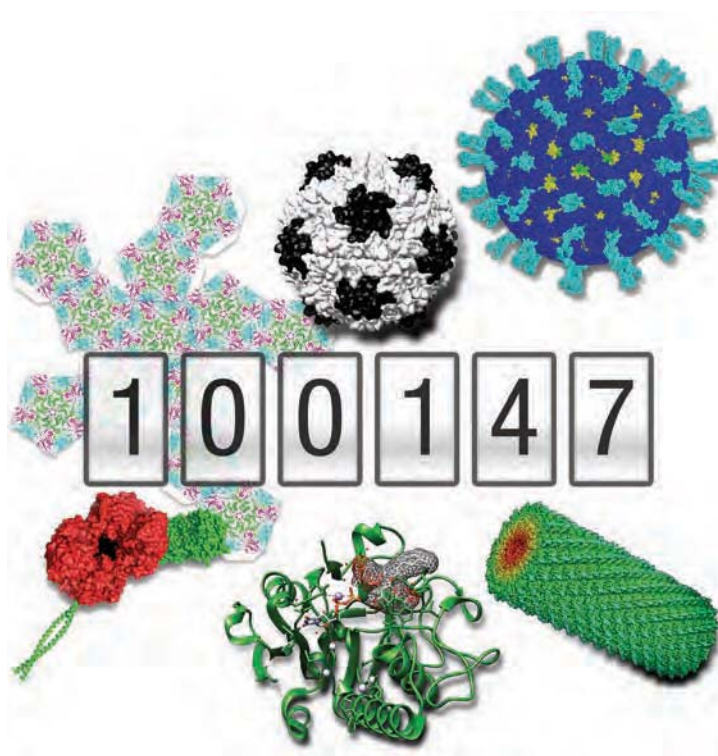
Visualisation and analysis of experimental data and validation information for NMR-derived entries in the PDB, using protein PA1076 from P. aeruginosa as an example (PDB entry 2k4v). Clockwise from top: NMR structure ensemble, coloured by domain; atoms with unusual chemical shift values shown as coloured spheres; violations of distance constraints shown as yellow lines between the corresponding hydrogen atoms; visualisation of experimental torsion-angle constraints and violations; visualisation of residual dipolar coupling data; representative model, with residues coloured according to the Red/Orange/Green quality score assigned by the NRG-CING resource.

Computational and evolutionary genomics

John Marioni

PhD in Applied Mathematics, University of Cambridge, 2008.

Postdoctoral research in the Department of Human Genetics, University of Chicago.

At EMBL-EBI since September 2010.

SELECTED REFERENCES

Achim K, *et al.* (2015) Single-cell expression profiling and spatial mapping into tissue of origin. *Nat Biotechnol* 33, 503-9

Kim JK, *et al.* (2015) Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation. *Cell Stem Cell* 17, 471-85

Kolodziejcyzk AA, *et al.* (2015) Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. *Nature Communications* 6, 8687

Vallejos CA, Marioni JC, Richardson S (2015) BASiCS: Bayesian Analysis of single-cell sequencing data. *PLoS Comput Biol* 11, e1004333

Our research focuses on developing the computational and statistical tools necessary to exploit high-throughput genomics data in order to understand the regulation of gene expression and to model developmental and evolutionary processes.

Within this context, we focus on work on three specific areas. First, we want to understand how the divergence of gene expression levels is regulated. By associating changes in expression with a specific regulatory mechanism, critical insights into speciation and differences in phenotypes between individuals can be obtained. Second, we want to use gene expression as a definition of the molecular fingerprint of individual cells to study the evolution of cell types. By comparing the molecular fingerprint associated with a particular tissue across species, it is possible to decipher whether specific cell types arise de novo during speciation or whether they have a common evolutionary ancestor. Thirdly, we want to model spatial variability in gene expression levels within a tissue or organism. By modelling such variability, heterogeneous patterns of expression within a cell-type can be identified, potentially allowing new cell-types, perhaps with novel functions, to be uncovered. Additionally, the extent of heterogeneity present across a tumour can also be studied using such approaches.

These three strands of research are brought together by single-cell sequencing technologies. By studying variability in gene expression (and other genome-wide characteristics) at a single-cell level, our ability to assay regulatory variation, molecular fingerprints and spatial patterns of expression will be revolutionised. As a key member of the Sanger-EBI Single Cell Genomics Centre we are closely involved in data generation and in using these data, especially single-cell RNA-sequencing, to answer numerous exciting biological questions. However, to exploit these data to the fullest extent, it is critical to develop the appropriate statistical and computational tools – this is one of the key challenges we face in the next few years.

The Marioni group develops statistical tools that exploit data generated using next-generation sequencing to understand the evolution and regulation of gene expression.

Future plans

Our group will focus on developing computational tools for understanding the regulation of gene-expression levels. We will develop methods for analysing single-cell RNA-sequencing data, which has the potential to reveal novel insights into cell-type identity and tumourigenesis. We will extend the model introduced by Buettner et al. to better order cells along a differentiation trajectory and further tease apart the contributions of different factors to heterogeneity in gene expression across cells. We will investigate how Bayesian approaches can be used to better identify highly variable genes across cell types, and develop robust computational approaches for assaying the degree of stochastic, allele-specific expression across single-cell populations. We will also use spatially-resolved single-cell transcriptomic data to identify cell types and examine heterogeneity in expression at the spatial level.

We will use our new methods to obtain insights into cell fate decisions during gastrulation – arguably the most important time in our lives – as part of our work on a project led by Wolf Reik at the Babraham Institute. We will continue to apply our models in biological contexts such as the study of heterogeneity in mouse embryonic stem cells, cancer biology and non-model systems to study evolution.

Classifying genes by their regulatory function. We used RNA-seq data generated from F0 mice and their F1 hybrids to classify genes into sets depending upon their regulatory mechanism (Goncalves *et al.*, *Genome Research*, 2012).

Statistical genomics and systems genetics

Oliver Stegle

PhD 2009 in Physics, University of Cambridge.

Postdoctoral Fellow, Max Planck Institutes Tübingen, 2009-2012.

Group leader at EMBL-EBI since November 2012.

SELECTED REFERENCES

Casale FP, *et al.* (2015) Efficient set tests for the genetic analysis of correlated traits. *Nature Methods* 12, 755-8

Sudmant PH, *et al.* (2015) An integrated map of structural variation in 2,504 human genomes. *Nature* 526, 75-81

Buettner F, *et al.* (2015) Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nature Biotechnology* 33, 155-60

Stephan J, Stegle O & Beyer A (2015) A random forest approach to capture genetic effects in the presence of population structure. *Nature Communications* 6, 7432

The Stegle group develops computational and statistical methods to study the genotype-phenotype map on a genome-wide scale.

Our interest lies in computational approaches to unravel the genotype – phenotype map on a genome-wide scale. How do genetic background and environment jointly shape phenotypic traits or causes diseases? How are genetic and external factors integrated at different molecular layers, and how variable are these molecular readouts between individual cells?

We use statistics as our main tool to answer these questions. To make accurate inferences from high-dimensional 'omics datasets, it is essential to account for biological and technical noise and to propagate evidence strength between different steps in the analysis. To address these needs, we develop statistical analysis methods in the areas of gene regulation, genome wide association studies (GWAS) and causal reasoning in molecular systems.

Our methodological work ties in with experimental collaborations and we are actively developing methods to fully exploit large-scale datasets that are obtained using the most recent technologies. In doing so, we derive computational methods to dissect phenotypic variability at the level of the transcriptome and the proteome and we derive new tools for single-cell biology.

Future plans

We will continue to develop innovative statistical approaches to analyze data from high-throughput genetic and molecular profiling studies. We are particularly interested in following up our recent efforts to model single-cell variation data. A major challenge in this area will be the integration of multiple modalities in single-cell genomics, for example linking single-cell epigenome variation with single-cell RNA-Seq. We are particularly interested in applying these methods to data from the Human Induced Pluripotent Stem Cell Initiative (HipSci), in which we are a partner.
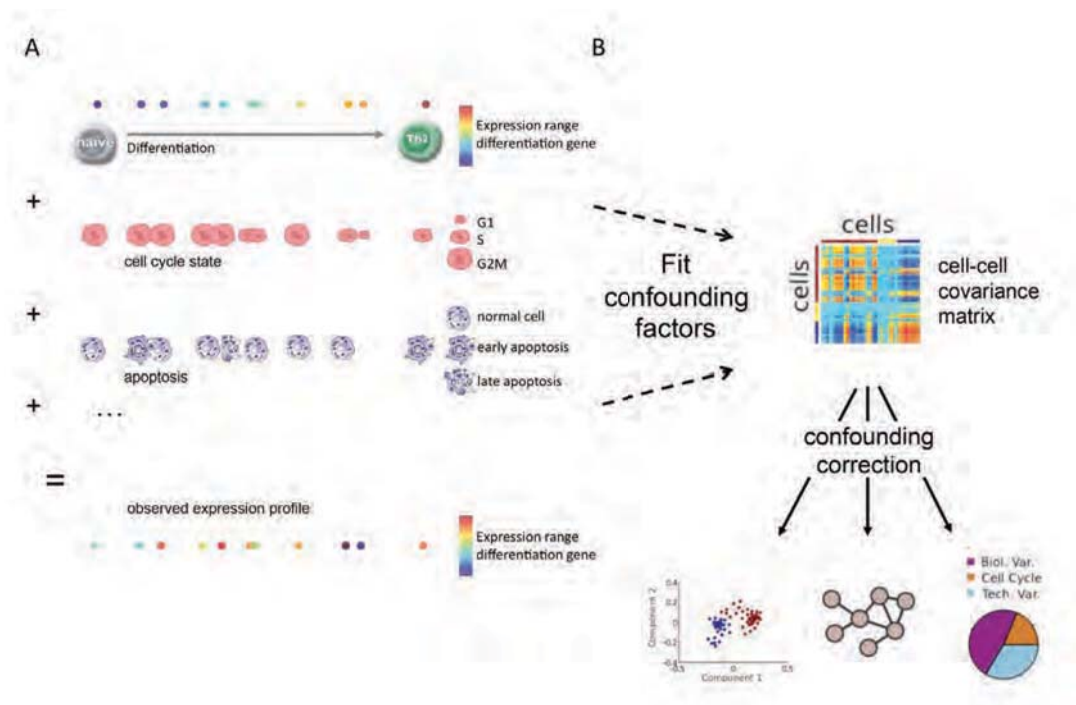
Illustration of statistical methodology to dissect transcriptional heterogenetiy in single-cell RNA-Seq datasets (adapted from Buettner *et al.* 2015).
Left: underlying source of variation in single-cell transcriptome data.
Right: Illustration of our scLVM approach to identify and account for such factors.

Cheminformatics and metabolism research

Christoph Steinbeck

PhD Rheinische Friedrich-Wilhelm-Universität, Bonn, 1995.

Postdoc at Tufts University, Boston, 1996-1997.

Group leader, Max Planck Institute of Chemical Ecology, Jena, 1997-2002.

Group leader, Cologne University 2002-2007.

Lecturer in Cheminformatics, University of Tübingen, 2007.

At EMBL-EBI since 2008.

SELECTED REFERENCES

Rocca-Serra P, *et al.* (2015) Data standards can boost metabolomics research, and if there is a will, there is a way. *Metabolomics* 12: 1, 1–13

Hastings J, *et al.* (2015) ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Research* 44, gkv1031

Salek RM, *et al.* (2015) COordination of Standards in MetabOlomicS (COSMOS): facilitating integrated metabolomics data access. *Metabolomics* 11:6, 1–11

Our team works on methods to decipher, organise and disseminate information about the metabolism of organisms. We develop and maintain MetaboLights, a metabolomics reference database and archive, and ChEBI, the database and ontology of chemical entities of biological interest. We also develop algorithms to: process chemical information; predict metabolomes based on genomic and other information; determine the structure of metabolites by stochastic screening of large candidate spaces; and enable the identification of molecules with desired properties. This requires algorithms based on machine learning and other statistical methods for the prediction of spectroscopic and other physicochemical properties for chemicals represented in chemical graphs.

Our research is dedicated to the elucidation of metabolomes, Computer-Assisted Structure Elucidation (CASE), the reconstruction of metabolic networks, biomedical and biochemical ontologies and algorithm development in cheminformatics and bioinformatics. The chemical diversity of the metabolome and a lack of accepted reporting standards currently make analysis challenging and time-consuming. Part of our research comprises the development and implementation of methods to analyse spectroscopic data in metabolomics.

The Steinbeck group studies small molecule metabolism in biological systems, including reconstruction of metabolic networks from genomic data, and structure elucidation and identification of metabolites.

Future plans

Our central theme of research is efficient methods and algorithms for the assembly, analysis and dissemination of information on small molecules of relevance for biological systems. This includes information about primary and secondary metabolites, and also on xenobiotics and other molecules of relevance, such as epitopes. To this end, we will continue our work in various related areas of ontology development, research on the computational representation of related data, inference of metabolomes from all types of available information, processing of metabolic and metabolomics information, and reconstruction of metabolic networks. We select these projects with an emphasis on applicability in our service foci. Here, our focus is on the extension of the ChEBI database towards greater usability for metabolism and natural products research, and the extension and establishment of our metabolomics database, MetaboLights. MetaboLights will be further enriched with more curated knowledge, such as more reference spectra, pathways, protocols and references to a larger number of existing resources. New online data analysis capabilities will strengthen MetaboLights position as an important research tool for the metabolomics community.

In 2015 we were successful in securing the Horizon 2020 EC grant – PhenoMeNal – with 13 Partners and funded with €8M. The PhenoMeNal project (http://phenomenal-h2020.eu/) aims to develop and deploy an integrated, secure, permanent, on-demand, service-driven, privacy-compliant, and sustainable e-infrastructure for the processing, analysis, and information-mining of the massive amount of medical molecular phenotyping and genotyping data that will be generated by metabolomics applications now entering research and the clinic.

Growth of data, by platform
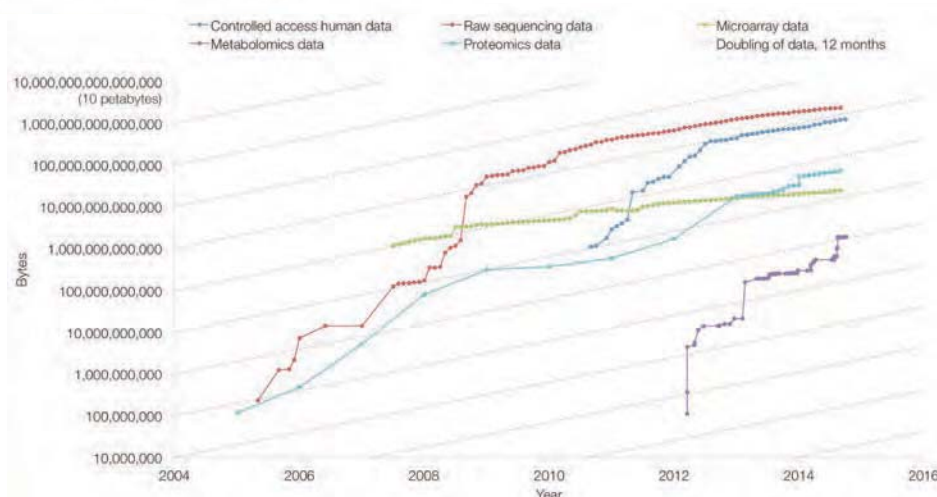
Metabolomics data is doubling faster than any other data platform at EMBL-EBI.

Proteins: structure, function and evolution

Janet Thornton

PhD King's College & NIMR, London, 1973.

Postdoctoral research, University of Oxford, NIMR & Birkbeck College, London. Lecturer, Birkbeck College, 1983-1989.

Professor of Biomolecular Structure, UCL since 1990.

Bernal Professor, Birkbeck College, 1996-2002.

Director of the Centre for Structural Biology at Birkbeck College and UCL, 1998-2001.

Director of EMBL-EBI, 2001 to 2015.

SELECTED REFERENCES

Cuesta SM, Rahman SA, Thornton JM (2016) Exploring the chemistry and evolution of the isomerases. *Proc. Nat. Acad. Sci.* 113, 1796-801

Furnham N, *et al.* (2015) Large-scale analysis exploring evolution of catalytic machineries and mechanisms in enzyme superfamilies. *J. Mol. Biol.* 428, 253-67

Tyagi N, *et al.* (2015) Comparisons of allergenic and Metazoan parasite proteins: allergy the price of immunity. *PLoS Comput Biol* 11, e1004546

Ivanov DK, *et al.* (2015) Longevity GWAS using the Drosophila genetic reference panel. *J Gerontol Series A: Biol Sci Med Sci*, glv047

The Thornton group aims to learn more about the 3D structure and evolution of proteins, or example by studying how enzymes perform catalysis, and how the insulin signalling pathway affects ageing.

The goal of our research is to understand more about how biology works at the molecular level, with a particular focus on proteins and their 3D structure and evolution. We use computational techniques to gain a theoretical understanding of how enzymes perform catalysis. We gather relevant data from the literature and develop novel software tools, which allow us to characterise enzyme mechanisms and navigate the catalytic and substrate space. In parallel, we investigate the evolution of these enzymes to discover how they can evolve new mechanisms and specificities. This involves integrating heterogeneous data with phylogenetic relationships within protein families, which are based on protein structure classification data derived by colleagues at University College London (UCL). The practical goal of this research is to improve the prediction of function from sequence and structure and to enable the design of new proteins or small molecules with novel functions.

We also explore sequence variation between individuals in different contexts and for different species. To understand more about the molecular basis of ageing in different organisms, we participate in a strong collaboration with experimental biologists at UCL. Our role is to analyse functional genomics data from flies, worms and mice and, by developing new software tools, relate these observations to effects on lifespan.

Future plans

Our work on understanding enzymes and their mechanisms using structural and chemical information will include a study of how enzymes, their families and pathways have evolved. We will study sequence variation in different individuals, including humans, flies and bacteria, and explore how genetic variations impact on the structure and function of a protein and sometimes cause disease. We will seek to gain a better understanding of reaction space and its impact on pathways, and to use this new knowledge to improve chemistry queries across our databases. Using evolutionary approaches, we hope to improve our prediction of protein function from sequence and structure. We will also improve our analyses of survival curves and combine data with network analysis for flies, worms and mice in order to compare the different pathways and ultimately explore effects related to human variation and age. This information should be of great use to mankind, although, as with any knowledge, will require careful consideration to avoid any possibility of abuse.

An overview of observed changes of enzyme function during evolution. The change of function data are extracted from the literature and from protein databases. Each circle represents one of the six primary 'classes' of enzymes. The width of the coloured ribbons show the frequency of changes from one class to another. For example, lyases most frequently evolve into tranferases. It is notable that all possible changes of class are observed to occur.
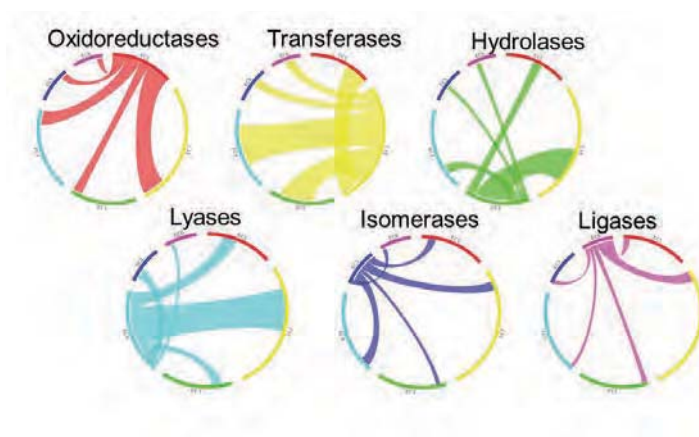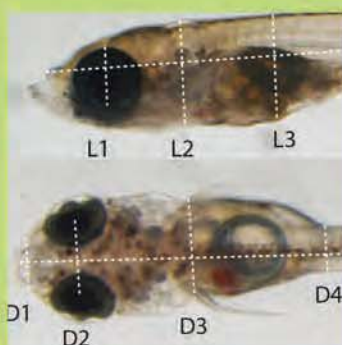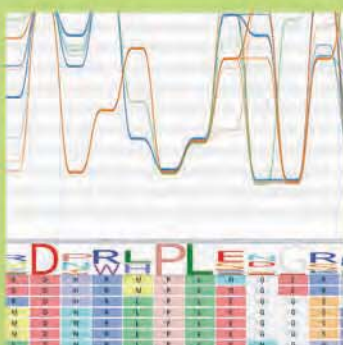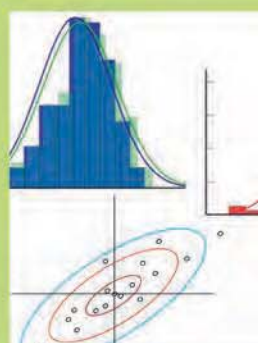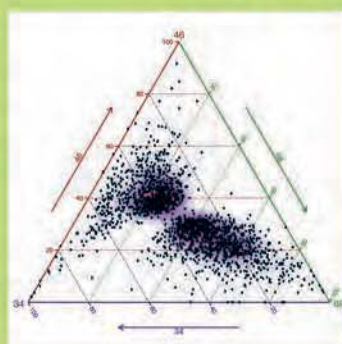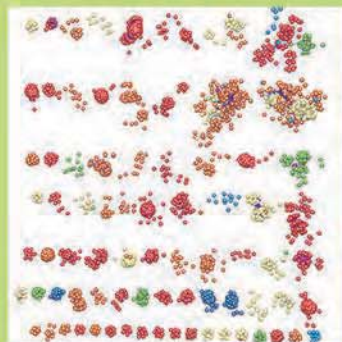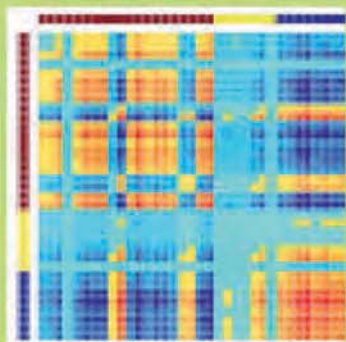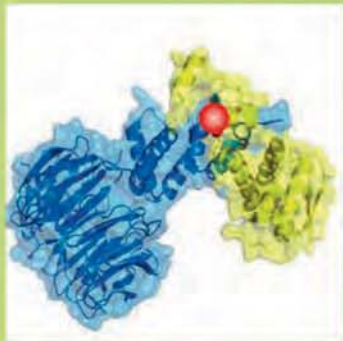
Bioinformatics Services

Service teams at EMBL-EBI focus on gathering, adding value to and presenting important collections of biological and chemical information for the benefit of the larger research community. Their work is enhanced by the input of basic researchers, some of whom are embedded in service teams. Master's students and other visiting scientists who conduct their research at EMBL-EBI have the opportunity to work in a unique environment, exploring how we can use an incredibly diverse range of information to understand life on a fundamental level.

Building on more than 20 years of experience in bioinformatics, EMBL-EBI maintains the world's most comprehensive range of molecular databases. We are the European node for globally coordinated efforts to collect and disseminate biological data. Many of our databases are part of the daily toolkit of biologists all over the world, for example the European Nucleotide Archive; the Ensembl genome explorer; the Gene Expression Atlas; UniProt, the universal protein resource; InterPro and the Protein Data Bank in Europe. IntAct (protein–protein interactions), Reactome (pathways), ChEBI and ChEMBL (small molecules), help researchers understand not only the molecular parts that go towards constructing an organism, but how these parts combine to create systems.

The details of each database vary, but they all uphold the same principles of service provision: accessibility, compatibility, comprehensive datasets, portability, and quality.

European Nucleotide Archive

Guy Cochrane

PhD University of East Anglia, 1999.

At EMBL-EBI since 2002.

Team leader since 2009.

The European Nucleotide Archive (ENA) team provides globally comprehensive primary data repositories for nucleotide sequencing information. ENA content spans the spectrum of data: from raw sequence reads through assembly to functional annotation of assembled sequences and genomes. Our team provides interactive and programmatic submission tools as well as curation support.

The ENA offers a broad palette of services over the web, via a powerful programmatic interface. Reflecting the centrality of nucleotide sequencing in the life sciences and the emerging importance of the technologies in applied areas such as healthcare, environmental and food sciences, ENA data and services form a core foundation upon which scientific understanding of biological systems has been assembled. With ongoing focus on data presentation, integration within ENA, integration with resources external to ENA, tools provision and services development, the team's commitment is to the utility of ENA content and achieving the broadest reach of sequencing applications.

In addition to the ENA resource itself, the team also provides core sequence archiving and compression technology, content, and services that are used in the construction and delivery of many other resources, including the European Genome-phenome Archive, the Metagenomics Portal, ArrayExpress and the UniProt Knowledgebase.

The ENA team provides globally comprehensive primary data repositories for nucleotide sequencing information. ENA content spans the full spectrum of molecular data.

Protein sequences, families and motifs

Rob Finn

PhD in Biochemistry, Imperial College London; Wellcome Trust Sanger Institute, 2001-2010; Janelia Research Campus, 2010-2013.

At EMBL-EBI since 2014.

The Protein Families team is responsible for the InterPro, Pfam and Rfam data resources, and coordinates the EBI Metagenomics project. InterPro integrates protein family data from 11 major sources, hierarchically classifying the different protein family, domain and functional site definitions to provide a unified view of the diverse data. A tool called InterProScan allows the identification of InterPro entries on protein sequences. Pfam, a member database of InterPro, generates new protein family entries and has the largest sequence coverage of any of the InterPro member databases. Both InterPro and Pfam have a number of important applications, including the automatic annotation of proteins for UniProtKB/TrEMBL and genome annotation projects. Rfam, a major contributor to RNAcentral, classifies non-coding RNA sequences into families using probabilistic models that take into account both sequence and secondary structure information, termed covariance models (CMs). Rfam is uniquely placed to annotate non-coding RNAs in genome projects.

Metagenomics is the study of the sum of genetic material found in an environmental sample or host species, typically using next-generation sequencing (NGS) technology. EBI Metagenomics enables researchers to submit sequence data and associated descriptive metadata to public nucleotide archives. Once deposited, our team helps ensure the data is functionally analysed using an InterPro-based pipeline, taxonomically analysed using the QIIME software package, and that the results can be visualised and downloaded via a web interface.

Molecular networks

Henning Hermjakob

Dipl. Inf (MSc) in Bioinformatics, University of Bielefeld, 1996.

Research assistant at the National Centre for Biotechnology (GBF), Braunschweig, Transfac Database team.

At EMBL-EBI since 1997.

The team develops tools and resources for the representation, deposition, distribution and analysis of proteomics and systems biology data.

The Hermjakob Team develops tools and resources for the representation, deposition, distribution and analysis of pathway (http://www.reactome.org) and systems biology (http://www.ebi.ac.uk/biomodels) data. We follow an open-source, open-data approach and are a major contributor to community standards, in particular the Proteomics Standards Initiative (PSI) of the international Human Proteome Organization (HUPO), and COMBINE systems biology standards. As a result of long-term engagement with the community, journal editors and funding organisations, data deposition in standards-compliant data resources such as IntAct, PRIDE, and BioModels is becoming a strongly recommended part of the publishing process. This has resulted in a rapid increase in the data content of public proteomics resources. Our curation teams ensure consistency and appropriate annotation of all data, whether from direct depositions or literature curation, to provide the community with high-quality reference datasets.

The Hermjakob team also contributes to the development of data integration approaches like BioJS visualisation widgets, semantic web technologies, dataset discovery (http://www.ebi.ac.uk/Tools/omicsDI), and provision of stable identifiers for biomolecular entities through identifiers.org. We offer a range of projects for interested Master's students.

Ensembl Genomes

Paul Kersey

PhD University of Edinburgh 1995.

At EMBL since 1999.

Team leader since 2008.

The non-vertebrate genomics group develops Ensembl Genomes and related resources.

The Non-vertebrate Genomics team develops Ensembl Genomes and related resources. Ensembl Genomes provides portals for bacteria, protists, fungi, plants and invertebrate metazoa, offering access to genome-scale data through a set of programmatic and interactive interfaces shared with the vertebrate-focused Ensembl project. Our major areas of interest include broad-range comparative genomics and the visualisation and interpretation of genomic variation, which is increasingly being studied in species throughout the taxonomy. Even small communities with little informatics infrastructure can now perform highly complex and data generative experiments that were once the sole domain of large, internationally coordinated sequencing projects. Through collaborating with EMBL-EBI and re-using our established toolset, such small communities can store, analyse and disseminate data more cheaply and powerfully than if they develop their own tools.

We collaborate closely with VectorBase, a resource focused on the annotation of invertebrate vectors; WormBase, a resource for nematode biology; and PomBase, focused on the fission yeast Schizosaccharomyces pombe. Plant science collaborations include Gramene in the US and several European groups in the transPLANT project. We have developed PhytoPath, a new portal for plant pathogen data, and are are currently working on the annotation of diverse genomes such as the biting midge, bread wheat and cassava whitefly.

UniProt Development

Maria J. Martin

BSc in Veterinary Medicine, Universidad Complutense, Madrid, 1990.

PhD in Molecular Biology (Bioinformatics), Universidad Autonoma, Madrid, 2003.

At EMBL-EBI since 1996.

Technical team leader since 2009.

The UniProt Development team provides the bioinformatics infrastructure for the databases and services of the Universal Protein Resource (UniProt). The team comprises software engineers and bioinformaticians who are responsible for the UniProt, the Gene Ontology Annotation and the Enzyme portal software and database development, and who study novel automatic methods for protein annotation and representation. Our user experience analyst coordinates the user request gathering process, which informs the design and development of the web site. The team is also responsible for the maintenance and development of tools for UniProt curation. We work in a fully complementary fashion with Claire O'Donovan's UniProt Content team to provide essential resources for the biological community, as the databases have become an integral part of the tools researchers use on a daily basis for their work.

The team provides bioinformatics infrastructure for the databases and services of the Universal Protein Resource.

Literature Services

Johanna McEntyre

PhD in Plant Biotechnology, Manchester Metropolitan University, 1990.

Editor, Trends in Biochemical Sciences, Elsevier, Cambridge, UK, 1990-1997.

Staff scientist, NCBI, National Library of Medicine, NIH, USA, 1997-2009.

At EMBL-EBI since May 2009.

Direct access to the scientific literature and the data that underlie it have become increasingly important as data-driven science continues to trend upwards. The Literature Services team addresses this in a number of ways, supporting the wider scientific research community and our data-provider colleagues at EMBL-EBI by providing valuable, multi-layer functionality in Europe PMC.

Europe PubMed Central, EMBL-EBI's literature resource, contains more than 30 million abstracts and 3.5 million full text articles. Abstracts include all of PubMed, agricultural abstracts from Agricola and patents from the European Patent Office. About 1.2 million of the full-text articles are open access, so they are free to read and to reuse in ways such as text mining. Europe PMC is supported by 27 European funding organisations, whose commitment supports their own Open Access mandates.

Our goal is to provide fast and powerful access to the literature, as well as features and tools that place the article narratives in the wider context of related data and credit systems that combine ORCID portfolios and article metrics. One of the key approaches we employ to meet this goal is to engage with individual scientists, text miners and database managers to understand how layers of value can be built upon the basic article content. We provide the infrastructure that enables the enrichment of the literature by individuals and computational methods developed throughout the community, and publish the results to maximise the usefulness of the core content and allow their widest possible reuse.

The Literature services team provides valuable, multi-layer functionality in Europe PMC.

UniProt content

Claire O'Donovan

BSc (Hons) in Biochemistry, 1992, University College Cork, Ireland.

Diploma in Computer Science, 1993, University College Cork, Ireland.

At EMBL since 1993.

At EMBL-EBI since 1994.

Technical team leader since 2009.

The team supports the development of databases and resources enabling the integration of scientific literature and large data sets.

The UniProt Content Team comprises biocurators working on the Universal Protein Resource (UniProt). Biocuration involves the translation and integration of information relevant to biology into a database or resource that enables integration of the scientific literature as well as large data sets. Accurate and comprehensive representation of biological knowledge, as well as easy access to this data for working scientists and a basis for computational analysis, are the primary goals of biocuration. These are achieved thanks to the convergent endeavors of biocurators and software developers and our team works in a fully complementary fashion with Maria Jesus Martin's UniProt Development group to provide essential resources to the biological community, such that databases have become an integral part of the tools researchers use on a daily basis for their work.

Molecular Interactions

Sandra Orchard

BSc (Hons) Biochemistry, University of Liverpool, 1982.

Team Leader, Inflammatory Diseases Dept, Roche Products Ltd UK, until 2001.

At EMBL-EBI since 2002

The Molecular Interactions team produces the IntAct molecular interaction database, making interaction data publicly available and accessible. IntAct is used to carry out a wide range of tasks, for example enabling network analysis of large-scale datasets, and facilitating the understanding of specific protein-binding interfaces. Our team also produces the Complex Portal, a reference resource for macromolecular complexes. We support the molecular interaction community standards published by the HUPO Proteomics Standards Initiative, and contribute actively to their further development. We develop and maintain tools such as the PSICQUIC web service, which improves data access and usability for diverse users in the research community. We work continually to improve interface usability and to make data more discoverable for our users. We also endeavour to assess the size and breadth of the unexplored aspects of the human interactome, and develop tools based on a novel Java library. We welcome Masters students who are interested in learning more about network analysis, data curation, Java development or the use of JavaScript for data visualisation.

Functional genomics production

Helen Parkinson

PhD in Genetics, 1997.

Research associate in Genetics, University of Leicester, 1997-2000.

At EMBL since 2000.

The Samples, Phenotypes and Ontologies team is organised into three themes: BioSamples and Semantic Data Integration, Mouse Informatics and the Gene Ontology Editorial Office. We provide ontologies, ontology tooling, and resources providing access to samples and ontologies both for EMBL-EBI resources and external users.

We develop international infrastructure for mouse data archiving, integration and dissemination, and are a partner on several international projects that share a theme of data integration of rich phenotypic data derived from models species to better understand gene function and how genetic variation contributes to disease.

Our team delivers the Gene Ontology, including adding terms and revising the Ontology's content and structure, alignment with other ontologies such as chemicals, anatomy and cell type. We also deliver the BioSamples database, a resource which identifies and describes samples across EMBL-EBI databases, and develop ontologies such as the Experimental Factor Ontology, EDAM, SWO and the Cell Line Ontology. We also deliver ontology tooling including the Ontology Lookup Service, which helps users build, browse and query ontologies. In collaboration with Paul Flicek's team, we also develop the NHGRI GWAS Catalog content and infrastructure.

We develop open-source software tools for managing data, developing and integrating ontologies and data, and work with semantic web technologies. We collaborate closely with Melissa Haendel and Chris Mungall of the Monarch Initiative, an integration project for genotype/phenotype data, by supplying human and mouse phenotypic data and using their tools in delivery of the IMPC portal.

The team focuses on metadata integration, ontology development and supporting tooling, development and delivery of content for EMBL-EBI's BioSample database and mouse data for the biomedical research community.

Functional genomics software development

Ugis Sarkans

PhD in Computer Science, University of Latvia, 1998.

Postdoctoral research at the University of Wales, Aberystwyth, 2000.

At EMBL-EBI since 2000.

Our team develops software for ArrayExpress, a database for functional genomics data, and the BioStudies database, a resource for biological datasets that do not have a dedicated home within EMBL-EBI services. As of April 2016, ArrayExpress holds data from almost two million assays such as microarray hybridisations and sequencing runs, and is one of the major data resources of EMBL-EBI. We are building and maintaining components of the ArrayExpress infrastructure, including internal data management tools, the data submission tool Annotare, and data access interfaces.

The BioStudies database was released in 2015. We are collaborating with the Literature team and load into BioStudies supplementary materials associated with publications, demonstrating the concept of a "data package" attached to a manuscript, and paving the way to tighter collaboration with publishers for the management of supplementary information. We also manage multi-omics datasets in BioStudies, for example, the "diXa" toxicogenomics data collection. Another purpose of BioStudies is dealing with types of data that are new for EMBL-EBI, such as imaging data.

Our team participates in several projects in a data management role. We believe that being close to large consortia that generate different types of high-throughput data places us in a better position to fulfil our main objective of developing the BioStudies infrastructure.

The team develops software for ArrayExpress and the BioStudies database.

Protein Data Bank in Europe

Sameer Velankar

PhD, Indian Institute of Science, 1997.

Postdoctoral researcher, Oxford University, UK, 1997-2000.

At EMBL-EBI since 2000.

The team accept and annotate world-wide depositions of biomacromolecular structures.

The Protein Data Bank in Europe (PDBe) manages the worldwide biomacromolecular structure archive, the Protein Data Bank (PDB) and is a founding member of the Worldwide Protein Data Bank (wwPDB). We accept and annotate worldwide depositions of biomacromolecular structures determined using X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy, 3D Electron Microscopy (EM) and other structure determination methods. PDBe is also a founding member of EMDataBank, which manages the deposition and annotation of electron microscopy data in EMDB.

Our goal is to ensure that PDBe truly serves the needs of the biomedical community. As part of that effort, we are constantly improving the web interface for existing tools and services and designing new tools to make structural data available to all. In the context of the SIFTS project, we integrate structural data with other biological data in the interests of facilitating discovery. These integrated data form the basis for many query interfaces that allow macromolecular structure data to be presented in its biological context. Our specific focus areas are: data integrity, data quality, integration and data dissemination to the non-expert biomedical community.
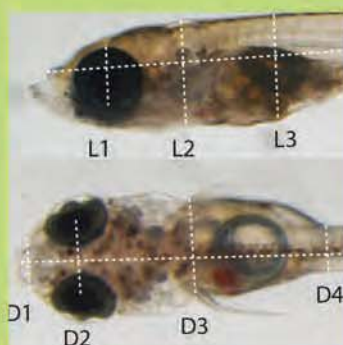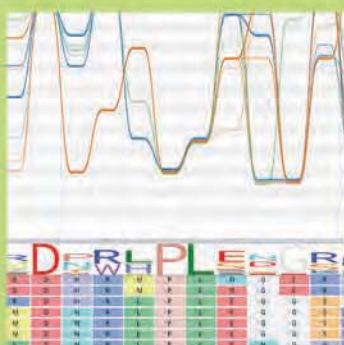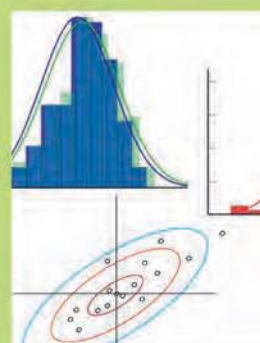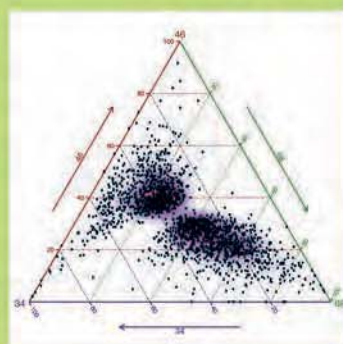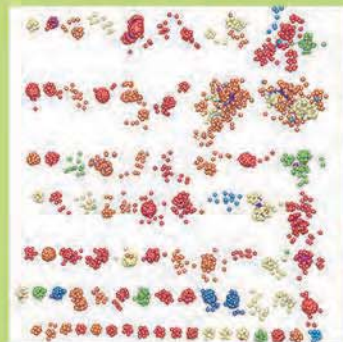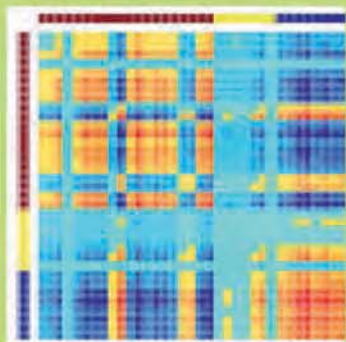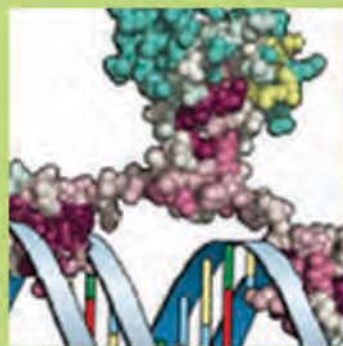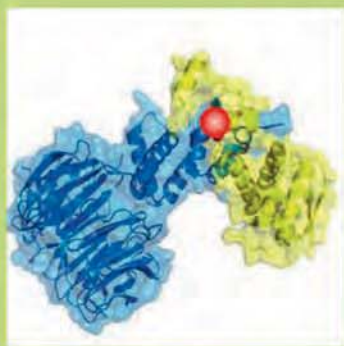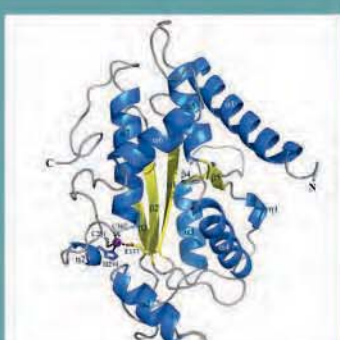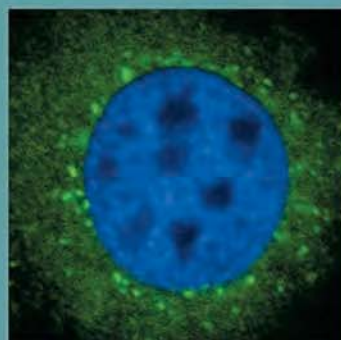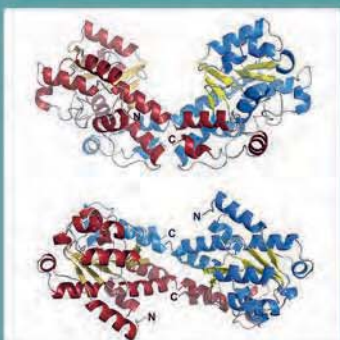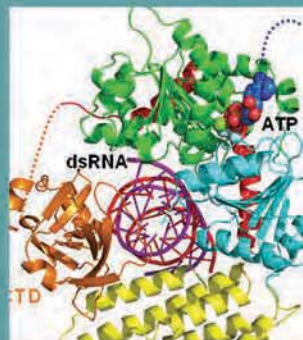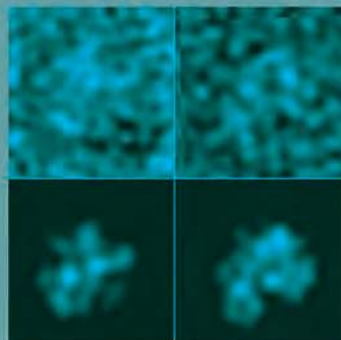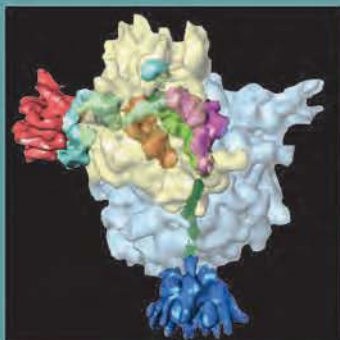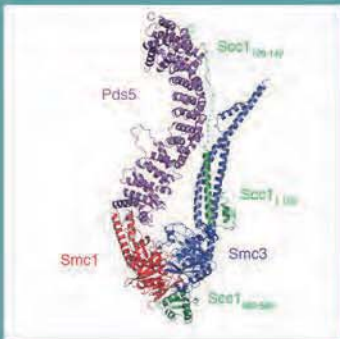
Structural Biology

EMBL Grenoble is a key player in the Partnership for Structural Biology (PSB) through which the Grenoble European Photon and Neutron (EPN) Campus institutes – EMBL, the European Synchrotron Radiation Facility (ESRF), the Institut Laue-Langevin, and the Institut de Biologie Structurale (IBS) – provide a uniquely comprehensive range of structural biology platforms for sample production, sample characterisation and structure determination for both in house research and external users.

A cornerstone of the Outstation's activities is the close interaction with the world-leading ESRF, which produces ultra-intense X–ray beams. EMBL staff collaborate with the ESRF in building and operating state-of-the-art X-ray beamlines, developing associated instrumentation and techniques, and providing expert help to visitors. Thanks to the ESRF Phase I upgrade programme, a new suite of highly sophisticated beamlines is now available for protein crystallography, with fast and automated data collection, and for small-angle X-ray scattering, including a high-throughput sample changer and online high performance liquid chromatography.

High throughput methods have also been introduced in upstream steps of the structure determination process. These include a highly successful robotic platform for nano-volume crystallisation run by the Outstation as well as a recently developed system called CrystalDirect, which enables automated crystal mounting and cryo-cooling. The combination of these technologies with the ESRF massively automated X-ray beamlines has enabled the development of a new generation of fully automated, remote controlled pipelines from protein to X-ray data, streamlining the structural analysis of complex biological targets. They also facilitate structure-guided drug design, through automated facilities for ligand and fragment screening. These platforms are now available to external users under the EU funded iNEXT project. Furthermore, the Outstation has a state-of-the-art Eukaryotic Expression Facility, which features expression of multiprotein complexes in insect cells using the MultiBac technology.

Concerning scientific research, the Outstation focusses on the structural biology of eukaryotic complexes, with a strong tradition in the study of systems involving protein-nucleic acid complexes and viruses. Structural work on aminoacyl-tRNA synthetases is particularly well known and has recently focussed on elucidation of the mode of action of novel boron-containing antibiotics, which

target leucyl-tRNA synthetase. Other projects involving protein-RNA interactions include cryo-EM studies of ribosome with the signal recognition particle and translocon and other proteins and complexes involved in RNA processing, transport and degradation, such as the nonsense-mediated decay (NMD) pathway. Further important areas include the analysis of mechanisms of transcriptional regulation, including at the epigenetic level. Groups are working on the structural analysis of eukaryotic transcription factor, chromatin-modification and nucleosome assembly complexes as well as elucidating the mechanisms by which piRNAs (small non-coding RNAs) protect the genome and trying to uncover the role of long non-coding RNAs (lncRNAs). Another focus is the study of segmented RNA viruses, particularly influenza and bunyaviruses, with the aim of understanding how their polymerases replicate and transcribe the viral genome. Complementary to this are studies on the innate immune receptors, which detect the presence of viral RNA in infected cells and activate interferon production.

Scientists at EMBL Grenoble have, via the PSB, access to a wide range of techniques, including molecular biology and biophysical instrumentation, negative stain and cryo-electron microscopy – including a top-end Polara electron microscope with K2 direct detector, and an F20 and T12 –, isotope labelling, nuclear magnetic resonance, neutron scattering, X-ray crystallography and small angle scattering. A confocal microscope with facilities for cross-correlation spectroscopy is available for live cell imaging.

Stephen Cusack
Head of EMBL Grenoble

Structural complexomics of eukaryotic gene expression

Imre Berger

PhD 1995, MIT Cambridge and Leibniz University, Hannover.

Postdoctoral research at MIT and the IMB, ETH Zürich. Habilitation 2005, ETH.

Group leader at IMB since 2005.

Group leader at EMBL Grenoble since 2007.

Professor, School of Biochemistry, University of Bristol UK since 2014.

Wellcome Trust Senior Investigator

SELECTED REFERENCES

Trowitzsch S, *et al.* (2015) Cytoplasmic TAF2-TAF8-TAF10 complex provides evidence for nuclear holo-TFIID assembly from preformed submodules. *Nat Commun* 6, 6011

Reich S, *et al.* (2014) Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* 516, 361–6

Bieniossek C, *et al.* (2013) The architecture of human general transcription factor TFIID core complex. *Nature* 493, 699-702

Imasaki T, *et al.* (2011) Architecture of the Mediator head module. *Nature* 475, 240-3

The Berger group studies eukaryotic multiprotein assemblies in transcription regulation, develops technologies to produce them recombinantly and subjects them to high-resolution structural and functional analyses.

Human gene transcription requires the controlled step-wise assembly of the pre-initiation complex (PIC), comprising a large ensemble of proteins and protein complexes including RNA Pol II and the general transcription factors. TFIID is the first general transcription factor to bind to gene promoters, and is a cornerstone of PIC assembly. However important, we lack detailed knowledge of its molecular architecture and interactions with cellular factors. Endogenous TFIID is scarce and heterogeneous. Therefore, we created new technologies to produce TFIID and similar multiprotein complexes recombinantly. Notably, our MultiBac system – a modular, baculovirus-based technology specifically designed for eukaryotic multi-protein expression – is now used in many labs worldwide. Recently, we determined the architecture of the 700 kDa heterodecameric human TFIID core complex by combining MultiBac-based production with cryo-EM, X-ray crystal analysis, homology modelling, and proteomics data.

We collaborate with groups from academia and industry to further automate labour-intensive steps in the multiprotein complex structure determination process, and have harnessed homologous and site-specific recombination methods for assembling multigene expression plasmids. We have developped ACEMBL, a proprietary automated suite for multigene recombination on our TECAN EvoII platform. It allowed us to produce numerous large multiprotein assemblies for structural studies, and to expand our multiprotein expression strategies to prokaryotic and mammalian hosts.

Future plans

We work towards entirely automating and miniaturising the production-process for eukaryotic multiprotein complexes including the entire human TFIID holocomplex, its various isoforms and other components of the preinitiation complex. In collaboration with the Schaffitzel Team and the Schultz Group (IGBMC Strasbourg), we subject the complex specimens produced to electron microscopic and X-ray crystallography analyses to understand their physiological function, and further our findings by *in vitro* and *in vivo* biochemical analysis.

Using state-of-the-art mass spectrometric methods we are developing MultiTRAQ, a new technology to address the challenge of defining crystallisable core assemblies of multiprotein complexes in a reasonable time frame. Another recent project line in our lab exploits synthetic biology techniques for genome engineering, with the aim of creating disruptive platforms for recombinant protein production, for both academic and industrial applications.

Recently, we implemented ComplexLink, our polyprotein-based technology to enable production of hitherto inaccessible protein complexes. We are applying ComplexLink to high-value targets such as viral polymerases and multisubunit kinases.
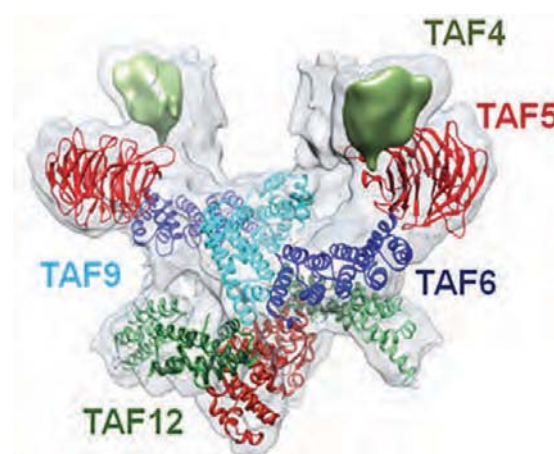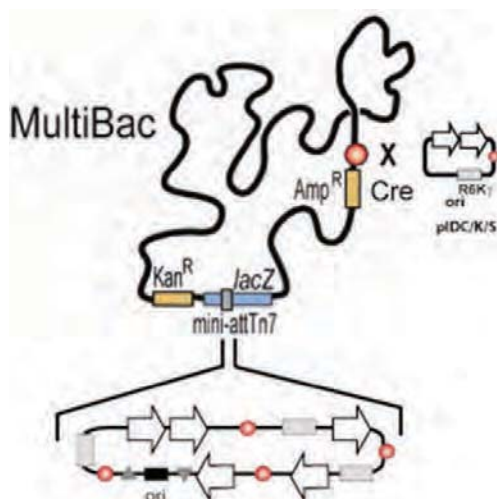
We develop and utilise advanced, automated technologies to produce eukaryotic multiprotein complexes for structural and functional analysis by a variety of methods, including X-ray crystallography.

Diffraction Instrumentation Team

Florent Cipriani

BSc 1974, Physics, University of Grenoble, France.

Senior engineer in nuclear and medical industries.

At EMBL Grenoble since 1991.

Team Leader since 2003.

Senior Scientist since 2011.

SELECTED REFERENCES

Bowler, M., *et al.* (2015). Automation and experience of controlled crystal dehydration: results from the European synchrotron HC1 Collaboration. *Cryst. Growth Des.* 15, 1043–54

Cipriani, F., *et al.* (2012). CrystalDirect: a new method for automated crystal harvesting based on laser-induced photoablation of thin films. *Acta Crystallogr. D Biol. Crystallogr.* 68, 1393-9

Pernot, P., *et al.* (2010). New beamline dedicated to solution scattering from biological macromolecules at the ESRF. *Journal of Physics, Conf. Ser.* 247

Sanchez-Weatherby, J., *et al.* (2009). Improving diffraction by humidity control: a novel device compatible with X-ray beamlines. *Acta Crystallogr. D Biol. Crystallogr.* 65, 1237-46

The core activity of our team is to develop instruments and methods for X-ray scattering experiments. The CrystalDirect™ harvester jointly conceived with the Márquez team is now in routine use at the High Throughput Crystallisation lab: crystals grown in CrystalDirect plates can be marked remotely using a web service and automatically mounted onto Spine pins. The machine is particularly suitable for harvesting micro-crystals, fragile crystals, as well as batches of micro-crystals for serial data collection.

A second harvester is being coupled with a FlexED3 cryo-storage system based on our Flex sample changer technology. With an initial capacity of three pucks, the system will perform up to 50 harvestings in less than an hour with minimal user intervention. The intrinsic low background of the CrystalDirect plates makes them ideal for *in situ* X-ray data collection, such as like that offered at the EMBL/ESRF BM14 beamline, the ESRF ID30B beamline, and the EMBL@PETRA-III P14 micro-focus beamline.

Evaluation kits were distributed to assess the design of the new miniSpine sample holder proposed as possible future European standard (Figure 1). The FlexED8 Edge Dewar sample changer installed on BM14 opens the beamline to UniPucks and is compatible with miniSpine (Figure 2). Additional test beamlines will be equipped with miniSpine compatible robotics, involving industrial partners. ID30B is equipped with a MD2s Diffractometer and a FlexHCD sample changer, and can process crystals *in situ* or frozen stored in SC3 pucks or in UniPucks. All these developments include contributions from the Márquez and McCarthy teams, the Schneider group, and the ESRF Structural Biology group. Supported by EMBLEM, most of our instruments are available to the scientific community worldwide.

Future plans

Next year a fully automated harvesting service based on the new CrystalDirect harvester and FlexED3 sample cryo-storage unit will be available at the HTX lab. A harvesting/data collection pipeline will be evaluated at BM14 using a harvester coupled to the FlexED8 sample changer.

Our main future challenge will be to exploit the new beam characteristics of the future ESRF-EBS MX beamlines. Sample delivery will be a central question and serial crystallography the rule. To prepare for this we will develop a new MD3 diffractometer optimised for high-speed rastered data collection as part of the ESRF ID23-2 micro-focus upgrade. We will also duplicate the FlexHCD sample changers to provide flexible robotics for all the MX beamlines.

The Cipriani team develops instruments and methods for X-ray scattering experiments and works with industry to make them available to scientists worldwide.

Figure 1: FlexED8 8 pucks EdgeDewar sample changer on BM14.

Figure 2: MiniSpine 36 slots puck and MiniSpine sample holder in a robot gripper
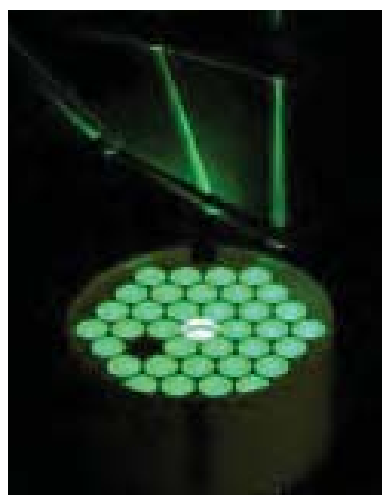
Structural biology of RNA-protein complexes in gene expression and host-pathogen interactions

Stephen Cusack

PhD 1976, Imperial College, London, UK.

Postdoctoral work at EMBL Grenoble.

Staff scientist at EMBL Grenoble 1980-1989.

Group Leader, Senior Scientist and Head of Outstation since 1989.

ERC Advanced Investigator

Fellow of the Royal Society.

SELECTED REFERENCES

Gerlach P, *et al.* (2015) Structural insights into bunyavirus replication and its regulation by the vRNA promoter. *Cell.* 16, 1267-79

Pflug A, *et al.* (2014) Structure of influenza A polymerase bound to the viral RNA promoter. *Nature*, 516, 355-60

Reich S, *et al.* (2014) Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* 516, 361-6

Kowalinski E, *et al.* (2011). Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell* 147, 423-35

The Cusack group uses X-ray crystallography to study the structural biology of protein-RNA complexes involved in RNA virus replication, innate immunity and cellular RNA metabolism.

We study the molecular mechanisms whereby the RNA of viruses such as influenza is, on the one hand, the template for transcription and replication of the viral genome by its RNA-dependent RNA polymerase and, on the other hand, an Achilles' heel, whose recognition as non-self can trigger an innate immune response to counter the viral infection. Molecular warfare between the virus and the host-cell occurs at many levels. Influenza has a unique mechanism of transcription priming called 'cap-snatching', which involves pirating short-capped oligomers from nascent cellular Pol II transcripts; this leads to shut-down of host cell gene expression. The cell counters RNA viruses with innate immune pattern-recognition receptors, such as the RNA helicase RIG-I, which recognise particular viral RNA structural motifs as non-self, thus activating a signalling pathway leading to interferon production and establishment of the anti-viral state. In response, viruses deploy proteins as counter-counter-measures to dampen the immune response, for instance, by interfering with the RIG-I signalling pathway.

In 2011 we published the first structure-based mechanism of activation of RIG-I, showing how RNA binding resulted in a major conformational change that liberated the N-terminal CARD domains for downstream signalling. In 2014 we published the first crystal structures of the complete heterotrimeric influenza polymerase and proposed a mechanism of how cap-snatching is performed.

Future plans

Our current goal is to obtain structural snapshots of influenza virus polymerase performing transcription of the viral genome (vRNA) by freezing states corresponding to cap-snatching, transcription initiation, elongation and poly-adenylation. In parallel we will do the same for viral replication, which is unprimed and occurs via an intermediate complementary RNA (cRNA). This will involve understanding how the behaviour of the polymerase bound to vRNA differs from that when bound to cRNA. We complement structural studies with in vitro polymerase enzymology and in-cell studies using mini-replicon systems, recombinant viruses, and live-cell imaging.

There are several other aspects of this project of particular interest. Firstly, study of how the influenza polymerase interacts with host proteins (e.g. transport factors, helicases, and splicing factors), which help it function efficiently in the cellular context, can give insight into the polymerase mutations required for an avian virus to adapt to be able to infect humans. Secondly, our structural work on influenza polymerase has opened up the area of structure-based drug design of novel anti-virals targeting multiple sites on the polymerase. To exploit this we co-founded a Vienna-based company called SAVIRA and the project is now being pursued by Roche in Basel. Thirdly, we have extended this work to polymerases of related segmented negative-strand RNA viruses such as the large family of bunyaviruses, which includes several emerging human pathogens. Finally, we are studying the nuclear cap-binding complex (CBC), which binds to the 5' cap of nascent Pol II transcripts and mediates interaction with nuclear RNA processing and transport machineries. One aim of this is to understand how influenza polymerase can compete with CBC for access to capped RNAs emerging from Pol II.
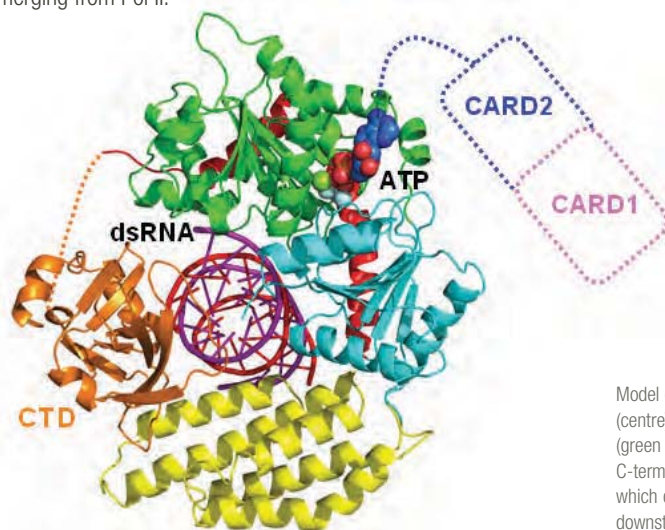
Model of the activated state of RIG-I with bound dsRNA (centre) and ATP (top-right). The helicase domains (green and cyan), the insertion domain (yellow) and the C-terminal domain (gold) all contribute to RNA binding, which displaces the CARD domains thus allowing downstream signalling and interferon expression.

Structure and function of lncRNA-protein complexes involved in transcription regulation

Marco Marcia

PhD 2010, Max Planck Institute of Biophysics and J.W. Goethe Universität, Frankfurt am Main, Germany.

Postdoctoral research at Yale University in New Haven, USA.

Group Leader at EMBL since 2014.

SELECTED REFERENCES

Zhao C, *et al.* (2015) Crystal structure of group II intron domain 1 reveals a template for RNA assembly. *Nat Chem Biol* 12, 967-72

Somarowthu S, *et al.* (2015) HOTAIR forms an intricate and modular secondary structure. *Mol Cell* 58, 353-61

Marcia M & Pyle AM (2012) Visualizing group II intron catalysis through the stages of splicing. *Cell* 151, 497-507

Marcia M, *et al.* (2009) The structure of Aquifex aeolicus sulfide:quinone oxidoreductase, a basis to understand sulfide detoxification and respiration. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9625-30

My group studies the mechanism of lncRNA recognition within nuclear ribonucleoproteins (RNPs) and the molecular bases for their cellular functions. RNPs are involved in several physiological processes, ranging from hormone-signalling to brain function and are thus implicated in severe pathologies, including neurodegenerative and vascular diseases, developmental disorders, and cancer. Despite their crucial importance, little is currently known about the structure and mechanisms of these RNPs.

We use mainly X-ray crystallography, but also electron microscopy, small angle scattering, analytical ultracentrifugation, biochemical techniques, spectroscopy, mass-spectrometry, bioinformatics and functional assays.

Our research expertise stems from my previous work on the structure and function of various classes of macromolecules, ranging from membrane proteins to large RNA enzymes, involved in fundamental metabolic pathways such as electron transport, signalling and splicing. I previously determined the crystal structures of the bacterial sulfide quinone oxidoreductase (SQR), a membrane flavoprotein conserved also in humans, where it is involved in serious neurodegenerative diseases and lethal encephalopathies. I also determined various structures of a self-splicing group II intron, revealing the molecular mechanism behind hydrolytic splicing, including the role of monovalent ions within an unprecedented catalytic metal ion cluster. My results also shed a new light on the more complex human spliceosome.

Currently, my research aims to answer the following questions: how can many thousands different lncRNAs form tight complexes with a relatively limited set of nuclear protein complexes? Which level of selectivity characterises the formation of such complexes? How is selectivity achieved? What structural motifs are involved in recognition? How complex is the structural architecture of the intervening lncRNAs and how is it maintained? How are chromatin-binding ability and enzymatic activity of the intervening proteins regulated by lncRNAs at a molecular level? Such studies will have direct medical implications and potentially lead to the development of new therapeutic approaches to cure some of the most invasive diseases of our modern societies

The Marcia group uses structural biology and biophysical approaches to study the molecular interactions between long non-coding RNAs (lncRNAs) and nuclear proteins and how their complexes regulate gene expression processes.

Future plans

- Identify the recognition motifs that guide formation of tight complexes between lncRNAs and nuclear proteins.
- Determine structures of such ribonucleoproteins.
- Building on structural insights, understand the molecular mechanism by which lncRNAs exert their cellular functions.
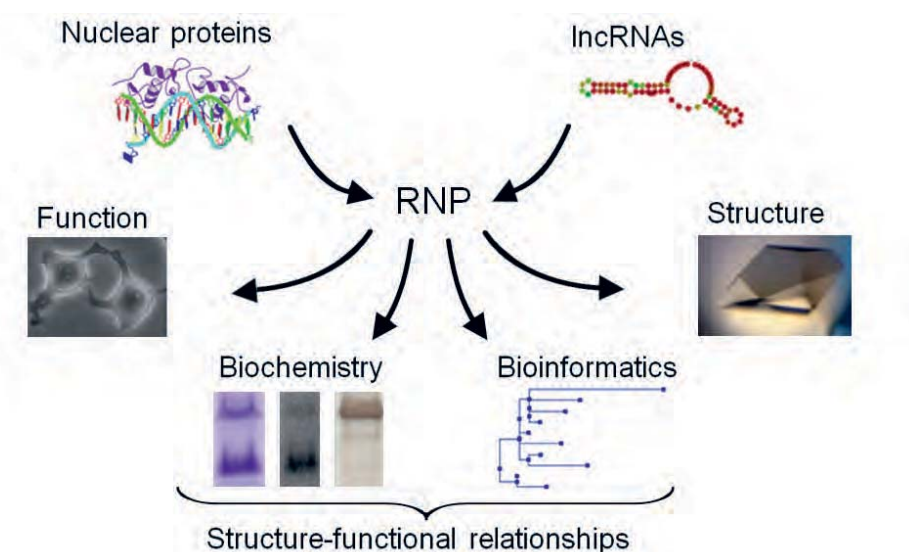
A variety of complementary approaches are used to derive structure-functional relationships for lncRNAs and nuclear protein complexes involved in transcription regulation.

High-Throughput Crystallisation Laboratory

José A. Márquez

PhD 1997, University of Valencia, Spain.

Postdoctoral research at EMBL.

Staff Scientist at EMBL Grenoble since 2003.

Team Leader since 2005.

Head of Crystallisation Facility since 2012.

SELECTED REFERENCES

Zander U. *et al.* (2016) Automated harvesting and processing of protein crystals through laser photoablation *Acta Cryst. D*72 (in press)

Rodriguez L. *et al.* (2014) C2-Domain Abscisic Acid-Related Proteins Mediate the Interaction of PYR/PYL/RCAR Abscisic Acid Receptors with the Plasma Membrane and Regulate Abscisic Acid Sensitivity in Arabidopsis. *Plant Cell* 26, 4802-20

Keller MA., *et al.* (2014) A gatekeeper helix determines the substrate specificity of Sjogren-Larsson Syndrome enzyme fatty aldehyde dehydrogenase. *Nat Commun.* 5, 4439

The Márquez team develops low volume, high-throughput, techniques to optimise protein crystallisation and uses them to study the structure of sensing and signalling molecules.

The High-Throughput Crystalisation (HTX) Laboratory is one of the major facilities for high-throughput, nanovolume, crystallisation screening in Europe and one of the major resources of Grenoble's Partnership for Structural Biology. It offers services to scientists working in European academic institutions through the EC-funded iNEXT and BioStructX projects and, at the same time, develops novel approaches in macromolecular crystallisation.

Integration of crystallisation and synchrotron data collection facilities through automated crystal harvesting and processing

In collaboration with the Cipriani group, we developed CrystalDirect™, a fully-automated method for harvesting and processing crystals (see figure). Crystals are grown on an ultrathin film in a vapour-diffusion crystallisation plate and recovered through laser-induced photo ablation. Advantages include: elimination of crystal fishing and handling; reduced mechanical stress during mounting; and compatibility with X-ray data collection. CrystalDirect was recently upgraded with an automated sample cry cooling and ligand soaking. The first prototype is now in operation at the HTX lab and is available to users through remote web interfaces (Marquez & Cipriani, 2014).

The Crystallisation Information Management System (CRIMS)

CRIMS tracks experiments and makes results available to users via the web in real-time, along with all experimental parameters. It has been licensed to 10 other laboratories in Europe, three of them at synchrotron sites. The analysis of the data stored in CRIMS has allowed us to develop a new method to determine the crystallisation likelihood of a protein sample based on a simple assay measuring thermal stability (Dupeux *et al.,* 2011). We constantly implement new functionalities to improve users' capabilities.

Molecular mechanisms in sensing and signalling

Our research focuses on the mechanisms of sensing and signalling at a structural level. We have recently contributed to the study of the function of plant members of the START family, including the PYR/PYL abscisic acid (ABA) receptors involved in the response to abiotic stress and the PR-10/Fra a proteins, which are involved in the control of secondary metabolic pathways during fruit ripening.

Future plans

In collaboration with other EMBL and ESRF groups we will work to develop a new generation of highly automated crystallography pipelines. These will integrate crystallisation, crystal harvesting and processing, and data collection in a single, automated workflow. This will enable us to implement fully automated pipelines for large scale ligand and fragment screening through macromolecular crystallography and to provide fully automated, remotely operated crystallography services to support the structural analysis of very challenging targets. Funded access to these new pipelines will be available through the EC-funded iNEXT project.
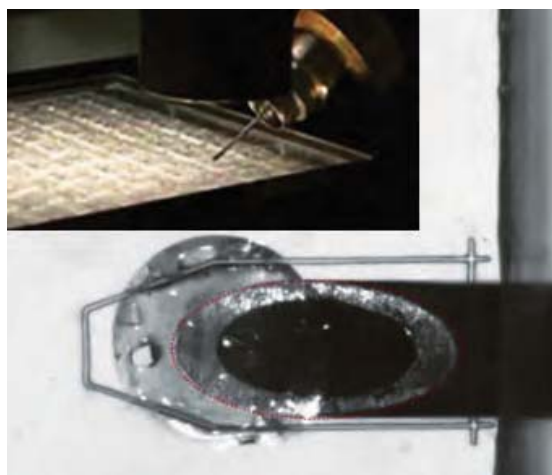
Figure 1: The CrystalDirect technique could benefit challenging structural biology projects, such as studies of membrane proteins or multi-protein complexes.
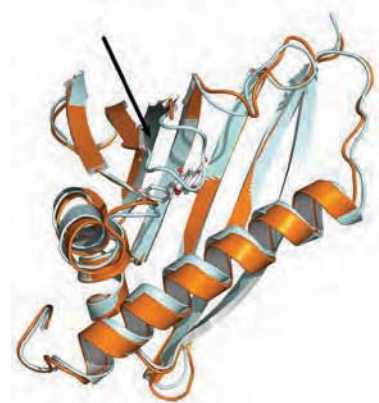
Figure 2: Detail of the Fra a 1 and Fra a 3 proteins in apo and ligand bound forms respectively.

Synchrotron Crystallography Team

Andrew McCarthy

PhD 1997, National University of Ireland, Galway.

Research associate, Utrecht University.

Postdoctoral research at Massey University and Auckland University.

Staff scientist at EMBL Grenoble.

Team leader since 2007.

SELECTED REFERENCES

Yang Z, *et al.* (2016) PIWI slicing and EXD1 drive biogenesis of nuclear piRNAs from cytosolic targets of the mouse piRNA pathway. *Mol. Cell* 61, 1-15

Delaforge E, *et al.* (2015) Large-scale conformational dynamics control H5N1 influenza polymerase PB2 binding to importin α. *J. Am. Chem. Soc.* 137, 15122-34

Bowler MW, *et al.* (2015) MASSIF-1: a beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules. *J. Synchrotron Radiat.* 22, 1540-7

Svensson O, *et al.* (2015) Fully automatic characterization and data collection from crystals of biological macromolecules. *Acta Crystallogr. D Biol. Crystallogr.* 71, 1757-67

The Synchrotron Crystallography Team works in close collaboration with the Structural Biology Group of the European Synchrotron Radiation Facility (ESRF) in the design, construction and operation of macromolecular crystallography (MX) and biological small angle X-ray scattering (bioSAXS) beamlines. Further information on these facilities can be found on the EMBL Grenoble services webpage. In addition we work in close collaboration with the Diffraction Instrumentation and High-Throughput Crystallisation teams on the implementation of new generation sample changers based on a 6-axis robot and *in situ* plate screening capabilities. We also actively contribute to the design and implementation of complex MX experimental workflows for advanced sample screening and data collection, as well as contributing to the development and expansion of ISPyB, a database for MX and bioSAXS experiments.

In the laboratory we study proteins involved in neuronal development, particularly the Slit-Robo signalling complex, proteins involved in piRNA biogenesis (Figure 2), and phosphoryl transfer proteins. Meanwhile, the BM14 group is actively involved in the structural studies of proteins involved in the *Toxoplasma gondii* epigenetic machinery in collaboration with Mohamed-Ali Hakimi of Grenoble Medical University.

The McCarthy team works on the design, construction and operation of MX and BioSAXS beamlines and studies proteins involved in neuron development.

Future plans

This year we will focus on expanding the capabilities of the new suite of MX-beamlines on ID30A (MASSIF) and ID30B. On BM14 and ID30B we will provide *in situ* screening and data collection for European users. Lastly, ID23-2 will be upgraded to provide an intense 1 μm^2 beamsize and equipped with a new high precision diffractometer designed by the Instrumentation Team. We will continue to develop existing facilities in order to optimise sample handling, and data collection and analysis functionalities. We hope that our combined efforts will directly impact on challenging structural biological problems.

In the laboratory, we will continue our research on the Slit-Robo signalling complex by trying to decipher how exactly Slit activates Robo on the cell surface and extend work on human kinase signal cascades. The BM14 group will use *in silico* screening methods to identify and test new potential inhibitors of the *Toxoplasma gondii* epigenetic machinery.

Figure 1: The new ID30B experimental end station at the ESRF.
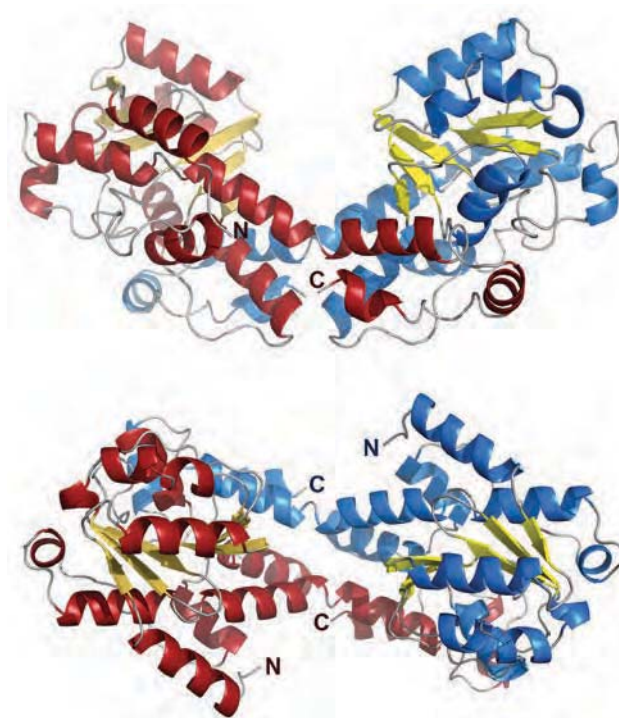
Figure 2: Crystal structure of Exd1 from *Bombyx mori*.

Structural biology of signal transduction and epigenetic gene regulation

Daniel Panne

PhD 1999, University of Basel.

Postdoctoral research at Harvard University, Boston.

Group Leader at EMBL Grenoble since 2007.

SELECTED REFERENCES

Muir K. *et al* (2016) Structure of the Pds5–Scc1 complex and implications for cohesin function. *Cell Rep.* 14(9):2116-26

Panne, D. (2013). Immunology: Cytosolic DNA sensing unraveled. *Nat. Chem. Biol.,* 9, 533-4

Delvecchio, M., *et al.* (2013). Structure of the p300 catalytic core and implications for chromatin targeting and HAT regulation. *Nat. Struct. Mol. Biol.,* 20, 1040-6

Larabi, A., *et al.* (2013). Crystal structure and mechanism of activation of TANK-binding kinase 1. *Cell Rep.,* 3, 734-46

The Panne group looks to understand how signalling pathways control gene expression, focussing on signalling in the innate immune system, epigenetic control of gene expression and chromatin assembly.

Our research has been focused on understanding how signalling pathways control gene expression by regulation chromatin assembly. We are particularly interested in pathways of the innate immune system that regulate gene expression. Pattern recognition receptors such as RIG-I or STING couple to IKK kinases such as TBK1 that regulate NFkB and IRF3/IRF7 activation. Once these transcription factors are activated, together with ATF2/c-Jun they form a higher-order regulatory structure called the 'enhanceosome' that regulates gene expression of interferon (figure 1). Higher-order transcription factor complexes such as the enhanceosome frequently recruit the co-activators CBP/p300. CBP/p300 are important to integrate the cellular signals by providing a scaffold function. CBP/p300 also modify chromatin and ultimately, in conjunction with remodellers and histone chaperones, makes chromatin permissive for gene transcription.

Understanding information transfer by such transient and dynamic complexes of cellular machinery is one of the most important challenges in biology today. One important first step toward characterising such dynamic processes is to determine the molecular architecture of essential components. We are using a combination of biophysical techniques including X-ray crystallography, electron microscopy, native mass spectrometry, and more to address the following questions concerning information transfer in this system: What is the architecture of signalling complexes that direct innate immune responses and control gene expression? How do these signalling pathways lead to assembly of higher-order regulatory complexes? How does assembly of such transcription factor complexes ultimately lead to chromatin modification? How does chromatin modification direct nucleosome remodelling and gene regulation? How does cohesin contribute to sister chromatid cohesion?

Future plans

Cellular signalling ultimately results in assembly of transcriptional regulatory complexes that direct chromatin modification, remodelling and gene expression. The enhanceosome has served as a paradigm for understanding signal integration on higher eukaryotic enhancers (figure 1). Assembly of the enhanceosome results in recruitment of enzymes such as CBP/p300 that acetylate chromatin. We aim to understand how recruitment of CBP/p300 allows cellular signal integration and chromatin acetylation. We ask how chromatin acetylation changes the structure of inhibitory nucleosomes and leads to a more permissive chromatin structure for gene expression. We also ask how cohesin, a topological ring complex that entraps sister chromatids is assembled. Answers to some of these questions are likely to contribute to our understanding of chromatin regulation and dysregulation in disease. This is not only of fundamental importance for cellular signalling, but also opens up opportunities for pharmacological targeting.
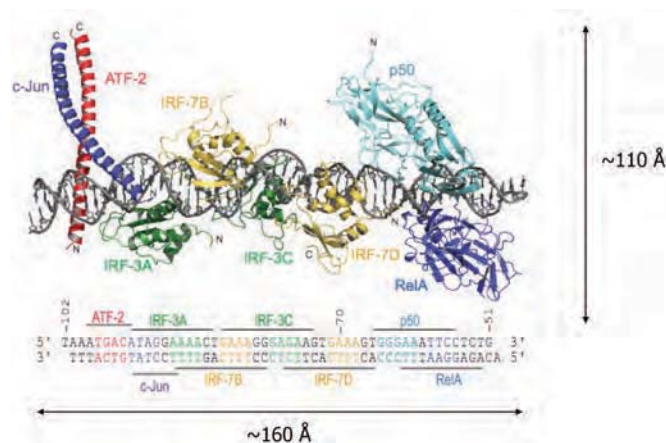
Figure 1: Atomic model of the INF-β enhanceosome.
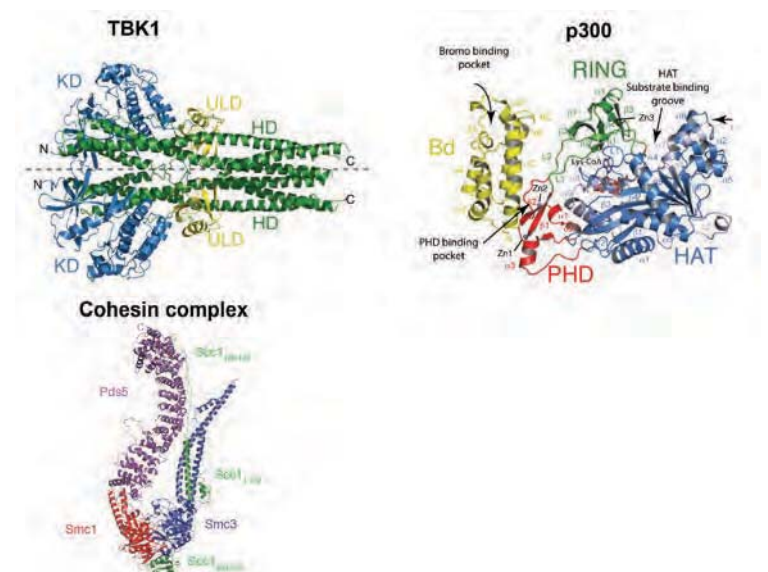
Figure 2: Gallery of structures.

Ribosomal complexes: targeting, translocation and quality control

Christiane Schaffitzel

PhD 2001, University of Zürich.

Habilitation 2008, ETH Zürich.

Team leader at EMBL Grenoble since 2007.

ERC Investigator

Professor of Biochemistry, University of Bristol, UK since 2014

SELECTED REFERENCES

Von Loeffelholz O. *et al.* (2015) The ribosome-SRP-FtsY co-translational targeting complex in the closed state. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3943-3948

Gaubitz C. *et al.* (2015) Molecular Basis of the Rapamycin Insensitivity of Target Of Rapamycin Complex 2, *Mol. Cell* 58, 977-988

von Loeffelholz O, *et al.* (2013) Structural basis of signal sequence surveillance and selection by the SRP-FtsY complex. *Nat. Struct. Mol. Biol.* 20, 604-10

Bieniossek C, *et al.* (2013) The architecture of human general transcription factor TFIID core complex. *Nature* 493, 699-702

Research in our laboratory combines molecular biology, *in vitro* and *in vivo* biochemistry and single-particle cryo-electron microscopy (cryo-EM) to study the structure and function of ribosomal complexes. Synthesis and folding of proteins require concerted interactions of the translating ribosome with translation factors, regulatory factors, molecular chaperones and factors involved in the export of proteins. Structures of translating ribosomes in complex with these factors provide critical insight into the interaction networks, stoichiometry and molecular mechanism of these megadalton-size complexes. Using cryo-EM, we can study the multicomponent translation machinery at close to physiological conditions. State-of-the-art electron microscopes and image processing, allow the determination of structures of prokaryotic and eukaryotic ribosomes at quasi-atomic resolution, demonstrating the power of this method.

A prerequisite for our functional and structural studies is the production of large amounts of homogenous, stable complexes in quantity and quality required for interaction assays, mass spectrometry and single-particle cryo-EM. In our laboratory, we established bacterial and eukaryotic cell-free translation systems for the in vitro generation of ribosomes displaying homogenous nascent polypeptide chains or stalled at a defined step in translation. We reconstitute the ribosomal complexes along the pathways of co-translational targeting and translocation and mRNA quality control. This approach was successfully applied in the case of the cryo-EM structures of the complex of the ribosome with the translocation machinery, of the translating ribosome-signal recognition particle (SRP) complex and of the ribosome in complex with SRP and SRP receptor (figure 1). The structural data supported by biochemical data, provide important and detailed snapshots of the mechanisms underlying these cellular processes ensuring correct folding, targeting and translocation of nascent proteins.

Future plans

We study ribosomal complexes involved in targeting, membrane protein integration, folding and assembly. We analyse the membrane protein complexes biochemically, by cross-linking / mass spectrometry (collaboration with Juri Rappsilber, TU Berlin, Germany) and single-particle cryo-electron microscopy. In collaboration with the groups of Matthias Hentze and Andreas Kulozik (Molecular Medicine Partnership Unit), we study mammalian ribosomal complexes involved in nonsense-mediated mRNA decay. We produce the eukaryotic factors involved by means of advanced recombinant eukaryotic technologies in collaboration with the Berger group. Moreover, our team collaborates with a number of groups to determine the structure of large macromolecular complexes in transcription, epigenetics and cellular signaling. In collaboration with Robbie Loewith (University of Geneva), we recently determined the first structure of Target of Rapamycin Complex 2 and the molecular basis of its rapamycin insensitivity (figure 2).

The Schaffitzel team combines molecular biology, biochemistry and cryo-electron microscopy to study large macromolecular complexes in protein targeting, secretion and membrane protein integration.
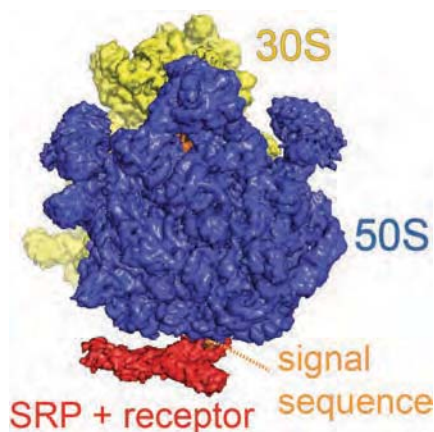
Figure 1: Cryo-EM structure of a bacterial co-translational targeting complex comprising the translating ribosome displaying a signal sequence (orange) bound to the signal recognition particle (SRP) and the SRP receptor (in red). The small ribosomal subunit (30S) is shown in yellow, the large ribosomal subunit (50S) in light blue.
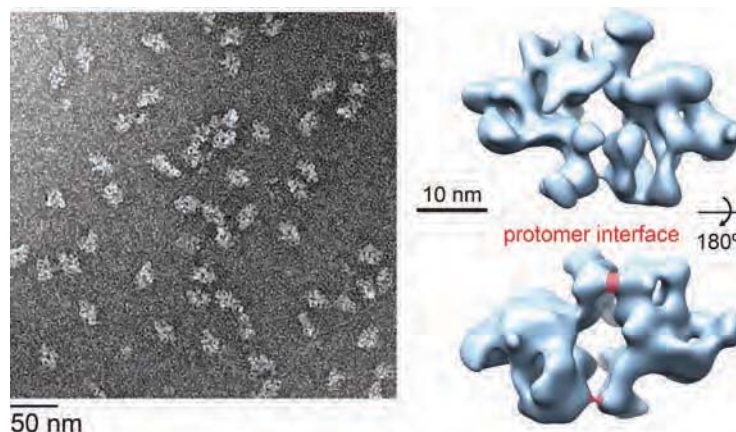
Figure 2: Structure of the Target of Rapamycin Complex 2. Left: Electron micrograph of negatively stained *S. cerevisiae* TORC2 prepared from endogenous source showing single particles. Right: The EM structure of the yeast TORC2 reveals a pseudo two-fold symmetry. The two protomers are related by a rotation of 180° around the z axis. The structure is shown in two different orientations (front and side view). The limited protomer interface of TORC2 is highlighted in red in the long side view.
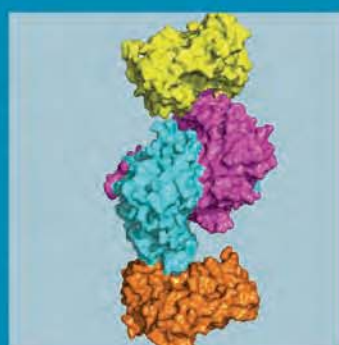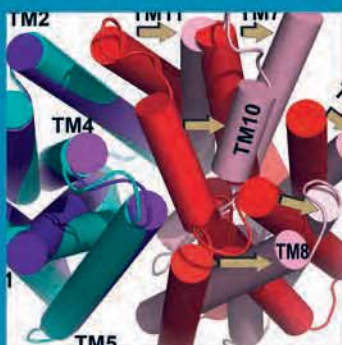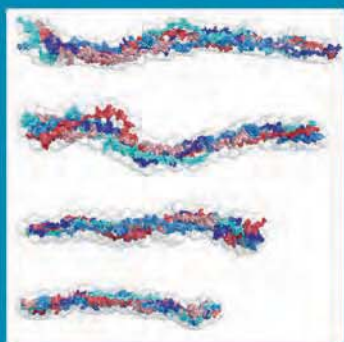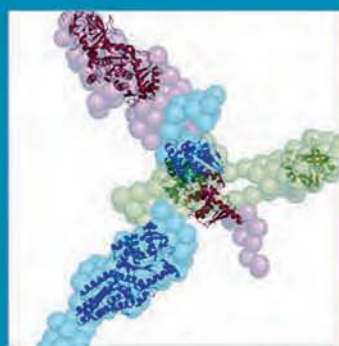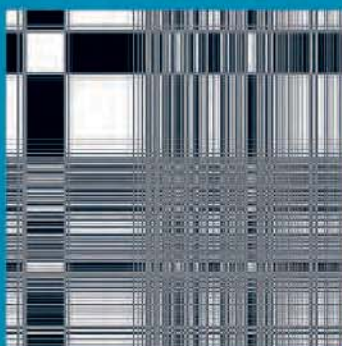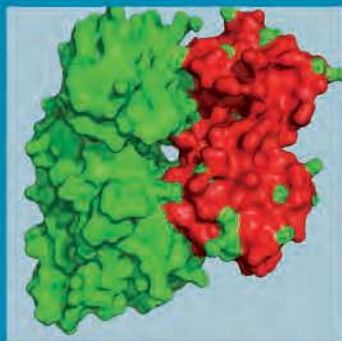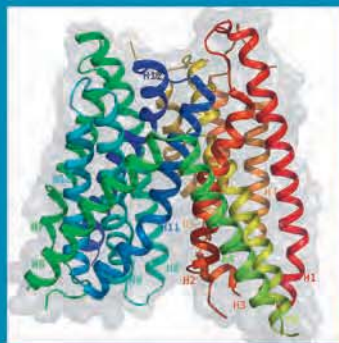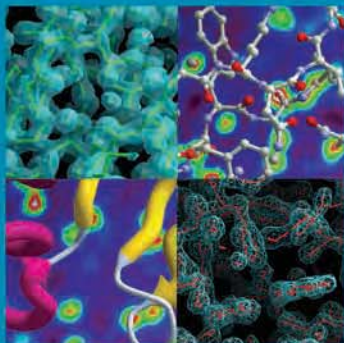
EMBL Hamburg

Structural Biology

Activities at EMBL Hamburg focus on state-of-the-art structural biology methods using synchrotron radiation, combining cutting-edge technology with an ambitious research programme for structures of multifunctional proteins and protein complexes of biomedical relevance.

EMBL Hamburg's laboratories are on the German Synchrotron Research Centre (DESY) campus, with synchrotron radiation (PETRA III) and laser (FLASH) facilities available. In addition, a powerful X-ray free electron laser is under construction. EMBL operates a new integrated facility, called EMBL@PETRA3, for applications in structural biology at the PETRA III ring. It comprises three state-of-the-art beamlines for macromolecular X-ray crystallography and small angle X-ray scattering of biological samples, complemented by facilities for sample preparation and characterisation, and data evaluation. EMBL Hamburg is also one of the main partners in the future Centre for Structural Systems Biology (CSSB) on the DESY campus.

EMBL Hamburg has set up an ambitious research programme for structures of multifunctional proteins and protein complexes of biomedical relevance. Present research interests of group leaders include cell surface receptors, protein assemblies in muscle cells, protein kinases, protein translocation into peroxisomes, and several projects relating to tuberculosis. Common to all projects is the goal to make optimum use of on-site high-brilliance synchrotron radiation and to explore novel opportunities of the X-ray Free Electron Laser. Beyond the tools in structural biology that are available on-site, EMBL Hamburg groups are engaged in many interdisciplinary collaborations with colleagues from other EMBL units, enabling access to a large variety of *in vitro* and *in vivo* functional techniques, including cellular imaging techniques.

EMBL Hamburg also has a well-established record for the development of novel, innovative technologies in structural biology. Leading software packages for the automation of data interpretation have been developed here and are used in a large number of projects across the world's research community. One example is the ARP/wARP package that allows automatic X-ray structure determination. Another package, ATSAS, allows the automatic interpretation of small angle X-ray scattering data for structural shape determination. Finally, there are two groups that focus on the development and construction of new equipment for experimental stations in structural biology, using synchrotron radiation. Present efforts focus on the installation of new robotics that allow automatic placement of biological samples into specialised synchrotron experiment facilities.

Matthias Wilmanns
Head of EMBL Hamburg

Synchrotron instrumentation for structural biology beamlines at PETRA III

Stefan Fiedler

PhD 1997, Johann-Wolfgang-Goethe-Universität, Frankfurt.

Postdoctoral fellow, then staff scientist at ESRF, Grenoble.

At EMBL Hamburg since 2004.

Team leader since 2006.

SELECTED REFERENCES

Siewert F, *et al*. (2012) High angular resolution slope measuring deflectometry for the characterization of ultra-precise reflective x-ray optics. *Meas. Sci. Technol.* 23, 074015

Schultke E, *et al*. (2011) Dual energy CT at the synchrotron: a piglet model for neurovascular research. *Eur. J. Radiol.* 79, 323-7

Keyrilainen J., *et al*. (2008) Toward high-contrast breast CT at low radiation dose. *Radiology* 249, 321-7

Roessle MW, *et al*. (2007) Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg. *Journal of Applied Crystallography* 40, S190-4

The Fiedler team focuses on the selection, customisation and integration of mechanics, control electronics and control software for X-ray based structural biology research.

EMBL has designed, built and operates three beamlines for structural biology at the PETRA III synchrotron radiation source on the DESY campus in Hamburg. Beamline facilities are dedicated to the leading techniques for X-ray-based structural research of biological samples: small angle X-ray scattering and macromolecular crystallography. Our team provides expertise in X-ray optics, precision mechanical engineering, robotics, control software and electronics and is in charge of the X-ray optical elements, experimental endstations, vacuum system, cryogenic system, control system, data acquisition system, technical infrastructure and parts of the civil engineering.

Recent major projects: All three beamlines reached regular user operation. In preparation for this, it was necessary to install and commission adaptive focusing optics at all beamlines, develop a multi-degree-of-freedom heavy-duty detector stage for the large-area pixel detectors at the MX beamlines (P13 instrument installed) and to develop and install white beam X-ray monitor systems for all beamlines. A cryogenic supply system for the beamline endstations has been installed and is in operation and a two-floor control hutch area with computing infrastructure was set-up in order to allow beamline control and users to perform experiments in a comfortable manner.

Multilayer optics: An ongoing in-house development is the construction of a double multilayer monochromator that is intended to boost the flux density at the P12 SAXS beamline – this enables time-resolved solution scattering experiments in the microsecond range. This instrument has been further developed from a prototype installed at a beamline of the former DORIS storage ring that can preserve the highly increased coherence of the radiation delivered by the PETRA III synchrotron. We are also working on the development of downstream experimental instrumentation with the level of synchronisation necessary for dynamic measurements.

Automation: For all beamline facilities, it is important to develop capabilities such as automatic operation and remote access. In this context, a robotic sample mounting system named MARVIN is being developed for the MX applications (see figure). This is characterised by high sample storage capacity, high sample mounting speed and flexibility. It is integrated, like all beamline elements, into a software-based control system which allows for a heterogeneous control environment and provides distributed access. The prototype has been in user operation on the BW7b beamline at DORIS and adapted versions are undergoing installation on the MX beamlines at PETRA.
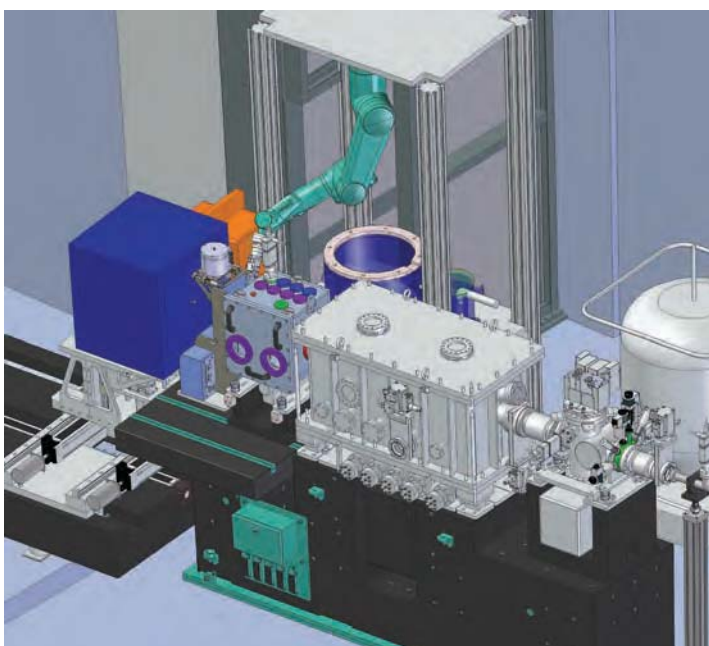
Design of the P14 experimental endstation with sample mounting system MARVIN.

Future plans

- Positional and intensity feedback and (active) vibrational stabilisation of the monochromatising optics.

- Automatic tuning of adaptive focusing mirror optics to vary beam size/shape.

- Integration of beamline elements into a global instrument protection system.

- Further automation of alignment and data acquisition and integration with sample preparation/crystallisation.

- Improved sample observation and positioning at or beyond the optical resolution limit for microcrystallographic investigations.

- Exploring preparative or bridging developments for the X-ray free electron laser with state-of-the-art synchrotron beamlines.

Integrative modelling for structural biology

Victor Lamzin

PhD 1987, Moscow State University.

Scientist, Inst. Biochemistry, Russian Academy of Sciences, Moscow, until 1991.

Postdoctoral research at EMBL Hamburg until 1995.

Staff scientist until 1997.

Group leader since 1997.

Senior scientist since 2003.

SELECTED REFERENCES

Giewekemeyer K, *et al.* (2015) Tomography of a cryo-immobilized yeast cell using ptychographic coherent X-ray diffractive imaging. *Biophys J.* 109, 1986-95

Carolan CG, Lamzin VS. (2014) Automated identification of crystallographic ligands using sparse-density representations. *Acta Crystallogr D Biol Crystallogr.* 70, 1844-53

Grigorenko VG, *et al.* (2013) Creation of recombinant strains of e-coli producing beta-lactamases. Study of catalytic properties and search for new beta lactamase inhibitors to overcome bacterial resistance. *RFBR Journal.* 4, 42-6.

Langer G, *et al.* (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc.* 3, 1171-9

The Lamzin group applies and develops cutting-edge computational methods and experimental approaches for sample quality control, experimentation and data interpretation in structural biology, with a major focus in macromolecular crystallography.

To fully understand the function of biological systems, accurate structural models of their components (small-molecule ligands, DNA, RNA, proteins and macromolecular assemblies) are required. Therefore, the focus of the group's activities is the development of the required methodologies, their application to projects of biomedical interest and their implementation in ARP/wARP (Langer *et al.,* 2008) – a world-leading software project. Given the breathtaking opportunities for structural biology arising with the availability of the European X-ray Free-Electron Laser (XFEL) from 2017, relevant research and development complement the group's portfolio.

Targets of biomedical interest: We integrate X-ray crystallography, lower resolution imaging, biochemistry, computational biology and biophysical methods in order to investigate targets of biomedical interest. These include: structural characterisation of components of the telomerase complex (Zvereva *et al.*, 2013) relevant to conditions with age disorder and cancer, inhibitor development for beta-lactamase (Grigorenko *et al.*, 2013) to combat antibiotic resistance, and studies of the nuclear pore complex. We also investigate the pathway of amyloid fibril formation via fragments of human gelsolin and class I hydrophobins (Kallio *et al.*, 2011).

Computer-aided drug design: We make use of various novel algorithms and, through their combination (Carolan *et al.*, 2014), develop new tools for drug discovery. Our ViCi software enables *in silico* screening of known ligands to provide new leads for drug design. Our interest in this direction is stimulated by our research into the biology of pathogenic species associated with human morbidity and mortality, and is focused on the probing of bacterial antibiotic resistance.

Biological imaging with FELs: We are developing protocols for the preparation and handling of biological samples and novel computational tools for the interpretation of measured diffraction data (Mancuso *et al.,* 2012). Focusing on the imaging of cellular nuclei, we explore the potential for single particle imaging experiments (Giewekemeyer *et al.*, 2015). In collaboration with the DESY Detector group and the P11 beamline staff, we probe the properties of the novel AGIPD detector, which will be used at the European XFEL.

Methods for biological structure determination: We develop a comprehensive range of algorithms for protein/ligand/DNA/RNA X-ray crystal structure determination and new procedures for dealing with challenging problems. We exploit the inherent properties of macromolecular structures and integrate additional information derived from a priori knowledge and dedicated databases.

Our main methodological focus is the ARP/wARP software project for macromolecular crystallography. For the automated interpretation of X-ray electron density maps we make use of sophisticated image and pattern recognition techniques, statistical analysis, data mining and bioinformatics tools. To provide users with easy access to quality assessment and model completion we have developed a user-friendly molecular viewer – ArpNavigator (Langer *et al.*, 2013).

Future plans

Together with international collaborators, we will undertake novel pilot projects aiming at the interpretation of structural data obtained from various sources and projects of medical or biotechnological importance. Driven by general academic interest, we will continue to focus on methods development for structural biology, addressing the challenges of limited resolution of the data and large size of macromolecular complexes. We will also continue contributing to the provision of computational services, synchrotron beamline facilities and applications for FEL-based diffraction.
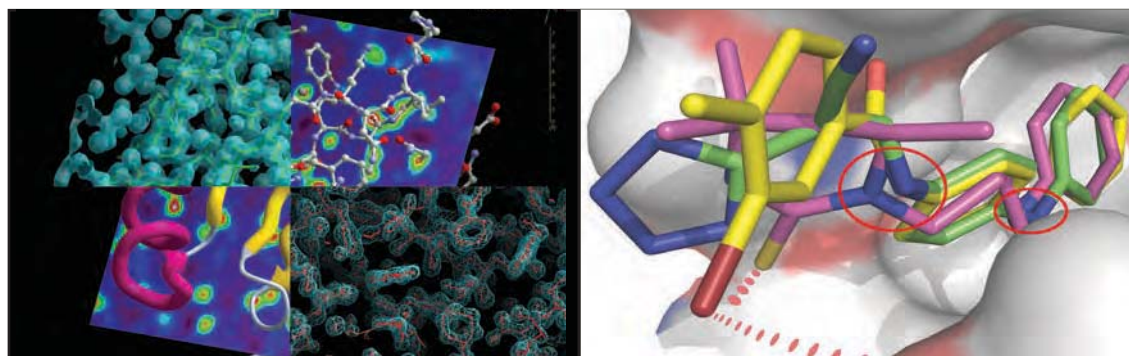
Figure 1: A selection of model viewing options in ArpNavigator. Shown clockwise from the top left: a stick representation in solid electron density; a ball-and-stick representation in planar density; a skeleton representation of the electron density shown as a mesh; the protein in cartoon representation in planar density.

Figure 2: A known inhibitor in green, aligned with two hits from the ViCi software, in the binding pocket of beta lactamase. Important interactions maintained are highlighted with red circles and those created for exploration with full red spots.

Structural and dynamic insights into nutrient uptake systems

Christian Löw

PhD 2008, Department of Physics at the University of Halle-Wittenberg.

Postdoctoral researcher at the Karolinska Institutet, Stockholm.

Group leader at EMBL Hamburg since February 2014.

SELECTED REFERENCES

Frauenfeld J, *et al.* (2016) A saposin-lipoprotein nanoparticle system for membrane proteins. *Nat Methods*, (Epub ahead of print)

Quistgaard EM, *et al.* (2016) Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nat Rev Mol Cell Biol.* 1, 123-32

Guettou F, *et al.* (2014) Selectivity mechanism of a bacterial homolog of the human drug-peptide transporters PepT1 and PepT2. *Nat. Struct. Mol. Biol.* 21, 728-31

Quistgaard EM, *et al.* (2013) Structural basis for substrate transport in the GLUT-homology family of monosaccharide transporters. *Nat. Struct. Mol. Biol.* 20, 766-8

Using biophysical and biochemical methods, the Löw group aim to enhance understanding of the structural basis for substrate recognition in peptide transporters.

Cell membranes compartmentalise metabolic processes and serve as selective barriers for permeation. Therefore, nutrient transport through the plasma membrane is essential to maintain homeostasis within the cell. Many proton-coupled secondary active transporters of the major facilitator superfamily (MFS) are involved in the accumulation of nutrients above extracellular levels. Structural and functional analyses of MFS transporters suggest an alternating-access mechanism for the transport of substrates across the membrane. Here the transporter adopts different conformational states, allowing the substrate binding site to face either side of the membrane. A full transport cycle involves at least three different conformational states – inward open, occluded and outward open –, with each of them in a ligand-bound and ligand-free state. Since MFS transporters are found in all branches of life and often with numerous gene copies, we believe that many if not all of these transporters follow a common transport mechanism. Proton coupled oligopeptide transporters of the PepT family (also known as the POT family) are responsible for the uptake of a range of different di- and tripeptides, derived from the digestion of dietary proteins, and are highly conserved in all kingdoms of life. The best studied members of this family include the two human peptide transporters, PepT1 and PepT2. Both are also of great pharmacological and pharmaceutical interest as they accept a number of drugs and amino acid-conjugated pro-drugs as substrates. A detailed understanding of the structural basis for substrate recognition can therefore help to convert pharmacologically active compounds into substrates for PepT1 and PepT2 and improve their absorption in the small intestine and subsequent distribution in the body. We therefore study the proton-dependent oligopeptide transporter (POT) family using a combination of biochemical and biophysical methods.

Future plans

- Characterisation of pro- and eukaryotic nutrient transporters in various states of the transport cycle using X-ray crystallography to decipher a common transport mechanism of MFS transporters.

- Structural and dynamic insights into the binding mode of POTs to peptides, drugs, and inhibitors.

- Molecular insights into structure and function of transport regulators of nutrient transporters.

Integral membrane proteins are a challenging class of proteins in terms of their structural and functional characterisation. Over the years we have developed and established new tools and a workflow for protein production and quality control of membrane proteins including functional assays. Furthermore, we are trying to find new ways to stabilise integral membrane proteins *in vitro* upon extraction from their natural environment.
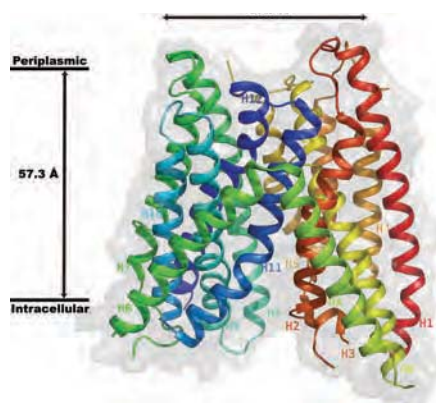
Figure 1: Structure of a proton dependent oligopeptide transporter (POT) in the inward open conformation (Guettou *et al,* 2013).
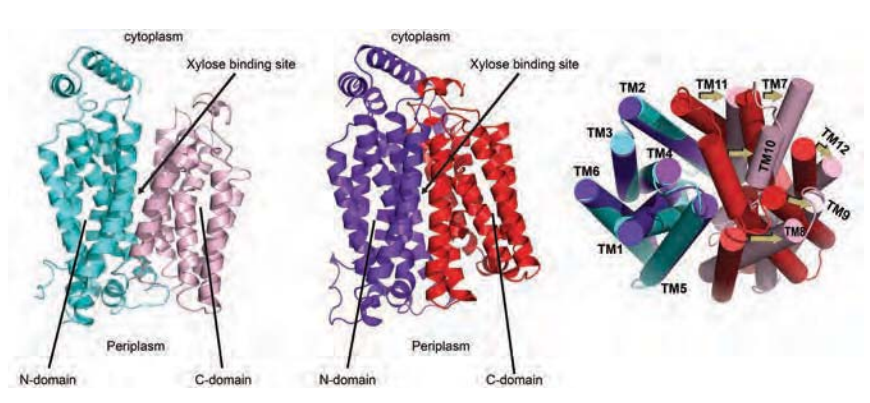
Figure 2: Structural differences between the inward open and occluded state structures of the sugar transporter XylE. The structures were overlaid and the transmembrane helices (TM) are labelled and arrows designate changes in the position of the helices upon opening of the transporter towards the cytoplasm.

Structural biology of cell surface receptors

Rob Meijers

PhD 2001, EMBL Hamburg/University of Amsterdam.

Postdoctoral research at the Dana Farber Cancer Institute, Boston.

Staff scientist at the Synchrotron Soleil, Saint Aubin, France, 2006-2009.

Group leader at EMBL Hamburg since 2009.

SELECTED REFERENCES

Finci LI, *et al.* (2014) The crystal structure of Netrin-1 in complex with DCC reveals the bifunctionality of Netrin-1 as a guidance cue. *Neuron* 83, 839-49

Dunne M, *et al.* (2014) The CD27L and CTP1L endolysins targeting clostridia contain a built-in trigger and release factor. *PLoS Pathog.* 10, e1004228

Meijers R *et al.* (2007) Structural basis of Dscam isoform specificity. *Nature* 449, 487-91

The Meijers group investigates signalling mechanisms across the cell membrane using integrated structural biology and biophysical techniques.

Receptor signalling in axon guidance: Many single-pass transmembrane receptors signal through association with other receptors, passing extra-cellular cues through the membrane to cytosolic signalling pathways. We have established mammalian expression technologies to produce large quantities of receptor fragments and ligands. Complexes between fragments of receptors and their ligands are characterised by crystallography, SAXS and electron microscopy, giving a holistic view of single-pass membrane receptor signalling.

We have determined the crystal structure of the guidance molecule netrin-1 in complex with a fragment of the deleted in colorectal cancer (DCC) receptor (Finci *et al*, 2014), revealing how the external cue netrin-1 brings two DCC receptors in close proximity, leading to the propagation of a signal across the cell membrane. This has dramatic effects, causing an axon to change its course or undergo apoptosis. Netrin-1 is capable of binding and combining several receptors, and we believe that this occurs through a generic receptor binding site that involves small molecules. These small molecules help to select certain receptors, thereby influencing signalling outcomes. We are in the process of identifying these small molecules and investigating how the other receptors bind to netrin-1.

Integrated sample preperation and characterisation (SPC) at the synchrotron beamlines: The SPC facility has been built to deliver a sample handling environment close to the beamlines. Users can process samples just before accessing the beamlines, using a range of biophysical techniques for sample purification and characterisation. Project-based access is available for extensive sample characterisation and optimisation. We provide a large state-of-the-art crystallisation facility, thermal shift assays for sample optimisation and standardised quality control protocols including mass spectrometry analysis.

Future plans

We will investigate other cell surface receptor signalling hubs to study how external morphogens can affect signalling across the cell membrane. We will focus on signalling systems that affect cell migration, since this affects fundamental processes in tissue development. Netrin-1, for instance, is thought to play an important role in the regulation of cell proliferation and the prevention of tumour metastasis. We would like to understand which factors contribute to netrin induced signaling, to help in the design of new therapies for controlled cell migration.

To further enhance the versatility in sample handling and characterisation at the PETRA III beamlines, we will develop new integrated sample environment setups. The miniaturisation and automation of the macromolecular crystallography and SAXS beamlines has led to an increase in sample throughput to such an extent, that it becomes worthwhile to integrate complementary sample manipulation on or near the beamlines. The standardisation of quality control, purification and sample delivery allows the users to handle more challenging samples on site, and to make better informed decisions about their experiments in real time.

Netrin-1 is a guidance molecule that steers axons to their targets and guides the branching of blood vessels to starving tissue. The crystal structure of netrin-1 in complex with its cell surface receptor DCC (Deleted in Colorectal Cancer) shown here provides a structural basis for the clustering of DCC by netrin-1. The structure also suggests that DCC can be replaced by other cell surface receptors, leading to different signalling outcomes.
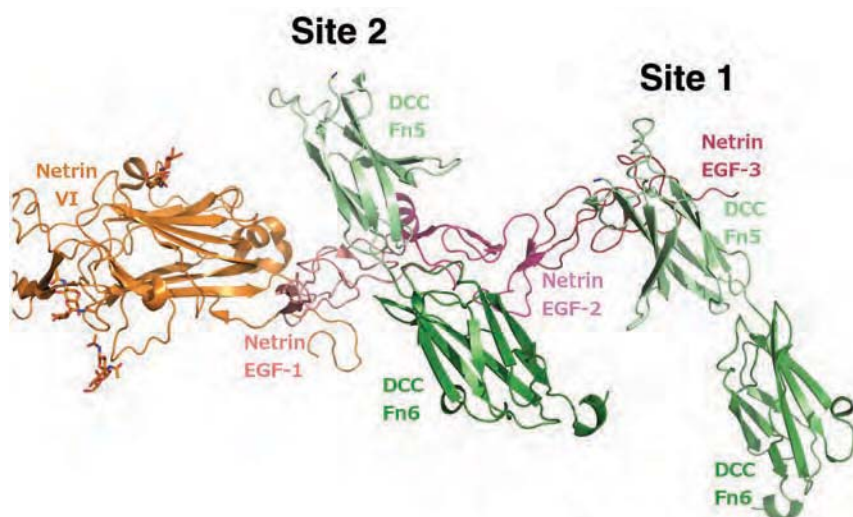
Tools for structure determination and analysis

Thomas Schneider

PhD 1996, Technical University of Munich/EMBL.

Postdoctoral research at the MPI for Molecular Physiology, Dortmund, and the University of Göttingen.

Group leader at the FIRC Institute of Molecular Oncology, Milan.

Group leader at EMBL since 2007.
Coordinator MX@PETRA3.

Senior scientist since 2015

SELECTED REFERENCES

Gati C, *et al.* (2014) Serial crystallography on in vivo grown microcrystals using synchrotron radiation. *IUCrJ* 1, 87-94

Mosca R & Schneider TR. (2008) RAPIDO: a web server for the alignment of protein structures in the presence of conformational changes. *Nucleic Acids Res.* 36, 42-6

Schneider TR. (2008) Synchrotron radiation: micrometer-sized x-ray beams as fine tools for macromolecular crystallography. *HFSP J* 2, 302-6

The Schneider group operates two beamlines for macromolecular crystallography at the PETRA III synchrotron and develops new methods for structure determination.

In structural biology a crystal offers two unique properties: firstly, by its very nature, a crystal creates a conformationally homogeneous ensemble. Secondly, through its periodicity, a crystal amplifies diffracted signals such that high-resolution features can be resolved. Our aim is to extract the best possible diffraction data from crystals and to support the evaluation of these data in terms of their molecular structure.

Since 2012 we have operated two beamlines for macromolecular crystallography: P13 and P14 on the PETRA III synchrotron at the DESY site in Hamburg. **http://www.embl-hamburg.de/services/mx**.

Inspired by the developments for crystallography at free-electron lasers, we have established the Serial Synchrotron Crystallography Methodology for data collection on micro-crystals on synchrotron beamlines in collaboration with Henry Chapman (CFEL Hamburg) and Lars Redecke (University of Lübeck). The 'serial helical scan approach' for data collection is now available as a routine protocol on the P14 beamline.

We also provide computational tools for phasing ('hkl2map') and structure analysis ('rapido') via the web **http://webapps.embl-hamburg.de**.

Future plans

We are pushing our beamlines towards adjustable X-ray beams and flexible data acquisition schemes in order to provide optimum conditions for crystals of different sizes and different diffraction quality. Serial crystallography approaches are most suitable for crystals with sizes in the sub 10-micron range. For larger crystals, systematic multiple crystal strategies need to be employed. Based on the versatile optics in place on the beamlines – i.e. adaptive X-ray optics on P13, adaptive optics plus compound reflective lenses (CRL) on P14 – we are developing protocols optimised for a wide range of different sample types. In collaboration with the Fiedler team at EMBL Hamburg and the Cipriani team at EMBL Grenoble, we are constantly improving the instruments on the beamline to obtain better (anomalous) data to enable experimental phasing of large molecular complexes. Once established, we implement experimental protocols in the MxCuBE-beamline user interface, so that these can be employed easily by the beamline users – see **https://github.com/mxcube/mxcube/wiki**.

In 2016, a CrystalDirect crystal harvesting system, developed by colleagues at EMBL Grenoble, will be deployed at EMBL Hamburg. The system will allow users to gently (and remotely) treat and harvest fragile crystals for synchrotron data collection, thereby enabling crystal structure determination for a wider range of samples. We will also use CrystalDirect to develop more rational and controlled approaches to derive macromolecular crystals for experimental phasing approaches.
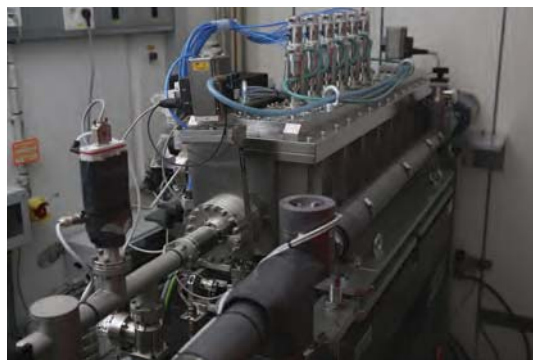
Figure 1: A 'transfocator' for the insertion of different combinations of CRLs into the X-ray beam on P14. Choosing different combinations of lenses allows users to match the size of the beam to the size of the crystal while maintaining the total X-ray flux constant.
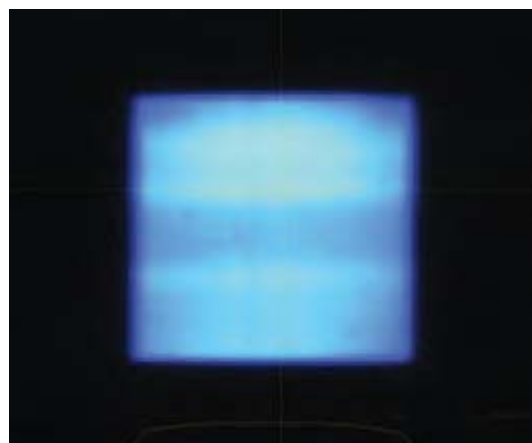
Figure 2: 2D profile of a homogeneous high-flux X-ray beam (ca. 150 μm x 150 μm) as formed by a set of 8 CRLs. Homogeneous illumination of a crystal is an important step for the collection of accurate data.

Small-angle X-ray scattering from macromolecular solutions

Dmitri Svergun

PhD 1982, Dr. of Science 1997, Institute of Crystallography, Moscow.

At EMBL since 1991.

Group leader since 2003.

Senior scientist since 2011.

SELECTED REFERENCES

Blanchet C, *et al.* (2015) Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J. Appl. Crystallogr.* 48, 431-43

Franke D, Jeffries CM & Svergun DI. (2015) Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. *Nature Methods* 12, 419-22

Graewert MA, *et al.* (2015) Automated pipeline for purification, biophysical and x-ray analysis of biomacromolecular solutions. *Scientific Reports* 5, 10734

Gruszka DT, *et al.* (2015) Cooperative folding of intrinsically disordered domains drives assembly of a strong elongated protein. N*ature Communications* 6, 7271

Small-angle X-ray scattering (SAXS) reveals low-resolution (1-2 nm) structures of biological macromolecules and functional complexes in solution. Recent experimental and methodical developments have significantly enhanced the resolution and reliability of the SAXS-based structural models, and the last decade saw a renaissance of biological SAXS worldwide.

Our group leads the development of novel computational methods for data analysis (Figure 1) and constructing structural models from the scattering data. Special attention is given to the joint use of SAXS with other methods including crystallography, NMR, electron microscopy and bioinformatics. We developed the world's most used program package, ATSAS, employed by more than 10,000 users from more than 50 countries.

Our group runs a dedicated high brilliance synchrotron beamline – P12 – at DESY's third generation storage ring, PETRA III. P12 has a robotic sample changer for rapid automated experiments, and possesses a data analysis pipeline for building structural models online. The beamline offers FedEx-style and remote data access options, as well as an in-line purification and biophysical characterisation setup using size exclusion chromatography (Malvern).

In collaborative projects with members of the ever-growing SAXS community, group members offer advice on sample preparation and provide extensive help with data collection, data interpretation and structural modelling. SAXS is employed to study overall structural organisation of macromolecules and conformational transitions and to quantitatively characterise oligomeric mixtures, intrinsically unfolded proteins, hierarchical systems and other objects of high biological and medical importance (Figure 2).

Future plans

The present and future work of the group includes:

- Further methods development for the reconstruction of macromolecular structure from X-ray and neutron scattering.

- Hybrid applications of SAXS with crystallography, NMR, electron microscopy and bioinformatics to construct and validate structural models.

- External user support and participation in collaborative SAXS projects at the P12 beamline

- Further extension of P12 capabilities including time-resolved and anomalous scattering techniques as well as ultra fast data acquisition combined with new multi-layer optics for high-brilliance X-ray experiments.

The Svergun group places special emphasis on hybrid approaches combining SAXS with X-ray crystallography, NMR spectroscopy and computational methods to elucidate macromolecular structure and conformational transitions in solution.
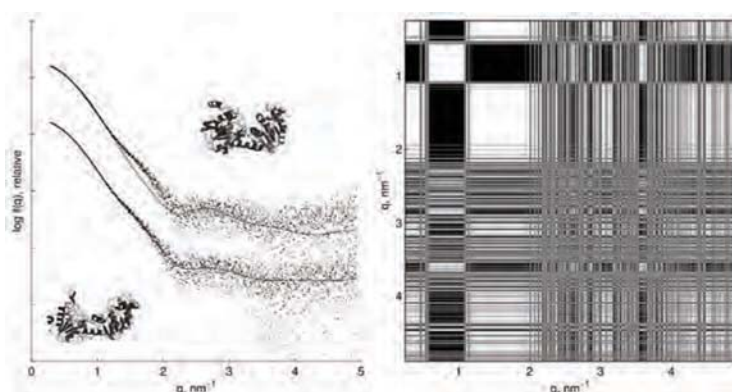
Figure 1: A new data analysis tool: Correlation Map (Franke, Jeffries & Svergun, 2015).
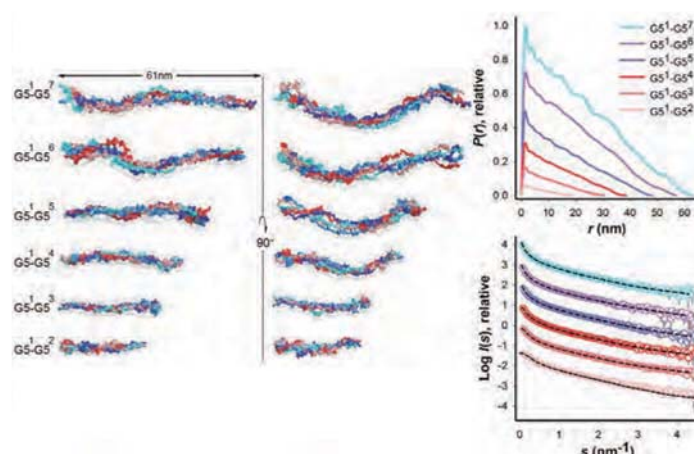
Figure 2: SasG Protein from *Staphylococcus aureus* (Gruszka *et al.*, 2015).

Structure and function of protein complexes in biological systems

Matthias Wilmanns

PhD 1990, University of Basel.

Postdoctoral research at the University of California, Los Angeles, 1991-1993, and EMBL Heidelberg 1993-1996

Group Leader, Senior Scientist and Head of EMBL Hamburg since 1997

Research Director of the Center for Structural Systems Biology (CSSB) since 2014

Professor at the University Hamburg Clinical Center Eppendorf (UKE) since 2015

SELECTED REFERENCES

Fodor K, *et al.* (2012) Molecular requirements for peroxisomal targeting of alanine-glyoxylate aminotransferase as an essential determinant in primary hyperoxaluria type 1. *PLoS Biol.*, 10, e1001309

Pinotsis N, *et al.* (2012) Superhelical architecture of the myosin filament-linking protein myomesin with unusual elastic properties. *PLoS Biol.*, 10, e1001261

Due AV, *et al.* (2011) Bisubstrate specificity in histidine/tryptophan biosynthesis isomerase from *Mycobacterium tuberculosis* by active site metamorphosis. *PNAS*, 108, 3554-9

de Diego I, *et al.* (2010) Molecular basis of the death-associated protein kinase-calcium/calmodulin regulator complex. *Sci Signal*, 3, ra6

The Wilmanns group investigates a variety of protein-ligand complexes within the context of biological systems, employing a broad range of molecular and structural biology techniques.

The architecture of the protein interactome in sarcomeric muscle cells: Many proteins found in muscle cells, when dysfunctional, are associated with cardiovascular diseases. We investigate how large protein filament systems forming the overall architecture of 'sarcomeric units' in muscle cells are connected and interact with each other, frequently mediated via small scaffold proteins. We have determined the structure and function of some key complexes, including telethonin-mediated assembly of the N-terminus of titin (Zou *et al.*, 2006) and the overall architecture of the elastic filament protein myomesin (Pinotsis *et al.*, 2008; 2012). Our future focus will be on novel protein interactions within the sarcomeric Z-disk and M-line region, and novel signalling functions of the protein partners involved.

Activity regulation in protein kinases: About 70 protein kinases in the human kinome share a common C-terminal autoregulatory domain. To investigate the mechanism of activity regulation in these kinases, we determined the structure of the kinase domain from the giant filament protein titin, in the inhibited apo-conformation (Mayans *et al.*, 1998) and unravelled the structure of the apoptotic Death Associated Protein Kinase-1, in the presence of the regulatory scaffold calcium/calmodulin (CaM) (figure 1). This structure provides insight into how CaM binding leads to kinase activation by withdrawing the autoregulatory domain from the kinase active site. Ongoing structural studies are complemented by *in vitro* and *in vivo* functional studies, to decipher underlying, general molecular mechanisms that regulate the activity of members of the CaM-dependent protein kinase family and ultimately promote drug discovery.

The architecture of the translocon of peroxisomes: Peroxisomes are cell organelles that allow sequestered metabolic processes that would interfere with other processes in the cytosol. Proteins involved in these processes are generally translocated as active and folded targets. We have unravelled the mechanism involved in the recognition of peroxisome protein targets by the peroxisome import receptor Pex5p, by determining the structure of the cargo-binding domain of the receptor in the absence and presence of the cargo protein sterol carrier protein 2 (Stanley *et al.*, 2006) and alanine-glyoxylate aminotransferase (figure 2). Our goal is to provide insight into the overall architecture of the peroxisomal translocon, using a broad range of structural biology, imaging, genetic and cell biology-oriented approaches.

Structural systems biology in *M. tuberculosis*: We have determined the X-ray structures of a number of protein targets. For instance, we were able to identity Rv2217 as a novel cysteine/lysine dyad acyltransferase, which allows activation of several important protein complexes by lipoylation (Ma *et al.*, 2006). Using available structural data and supported by European research network systeMTb, we aim to use systems biology-orientated approaches to investigate functional processes in living mycobacteria, with the aim of making data available to promote the development of new drugs, vaccines and diagnostic markers.
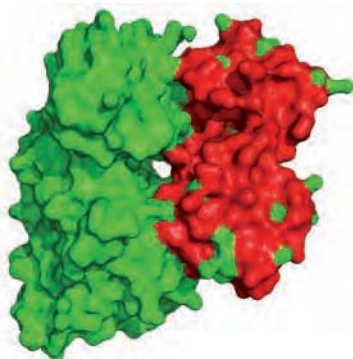
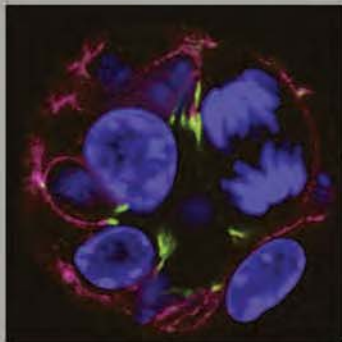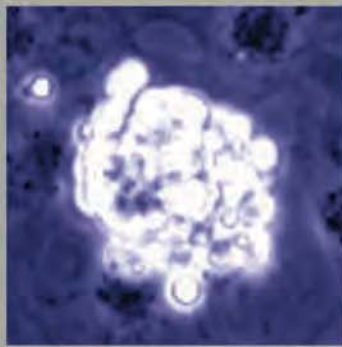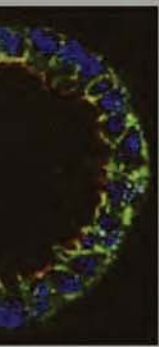Figure 1: DAPK-CaM complex. Colour codes: DAPK, green; CaM, red. (de Diego *et al.*, 2010)
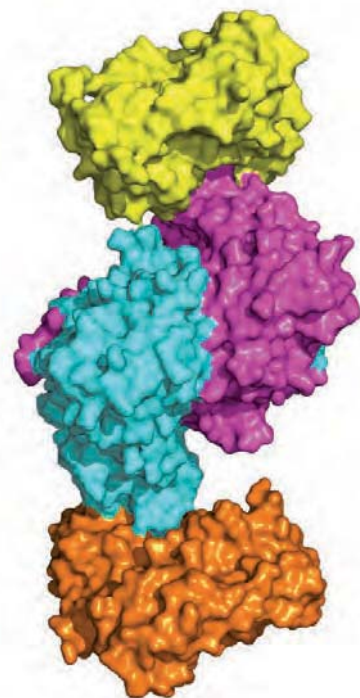
Figure 2: AGT-(Pex5p)2-AGT complex. Colour codes: AGT, yellow, orange; Pex5p, cyan, magenta. (Fodor *et al.*, 2012)

EMBL Monterotondo
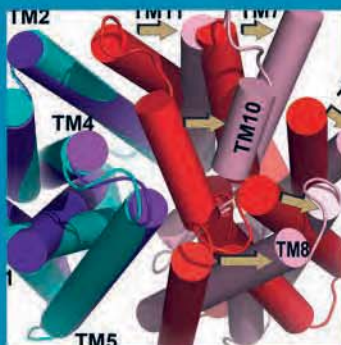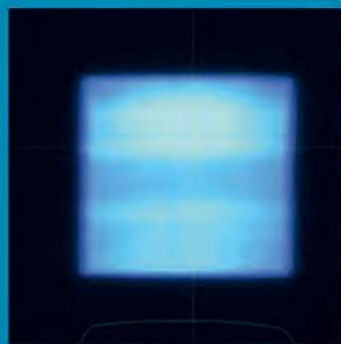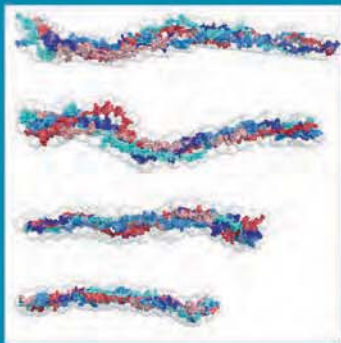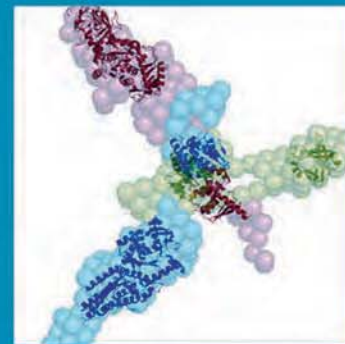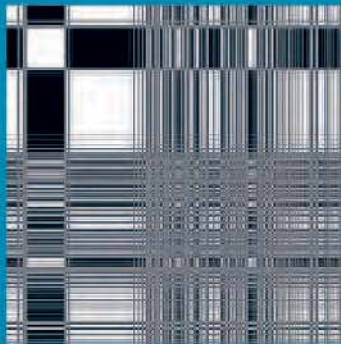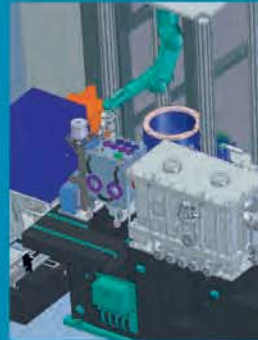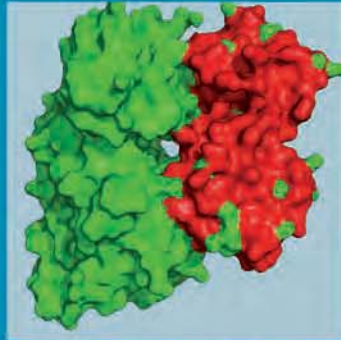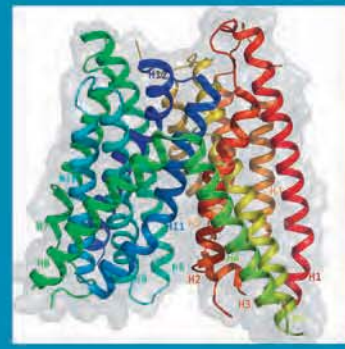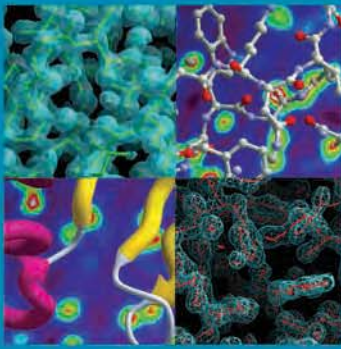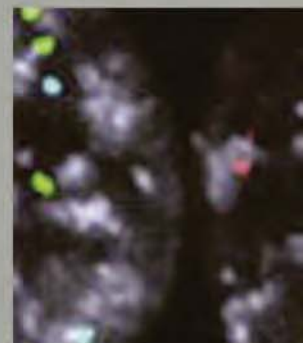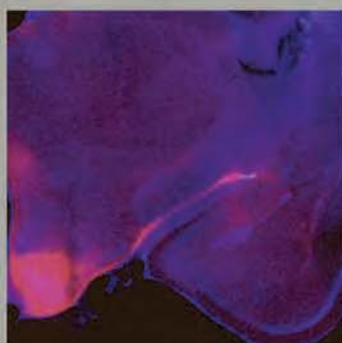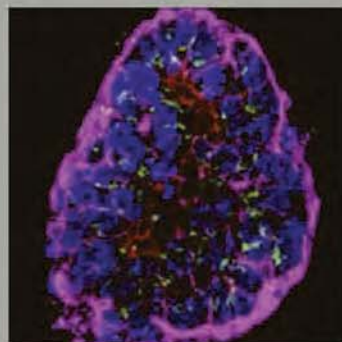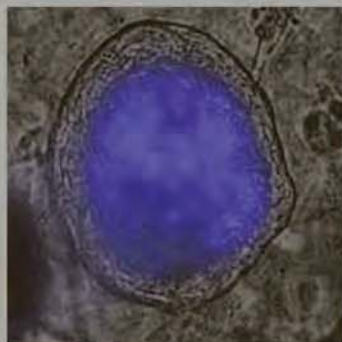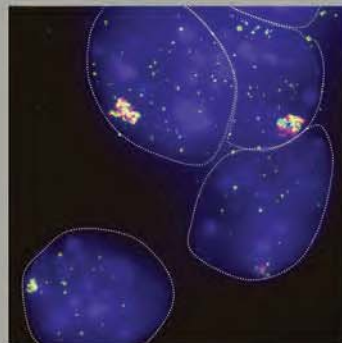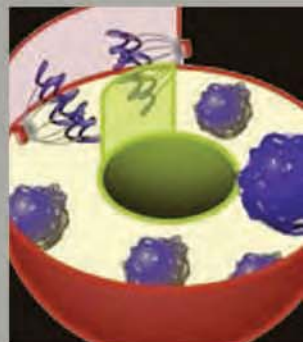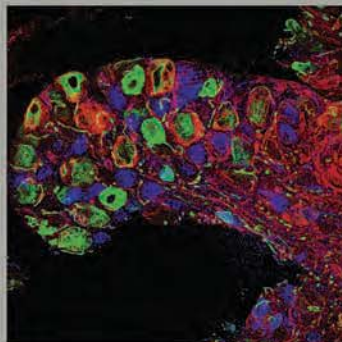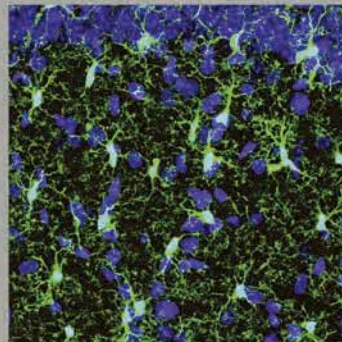
Mouse Biology

Our research is focused on the study of critical mammalian physiological phenomena from a molecular perspective in the context of the whole organism.

The Unit is composed of independent research groups whose work is bound together by the study of the mouse *in vivo* as the principle experimental organism and the application of state-of-the-art methods for genetic and epigenetic perturbation of cellular and physiological function. Recent advances in gene delivery and modification technology have dramatically increased the ease and rapidity with which genetic and epigenetic perturbations can be introduced into the mouse genome allowing increasingly sophisticated causal experimental approaches in this organism.

Currently research groups at EMBL Monterotondo are investigating the epigenetic control of early development, blood cell development and differentiation, neural circuits and behavior, neural computation, and somatosensation. A major and continuing effort is underway to increasingly centre research activities at the Unit around neurobiology and epigenetics.

A state-of-the-art animal facility provides access to a full range of mouse transgenic and genetic engineering technologies, and researchers are further supported by on site facilities for histology, flow cytometry, and microscopy. Additional support is provided by EMBL Core Facilities at the Heidelberg site (Genomics, Electron Microscopy, Metabolomics, Protein Expression, Advanced Light Microscopy, and Proteomics). Efficient and flexible administrative and generous financial support provides an environment where EMBL researchers can focus their undivided attention on the pursuit of ambitious research questions. Researchers at EMBL take advantage of close links with Units at EMBL sites in Hinxton (Bioinformatics), Heidelberg (Developmental Biology, Genome Biology, Cell Biology & Biophysics, Structural & Computational Biology), and Hamburg and Grenoble (Structural Biology). Interdisciplinary collaborations between EMBL groups and non-EMBL researchers are facilitated by the EI$_3$POD competitive postdoctoral funding mechanism. The Mouse Biology Unit is particularly committed to continuing to foster collaborations and strengthening links with other major Italian research centres.

Philip Avner
Head of EMBL Monterotondo

Visual systems neuroscience

Hiroki Asari

PhD 2007, Cold Spring Harbor Laboratory, USA.

Postdoctoral research at Harvard University and the California Institute of Technology, USA.

Group leader at EMBL since 2016.
Joint appointment with the Developmental Biology Unit.

SELECTED REFERENCES

Asari H & Meister M. (2014) The projective field of retinal bipolar cells and its modulation by visual context. *Neuron* 81, 641-52

Asari H & Meister M. (2012) Divergence of visual channels in the inner retina. *Nat. Neurosci.* 15, 1581-1589

Asari H, Zador AM. (2009) Long-lasting context dependence constrains neural encoding models in rodent auditory cortex. *J. Neurophysiol.* 102, 2638-56

Asari H, Pearlmutter BA & Zador AM. (2006) Sparse representations for the cocktail party problem. *J. Neurosci.* 26, 7477-90.

The Asari group combines experimental and computational approaches to study the principles and the function of neuronal circuits, specifically in the early visual system in mice.

How do neurons interact with each other to perform specific neurological functions? To address this question, my postdoctoral research focused on the circuit functions of the retina because it is an experimentally most accessible part of the central nervous system that performs considerable computations with well-defined input stimuli and output responses. Specifically, I explored how signals flow in the inner retinal circuits, and showed that a given principal neuron (bipolar cell) can send dramatically different signals to its various postsynaptic partners (ganglion cells) by forming distinct local microcircuits with inhibitory interneurons (amacrine cells). This indicates that each of these microcircuits serves as a computational unit for visual processing in the retina.

Future plans

As we have learned more about local circuit functions of individual brain areas, it becomes increasingly important to understand (a) how these areas interact with each other to modulate their local circuit functions; and (b) how such interactions help process sensory and motor signals to organize an animal's behaviour. An excellent model system to address these questions is the mouse retina. First, the retina is one of the best understood circuits in the brain, and the physiological functions are known in detail from the molecular to the cellular circuit level. Second, various tools are available in mice to label, monitor, and manipulate specific cell types and circuits. Third, although the retina is often thought to make only feed-forward connections to the brain, there is an anatomical substrate of efferent inputs from multiple brain areas to the retina across species.

We will thus focus on the bidirectional interactions between the retina and the brain under different behavioural and internal states of an animal, and analyse the functional role of the retinal efferents and their impact on the retinal afferents from the viewpoint of visual computation (Figure 1). The results will clarify how visual processing in the retina is dynamically modulated by efferent inputs under different behavioural conditions, and how a diverse set of retinal outputs serves as a basis for visual computation along the afferent visual pathways. This will help explain what each stage of visual processing is for, and also help refine an input-output model to better describe visual responses at each stage of the visual system (Figure 2). Ultimately, the outcome of our research will support a future development of visual prosthetic devices by faithfully emulating the function of the early visual system, in particular, the retina.
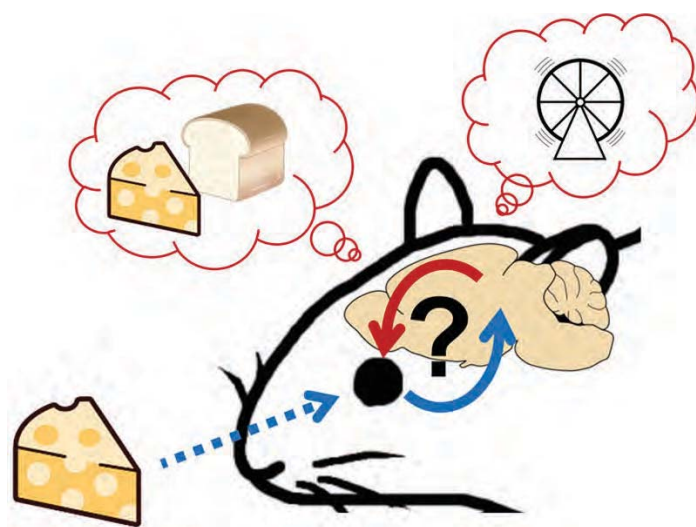
Figure 1: Big question: How does the retina take part in the brain-wide complex feedback circuits?
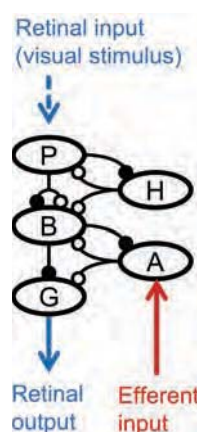
Figure 2: Retinal circuit diagram. Main cell types: P, photoreceptor; H, horizontal cell; B, bipolar cell; A, amacrine cell; G, ganglion cell. Major synaptic connections: closed circles, excitatory; open circles, inhibitory. Certain amacrine cells receive efferent inputs from the brain (red).

Dynamics of epigenetic regulation

Philip Avner

PhD in yeast genetics, University of Warwick.

CNRS Director of Research.

Head of the Mouse Molecular Genetics Unit
at the Institut Pasteur 1990-2011.

Head of the Developmental Biology Department
at the Institut Pasteur 2006-2011.

Emeritus Director CNRS and Institut Pasteur since 2012

Head of EMBL Monterotondo since 2012.

SELECTED REFERENCES

Prudhomme J, *et al.* (2015). A rapid passage through a two-active-X-chromosome state accompanies the switch of imprinted X-inactivation patterns in mouse trophoblast stem cells. *Epigenetics Chromatin* 8, 52

Cerase A., *et al.* (2015). Xist localization and function: new insights from multiple levels. *Genome Biol.* 15, 166

Guyochin A, *et al.* (2014). Live cell imaging of the nascent inactive X chromosome during the early differentiation process of naive ES cells towards epiblast stem cells. *PLoS One,* 9, e116109

Merzouk S, *et al.* (2014). Lineage-specific regulation of imprinted X inactivation in extraembryonic endoderm stem cells. *Epigenetics Chromatin* 11

The genetic material of the cell is not all equally available for transcription and this availability, which varies with cell type and developmental stage, is mediated largely by epigenetic modifications to the genome playing out mainly at the level of the chromatin. The double focus of our research has been on mouse genetics and epigenetics, with a particular emphasis on the interface between genetics and epigenetics, as revealed using X-inactivation as an experimental paradigm. X-inactivation, which occurs early during development in female mammalian embryos, ensures the dosage compensation between females carrying two copies of the X chromosome and males with a single copy of the X. Parts of the process in the mouse can be modelled *ex vivo* using female embryonic stem cells. We have been at the forefront of research into the characterisation and functional analysis of the different components of the X-inactivation centre, the key complex on the X chromosome for the initiation of X-inactivation. Ongoing X-inactivation research involves the study of the Xce locus (X-controlling element), a classically defined genetic locus existing in different forms, which appears to influence which of the two X chromosomes will be chosen to undergo X-inactivation, and studies bearing on the stability and nature of the imprinted X-inactivation process occurring in the different extra-embryonic lineages of the mouse.

By carrying out this research we are not only discovering the multiplicity and plasticity of mechanisms that feed into this process of epigenetic control but also providing insights into the links between epigenetic processes and development. Our approach involves a combination of genetics, genomics, biochemistry and cell biology and both *ex vivo* and *in vivo* experimental approaches.

The Avner group combines genetics, genomics, biochemistry, and cell biology to study the nature of the X-inactivation process and the role of epistasis in genetic regulation

Future plans

Our knowledge of the epigenetic processes underlying the X-chromosome process will be used to inform our approach to defining the contributions of epigenetic regulation to allele-specific epistasis, the process where the effects of one gene are modified in a non-additive allele-specific manner by one or several other genes. The importance of such epistasis for some developmental processes such as haemopoiesis, and its molecular basis is being explored using mouse inter-specific chromosome substitution strains.

Mouse female trophoectoderm stem cells: The XIST non-coding RNA (green) shows partial overlap with the repressive H3K27Me3 histone mark on the inactive X chromosome (Morey et al.)

Neural Control of Instinctive Behaviour

Cornelius Gross

PhD 1995, Yale University.

Postdoctoral research at Columbia University.

Group leader at EMBL Monterotondo since 2003.

Deputy Head of Outstation and Senior Scientist since 2009.

ERC Advanced Investigator.

SELECTED REFERENCES

Zhan Y, *et al.* (2014) Deficient neuron-microglia signaling results in impaired functional brain connectivity and social behavior. *Nat. Neurosci.* 17, 400-6

Silva BA, *et al.* (2013) Independent hypothalamic circuits for social and predator fear. *Nat. Neurosci.* 16, 1731-3

Gross CT, Canteras NS (2012) The many paths to fear. *Nat. Rev. Neurosci.* 13, 651-8

Paolicelli RC, *et al.* (2011) Synaptic pruning by microglia is necessary for normal brain development. *Science* 333, 1456-8

The Gross group uses pharmacological, histochemical, electrophysiological and behavioural genetic approaches to study the neural circuits underlying behaviour in mice.

The laboratory is interested in understanding, at a molecular and neural circuit level, how early life events influence brain development in order to establish behavioural traits in adulthood, with a particular focus on fear and anxiety. We are currently pursuing two areas of research:

Neural circuits encoding fear and anxiety

Fear is a mental state elicited by exposure to threats or cues that signal those threats that is part of our natural defense mechanism. However, in its pathological form fear can become excessive or inappropriate – features associated with anxiety disorders. The amygdala plays a central role in processing threat stimuli that are then integrated by downstream hypothalamic and brainstem circuits to produce appropriate defensive behaviours. Our team has showed that distinct amygdala outputs and downstream circuits are recruited in response to different types of fear with defensive responses to painful stimuli, predators, and bullies mediated by distinct pathways (Gross and Canteras, 2012; Silva *et al.*, 2013; Figure 1). These data suggest that pathological fear comes in different flavours and may be amenable to selective therapeutic treatment. Current work in the lab combines molecular genetic, electrophysiological, and behavioural methods in mice to understand how amygdala, hypothalamic, and brainstem circuits support and adapt fear responses to diverse threats.

Developmental programming of brain wiring by microglia

Microglia are non-neuronal cells of the myeloid lineage that infiltrate the brain during development and are thought to play a role in brain surveillance. Recent studies from our group and others have shown that microglia play a key role in the elimination of synapses during postnatal brain development, a phenomenon called 'synaptic pruning' (Paolicelli *et al.*, 2011). Mice with deficient synaptic pruning show weak functional brain connectivity, poor social behaviour, and increased repetitive behaviour – all hallmarks of autism – suggesting that some features of this neurodevelopmental disorder may depend on a deficit in synaptic pruning (Yang, Paolicelli *et al.*, 2014). We are currently using a variety of tools to identify the 'eat me' and 'spare me' signals that regulate pruning and understand how synapse elimination remodels neural circuits during development.

Future plans

We aim to discover the neural circuits and molecular mechanisms that support individual differences in behavioural traits in health and disease. On the long term, this should allow us to form specific hypotheses about how human behaviour is determined and lead to improved diagnostic and therapeutic tools for mental illness.
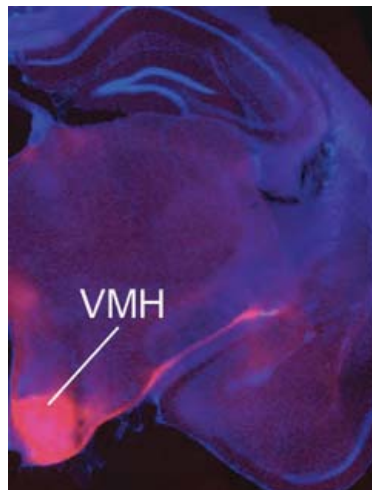
Figure 1: Expression of the hM4D/CNO pharmacogenetic inhibition tool in cell bodies and projections of the VMH in Nr5a1::hM4D-2A-tomatoF transgenic mice as revealed by farnesylated tomato (red) reporter protein (B. Silva).
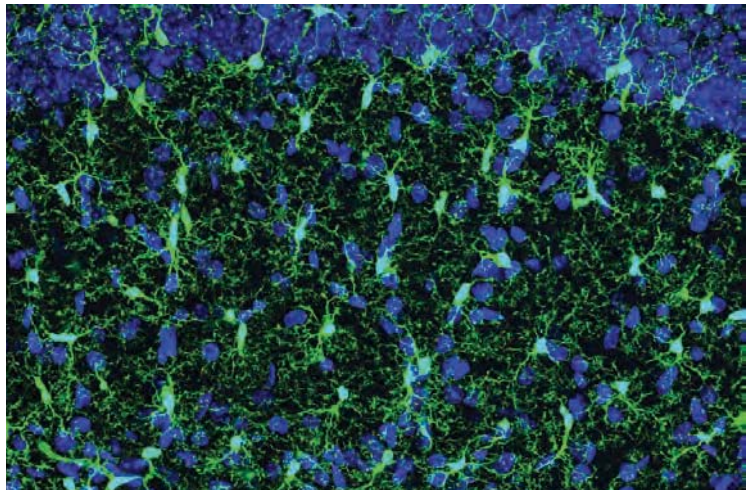
Figure 2: Microglia (green cells) visualised in the hippocampus of Cx3cr1GFP transgenic mice (cell nuclei labelled in blue, R. Paolicelli).

Molecular physiology of somatosensation

Paul Heppenstall

PhD 1997, University of Edinburgh.

Postdoctoral work at the Max Delbrück Centrum, Berlin.

Junior Professor at the Charité, Berlin.

Group leader at EMBL Monterotondo since 2008.

Group leader in the Molecular Medicine Partnership Unit.

SELECTED REFERENCES

Yang G, *et al.* (2015) Genetic targeting of chemical indicators *in vivo*. *Nat. Methods* 12, 137–9

Nockemann D, *et al.* (2013) The K(+) channel GIRK2 is both necessary and sufficient for peripheral opioid-mediated analgesia. *EMBO Mol Med* 5, 1263-77

Kalebic N, *et al.* (2013) αTAT1 is the major α-tubulin acetyltransferase in mice. *Nat Commun* 4, 1962

Zurborg S, *et al.* (2007) Direct activation of the ion channel TRPA1 by Ca2+. *Nat. Neurosci.* 10, 277-9

Somatosensation is the process by which we sense touch and pain. It is dependent upon specialised sensory neurons which extend from the skin to the spinal cord and are tuned to detect mechanical, thermal and chemical stimuli. Surprisingly, the mechanisms that transduce these forces into electrical signals at the peripheral endings of sensory neurons are not well understood. Our research focuses on identifying and characterising these transduction components and exploring how they are altered during chronic pain states.

We use a combination of molecular, imaging and electrophysiological techniques to examine functional properties of sensory neurons at their peripheral and central terminals. At the molecular level, we are interested in mechanisms of touch sensitivity of sensory neurons. Normal mechanical sensitivity is dependent upon a complex of proteins that are localised at the peripheral endings of sensory neurons. Evidence supports a central role for the cytoskeleton in regulating the composition and function of this complex. Using cellular, electrophysiological and molecular imaging techniques we are characterising the contribution of the cytoskeleton, in particular microtubules to mechanotransduction.

Another focus of the group is to understand the biophysical properties of ion channels involved in sensory transduction. Much of our work has concentrated on the ion channel TRPA1, a member of the Transient Receptor Potential (TRP) family of channels. In mammals, TRPA1 is expressed by nociceptors and plays a key role in detecting noxious chemicals. We demonstrated that intracellular calcium ions directly activate TRPA1 via an EF-hand domain in the N-terminus of the protein and that calcium is essential for normal activation of the channel by noxious chemicals. We are now interested in how TRPA channels have evolved to sense diverse stimuli across different phyla – for example, in snakes and insects TRPA1 orthologues are activated by warm temperatures. Using a combination of molecular and electrophysiological techniques we have mapped the regions in *Drosophila* TRPA1 that are responsible for sensing temperature and described how single TRPA1 channels are activated by heat.

The Heppenstall group combines molecular, imaging and electrophysiological techniques to examine how sensory neurons turn information about touch and pain into electrical signals.

Future plans

- Identification of novel genes involved in touch and pain.

- Mutagenesis of transduction channels and associated proteins to determine their mechanism of action.

- Tissue-specific and conditional mutagenesis of sensory-related genes in defined subpopulations of sensory neurons.

- Development of new techniques to measure functional properties of sensory neurons at their terminals.

The major focus of the laboratory is to correlate cellular studies on somatosensation with observations made at the physiological level. To this end we are developing genetic approaches that, combined with electrophysiological and molecular imaging techniques, will enable us to characterise sensory neurons *in situ*. A better understanding of sensory neuron function may ultimately lead to improved therapies for the treatment of chronic pain.

SNAP tag labelling of sensory neuron cell bodies in the Dorsal Root Ganglia.

Generation of hematopoietic progenitor and stem cells during development

Christophe Lancrin

PhD 2003, Université Pierre et Marie Curie (Paris VI), Paris, France.

Postdoctoral research at the Paterson Institute for Cancer Research, Manchester, United Kingdom.

Group Leader at EMBL since January 2011.

SELECTED REFERENCES

Vargel Ö, *et al.* (2016) Activation of the TGFβ pathway impairs endothelial to haematopoietic transition. *Scientific Reports* 6, 21518

Lancrin C, *et al.* (2012) GFI1 and GFI1B control the loss of endothelial identity of hemogenic endothelium during hematopoietic commitment. *Blood* 120, 314-22

Ferreras C, *et al.* (2011) Identification and characterization of a novel transcriptional target of RUNX1/AML1 at the onset of hematopoietic development. *Blood* 118, 594-7

Lancrin C, *et al.* (2010) Blood cell generation from the hemangioblast. *J. Mol. Med.* 88, 167-72

The Lancrin group studies the haematopoietic system and looks to develop strategies to improve methods for generating blood cells from stem cells.

The continuous generation of blood cells throughout life relies on the existence of haematopoietic stem cells (HSC) generated during embryogenesis. They have the ability to self-renew and to generate all types of blood cells. Any pathology affecting these cells could lead to the development of serious diseases such as leukaemia and anaemia.

The origin of blood cells has been the subject of an intense scientific debate during the last decade. A first theory suggests that haematopoietic cells arise from a mesodermal progenitor with smooth muscle, endothelial, and haematopoietic potential called the haemangioblast. However, a conflicting theory associates the first haematopoietic cells with a phenotypically differentiated endothelial cell with haematopoietic potential (i.e. a haemogenic endothelium).

We used a model of early haematopoiesis based on embryonic stem cells (ESC) – that have the capacity to generate any cell types. Using this system coupled with time-lapse microscopy, clonogenic assays and flow cytometry analysis, we have demonstrated that the haemangioblast generates haematopoietic progenitors through the formation of a haemogenic endothelium stage, providing the first direct link between these two precursor populations. Our results merge the two a priori conflicting theories on the origin of haematopoietic development into a single linear developmental process, which makes the haemogenic endothelium the immediate precursor of blood cells (figures 1 and 2).

Future plans

Recently, the generation of the ESC-like induced pluripotent stem cells (iPSC) from fully differentiated cell types, such as skin fibroblast, provided a major breakthrough in the field of regenerative medicine. However, important work has to be done to differentiate efficiently iPSC or ESC toward specific cell types including blood cell progenitors such as HSC.

Consequently, the focus of our research is to unravel the mechanisms underlying the generation of blood cells from the haemogenic endothelium. Combining single cell transcriptomics, lineage tracing, time-lapse microscopy, in vitro and in vivo, we plan to identify and study the genes responsible for the generation of blood progenitors and blood stem cells during embryonic life. Our research will further the understanding of the mechanisms of cell fate decisions leading to the production of the first haematopoietic cells and enable the development of new strategies to improve methods of blood cell generation from ESC or iPSC for regenerative medicine.
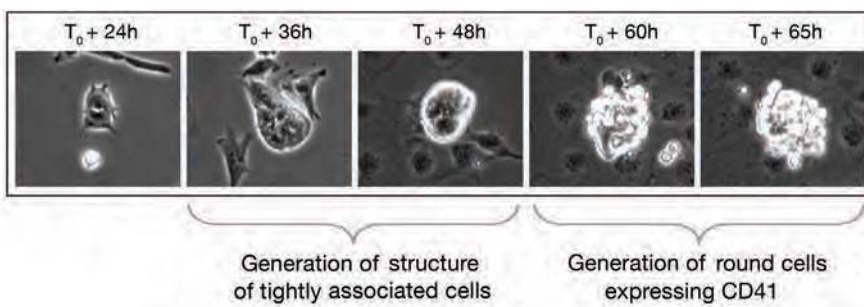
Figure 1: Time-lapse microscopy analysis of haemangioblast differentiation. The formation of a blast colony from the haemangioblast can be retrospectively divided in two consecutive phases: a generation of a structure of tightly associated endothelial cells and the production of round non adherent cells expressing the haematopoietic marker CD41.
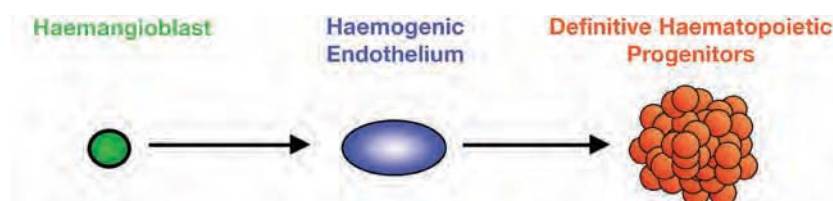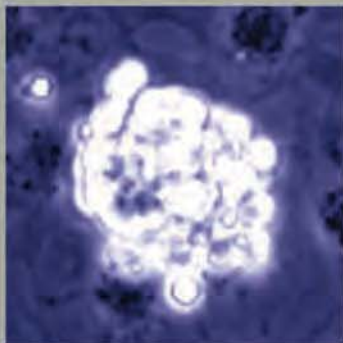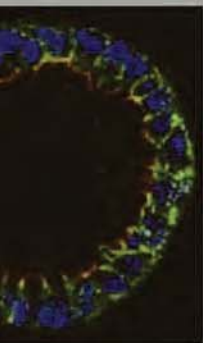

Figure 2: New model of blood cell origin: The haemangioblast and the haemogenic endothelium are part of the same developmental pathway to generate blood cell progenitors during embryonic life.

Index

Imprint

© EMBL 2016

European Molecular Biology Laboratory

Heidelberg ۱ Hinxton ۱ Grenoble ۱ Hamburg ۱ Monterotondo

EMBL member states

Austria | Belgium | Croatia
Czech Republic | Denmark
Finland | France | Germany
Greece | Iceland | Ireland
Israel | Italy | Luxembourg
Malta | the Netherlands
Norway | Portugal | Spain
Sweden | Switzerland
United Kingdom

Associate member states

Argentina | Australia

Prospect member states

Hungary | Lithuania | Poland
Slovak Republic

www.embl.org

European Molecular Biology Laboratory