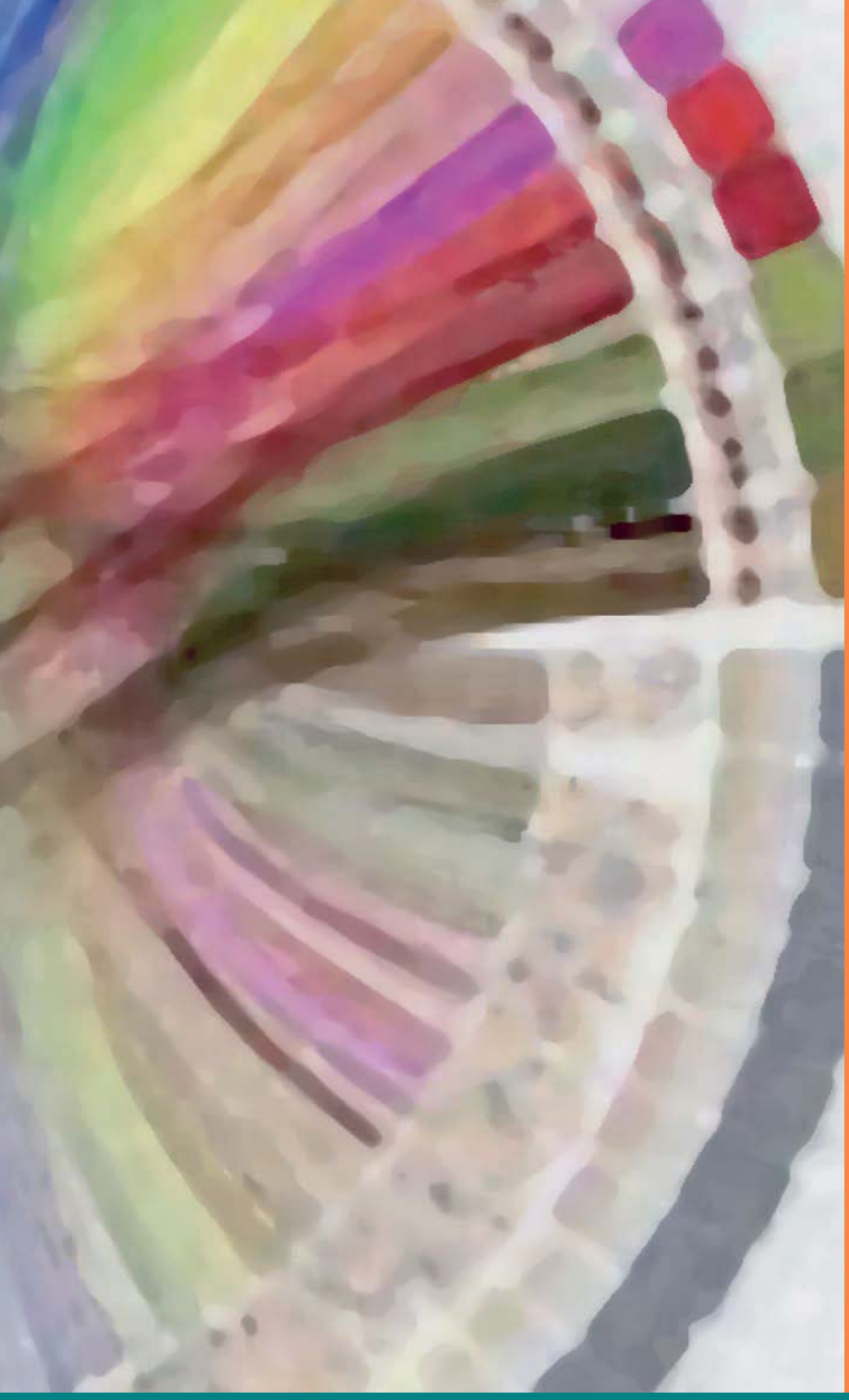


2015

Research at a Glance

EMBL

Contents

EMBL was established forty years ago to create a European centre of excellence for highly talented young scientists. Today, it is Europe's leading institution in the molecular life sciences, and one of the highest-ranked research institutes in the world. In recognition of its unique value, EMBL – Europe's only intergovernmental laboratory in the life sciences – enjoys continued support from its member states, which have now increased in number to twenty-one member states, two associate members outside of Europe, and three prospect member states that committed to joining EMBL in the next three years.

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 well beyond its research mission. EMBL is a major provider of research infrastructures and services for the life sciences, and offers training programmes for scientists, regularly used as a model of best practice by other research organisations. EMBL is broadly engaged in technology development, and drives innovation through a successful technology transfer programme, thereby allowing scientists and society at large to benefit from its inventions and discoveries. Finally, EMBL contributes to shaping European science policy and strategy, and promotes the integration of research activities in Europe and worldwide.

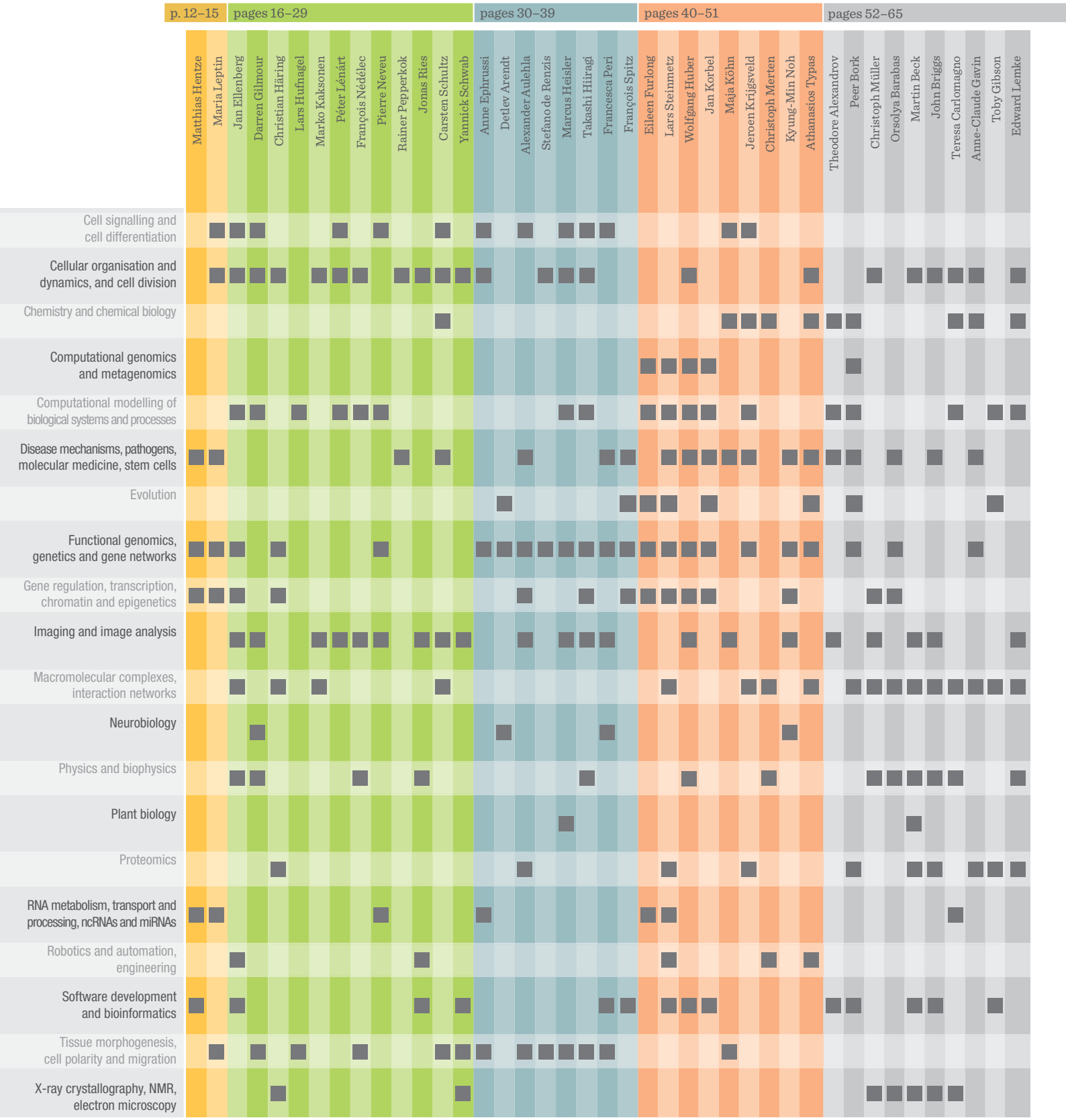
The unique mix and seamless integration of these research-related activities at EMBL is complemented by a variety of successful organisational principles, such as international recruitment of the most talented scientists, regular staff turnover, and rigorous peer review of the Laboratory's activities.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

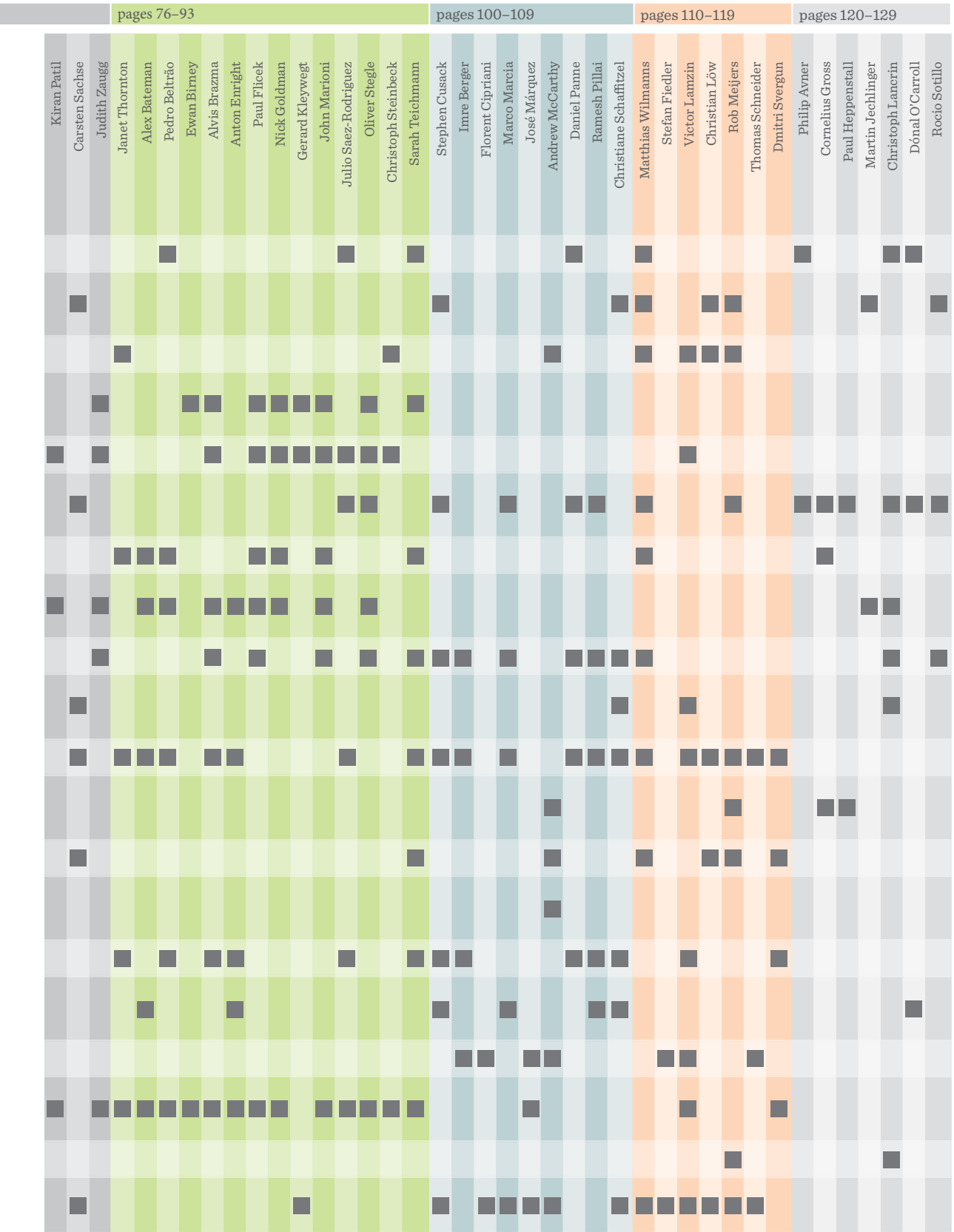
In *Research at a Glance*, you will find a concise overview of the work of our research groups and core facilities, addressing some of the most challenging and pressing questions in the molecular life sciences. The overarching goal of the Laboratory's research is to gain a comprehensive mechanistic understanding of biological systems by navigating across scales – from single molecules over cells and tissues to entire organisms. What truly distinguishes EMBL, however, is its interdisciplinary and collaborative approach to science, whereby researchers with complementary expertise from different disciplines work together to tackle specific biological problems. When researchers leave to assume key positions in other institutes, the skills they have developed in the dynamic, interdisciplinary and international EMBL environment are exported to our member states.

The critical mass of expertise and resources concentrated at EMBL has produced many important achievements. The value of the laboratory's pioneering work to the scientific community was illustrated in 2014 by a report published in the journal *Nature* on the 100 most highly cited papers of all time, which included three papers produced at EMBL. The impact of EMBL's scientific output, its attractiveness to world-leading young scientists, and the continuous support from our member states are a testimony to the Laboratory's success, and indicate that we will remain at the forefront of the life sciences and consolidate our leadership position in the future European research landscape.

Iain Mattaj
EMBL Director General

Research topics




pages 76–93
pages 100–109
pages 110–119
pages 120–129
Kiran Patil
Carsten Sachse
Judith Zaugg
Janet Thornton
Alex Bateman
Pedro Beltrão
Ewan Birney
Alvis Brazma
Anton Enright
Paul Flicek
Nick Goldman
Gerard Kleywegt
John Marioni
Julio Saez-Rodriguez
Oliver Stegle
Christoph Steinbeck
Sarah Teichmann
Stephen Cusack
Imre Berger
Florent Cipriani
Marco Marcia
José Márquez
Andrew McCarthy
Daniel Panne
Ramesh Pillai
Christiane Schaffitzel
Matthias Wilmanns
Stefan Fiedler
Victor Lamzin
Christian Löw
Rob Meijers
Thomas Schneider
Dmitri Svergun
Philip Avner
Cornelius Gross
Paul Heppenstall
Martin Jechlinger
Christoph Lancrin
Dónal O'Carroll
Rocío Sotillo
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

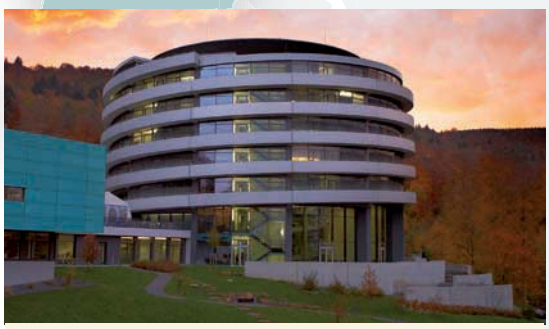
EMBL was founded in 1974 to create a central European laboratory in the emerging field of molecular biology.

It remains the only intergovernmental research organisation in Europe that performs research in the molecular life sciences, and is directly supported by 21 member states, two associate members outside of Europe, and three 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.

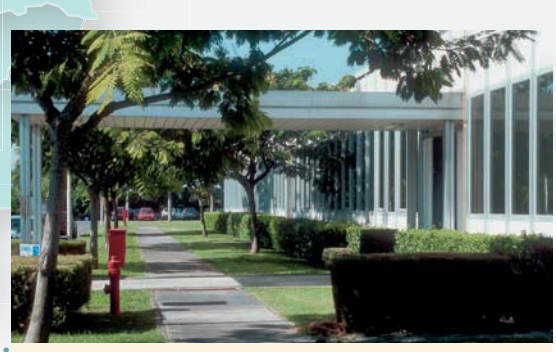
Hinxton

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).

EMBL Heidelberg

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 close to 200 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, interdisciplinary and collaborative workplace

EMBL's staff comprises more than 1700 people from more than 60 different countries – this internationality creates an atmosphere that is creative, interdisciplinary and collaborative, with an unparalleled breadth of expertise and complementary skills.

EMBL is an equal opportunity employer offering attractive conditions and benefits appropriate to an international research organisation. All employees benefit from excellent working conditions, a young and international atmosphere and a high-quality infrastructure of social services.

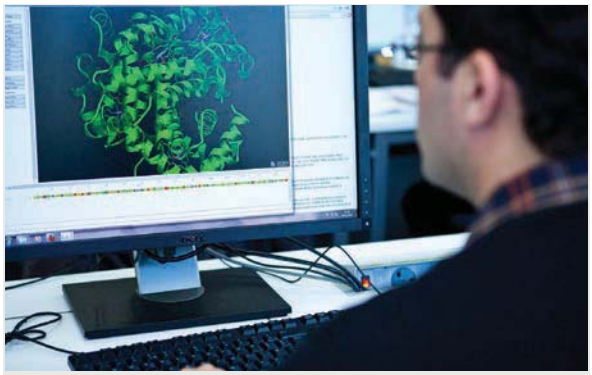
On-site childcare is available at some of EMBL's locations, helping staff to combine professional and family life.

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.

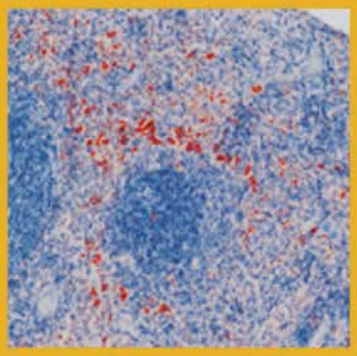
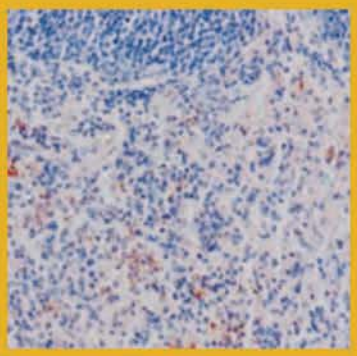
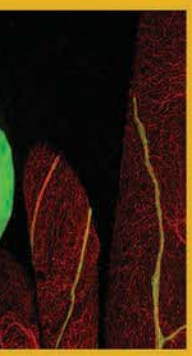
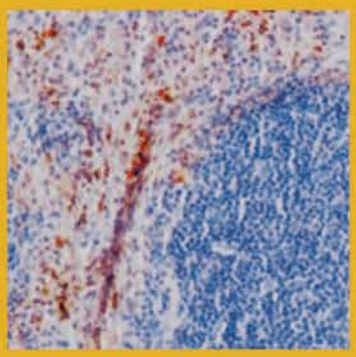
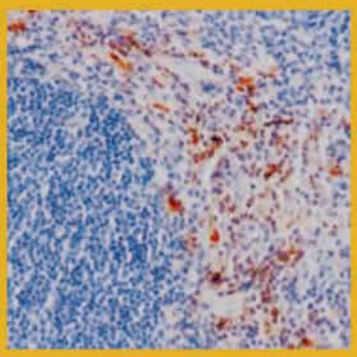
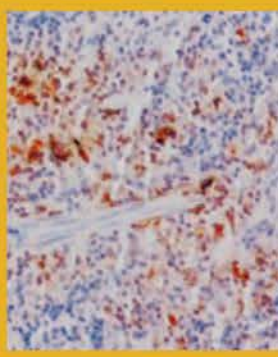
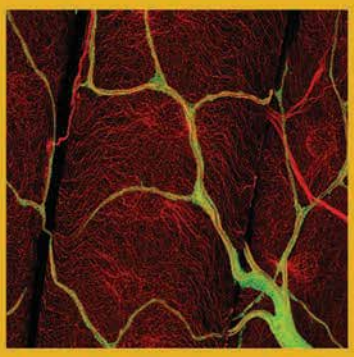
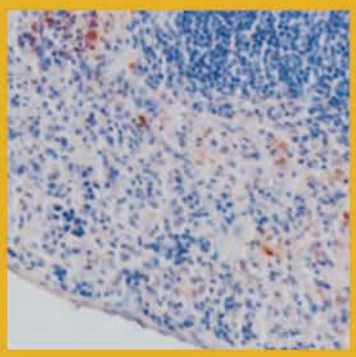
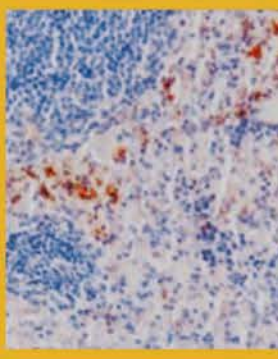
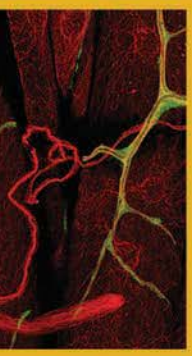
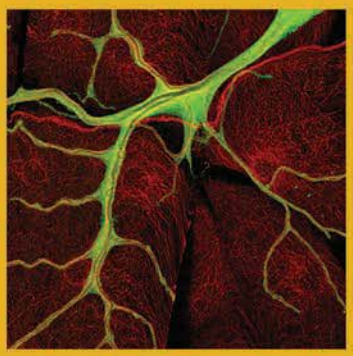
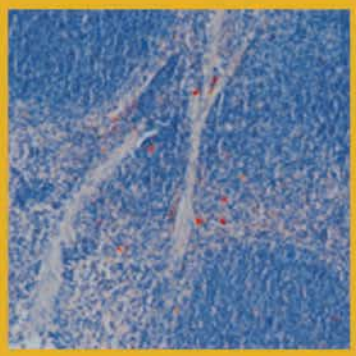
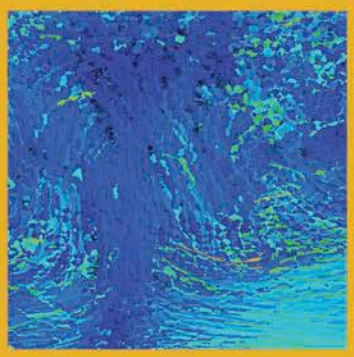
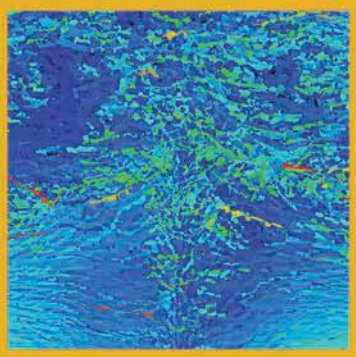
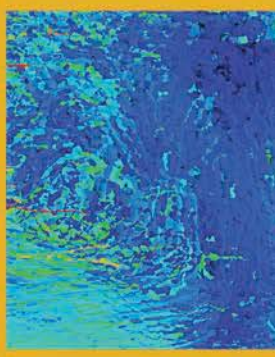
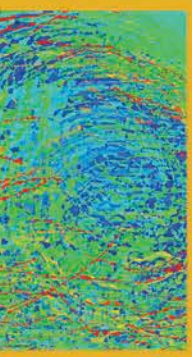
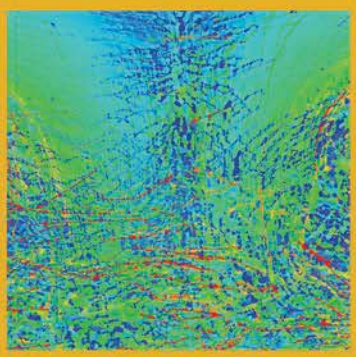
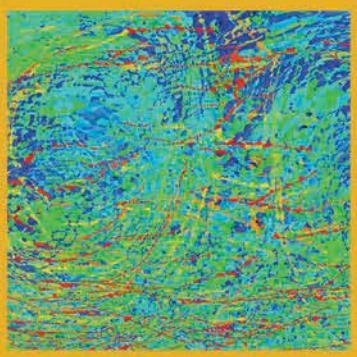
In addition to advanced scientific development, EMBL offers vocational training to improve skills in areas such as coaching, team management and communication. Establishing a good work-life balance is emphasised at every career stage.



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.

EMBL Heidelberg

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 of EMBL 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

Kwon SC, *et al.* (2013) The RNA-binding protein repertoire of embryonic stem cells. *Nat. Struct. Mol. Biol.* 20, 1122-30

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.

Previous and current research

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 connect cell metabolism and gene expression in previously unrecognised ways (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 projects and goals

- To uncover the basic mechanisms underlying protein synthesis and its regulation by miRNAs and 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

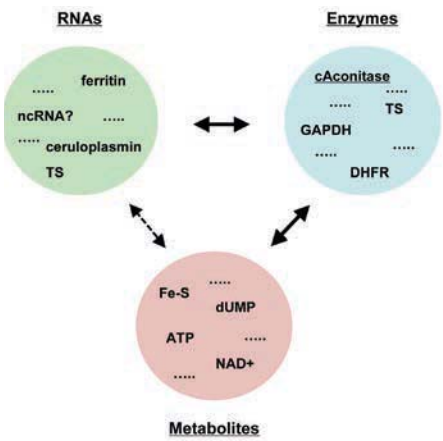


Figure 1: Exploring REM networks.

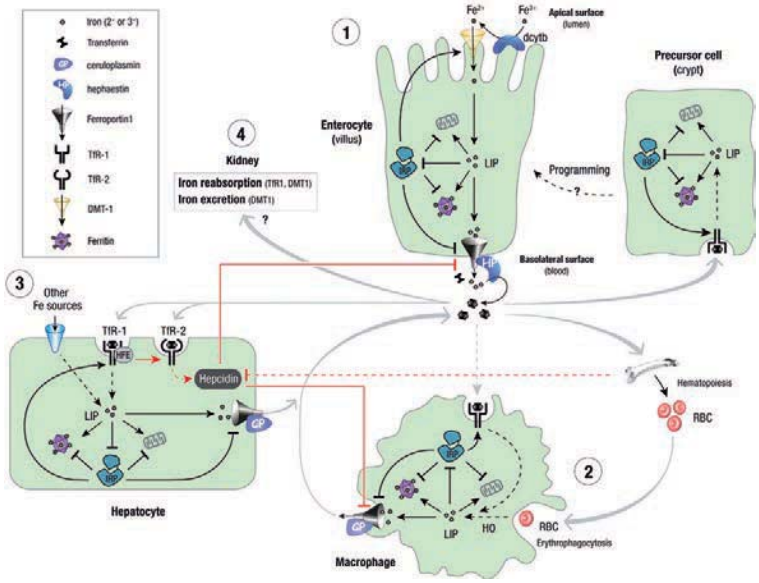


Figure 2: Systems biology of mammalian iron metabolism.

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

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

Banerjee, S. & Leptin, M. (2014) Systemic response to UV involves induction of leukocytic IL-1beta and inflammation in zebrafish. *J Immunol* 193, 1408-15

Jayanandanan N, Mathew R, & Leptin M. (2014) Guidance of subcellular tubulogenesis by actin under the control of a synaptotagmin-like protein and Moesin. *Nat Commun* 5, 3036

Rauzi M, *et al.* (2013) Physical models of mesoderm invagination in *Drosophila* embryo. *Biophys J* 105, 3-10

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

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.

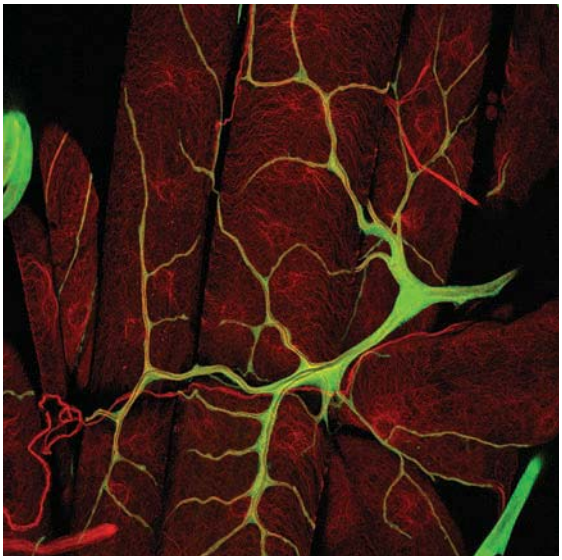


Figure 1: Tracheal cell (green) ramifying on a set of muscles (microtubules stained in red) in a *Drosophila* larva.

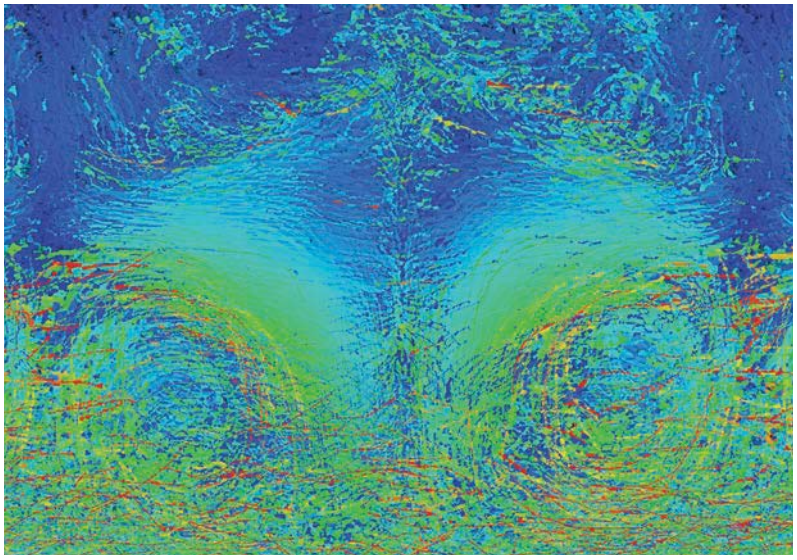
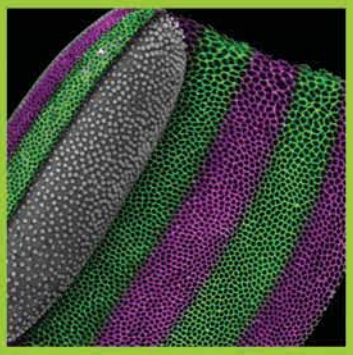
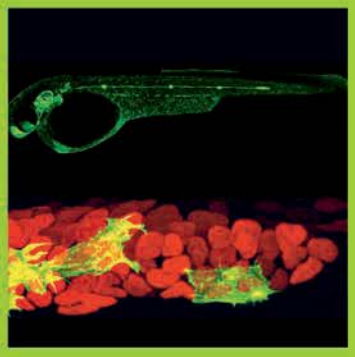
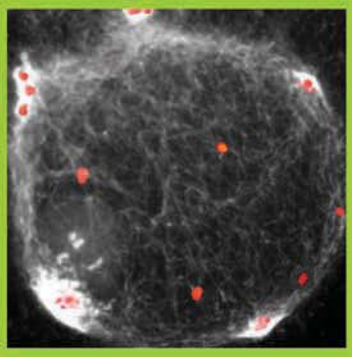
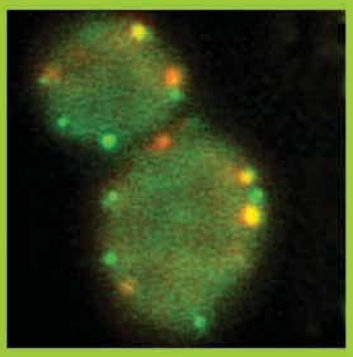
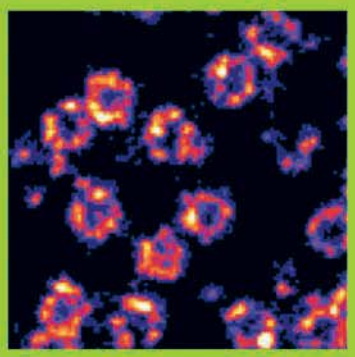
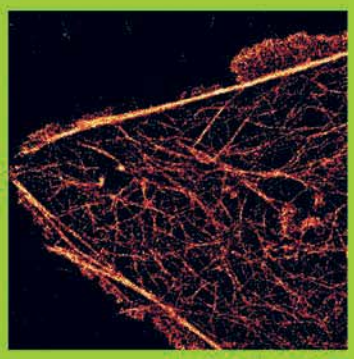


Figure 2: A flat projection of the entire surface of a *Drosophila* embryo in which the position and speed of 6000 cells is followed over a 40 minute period. The head of the embryo is at the top, the center of the image is the ventral midline towards which the lateral cells are moving. Image by Matteo Rauzi.

1974 - 2014 Research Highlights

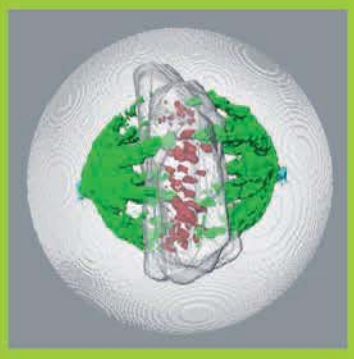


Dynamic reorganisation of Microtubules is required for the repositioning of organelles and cell components as an epithelium forms.
Bacallao R, *et al.* (1989) The subcellular organization of Madin-Darby canine kidney cells during the formation of a polarized epithelium. *J Cell Biol.* 109, 2817-32

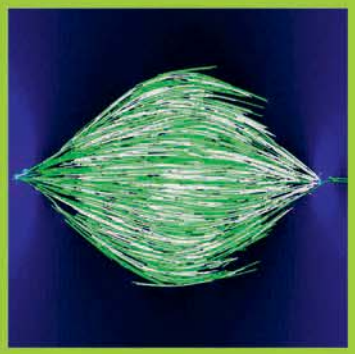
Proteins rab2, rab5 and rab7 are differentially associated with specific cellular compartments dedicated to transporting material into and out of the cell and thus control different stages of those processes.
Chavrier P, *et al.* (1990) Localization of low molecular weight GTP binding proteins to exocytic and endocytic compartments. *Cell* 62, 317-29
AND
Bucci C, *et al.* (1992) The small GTPase rab5 functions as a regulatory factor in the early endocytic pathway. *Cell* 70, 715-28

Systematic screen for all the genes involved in cell division on chromosome III of the worm *C. elegans*, one of the first systematic RNAi screens.
Gönczy P, *et al.* (2000) Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* 408, 331-6

For the meiotic spindle to form, only microtubules, chromatin and associated factors are required – centrosomes and kinetochores are not needed for this process.
Heald R, *et al.* (1996) Self-organization of microtubules into bipolar spindles around artificial chromosomes in *Xenopus* egg extracts. *Nature* 382, 420-5



Decade-old controversy over structure of nuclear pore solved by new super-resolution microscopy method.
Szymborska A, *et al.* (2013) Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging. *Science* 341, 655-8



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, 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 superresolution 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 microscopy methods now allow us to carry out cell biology in developing organisms to understand how cell organisation and collective cell behaviour leads to organ formation.

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, or 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, when progenitor cells differentiate into new cell types, the genetic programme is changed and a reorganisation of cellular architecture takes place, guided by rules that we begin to unravel. Understanding these rules and principles is our challenge in the years to come.

Jan Ellenberg
Head of the Cell Biology and Biophysics Unit

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

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

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

Neumann B, *et al.* (2010) Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature*, 464, 721-7

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

Previous and current research

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 are now studying in live cells – in high throughput – protein function, protein-protein interactions and protein networks during somatic mitosis by automating advanced fluorescence imaging and single molecule techniques, such as fluorescence (cross) correlation spectroscopy.

We also recently determined the positions of various nuclear pore complex (NPC) components and directly resolved the ring-like structure of the NPC by light microscopy, combining stochastic super-resolution microscopy (SRM) with single particle averaging (figure 1). Currently we elucidate the assembly mechanism of the NPC and chromatin dynamics over the cell cycle.

By complete kinetochore tracking we demonstrated that meiotic spindle assembly and asymmetric positioning rely on novel mechanisms and that meiotic chromosome biorientation is highly error prone. We are now developing gentle light-sheet-based imaging systems for high-throughput imaging of mouse oocytes and embryos to allow systematic molecular analysis of meiosis and early embryonic mitosis.

Future projects and goals

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 as well as 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 and 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 to establish a physiological molecular model for early mammalian development and infertility.

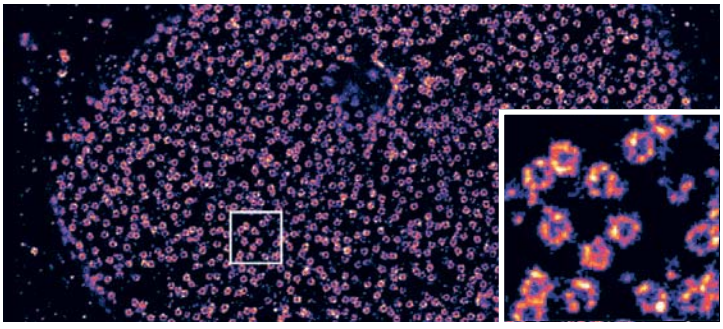


Figure 1: Super-resolution image of a nuclear pore labelled with an antibody against one protein component of the *NPC* (Nup160). An image is gradually built up by localising the centers of individual fluorophores switching between light-emitting and dark state (Szymborska A, *et al,* 2013).

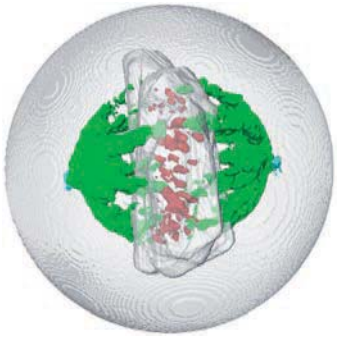


Figure 2: A 3D metaphase model reconstructed from confocal images of HeLa cells. The colours identify different elements/proteins important for mitosis (centrosome, cyan; chromatin, green; kinetochores/centromeres, red).

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

Previous and current research

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.

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.

Future projects and goals

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.

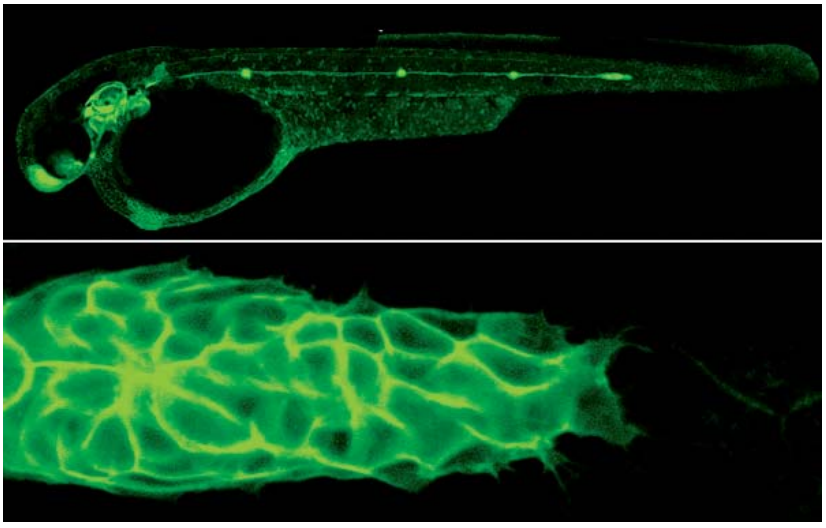


Figure 1: The zebrafish migrating lateral line organ allows collective migration to be easily studied *in vivo*.

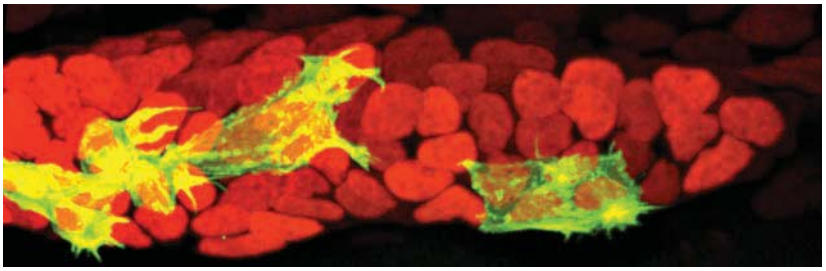


Figure 2: Visualising actin dynamics (LifeAct-GFP) within migrating primordium.

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.

Previous and current research

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 projects and goals

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.

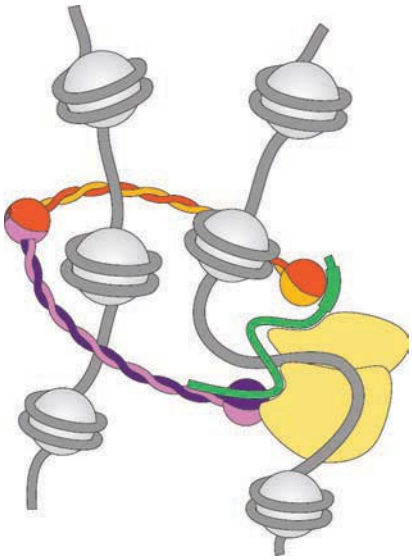


Figure 1: Model for the organisation of mitotic chromosomes by condensin rings.

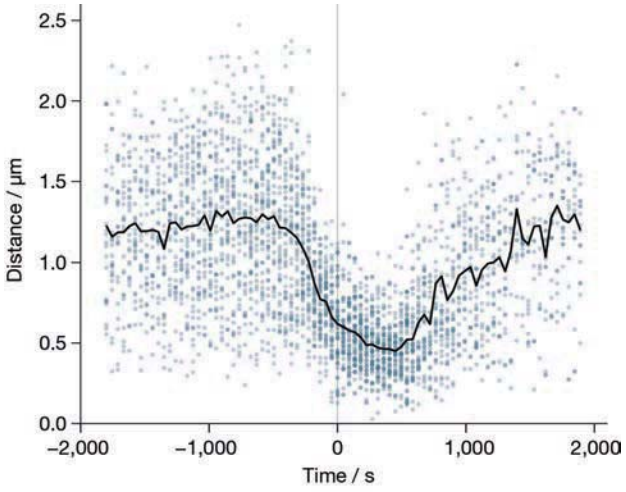


Figure 2: The live-cell chromosome condensation assay tracks the distances between two fluorescently labelled chromosome loci over time. Alignment of a large number of single cell tracks (circles) to the time of anaphase onset (t = 0) generates an average distance plot (line) as a quantitative read-out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

Previous and current research

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).

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

Future projects and goals

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.

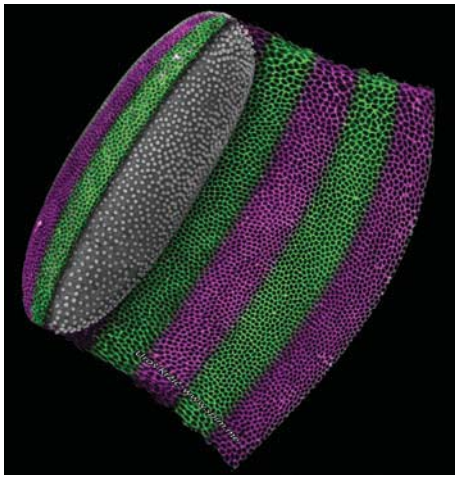


Figure 1: MuVi-SPIM image of a *Drosophila* embryo. Eight views were fused to yield an *in toto* reconstruction of the embryo (one side membrane unrolled). The high speed of the microscope enables a detailed reconstruction of cell lineage and shape changes over extended periods of development.

Dynamics of membrane trafficking

Marko Kaksonen

PhD 2002, University of Helsinki, Finland.

Postdoctoral research at the University of California, Berkeley, USA.

Group leader at EMBL since 2006.

SELECTED REFERENCES

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

Skruzny M, *et al.* (2012). Molecular basis for coupling the plasma membrane to the actin cytoskeleton during clathrin-mediated endocytosis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 2533-42

Brach T, Specht T, & Kaksonen M. (2011). Reassessment of the role of plasma membrane domains in the regulation of vesicular traffic in yeast. *J. Cell. Sci.* 124, 328-37

Using budding yeast as a model, the Kaksonen group wants to understand how complex molecular machineries drive vesicle trafficking.

Previous and current research

Many biological processes at the cellular level are based on complex networks of macromolecular interactions. These networks have a modular organisation, where the different modules form dynamic molecular machines that drive processes such as signalling, cell motility, cytokinesis, and vesicle trafficking. Our group's long-term goal is to contribute to the understanding of the general principles governing the assembly and function of these supramolecular machines.

More specifically, we are interested in the formation of cargo-loaded transport vesicles, such as endocytic vesicles, whose formation is driven by highly dynamic molecular machinery composed of more than 50 different protein species and of several thousand individual protein molecules. We aim to understand the processes that regulate the assembly of the endocytic machinery, the recruitment of the cargo molecules, and the selection of the location and timing of endocytic events in the cell.

Our main experimental organism is the budding yeast, *Saccharomyces cerevisiae*. In our studies we use quantitative live-cell imaging methods – such as particle tracking, FRAP, FCS/FCCS, and high-throughput microscopy – in combination with powerful yeast genetics. We also use correlated light and electron microscopy to gain nanometer-scale information about the endocytic structures, and biochemistry to characterise protein-protein and protein-lipid interactions.

Future projects and goals

We are interested in the mechanisms that initiate the assembly of the endocytic machinery and regulate the precise timing of the sequential stages of the assembly. The spatial distribution of the endocytic events is tightly coupled to the cell cycle and to the overall polarity of the cell. The spatially regulated initiation of endocytic events is critical for determining the cellular distribution of endocytosis.

We are also studying the mechanisms of selective recruitment of cargo molecules into the endocytic vesicles. The recruitment of cargo proteins is tightly regulated by a family of endocytic adaptors. We want to learn how this adaptor system integrates environmental and intracellular signals in deciding which cargoes to recruit. In yeast, endocytosis is strictly dependent on actin polymerisation, but the mechanisms by which actin drives vesicle budding are not well understood. We are currently studying the molecular basis of the coupling between the actin cytoskeleton and the endocytic membrane. We have also started to investigate the evolution of the membrane–actin coupling in animals and fungi using a phylogenetic comparative approach.

The core membrane trafficking events, such as the clathrin-mediated endocytosis, are elemental cellular processes that are involved in multiple biological phenomena ranging from cell polarisation to neural plasticity. As most of the yeast trafficking proteins are widely conserved in eukaryotes, we believe that mechanisms that we unravel in yeast cells will be applicable to eukaryotes in general.

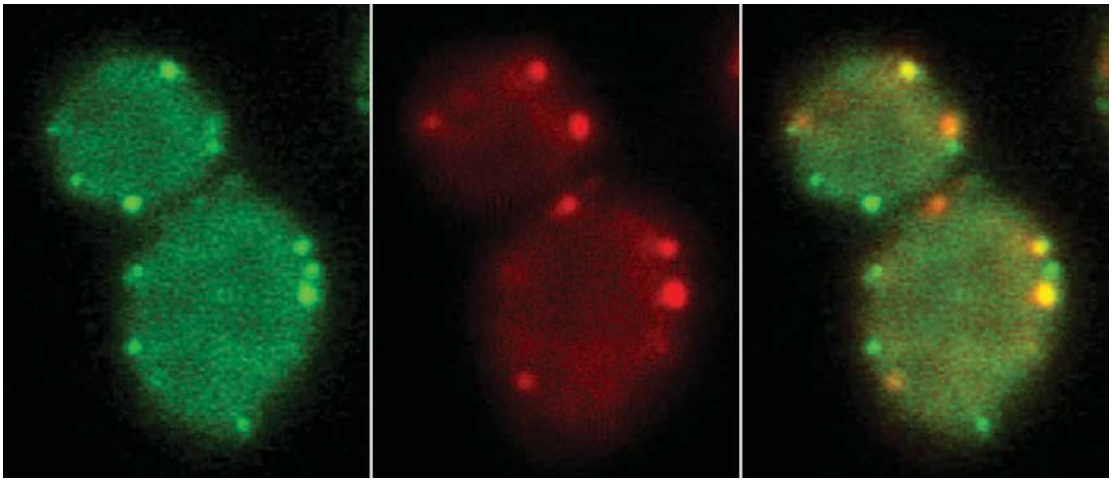


Figure 1: A yeast cell expressing fluorescently labelled endocytic proteins. The first two images show a coat protein Sla1 (green) and an actin-binding protein Abp1 (red). The last image shows both channels merged. The spots at the cell surface reveal the transient accumulation of the proteins at endocytic sites during vesicle budding.

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

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

Field CM & Lénárt P. (2011) Bulk cytoplasmic actin and its functions in meiosis and mitosis. *Curr. Biol.* 21, 825-30

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

Previous and current research

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.

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

Future projects and goals

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.

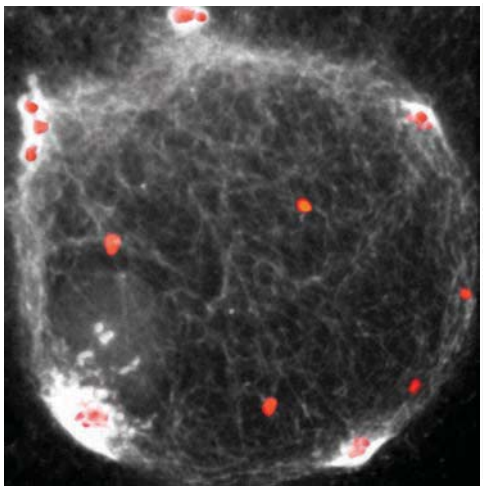


Figure 1 (far left): The actin filament network (gray) embedding the chromosomes (red).

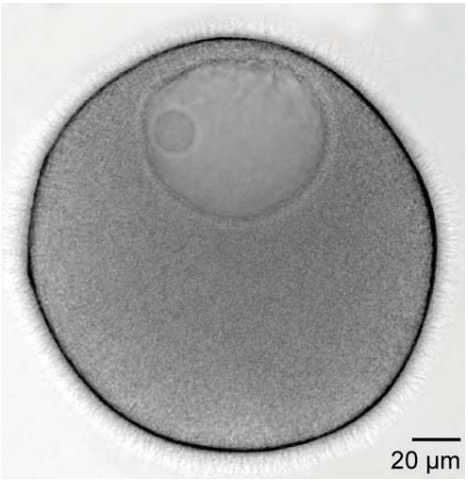


Figure 2 (near left): Transparent starfish oocytes are uniquely suited for imaging meiotic divisions.

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.*

Previous and current research

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 projects and goals

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.

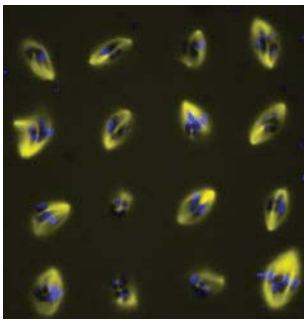
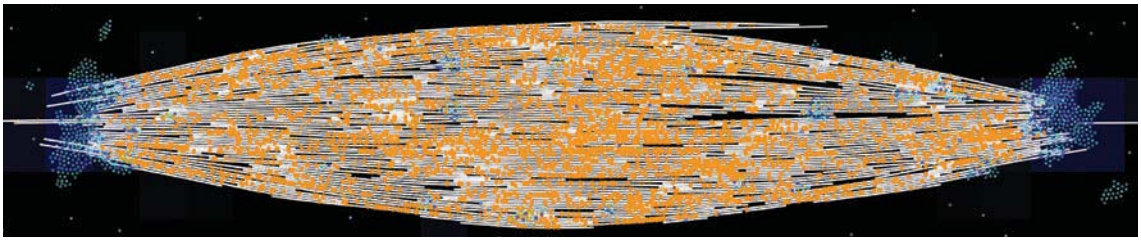


Figure 1: An array of mitotic spindles obtained *in vitro* with Xenopus laevis egg extracts (Dinarina *et al.*, 2009).

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 REFERENCE

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

Previous and current research

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. This reductionist approach still allows us to quantitatively and robustly discriminate pluripotency, cancer and lineage commitment. More importantly, it suggests 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.

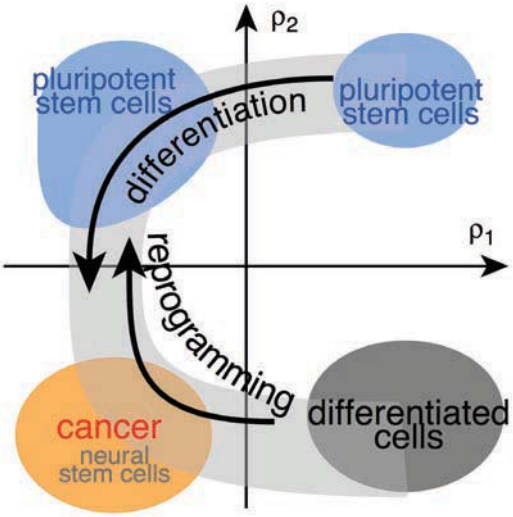
The Neveu group takes an integrated systems biology approach to investigate the molecular changes that determine what a stem cell becomes.

Future projects and goals

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?



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.

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 REFERENCES

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

Tängemo C, *et al.* (2011) A novel laser nanosurgery approach supports de *novo* Golgi biogenesis in mammalian cells. *J. Cell. Sci.* 124, 978-87

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

Previous and current research

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.

Future projects and goals

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.

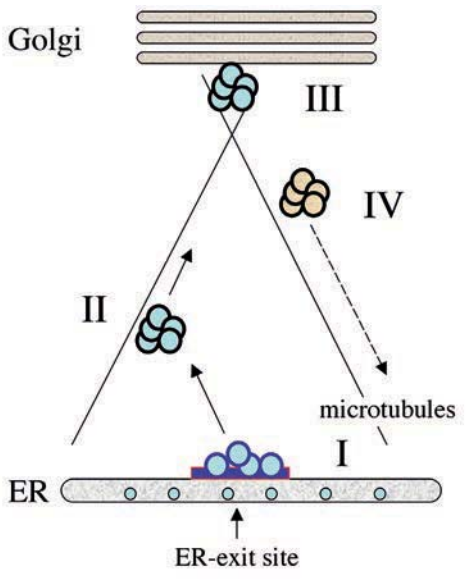


Figure 1: The four steps involved in ER to Golgi transport in mammalian cells. (I): Biogenesis of COPII coated vesicles occurs at specialised ER exit sites of the ER.(II): COPII vesicles homotypically fuse to form larger vesicular tubular transport carriers (VTCs) that are transported to the Golgi complex 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.

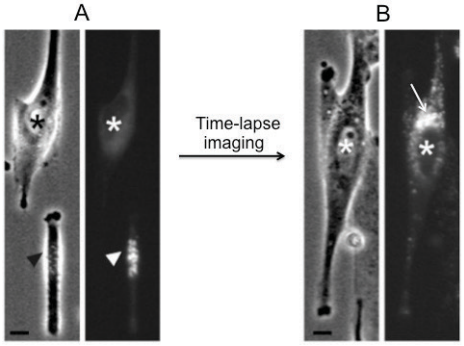


Figure 2: (A) Cells are cut by laser nano-surgery to generate a Golgiless karyoplast and Golgi containing Golgiplasts (arrowhead). Karyoplasts are then followed by time-lapse microscopy to monitor *de novo* Golgi biogenesis in living cells (B). The arrowhead points to the Golgi-like structure reforming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

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

Previous and current research

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.

The Ries group develops cutting -edge super-resolution microscopy methods to determine structures of multi-protein assemblies in the cellular context.

Future projects and goals

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.

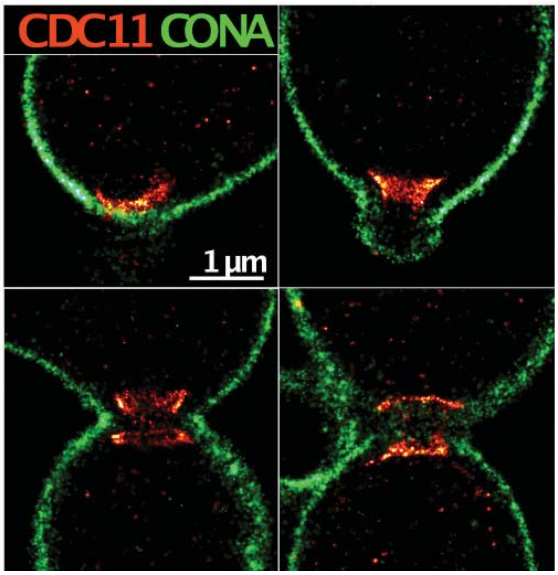


Figure 1: Dual-colour super-resolution images of Cdc11 (red) and the cell-wall marker ConA (green) show the formation and disassembly of the Cdc11 ring.

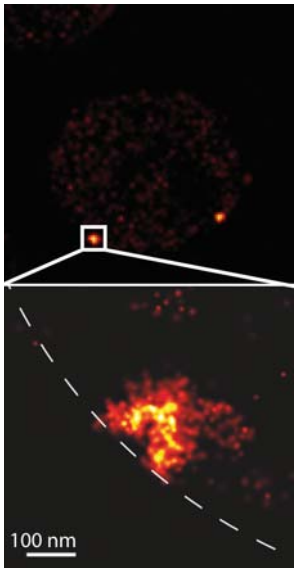


Figure 2: Actin in yeast. Yeast expressing Abp1-mMaple imaged by localisation microscopy. Five fixed example sites at different endocytic time points.

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

Feng S, *et al.* (2014) A rapidly reversible chemical dimerizer system to study lipid signaling in living cells. *Angew. Chem. Int. Ed. Engl.* 53, 6720-3

Gehrig S, *et al.* (2014) Lack of neutrophil elastase reduces inflammation, mucus hypersecretion and emphysema, but not mucus obstruction, in mice with CF-like lung disease. *Am. J. Respir. Crit. Care Med.* 189, 1082-92

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

Laketa V, *et al.* (2014) PIP(3) induces the recycling of receptor tyrosine kinases. *Sci Signal* 7, ra5-ra5

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.

Previous and current research

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 projects and goals

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.

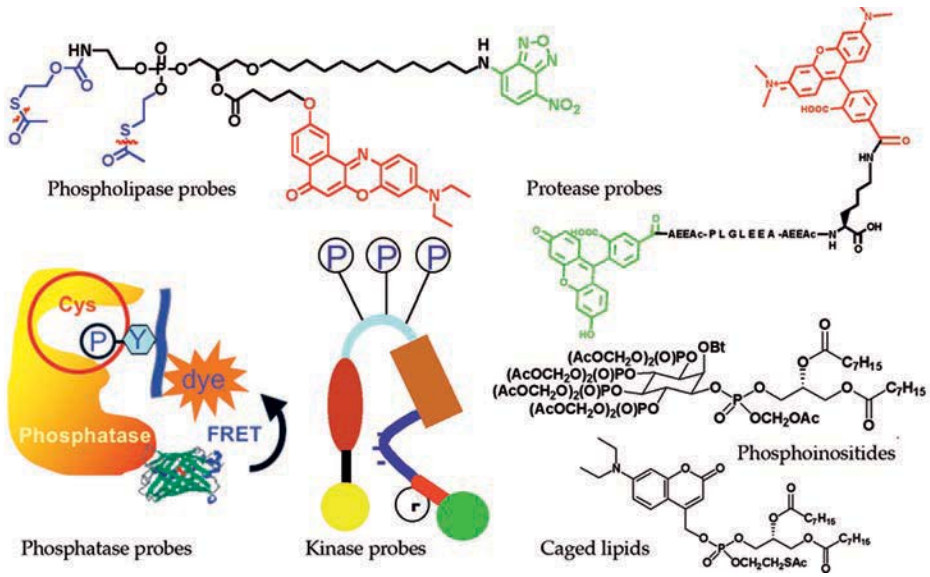


Figure 1: Several reporter and modulator molecules have been developed in our lab, including: small molecule sensors for lipases and proteases; genetically encoded reporters for kinase and phosphatase activities; membrane-permeant and photoactivatable lipid molecules; and lipid derivatives that can be 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.* (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

Kolotuev I, *et al.* (2013) A pathway for unicellular tube extension depending on the lymphatic vessel determinant Prox1 and on osmoregulation. *Nat. Cell Biol.* 15, 157-68

Previous and current research

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, and one that reaches 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).

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

Future projects and goals

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 transmission electron microscopy (TEM) and scanning electron microscopy (SEM) tomography. Our team, in collaboration with other research teams at EMBL 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.

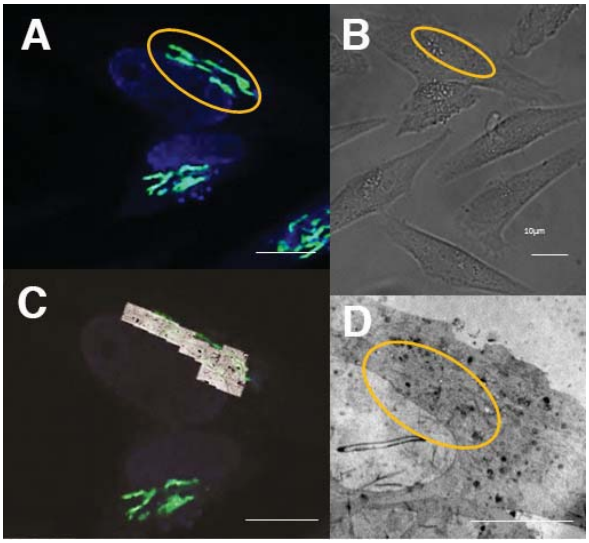


Figure 1: CLEM on cultured cells – A and B: the Golgi apparatus is tagged with a GFP marker and imaged by light microscopy (in collaboration with the Pepperkok Team). Using a CLEM workflow, the same cell is tracked and the region of interest studied by electron microscopy.

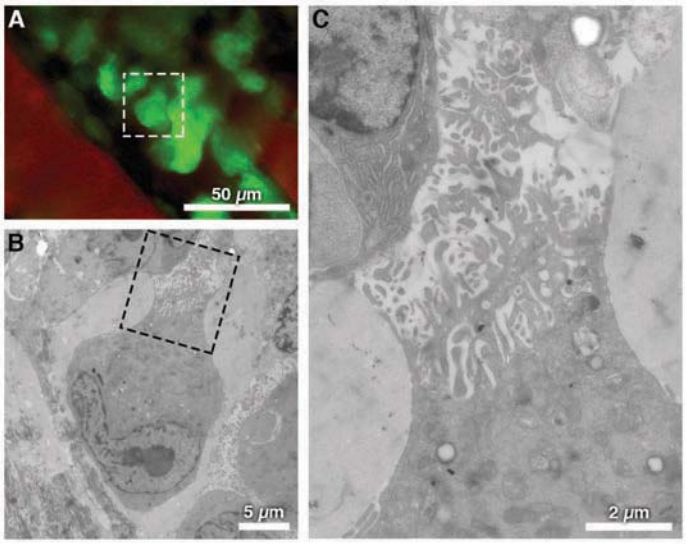
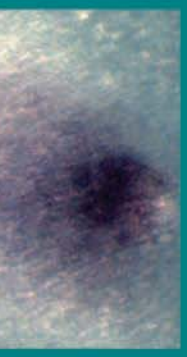
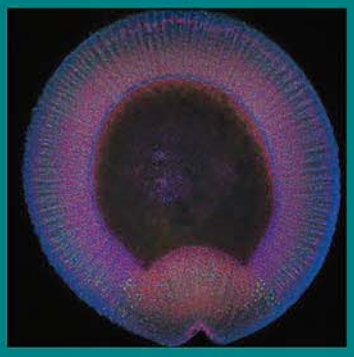
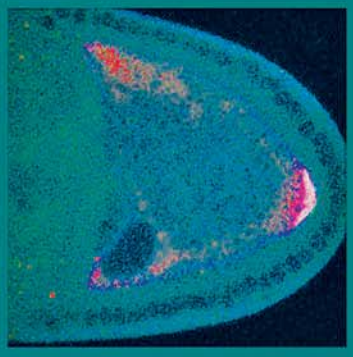
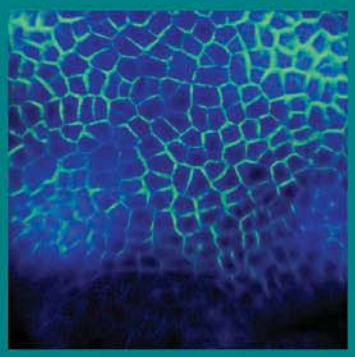
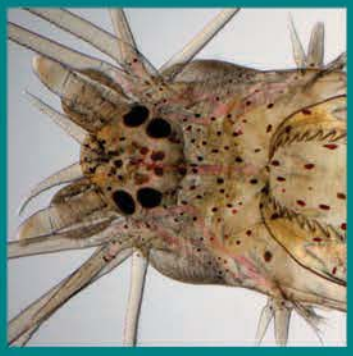
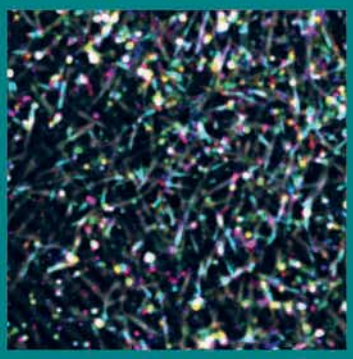


Figure 2: CLEM on cancer cells – from intravital imaging to ultrastructure. Fluorescent cancer cells were injected in mouse skin and imaged with multi photon microscopy (J. Goetz and L. Mercier, Inserm, France), enabling the visualisation of both the invasive cells (green) and the vasculature (red, stained with Nile blue) (A). (B and C) Following EM sample preparation, the cell of interest can be retrieved and imaged at high resolution with transmission electron microscopy.


1974 - 2014 Research Highlights

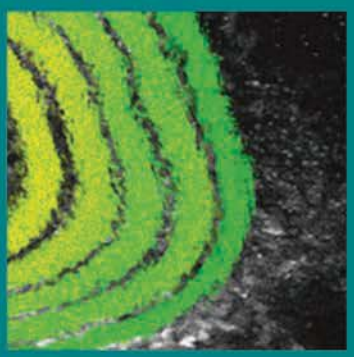
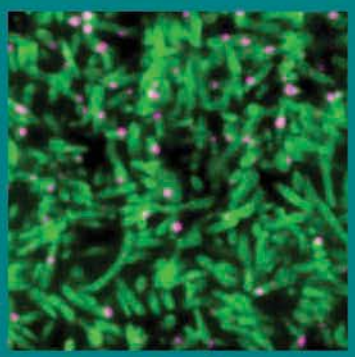

Nobel-prize winning work: identification of 15 genes that control how a fruit fly's body parts are initially specified; similar genes are later shown to exist in humans.
Nüsslein-Volhard C & Wieschaus E (1980) Mutations affecting segment number and polarity in *Drosophila*. *Nature* 287, 795-801

At least two oncogenes must act in concert to cause leukaemia.
Beug H, *et al.* (1984) Ts mutants of E26 leukemia virus allow transformed myeloblasts, but not erythroblasts or fibroblasts to differentiate at the nonpermissive temperature. *Cell* 39, 579-88
AND
Kahn P, *et al.* (1986) v-erbA cooperates with sarcoma oncogenes in leukemic cell transformation. *Cell* 45, 349-56

The sequence of Hox genes along a chromosome corresponds to the roles of the genes in specifying the vertebrate body plan: the first genes define which part becomes the head, the next group define the torso, and so on.
Izpisua-Belmonte JC, *et al.* (1991) Murine genes related to the *Drosophila* AbdB homeotic genes are sequentially expressed during development of the posterior part of the body. *The EMBO journal* 10, 2279

Linking cell proliferation and cell death: in the fruit fly embryo, a microRNA called bantam controls cell proliferation and turns off a gene that promotes cell death.
Brennecke J, *et al.* (2003) bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in *Drosophila*. *Cell* 4, 25-36

First complete developmental blueprint of a vertebrate: using the DSLM microscope they developed, EMBL scientists tracked the movements of all the cells in a zebrafish embryo for the first 24 hours of its life.
Keller PJ, *et al.* (2008) Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. *Science* 322, 1065-9

EMBL Heidelberg

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

RNA localisation and regulation 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

Gaspar I, *et al.* (2014) Klar ensures thermal robustness of *oskar* localization by restraining RNP motility *J. Cell Biol* 206, 199-215

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

Hachet O & Ephrussi A. (2004) Splicing of *oskar* RNA in the nucleus is coupled to its cytoplasmic localization. *Nature* 428, 959-63

The Ephrussi group aims to understand the mechanisms underlying RNA transport and localised translation – fundamental processes that promote the functional polarisation of cells during development.

Previous and current research

Intracellular RNA transport coupled with localised translation is a powerful and widespread mechanism that promotes the functional polarisation of cells, from yeast to man. Asymmetric localisation of messenger RNAs within cells has key roles in cell fate decisions, cell migration, cell morphology and function. mRNA targeting is particularly evident in large cells, such as eggs and neurons, where it allows rapid and localised deployment of protein activities in response to extrinsic signals.

An ideal model for the study of RNA transport is the large *Drosophila* oocyte, in which asymmetrically localised cell fate determinants specify the body axes and pattern the future embryo. During oogenesis, mRNAs encoding these embryonic axis determinants 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 cell. We use these RNAs as models to understand how mRNA localisation and translational control are regulated in space and time.

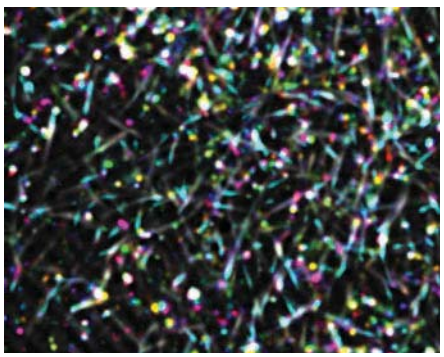
One RNA of particular interest is *oskar*, which encodes the posterior determinant of *Drosophila*. Oskar protein is uniquely endowed with the capacity to induce germ cell formation in the embryo, which it does by nucleating formation of the germ plasm and its germline determining RNP complexes, called polar granules. How *oskar* mRNA is transported and anchored at the posterior pole of the oocyte and its translation regulated is one focus of research in the lab.

We are also investigating the roles of other classes of RNAs, including long non-coding RNAs and piRNAs, and of non-canonical RNA binding proteins, in *Drosophila* embryonic development and neurogenesis. *Drosophila*, with its exceptional genetic tools, is also well suited to biochemical and cell biological investigation, including live imaging, of the processes of cell polarisation, mRNA localisation and translational control.

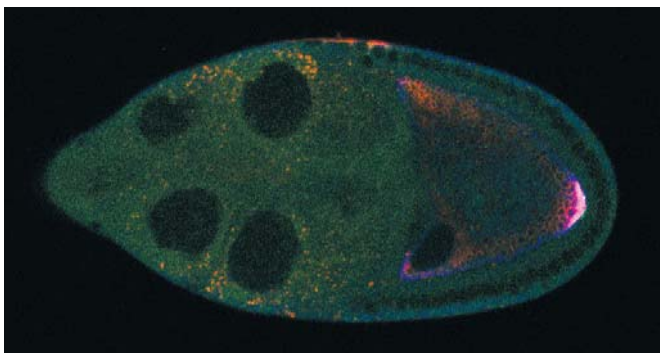
Future projects and goals

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

- Polarisation of the cytoskeleton.
- The roles and regulation of cytoskeletal motors in RNA localisation.
- Assembly of transport-competent RNPs: the *cis*-acting RNA targeting elements and interacting proteins, how they assemble and associate with motor proteins.
- Translational regulation of localised mRNAs.
- Germ plasm assembly and function.



oskar mRNA on the move. Time projection of a squash of ooplasm from a stage 9 oocyte imaged with TIRF microscopy. *oskar* mRNA (labelled with MS2-MCPGFP, shown in rainbow colours) utilises microtubules (labelled with mCherrya1-tubulin and EB1-Cherry, shown in gray with cyan tips, indicating plus ends) to take fast, long linear runs.



A *Drosophila* egg-chamber, showing colocalisation of *oskar* mRNA, Staufen protein and a microtubule polarity marker at the posterior of the oocyte.

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

Previous and current research

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.

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

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 projects and goals

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.

SELECTED REFERENCES

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

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

Previous and current research

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.

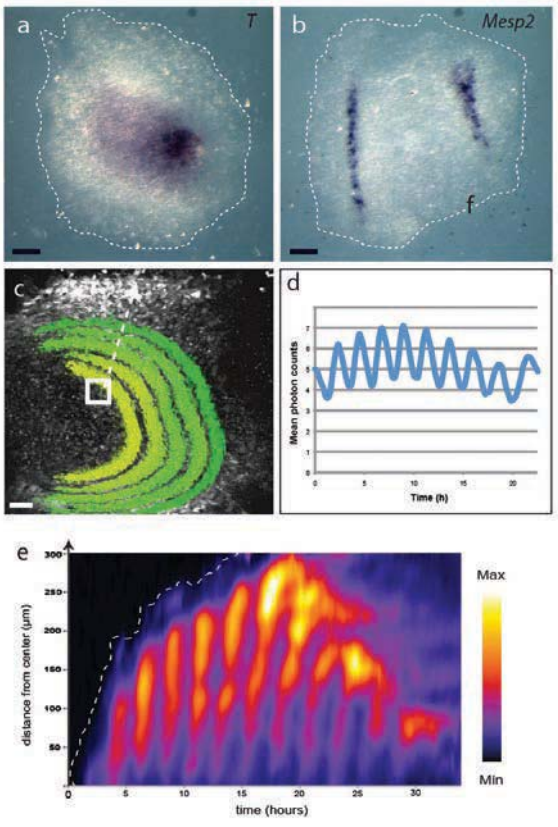


Figure 1: *Ex vivo* cell culture model for mesoderm patterning and oscillations. a, b) Primary mesoderm cells retain undifferentiated PSM identity in the centre of the culture, before initiating a differentiation program in the periphery. c) Snapshot of time-series using fluorescent lunatic fringe reporter 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.

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 projects and goals

- 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.

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

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

De Renzis S, *et al.* (2006) Dorsal-ventral pattern of Delta trafficking is established by a Snail-Tom-Neuralized pathway. *Dev. Cell* 10, 257-64

Previous and current research

Tissue morphogenesis is triggered by shape changes in single cells or groups 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). We want to understand how membrane trafficking and cytoskeletal dynamics are regulated during morphogenesis and how this, in turn, impacts on specific cell and tissue behaviour. To this end, we combine high-resolution imaging methods with genetics and biochemistry using the early Drosophila embryo as model system (see figure 1 and online video).

We have recently developed a modified form of total internal reflection fluorescence (TIRF) microscopy to follow apical surface dynamics in live embryos with unprecedented spatio-temporal resolution. Using this approach we have identified a novel endocytic pathway controlling the morphology of the apical surface during epithelial morphogenesis (figure 2), thus demonstrating for the first time that endocytosis directly controls cell and tissue shape. We are now using similar high-resolution imaging methods in combination with electron tomography to study the involvement of endocytosis in the regulation of cell signalling and membrane remodelling during tissue morphogenesis.

We are also interested in characterising the impact of lipid metabolism during morphogenesis. Using a chemical genetic approach we uncovered an exciting link between so-called 'lipid induced phenotypes' and developmental gene activities underlying the regulation of cell and tissue shape. Finally, we are developing new optogenetic tools to control protein activity with light during tissue morphogenesis with high spatio-temporal precision.

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.

Future projects and goals

Using a combination of imaging, genetics and optogenetic 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.

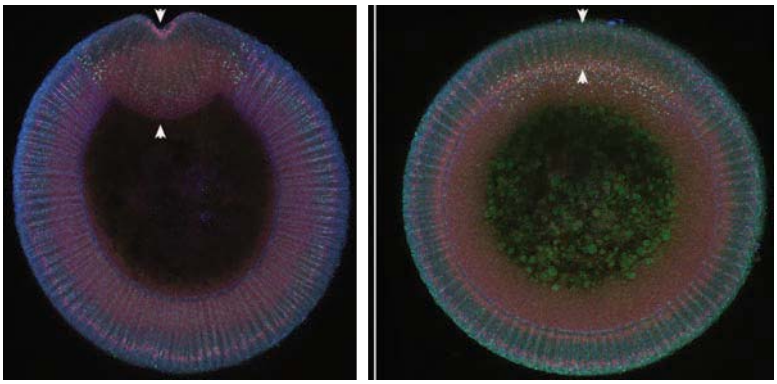


Figure 1: Cross-section of a Drosophila embryo during late cellularisation (left panel) and ventral furrow formation (right panel) stained with antibodies against b-catenin (red), Notch (green) and Delta (blue). Embryos are oriented with the ventral side facing down and dorsal up. Cells on the ventral side are elongated along the apico-basal axis (left panel arrowhead) compared to their dorso-lateral neighbours. Endocytosis of Notch and Delta is specifically up-regulated in ventral cells during invagination.

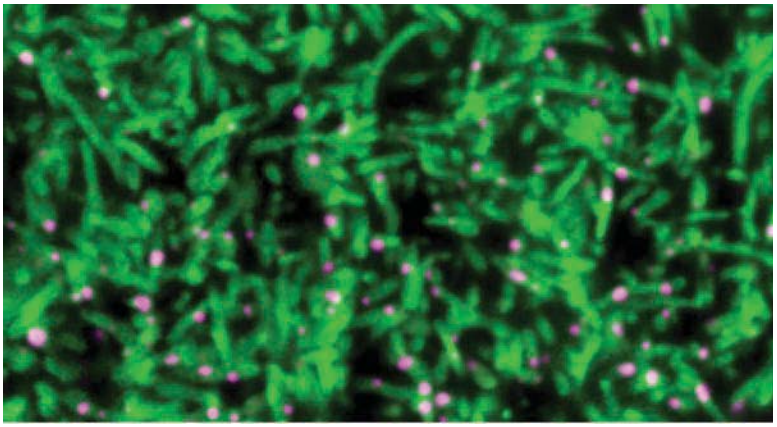


Figure 2: Application of TIRF-M imaging to early *Drosophila* embryo allowed uncovering a prominent endocytic pathway controlling the morphology of the apical surface during epithelial development. Rab5 endosomes (purple), plasma membrane (green).

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.*

Previous and current research

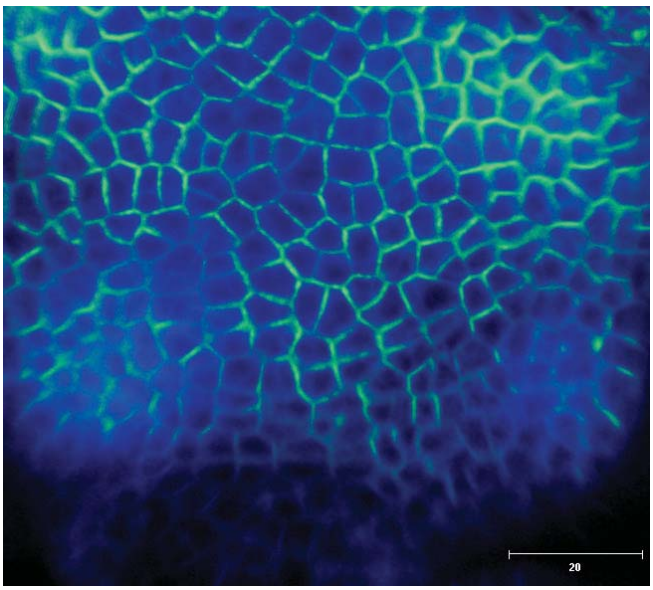
In addition to providing us with the air we breathe, the food we eat and much of the energy and materials we use, plants exhibit a unique beauty associated with their strikingly symmetrical patterns of development. Multicellularity also evolved independently in plants giving us an opportunity to compare 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 such questions, since organ formation involves the coordination of cell polarity, gene expression and morphogenesis. Our recent 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. Interestingly, we have also found that the patterns of cell polarity associated with organogenesis correlate spatially with particular patterns of gene expression normally associated with the dorsal and ventral cell types of lateral organs. This raises the question of whether these expression domains play a causal role in organising cell polarity patterns and, in turn, whether these polarity patterns influence dorsiventral gene expression. This rich interplay is one of our prime focuses.

Future projects and goals

Our ERC-funded project focuses on the establishment and function of dorsiventral boundaries. Previously we developed confocal-based methods for imaging growing plant tissues, enabling us to obtain dynamic high-resolution data for protein localisation and gene expression (making full use of the different GFP spectral variants). By incorporating such data directly into mathematical models we aim to develop an explicit understanding of the complexity underlying patterning processes associated with dorsiventral cell type specification. Our main questions include: How do dorsiventral gene expression boundaries regulate organ morphogenesis and positioning – for example cell polarity patterns? How are dorsiventral gene expression boundaries established and regulated?

Like animals, plants can also re-pattern their tissues in response to wounding. Wounding also causes dramatic changes to dorsiventral patterning, although the mechanisms by which this occurs remain unknown. Our recent results show that cell polarity patterns respond dramatically to wounds, suggesting this cellular response may play an important role in tissue reorganisation. We aim to investigate this possibility using two-photon induced ablation and DSLM microscopy.



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 in mammalian development

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

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

Courtois A, *et al.* (2012) The transition from meiotic to mitotic spindle assembly is gradual during early mammalian development *J. Cell Biol.* 198, 357-70

Dietrich JE & Hiiragi T (2007) Stochastic patterning in the mouse pre-implantation embryo. *Development* 134, 4219-31

Previous and current research

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). What 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 could establish 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 stabilizes 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 will allow 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 projects and goals

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 the mouse embryo;
- molecular characterisation of the *de novo* formation of epithelial polarity;
- understanding the role of cell mechanics in embryogenesis;
- identification of the trigger and mechanism for centriole biogenesis *in vivo*.

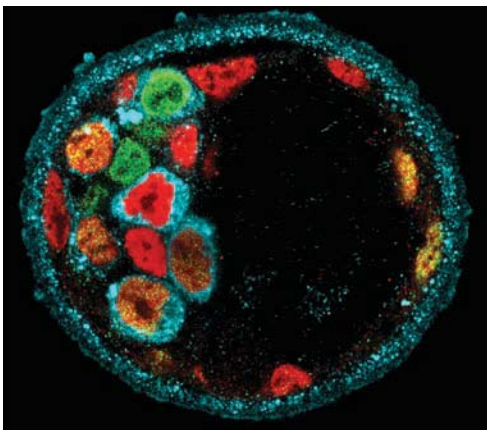


Figure 1: Molecular heterogeneity during mouse blastocyst patterning. Cells expressing Nanog (green), Gata6 (red) or Serpinh1 (blue).

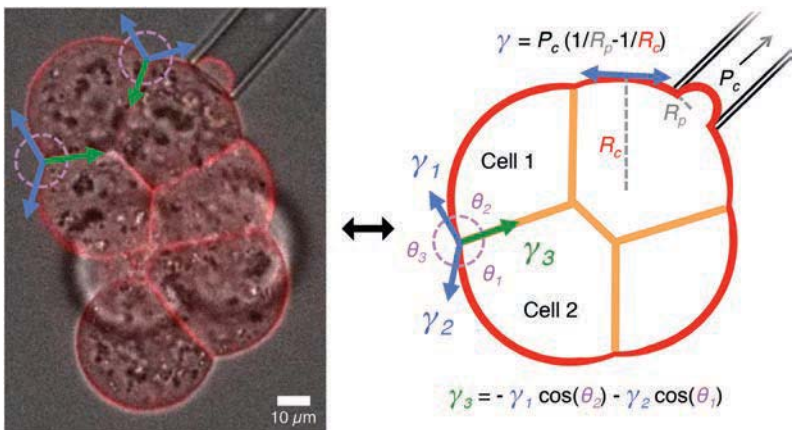


Figure 2: Mapping of surface tensions in a developing mouse embryo.

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.

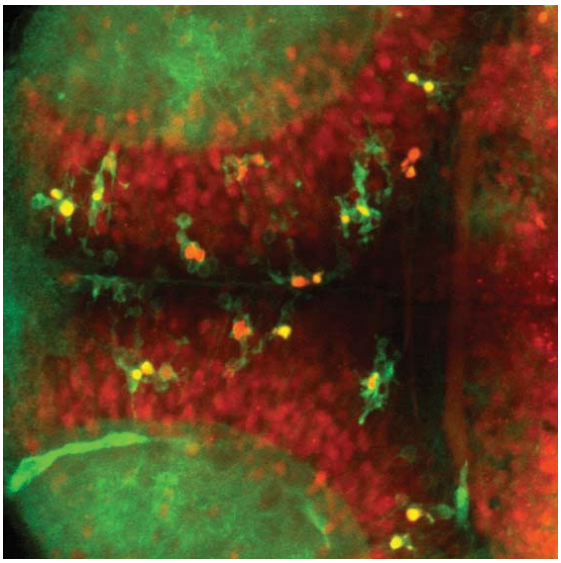
Previous and current research

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 projects and goals

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.

Gene regulation and genome architecture

Francois Spitz

PhD 1997, Institut Cochin de Génétique Moléculaire, Paris.

Postdoctoral research at the University of Geneva.

Group leader at EMBL since 2006.

SELECTED REFERENCES

Symmons O, *et al.* (2014) Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* 24, 390-400.

Andrey G, *et al.* (2013) A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* 340, 1234167

Marinić M, *et al.* (2013) An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Dev. Cell* 24, 530-42

Ruf S, *et al.* (2011) Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nat. Genet.* 43, 379-86

Previous and current research

The patterning of the embryo and the specification of its different cell types are driven by the implementation of cell-specific gene expression programs. In vertebrates, the *cis*-acting elements that regulate transcription can be located hundreds of kilobases away from the genes they control, particularly for genes with important functions during development. Because of this, the genome appears to be composed of intermingled arrays of unrelated genes and *cis*-regulatory elements. Therefore, the mechanisms that regulate enhancer-promoter interactions are essential to transform this apparent genomic and regulatory conundrum into gene- and tissue- specific expression programs. Recent data reveal that genomic loci adopt specific chromatin structures and conformations in the nuclei of different cell types, correlating with differential gene activity. Yet, the *cis*-acting genomic elements that determine how a genomic locus folds into specific structural and regulatory architectures, and the precise roles of the chromatin, protein-complexes and non-coding RNAs suggested to contribute to this process, are still unclear.

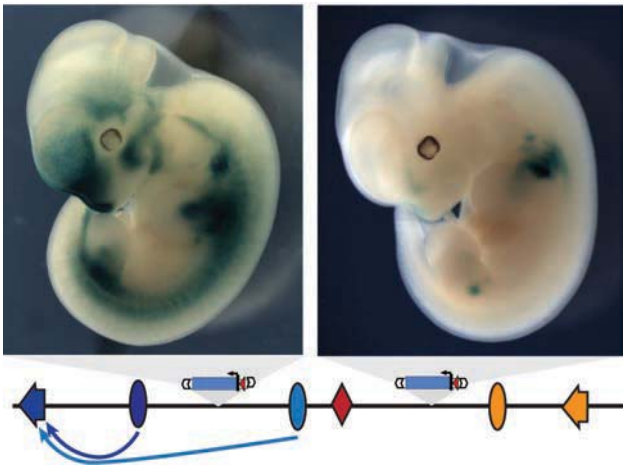
The Spitz group aims to understand how the intricate distribution of regulatory elements along the genome is transformed into specific gene expression profiles.

Our lab has developed several experimental approaches to explore the regulatory architecture of the mouse genome and characterise functionally the mechanisms that organise it. Towards this aim, we have established an efficient *in vivo* system that, through the combined used of transposases and recombinases, allows the reengineering, in a systematic manner, of the mouse genome. With this approach, we generated a unique resource comprising hundreds of mouse strains carrying regulatory sensors throughout the genome, and series of specific chromosomal rearrangements in selected loci. This genomic resource enables us to dissect functionally the genomic information and the mechanisms that organise a linear genome into structurally distinct domains and chromatin loops, so as to implement long-range specific regulatory interactions.

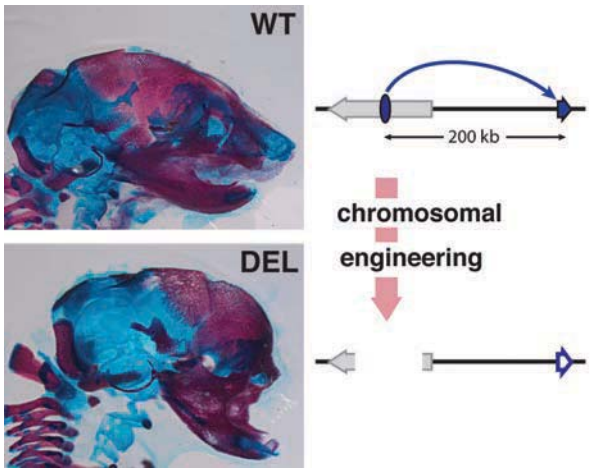
Future projects and goals

Structural conformation and regulatory organisation of the genome: By combining advanced genomic engineering with chromatin profiling (ChIP-Seq) and conformation analyses (4C-chromatin conformation capture, super-high resolution FISH), we aim to learn how the genomic organisation of a locus determines the specific chromatin structures and conformations that it adopts in the nucleus, and determines their functional significance in the context of a developing embryo.

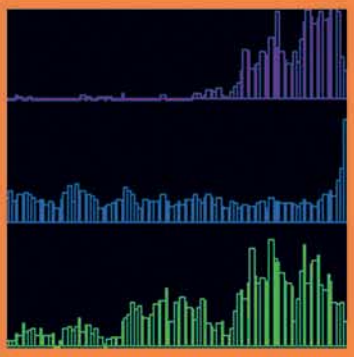
Regulatory architecture, disease and evolution: Our mouse models provide insights into the consequences of structural variations or chromosomal aneuploidies found in humans, both at the phenotypic and molecular level. Comparison of the regulatory architecture of developmental gene loci between different species can reveal how large-scale changes in chromosomal organisation may have contributed to evolution of body forms.

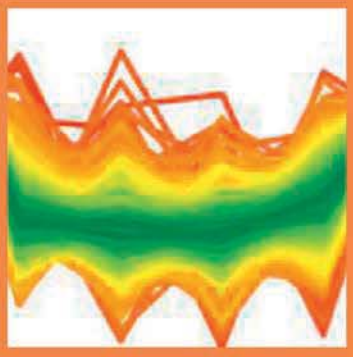


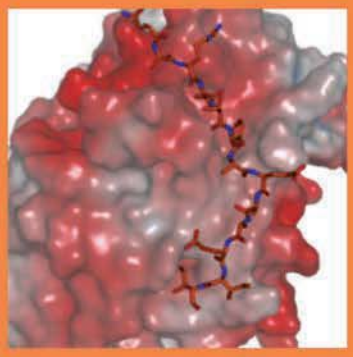
Adjacent insertions of a sensor gene showed different activities, highlighting the regulatory architecture of the corresponding locus (see Ruf *et al.*, 2011).

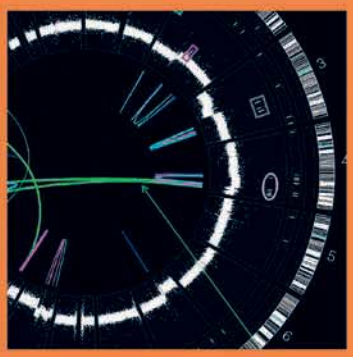


Abnormal skull development in mice with a deletion of distal enhancers engineered by *in vivo* recombination (see Marinic, *et al.* 2013).

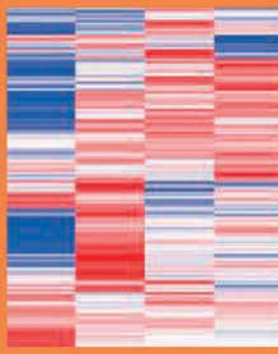






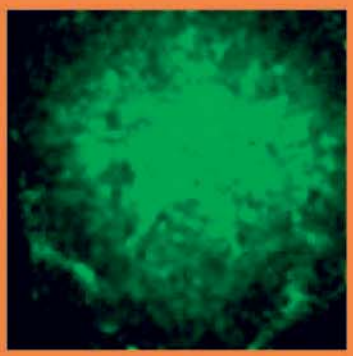


1974 - 2014 Research Highlights



Identification of CRM1 as the first known protein that exports cargo macromolecules from the nucleus.
Fornerod M, *et al.* (1997) CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* 90, 1051-60

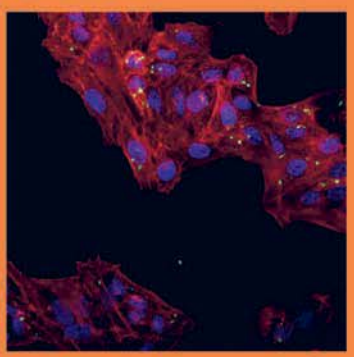
Contrary to what was previously thought, the majority of promoters – DNA sequences that tell the cellular machinery to start transcribing a gene – drive transcription in both directions.
Xu Z, *et al.* (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457, 1033-7



An inherited mutation in gene p53 is likely the link between "exploding chromosomes" (chromothripsis) and the paediatric brain tumour medulloblastoma.
Rausch T, *et al.* (2012) Genome Sequencing of Pediatric Medulloblastoma Links Catastrophic DNA Rearrangements with TP53 Mutations. *Cell* 148, 59-71

Systematic identification of genetic switches called enhancers and the molecules that activate them – transcription factors – can be used to draw a cell's family tree.
Junion G, *et al.* (2012) A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* 148, 473-86

New method for identifying proteins in a sample by identification of peptides – fragments of proteins – in mass spectrometry experiments.
Mann M & Wilm M (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Analytical Chemistry* 66, 4390-9



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 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 2009-12.
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, *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, *et al.* (2012) Tissue specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature Genetics* 44, 148-56

Previous and current research

The Furlong group dissects fundamental principles of transcriptional regulation, and how that drives cell fate decisions during development, focusing on functional and organisational 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 initially established through the integration of environmental cues (signalling pathways) with transcriptional networks, which converge on cis-regulatory elements called enhancers. Enhancers therefore 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.

Much of our research focuses on mechanisms of enhancer function, including how the cis-regulatory genome is organised with 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 multicellular embryo, *Drosophila* mesoderm specification.

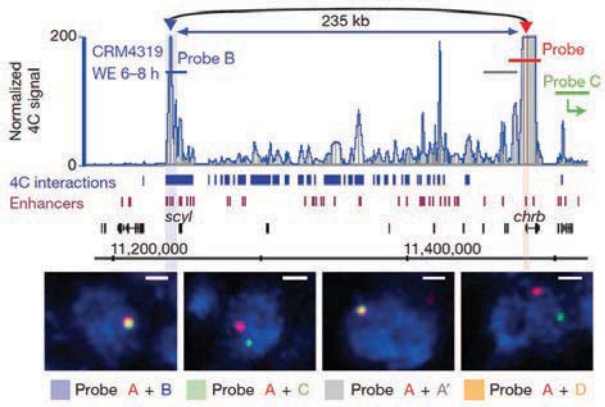


Figure 1: Enhancers interact with genes over very long distances within the *Drosophila* genome, as shown by 4C-Seq (top) and DNA FISH (bottom) during embryogenesis (Ghavi-Helm Y, *et al. Nature* 2014).

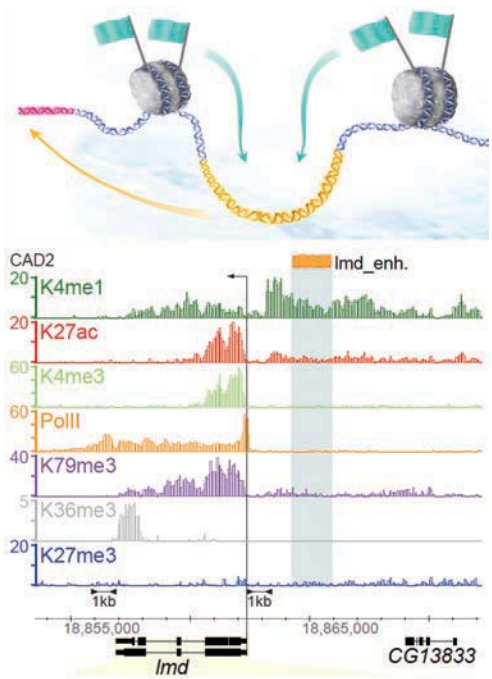


Figure 2: Chromatin state and Pol II occupancy on enhancers (yellow) is highly predictive of enhancers' activity, with Pol II being predictive for the precise timing during development (Bonn, *et al. Nature Genetics* 2012).

Future projects and goals

Chromatin remodelling during cell fate decisions: To uncover general properties of enhancer function during embryogenesis, 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)). Using this method, we are currently dissecting the interplay between changes in chromatin remodelling, transcription factor and Pol II occupancy with dynamic changes in developmental transitions.

Enhancer looping and 3D Genome Topology: For enhancers to function, they must come in proximity to their target gene's promoter. This often results in the 'looping out' of intervening DNA. We have recently examined this within two time-points of development (*Nature* 2014), and discovered extensive long-range interactions within the compact *Drosophila* genome, and a surprising stability of these interactions during these stages of development. We will build on this, looking at enhancer topology during a much longer developmental time-span, integrating high-resolution imaging to understand the relationship between proximity and transcriptional regulation.

Variation and plasticity in cis-regulatory networks: Variation in cis-regulatory elements can affect gene expression and account for individual differences in phenotypes. However, little is known about how much variation is tolerated during essential developmental processes during embryonic development. We are investigating this by determining the extent to which natural sequence variation among wild isolates, and nearby species, affects embryonic development at a transcriptional and genome organisational level.

Systems genetics

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

Aiyar R, *et al.* (2014) Mitochondrial protein sorting as a therapeutic target for ATP synthase disorders. *Nature Commun.* 5, 5585

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

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

Xu Z, *et al.* (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457, 1033-7

Previous and current research

One of the most daunting obstacles in biomedicine is the complex nature of most phenotypes (including cancer, diabetes, heart disease) due to interactions between multiple genetic variants and environmental influences. A central challenge is to understand how genetic and environmental perturbations affect health and disease. Our research addresses this challenge by developing novel genomic approaches to investigate the molecular processes that link genotype to phenotype, identifying the causal underlying factors, and quantifying their contributions. We investigate inter-individual variation at the level of the genome, transcriptome, and proteome, which we integrate with higher-level phenotypes.

The Steinmetz group bridges diverse domains of genome science, from deciphering the structure and function of genomes to the application of these insights in understanding diseases.

Our projects are in the following areas:
Functions and mechanisms of transcription: 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, and the molecular phenotypes that arise from pervasive transcription. Recently, we discovered extensive variation in the start and end sites of transcript molecules produced by each gene by developing a novel technique to map full-length transcript isoforms (figure 1).

Quantitative genetics: We piloted new technologies to dissect the genetic basis of complex, multifactorial phenotypes. We use high-throughput, quantitative approaches to study how genetic variation is inherited through recombination and the consequences of genetic variation. By integrating multiple layers of molecular data, we aim to predict phenotype from genotype and define intervention points that can be targeted to modulate phenotypes of interest (Figure 2).

Disease models: We use multiple model systems, primarily yeast and human cells, to characterise the genetic and cellular systems affected in particular diseases, and assess potential therapeutic strategies. We are applying personalised functional genomics to study diseases in patient-derived cells using systematic and targeted approaches, to unravel mechanisms and discover novel treatments (watch: **http://bit.ly/1AmUA0W**).

Future projects and goals

We are integrating multiple layers of molecular data in order to understand how the genome is read for function. Using novel algorithms, intervention points can be identified from such data that can be targeted to modulate phenotypes of interest. We are also following up on our studies of transcriptional regulation through targeted investigations of the interplay between epigenetics and transcription, the functional consequences of complex transcriptome architecture, and its contribution to single-cell heterogeneity. Ultimately, by integrating genetics, genomics, systems biology, and computational modelling, we aim to develop approaches that unravel disease mechanisms and predict effective therapeutics, enabling personalised and preventive medicine.

Our lab operates in an integrated manner across sites in Heidelberg, Germany, and at Stanford University in the US.

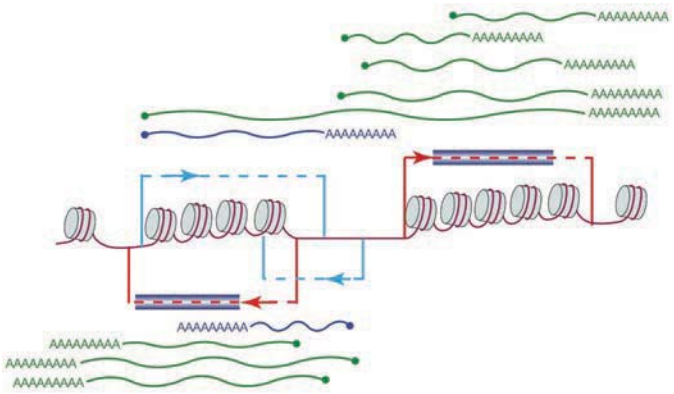


Figure 1: Extensive variation in transcript start and end sites revealed by TIF-Seq, a novel technique for transcript isoform profiling (Pelechano *et al.*, *Nature* 2013).

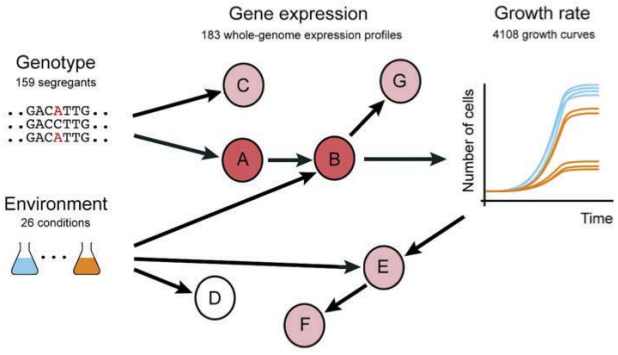


Figure 2: Gene-environment interactions reveal causal pathways (A-B) that mediate genetic effects on phenotype (Gagneur *et al.*, *PLOS Genetics* 2013).

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

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

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

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

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

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

Previous and current research

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 (e.g. genetic variation) or their environment (e.g. drugs). 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. More specifically, we use large-scale data acquisition and quantitative modelling of phenotypes and molecular profiles, systematic perturbations (ie: drugs or RNAi screens) 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.

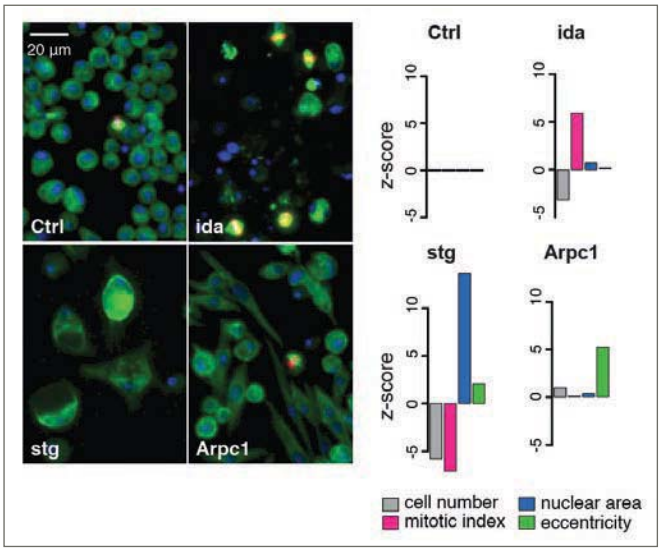


Figure 1: Automated multivariate phenotyping of cells by combinatorial RNAi and automated image analysis.

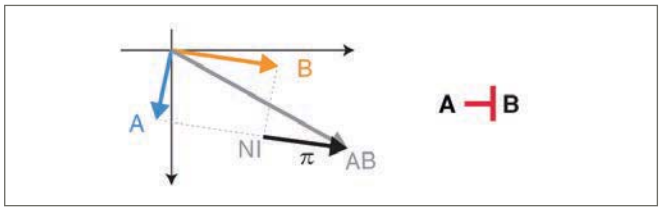


Figure 2: Epistatic interactions between genetic or drug perturbations are mapped from high-throughput microscopy data by multivariate phenotyping and vector space modelling (Fisher B *et al.* 2015).

Future projects and goals

We aim to develop the computational techniques needed to analyse current and novel raw biological data:

- Precision oncology: we work with clinical researchers to develop predictive assays and algorithms based on genome sequencing and other levels of molecular profiling.
- Many powerful mathematical and computational 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 sometimes term 'Translational Statistics'.
- Transcriptomics, gene regulation and 3D nuclear organisation.
- Quantitative proteomics and *in vivo* drug-target mapping.
- Single-cell and single-molecule data modelling.
- 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.

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

Previous and current research

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.

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.

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.

Future projects and goals

- 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.

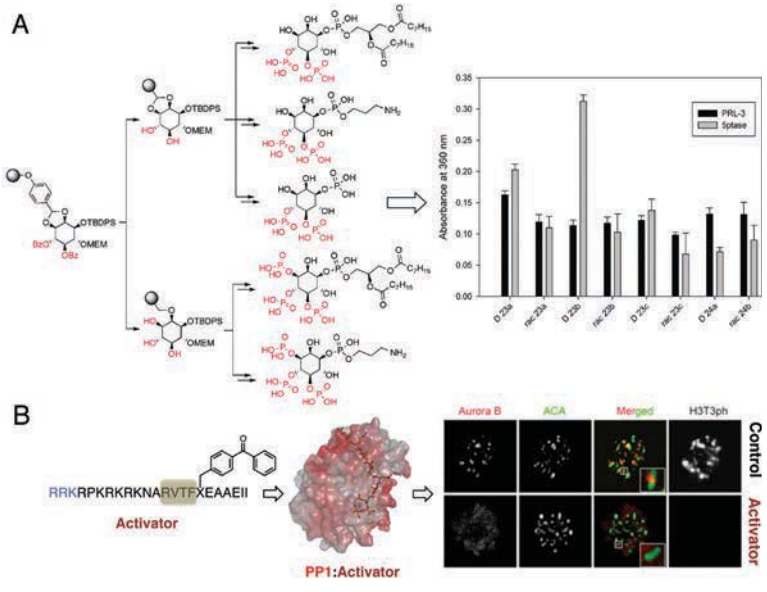


Figure 1: (A) Solid phase synthesis of phosphoinositides for the preparation of libraries for SAR studies with lipid phosphatases (Bru *et al.*, *Chem. Sci.* 2012). (B) Selective activators of PP1 in cells enable us to gain new insights into PP1 biology (Chatterjee *et al.*, *Angew. Chem. Int. Ed.* 2012;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

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

Origin and function of genetic variation

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

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

Korbel JO, Campbell PJ (2013) Criteria for inference of chromothripsis in cancer genomes. *Cell* 152, 1226-36

Weischenfeldt, J. *et al.* (2013) Integrative genomic analyses reveal an androgen-driven somatic alteration landscape in early-onset prostate cancer. *Cancer Cell* 23, 159-70

Rausch, T. *et al.* (2012) Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell* 148, 59-71

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

Previous and current research

Genetic variation is a principal reason why we differ from one another and can be used as starting point to unravel disease mechanisms. Advances in DNA sequencing technology have facilitated the characterisation of genetic variation at genome-wide scale. Our group is investigating the extent, origin, and functional consequences of DNA variation, with a particular focus on genomic structural variants (SVs) such as deletions, duplications, inversions and translocations – the most consequential type of heritable genetic variation in humans in terms of basepairs affected. Germline and somatic SV classes have been linked to numerous heritable diseases and cancers. Our laboratory uses a 'hybrid' approach, integrating laboratory and computational biology techniques, to combine data generation and analysis with hypothesis generation and testing in experimental model systems.

A cancer genome study that we recently performed revealed that the development of medulloblastoma, the most common malignant brain tumour in children, frequently involves a remarkable process known as chromothripsis, where localised chromosomal shattering and repair occur in a one-off massive DNA rearrangement event (figure 1). We also recently made progress in understanding the etiology of early onset prostate cancer, the initiation of which we found to be largely driven by androgen-mediated somatic SVs, and we further uncovered that enhancer hijacking drives oncogene expression in medulloblastoma (figure 2). Our group also plays crucial roles in international research consortia such as the 1000 Genomes Project, where we are generating fine-resolution genetic variation maps in humans and relate these to functional genomics data.

Within the Pan-Cancer Analysis of Whole Genomes (PCAWG) initiative of the International Cancer Genome Consortium (ICGC) we have begun investigating whole genome, DNA methylome, and transcriptome sequencing data of ~2500 cancer patients. Using integrative computational and statistical approaches we aim to unravel commonalities and discrepancies between cancer types at the molecular level and study determinants of disease-susceptibility, to unravel causalities and to facilitate the molecular classification of malignancies with a potential impact on diagnostics and treatment.

Future projects and goals

- Uncovering genetic determinants for the development and progression of cancer in humans, and studying commonalities and differences between tumour types.
- Combining genomic and epigenetic studies, including chromatin state and conformation analyses, to identify determinants of DNA rearrangement formation and selection.
- Constructing near-complete human genome variation maps using third generation sequencing technologies.
- Development of *in vitro* and *in silico* approaches for deciphering the molecular origin and function of SVs in humans and model organisms.
- Deciphering the mechanistic basis of chromothripsis, an SV process particularly abundant in highly aggressive malignancies.

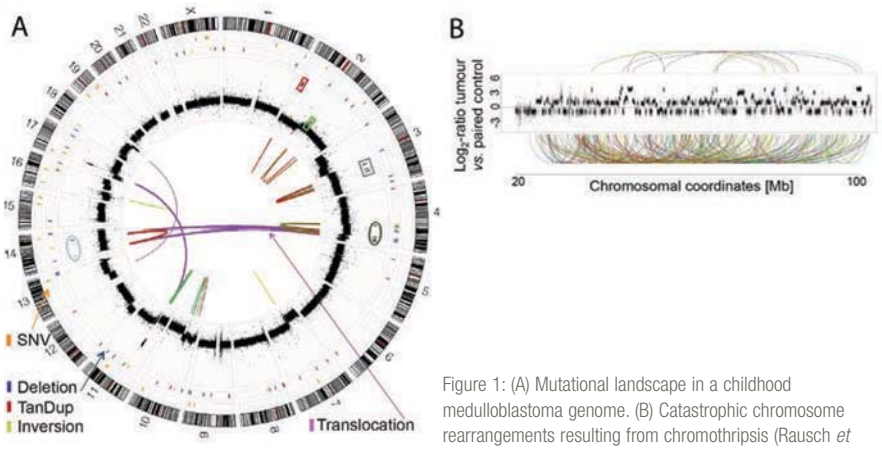


Figure 1: (A) Mutational landscape in a childhood medulloblastoma genome. (B) Catastrophic chromosome rearrangements resulting from chromothripsis (Rausch *et al.*, *Cell* 2012).

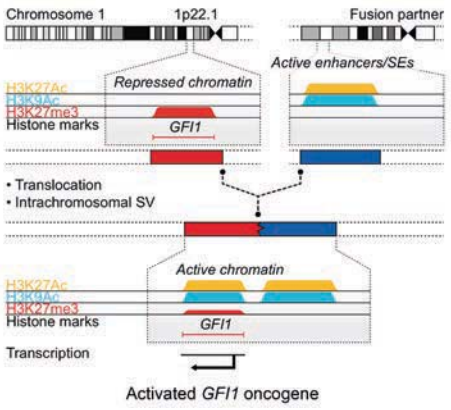


Figure 2: Oncogene activation by enhancer hijacking (Northcott *et al. Nature* 2014).

Functional proteomics

Jeroen Krijgsveld

PhD 1999, University of Amsterdam, The Netherlands.

Postdoc at Utrecht University, The Netherlands and Harvard Medical School, Boston, USA.

Assistant Professor, Utrecht University, The Netherlands.

Team leader at EMBL since 2008.

SELECTED REFERENCES

Hughes CS, *et al.* (2014) Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.* 10, 757

Hansson J, Krijgsveld J (2013) Proteomic analysis of cell fate decision. *Curr. Opin. Genet. Dev.* 23, 540-7

Hansson J, *et al.* (2012) Highly coordinated proteome dynamics during reprogramming of somatic cells to pluripotency. *Cell Rep* 2, 1579-92

Eichelbaum K, *et al.* (2012) Selective enrichment of newly synthesized proteins for quantitative secretome analysis. *Nat. Biotechnol.* 30, 984-90

Previous and current research

Proteins fulfill most of the functions that are crucial in establishing cellular phenotypes. In addition, it is becoming increasingly clear that proteins rarely act alone, but that they constitute intricate networks, both among themselves and with other biomolecules. This system is both robust and dynamic, allowing a cell to respond to external cues, or to develop from an embryonic to a mature state. Our interest is in understanding cellular properties from this perspective, realising that one needs to study proteins collectively rather than in isolation, and dynamically rather than under static conditions.

Our research is centred on quantitative proteomics, combining biochemistry, mass spectrometry, analytical chemistry, and bioinformatics, and applies to various biological systems (yeast, *Drosophila*, mammalian cells). Our main interest is to understand how changes in protein expression, localisation and interaction underlie processes of stress-response, differentiation and reprogramming.

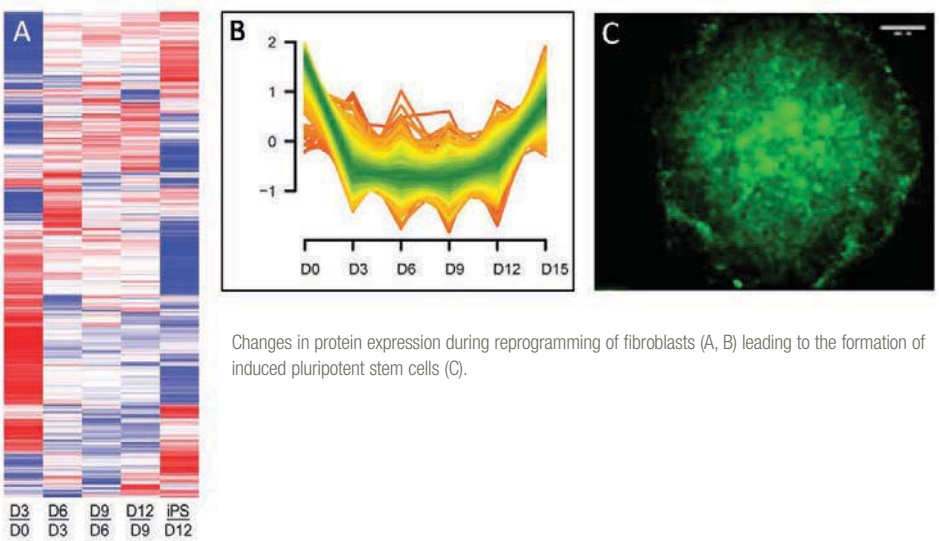
For instance, large-scale proteomic experiments enable us to characterise the proteomes of highly purified mouse hematopoietic stem cells and progenitor populations obtained by fluorescence-activated cell sorting, generating novel insights in the initial steps of hematopoiesis *in vivo.* Furthermore, we have performed time course analyses quantifying the proteome changes in fibroblasts during their reprogramming to induced pluripotent stem cells (iPSCs), identifying and functionally validating proteins that are key in the gain of pluripotency.

Apart from these large-scale analyses of intracellular proteomes, we have developed new tools to study secretory proteins and their role in cell signalling and communication. Furthermore, we are interested in regulatory principles of transcriptional activation and protein turnover in the face of developmental processes or response to stress. We are therefore developing novel techniques to identify proteins that interact with regulatory domains in the genome, both *in vivo* and *in vitro.* We aim to identify proteins that drive (or inhibit) transcription in a gene- and condition-specific manner, for example to understand how transcription of developmentally important genes is controlled. To further explore the link between genome regulation and protein output, we study protein turnover taking yeast as a model system. By determining protein synthesis and degradation proteome-wide and across a range of growth conditions, we aim to construct models of how protein homeostasis is maintained.

The Krijgsveld team uses biochemical and mass spectrometric approaches to understand the dynamics of protein expression and interaction in the context of cellular differentiation and stress response.

Future projects and goals

- Develop new tools to study protein-DNA and protein-RNA interactions to identify and functionally characterise proteins that regulate transcription and translation.
- Integration of proteomics and next-generation sequencing to understand the molecular basis of protein homeostasis.
- Study cellular communication via secretory proteins.



Changes in protein expression during reprogramming of fibroblasts (A, B) leading to the formation of induced pluripotent stem cells (C).

Miniaturising biology and chemistry in microfluidic systems

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

El Debs B, *et al.* (2012) Functional singlecell hybridoma screening using droplet-based microfluidics. *PNAS* 109, 11570-75

Vyawahare S, *et al.* (2010) Miniaturization and parallelization of biological and chemical assays in microfluidic devices. *Chem. Biol.* 17, 1052-65

Granieri L, *et al.* (2010) High-throughput screening of enzymes by retroviral display using droplet-based microfluidics. *Chem. Biol.* 17, 229-35

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 novel approaches in microfluidic technology to address complex, multidisciplinary questions at the interface of biology, chemistry and engineering.

Previous and current research

Working on the micro-scale offers some unique advantages:

- Drastically increased throughput (processing up to a million samples an hour).
- Superb spatio-temporal resolution (assays can be carried out on micrometre length scales and sub-millisecond timescales).
- Low material consumption, enabling single-organism, single-cell, or even single-molecule assays.

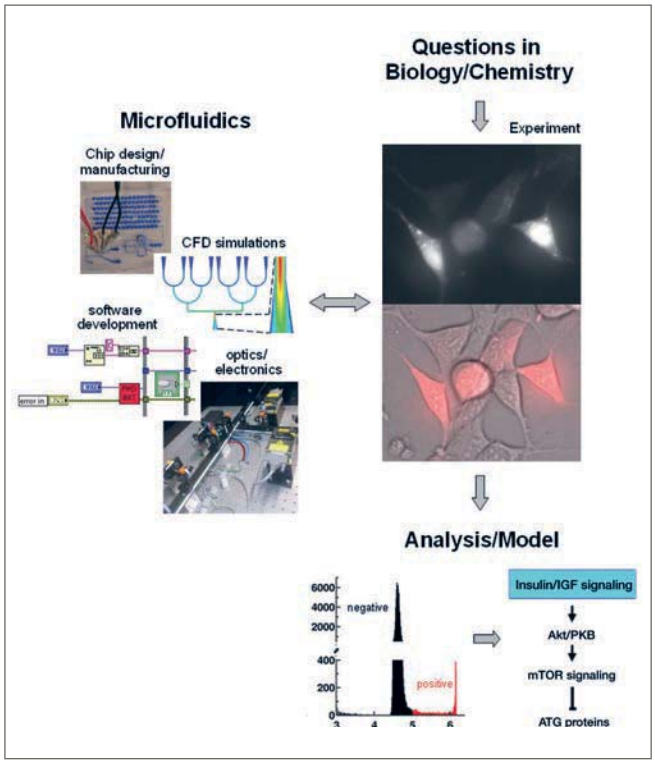
During the past couple of years we have developed powerful microfluidic platforms for cell-based and (bio)chemical assays. We perform all steps ranging from the design and manufacturing of microfluidic chips and detection systems, to the cultivation and study of human cells and multicellular organisms. Furthermore, we are interested in combinatorial chemistry, perform computational fluid dynamics simulations, and develop novel software controlling our microfluidic systems.

For many 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. We also use continuous-phase microfluidics to generate laminar flow patterns, where we expose cells and organisms (or even small parts thereof) to different chemical environments. Amongst other applications, this allows the analysis of signalling events in developing embryos.

Future projects and goals

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

- **Biomedical applications**: Droplet-based microfluidics enables functional antibody screening at very high throughput. We aim to use this technique for several applications: to identify therapeutic antibodies, to identify potential HIV vaccine candidates, and to develop novel approaches for personalised cancer therapy.
- **Cell biology**: With large-scale chemical perturbations we want to identify pathway interactions in stem cell differentiation and carcinogenesis. These microfluidic chemical genetics approaches require only small numbers of cells and are hence compatible with primary cells or even patient biopsies.
- **Genomics**: We are developing microfluidic modules for single-cell barcoding and sequencing. Furthermore, we are setting up integrated microfluidic ChIPseq platforms allowing for the analysis of less than 5000 cells. Notably, some of our modules have already been commercialised.



Microfluidic approaches in biology and chemistry

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.* (2014) ATRX tolerates activity-dependent histone H3 methyl/phos switching to maintain repetitive element silencing in neurons. *Proc Natl Acad Sci USA,* e-pub ahead of print

Maze I, *et al.* (2014) Every amino acid matters: essential contributions of histone variants to mammalian development and disease. *Nature Reviews Genetics* 15, 259-71

Noh KM, *et al.* (2012) Repressor element-1 silencing transcription factor (REST)-dependent epigenetic remodeling is critical to ischemia-induced neuronal death. *Proc Natl Acad Sci USA* 109, E962-71

Goldberg AD, *et al.* (2010) Distinct factors mediate histone variant H3.3 deposition at specific genomic regions. *Cell* 140, 678-91

Previous and current research

Chromatin, the faithful association of genomic DNA with histone proteins, exists as the physiological form of our genome and the substrate for processes that regulate cellular gene expression. Numerous diseases are associated with mutations in genes that encode for chromatin-binding and/or chromatin-modifying enzymes, which together act as epigenetic regulators. Combining neurobiology and chromatin biology, we aim to study the molecular mechanisms that link genetic mutations encoded in epigenetic regulators to the widespread chromatin alterations associated with brain diseases. A central question grounding our research is how the chromatin modification network engages in brain development, function and disease.

The Noh group studies chromatin links vital for neurodevelopment and disease.

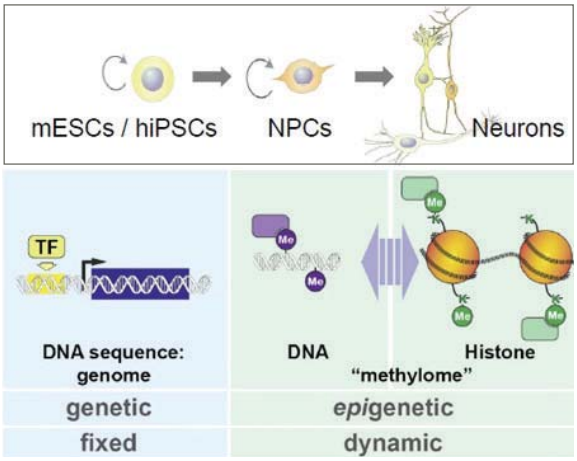
Previously, our team studied a cellular pathway of ischemia-induced neuronal death, and showed that the transcriptional repressor REST causes epigenetic remodelling and repression of multiple target genes including AMPA receptor in postischemic neurons. We further demonstrated that REST knockdown prevents neuronal death in a clinically relevant in vivo model of ischemia. During the past few years, 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. In addition, we revealed the molecular mechanisms that underlie DNA methylation, specifically the interplay between histone post-translational modifications and DNA methylation, and identified the biological function of this interaction in cell lineage specification.

Future projects and goals

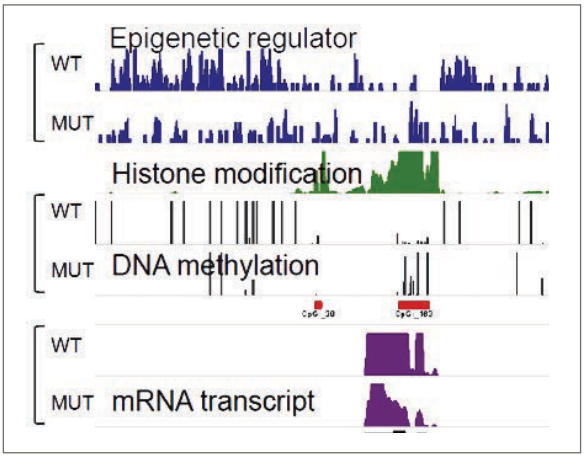
We aim to study chromatin regulation, its interpretation during brain development, and its misinterpretation in relation with brain cognitive and developmental diseases. We will use differentiating neurons from mESCs and human induced pluripotent cells (hiPSCs) to model developmental stages and facilitate the necessary genetic manipulations / engineering. Defining the 'epigenetic landscape' – both in normal and abnormal brain cells – will help provide novel targets for therapeutic intervention for cognitive and developmental diseases of the brain.

Our research projects are to:

- Determine combinatorial histone modifications that link *de novo* DNA methyltransferase location/function during neuronal lineage commitment.
- Identify the location/function of mutated histones and epigenetic regulators specific to cognitive deficits, and explore alterations of the epigenetic landscape in developing neurons.
- Investigate PHD (Plant Homeo Domain)-containing epigenetic regulators that integrate specific signalling pathways into developmental transcription programs.



Top: biological system of interest, bottom: concept of epigenetic mechanisms.



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

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

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

Egan AJ, *et al.* (2014) Outer-membrane lipoprotein LpoB spans the periplasm to stimulate the peptidoglycan synthase PBP1B. *Proc. Natl. Acad. Sci. USA* 22, 8197-202

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.

Previous and current research

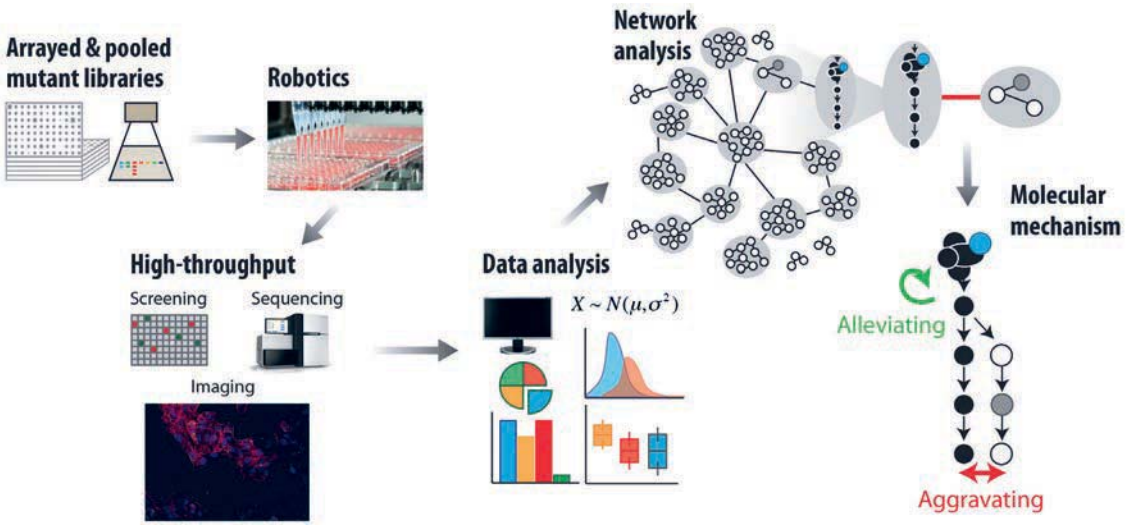
The recent explosion of genomic sequence information provides a first step towards better understanding diverse bacteria, but also makes it crucial to develop large-scale phenotyping approaches to characterise functions of novel genes and to map them within pathways. We are developing such high-throughput, 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.

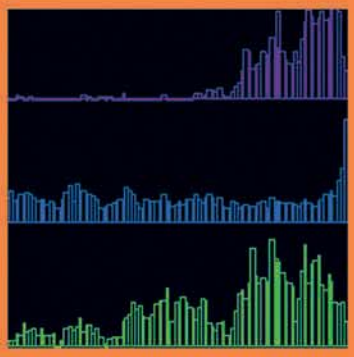
We are also developing large-scale automated platforms for elucidating the mode-of-action of new antibacterials, for large-scale profiling of combinatorial drug therapies and for dissecting the underlying mechanism(s), and for identifying adjuvants that re-sensitise multi-resistant bacterial pathogens 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.

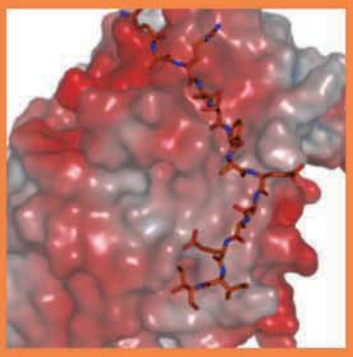
Future projects and goals

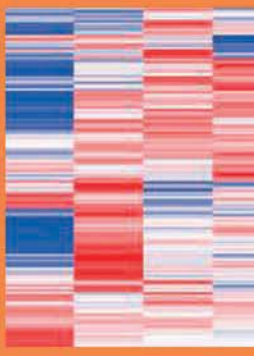
We are now expanding our efforts in two directions. First, we are introducing our high-throughput screening approaches into abundant and prevalent species of the human gut microbiome. In collaboration with the Alexandrov, Bork, and Patil groups, and utilising a plethora of complementary technologies – such as imaging mass spectrometry, cutting-edge microscopy 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.

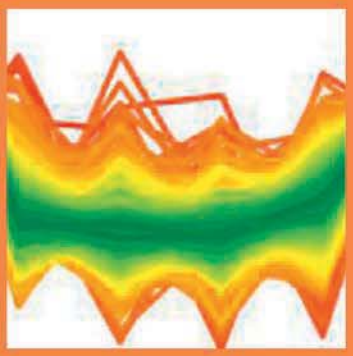


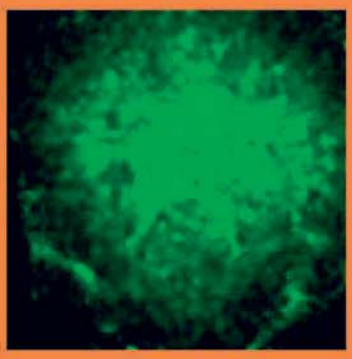
High-throughput gene-gene, gene-drug and drug-drug interaction profiling provides novel mechanistic insights into the cellular network architecture and into drug mode-of-action.

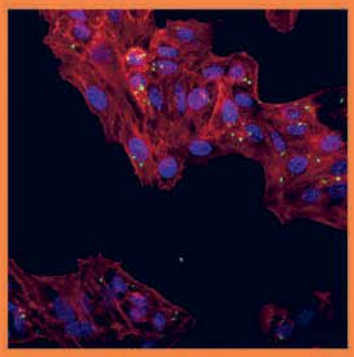


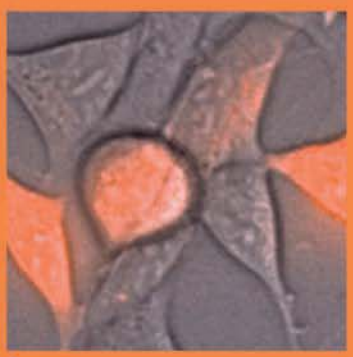


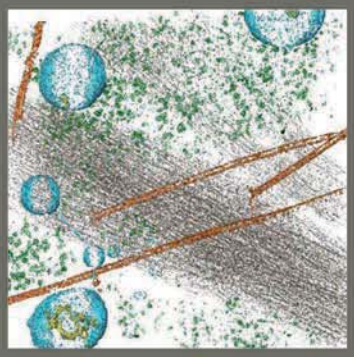
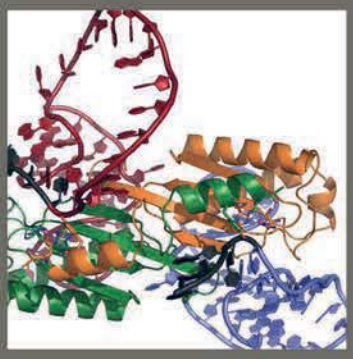


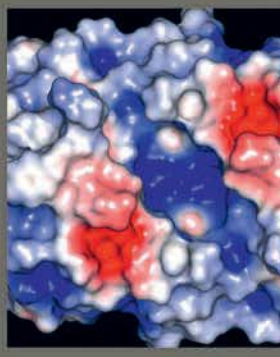








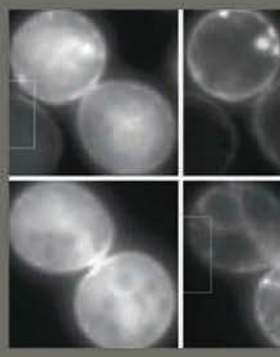
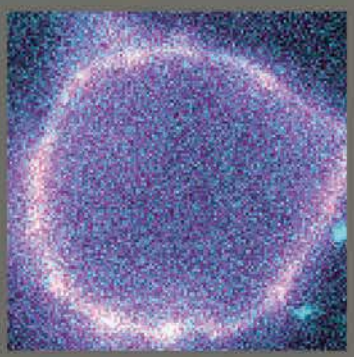
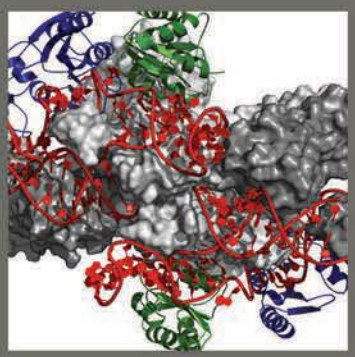
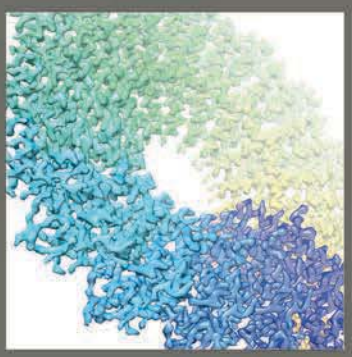
1974 - 2014 Research Highlights

Method for preparing and observing unfixed and unstained frozen biological samples using cryo-electron microscopy.
Dubochet J, *et al.* (1982) Electron microscopy of frozen water and aqueous solutions. *Journal of Microscopy* 128, 219-37
AND
Adrian M, *et al.* (1984). Cryo-electron microscopy of viruses. *Nature* 308, 32-6

3D structure of the molecular machine that collects energy from light in green plants.
Kühlbrandt W & Wang DN (1991) Three-dimensional structure of plant light-harvesting complex determined by electron crystallography. *Nature* 350, 130-47

First comprehensive map of protein interactions in yeast cells highlighted that most tasks are performed by networks of proteins.
Gavin AC, *et al.* (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141-7
AND
Gavin AC, *et al.* (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440, 631-6.

The combination of microbes in each person's intestine falls into one of three 'gut types'; also the identification of microbial genetic markers related to age, gender and body-mass index.
Arumugam M, *et al.* (2011) Enterotypes of the human gut microbiome. *Nature* 473, 174-80.

EMBL Heidelberg

Structural and Computational Biology

The Unit pursues an ambitious research programme with a strong basis in integrated structural systems biology and a far-reaching computational component that bridges into various areas of biology.

A wide spectrum of expertise allows the Unit to tackle problems at different ranges of spatial resolution, connecting atomic structures and dynamic information obtained by X-ray crystallography and NMR with medium-range resolution from single particle electron microscopy, and cellular imaging obtained by electron tomography and light microscopy. Dedicated large scale biochemistry, proteomics, chemical biology, biophysics, and cell biology approaches complement the structural biology activities and, in conjunction with a wide range of innovative computational biology activities, are integrated into a comprehensive description of biological function.

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. One example is the comprehensive structural and temporal description of an entire cell at almost molecular resolution. It goes hand in hand with the application of and integration of various 'omics' approaches to the small bacterium *Mycoplasma pneumoniae*, by characterising its dynamic protein organisation and merging this molecular information to cellular, high-resolution tomograms. In the thermophilic fungus Chaetomium thermophilum spatial and temporal networks will be deduced using multidisciplinary approaches including structural studies, large scale biochemistry and computational biology. Together, they will provide insight into eukaryotic thermophily at the molecular and cellular level.

Currently, the Unit consists of twelve research groups covering a broad methodological spectrum. The core technologies include electron microscopy (three groups), X-ray crystallography (two groups), NMR (one group), chemical biology (two groups) and computational biology (four groups). However, 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.

The Unit is very well equipped for experimental and computational work. Experimental facilities include: a crystallisation robot and automated crystal visualisation; rotating anode and image plate detector for the collection of X-ray diffraction data; 800 MHz, 700 MHz, 600 MHz and 500 MHz NMR spectrometers; and several transmission electron microscopes, including a high-throughput Titan Krios microscope for single particle cryo-electron microscopy and cryo-electron tomography. 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 3000 CPU cores, whereby large central clusters and separate workstations are conveniently networked.

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

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 biofinformatics at EMBL Heidelberg since 2011.

Group leader in the Molecular Medicine Partnership Unit.

ERC Advanced Investigator.

SELECTED REFERENCES

Zeller, G. *et al.* (2014) Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol.* 10, 766

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

Qin J, *et al.* (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59-65

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.

Previous and current research

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.*, 2001), foundational work on the study protein interaction networks (von Mering *et al.*, 2002) and comparative metagenomics (Tringe *et al.*, 2005), and exploration of drug-target interactions using global human "readouts" such as side effects (Campillos *et al.*, 2008).

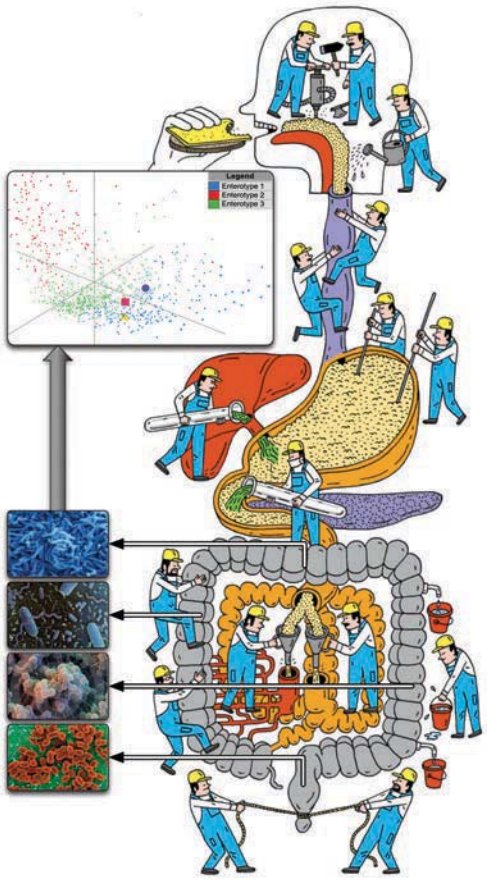
Although we currently have a number of ocean microbiome projects in the context of the TARA Oceans expedition, we mainly focus on the human gut microbiome. We employ metagenomics to uncover the principles of microbial communities in healthy and diseased humans. We identified three main "enterotypes" – or gut microbial community compositions – in developed countries. (Arumugam *et al.*, 2011), and showed that each human appears to carry individual strains (Schloissnig *et al.*, 2013). We are finding microbial markers for a number of diseases such as obesity (Le Chatelier *et al.*, 2013) and colon cancer (Zeller *et al.*, 2014).

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.*, 2013, Iskar *et al.*, 2013) and Iso study the impact of drugs on the microbiome.

Future projects and goals

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. 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.

We also contribute to EMBL's Bio-IT community and the development of the internal EMBL Bio-IT portal to help network and facilitate the work of bioinformaticians in the EMBL community.



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.

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

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.* 21, 2367-79

Fernández-Tornero C, *et al.* (2013) Crystal structure of the 14-subunit RNA polymerase I. *Nature* 7473, 644-9

Taylor NM et al. (2013) RNA polymerase III-specific general transcription factor IIIC contains a heterodimer resembling TFIIF Rap30/Rap74. *Nucleic Acids Res.* 19, 9183-96

Previous and current research

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 inner-working 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 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 – ie: 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 synthesizes 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 projects and goals

- 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.

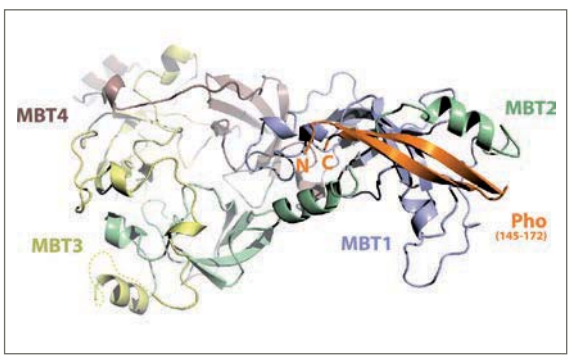


Figure 1: Interaction between the Polycomb group protein Sfmbt 4MBT domain and the Pho spacer region (Alfieri *et al.*, 2013).

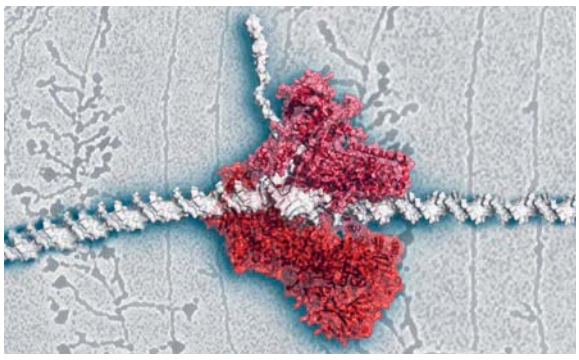


Figure 2: Crystal structure of 14-subunit yeast RNA polymerase I. The background shows an electron micrograph of Miller chromatin spreads where nascent pre-rRNA transcripts form tree-like structures (Fernández-Tornero *et al.*, 2013).

Spatial metabolomics

Theodore Alexandrov

PhD 2007, St. Petersburg State University.

Postdoctoral Researcher, University of Bremen.

Group Leader, University of Bremen.

Visiting Researcher, University of California San Diego.

Team leader at EMBL since 2014.

SELECTED REFERENCES

Alexandrov T, Bartels A (2013) Testing for presence of known and unknown molecules in imaging mass spectrometry. *Bioinformatics* 29, 2335–42

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.

Previous and current research

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 generating 100 gigabytes of information-rich data for one sample only. 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 projects and goals

- 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.

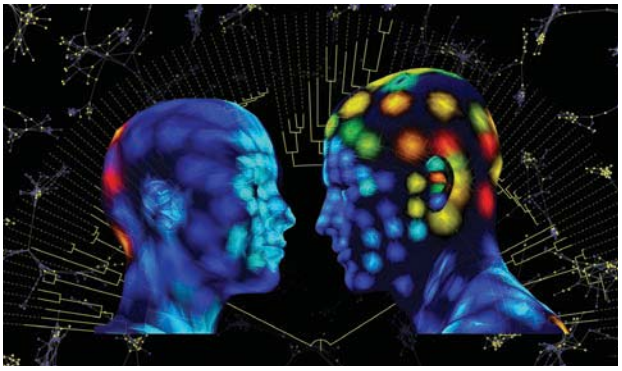


Figure 1: Surface mapping of two metabolites on the skin of female and male individuals. The models are overlaid with a molecular network in background showing the structural relations between these and hundreds of other detected metabolites.

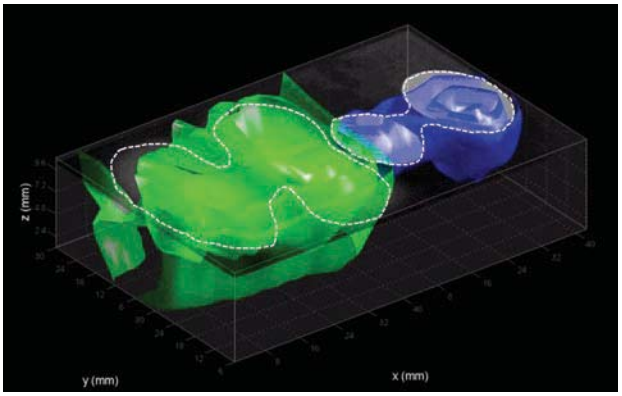


Figure 2: 3D spatial localisation of two metabolites (green, a rhamnolipid with an inhibitory function, and blue) secreted within the agar medium by the interacting colonies of *P. aeruginosa* and *C. albicans* (Watrous *et al.*, *ISME J.*, 2013).

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

Previous and current research

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 (mainly X-ray crystallography), molecular biology, biochemistry, biophysics, 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, and their regulation in the cell.

Sleeping Beauty: This transposon is a prime tool in vertebrate genetics. We study its structure and mechanisms and, in collaboration with the Gavin and Beck groups, we also investigate how it interacts with other components of human host cells.

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 study the mechanisms for two of them 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. In collaboration with the Carlomagno and Pillai groups, we investigate 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.

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.

Future projects and goals

- 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.

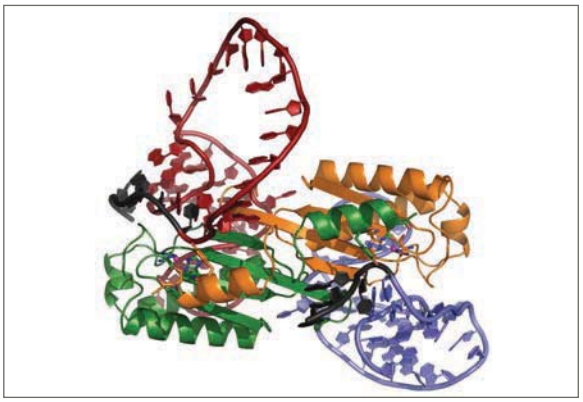


Figure 1: Five crystal structures of the IS608 transpososome together with associated biochemical data elucidate the entire pathway of single-stranded DNA transposition and show how it selects its integration site.

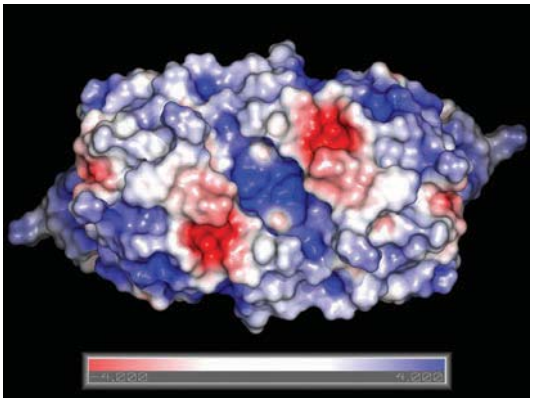


Figure 2: Crystal structure of the primary piRNA biogenesis factor Zucchini reveals its endonuclease function.

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

Bui KH, *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. *Mol. Syst. Biol.* 9, 648

Malmstrom J, *et al.* (2009) Proteome-wide cellular protein concentrations of the human pathogen *Leptospira interrogans*. *Nature* 460, 762-5

Beck M, *et al.* (2007) Snapshots of nuclear pore complexes in action captured by cryo-electron tomography. *Nature* 449, 611-5

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

Previous and current research

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 projects and goals

- 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.

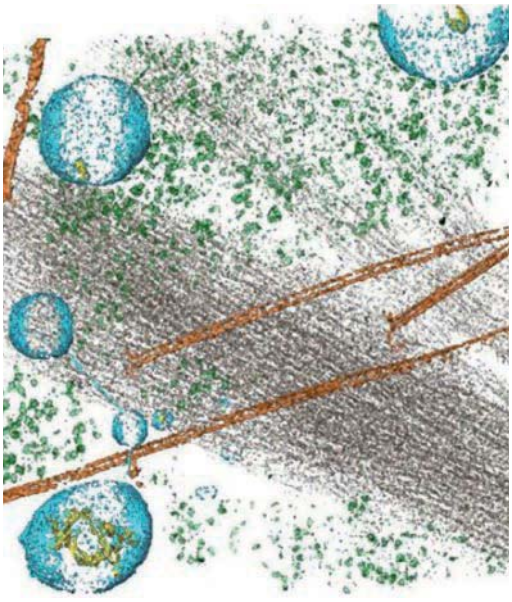


Figure 1. Cryo-electron tomogram of a fraction of the cytoplasm of a human cell. Microtubules are coloured in orange, stress fibres in grey, protein complexes in green, membranes in cyan and vesicular contents in yellow.

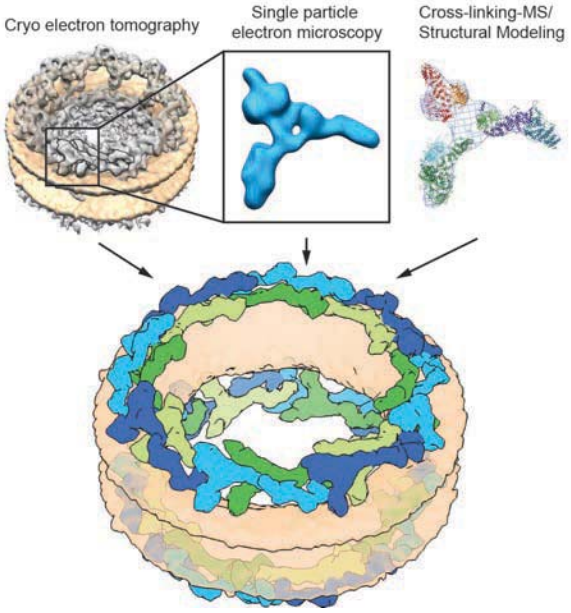


Figure 2. Model of the scaffold arrangement of the human Nuclear Pore Complex revealed by an integrated approach consisting of cryo-electron tomography, single particle EM, cross-linking MS and structural modelling (Bui, von Appen *et al.*, *Cell*, 2013).

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

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

Bharat TA, *et al.* (2012) Structure of the immature retroviral capsid at 8A resolution by cryo-electron microscopy. *Nature* 487, 385-9

Faini M, *et al.* (2012) The structures of COPI-coated vesicles reveal alternate coatomer conformations and interactions. *Science* 336, 1451-4

Previous and current research

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.

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.

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.

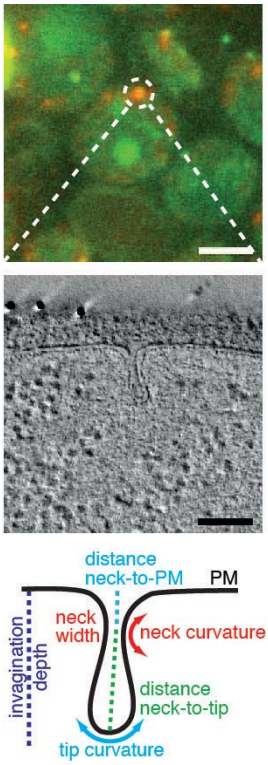


Figure 1: Correlated fluorescence and electron microscopy can be used to locate a defined intermediate stage in endocytosis and extract quantitative information. This can be applied to multiple stages to understand the whole process (Kukulski *et al.* 2012).

Future projects and goals

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.

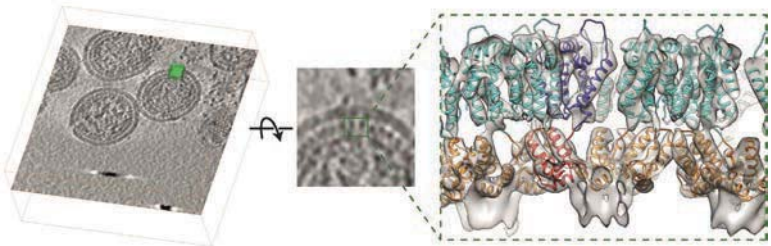


Figure 2: An optimised combination of cryo-electron tomography and image processing, developed in the lab, allowed the structure of the immature HIV-1 capsid to be resolved *in situ* - within heterogeneous virus particles. (Schur *et al.* 2015).

Functional mechanisms of complex enzymes involved in RNA metabolism and methodology development for drug design

Teresa Carlomagno

PhD 1996, University of Naples Federico II.

Postdoctoral research at Frankfurt University and Scripps Research Institute.

Group leader at the Max Planck Institute for Biophysical Chemistry, Göttingen, 2002-2007.

Group leader at EMBL since 2007. Joint appointment with the Genome Biology Unit.

SELECTED REFERENCES

Lapinaite A, *et al.* (2013) The structure of the box C/D enzyme reveals regulation of RNA methylation. *Nature* 502, 519-23

Marchanka, A, *et al.* (2013) A suite of solid-state NMR experiments for RNA intranucleotide resonance assignment in a 21 kDa protein-RNA complex. *Angew Chem. Int. Ed. Engl.* 125, 10180-5

Skjaerven L, *et al.* (2013) Accounting for conformational variability in protein-ligand docking with NMR-guided rescoring. *J. Am. Chem. Soc.* 135, 5819-27

Ballare C, *et al.* (2012) Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nat. Struct. Mol. Biol.* 19, 1257-65

The Carlomagno group uses NMR spectroscopy in combination with biochemical and biophysical techniques to study the structure and dynamics of biomolecular complexes.

Previous and current research

Our group focuses on studying: i) structure-activity relationships of RNP complexes involved in RNA processing; and ii) the interaction of small drugs with cellular receptors.

Our work aims at describing the features of RNA-protein recognition in RNP complex enzymes and at characterising the structural basis for their function. Recently, we investigated the nucleolar multimeric Box C/D RNP complex responsible for the methylation of the 2'-O-position in rRNA. During the biosynthesis and processing of the pre-rRNA transcripts, post-transcriptional modifications of ribonucleotides occur in functionally important regions, such as at intersubunit interfaces, decoding and peptidyltransferase centers. Among the possible modifications, 2'-O-ribose methylation was shown to protect RNA from ribonucleolytic cleavage, stabilise single base pairs, serve as chaperone, and impact folding at high temperatures. We solved the structures of the 400 kDa enzyme in solution. A large conformational change is detected upon substrate binding, revealing an unexpected 3D organisation of the catalytic RNP (figure 1). In addition, the structure revealed an unsuspected mechanism of sequentially controlled methylation at dual sites of the rRNA, which might have important implications for ribosome biogenesis.

Conformational switches occur in macromolecular receptors at all cellular levels, dependent on the presence of small organic molecules that are able to trigger or inhibit specific cellular processes. In a second area of research, we develop both computational and experimental tools to access the structure of large receptors in complex with function regulators. We are the developers of INPHARMA, a novel approach to structure-based drug design that does not require crystallographic structures of the receptor-drug complex (figure 2). We apply our methods to study the functional mechanisms of anti-cancer drug-leads, designed as inhibitors of kinases, proteasome and membrane receptors.

Future projects and goals

My team uses a multidisciplinary approach combining nuclear magnetic resonance spectroscopy (NMR), and biochemical, biophysical and computational methods. Our philosophy is to tackle the structure of high molecular weight complexes, whose large size impedes a detailed structural description by NMR only, with an array of different complementary methodologies, such as segmental and specific labelling of both proteins and RNAs, small angle scattering (SAS), electron microscopy (EM), electron paramagnetic resonance (EPR), fluorescence resonance energy transfer (FRET), mutational analysis and biochemical experiments (e.g. cross-link). With our complementary approach it is possible to examine RNP particles in solution, in their native environment, where they preserve both their structure and dynamic properties.

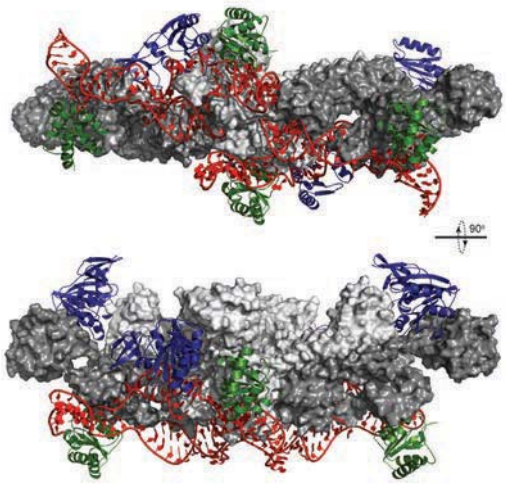


Figure 1: Structure of the RNA-methylating machinery Box C/D RNP shows that only one pair of proteins (blue) can add methyl groups to the RNA (red) at a time (Lapinate *et al.*, 2013).

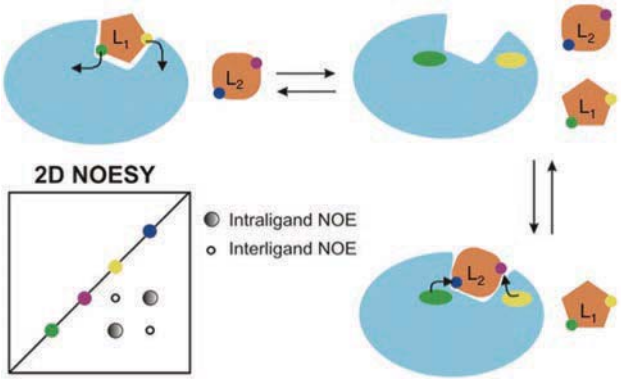


Figure 2: Schematic representation of the principle of the INPHARMA NOEs.

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

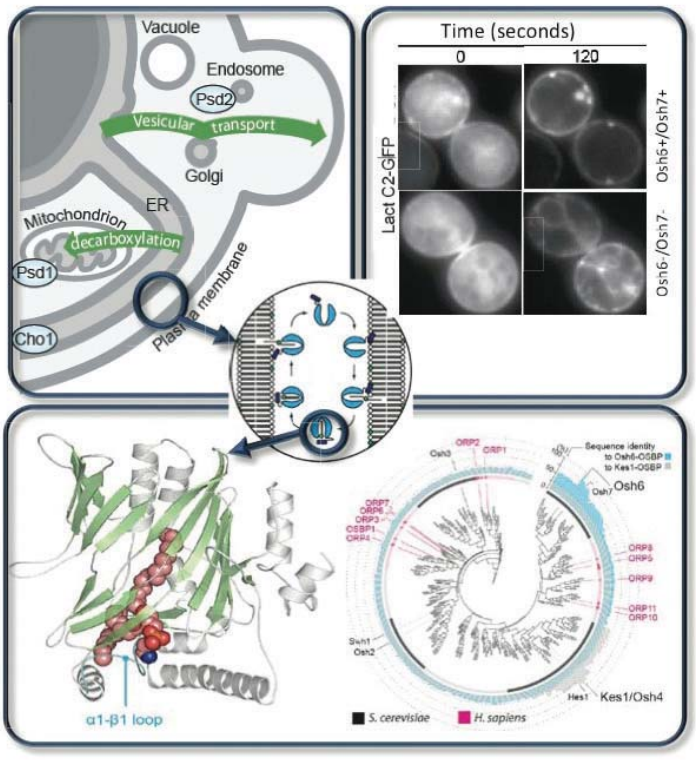
Previous and current research

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.

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

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.



Future projects and goals

- 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

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

Dinkel H, *et al.* (2013) The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res.* 42, 259-66

Gibson, TJ, *et al.* (2013) The transience of transient overexpression. *Nat. Methods* 10, 715-21

Van Roey K, *et al.* (2013) The switches.ELM resource: a compendium of conditional regulatory interaction interfaces. *Sci. Signal* 6, rs7

The Gibson group investigates protein sequence interactions, undertakes computational analyses of macromolecules, and develops tools to enhance sequence analysis research.

Previous and current research

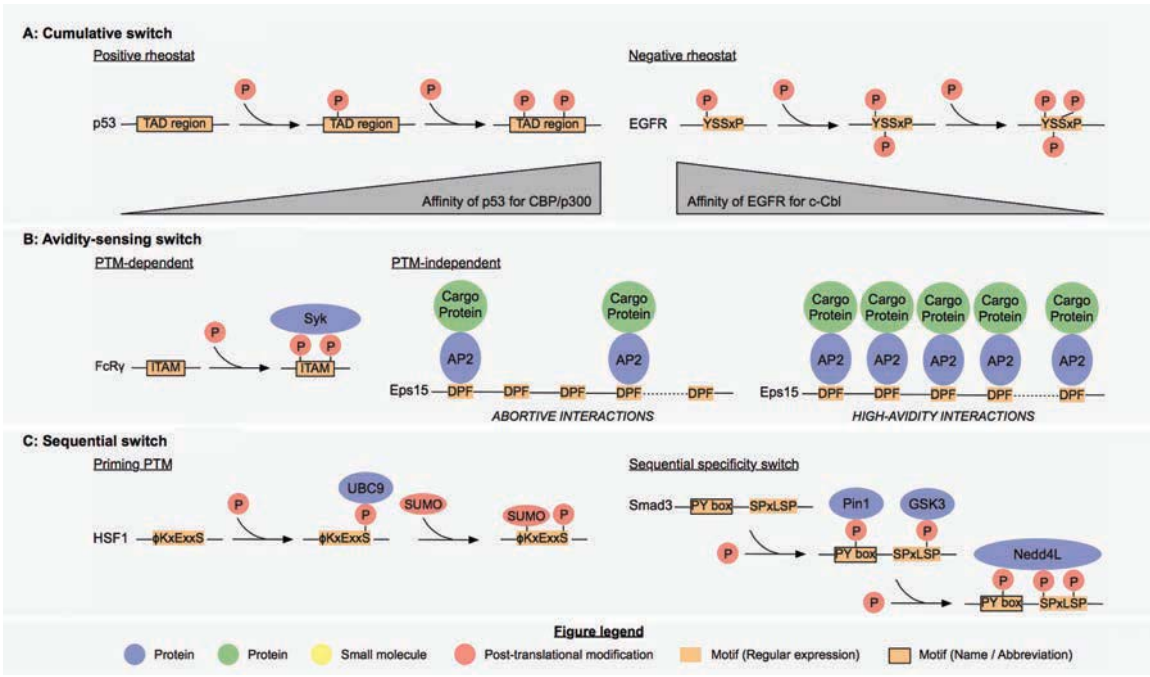
Regulatory decisions during eukaryotic cell signalling are made within large dynamic protein complexes by in-complex molecular switching (see Van Roey *et al.*, 2012, 2013). Cell regulation is networked, redundant and, above all, cooperative: the proteins involved make remarkable numbers of interactions, and thus have highly modular architectures. This goes against the traditional but misleading 'kinase cascade' metaphor. Regulatory proteins make remarkable numbers of interactions, with the corollary that they also have highly modular architectures.

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 – for example, in a recent interdisciplinary collaboration we performed bioinformatics analyses of the SxIP motif that is critical for the regulation of microtubule ends.

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 projects and goals

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 are looking at interaction networks and systems in stem cells and primary cilia.



Schematic of cumulative and sequential regulatory switches involving linear motif interactions (see Van Roey *et al.*, 2012).

Structural light microscopy, single molecule spectroscopy

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.

SELECTED REFERENCES

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

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

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

Milles S & Lemke EA (2011) Single molecule study of the intrinsically disordered FG-repeat nucleoporin 153. *Biophys. J.* 101, 1710-9

Previous and current research

The Lemke group uses an interdisciplinary approach to elucidate the nature of naturally 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 disorderedunfolded 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.

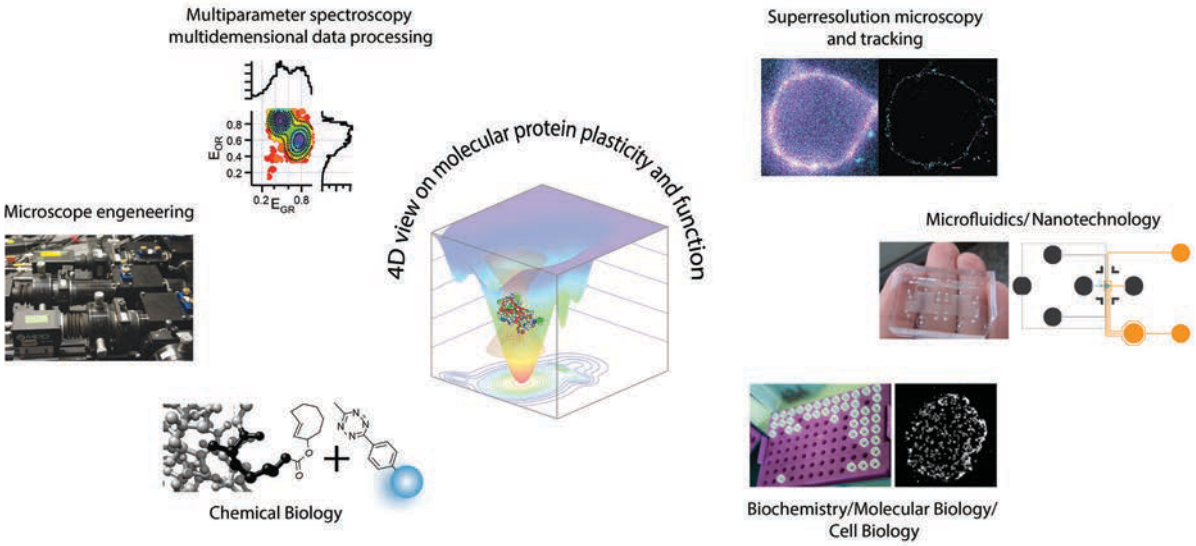
Most common strategies for probing protein structure are incompatible with the highly dynamic nature of molecular disorder. In contrast, single molecule and super-resolution techniques, which directly probe the distribution of molecular events, can reveal important mechanisms that otherwise remain obscured. 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.

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 posttranslational 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 projects and goals

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.



We interface a large set of tools 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. We also aim to explore the potential of these IDP biopolymers for novel applications in the life and material sciences.

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.* (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.* (2012) Impact of stoichiometry representation on simulation of genotype-phenotype relationships in metabolic networks. *PLoS Comput. Biol.* 8, e1002758

Brochado AR, *et al.* (2010) Improved vanillin production in baker's yeast through *in silico* design. *Microb. Cell Fact.* 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.

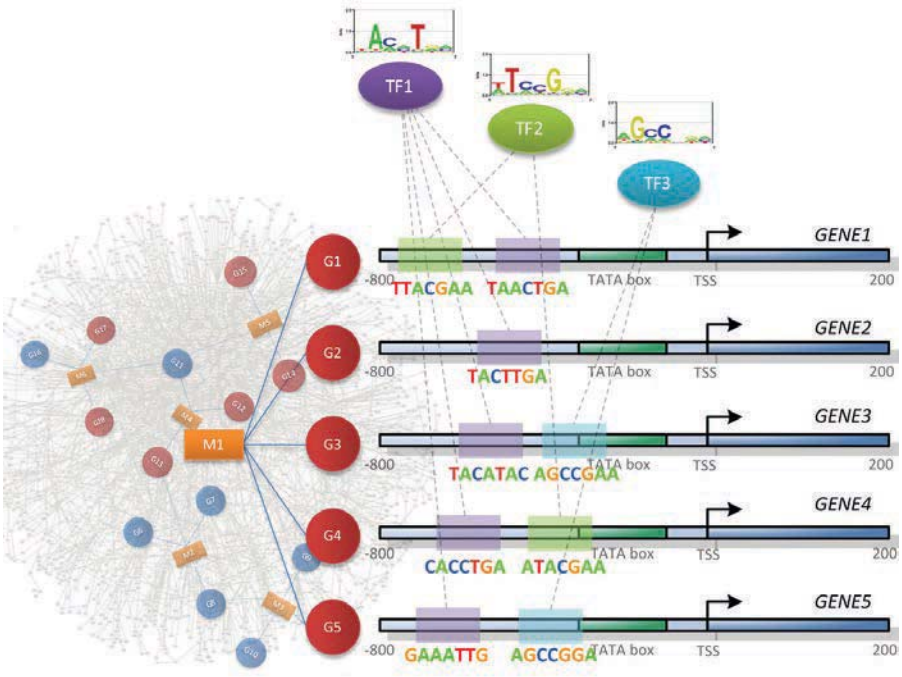
Previous and current research

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 maximise fitness, cells respond to environmental or genetic perturbations through a highly coordinated regulation of metabolism. The research in our group focuses on understanding the basic principles of operation and regulation of metabolic networks. We are particularly interested in developing models connecting genotypes to metabolic phenotypes (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 and more). This combination of computational and experimental approaches has previously enabled us to improve yeast cell factories producing vanillin – a popular flavouring agent. Currently we are 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 bacterial 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 redesign strategies towards improving the productivity of cell factories. These strategies guide our experimental implementation, which in turn help us to further improve the design algorithms in an iterative fashion.



Reporter algorithm integrates omics data with metabolic network and thereby identifies metabolic regulatory hotspots. M1 – metabolite; G1-5 – upregulated genes; purple/green/blue circles & squares – transcription factors and corresponding binding motifs.

Future projects and goals

We are interested in expanding the scope of our computational and experimental models to gain mechanistic insight into 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.

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

Fromm SA, *et al.* (2015) Seeing tobacco mosaic virus through direct electron detectors. *J Struct Biol* 189, 87-97

Desfosses A, *et al.* (2014) SPRING - an image processing package for single-particle based helical reconstruction from electron cryomicrographs. *J. Struct. Biol.* 185, 15-26

Guichard P, *et al.* (2012) Cartwheel architecture of *Trichonympha* basal body. *Science* 337, 553

Previous and current research

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.

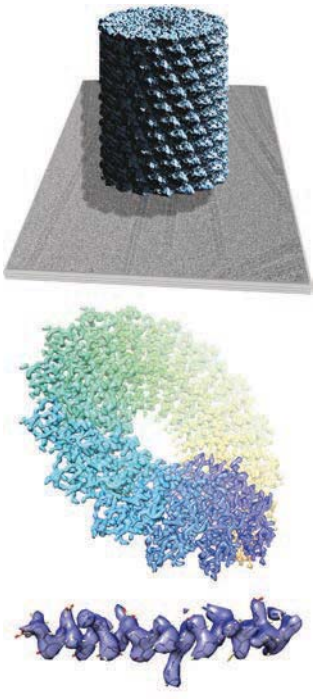
We study the structures of molecular assemblies using biochemical and biophysical techniques, and subsequently visualise them by electron cryomicroscopy (cryo-EM). By this technique, large macromolecular structures and multi-protein complexes can be studied in their near-native environment without the 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.

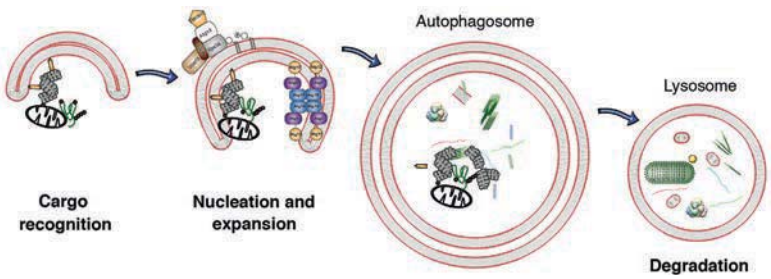
Future projects and goals

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.

Group leader at EMBL since 2014.

SELECTED REFERENCES

Castelnuovo M, *et al.* (2014) Role of histone modifications and early termination in pervasive transcription and antisense-mediated gene silencing in yeast. *Nucleic Acids Res.* 42, 4348-62

Kasowski M, *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 *et al.* (2012) Gene loops enhance transcriptional directionality. *Science* 338, 671-5

The Zaugg group uses computational approaches to investigate 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.

Previous and current research

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.

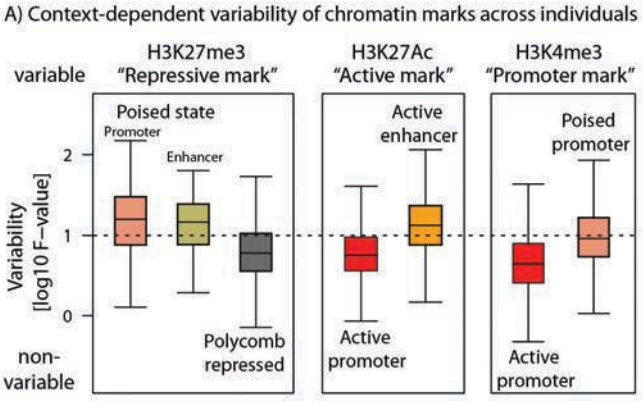
To gain a better mechanistic insight into potential causes of known genotype-disease associations our lab is developing and applying computational tools to investigate the variability of molecular phenotypes across individuals and linking them to genetic variation. In addition, since many of the disease-associated SNPs are located in regulatory elements, we have a strong interest in mining functional genomics data to further our understanding of gene regulatory mechanisms.

Our recent findings suggest a genetic basis of chromatin states, challenging the traditional view of chromatin being an epigenetic mark. Interestingly, there is a dramatic discrepancy in variability among individuals between enhancer elements (most variable) and gene expression (least variable). We further found that regulatory elements that are variable among individuals are enriched for SNPs that have previously been found to be associated 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.

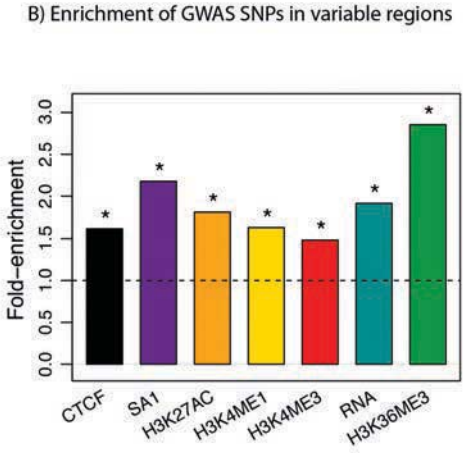
Future projects and goals

In the future we will continue our work towards the understanding of complex traits and diseases along three lines of research:

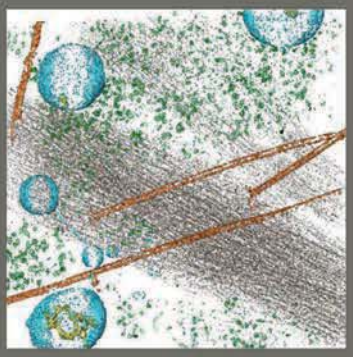
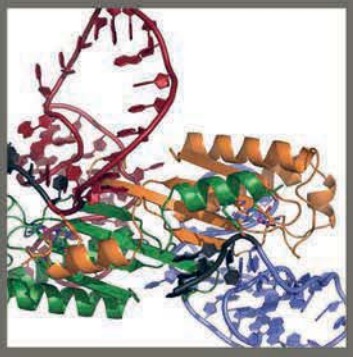
- We will apply our computational models to current genome-wide association studies to further our understanding about the known associations between genetic variants and complex diseases.
- We will expand our approach to include data from more downstream molecular phenotypes, such as protein expression and complex composition, to estimate the impact of genetic variation on the activity of complete biological pathways.
- We will use response to drugs as a model system to investigate the role of chromatin in mediating genotype-environment interactions across individuals.

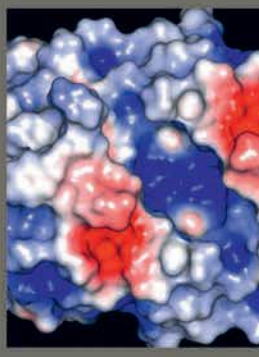


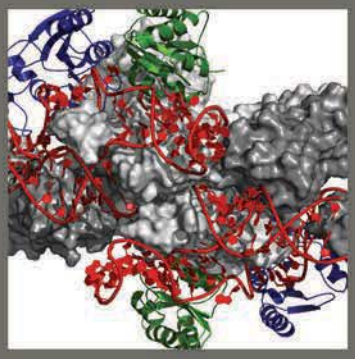
A) Chromatin marks in enhancers and poised states are highly variable among individuals whereas promoters tend to be consistent.

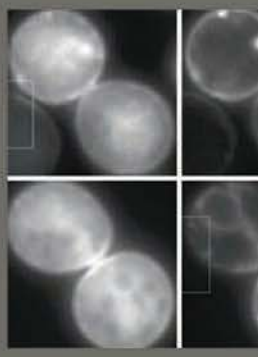
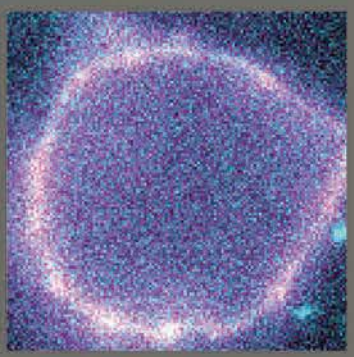
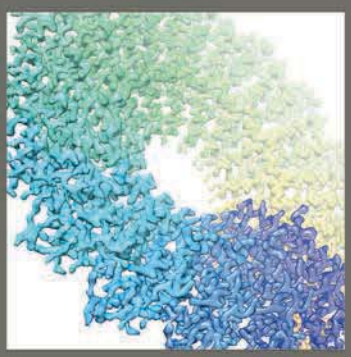


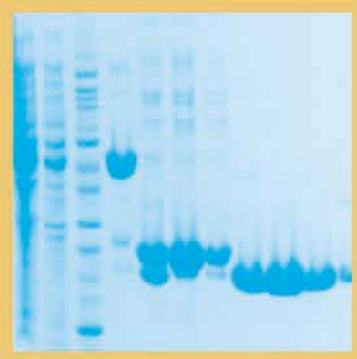
B) Variable regions are highly enriched for SNPs that have been associated with complex traits and diseases in genome-wide association studies (GWAS).

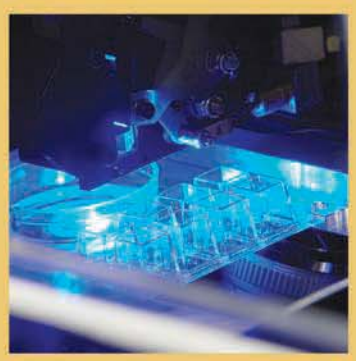



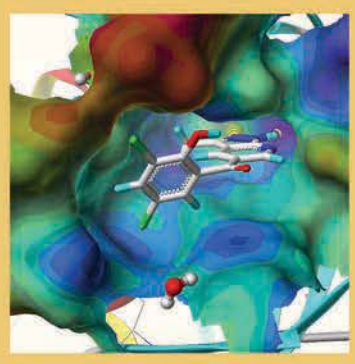



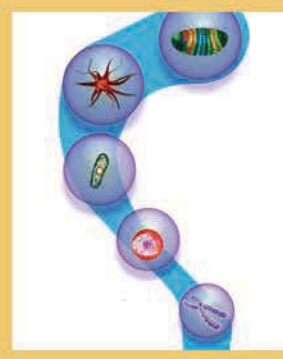


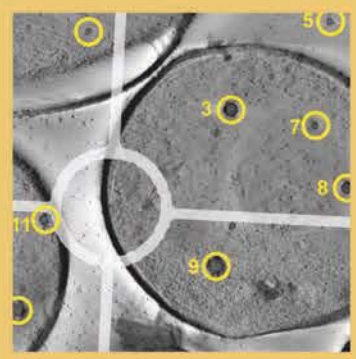




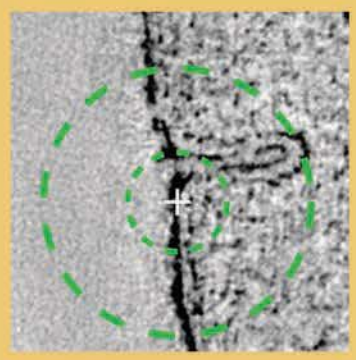


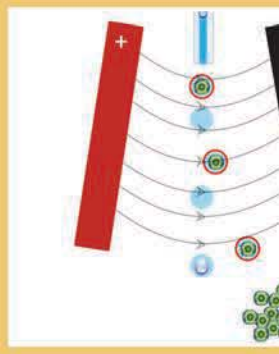


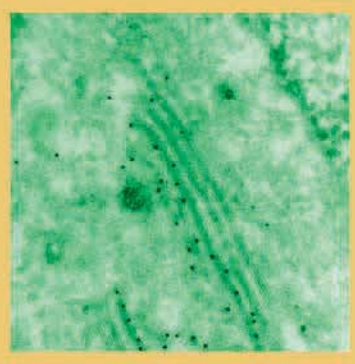


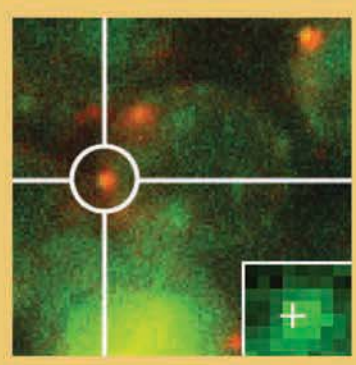


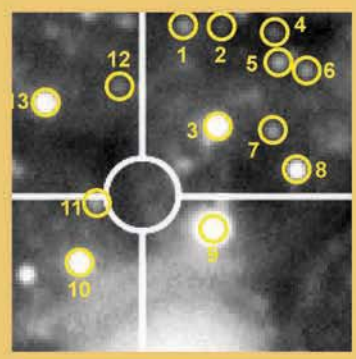

5
3
7
8
11
9




+






1
2
4
5
6
12
3
3
7
8
11
9
10

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, 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

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

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

Neumann B, *et al.* (2010) Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* 464, 721-7

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

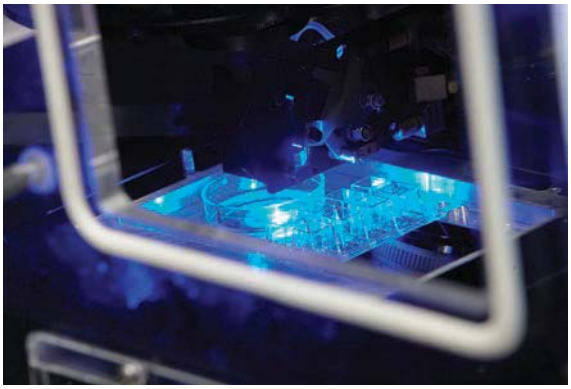
- The ALMF presently manages about 20 state-of-the-art microscope systems and 10 high-throughput microscopes from leading industrial companies.
- Several workstations for image analysis are provided.
- More than 50 visitors per year come to carry out their own experiments in the ALMF or to evaluate microscopy 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 manages 18 advanced microscope systems and eight 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.

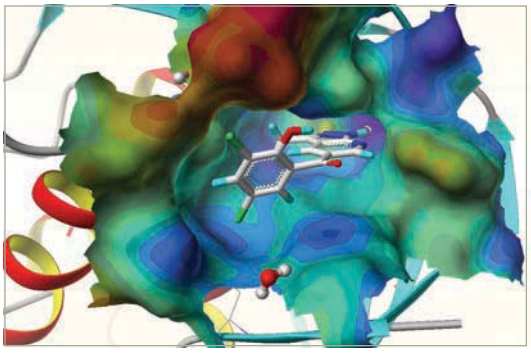
Services provided

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

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

Romero-Brey, I., *et al.* (2012) Three-dimensional architecture and biogenesis of membrane structures associated with hepatitis C virus replication. *PLoS Pathog.* 8, e1003056

Kukulski, W., *et al.* (2011) Correlated fluorescence and 3D electron microscopy with high sensitivity and spatial precision. *J Cell Biol.* 192, 111-9

The facility provides advanced expertise in electron microscopy – from sample preparation to image analysis – and for a large variety of biological samples ranging from macromolecules to tissues.

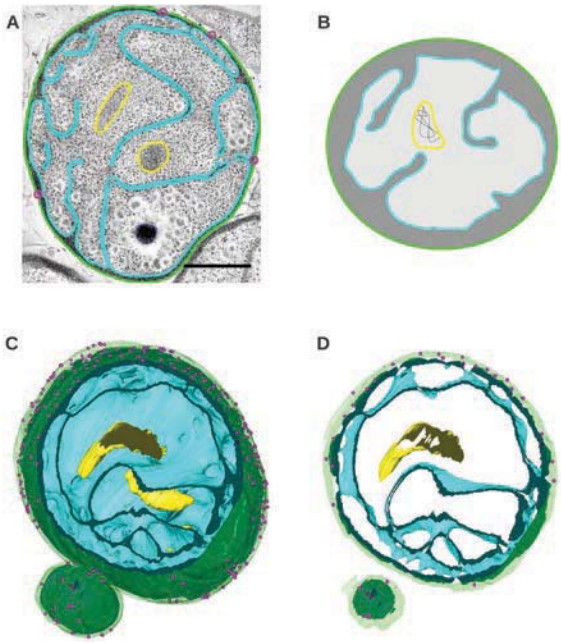
The EMCF activities cover a large spectrum of EM techniques with a major focus on sample preparation, immuno-localisation of proteins, ultrastructural analysis in 2D and 3D, and data processing. Staff in the facility can help you to define optimal experimental conditions for your project – we have experience 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: With the implementation of the Automated Serial Imaging with FIB-SEM (in 2014) the facility will develop its portfolio of 3D imaging applications (serial section, electron tomography (ET), serial ET). The technology presents new opportunities for the understanding of the cellular fine architecture as it offers high-resolution 3D imaging (5 nm, isotropic) of volumes containing up to several cells. With the Crossbeam Auriga 60 from Zeiss we can image cultured cells and tissues from different types of model organisms.



3D reconstruction of bacteria with a complex endomembrane system (from Santarella-Mellwig R *et al*, *Plos Biol* 2013).

Services provided

Techniques:

- Sample preparation for single particle analysis.
- Chemical fixation, high pressure freezing of cells and multi-cellular specimens.
- Resin embedding.
- Ultramicrotomy and cryo-ultramicrotomy (Tokuyasu technique).
- Immuno-labelling and TEM imaging.
- TEM tomography and FIB-SEM imaging.
- Correlative light and electron microscopy.
- Image analysis and 3D cellular modelling.

Teaching and training:

- Organisation of basic and advanced courses.
- Personalised training for internal and external users and visitors.

Technology partners

FEI Company (our Transmission Electron Microscopes), Carl Zeiss (the Crossbeam), Leica Microsystems (ultramicrotomes, high pressure freezing and freeze substitution), AbraFluid (high pressure freezing).

Flow Cytometry Core Facility

Alexis Pérez González

PhD 2003, Center of Molecular Immunology & University of Havana, Cuba.

Postdoctoral research and cytometry lab manager at Gulbenkian Institute of Science, Oeiras, Portugal.

At EMBL since 2006. Facility Manager since 2012.

SELECTED REFERENCES

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

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

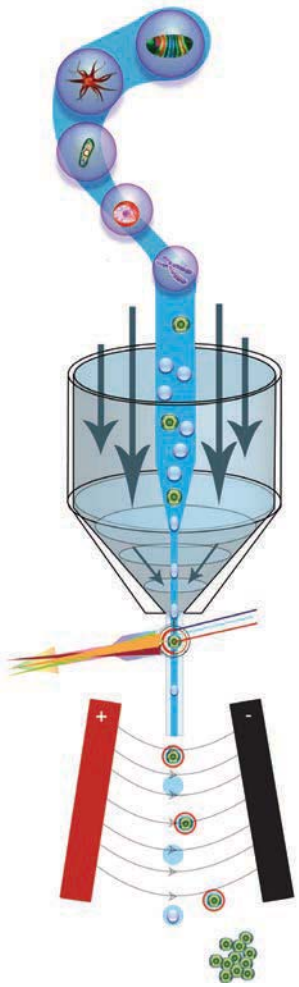
We offer a wide range of flow cytometric techniques. Our equipment adds flexibility in the preparation and execution of experiments, allowing different approaches to addressing scientific problems. Our facility strives to meet researchers' needs and enable the highest possible resolution in terms of 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.
- Chromosome karyotyping and sorting for DNA sequencing and proteomics studies.
- High-resolution analysis through photo saturation of dimly fluorescent bistable states in reworked bacterial signalling cascades.
- Establishment of clonal cell lines carrying fluorescent protein-tagged genome-edited genes for 4D life cell imaging, protein interactions assessment and protein concentrations during mitosis.
- Rmax: Development of a universal and sensitive method to evaluate cell sorters performance via sort recovery.
- Single cell sorting of high efficiencies for single cell quantitative genome, transcriptome and *in vitro* studies.



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

Services provided

- Complex multi-colour analysis of cell populations based on light scatter, fluorescent probes content and light intensities (including polarisation).
- Sorting of rare populations out of a heterogeneous particle mix. 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 training 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.

Genomics Core Facility

Vladimír Beneš

PhD 1994, Czech Academy of Sciences, Prague.

Postdoctoral research at EMBL.

Facility head since 2001.

SELECTED REFERENCES

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

Gupta I, *et al.* (2014) Alternative polyadenylation diversifies post-transcriptional regulation by selective RNA-protein interactions. *Mol. Syst. Biol.* 10, 719

Spornraft M, *et al.* (2014) Optimization of Extraction of Circulating RNAs from Plasma - Enabling Small RNA Sequencing. *PLoS ONE* 9,e107259

Zeller G, *et al.* (2014) Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol. Syst. Biol.* 10, 766

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 and cBot instruments, as well as MiSeq, NextSeq and Ion Torrent sequencers. 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 (including strand-specific libraries).
- 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.

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.

Protein Expression and Purification Core Facility

Hüseyin Besir

PhD 2001, Max Planck Institute of Biochemistry, Munich.

Postdoctoral research at Roche Diagnostics, Penzberg, and the Max Planck Institute of Biochemistry, Munich.

Facility head at EMBL since 2006.

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

Following each purification, we can perform biophysical analyses to ensure the quality of the purified sample in terms of correct folding and stability. Our facility also develops or evaluates new techniques and advanced protocols for protein production and purification and there is significant focus on developing time-saving solutions for these activities. Moreover, we are keeping stocks of a large number of expression vectors and bacterial strains for the users as well as preparing a collection of frequently used proteins for general use, which helps to considerably reduce the expenses of our users.

The facility produces and purifies proteins from E. coli*, insect, mammalian cells and sera, using a variety of chromatographic methods and provides support for biophysical characterisation of purified proteins.*

Major projects and accomplishments

We have evaluated new variants of our pETM-series expression vectors for *E. coli* that can now be used for sequence and ligation-independent cloning (SLIC). We have adapted vectors for insect and mammalian cells for the same cloning protocol. Using a single PCR product with the gene of interest, it is possible to integrate the insert into all of the vectors due to the universal overlaps that are present in the linearised vectors and the PCR product. A lethal ccdB gene in the original template vectors inhibits the growth of false positive colonies, which reduces the number of clones to test for the correct insert. With this new vector set, one can test the expression of a gene in different expression systems in parallel and avoid the redesigning of inserts for restriction-based cloning.

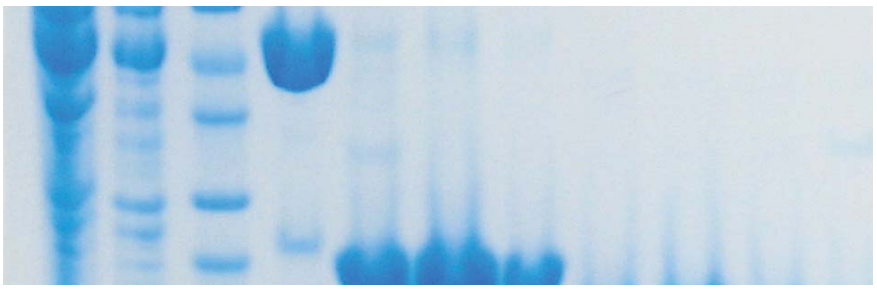
We have established a generic protocol for expression of fusion protein based on the small SUMO proteins and their highly specific protease SenP2. In most of our expressions, SUMO-fusion proteins showed high expression yields. In cases of insoluble product, we developed a protocol for proteolytic cleavage of the urea-denatured fusion protein, with the robust protease under conditions where other proteases show a poor performance. We can obtain pure, untagged proteins that are otherwise difficult to express or purify which include cytokines for cell culture or antigens for immunisation.

Services provided

- Expression and purification: proteins in *E. coli*, insect and mammalian cells.
- Scientific and technical advice to users at EMBL and external researchers.
- Preparing injection material for immunisations and purification of antibodies from serum and hybridoma supernatants.
- Providing quality analysis and biophysical characterisation of purified proteins (ITC, analytical ultracentrifugation, CD).
- Maintaining collections of expression vectors and bacterial strains.
- Producing frequently used enzymes and protein molecular weight markers for general use within EMBL.
- Developing and testing new vectors and protocols.
- Access to protocols and vector sequence information on the website.
- Caring for equipment for protein production and analysis.

Technology partners

We are open to collaborations with academic or industrial partners to evaluate new products or technological developments. Furthermore, we have initiated a network of protein facilities across Europe called P4EU (Protein Production and Purification Partnership in Europe (P4EU)) to improve information exchange and evaluation of new technologies.



SDS-PAGE analysis after purification of LIF by ion-exchange chromatography.

Proteomics Core Facility

Jeroen Krijgsveld

PhD 1999, University of Amsterdam, The Netherlands.

Postdoc at Utrecht University, The Netherlands and Harvard Medical School, Boston, USA.

Assistant Professor, Utrecht University, The Netherlands.

Team leader at EMBL since 2008.

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

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

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).

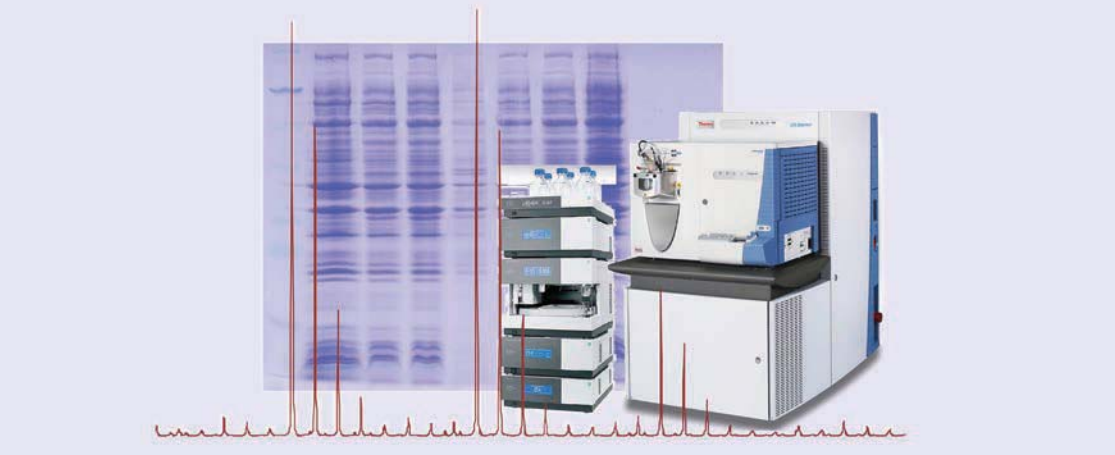
Services provided

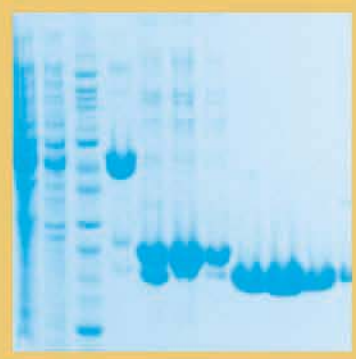
Proteomics:

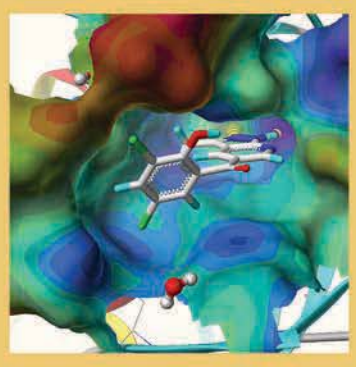
- Protein identification from gel or in solution.
- High resolution and high mass-accuracy MS, MSMS, and LC-MSMS (Thermo Orbitrap Velos Pro, Q-Exactive, and Orbitrap Fusion) for identification and quantification of proteins in complex mixtures.
- Ion trap (Bruker HCT) LC-MSMS for routine identification of proteins from coomassie and silver-stained gels.
- Triple-quad mass spectrometry (Thermo Vantage) for targeted protein analysis.
- Protein quantification by stable-isotope labelling (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).

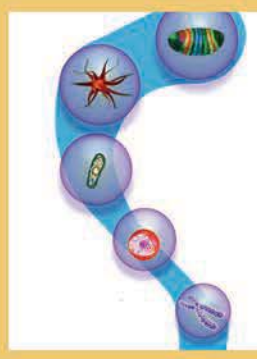
Analysis of intact proteins:

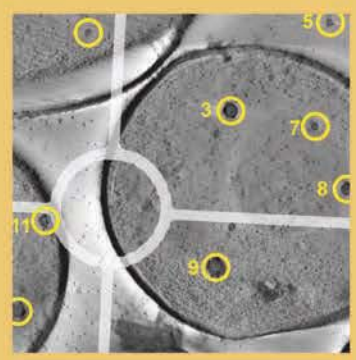
- Molecular weight determination of intact proteins by ESI mass spectrometry (high-mass Q-TOF).
- Determination of N- and C-termini of proteins and products of limited proteolysis.
- Verification of incorporation of non-natural amino acids.

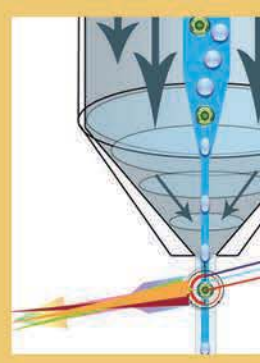


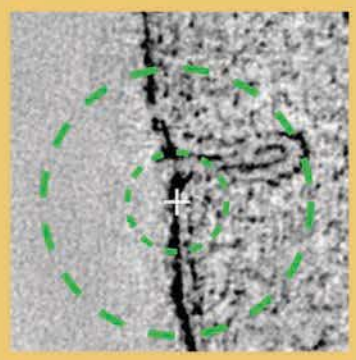


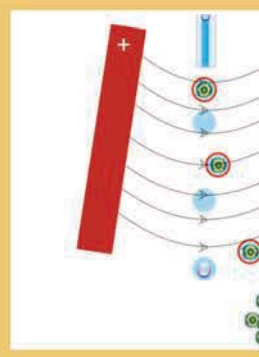


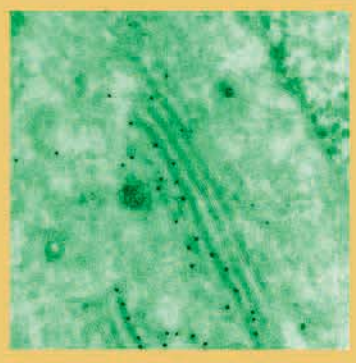


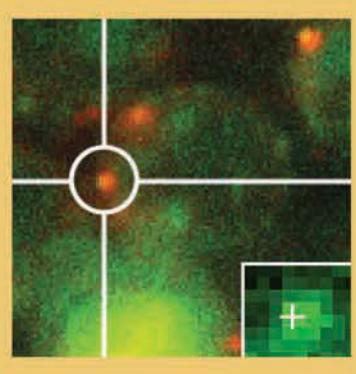

5
3
7
8
11
9

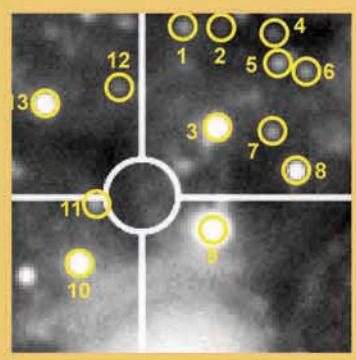


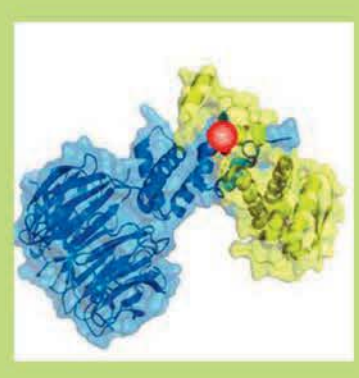


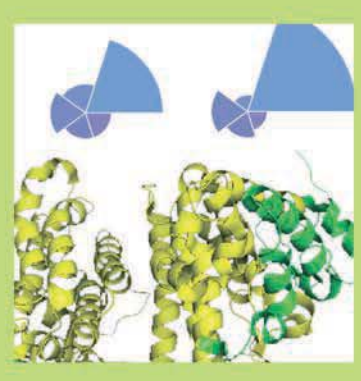

+

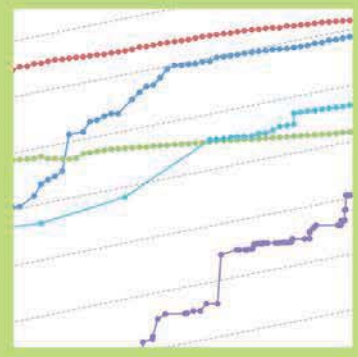




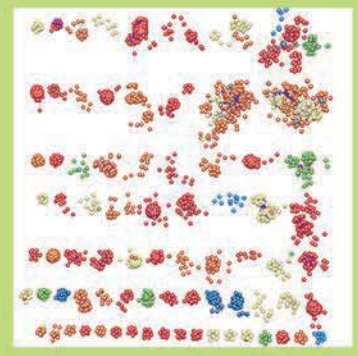

12
1
2
4
5
6
3
3
7
8
11
9
10





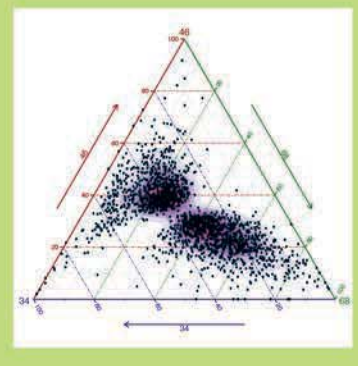


1974 - 2014 Research Highlights



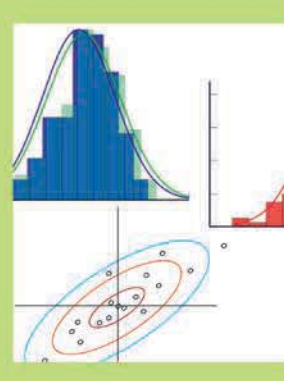
A new way to predict protein interactions by virtue of the fact that sometimes proteins working together will fuse into a single, multifunctional polypeptide. Their algorithm is still used today to determine co-functionality for thousands of pairs of proteins.
Enright AJ, *et al.* (1999) Protein interaction maps for complete genomes based on gene fusion events. *Nature* 402, 86-90

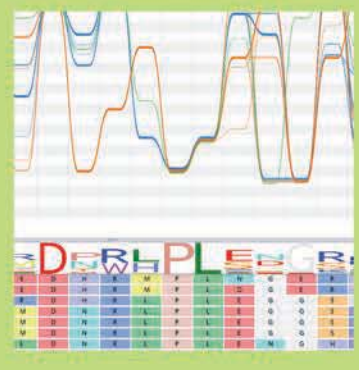
An international scientific collaboration produced a draft sequence of the human genome and made it freely available in the public domain. This act had a profound impact on the advance of biology, as it allowed scientists the world over a to freely explore this extraordinary trove of information about human development, physiology, medicine and evolution.
Lander ES, *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860-921

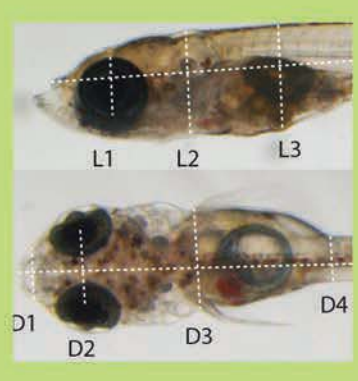


A detailed map of genome function that identifies four million gene 'switches'. The ENCODE project published over 30 papers under open-access license in several different journals, with the contents linked by topic and united for optimum exploration in a single interface provided by ***Nature***. A virtual machine allows readers to explore the data in context and reproduce the experimental conditions.
ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57-74

A novel, scalable approach to the long-term archiving of data, using the 'natural' storage archive provided by DNA itself. The new method involves translating binary digital files into non-repeating strings of A, T, G and C and – crucially – applying an error-correction algorithm similar to those applied in everyday technologies such as mobile phone transmission.
Goldman N, *et al.* (2013) Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* 494, 77–80







EMBL-EBI Hinxton

European Bioinformatics Institute

EMBL-EBI is probably best known worldwide for its provision of biological information and bioinformatics services. However, about 20% of the institute is devoted to curiosity-driven research using computational approaches to unravel the secrets of life.

The development of new technologies provides a constant driver for innovative research into processing and analysing the data generated. For example, the wide uptake of next-generation sequencing by life scientists has led to unprecedented growth in sequence data. These data require novel algorithms to turn them into reliable information, and perhaps even more challenging is to use these new data to obtain novel insights into biological processes.

Research at EMBL-EBI is carried out both in groups devoted solely to research and in some of the larger service teams that have associated research activities. All researchers have computational approaches as their major focus, but most also collaborate closely with experimentalists and often generate experimental data themselves. Our research is highly collaborative within EMBL as well as with many external colleagues. We are highly interdisciplinary; our faculty comprises scientists who originally trained in biology, physics, chemistry, engineering, medicine and mathematics. We develop novel algorithms and protocols for handling data, such as checking the quality of the data; interpret data; and integrate data to generate new knowledge. We use this information to develop novel hypotheses about the basic molecular processes of life.

Although we are united in using computers, the biological questions we address and the algorithms we develop and use are very diverse. We explore biological questions spanning genome evolution, transcriptional regulation and systems modelling of basic biological processes and disease. For example, different groups are investigating the molecular basis of ageing; the differentiation of stem cells; the basis for neuronal plasticity; and the early development of brain structure. Others are exploring regulation through epigenetics or RNA processing; how phenotype is related to genotype both in mice and humans; and how new enzyme reactions appear during evolution. All our discoveries are published in peer-reviewed journals but in addition, as part of these studies, our researchers often develop novel bioinformatics services, which are usually made freely available for all users so that our work helps facilitate new discoveries throughout the global scientific community.

Increasingly, much of our work is related to problems of direct medical significance, and with the emergence of personal genomes we are very conscious of the need to contribute to the translation of this new knowledge into medicine and the environment. This process is just beginning and will provide many challenges to computational biologists over the coming years.

Janet Thornton
Director, EMBL-EBI

Proteins: structure, function and evolution

Janet Thornton

PhD 1973, King's College, London, and NIMR.

Postdoc at Oxford University, NIMR and Birkbeck College. Lecturer, Birkbeck College, 1983-1989.

Professor of Biomolecular Structure, UCL 1990-2001.

Bernal Professor at Birkbeck College, 1996-2001. Director, Centre for Structural Biology, Birkbeck College and UCL, 1998-2001.

Director of EMBL-EBI since 2001.

SELECTED REFERENCES

Rahman SA, *et al.* (2014) EC-BLAST: a tool to automatically search and compare enzyme reactions. *Nat Methods* 11, 171-4

Tullet JM, *et al.* (2014) DAF-16/FoxO directly regulates an atypical AMP-activated protein kinase gamma isoform to mediate the effects of insulin/IGF-1 signaling on aging in *Caenorhabditis elegans*. *PLoS Genet* 10, e1004109

Papatheodorou I, *et al.* (2014) Comparison of the mammalian insulin signalling pathway to invertebrates in the context of FOXO-mediated ageing. *Bioinformatics* 30, 2999-3003

Martinez Cuesta S, *et al.* (2014) The evolution of enzyme function in the isomerases. *Curr Opin Struct Biol* 26, 121-30

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.

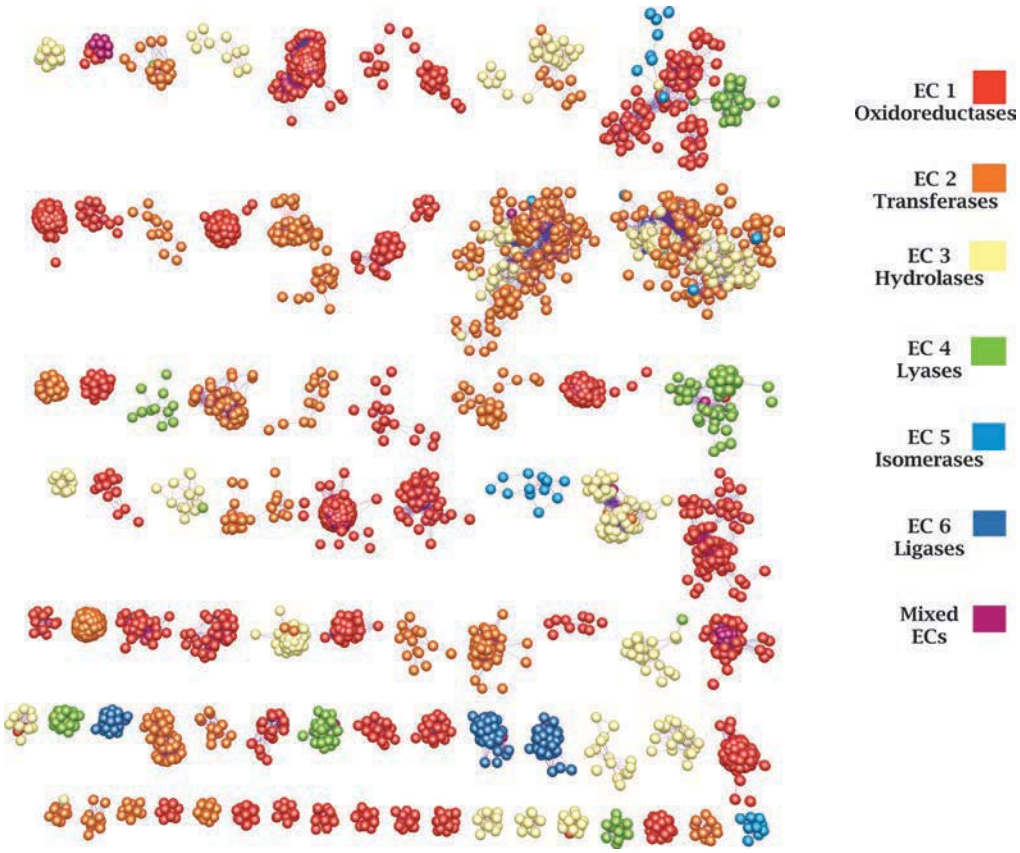
Previous and current research

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 explore how enzymes perform catalysis by gathering relevant data from the literature and developing 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 especially those variants related to diseases. 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 life span.

Future projects and goals

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 continue our study of reactions and use this new knowledge to improve chemistry queries across our databases. We will study sequence variation in different individuals and explore how genetic variations impact on the structure and function of a protein and sometimes cause disease. Using evolutionary approaches, we hope to improve our prediction of protein function from sequence and structure. We will continue our ageing studies, exploring longevity sub-phenotypes and trying to identify small molecules that might modulate lifespan in the model organisms.



Characterising the universe of enzyme reactions using EC-BLAST. Clustering of 5,073 representative reactions, using a combination of bond and reaction-center similarity scores. Each sphere represents one reaction, colored by primary IUBMB EC (Enzyme Commission) class. All reaction similarity clusters with P <0.01 and cluster size of more than ten reactions are shown.

Note that some clusters include enzymes of different functions and primary EC classes (shown in different colours). Despite this, their clustering illustrates the similarity of the reactions they catalyse.

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

Finn RD, *et al.* (2014) Pfam: the protein families database. *Nucleic Acids Res.* 42, 222-30

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,*et al.* (2010) Quantifying the mechanisms of domain gain in animal proteins. *Genome Biol.* 11, R74

Bateman A, *et al.* (2009) Phospholipid scramblases and Tubby-like proteins belong to a new superfamily of membrane tethered transcription factors. *Bioinformatics* 25, 159-62

Previous and current research

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.

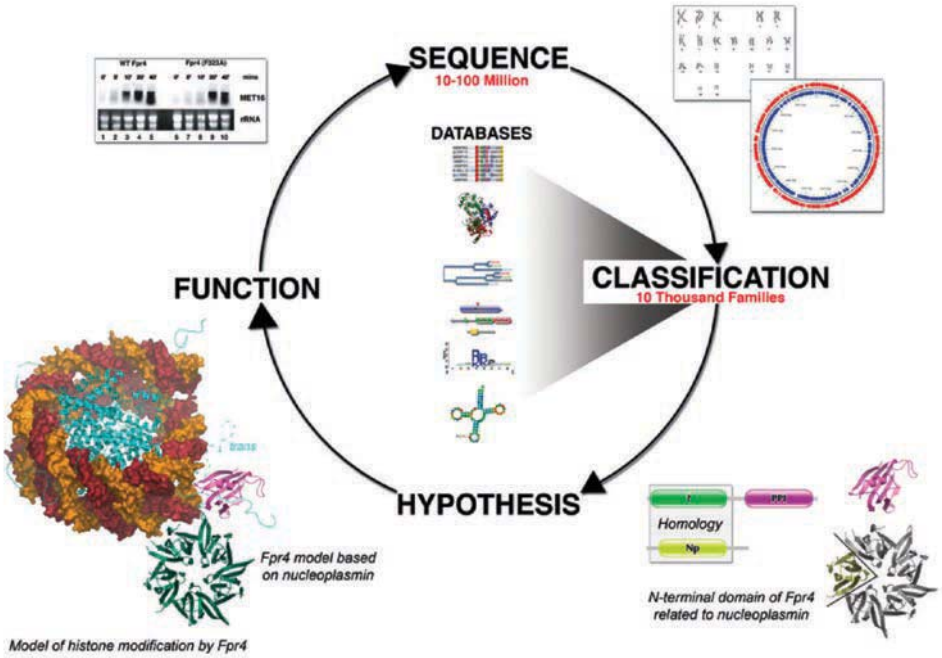
The Bateman group endeavors to classify proteins and certain RNAs into functional families with a view to producing a 'periodic table' of these molecules.

Our 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 it. Current research is looking at identification of non-coding RNAs and understanding the function through computational analysis.

Future projects and goals

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 in two directions. Firstly, we will investigate the plasticity of the protein interaction network between individuals. To do this we will identify natural human variation such as SNPs and CNVs that rewire the protein interaction network.

We will also 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.



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 Beltrao

PhD 2007 in Biology, 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

Rinschen MM, *et al.* (2014) Phosphoproteomic analysis reveals regulatory mechanisms at the kidney filtration barrier *J Am Soc Nephrol* 25,1509-1522

Beltrao P, *et al.* (2013). Evolution and functional cross-talk of protein post-translational modifications *Mol. Sys. Biol.* 9, 714

Swaney DL, *et al.* (2013) Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation *Nat. Methods* 10, 676-82

Beltrao P, *et al.* (2012) Systematic functional prioritization of protein post-translational modifications *Cell* 150, 413-25

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.

Previous and current research

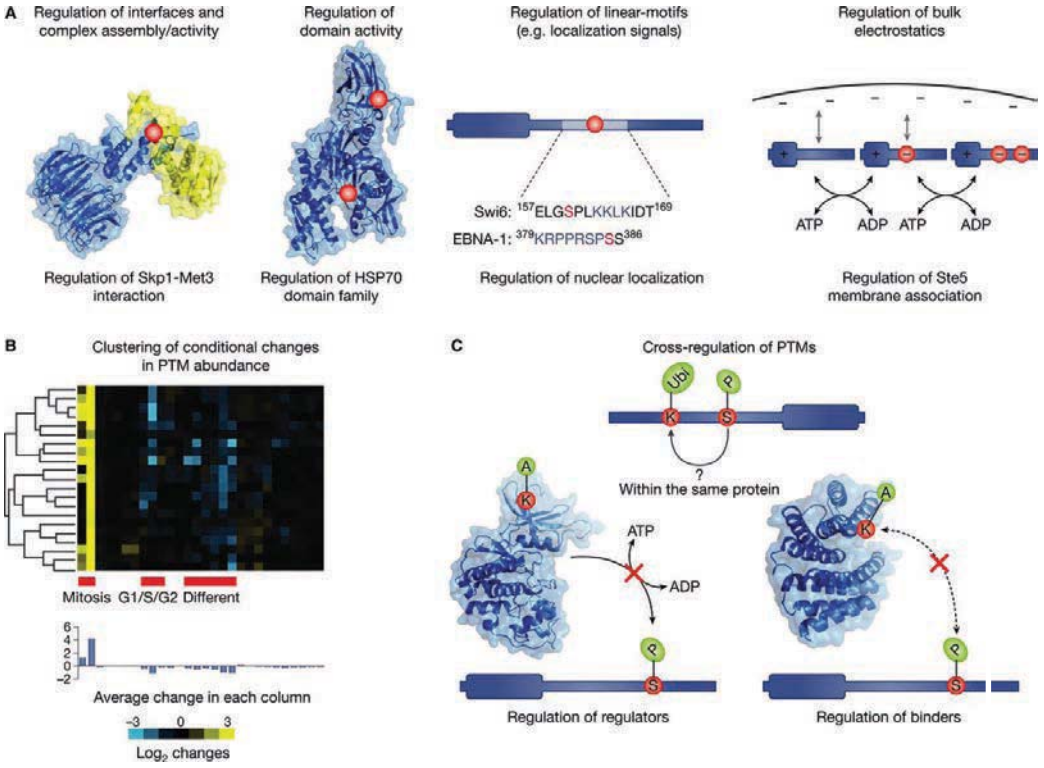
Our group is interested in understanding how novel cellular functions arise and diverge during evolution. We study the molecular sources of phenotypic novelties, exploring how genetic variability that is introduced at the DNA level is propagated through protein structures and interaction networks to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

Changes in cellular interaction networks underpin variation in cellular responses and sensitivity to environmental perturbations or small molecules. As we model and study the evolution of cellular interaction networks, we begin to see how different individuals or species diverge in their response to drugs. Understanding this relationship will enable us to develop methods to predict how genetic changes result in specific sensitivity to drug combinations.

Future projects and goals

In 2015 we will continue to study the evolution of cellular interaction networks with a specific focus on post-translational regulatory networks, genetic and chemical–genetic networks. Our PTM-related work has been very focused on the study of conditional dependent regulation as well as phylogenetic based approaches to determine the age of a PTM. The genetic studies have primarily studied the condition dependence of genetic-interaction networks.



Functional role of post-translational modifications. PTMs act to change the activity of proteins through different mechanism and in response to different conditions. (A) Different mechanism used by PTMs to regulate protein activity. (B) Example of conditional regulation of phosphorylation sites. (C) Mechanism of cross-regulation between different PTM types. (Beltrao *et al. Mol Sys Bio* 2013).

Sequence algorithms and intra-species variation

Ewan Birney

PhD 2000, Sanger Institute.

At EMBL since 2000.

Associate Director of EMBL-EBI since 2012

SELECTED REFERENCES

Spivakov M, *et al.* (2014) Genomic and phenotypic characterization of a wild medaka population: towards the establishment of an isogenic population genetic resource in fish. *G3* 4, 433-45

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

Marti-Solano M, *et al.* (2014) Integrative knowledge management to enhance pharmaceutical R&D. *Nat Rev Drug Discov* 13, 239-40

Previous and current research

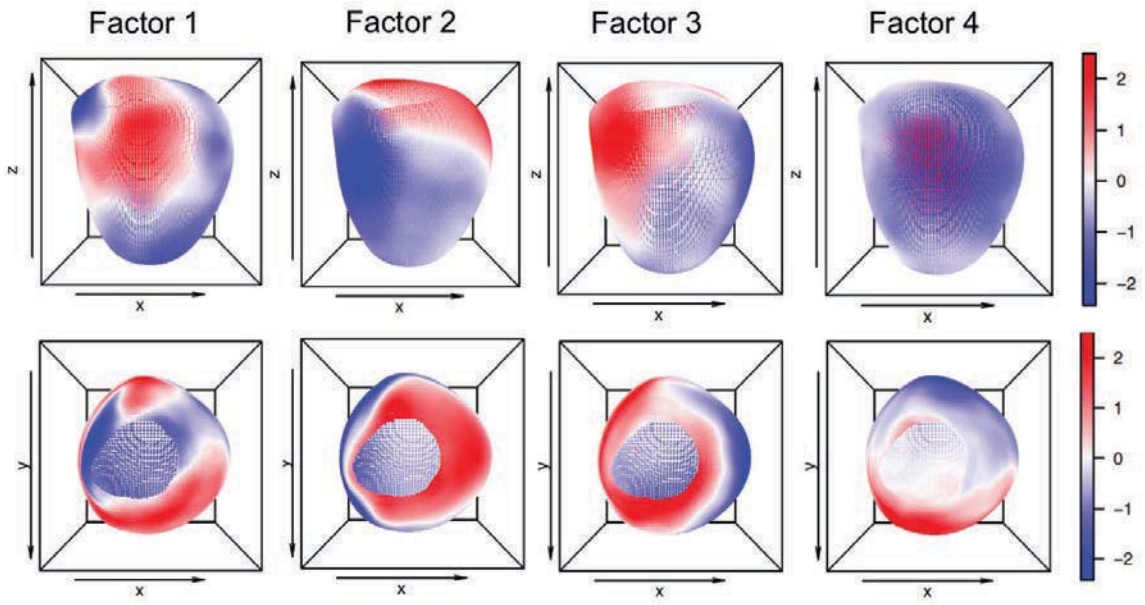
DNA sequence is at the heart of molecular biology and bioinformatics. Our group has a long-standing interest in developing sequencing algorithms. Over the past four years a considerable focus has been on compression, with theoretical and now practical implementations of compression techniques. Our 'blue skies' research includes collaborating with Dr Nick Goldman on a method to store digital data in DNA molecules. Both our groups continue to be involved in this area as new opportunities arise - including the application of new sequencing technologies.

We are also interested in the interplay of natural DNA sequence variation with cellular assays and 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 pursue association analysis for a number of molecular (e.g. RNA expression levels, chromatin levels) and basic biology traits in a number of species in which favourable populations are available, including human and *Drosophila*. We hope to expand this to a variety of basic biological phenotypes in other species, for example by establishing the first vertebrate near-isogenic wild panel in Japanese Rice Paddy fish (Medaka, *Oryzias latipes*).

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

Future projects and goals

In 2015 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.



Visualisations of low dimensional projections of sources of variance in health human hearts on idealised 3D model of the human heart.

Functional genomics research

Alvis Brazma

MSc in Mathematics 1982, University of Latvia.

PhD in Computer Science 1987, Moscow State University.

At EMBL-EBI since 1997.

SELECTED REFERENCES

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

Lappalainen T, *et al.* (2013) Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 501, 506-11

Gonzàlez-Porta M, *et al.* (2013) Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. *Genome Biol* 14, R70

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.

Previous and current research

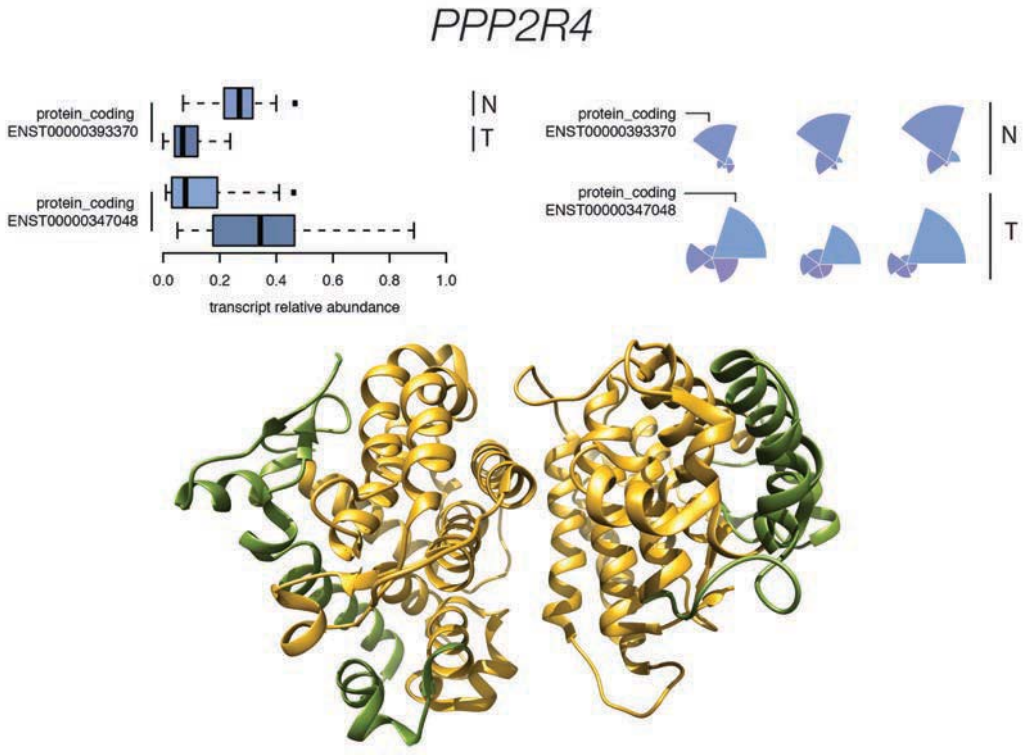
Our team complements the Functional Genomics service team, and focuses on developing new methods and algorithms and integrating new types of data across multiple platforms. We are particularly interested in cancer genomics and relationships between transcriptomics and proteomics. We collaborate closely with several groups at EMBL-EBI, including the Marioni, Stegle and Saez-Rodriguez groups.

Our group led the analysis of transcript isoform use and fusion gene discovery from lymphoblastoid cell lines of 465 individuals who participated in the 1000 Genomes Project, as a part of the GEUVADIS project. The human transcriptome contains in excess of 100,000 different transcripts. We analysed transcript composition in 16 human tissues and five cell lines to show that, in a given condition, most protein coding genes have one major transcript expressed at significantly higher level than others, and that in human tissues the major transcripts contribute almost 85% to the total mRNA.

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.

Future projects and goals

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 ICGC, 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 what implication this has on proteome.



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 of non-coding RNAs

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

Cossetti C, *et al.* (2014) Extracellular vesicles from neural stem cells transfer IFN-γ via Ifngr1 to activate Stat1 signaling in target cells. *Mol. Cell* 56,193-204

Camps C, *et al.* (2014) Integrated analysis of microRNA and mRNA expression and association with HIF binding reveals the complexity of microRNA expression regulation under hypoxia. *Mol. Cancer* 13, 28

Murray MJ, *et al.* (2014) Serum levels of mature microRNAs in DICER1-mutated pleuropulmonary blastoma. *Oncogenesis* 3, e87

De Fazio S, *et al.* (2011) The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature* 480, 259-63

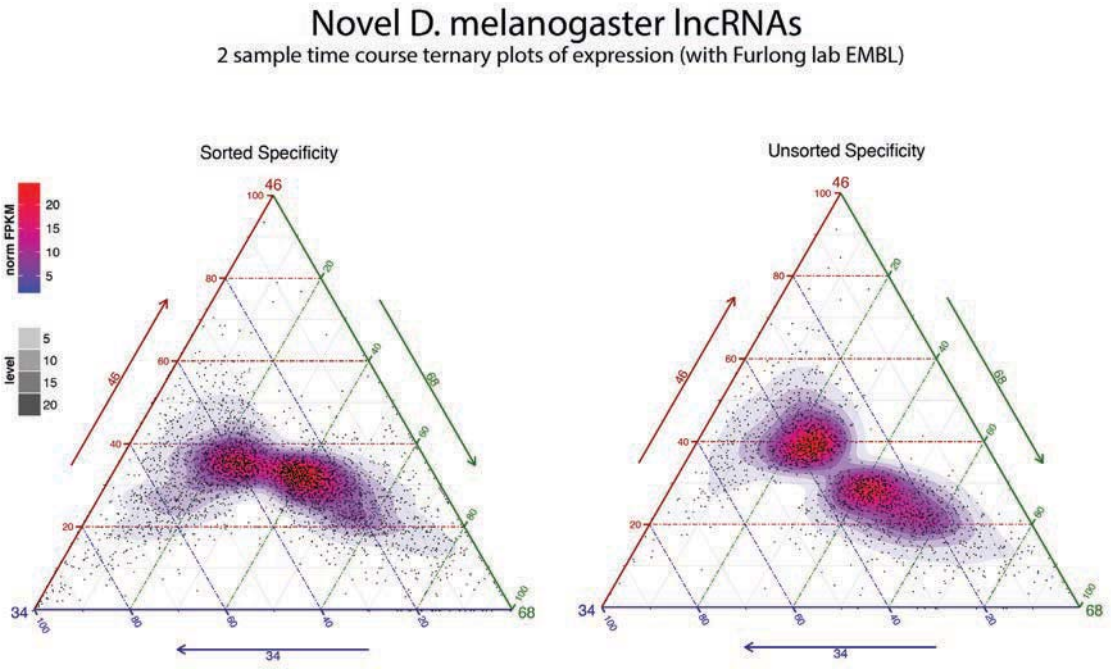
Previous and current research

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. There are many more layers of gene regulation than previously thought and it is likely that there are far more functional regulatory RNAs than proteins. Our work involves the development of algorithms, protocols and datasets for functional genomics to better understand non-coding RNAs. 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.

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

Future projects and goals

Our long-term goal is to combine regulatory RNA target prediction, secondary effects and upstream regulation into complex regulatory networks. Some current research foci in the lab include: prediction and functional analysis of long non-coding RNAs, prediction of microRNAs in newly sequenced species, phylogenetic analysis of long non-coding RNAs and an exploration of how non-coding RNAs are involved in germline development. We have close links to the clinical school at Addenbrookes hospital to assess the roles for non-coding RNAs in diagnosis of cancer. We will continue to build strong links with experimental laboratories that work on miRNAs 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).

Vertebrate Genomics

Paul Flicek

DSc 2004 Washington University.

Honorary Faculty Member, Wellcome Trust Sanger Institute since 2008.

At EMBL-EBI since 2005.

Team Leader since 2007, Senior Scientist since 2011.

SELECTED REFERENCES

Flicek P, *et al.* (2014) Ensembl 2014. *Nucleic Acids Res* 42, D749-55

Ritchie GR, *et al.* (2014) Functional annotation of noncoding sequence variants. *Nat Methods* 11, 294-6

Ballester B, *et al.* (2014) Multi-species, multi-transcription factor binding highlights conserved control of tissue-specific biological pathways. *Elife* 3, e02626

Eckersley-Maslin MA, *et al.* (2014) Random monoallelic gene expression increases upon embryonic stem cell differentiation. *Dev Cell* 28, 351-65

The Flicek team develops large-scale bioinformatics infra structure, explores the evolution of transcriptional regulation and develops algorithms to study epigenomic data.

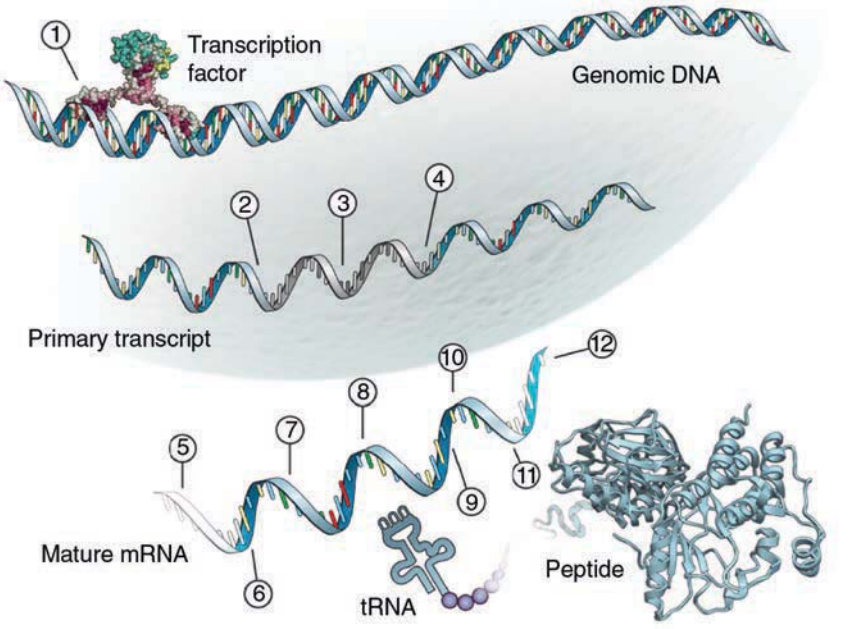
Previous and current research

The Vertebrate Genomics team creates and maintains the genomic resources of the Ensembl project and is responsible for data management for a number of large-scale projects including BLUEPRINT. Other resources include the International Genome Sample Resource (IGSR), which incorporates the data of the 1000 Genomes Project, and the curated GWAS catalog. 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 genome annotation as well as the evolution of transcriptional regulation and understanding tissue specificity. Based on comparative regulatory genomics techniques, our work provides some of the most definitive results on how transcription factor binding evolves across the vertebrate lineage. Our studies of tissue-specific gene regulation included clarifying the role of CTCF, cohesin and other genomic structural proteins.

Future projects and goals

In 2015 we will establish the International Genome Sample Resource (IGSR). The IGSR will maintain the data collections from the 1000 Genomes Project, and enable the addition of new samples and data types. Our research projects will continue to expand in the number of species, tissues and specific DNA–protein interactions explored. We will address these areas of research both in the context of our established collaborative projects with the Odom group at the University of Cambridge and as part of other collaborations. Our quest to understand the differentiation process and components of cell- and tissue-specific regulation will be enhanced by new collaborations with the GTEx Consortium, which aims to create a comprehensive public atlas of gene expression and regulation across multiple human tissues, and with the Centre for Therapeutic Target Validation.



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 varian
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.

Evolutionary tools for genomic analysis

Nick Goldman

PhD 1992 University of Cambridge.

Postdoctoral work at NIMR, 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

Parks SL, Goldman N (2014) Maximum likelihood inference of small trees in the presence of long branches. *Systematic Biology* 63, 798–811

Tamuri AU, *et al.* (2014) A penalized-likelihood method to estimate the distribution of selection coefficients from phylogenetic data. *Genetics* 197, 257–71

Behjati S, *et al.* (2014) Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* 513, 422-5

Schwarz RF, *et al.* (2014) Phylogenetic quantification of intra-tumour heterogeneity. *PLoS Computational Biology* 10, e1003535

Previous and current research

Evolution is the historical cause of the diversity of all life. Our group's research focuses on the development of data analysis methods for the study of molecular sequence evolution and for the exploitation of evolutionary information to draw powerful and robust inferences about phylogenetic history, molecular evolutionary processes and genomic function. The evolutionary relationships between all organisms require that we analyse molecular sequences with consideration of the underlying structure relating those sequences.

We develop mathematical, statistical and computational techniques to reveal information present in genome data, to draw inferences about the processes that gave rise to that data and to make predictions about the biology of the systems whose components are encoded in those genomes.

Our three main research activities are: developing new evolutionary models and methods; providing these methods to other scientists via stand-alone software and web services; and applying such techniques to tackle biological questions of interest. We participate in comparative genomic studies, both independently and in collaboration with others; this typically involves the analysis of next-generation sequencing (NGS) data. This vast source of new data is providing great gains in understanding genomes, and brings with it many new challenges.

The Goldman group is developing improved mathematical and statistical methods for analysing DNA and amino acid sequences to study how these sequences evolve.

Future projects and goals

We remain 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 continue to concentrate on linked areas that draw on our expertise in phylogenetics, genomics and NGS. Basic to all our work are the fundamentals of phylogenetic analysis, where we are continuing to devise improved tests for detecting positive selection, investigating the use of non-reversible models of sequence substitutions and developing data analysis methods to detect and represent the discordant evolutionary histories of different genes in an organism's genome. We remain committed to keeping abreast of evolving NGS technologies and exploiting them for new experiments — particularly intriguing is the new possibility of sequencing single cells, opening the way to studies that can trace the development of the different parts of an individual living organism. We will continue to look to medical applications of NGS and phylogenetics as a source of inspiring collaborations, and have new collaborations underway bringing molecular evolutionary methods and whole genome NGS of pathogens into a clinical setting where they may be applicable in 'near real time' to help inform doctors' decisions and treatment choices.

Alvis Sequence Bundle visualisation of the HAD family. The Bundle shows the different sequence groups in different colour. Horizontal dependencies are immediately visible. For example, all Ciona sequences (green) have Met in position 241 and also exclusively have a Tyr residue in position 232 and a Val at position 249. This information is not available from the standard sequence logo (shown above the alignment). Above the Bundle, green shaded markers indicate which sites are most likely responsible for the grouping. Here, in agreement with the original paper (Seifreid *et al.* 2014), site 241 (marked with an orange triangle) is found as being most significant.

Gerard Kleywegt

PhD 1991 University of Utrecht.

Postdoctoral researcher, then independent investigator, University of Uppsala, 1992-2009.
Co-ordinator, then Programme Director of the Swedish Structural Biology Network, 1996-2009.
Research Fellow of the Royal Swedish Academy of Sciences, 2002-2006.
Professor of Structural Molecular Biology, University of Uppsala, 2009.

At EMBL-EBI since 2009.

SELECTED REFERENCES

Berman HM, *et al.* (2014) The Protein Data Bank archive as an open data resource. *J Comput Aided Mol Des* 28, 1009-14

Dutta S, *et al.* (2014) Improving the representation of peptide-like inhibitor and antibiotic molecules in the Protein Data Bank. *Biopolymers* 101, 659-68

Gutmanas A, *et al.* (2014) PDBe: Protein Data Bank in Europe. *Nucleic Acids Res* 42, D285-D291

Patwardhan A, *et al.* (2014) A 3D cellular context for the macromolecular world. *Nat Struct Mol Biol* 21, 841-5

We aim to transform the structural archives into a truly useful resource for biomedical and related disciplines.

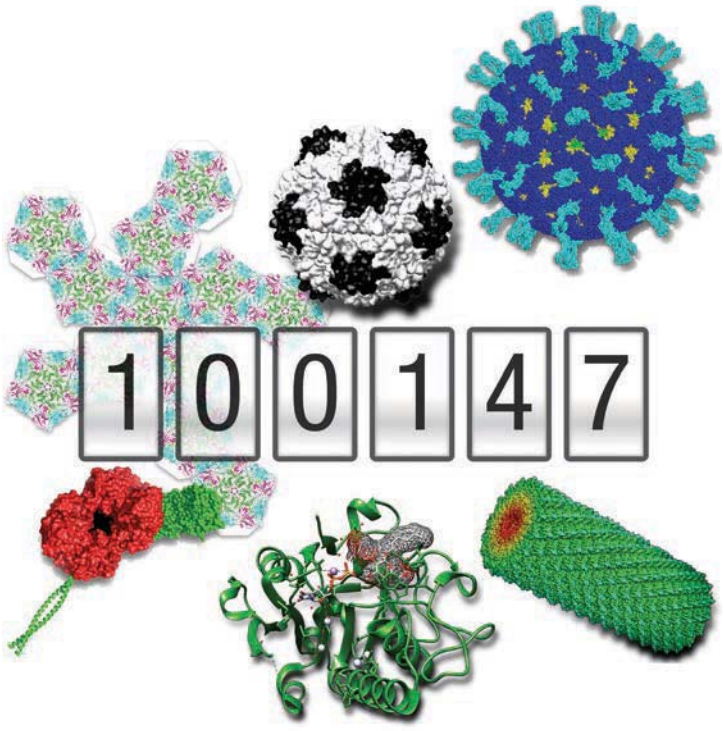
Previous and current research

The Protein Data Bank in Europe (PDBe) is an integrated structural data resource that aims to evolve with the science of structural biology and with the needs of biologists. PDBe handles the deposition and annotation of structural data, provides integrated, high-quality macromolecular and (sub-) cellular structures and related data, and maintains in-house expertise in X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy and 3D cryo-Electron Microscopy (3DEM). We focus on providing advanced services, integration of structural and other information, and providing ligand-related, validation and experimental data. Our mission is to bring structure to biology, and our goal is to make PDBe the logical first stop on any quest for information about 3D molecular and cellular structure.

Future projects and goals

To transform the structural archives into a truly useful resource for biomedical and related disciplines, we will focus on providing advanced services such as PDBePISA, PDBeFold, PDBeMotif and on developing powerful new search and browse facilities. We will devote considerable efforts to the annotation, validation and visualisation of ligand data; integration with other 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 the experimental data support the structural models and inferences.

Over the next few years, we expect the field of cellular structural biology to expand rapidly and have an increasing impact on biology and related fields. Furthermore, hybrid approaches to structure determination will become much more common. We are collaborating actively in this area so we can meet the challenges and embrace the opportunities afforded by exciting developments ahead. These efforts will be supported by a number of substantial new grants from the Wellcome Trust, MRC and BBSRC.



A collage of images signifying events in 2014. "100 147" was the number of entries in the PDB on 14 May 2014, when it exceeded 100 000 for the first time. The other images and events (clockwise from top-left):

- On Virus Appreciation Day, PDBe released a "fold your own virus" activity for children and students, which allows them to construct a 3D paper model of rhinovirus (PDB entry 4rhv) using scissors and glue;
- The football World Cup (which saw EMBL member states grab the gold and bronze medals) inspired this image of ryegrass mottle virus (PDB entry 2izw);
- Since 10 December, even very large structures, such as this rotavirus particle (PDB entry 4v7q), can be represented as single entries using the versatile mmCIF format;
- New detector technology makes it possible to determine 3DEM structures with an unprecedented level of detail, rivalling that attainable with X-ray crystallographic methods. Shown here is the 3.35Å-resolution structure of tobacco mosaic virus (EMDB entry 2842);
- To mark WHO Blood Donor Day, PDBe paid attention to the enzyme "histo-blood group ABO system transferase", which synthesises the ABO blood group antigen;
- A composite image of a poppy, made up of 3 PDB structures (entries 3zc9, 3bas and 4bqm), was used in PDBe social media outreach on Armistice Day.

Computational and evolutionary genomics

John Marioni

PhD 2008 University of Cambridge.

Post-doctoral research at the University of Chicago, 2008-2010.

Group Leader at EMBL-EBI since 2010.

Associate Faculty Member at the Wellcome Trust Sanger Institute since 2014.

SELECTED REFERENCES

Buettner F, *et al.* (2015) Computational analysis of cell-to-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol* 33, 155–60

Pettit JB, *et al.* (2014) Identifying cell types from spatially referenced single-cell expression datasets. *PLoS Comput Biol* 10, 1003824

Schmitt BM, *et al.* (2014) High-resolution mapping of transcriptional dynamics across tissue development reveals a stable mRNA-tRNA interface. *Genome Res* 24,1797-807.

Stegle O, *et al.* (2015) Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet* 16, 133-45

The Marioni group develops statistical tools that exploit data generated using next-generation sequencing to understand the evolution and regulation of gene expression.

Previous and current research

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.

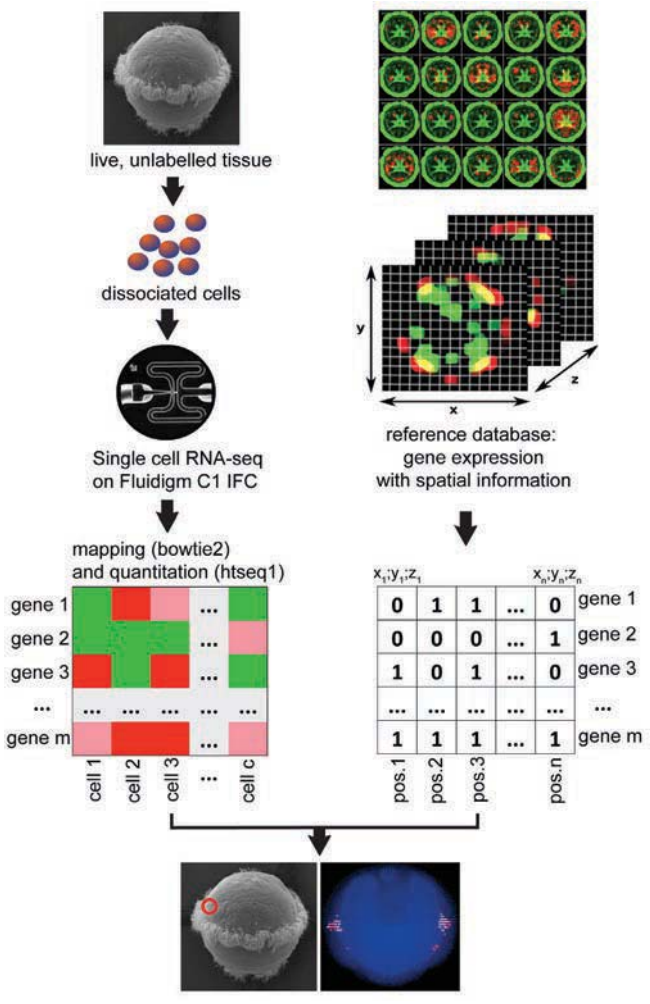
To understand how the divergence of gene expression levels is regulated, we associate changes in expression with a specific regulatory mechanism. In so doing, we gain critical insights into speciation and differences in phenotypes between individuals. We study the evolution of cell types by using gene expression as a definition of the molecular fingerprint of individual cells. Comparing the molecular fingerprint associated with a particular tissue across species allows us to decipher whether specific cell types arise *de novo* during speciation, or whether they have a common evolutionary ancestor. We model spatial variability in gene-expression levels within a tissue or organism to identify heterogeneous patterns of expression within a cell type. This potentially allows us to uncover new cell types, perhaps with novel functions. We use similar approaches to study the extent of heterogeneity present throughout a tumour.

These strands of research are brought together by single-cell sequencing technologies. As founding members of the Sanger Institute–EMBL-EBI Single Cell Genomics Centre and we are closely involved in the centre's efforts to improve data generation and analysis methods, especially single-cell RNA-sequencing, and in using them to answer numerous exciting biological questions. We see the development of appropriate statistical and computational tools as critical to the full exploitation of these data, and will focus on these challenges over the next few years.

Future projects and goals

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.



An approach for integrating single-cell RNA-sequencing with spatial information. We used an existing gene expression atlas (right) to link single-cell RNA-seq data (left) from cells extracted from the developing brain of *P. dumerilii* (top left) with the spatial coordinates. The gene expression atlas was binarized, resulting in a matrix of n positions that each comprise of presence/absence values for m genes. For each sequenced cell c, expression data for m genes was compared to the n positions in the reference matrix and matched with the respective position(s) based on highest similarity. An example of the likely position for one cell is indicated in the bottom panel: by the red circle in the bottom left (ventral view of *P. dumerilii* larva) and red voxels in the apical view at bottom right.

Julio Saez-Rodriguez

PhD 2007 University of Magdeburg.

Postdoctoral work at Harvard Medical School and M.I.T.

At EMBL-EBI since 2010.
Joint appointment at Genome Biology Unit (EMBL-HD).

SELECTED REFERENCES

Costello JC, *et al.* (2014) A community effort to assess and improve drug sensitivity prediction algorithms. *Nat Biotechnol* 32, 1202-12

Egea JA, *et al.* (2014) MEIGO: an open-source software suite based on metaheuristics for global optimization in systems biology and bioinformatics. *BMC Bioinf* 15, 136

Gobbi A, *et al.* (2014) Fast randomization of large genomic datasets while preserving alteration counts. *Bioinformatics* 30, i617-i623

Vaga S, *et al.* (2014) Phosphoproteomic analyses reveal novel cross-modulation mechanisms between two signaling pathways in yeast. *Mol Sys Biol* 10, 767

The Saez-Rodriguez group develops computational methods and tools to analyse signalling networks and mathematical models to better understand how signalling is altered in human disease.

Previous and current research

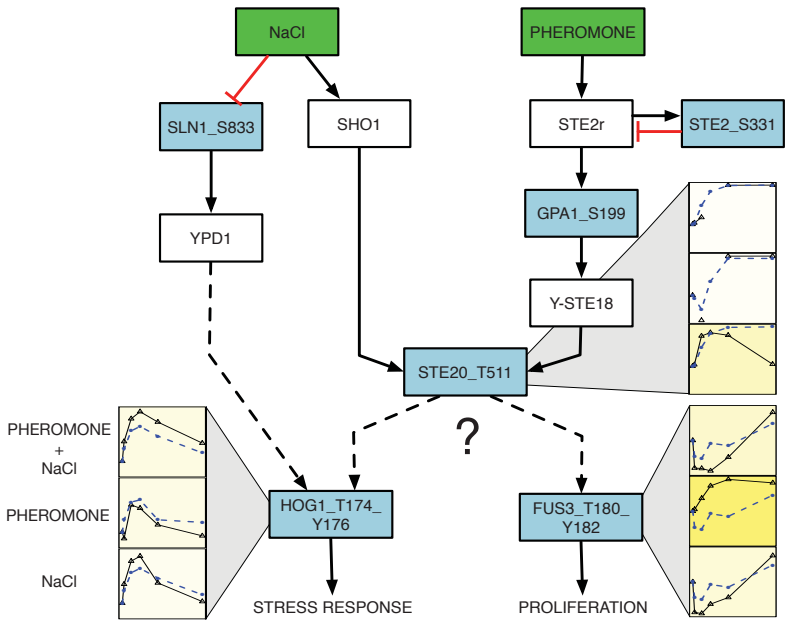
Human cells are equipped with complex signalling networks that allow them to receive and process the information encoded in myriad extracellular stimuli. Understanding how these networks function is a compelling scientific challenge and has practical applications, as alteration in the functioning of cellular networks underlies the development of diseases such as cancer and diabetes. Considerable effort has been devoted to identifying proteins that can be targeted to reverse this deregulation; however, their effect is often unexpected. It is hard to assess their influence on the signalling network as a whole and thus their net effect on the behaviour of the diseased cell. Such a global understanding can only be achieved by a combination of experimental and computational analysis.

Because our research is hypothesis-driven and tailored towards producing mathematical models that integrate diverse data sources, we collaborate closely with experimental groups. Our models integrate a range of data, from genomic to biochemical, with various sources of prior knowledge and with an emphasis on providing both predictive power of new experiments and insights into the functioning of the signalling network. We combine statistical methods with models describing the mechanisms of signal transduction, either as logical or physico-chemical systems. To do this, we develop tools and integrate them with existing resources. We then use these models to better understand how signalling is altered in human disease and predict effective therapeutic targets.

Future projects and goals

In 2015 we will continue to develop methods and tools to understand signal transduction in human cells and its potential to yield insights of medical relevance. Our main focus will be on modelling signalling networks using phospho-proteomics data, and integrating these networks with gene regulation and metabolism. An area of particular interest for us will be single-cell signalling data. We will also develop methods to identify drug targets by integrating genomic and transcriptomic data with information on signalling pathways.

Using these novel methods we will address questions such as: What are the origins of the profound differences in signal transduction between healthy and diseased cells and, in the context of cancer, between normal and transformed cells? What are the differences in signal transduction among cancer types? Can we use these differences to predict disease progression? Do these differences reveal valuable targets for drug development? Can we study the side effects of drugs using these models?



Logic modelling to understand pathway crosstalk in yeast. We systematically evaluated various mechanisms proposed in the literature by converting them into logic models and testing them against time-courses of phosphopeptides. Measurements are shown in black, simulations in blue and disagreement between data and simulation in background color. (Adapted from Vaga *et al.* 2014).

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

Buettner F, *et al.* (2015) Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nature Biotech* 33, 155-60

Cubillos FA, *et al.* (2014) Extensive cis-regulatory variation Robust to environmental perturbation in *Arabidopsis*. *Plant Cell* 26, 4298-310

Fusi N, *et al.* (2014) Warped linear mixed models for the genetic analysis of transformed phenotypes. *Nat Commun* 5, 4890

Smallwood SA, *et al.* (2014) Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods* 11, 817-20

Previous and current research

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.

The Stegle group develops computational and statistical methods to study the genotype-phenotype map on a genome-wide scale.

Future projects and goals

In 2015 we will continue to develop innovative statistical approaches to analyse 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.

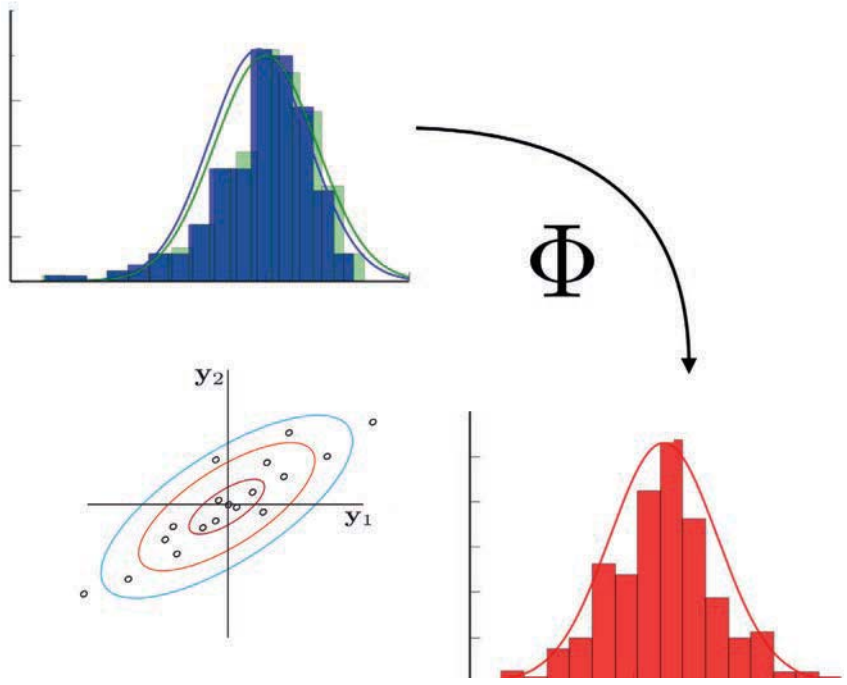


Illustration of the WarpedLMM approach for improved genetic mapping of quantitative traits. This integrative statistical model jointly discovers a non-linear parameterisation of the phenotype data while testing for genetic associations. This joint modeling approach helps to greatly increase power and interpretation of genome-wide association studies.

Small molecule metabolism in biological systems

Christoph Steinbeck

PhD 1995 Rheinische Friedrich-Wilhelm-Universität, Bonn.

Postdoctoral research at Tufts University, Boston, 1996-1997.

Group leader, MPI 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

Rueedi R, *et al.* (2014) Genome-wide association study of metabolic traits reveals novel gene-metabolite-disease links. *PLoS Genet* 10, e1004132

Griss J, *et al.* (2014) The mzTab data exchange format: communicating mass-spectrometry-based proteomics and metabolomics experimental results to a wider audience. *Mol Cell Proteomics* 13, 2765-75

Venkata C, *et al.* (2014) The potential utility of predicted one bond carbon-proton coupling constants in the structure elucidation of small organic molecules by NMR spectroscopy. *PLoS One* 9, e111576

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.

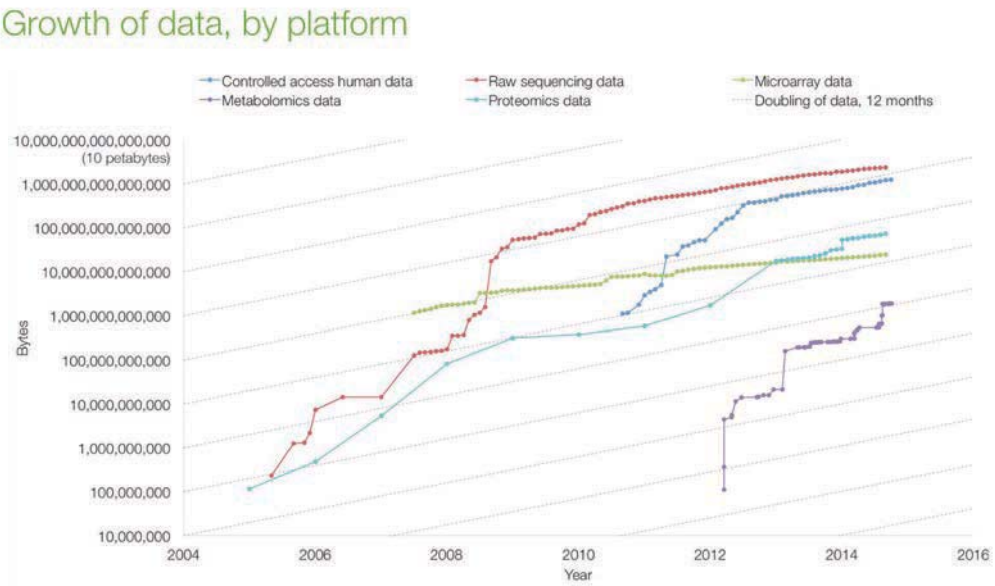
Previous and current research

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.

Future projects and goals

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. Last but not least, our MRC-funded work with the Phenome Centres at Imperial College London will start in 2015, where we will be instrumental to make information on the metabolomes of a large number of human cohorts available to the public.



Metabolomics data is doubling faster than any other data platform at EMBL-EBI.

Gene expression regulation and protein complex assembly

Sarah Teichmann

PhD 2000, University of Cambridge and MRC Laboratory of Molecular Biology.

University College London, 2000-2001. MRC Laboratory of Molecular Biology, 2001-2012.

Fellow and Director of Studies, Trinity College, since 2005.

Department of Physics/Cavendish Laboratory, University of Cambridge, 2013-2016.

Group Leader at EMBL-EBI and Wellcome Trust Sanger Institute since 2013.

SELECTED REFERENCES

Mahata B, *et al.* (2014) Single-cell RNA sequencing reveals T helper cells synthesizing steroids *de novo* to contribute to immune homeostasis. *Cell Rep* 7, 1130-42

Marsh JA, Teichmann SA (2014) Protein flexibility facilitates quaternary structure assembly and evolution. *PLoS Biol* 12, e1001870

Perica T, *et al.* (2014) Evolution of oligomeric state through allosteric pathways that mimic ligand binding. *Science* 346, 1254346

Previous and current research

Our group seeks to elucidate general principles of gene expression and protein complex assembly. We study protein complexes in terms of their 3D structure, structural evolution and the principles underlying protein-complex formation and organisation. We also explore the regulation of gene expression during switches in cell state, and use mouse T-helper cells as a model of cell differentiation. We combine computational and wet-lab approaches at both EMBL-EBI and the Wellcome Trust Sanger Institute.

Combining computational and wet lab approaches, the Teichmann group aims to understand general principles of gene expression and protein complex assembly.

The wealth of genome-scale data available for sequences, structures and interactions provides an unprecedented opportunity to investigate systematically principles of gene and protein interactions. We focus on the evolution and dynamics of regulatory and physical interaction networks, combining computational and mathematical approaches with genome-wide and gene/protein experiments. Our two main areas are transcription factors and the regulation of gene expression; and physical protein–protein interactions and protein complexes.

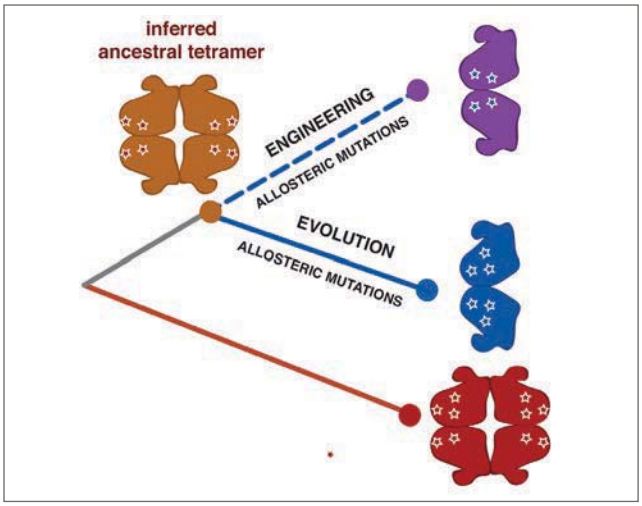
Using the T-helper cell system, we explore the hierarchy and kinetics of molecular events that contribute to changes in gene expression, and whether the kinetics of these interactions graded or switch-like. In 2014 we discovered a steroid-producing T-cell subtype that can inhibit T-cell proliferation and B-cell class switch, and proposed that de novo steroid biosynthesis in immune cells has a role in immune homeostasis.

Our group investigates principles that govern the folding and assembly of protein complexes. Using the informative power of genomic, proteomic and structural data, we capture the critical changes in sequence and structure that distinguish protein-complex formation from the sea of functionally neutral changes. The 3DComplex.org database is a research tool for our work in this area. Our *in silico*, phylogeny-based methods predict critical ancestral mutations involved in changing protein complexes, and we test these using wet-lab biophysical and biochemical techniques.

Future projects and goals

We will continue our projects in structural bioinformatics of protein complex assembly and expand our programme in genomics of gene expression. We will focus on single-cell transcriptomics of the dynamics of immune response to pathogens. This will reveal the full spectrum of CD4+ T-cell types, and lead to new discoveries. Further, we will explore the evolution of an infection time course in terms of expansion and contraction of diverse cell types, and their cell-cell interactions. These *in vivo* experiments, together with in vitro T-cell and ES-cell experiments, will inform us about the cellular circuitry and decision-making in switching from one cell type to another. To gain more insight into cellular switches, we will integrate high-throughput high-content imaging with single-cell RNA sequencing.

We will also pursue methods development in single-cell bioinformatics approaches. Together with the Marioni and Stegle groups at EMBL-EBI, we are keen to find new ways to dissect technical from biological cell-to-cell variation in gene expression, predict regulatory relationships, gene expression modules and cell states from the new flood of single cell RNA-sequencing data.



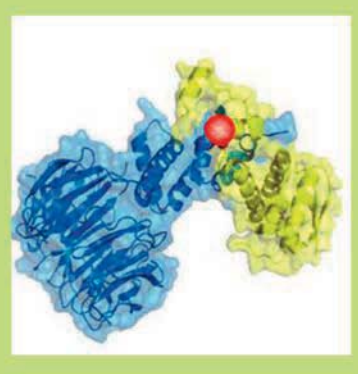
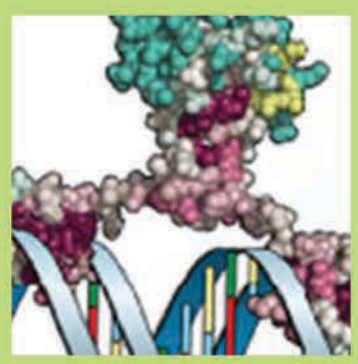
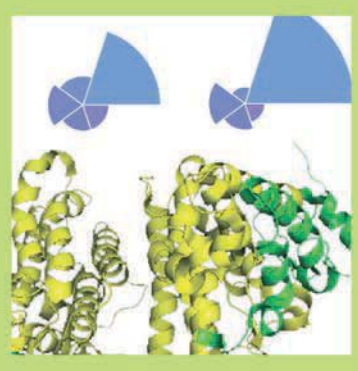
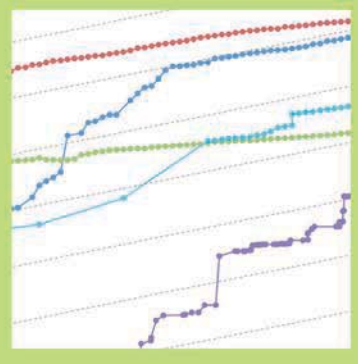
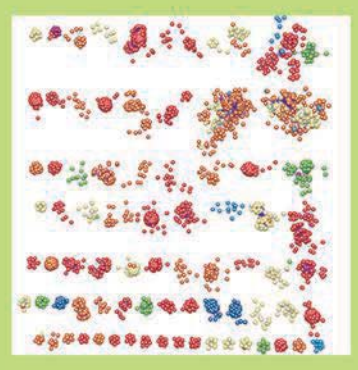
Allosteric mutations can change oligomeric state by employing the same conformational dynamics as the allosteric ligands. PyrR homologues differ by mutations all of which are outside of the tetrameric interface. A subset of these allosteric mutations can be used to engineer a shift in oligomeric state in the ancestral PyrR. Allosteric mutations act by introducing conformational change in a manner analogous to the allosteric ligands.

EMBL-EBI Hinxton

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' 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 databses 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 1999, University of East Anglia.

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. All ENA services are supported with a helpdesk and training programme. 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 technology, content and services that are used in the construction and delivery of many other EBI 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 spectrum of molecular data.

Proteomics Services Team

Henning Hermjakob

MSc Bioinformatics University of Bielefeld, Germany, 1995.

Research assistent at German National Centre for Biotechnology (GBF) 1996.

At EMBL-EBI since 1997.

Our team develops tools and resources for the representation, deposition, distribution and analysis of proteomics and systems biology data. The team follows an open-source, open-data approach and is a major contributor to community standards, in particular the Proteomics Standards Initiative (PSI) of the international Human Proteome Organization (HUPO), and systems biology standards (COMBINE network).

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 EMBL-EBI's proteomics resources. In addition, the 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 team also contributes to the development of data integration technologies using protocols like Distributed Annotation System (DAS) and semantic web technologies, and providing stable identifiers for biomolecular entities through identifiers.org.

The Proteomics Services team develops tools and resources for the representation, deposition, distribution and analysis of proteomics and systems biology data.

Non-vertebrate genomics

Paul Kersey

PhD University of Edinburgh 1992.

Postdoctoral work at University of Edinburgh and MRC Human Genetics Unit, Edinburgh.

At EMBL-EBI since 1999.

The non-vertebrate genomics group develops Ensembl Genomes and related resources.

The Ensembl Genomes team provides services based on the genomes of non-vertebrate species. The development of next-generation sequencing technologies has led to an explosion of reference genome sequences and genome-wide measurements and interpretation. Our team provides portals for bacteria, protists, fungi, plants and invertebrate metazoa, offering access to these data through a set of programmatic and interactive interfaces, exploiting developments originating in the vertebrate-focused Ensembl project. Collectively, the two projects span the taxonomic space.

The falling costs of DNA sequencing have led to the availability of such data even in species studied only by small communities with little informatics infrastructure. 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. Our leading collaborators include VectorBase (Megy *et al.*, 2011), a resource focused on invertebrate vectors of human disease, WormBase (Harris *et al.*, 2013), a resource for nematode biology, and PomBase (McDowall *et al.*, 2014), focused on the fission yeast *Schizosaccharomyces pombe*. In the plant domain, we collaborate closely with Gramene in the US and a range of European groups in the transPLANT project. Our major interests include broad-range comparative genomics and the visualisation and interpretation of genomic variation, which is being increasingly studied in species throughout the taxonomy. We have developed a new portal for plant pathogen data, PhytoPath (launched in early 2012), and are involved in the annotation of the wheat genome, which is large, hexaploid and of potentially massive importance in enabling the development of improved crops to meet humanity's growing need for food.

UniProt development

Maria J. Martin

BSc In Veterinary Medicine.

PhD in Molecular Biology (Bioinformatics), 2003.

At EMBL-EBI since 1996.

Team Leader since 2009.

The team provides bioinformatics infrastructure for the databases and services of the Universal Protein Resource

Our 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 and Gene Ontology Annotation software and database development, and who study novel automatic methods for protein annotation and representation. The team's user experience analyst coordinates the user request gathering process, which informs the design and development of the website. We are also responsible for the maintenance and development of tools for UniProt curation.

Our team works 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.

Literature services

Johanna McEntyre

PhD in plant biology, Manchester Metropolitan University 1990.

Editor, Trends in Biochemical Sciences, Elsevier, Cambridge, UK, 1997.

Staff Scientist, NCBI, National Library of Medicine, NIH, USA, 2009.

Team Leader at EMBL-EBI since 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, now the sole literature database offered by EMBL-EBI, contains more than 30 million abstracts and 3 million full text articles. The abstracts component includes all of PubMed, agricultural abstracts from Agricola and patents from the European Patent Office. About 900,000 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 26 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 such as article citations. 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.

Team Leader since 2009.

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.

The team supports the development of databases and resources enabling the integration of scientific literature and large data sets.

Functional genomics production

Helen Parkinson

PhD in Genetics, 1997.

Research associate in Genetics, University of Leicester, 1997-2000.

At EMBL-EBI since 2000.

The team focusses 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.

Our team comprises the Gene Ontology Editorial Office, Mouse Informatics and the former Functional Genomics Production teams. We provide ontology resources such as the Gene Ontology and Experimental Factor Ontology, and sample/phenotype resources such as the BioSamples database, Infrafrontier and the International Mouse Phenotyping Consortium (IMPC), among others. The team works with external collaborators on projects funded by the European Commission, the BBSRC, the Wellcome Trust, the National Institutes of Health and the National Science Foundation. Examples include PhenoImageShare, which addresses the indexing of image annotation in the context of genomic data, so that images are accessible and queryable with biomolecular datasets. Another of our projects, Embryonic Phenotyping, captures and integrates mouse embryonic images with genomic and phenotypic data. DIACHRON addresses both the use of RDF technology to represent scientific data and strategies for exploring data life cycles in the interests of enabling data preservation. The inception of these projects expands our efforts in RDF generation and application building, ontology delivery and provision of community access to data.

In addition to microRNA data analysis, our research activities have focused on analyses (in R) of RNA-seq data in collaboration with the Geuvadis consortium and genotype imputation for individuals from the 1000 Genomes Project. Semantic web projects include development of an RDF representation of Gene Expression data from the Gene Expression Atlas and ArrayExpress.

Variation team

Justin Paschall

MA 2008, Washington University Saint Louis.

Team Leader at EMBL-EBI since 2012

The Variation team develops and maintains database archive germline and somantic resources for genetic variation in all species.

Variation data is the primary analysis product of the sequencing, alignment and variant-calling pipeline to studies of population genetics, genotype-to-phenotype association and functional analysis linking the genome to molecular pathways. The European Variation Archive (EVA), the global reference catalogue of genetic variation, provides a basis for interpreting each new genome and variant observed in research and clinical studies. It provides a primary archive service for genetic variation data and builds on EMBL-EBI's sequence-level archives, supporting value-added analysis and visualisation resources. Together with international partners, the EVA provides a stable, accessioned database that catalogues and provides access to genetic variation in all species. This is a powerful tool for researchers working in clinical, agricultural, biotechnology and ecology research.

Human genetic data presents particular challenges in terms of protecting participant privacy when individually unique genomes are archived for scientific research, often requiring controlled-access approval systems to ensure compliance with data access policies. The European Genome-phenome Archive (EGA) supports secure controlled-access data management for human genomes and variation data, providing a standard mechanism to provide access to data to a wide set of research users in a secure manner.

Functional genomics software development

Ugis Sarkans

PhD in Computer Science 1998, University of Latvia.

Postdoctoral research at the University of Wales, Aberystwyth, 2000.

At EMBL-EBI since 2000.

Our team develops software for ArrayExpress, a core EMBL-EBI resource, and the BioStudies database, a resource for biological datasets that do not have a dedicated home within EMBL-EBI services. We also contribute to the development of EMBL-EBI's BioSamples database, which centralises biological sample data.

Together with the Expression Atlas team we build and maintain data management tools, user interfaces, programmatic interfaces, and annotation and data submission systems for functional genomics resources. Our team participates in several European 'multi-omics' and medical informatics projects: provide data management solutions for the EU-AIMS project on autism spectrum disorder research, integrate our "R cloud" scientific computation infrastructure into medical data exploration systems, and work towards a generic data security infrastructure in the BioMedBridges project, which facilitates data sharing across the life sciences. We believe that being close to large consortia that generate different types of high-throughput data places us in a better position to fulfill our main objective of developing ArrayExpress and BioStudies infrastructures.

The team develops software for Array-Express and the BioStudies database.

Protein Data Bank in Europe: content and integration

Sameer Valenkar

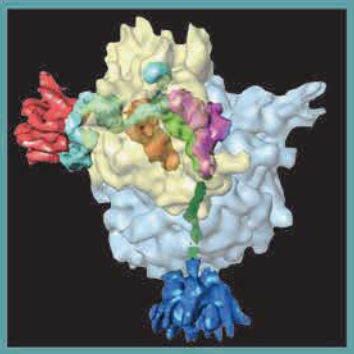
PhD 1997, Indian Institute of Science. Postdoctoral researcher, Oxford University, UK, 1997-2000.

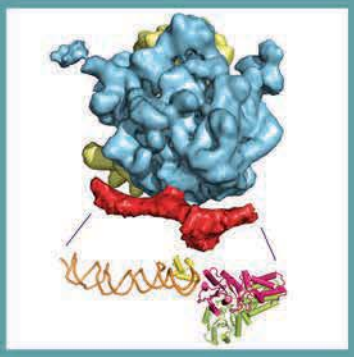
At EMBL-EBI since 2000.

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.

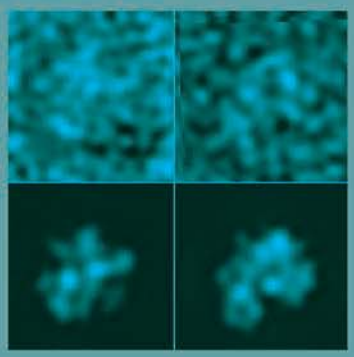
Our goal is to ensure that PDBe truly serves the needs of the biomedical community. As part of that effort, we are constantly improving the web interface for existing tools and services and designing new tools to make structural data available to all. In the context of the SIFTS project, we integrate 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.

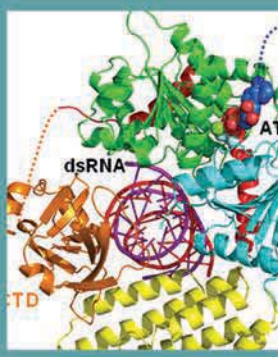
The team accept and annotate world-wide depositions of biomacromolecular structures.

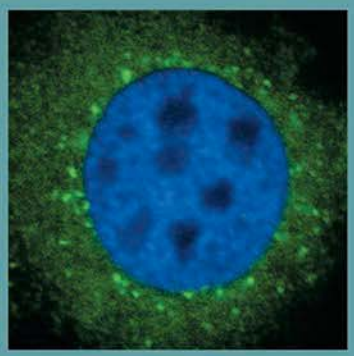




1974 - 2014 Research Highlights





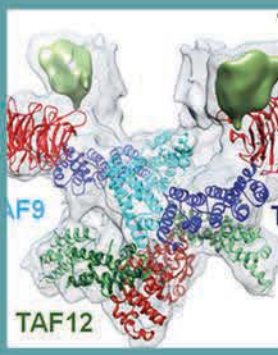


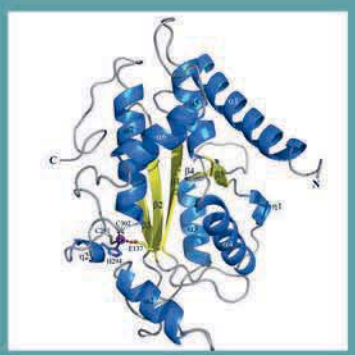
Proteins that load amino acids – protein building blocks – onto tRNA come in at least two types, implying that their evolution was more complex than previously thought.
Cusack S, *et al.* (1990) A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 Å. *Nature* 347, 249-55

First 3-dimensional structure of the proteins EF-Tu and EF-Ts, involved in the elongation of protein chains by ribosomes, bound to each other showed how EF-Ts acts as a reset button to allow EF-Tu to transport amino acids to the ribosome that will assemble them into proteins.
Kawashima T, *et al.* (1996) The structure of the *Escherichia coli* EF-Tu· EF-Ts complex at 2.5 Å resolution. *Nature* 379, 511-8

New microdiffractometer for handling protein microcrystals in crystallography experiments, designed for maximum precision and ease of use.
Perrakis A, *et al.* (1999) Protein microcrystals and the design of a microdiffractometer: current experience and plans at EMBL and ESRF/ID13. *Acta Crystallographica. Section D: Biological Crystallography* 55, 1765-70

First 3-dimensional structure of the transport protein importin-ß bound to part of importin-α showed that the former wraps around the latter, implying that, after entering the cell's nucleus, importin-ß has to undergo a dramatic change to uncoil and release the cargo it has carried inside.
Cingolani G *et al.* (1999) Structure of importin-β bound to the IBB domain of importin-α. *Nature* 399, 221-9





EMBL Grenoble

Structural Biology

EMBL Grenoble is a key player in the Partnership for Structural Biology (PSB) which integrates the activities of several Grenoble institutes involved with structural biology and thus is able to provide a uniquely comprehensive range of platforms for both inhouse research and external users.

A cornerstone of the PSB is the close interaction of EMBL Grenoble with European Synchrotron Radiation Facility (ESRF), which involves collaboration on building and operating beamlines for macromolecular crystallography and small-angle scattering, developing instrumentation and techniques, and providing expert help to visitors. The highly automated ESRF crystallography beamlines are equipped with EMBL-designed high-precision micro-diffractometers and frozen crystal sample changers. A new X-ray small-angle scattering instrument built by ESRF and EMBL is now operational with a custom designed small-volume automatic sample changer.

High-throughput methods have also been introduced in other steps of the structure determination process. These include a very successful robotic system for nanovolume crystallisation and a novel high throughput screening method, ESPRIT, which enables soluble protein domains to be identified in otherwise badly expressed or insoluble proteins. More recently, a Eukaryotic Expression Facility (EEF) has been established specialising in the expression of multi-subunit complexes in insect cells, building on and further developing the MultiBac method. All these platforms are now available to external users under the EU funded BioStruct-X projects.

A strong tradition at the outstation is 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. Projects involving protein-RNA interactions also include cryo-EM studies of the signal recognition particle and holo-translocon with the ribosome and other proteins and complexes involved in RNA processing, transport and degradation, such as the nonsense-mediated decay (NMD) pathway.

Other 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 and chromatin-modification complexes as well as elucidation of the mechanisms by which piRNAs (small non-coding RNAs) protect the genome. 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 viral RNA in infected cells and activate interferon production. The first crystal structures of domains of the influenza virus polymerase, determined in collaboration with other members of the CNRS-Grenoble University-EMBL Unit of Virus Host Cell Interactions (UVHCI), which is situated next to the outstation, have led to an important programme of anti-influenza drug design in collaboration with pharmaceutical companies.

Scientists at EMBL Grenoble have access to a wide range of techniques, including molecular biology and biophysical techniques, cryo-electron microscopy, isotope labelling, NMR, neutron scattering, X-ray crystallography and small angle scattering. A confocal microscope with facilities for cross-correlation spectroscopy is available for the study of complex formation in cells, as well as a top-end Polara electron microscope with direct detector and cryo-tomography capability.

Stephen Cusack
Head of EMBL Grenoble

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.

Director of EMBL-Grenoble University-CNRS Unit for Virus Host Cell Interactions (UVHCI) since 2007.

ERC Advanced Investigator.

SELECTED REFERENCES

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, 316-66

Palencia A, *et al.* (2012) Structural dynamics of the aminoacylation and proofreading functional cycle of bacterial leucyl-tRNA synthetase. *Nat. Struct. Mol. Biol.* 19, 677-84

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 metabolism, translation, RNA virus replication and innate immunity.

Previous and current research

Aminoacyl-tRNA synthetases play an essential role in protein synthesis by charging specifically their cognate tRNA(s) with the correct amino acid. Current structural studies aim to understand their mechanism and substrate specificity. We have focused on leucyl-tRNA synthetase: it contains a large 'editing' domain that hydrolyses mischarged amino acids, a proof reading activity that maintains translational fidelity. Our group structurally characterised the large conformational changes required to switch from the aminoacylation to the editing configurations. This led to the understanding of the mechanism of action of a new anti-fungal compound, and to the design of new antibiotics that target multiresistant gram negative bacteria and tuberculosis.

The nuclear cap-binding complex (CBC) binds to the m7Gppp cap at the 5' end of RNA Pol II transcripts and mediates interaction with nuclear RNA processing machineries for splicing, poly-adenylation and transport. We determined the structure of the human CBC, a 90 KDa heterodimeric protein. We also studied several other proteins involved in cap-dependent processes, such as the three conserved Upf proteins (Upf1, Upf2 and Upf3) that are essential for a RNA quality-control check called nonsense-mediated decay (NMD).

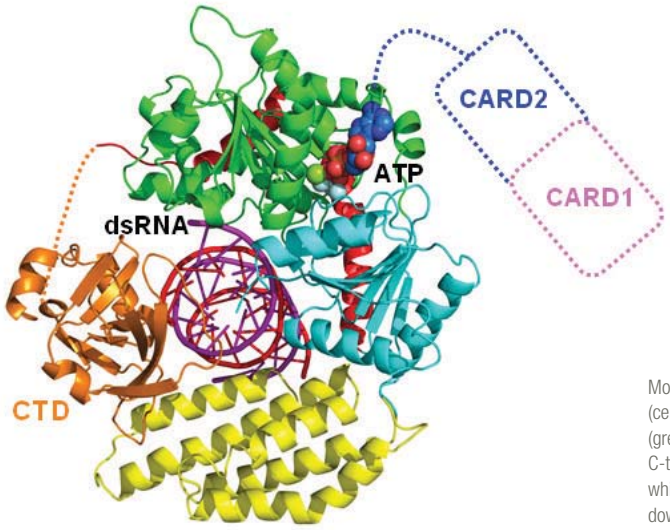
Future projects and goals

Ongoing projects related to RNA metabolism include continued studies on PHAX and ARS2, both of which bind CBC and are linked to the metabolism of small RNAs.

In 2011 we published the first structure-based mechanism of activation of RIG-I, a receptor involved in the innate immunity, and are continuing to study this signalling pathway.

The major focus is structure determination of the influenza virus RNA-dependent RNA polymerase, the viral replication machine. We have determined the structure of four distinct domains from the polymerase, including the two key domains involved in the 'cap-snatching' process of viral mRNA transcription: the cap-binding site in PB2 and the endonuclease in PA. These results give some insight into the polymerase mutations required to adapt an avian virus to be able infect humans and permit structure-based antiviral drug design. To pursue this we have co-founded a Vienna-based company called SAVIRA and the project is now in the hands of Roche in Basel. This work is now being extended to the polymerases of other segmented RNA viruses such as bunyaviruses, which also perform cap-snatching. In collaboration with the Ellenberg and Briggs groups we studied the assembly and trafficking of the influenza polymerase and RNPs in living, infected cells.

In 2014 we published the first crystal structures of the complete influenza polymerase and proposed a mechanism of how cap-snatching is performed. We now try to obtain structural snapshots of the polymerase in action.



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.

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.

Previous and current research

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.

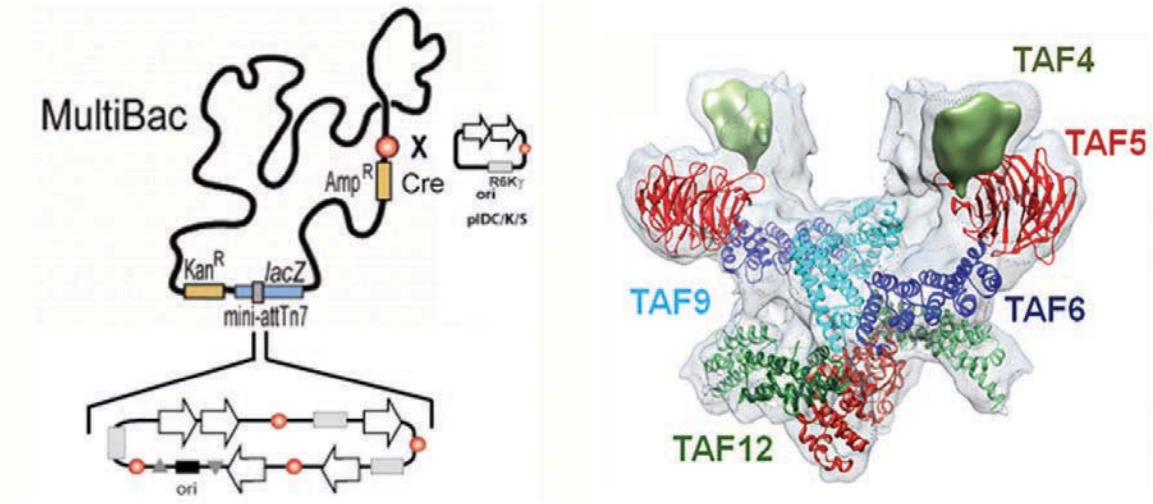
We collaborate with groups from academia and industry to further automate labour-intensive steps in the multiprotein complex structure determination process, and have harnessed homologous and site-specific recombination methods for assembling multigene expression plasmids. We have developped ACEMBL, a proprietary automated suite for multigene recombination on our TECAN Evoll 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 projects and goals

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

Round A *et al.* (2015) BioSAXS Sample Changer: a robotic sample changer for rapid and reliable high-throughput X-ray solution scattering experiments. *Acta Crystallogr. D Biol. Crystallogr.* 71, 67-75

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. A*cta Crystallogr D Biol Crystallogr* 65, 1237-46

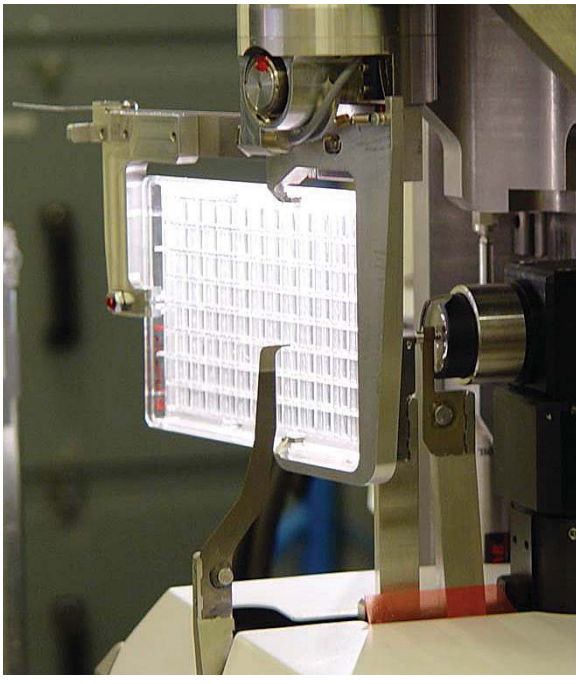
The Cipriani team develops instruments and methods for X-ray scattering experiments and works with industry to make them available to scientists worldwide.

Previous and current research

The core activity of our team is to develop instruments and methods for X-ray scattering experiments. CrystalDirect™, the system jointly conceived with the Márquez team is now available to users of the High Throughput Crystallisation lab: crystals grown in CrystalDirect (CD) plates can be identified and marked remotely using a web service for automatic harvesting. Optimal crystal freezing is ensured by removing the excess solution that surrounds the crystals and, if necessary, by adding cryo-protectant directly in the CD plate, before harvesting. With a harvesting time of 90 seconds CD potentially allows the mounting of several hundred crystals per day. The machine is particularly suitable for harvesting micro-crystals, fragile crystals, and to extract thin needles from clusters. The intrinsic low background of the CD plates makes them ideal for *in situ* X-ray data collection, such as what is offered on request at the EMBL/ESRF BM14 beamline. The first *in situ* rastered data collections were made at the EMBL@PETRA-III P14 beamline on batches of 5-20 micron sized crystals of model proteins. Several hundred crystals in the same crystallisation drop were shot, producing thousands of diffraction images, from which high resolution structures were obtained. “MiniSpine”, a new compact and precise sample holder for frozen crystallography, is being developed within the framework of the BioStruct-X programme. Evaluation kits will be proposed early next year and test sites established with the support of industrial partners. All these developments include contributions from the ESRF Structural Biology group, Schneider group as well as the McCarthy and Márquez teams. Supported by EMBLEM, most of our instruments are available to the scientific community worldwide.

Future projects and goals

A second CD harvester with automated crystal storage will soon be built to satisfy the demand of projects involving ligands screening. The current harvester will be moved at a beamline to pipeline crystal harvesting and data collection. We will equip the ESRF ID30B beamline with a MD2S diffractometer and new 6 axis robotic sample changing solution to offer the ESRF user community access to serial *in situ* and frozen data collection in a flexible robotized environment. The future ESRF Phase-II MX beamlines will require only a few tens of milliseconds to extract the information contained in a crystal, making sample delivery a central question to beamline efficiency once again – one we should address when considering the future operation of the European XFEL facility in Hamburg.



Micro-crystals in situ data collection – CD plate mounted on a MD3 diffractometer, at the EMBL@PETRA3 P14 beamline.

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

Marcia M, Pyle AM (2014) Principles of ion recognition in RNA: insights from the group II intron structures. *RNA* 20, 516-27

Marcia M, Pyle AM (2012) Visualizing group II intron catalysis through the stages of splicing. *Cell* 151, 497-507

Marcia M *et al.* (2010) A new structure-based classification of sulfide:quinone oxidoreductases. *Proteins* 78, 1073-83

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

Previous and current research

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.

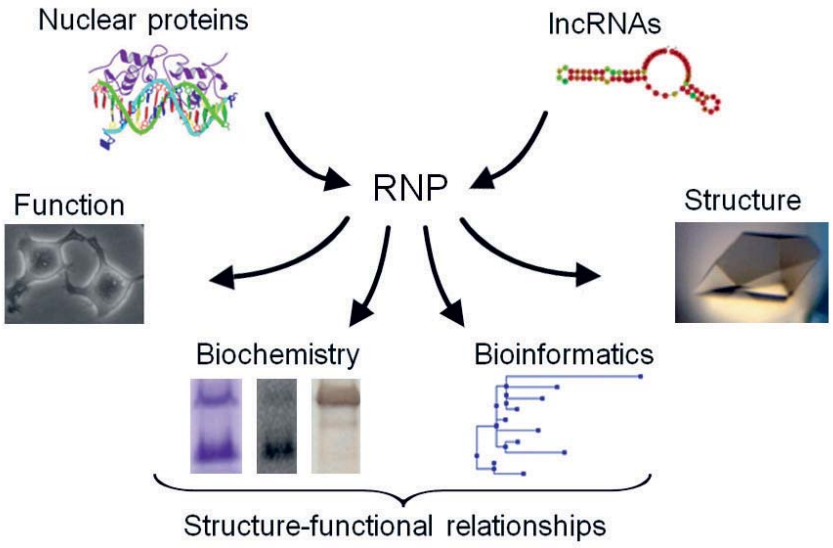
Our research expertise stems from my previous work on the structure and function of various classes of macromolecules, ranging from membrane proteins to large RNA enzymes, involved in fundamental metabolic pathways such as electron transport, signalling and splicing. I previously determined the crystal 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 projects and goals

- 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

Gueneau E, *et al.* (2013) Structure of the MutLαC-terminal domain reveals how Mlh1 recruits the exonuclease I and contributes to Pms1 endonuclease site. *Nat. Struct. Mol. Biol.* 20, 461-8

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

Dupeux F, *et al.* (2011) A thermal stability assay can help to estimate the crystallization likelihood of biological samples. *Acta Crystallogr. D Biol. Crystallogr.* 67, 915-9

Santiago J, *et al.* (2009) The abscisic acid receptor PYR1 in complex with abscisic acid. *Nature* 462, 665-8

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.

Previous and current research

The HTX lab 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 P-CUBE 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 Crystal Direct™, 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. Crystal Direct 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 (see figure).

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 projects and goals

In collaboration with other EMBL and ESRF groups we will work to develop a new generation of highly automated pipelines that integrate crystallisation, crystal harvesting and processing, and data collection in a single, automated workflow. This involves developing new CRIMS interfaces to allow remote access to the CrystalDirect harvester, as well as integrating these resources with the newly developed MASSIF beamlines at the ESRF. We will also explore the potential of the CrystalDirect system to support and facilitate serial crystallography experiments.

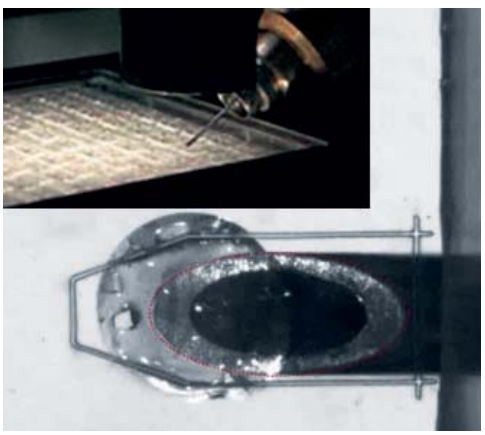


Figure 1: The CrystalDirect technique could benefit challenging structural biology projects, such as studies of membrane proteins or multi-protein complexes.

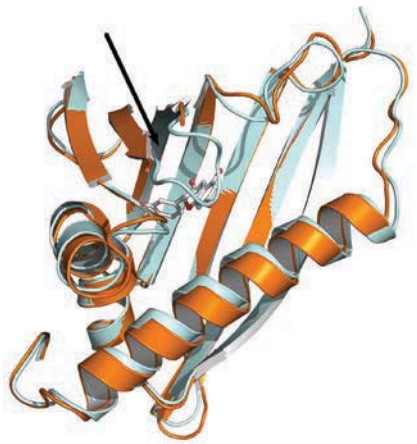


Figure 2: Detail of the Fra a 1 and Fra a 3 proteins in apo and ligand bound forms respectively.

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

De Maria Antolinos, *et al.* (2015) ISPyB for BioSAXS, the gateway to user autonomy in solution scattering experiments. *Acta Crystallogr. D Biol. Crystallogr.* 71, 76-85

Cora E *et al.* (2014) The MID-PIWI module of Piwi proteins specifies nucleotide- and strand-biases of piRNAs. *RNA* 20, 773-781

van Rooyen *et al.* (2014) Assembly of the novel five-component apicomplexan multi-aminoacyl-tRNA synthetase complex is driven by the hybrid scaffold protein Tg-p43. *PLoS One* 20, e89487

Sayou C, *et al.* (2014) A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. *Science* 343, 645-8

Previous and current research

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 found on the EMBL Grenoble services webpage. In addition to being responsible for a number of beamlines, we collaborate with the Cipriani group on the development of a new type of sample changer 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 methods as well as contributing to the development and expansion of ISPyB, a database for MX and bioSAXS experiments.

In addition we study proteins involved in neuronal development, particularly the Slit-Robo signalling complex, proteins involved in piRNA biogenesis (figure 2), as well as plant transcription factors and their regulators. Meanwhile, the BM14 group is actively involved in the structural studies of proteins involved in the *Toxoplasma gondii* epigenetic machinery.

The McCarthy team works on the design, construction and operation of MX and BioSAXS beamlines and studies proteins involved in neuron development.

Future projects and goals

This year will be particularly exciting with the completion of the UPBL10 project with our ESRF colleagues and full user operation of the new suite of MX-beamlines on ID30A (MASSIF) and ID30B. This ambitious project is part of the ESRF upgrade program and will ensure that European users will have continued access to state-of-the-art structural biology beamlines for the next five years. We will continue to develop and improve the available facilities in order to optimise sample handling, and data collection and analyses functionalities. We hope that all our combined efforts will push the boundaries of structural biology currently available to better understand how complex biological systems work at the molecular level.

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 phosphoryl transfer into human kinase signal cascades. The BM14 group will over-express several *Toxoplasma* proteins in *E. Coli* to be structurally and biochemically characterised using the PSB facilities. Lastly, our work on piRNA biogenesis and plant transcription factors will be continued.

Figure 1: The new MASSIF-1 experimental end station at the ESRF.

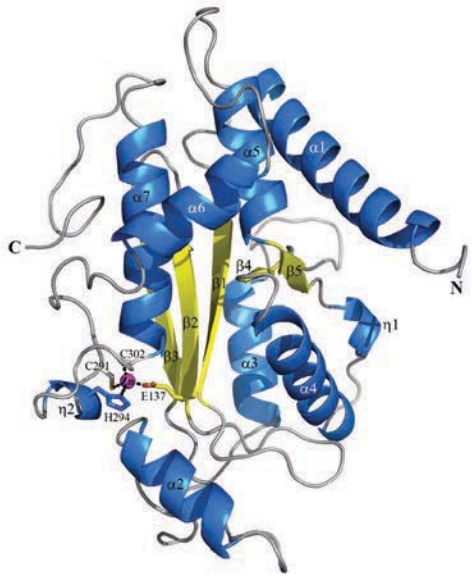


Figure 2: Crystal structure of the Maelstrom domain from Bombyx mori.

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

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

Panne D, *et al.* (2007) An atomic model of the interferon-beta enhanceosome. *Cell* 129, 1111-23

The Panne group looks to understand important signalling processing pathways in the cell, which could help in the discovery of anti-viral drugs.

Previous and current research

Cellular control logic is ultimately embedded in the molecular architecture of the molecular machines that make up the living cell. Many molecular machines, especially complexes involved in cellular signalling, are transient, with a variety of states and a succession of structures. For example, transcription factors such as NFkB, IRF3, IRF7 and ATF2/c–Jun bind as complexes to enhancers ('enhanceosomes') in a combinatorial and dynamic fashion to regulate expression of genes (figure 1). Enhanceosome assembly is mediated by a set of IKK kinases such as TBK1 that regulate NFkB and IRF3/IRF7 activation. These signalling components interact with each other and with other molecules in highly structured but complex ways.

Understanding such transient and dynamic complexes of the 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: What is the architecture of signalling complexes that direct innate immune responses? 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?

Future projects and goals

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 2). 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 CBP/p300 read out histone modification patterns, and how chromatin recognition and modification are coupled. Answers to some of these questions are likely to contribute to our understanding of epigenetic gene regulation and dysregulation in disease. This is not only of fundamental importance for cellular signalling, but also opens up opportunities for pharmacological targeting.

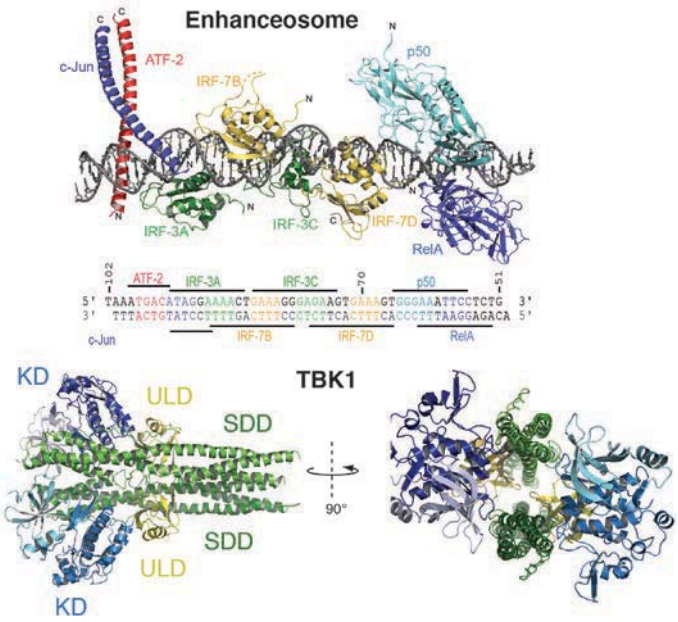


Figure 1: We employ a number of different resolution techniques to visualise the architecture of cellular components.

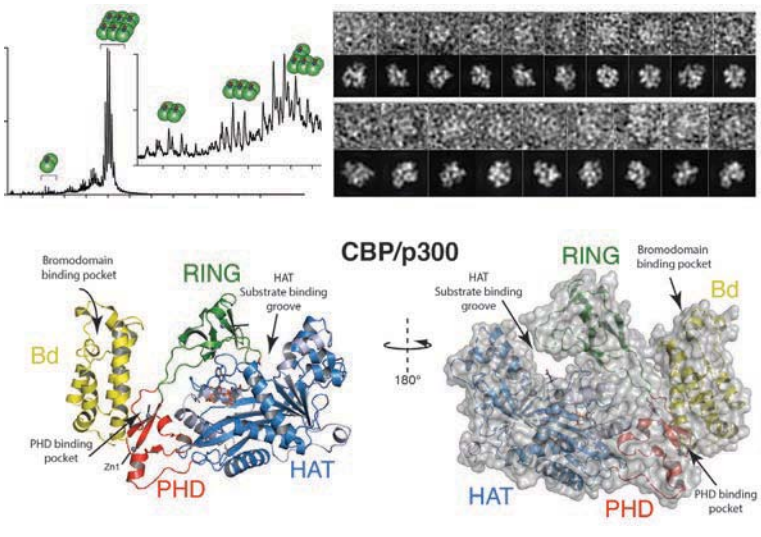


Figure 2: Atomic model of the INF-β enhanceosome.

Regulation of gene expression by non-coding RNAs

Ramesh Pillai

PhD 2002, University of Bern.

Postdoctoral research at the Friedrich Miescher Institute, Basel.

Group leader at EMBL Grenoble since 2006.

ERC Investigator.

SELECTED REFERENCES

Xiol J, *et al.* (2014) RNA clamping by Vasa assembles a piRNA amplifier complex on transposon transcripts. *Cell* 157, 1698-711

Cora E, *et al.* (2014) The MID-PIWI module of Piwi proteins specifies nucleotide- and strand-biases of piRNAs. *RNA* 20, 773-81

Xiol J, *et al.* (2012) A role for Fkbp6 and the chaperone machinery in piRNA amplification and transposon silencing. *Mol. Cell* 47, 970-9

Reuter M, *et al.* (2011) Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature* 480, 264-7

Previous and current research

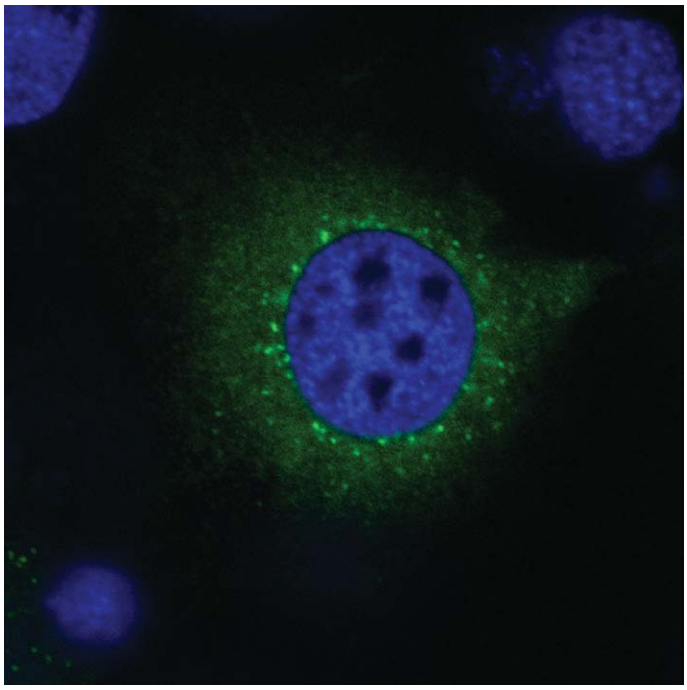
Past invasion events from mobile genetic elements have left eukaryotic genomes littered with repeats and other transposon sequences. Much of these are inactive fossils, but some could still get activated and cause genome instability, despite being silenced in the germlines. This silencing is mediated, in animal germ cells, by a specialised class of ~30 nt small non-coding RNAs called piwi-interacting RNAs (piRNAs). They constitute an epigenetic component of the genome defense mechanism. In mammals, they are believed to recruit DNA methyltransferases to transposon sequences. In *Drosophila*, maternally produced piRNAs are deposited in the egg and they contribute to protection from new transposons brought in by the paternal genome.

Our lab is interested in understanding the molecular mechanisms involved in piRNA biogenesis and function, both in mammals and in *Drosophila*. A striking feature of piRNAs is their clustered genomic origins: it is believed that a long single-stranded transcript arising from a cluster is processed into thousands of piRNAs. This mechanism and the factors involved are still unknown, so we have taken a biochemical approach to identify them in mice. We identified the Tudor domain-containing protein 1 (Tdrd1), which recognises symmetrical dimethyl arginine modification marks on Piwi proteins, and the putative helicase Mov10l, which is an essential piRNA biogenesis factor. In all these studies, we have used a variety of techniques such as protein biochemistry, cellular imaging, small RNA bioinformatics, and mouse mutants. We also collaborate with structural biologists to obtain atomic resolution images of the identified pathway components: we recently unveiled the structure of the recognition pattern between the 2'-O-methyl mark on piRNAs and the PAZ domain of a Piwi protein.

The Pillai group seeks to understand the molecular mechanisms involved in piRNA biogenesis and its function in protecting the genome from instability.

Future projects and goals

We will continue to analyse additional factors identified in our complex purifications. We aim to understand what features define genomic regions as piRNA clusters, and whether there is a link between transcription from the clusters and piRNA biogenesis. Via live cell imaging techniques we want to study assemblies of small RNPs *in vivo* and define the contribution of the individual constituents of the complex to this process. We will intensify the collaborations on structural biology of Piwi complexes, adding another dimension to our understanding of germline small RNAs.



Localisation of a tagged insect Piwi protein to perinuclear cytoplasmic granules in insect cell cultures. These are putative piRNA biogenesis sites, similar to the nuage in germ cells.

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.* (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

Estrozi LF, *et al.* (2011) Cryo-EM structure of the *E. coli* translating ribosome in complex with SRP and its receptor. *Nat. Struct. Mol. Biol.* 18, 88-90

Schaffitzel C *et al.* (2006) Structure of the *E. coli* signal recognition particle bound to a translating ribosome. *Nature* 444, 503-6

The Schaffitzel team combines molecular biology, biochemistry and cryo-electron microscopy to study large macromolecular complexes in protein targeting, secretion and membrane protein integration.

Previous and current research

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 mechanisms of these megadalton-size complexes. Using cryo-EM, we can study the multi-component translation machinery at close to physiological conditions. By using state-of-the-art electron microscopes and by image processing of large data sets, EM structures of prokaryotic and eukaryotic ribosomes have been obtained at subnanomolar 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 the quantity and quality required for interaction assays, mass spectrometry and single-particle cryo-EM. We have established bacterial and eukaryotic cell-free translation systems for the *in vitro* generation of ribosomes that display homogenous nascent polypeptide chains or have 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 SecYEG (figure 1), of the translating ribosome-signal recognition particle (SRP) complex, and of the ribosome in complex with SRP and SRP receptor (figure 2). The data from intermediate resolution structures derived from cryo-EM, in conjunction with high-resolution structures of the ribosome and of the isolated factors, were combined in a hybrid approach to generate quasi-atomic models of the ribosomal complexes involved. The structural data, supported by biochemical data, provides important and detailed snapshots of the mechanisms underlying these cellular processes, ensuring correct folding, targeting and translocation of nascent proteins.

Future projects and goals

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) and single-particle cryo-electron microscopy. In collaboration with the Cusack, Hentze, and Kulozik (MMPU) groups, we study mammalian ribosomal complexes involved in nonsense-mediated mRNA decay. We produce the eukaryotic factors involved using advanced recombinant eukaryotic technologies in collaboration with the Berger group. Finally, our team collaborates with a number of groups to solve the structures of large macromolecular complexes in transcription, epigenetics, and cellular signalling.

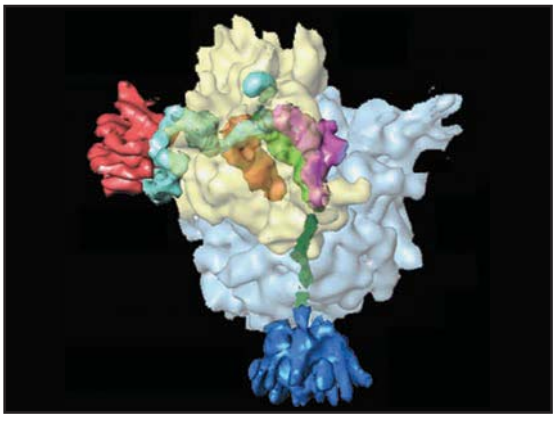


Figure 1: EM reconstruction of the ribosome nascent chain complex and the translocon. The 30S ribosomal subunit is shown in yellow, the 50S subunit in light blue, the A, P and E site tRNAs are in magenta, green and orange respectively. The mRNA is depicted in cyan and the nascent chain in the polypeptide exit tunnel in green. The active translocon in dark blue is located at the exit of the ribosomal tunnel. A second, non-translocating translocon (red) is bound to the mRNA.

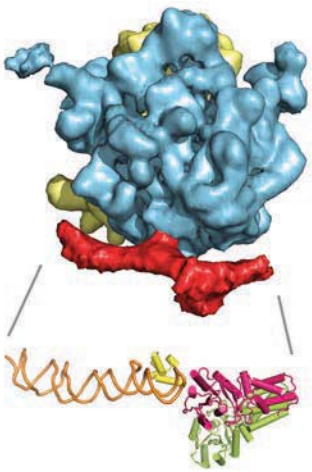
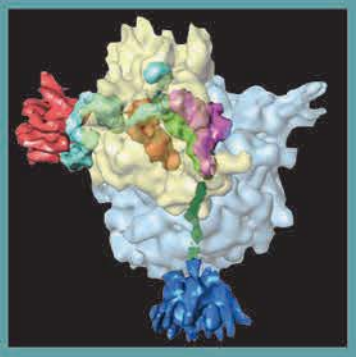
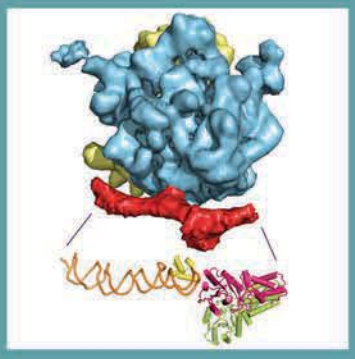
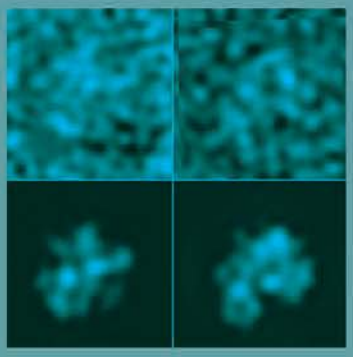


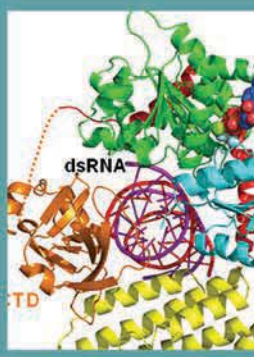
Figure 2: Cryo-EM structure of the ribosome (yellow/blue) bound to the signal recognition particle (SRP) and the SRP receptor (both in red). Below it is an atomic model of SRP (green-yellow/ orange) and its receptor (pink).

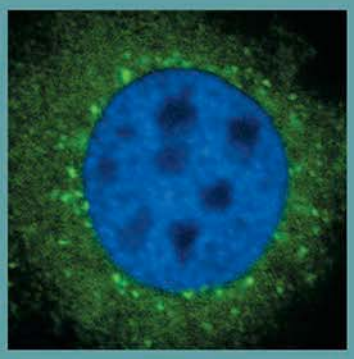

IRF-3A
AGGAAAAC
TCCTTTTG

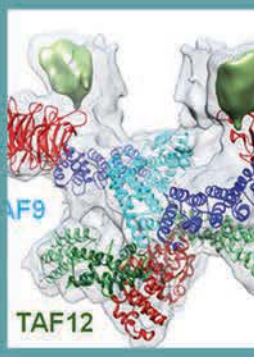


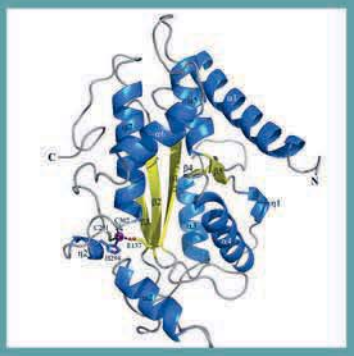


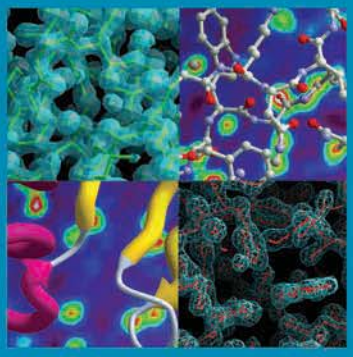


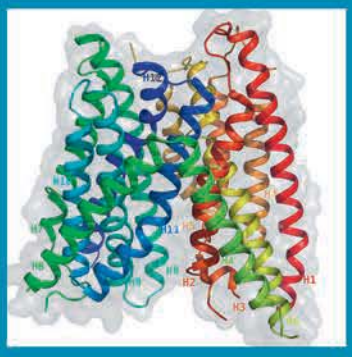

dsRNA
CTD

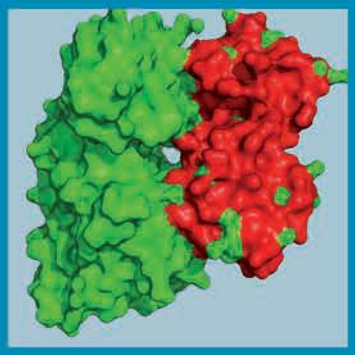



AF9
TAF12

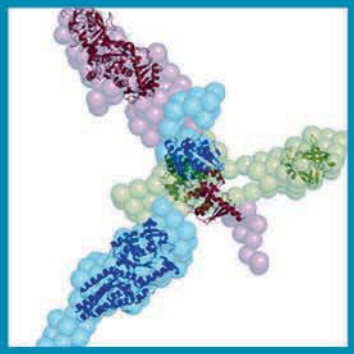


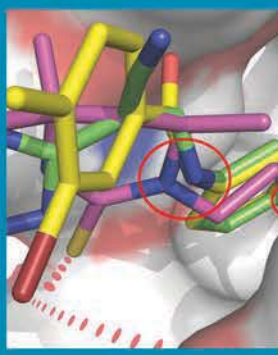






1974 - 2014 Research Highlights



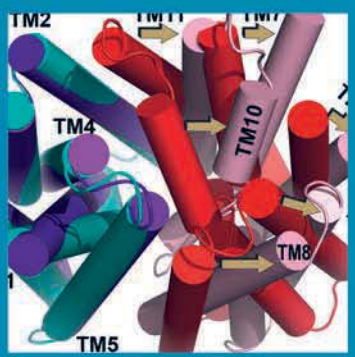


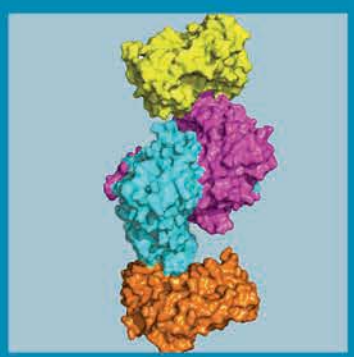
Pioneering use of synchrotron radiation for X-ray crystallography.
Rosenbaum G, & Holmes KC (1971) Synchrotron radiation as a source for X-ray diffraction. *Nature* 230, 434-7

Developed a versatile data processing system for multi-element detectors that enabled made accurate measurements for longer-running X-ray crystallography experiments possible, and which later impacted small-angle scattering (SAXS).
Hendrix J, *et al.* (1982) A wire per wire detector system for high counting rate X-ray experiments. *Nuclear Instruments and Methods in Physics Research* 201, 139-44

Three-dimensional structure showed that Titin – a protein that plays an important role in muscle development and contraction – is activated in a two-step process when muscle fibres are forming in the embryo.
Mayans O, *et al.* (1998) Structural basis for activation of the titin kinase domain during myofibrillogenesis. *Nature* 395, 863-9

New software for generating better models of the 3-dimensional structure of molecules based on crystallography data: faster, more objective and more reliable than previous methods.
Perrakis A, *et al.* (1999) Automated protein model building combined with iterative structure refinement. *Nature Structural & Molecular Biology* 6, 458-63





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

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.

Group leader at EMBL Heidelberg 1993-1997.

Head of EMBL Hamburg since 1997.

Research Director of the Center for Structural Systems Biology (CSSB) since 2014.

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.

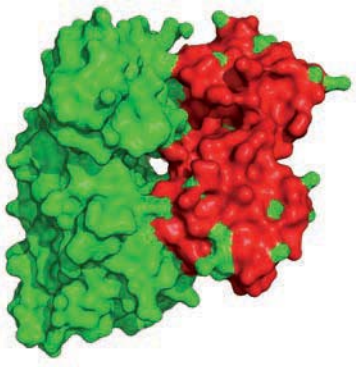


Figure 1: DAPK-CaM complex. Colour codes: DAPK, green; CaM, red. (de Diego *et al.*, 2010)

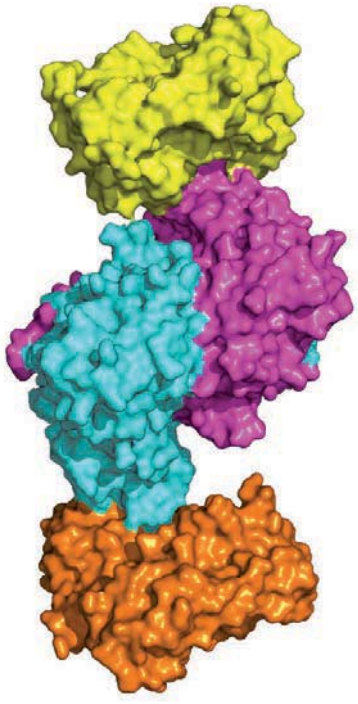


Figure 2: AGT-(Pex5p)2-AGT complex. Colour codes: AGT, yellow, orange; Pex5p, cyan, magenta. (Fodor *et al.*, 2012)

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

Previous and current 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.

The Fiedler team focuses on the selection, customisation and integration of mechanics, control electronics and control software for X-ray based structural biology research.

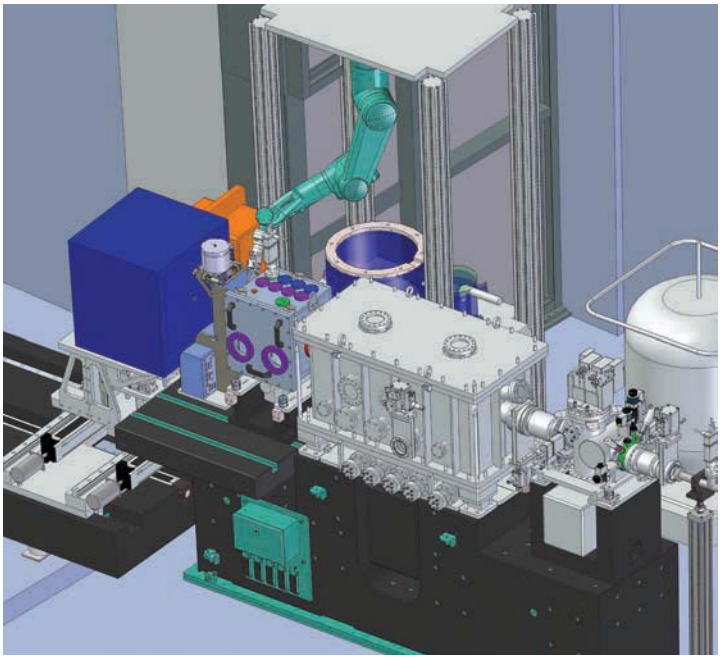
Recent major projects: All three beamlines reached regular user operation. In preparation for this, it was necessary to install and commission adaptive focusing optics at all beamlines, develop a multi-degree-of-freedom heavy-duty detector stage for the large-area pixel detectors at the MX beamlines (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.

Future projects and goals

- 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.



Design of the P14 experimental endstation with sample mounting system MARVIN.

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

Carolan CG & Lamzin VS. (2014) Automated identification of crystallographic ligands using sparse-density representations. *Acta Crystallogr D Biol Crystallogr.* 70, 1844-53

Zvereva M, *et al.* (2013).Towards atomic structure of telomerase complex components in order to create regulators to control growth of cancer cells. *RFBR Journal* 4, 33-7

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 GG, *et al.* (2013) Visual automated macromolecular model building. *Acta Crystallogr. D Biol. Crystallogr.* 69, 635–41

The Lamzin group applies and develops cutting-edge computational methods and experimental approaches for sample quality control, experimentation and data interpretation in macromolecular crystallography.

To fully understand the function of biological systems, accurate structures of their components – DNA, RNA, proteins and macromolecular assemblies – are required. Given the breath-taking opportunities for structural biology arising with the availability of the European X-ray Free Electron Laser (FEL) from 2017, relevant research and development are becoming a focus of the group's activities.

Previous and current research

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 class I hydrophobins (Kallio *et al*, 2011) and fragments of human gelsolin, which may be used in drug delivery to solubilise hydrophobic pharmaceuticals.

Structure-based 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 preparation and handling of biological samples and novel computational methods for the interpretation of measured FEL data (Mancuso *et al*, 2012). Our focus is on the imaging of cellular organelles like nuclei or mitochondria, exploiting the potential for single particle imaging experiments of FELs.

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 inherent properties of macromolecular structures (Grigorenko *et al*, 2013) 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, which is based on the use of pattern-recognition methods. An intuitive and user-friendly molecular viewer – the ArpNavigator (Langer *et al*, 2013) – enables user control of the model building process and provides easy access to a range of methods for quality assessment and model completion.

Future projects and goals

Together with international collaborators, we will undertake novel pilot projects aiming at interpretation of structural data obtained from various sources and projects of medical or biotechnological importance. We will continue to focus on software development for structural biology driven by general academic interest and continue contributing to provision of computational services, beamline facilities and applications for FEL-based diffraction.

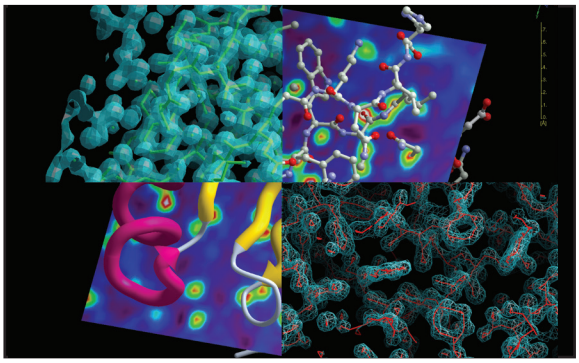


Figure 1: A selection of model-viewing options in ArpNavigator. Shown clockwise from the top left are a stick representation in solid electron density, a ball-and-stick representation in planar density, a skeleton representation of the electron density shown as a mesh and the protein in cartoon representation in planar density.

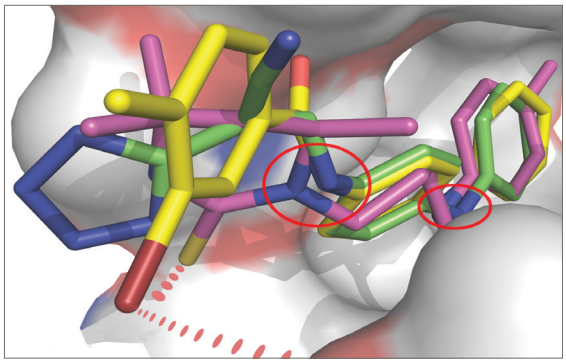


Figure 2: A known inhibitor in green, aligned with two hits from the ViCi software, in the binding pocket of beta lactamase. Important interactions maintained are highlighted with red circles and those created for exploration with full red spots.

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.

Group leader at EMBL Hamburg since February 2014.

SELECTED REFERENCES

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

Guettou F, *et al.* (2013) Structural insights into substrate recognition in proton-dependent oligopeptide transporters. *EMBO Rep.* 14, 804-10

Previous and current research

Cell membranes compartmentalise metabolic processes and present a selective barrier 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 in the cell. 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.

Using biophysical and biochemical methods, the Löw group aim to enhance understanding of the structural basis for substrate recognition in peptide transporters.

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 thus improve their absorption in the small intestine and subsequent distribution in the body. To this end we will study the proton-dependent oligopeptide transporter (POT) family using a combination of biochemical and biophysical methods. POTs share the canonical fold of MFS transporters with 12 predicted transmembrane helices each. There are currently no crystal structures available for any of the human PepT transporters, but the first bacterial PepT structures have recently been reported (including one from our lab).

Future projects and goals

- Structural and functional studies of the reaction cycle of bacterial POTs using X-ray crystallography.
- Structural and dynamic insights into the binding mode of POTs to peptides, drugs, and inhibitors.
- Functional expression, purification, biochemical characterisation and crystallisation of eukaryotic POTs.

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 (in whole cells or in reconstituted systems) with the major focus on nutrient uptake systems. The PETRA III synchrotron presents an excellent facility for us to continue this work.

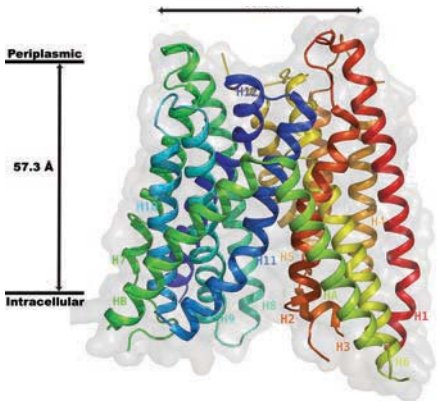


Figure 1: Structure of a proton dependent oligopeptide transporter (POT) in the inward open conformation (Guettou *et al,* 2013).

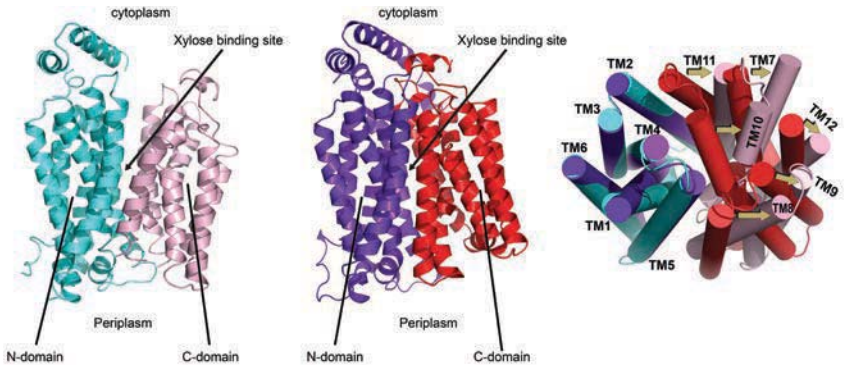


Figure 2: Structural differences between the inward open and occluded state structures of the sugar transporter XylE. The structures were overlaid and the transmembrane helices (TM) are labelled and arrows designate changes in the position of the helices upon opening of the transporter towards the cytoplasm.

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.

Previous and current research

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.

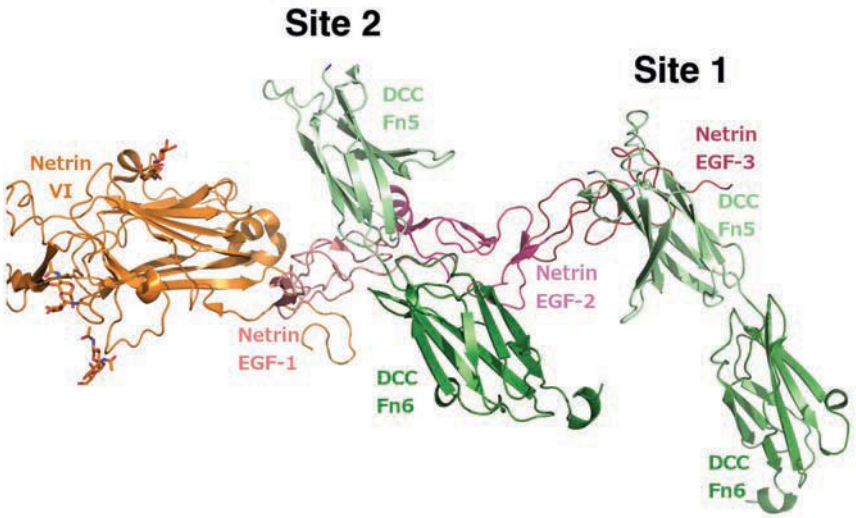
We have determined the crystal structure of the guidance molecule netrin-1 in complex with a fragment of the deleted in colorectal cancer (DCC) receptor (Finci *et al*, 2014), revealing how the external cue netrin-1 brings two DCC receptors in close proximity, leading to the propagation of a signal across 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 projects and goals

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.

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

Penengo L, *et al.* (2006) Crystal structure of the ubiquitin binding domains of rabex-5 reveals two modes of interaction with ubiquitin. *Cell*, 124, 1183-95

Previous and current research

During the past six years, we built three beamlines at the new PETRA III synchrotron in close collaboration with the Cipriani and Fiedler teams. Since 2012, the beamlines for small angle X-ray scattering on solutions (SAXS) and X-ray crystallography on crystals of biological macromolecules (MX) are open for users. The beamlines are embedded in an Integrated Facility for Structural Biology that supports non-specialists in taking a project from producing a suitable sample to determination of the structure by SAXS and/or MX.

Partly due to the enormous progress in synchrotron radiation-based structural biology, structural data on biological macromolecules are produced at an ever-increasing rate, creating the need to develop tools for efficient mining of structural data. We are developing tools for which the central concept is to use coordinate errors throughout all calculations. The necessity of this approach becomes clear when one considers that in contrast to sequence data, where a nucleotide entry can only be right or wrong, the precision in the location of an atom in a crystal structure can vary over several orders of magnitude. While the position of an atom in a rigid region of a protein giving diffraction data to high resolution may be known to within 0.01 Å, for an atom in a flexible region of a poorly diffracting protein the coordinate error may reach more than 1.0 Å.

From a technical point of view, extracting information from large amounts of raw structural data (as many as hundreds of structures containing thousands of atoms each) is a very complex task and requires sophisticated algorithms, both for the analysis and for the presentation and 3D visualisation of the results.

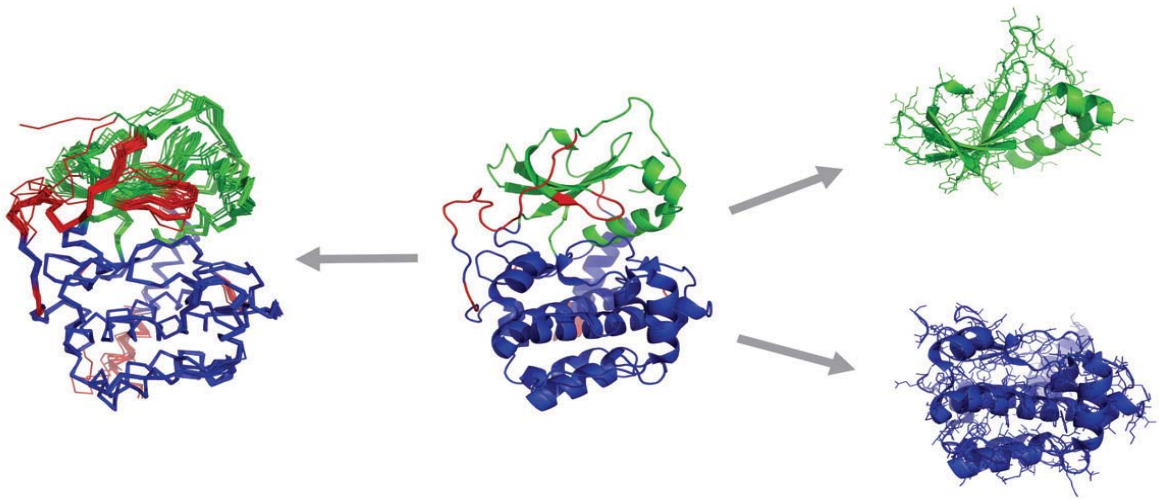
The Schneider group is operating two beamlines for macromolecular crystallography at the PETRA III synchrotron and develops new methods for structure determination.

Future projects and goals

We will continue the commissioning work on the new beamlines and strengthen the user programme to full capacity. The two beamlines for macromolecular crystallography will offer excellent conditions for data collection on small (micron-sized) crystals, crystallographic phasing, and the development of new experimental protocols for challenging systems.

On the computational side, we will work on improving the error models underlying our methods and on expanding our computational framework using genetic- and graph-based algorithms. We also plan to use recurrent structural fragments extracted from ensembles of structures as search models in molecular replacement and for the interpretation of low-resolution electron density maps. This aspect of our computational work will be very helpful in the interpretation of experimentally phased electron density maps obtained on the PETRA III beamlines.

For further information, see: **www.embl-hamburg.de/facilities/mx**



Rigid regions (centre; blue and green) identified from ensembles of structures can be used for superposition and subsequent analysis (left) or as fragments to interpret experimental data from methods with lower resolution than X-ray crystallography (right).

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

Valentini E, *et al.* (2015) SASBDB, a repository for biological small-angle scattering data. *Nucleic Acids Res.* 43, D357-63

Hagelueken G, *et al.* (2015) A coiled-coil domain acts as a molecular ruler to regulate O-antigen chain length in lipopolysaccharide. *Nat. Struct. Mol. Biol.* 22, 50-6

Ribeiro ED Jr, *et al.* (2014) The structure and regulation of human muscle α-Actinin. *Cell* 159, 1447-60

Petoukhov MV, Weissenhorn W & Svergun DI (2014) Endophilin-A1 BAR domain interaction with arachidonyl CoA. *Front Mol. Biosci.* 1, 20

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.

Previous and current research

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 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, group members help users not only with data collection, but also with 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 1 illustrates the study of the structure and regulation of human muscle α-Actinin by a combination of crystallography and SAXS. Molecular modelling revealed structural flexibility is essential for the crosslinking of antiparallel F-actin.

In 2014, the group launched a Small Angle Scattering Biological Data Bank (www.sasbdb.org, figure 2) a curated repository for small angle X-ray and neutron scattering data and models for the biological community.

Future projects and goals

- 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.
- Participation in collaborative SAXS projects at the P12 beamline.
- Further extension of P12 including time-resolved and anomalous scattering approaches.

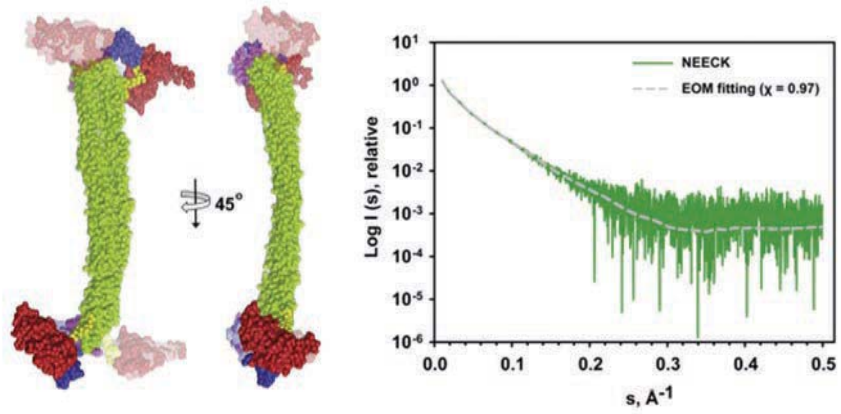


Figure 1: Structural plasticity of 200 kDa α-Actinin-2 in solution studied by SAXS (Ribeiro *et al.*, 2014).

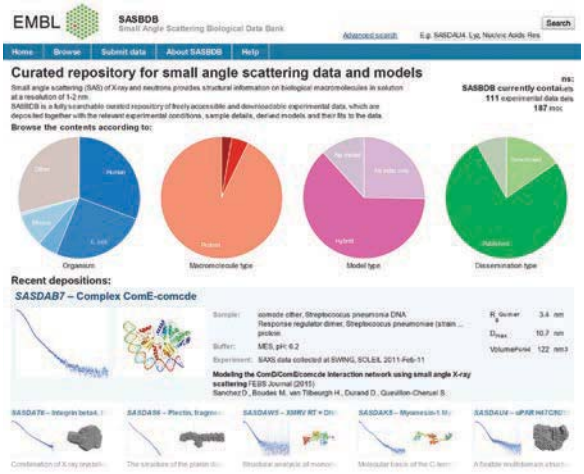
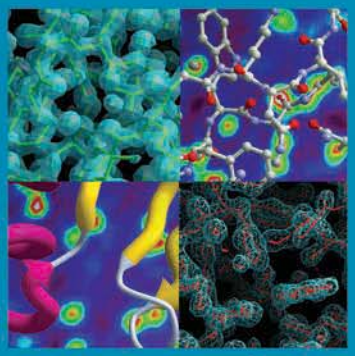
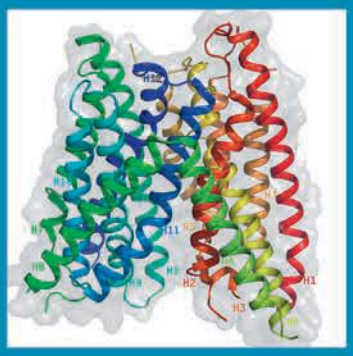
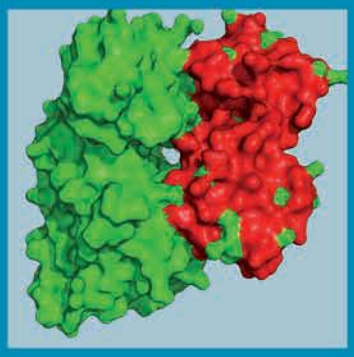
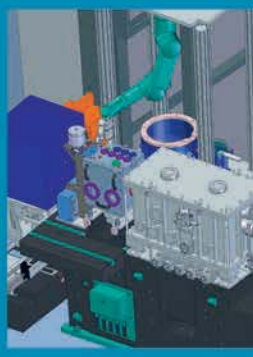


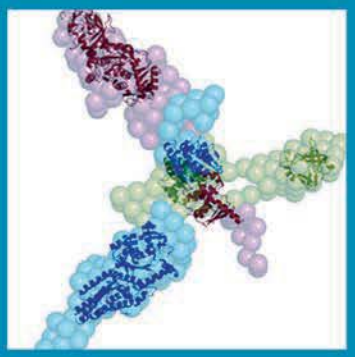
Figure 2: Web interface of SASBDB (Valentini *et al.*, 2015).



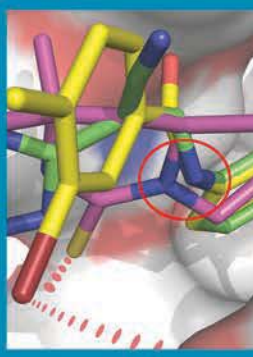


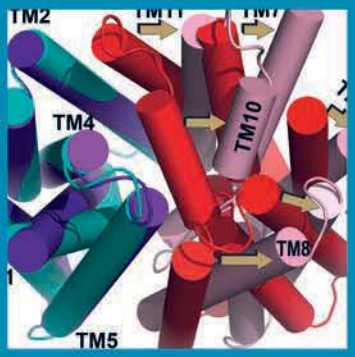


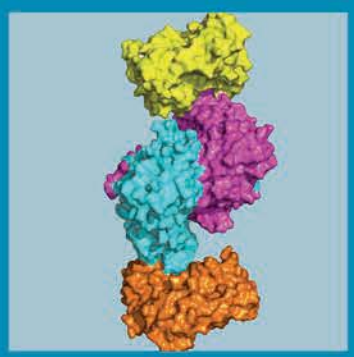


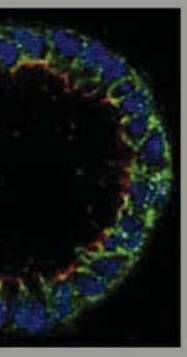
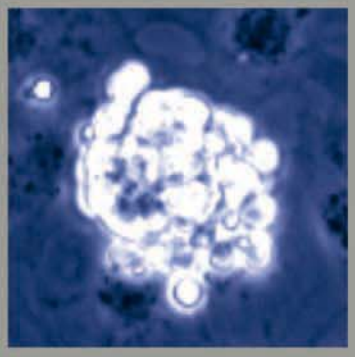
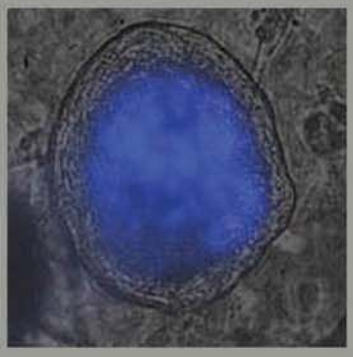
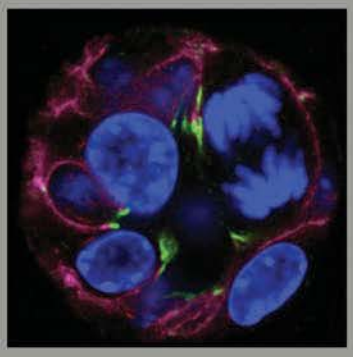
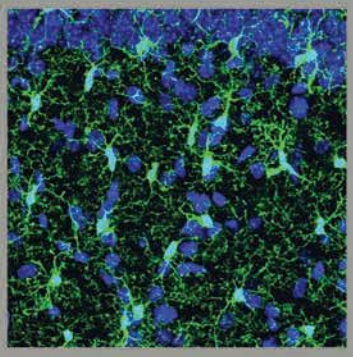



YY
VE
IN
NK
PD
QQ
DP
KN

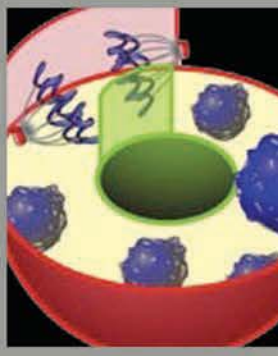



TM2
TM4
TM10
TM8
TM5



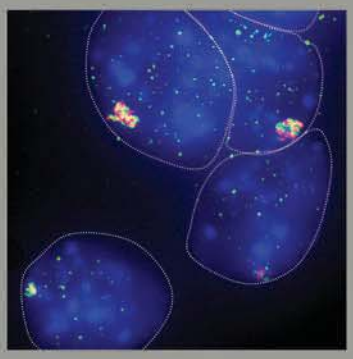






1974 - 2014 Research Highlights



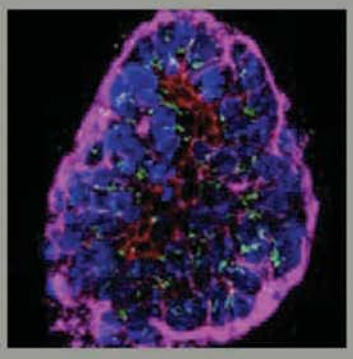
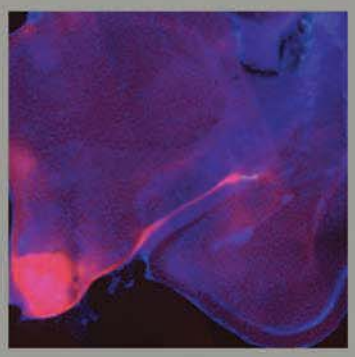
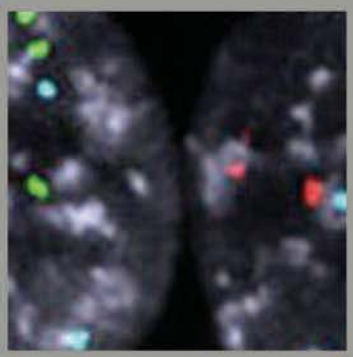
Identification of a crucial genetic switch for muscle repair, which switches white blood cells called macrophages from clean-up mode to promoting regeneration.
Ruffell D, et al. (2009) A CREB-C/EBPβ cascade induces M2 macrophagespecific gene expression and promotes muscle injury repair. *Proceedings of the National Academy of Sciences* 106, 17475-80

Revised model of how transposons – fragments of DNA that copy-and-paste themselves in the genome, with potentially hazardous consequences – are silenced: a protein called Mili generates piRNA molecules that guide another protein, Miwi2, to the transposon DNA to inactivate it.
De Fazio S, *et al.* (2011) The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature* 480, 259-63



Cells called microglia play a crucial role in brain development: they trim connections between neurons, shaping how the brain is wired.
Paolicelli RC, *et al.* (2011). Synaptic pruning by microglia is necessary for normal brain development. *Science* 333, 1456-8

Mouse model of cot death (sudden infant death syndrome, or SIDS): deficits in serotonin in the brainstem can be sufficient to cause cot death, supporting the idea that a congenital defect in serotonin function could be involved in the condition.
Audero E, *et al.* (2008) Sporadic autonomic dysregulation and death associated with excessive serotonin autoinhibition. *Science* 321, 130-3

EMBL Monterotondo

Mouse Biology

Dynamic partnerships and exchanges with other international academic research and clinical centres, and participation in multiple EU-wide mouse research and informatics initiatives, are integral parts of our discovery process into genetics and genomics, cell biology and pathology.

The continued refinement and diversification of methodologies for introducing genetic and epigenetic perturbations of cellular and physiological function is allowing the outstation to generate ever more accurate mouse models of human disease and multigenic disorders. The use of engineered nucleases in particular is allowing the 'democratisation' of experimental perturbation to virtually any mouse strain, allowing much more sophisticated approaches to be applied to understanding the complex interactions occurring between genes and across the genome- interactions which will likely have both epigenetic and environmental components.

Research groups at EMBL Monterotondo are using these powerful tools to investigate wide-ranging aspects of mammalian biology, including gene expression during early embryogenesis, haemopoietic development and differentiation, cancer and regeneration, behaviour and sensory perception.

A state-of-art animal facility provides a full range of mouse transgenic and gene knock-out production, embryo rederivation and cryopreservation services, together with a specialised phenotyping suite. Other centralised facilities include histology, confocal microscopy and flow cytometry.The many interactions with groups on other EMBL campuses including through the Interdisciplinary Postdoctoral (EIPOD) Fellowship Programme provides exciting and continuing prospects for joint projects. Ongoing international collaborations include those with groups at the University of Florida, the MPI Freiburg, the EPFL in Lausanne, the Gordon Institute in Cambridge, the Institut Pasteur in Paris, and the CNIO in Madrid Spain.

Philip Avner
Head of EMBL Monterotondo

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.

Head of EMBL Monterotondo since 2012.

SELECTED REFERENCES

Dubois A, *et al.* (2014) Spontaneous reactivation of clusters of X-linked genes is associated with the plasticity of X-inactivation in mouse trophoblast stem cells. *Stem Cells* 32, 377–90

Navarro P, *et al.* (2010) Molecular coupling of Tsix regulation and pluripotency. *Nature* 468, 457-60

Navarro P, *et al.* (2008) Molecular coupling of Xist regulation and pluripotency. *Science* 321, 1693-5

Duret L, *et al.* (2006) The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science*, 312, 1653-5

The Avner group combines genetics, genomics, biochemistry, and cell biology to study the nature of the key complex on the X chromosome responsible for X-inactivation.

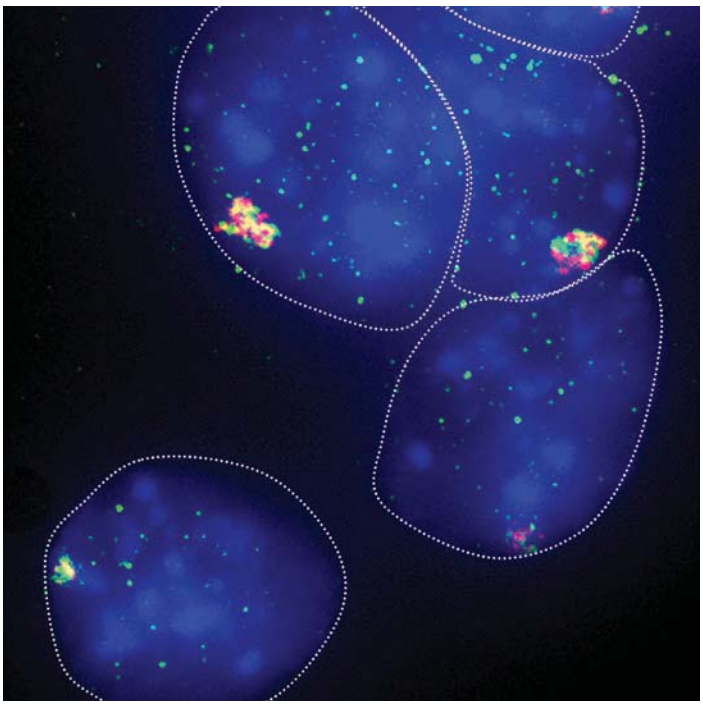
Previous and current research

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 research involves the study of the Xce locus (X-controlling element), a 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 on the basis of the differences in the stability of the 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.

Future projects and goals

Objectives of our future work will be to gain further mechanistic insights into variations in the X-inactivation process that occur in different cell lineages and the sensitivity of the process to trans-acting genetic factors. We will use our knowledge of the X-chromosome to explore more widely the contribution of epigenetic regulation to allele-specific epistasis, the process where the effects of one gene are modified in an allele-specific manner by one or several other genes.



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.*).

Developmental programming of 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

Previous and current research

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:

The Gross group uses pharmacological, histochemical, electrophysiological and behavioural genetic approaches to study the neural circuits underlying behaviour in mice.

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 projects and goals

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.

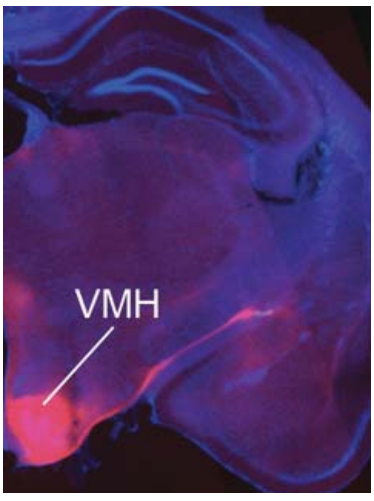


Figure 1: Expression of the hM4D/CNO pharmacogenetic inhibition tool in cell bodies and projections of the VMH in Nr5a1::hM4D-2A-tomatoF transgenic mice as revealed by farnesylated tomato (red) reporter protein (B. Silva).

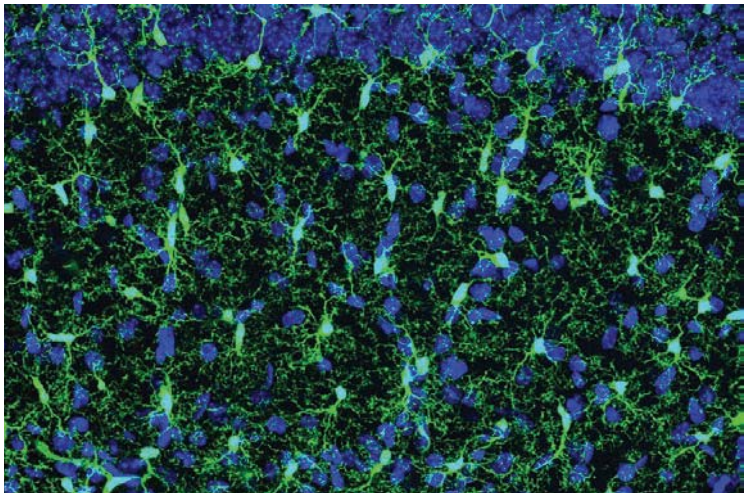


Figure 2: Microglia (green cells) visualised in the hippocampus of Cx3cr1GFP transgenic mice (cell nuclei labelled in blue, R. Paolicelli).

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

The Heppenstall group combines molecular, imaging and electrophysiological techniques to examine how sensory neurons turn information about touch and pain into electrical signals.

Previous and current research

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.

Future projects and goals

- 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.

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 since 2010.

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

Previous and current research

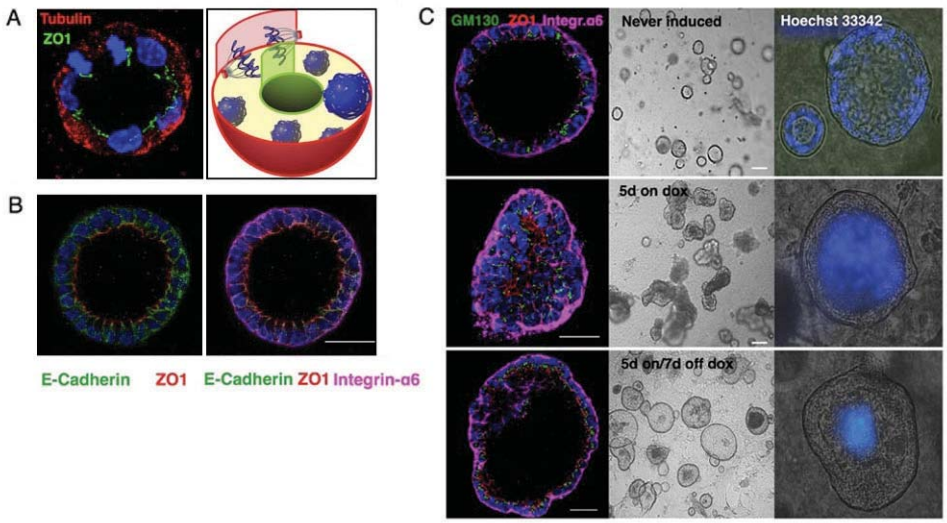
Extensive evidence now supports the concept of oncogene addiction (the dependence of tumour cells on their initiating lesion for survival). In patients and mouse models, interference with the activity of cancer-initiating oncogenes can result in tumour regression. However, novel therapies that target 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, exhibited characteristics of mammary gland progenitors and could efficiently re-populate the mammary fat pads of immundeficient mice. The successful isolation of a pure population of surviving cells after oncogene withdrawal will allow us to characterise these residual 'dormant' tumour cells in detail.

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

Future projects and goals

- Determine at which point during tumourigenesis cells acquire the ability to survive oncogene withdrawal.
- Identify the molecular properties that distinguish surviving-residual cells, from naïve 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.

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

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

Lancrin C, *et al.* (2009) The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage. *Nature* 457, 892-95

The Lancrin group studies the haematopoietic system and looks to develop strategies to improve methods for generating blood cells from stem cells.

Previous and current research

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 projects and goals

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 haemogenic endothelium from its precursor, the haemangioblast, and its subsequent commitment to haematopoiesis. Combining genomics, time-lapse microscopy, and loss and gain of function experiments *in vitro* and *in vivo*, we plan to identify and study the genes responsible for the generation of the first blood progenitors during embryonic life. Our research will further 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.

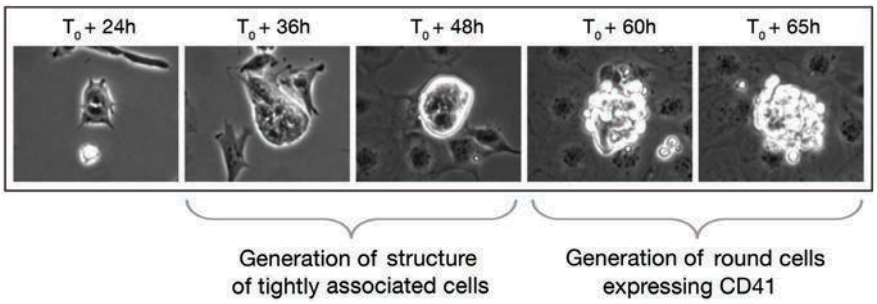


Figure 1: Time-lapse microscopy analysis of haemangioblast differentiation. The formation of a blast colony from the haemangioblast can be retrospectively divided in two consecutive phases: a generation of a structure of tightly associated endothelial cells and the production of round non adherent cells expressing the haematopoietic marker CD41.

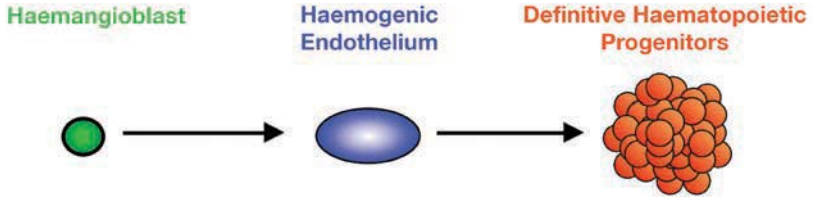


Figure 2: New model of blood cell origin: The haemangioblast and the haemogenic endothelium are part of the same developmental pathway to generate blood cell progenitors during embryonic life.

RNA function in germ and stem cell biology

Dónal O'Carroll

PhD 1999, Research Institute of Molecular Pathology, Vienna.

Postdoctoral research at The Rockefeller University, New York.

Adjunct member of faculty, The Rockefeller University, since 2007.

Group leader at EMBL Monterotondo since 2007.

ERC Investigator since 2012.

SELECTED REFERENCES

Comazzetto S, *et al.* (2014) Oligoasthenoteratozoospermia and infertility in mice deficient for miR-34b/c and miR-449 loci. *PLoS Genet.* 10, e1004597

Di Giacomo M, *et al.* (2013) Multiple epigenetic mechanisms and the piRNA pathway enforce LINE1 silencing during adult spermatogenesis. *Mol. Cell* 50, 601-8

De Fazio S, *et al.* (2011) The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature* 480, 259-63

Yang JS, *et al.* (2010) Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15163-8

Previous and current research

The integrity of the genome transmitted to the next generation intrinsically relies on cells of the germ line. Processes that ensure germ cell development, genomic stability, and reproductive lifespan are essential for the long-term success of a species. We are interested in characterizing spermatogonial stem cell (SSC) populations that support fertility as well the regenerative capacity of the testis throughout adult life. In addition, we tackle fundamental questions regarding the mammalian male germ line and heredity from an RNA perspective. Specifically, our research explores the contribution of non-coding RNA (miRNA, piRNA and lncRNA) and RNA modification pathways within germ cell development as well as testicular homeostasis/regeneration. Our research objectives focus on the contribution of these emerging pathways on the underlying circuitry of self-renewal that underpins the SSC, as well as the coordination of the various cellular/differentiation processes of spermatogenesis.

The acquisition of both pluripotency and totipotency is associated with the deregulation of transposable elements – our goal is understand the mechanisms by which germ cells manage this formidable threat to the gametes, and thus transgenerational genome stability. Specifically, we explore transposon silencing in the germ line by the Piwi-interacting RNA (piRNA) pathway as well as epigenetic mechanisms.

The O'Carroll group studies germ cell development and spermatogonial stem cell populations using state-of-the-art genetic strategies coupled with high-throughput sequencing approaches.

Future projects and goals

- Identification and characterisation of the spermatogonial stem cell populations.
- Dissection of the pathways required for spermatogonial stem cell self-renewal and testicular regeneration.
- Post-transcriptional RNA modification function in germ cell development.
- Establishment and maintenance of epigenetic transposon silencing in the male germ line.

Mitotic chromosomal instability and oncogene dependence

Rocio Sotillo

PhD 2002, Autónoma University (Madrid) and Spanish National Cancer Center (CNIO), Spain.

Posdoctoral work at Memorial Sloan Kettering Cancer Center, New York, USA.

At EMBL since September 2010.

ERC investigator since 2011.

HHMI International Early Career Scientist.

Professor University of Heidelberg since 2015

SELECTED REFERENCES

Rowald K, Sotillo R. (2014) Is aneuploidy deciding your fate? *Cell Cycle* 13, 1370

Schvartzman JM, *et al.* (2011) Mad2 is a critical mediator of the chromosome instability observed upon Rb and p53 pathway inhibition. *Cancer Cell* 19, 701-14

Sotillo R, *et al.* (2010) Mad2-induced chromosome instability leads to lung tumour relapse after oncogene withdrawal. *Nature* 464, 436-40

Schvartzman JM, *et al.* (2010). Mitotic chromosomal instability and cancer: mouse modelling of the human disease. *Nat. Rev. Cancer* 10, 102-15

Sotillo's research uses genetics and 3D cell culture systems to study the molecular mechanisms that lead to chromosomal instability and its consequences in tumour initiation and suppression.

Previous and current research

Chromosomal instability (CIN), the inability to correctly segregate sister chromatids during mitosis, is a hallmark of cancer cells. Overexpression of the mitotic checkpoint protein Mad2, commonly found in human tumours, leads to CIN and the development of aneuploid tumours in mouse models. Moreover, CIN can facilitate escape from oncogene addiction (the dependence of tumour cells on their initiating lesion for survival) and may be responsible for tumour relapse after targeted therapies. Very little is known about the mechanism of how and when CIN promotes tumour relapse. Our lab focuses on understanding the molecular mechanisms that lead to CIN and the consequences it may have in tumour initiation, suppression and relapse. We hope that the genes or proteins identified could be targeted therapeutically. We will use a combination of mouse genetics and highly innovative 3D *in vitro* culture systems.

Future projects and goals

- To study the dependence of tumour cells on the mitotic checkpoint *in vivo* and evaluate the potential for therapeutic interference with mitotic checkpoint genes.
- To study the effects of tumour regression and recurrence in chromosomally unstable tissues.
- To analyse the dual role of chromosome instability (tumour initiating and tumour suppressive) dependent on levels of aneuploidy, tissue type and molecular nature of the cooperating lesion(s).

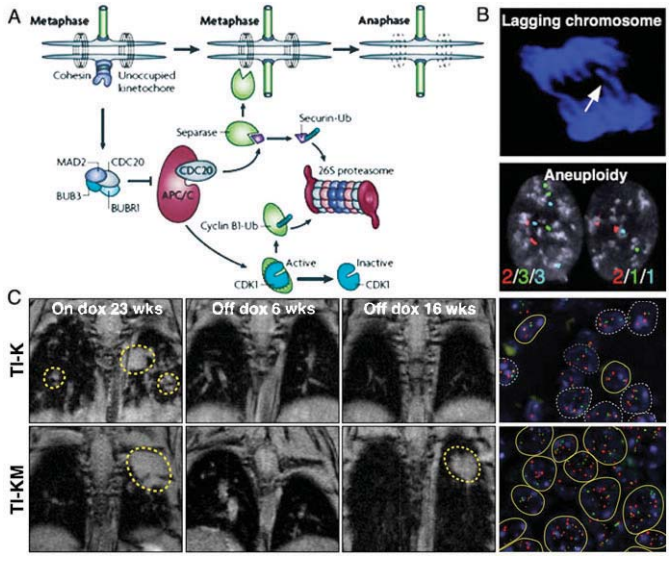


Figure 1:

A) The mitotic checkpoint complex (Mad2, Cdc20, BubR1 and Bub3) is shown to inhibit the anaphase promoting complex/cyclosome (APC/C) until the last kinetochore is correctly attached to microtubules.

B) Evidence of lagging chromosomes and aneuploidy measured by FISH analysis on cells overexpressing Mad2.

C) MR images of Kras transgenic mice (TI-K) and Kras+Mad2 (TI-KM) mice on doxycycline at indicated times showing lung tumours (yellow circles) (left panel), after 2-6 weeks of doxycycline withdrawal showing complete regression (middle panel) and after 14-16 weeks off doxycycline (right panel) showing recurrent tumours in mice that overexpress Mad2. Moreover recurrent tumours from these mice are highly aneuploid.

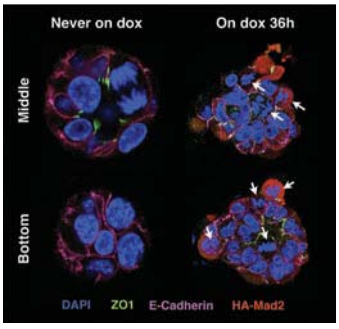
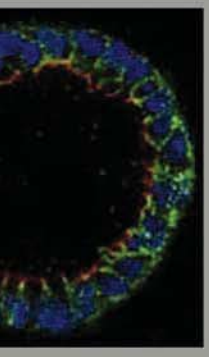
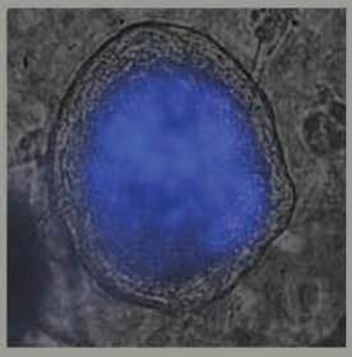
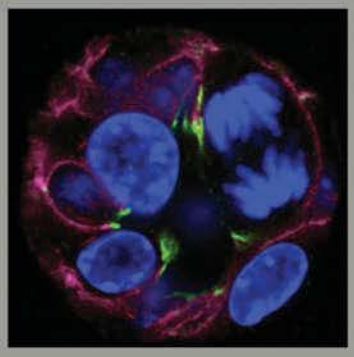
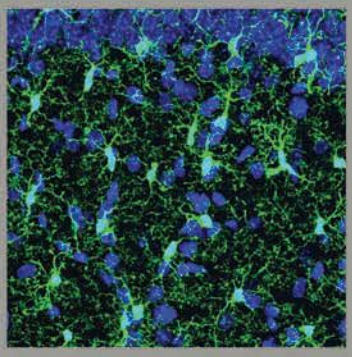


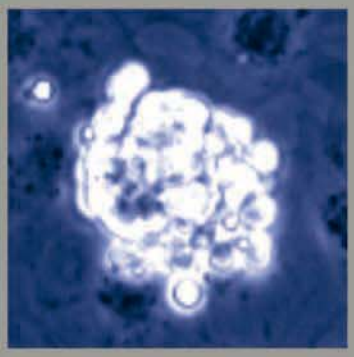
Figure 2: Primary mammary cells from TetO-Mad2/TetO-Myc/MMTV-rtTA mice grown in 3D culture. Left panel shows never induced cells that are beginning to form a polarized acinus. Right panel: loss of epithelial cell polarity in an acinus grown from tritransgenic cells and exposed to doxycycline for 36h. White arrows show abundant mitotic cells and lagging chromosomes after Mad2 overexpression.

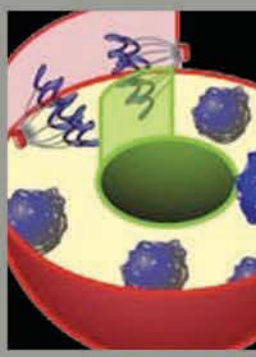


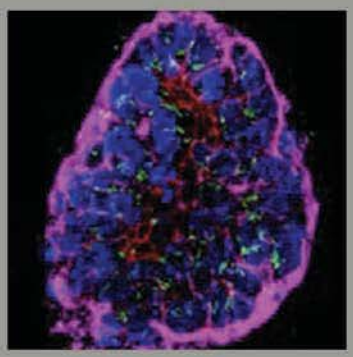


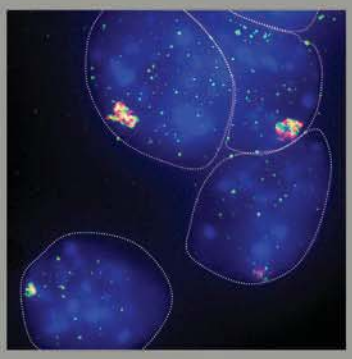


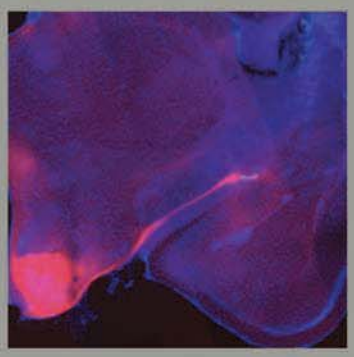


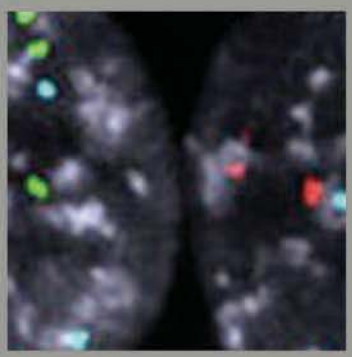












Index

Imprint

© EMBL 2015

Produced by Manuela Beck, Adam Gristwood, Isabelle Kling, Petra Riedinger, Mary Todd Bergman

EMBL Heidelberg
Meyerhofstraße 1
69117 Heidelberg
Germany
Tel. +49 (0)6221 387 0, Fax +49 (0)6221 387 8306
www.embl.org
info@embl.org

EMBL-EBI (European Bioinformatics Institute)
Wellcome Trust Genome Campus, Hinxton
Cambridge CB10 1SD
United Kingdom
Tel. +44 (0)1223 494444, Fax +44 (0)1223 494468
www.ebi.ac.uk
comms@ebi.ac.uk

EMBL Grenoble
6, rue Jules Horowitz, BP181
38042 Grenoble, Cedex 9
France
Tel. +33 (0)4 76 20 72 69, Fax +33 (0)4 76 20 71 99
www.embl.fr
info@embl.fr

EMBL Hamburg
c/o DESY
Notkestraße 85
22607 Hamburg
Germany
Tel. +49 (0)40 89 90 20, Fax +49 (0)40 89 90 21 04
www.embl-hamburg.de
info@embl-hamburg.de

EMBL Monterotondo
Adriano Buzzati-Traverso Campus
Via Ramarini, 32
00015 Monterotondo (Rome)
Italy
Tel. +39 06 90091285, Fax +39 06 90091272
www.embl.it
admin@embl.it

European Molecular Biology Laboratory

Heidelberg | Hinxton | Grenoble | Hamburg | Monterotondo

EMBL member states

Austria | Belgium | Croatia
Czech Republic | Denmark
Finland | France | Germany
Greece | Iceland | Ireland
Israel | Italy | Luxembourg
the Netherlands | Norway
Portugal | Spain | Sweden
Switzerland | United Kingdom

Associate member states

Argentina | Australia

Prospect member states

Hungary | Poland | Slovakia

www.embl.org

European Molecular Biology Laboratory