CRG
Centre for Genomic Regulation

Centre for Genomic Regulation
Annual Report 2011

Centres de recerca de Catalunya

Produced by: Department of Communication & Public Relations
Centre for Genomic Regulation (CRG)
Dr. Aiguader, 88
08003 Barcelona, Spain
www.crg.eu

Texts and graphics: CRG scientists, CRG members of the management team, Cristina Sáez, Anabel Herrera, Department of Communication & Public Relations

Graphic Design: Genoma ArtStudio SCP (www.genoma-artstudio.com)

Photography: Ivan Martí and CRG scientists

Cover photography: Luis Bejarano

Printing: Escardó, S.L.

Legal deposit: B-17275-2012

CRG
Centre
for Genomic
Regulation

CONTENTS

MESSAGE FROM THE FORMER DIRECTOR

As in previous years, in 2011 we once again look back and highlight the events that marked life at the CRG. This task is now Luis Serrano's, but since I was the director for the 6 first months of the year, Luis wanted me to use this occasion to express my feelings to the CRG community. I am grateful to Luis for the opportunity to document my thanks to all of you for your strong commitment to the CRG project and your continuous efforts towards making it successful. We can be proud of what we have achieved in the first 10 years, as the CRG approaches adolescence. And by we, I mean everybody at the CRG, because we have all joined forces with a clear common goal, namely to excel in science. This team spirit is one of our main strengths and is probably the reason for the survival of our high standards despite the dramatic economic crisis that threatens the country. For me it was difficult to leave the direction of the CRG at such a problematic time, given that the selection of the new director was a procedure controlled by the Scientific Advisory Board (SAB) and I had no say in it. But now it is a very good feeling for all of us to know that in these troubled times we have a new and strong colleague at the helm of the CRG and a powerful supporter of good science as head of our board of trustees. I am confident that the support from the Catalan administration will continue even if the economic crisis worsens. And I am sure that all of you will help Luis to exploit his, and your, imagination and creativity for ensuring that the CRG continues to be such a success. I will contribute my part in the next couple of years, but after leaving active research, I hope to be around in 2022 to celebrate with you the full development of the adult CRG on its 20th birthday.

Miguel Beato
Former director (2000-2011)

FOREWORD

"The CRG is undisputedly one of the biggest successes of Spanish science and an example of how a regional government can do things right"

Spain, whilst being considered a country that has contributed to culture enormously through art and literature, is not considered a place for innovation and research. Recapitulating the famous phrase of the Spanish intellectual, Miguel de Unamuno "Que inventen ellos" (let them make the discoveries), in Spain science and technology have generally played a marginal role. However, this stereotype is changing due to the strong support that some Spanish regional governments are giving to science. In this respect, 10 years ago the government of Catalonia decided to position the community amongst the strongest European regions in research and development. Instead of reinventing the wheel they sent scouts all over Europe to see what was working, to copy and adapt it to the national legislation. The result of this exercise was the creation of ICREA which ensures the hiring of top national and international scientists as well as a series of research institutes. A landmark of the new institutes is the absence of civil servants, generous funding and budget autonomy.

To someone that did his PhD in the Centre for Molecular Biology (CBM) in Madrid more than 26 years ago, the incredible success of the Catalan initiative amazes me. If someone had told me at that time that we would have institutes in Catalonia and Spain that are, according to objective criteria, among the top 25 in the world, and that we would have research centres with more than 70% of the PhDs and posdocs being international, I would have laughed. But this is the reality. A reality that we need to preserve even in the midst of the economic crisis.

Looking at the CRG 10 years ago and where it is now, I realise the titanic work of our previous Director, Miguel Beato. Thanks to his vision and the support of the regional and central authorities he has managed to create an institute that competes in the international league and shows that Spain and Catalonia do not only have sun and culture, but provided with the right vision and investment they can be as innovative and competitive as any other country in the world.

It is my hope that in my mandate as CRG Director I will contribute to the development of the CRG as a top research institute. My intention is not only to have a place where great international science is carried out, but also to help the development of the regional and national economy, educate society as to the importance of science and expand the influence of the CRG not only in Europe, but worldwide. Time will tell if I am successful.

Luis Serrano
Director

A LOOK BACK AT THE YEAR

A LOOK BACK AT THE YEAR

The most significant change in 2011 was the replacement of Miguel Beato as Director of the CRG in June. Miguel Beato was CRG director for 10 years and following CRG rules he stepped down after an international search for a new Director. Following the recommendations of the SAB, Luis Serrano was appointed as new CRG Director.

All together the researchers and management team are contributing to make the CRG an attractive place for young scientists from around the world.

> *The SCImago Institution Rankings (SIR) World Report 2011*, covering the years 2005-2009, classifies the CRG in 16th position (according to the Q1 indicator) out of over 3,042 research institutions in the health sector around the world. In Europe only four other research centres have a higher Q1 indicator. Given that the CRG was officially inaugurated in the fall of 2002, this is a remarkable success.

> The number and quality of the papers published by CRG scientists have reached a plateau since 2009 reflecting the growth freezing experienced in that year. In 2011, 185 papers were published in international journals with an average impact factor of 9.32, and 118 seminars were held by very high quality invited speakers. Many of these activities found their way to the public media and the CRG hit the news (newspapers, radio, TV) on 670 occasions. We also made an institutional video that can be seen on our website and on the CRG YouTube channel.

> Finally some of our young and senior scientists have obtained different awards in recognition of the excellence of their science. Ben Lehner obtained the City of Barcelona Award 2011, Xavier Estivill the Preclinical Biomedical Research Lilly Foundation Prize, Miguel Beato the National Research Award from the Catalan Institution for Research Support Foundation and Marc Bataller the PhD Extraordinary Prize from the University of Barcelona.

Xavier Estivill receiving the Lilly Foundation Prize

RESEARCH

Systems Biology

After his appointment as the new director of the institute, Luis Serrano stepped down as the Systems Biology programme coordinator. As a result, James Sharpe was nominated acting coordinator of both this programme and the EMBL/CRG Systems Biology Research Unit.

The 6 groups in the programme cover a wide range of topics: from dynamic gene regulatory networks to systems neuroscience, and employ a wide range of model systems to address these issues, including prokaryotes, *C. elegans, Drosophila* and mice. Underlying this diversity, however, is a common goal to combine systematic and quantitative data collection, using computational models, going beyond molecular descriptions and arriving at a deeper dynamic understanding of complex biological processes. To achieve these goals the programme is strongly interdisciplinary, comprising an increasing number of physicists, mathematicians and computer scientists, in addition to molecular biologists. In 2011 the programme was awarded many new grants (including BioPreDyn and FLiACT, coordinated by Yogi Jaeger and Matthieu Louis respectively, and 4D Cellfate. coordinated by Luciano Di Croce). Given the wide range of biological questions addressed in the programme, a very interactive 2-day retreat was organised this year, in which postdocs and students had to propose ways to tackle the scientific questions from each other's labs. The programme also offered a very successful 1 week summer school on modelling in systems biology, with an international team of lecturers. Publication highlights during 2011 included papers on signal transduction, protein structure, gene regulatory networks, multicellular patterning, morphometrics, larval chemotaxis, and the impact of stochastic molecular noise at the whole organism level.

Gene Regulation

Throughout 2011 the three groups of this programme continued to combine genome-wide methodologies with detailed mechanistic analysis to study different aspects of gene regulation, including various levels of chromatin structure, from nucleosome positioning to chromosomal interactions, histone and DNA epigenetic marks, genome and transcriptome binding sites for transcription and RNA binding factors, to understanding the regulation of transcription, pre-mRNA splicing and mRNA translation. As a consequence, the number of bioinformaticians in the programme has continued to increase. Apart from the Genomics Unit, other core facilities, in particular Proteomics and Advanced Light Microscopy, have been used heavily. Groups working on splicing and chromatin in the programme have continued collaborating to identify chromatin features that influence alternative splicing. The "Chromatin Group" has continued to meet on a monthly basis, along with groups from the Cell Differentiation and Cancer and Bioinformatics and Genomics Programmes.

The programme's search for a new group leader was finally successful and Guillaume Filion, from the Dutch Cancer Centre in Amsterdam, will start in January 2012, just in time for the second external evaluation of the programme that will take place from January 11-13, 2012.

Differentiation and Cancer

The areas of research covered within this programme are stem cell biology and regeneration as well as epigenetic mechanisms of cell fate instruction, cancer and ageing.

The year 2011 saw a number of important publications by the programme. Three highlights will be mentioned briefly. Firstly, Luciano Di Croce's group discovered an unusual mode of action of the Myc oncogene that is relevant for leukemia formation. They found that the oncogene binds to the retinoic acid receptor in leukemia cells, thereby helping to repress myeloid differentiation genes. However, when phosphorylated, Myc turns into an activator of differentiaton genes, thus switching from an oncogene to a tumour suppressor. This switch is not only interesting mechanistically but may also represent an Achilles heel for novel drug approaches to the treatment of acute myeloid leukemia.

In a second report Salvador Aznar Benitah and co-workers dissected the role of CBx4, a component of the polycomb repressor complex one, in the regulation of epidermal stem cell (EpiSC) functions. They discovered that, surprisingly, different domains of Cbx4 play a role in protecting EpiSCs from entering into differentiation on one hand and senescence on the other.

In the third publication, Salvador's group made another important discovery. They found an unexpected link between the circadian clock machinery and EpiSC heterogeneity. Using various genetically modified mouse models they showed that perturbations of clock genes changes the equilibrium between dormant and activated (cycling) EpiSCs. These data suggest that the responsiveness of EpiSCs to environmental stimuli triggering their differentiation is influenced by daylight. In addition, they discovered that circadian clock genes also control the efficiency of tumour formation in the skin. These experiments open a whole new avenue of research linking the circadian clock machinery with both adult stem cell homeostasis and cancer formation.

In addition, Luciano Di Croce was appointed associate member of the Network of Excellence EpiGenSys.

Genes and Disease

The global objective of the Genes and Disease programme is to carry out research on the molecular basis of human disease, from the discovery of genes involved in human disorders to the development of preventive and therapeutic strategies. The efforts of the Genes and Disease programme researchers are focused on analysing sequence, structural and epigenetic variants of the human genome that could participate in the predisposition and evolution of disease. Specific collective work within the programme is focused on understanding the function of genes with potential implications for mental retardation, psychiatric disorders and neurodegeneration, by using bioinformatic, cellular and animal model approaches.

Investigators from the programme (Estivill's group), with their analytical pipeline for structural variations (SV), have contributed to the characterisation of the genome of chronic lymphocytic leukaemia (Puente *et al. Nature* 2011; *Quesada et al. Nature Genetics* 2011) and body mass index phenotypic disorders (Jacquemont *et al*. *Nature* 2011). Estivill's group has identified that a common SV, involving the deletion of two genes (LCE3B and LCE3C) is associated with psoriasis, psoriatic arthritis and rheumatoid arthritis predisposition (Docampo *et al*. *Arthritis Rheum* 2011; Riveira-Munoz *et al. J Invest Dermatol* 2011). This group has also identified the role of several non-coding RNAs in psychiatric and neurodegenative disorders, detecting alterations in several miRNAs in the early stages of Parkinson's disease (Muiños-Gimeno *et al. Biol Psychiatry* 2011; Miñones-Moyano *et al. Hum Mol Genet* 2011). The work of Estivill has received the Lilly Foundation Biomedicine and the Fibromyalgia Foundation awards.

The group of Susana de la Luna and Mariona Arbonés has shown that dysregulation of the apoptotic response in differentiating neurones participates in the neuropathology of diseases that display DYRK1A gene-dosage imbalance effects, such as Down's syndrome (Calvo *et al. Genes Dev* 2011). They have shown that mouse lineages overexpressing Dyrk1A present alterations in cellularity, and altered electrical responses in the mature retina. The group has studied other members of the DYRK family of protein kinases and has characterised the less known member of the family, DYRK4, and found that alternative splicing is used to generate functional diversity, and that substrate specificity may be a critical factor governing biological specificity among DYRK protein kinases (Papadopoulos *et al. J Biol Chem* 2011).

The group of Cristina Fillat has nicely demonstrated the implication of E-cadherin in the modulation of the gap-junction mediated TK/GCV bystander effect (Garcia-Rodríguez *et al. Gene Ther* 2011). They have also shown the key role of connexin-26 in the bystander effect of gemcitabine toxic metabolites in pancreatic adenocarcinoma (Garcia-Rodríguez *et al. Mol Cancer Ther* 2011). Fillat's group has shown that TK-expressing oncolytic adenoviruses can be traced by PET imaging providing real time information on the activity of the virus (Abate *et al. PLoS One* 2011). The groups of Cristina Fillat and Mara Dierssen have shown that overexpression of RCAN1 in a transgenic mouse results in marked cognitive defects (Dierssen *et al. PLoS One* 2011).

The group of Mara Dierssen has characterised a mouse model that overexpresses the CHRNA5/A3/B4 genomic cluster (Gallego *et al. Amino Acids* 2011; Gallego *et al. Alcohol* 2011; Viñals *et al. Drug Alcohol Depend* 2011).

The group of Stephan Ossowski has developed several bioinformatic tools for the Next Generation Sequencing (NGS) analysis of genome, transcriptome and epigenome data. His group has integrated several NGS-based analysis methods including whole genome and exome sequencing, ChIP-seq, RNA-seq and DNaseI-seq in order to detect genomic, genic and epigenomic variation related to disease or intolerance to specific treatments. He is improving computational analysis algorithms for each sequencing method and integrating and correlating the resulting marks and measures, *e.g.*, genomic variants, gene

expression, histone modifications, DNA methylation and chromatin conformation (Guo *et al. Plant Physiol* 2011; Schneeberger *et al. Proc Natl Acad Sci* 2011; Hu *et al. Nat Genet* 2011).

At the end of the year, Cristina Fillat and Mariona Arbonés and their labs left the CRG. Cristina and her Gene Therapy lab joined the August Pi i Sunyer Biomedical Research Institute (IDIBAPS), in Barcelona, within the Paediatrics and Developmental Medicine programme. Mariona and her lab moved to the Molecular Biology Institute of Barcelona (IBMB-CSIC).

Mara Dierssen was elected president of the Spanish Society of Neurosciences (SENC).

Bioinformatics and Genomics

Throughout 2011, the groups in this programme continued to use computational methods to understand the forces shaping genome sequence and function, but they have increasingly combined computational approaches with high throughput interrogation of genome activity and evolution - mostly through the use of Massively Parallel Sequencing instruments. Moreover, the efforts of some of the groups towards more classical experimental validation have also grown substantially, and, in consequence, the size of the programme devoted to experimental work has grown in parallel. It is in this interface between experimentation and computation where the most relevant contributions of the programme will occur in 2012 - notably the work in process by the group of Gian Gaetano Tartaglia in the development of methods to predict interactions between long non-coding RNAs (lncRNAs) and proteins, and the work of Cedric Notredame, also in process during 2011, on the use of high throughput expression data together with evolutionary conservation to discover and annotate long non-coding RNAs.

Cell and Developmental Biology

The Cell and Developmental Biology programme specialises in the study of mechanisms involved in cell compartmentation, shape, structure and division.

A combination of *in vitro* biochemical analysis, yeast and zebra fish genetics is used, as well as mammalian cell- and *Xenopus* egg-based assays to address the specific processes of membrane trafficking, organelle dynamics, protein translocation, cell shape, spindle dynamics and cytokinesis.

Julia von Blume from the Malhotra lab was recruited to the Max Planck Institute for Biochemistry in Martinsried, Munich as a group leader, and Hernán López-Schier, one of the group leaders of the programme, and his lab left the CRG and joined the Helmholtz Zentrum München, in Germany, where he will run the Sensory Biology & Organogenesis Unit.

Additionally, Isabelle Vernos, senior group leader, was elected to the council of ERC Research Council.

CORE FACILITIES

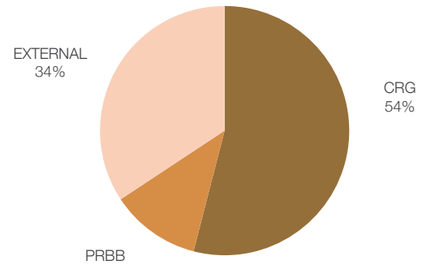
The core facilities and the technology they offer continue to be a valued support for the research performed at the CRG. The programme currently comprises eight Core Facility Units, Genotyping, Microarrays, Ultrasequencing, Proteomics, Advanced Light Microscopy, Biomolecular Screening & Protein Technologies, FACS, and Bioinformatics, as well as the Histology Service that is only accessible to internal users.

In 2011 the Core Facilities experienced some remodelling, change and improvement. In January Renza Roncarati joined the CRG as Head of the High-throughput Screening Unit. Renza has long-standing experience in biology, assay development and automation, compound and siRNA screening, and has worked both, in academic as well as industrial environments. In October the High-throughput screening Unit and the previously only internally accessible Protein Service were merged, and Renza is now running the “Biomolecular Screening & Protein Technologies Unit”. Taking advantage of already existing collaborations between the two Units, by integrating expertise and technologies this fusion provided a platform for synergistic developments in the fields of medium to high-throughput RNA interference assays, chemical screening, protein engineering and biophysical characterisation of proteins and nucleic acids.

The Proteomics and Bioinformatics Units also experienced personnel changes. Guglielmo Roma left the CRG in July and Ernesto Lowy has been managing the Bioinformatics Core since then. Henrik Molina left the CRG in August, Cristina Chiva took over as interim Manager, and Eduard Sabidó was recruited from the ETH Zurich to take over the Head of Unit position at the beginning of 2012. In 2011 for the first time, the IMIM has been contributing personnel to the Core Facilities: Sabrina Bascones joined the FACS Unit and Ines Guimaraes the Microarrays Unit. The major equipment upgrades in 2011 have been 1) the purchase of a triple quadrupole mass spectrometer, which now enables the Proteomics Unit to perform targeted, quantitative proteomics analysis, and 2) the installation of the Roche-NimbleGen platform in the Microarrays Unit.

As in previous years, the overall activity in Core Facilities continued to increase. Compared to 2010, the invoicing for core facility services increased by 30%, and the users contributed significantly to covering the fixed cost of the platforms (equipment depreciation and maintenance, personnel and structural cost). The CRG core facilities are not only well-established locally, with users coming from different institutions in Barcelona and Spain (as well as from abroad), but are also a recognised partner in European initiatives. The Advanced Light Microscopy Unit is a partner in the ESFRI initiative EuroBioimaging, the Proteomics Unit is a transnational access and research site with the European Infrastructure network PRIME-XS, and the Genotyping Unit is a transnational access site in the European infrastructure network ESGI.

Core Facilities - Services Use % 2011



ADVANCED TRAINING

In 2011, the **International PhD Programme** continued to attract much young talent from across the globe, supported both by internal and external competitive funds and the **“la Caixa” International PhD Fellowship Programme**. The 10 candidates selected in the “la Caixa” call started their PhD theses in September 2011, after a highly competitive selection process with 258 candidates from over 70 different countries competing. In 2011, the specific training offered to PhD students included 18 *Advanced Seminars in Biomedical Research*, in partnership with the *Universitat Pompeu Fabra*, as well as 29 practical scientific and technical courses organised by the CRG faculty and Core Facilities. Also in 2011, the CRG PhD community actively promoted a number of initiatives, including the annual PhD symposium (24-25 November 2011), a PhD retreat (6-7 November 2011), and a joint retreat with the Friedrich Miescher Institute for Biomedical Research in Switzerland (20-22 May 2011).

The **International Postdoctoral Programme** at the CRG currently hosts around 100 postdocs supported by internal and competitive funding from highly prestigious institutions and the **INTERPOD Postdoctoral Fellowship Programme**, co-funded by the European Commission under the COFUND scheme. In 2011, 19 external fellowships and postdoctoral contracts were awarded to CRG postdocs, including 5 EMBO long-term, 3 Marie Curie fellowships and 3 postdoctoral fellowships from the Swiss National Foundation. The second INTERPOD call in 2011 attracted 30 applications for 6 positions. In 2011, postdocs at the CRG also organised their own symposium (10 June 2011) and a retreat (28 October 2011).

In addition to the broad range of high profile CRG/PRBB seminars and conferences with invited speakers and internal seminars (such as data and journal clubs), since 2011 the CRG has offered **practical training courses**. The first course was a Summer School in Systems Biology (July 2011) addressed to both PhD students and postdocs. **Complementary skills and career development courses** are provided by the successful Career Development Programme INTERVALS open to PRBB residents. In 2011, 29 courses were offered ranging from grant and scientific writing to leadership, methodology of biomedical publications, job opportunities in science related sectors and intellectual property. Attendees from the CRG represent 26% of the total attendees of the PRBB with 71 participants; most of them being PhD students and postdocs, but young Principal Investigators and other staff also benefited from the INTERVALS training courses in 2011. **Additional non-scientific courses** were also offered by the CRG in 2011: language courses, biosafety training and the first edition of a course on scientific poster making. In 2012, the CRG plans to expand the offer of practical training courses, covering scientific and technological courses, as well as career development and complementary skills.

Currently, there are several **mobility agreements** in place at the CRG: *Università* Sapienza in Rome within a *Leonardo da Vinci* Project for diploma students (3 placements in 2011), *Universitat Politècnica de Catalunya* within the Master Erasums Mundus in Data Mining and Knowledge Management (1 placement in 2011) and *Università San Raffaelle* (leading to a joint retreat in 2012). Finally, the CRG has initiated efforts with top-level biomedical institutes worldwide to establish **training and mobility agreements**. The most remarkable negotiations are with the biomedical research institute in India, *the National Center for Biological Sciences* (NCBS) for the establishment of a joint NCBS-CRG Postdoctoral Fellowship Programme.

Advanced Training figures

PhD students: 98 (81% foreigners)
External fellowships obtained by PhD students: 9 +10 “la Caixa” fellowships
Advanced Seminars offered to Masters and PhD students: 18
Practical courses offered to PhD students: 29
PhD theses defended: 12
Postdoc researchers: 99 (70% foreigners)
External fellowships obtained by postdoctoral researchers: 19

PhD Theses Defended in 2011

Name	Programme	Date	University	Thesis Project
Andreia Carvalho	Systems Biology	03-03-11	UPF	*Engineering synthetic spatial patterning sytems in mammalian cells*
Carla Obradors	Genes & Disease	12-04-11	UB	*Impacte de la desregulació in vivo de l’expressió del gen ntrk3 (trkc) sobre el desenvolupament cortical d’un model murí del trastorn de pànic i ansietat*
Lars Bussmann	Differentiation & Cancer	06-05-11	University of Heidelberg	*Insights into C/EBPa and C/EBPb Mediated Direct Reprogramming of B Cells into Macrophages*
Bernd Boehm	Systems Biology	18-05-11	UPF	*Are isotropic cellular behaviours sufficient to explain limb morphology?*
Hagen Tilgner	Bioinformatics & Genomics	02-06-11	UPF	*Modelling Splicing*
Martin Schutz	Cell Biology	01-07-11	UPF	*The role of NIMA-like kinase Nek9 in mitosis*
Indra Wibobo	Cell Biology	04-07-11	UPF	*Zebrafish lateral line system: The roles of Eya 1 in migrating primordium and notch signalling in hair cell development and regeneration.*
Erik Verschueren	Systems Biology	17-08-2011	University of Brussels	*Evolution of specificity in peptide recognition domains*
Lorena Pantano	Genes & Disease	26-09-2011	UPF	*Full characterisation of the small RNA transcriptome using novel computational methods for high-throughput sequencing data: study of miRNA variability in eukaryotic organisms*
Elena Miñones	Genes & Disease	03-11-2011	UB	*Characterisation of the contribution of small non-coding RNA deregulation to Parkinson’s disease pathogenesis*
Nuno Miguel Luis	Differentiation & Cancer	07-11-2011	UPF	*The role of Cbx4/Polycomb-2 in epidermal stem cell homeostasis*
Marco Constante	Systems Biology	01-12-2011	UPF	*Development of synthetic biology devices for iron metabolism research*

ALUMNI

CRG Alumni began in 2011, at the request of some of the researchers from the centre with vast scientific experience, and who have come from other research institutes in various different countries where this initiative already exists.

With the aim of continuing its growth and working to become an institute of excellence and an international benchmark, the CRG, since its foundation, has trained and produced a considerable number of excellent scientists who, after leaving the CRG, have continued to contribute their knowledge and energy into social welfare in various sectors.

Initially, CRG Alumni was conceived as a web project, which could become a platform for all of the researchers and employees who have been part of the CRG at some point in their professional careers, where they could share both professional and personal experiences.

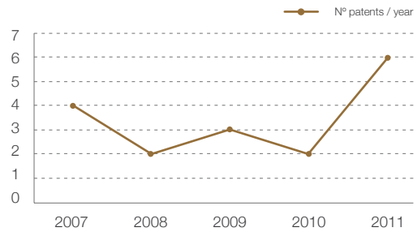
What does CRG Alumni offer?

> Access to a network of contacts who have a shared experience in the field of research excellence
> Priority registration for symposia, conferences, workshops and free courses organised by the CRG, and special registration fees for those activities that are not free
> A job offer section, contributed to by the CRG Alumni members themselves
> The opportunity to create or develop new links to or projects with network members.
> The ability to offer to the members any other initiative, which arising from CRG Alumni that could be of interest to the community.
> The possibility, in the future, of being able to participate in activities or meetings of former CRG members

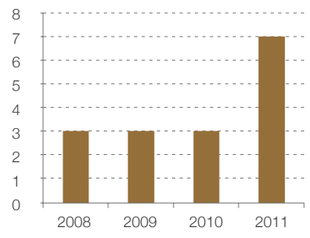
TECHNOLOGY TRANSFER

Compared with previous years, in 2011 the Technology Transfer Office (TTO) doubled the number of patents filed and the number of research and license agreements signed with companies. Revenues from software licences and antibodies licenses have been constantly increasing. The TTO won two assessment grants of more than 500,000€ to support the valuing of patents with potential interest for the companies. The TTO has also signed a strategic research agreement with Sanofi, including, from the very beginning, 4 specific research projects with a global economic value of more than 1M €. This strategic agreement is the first Sanofi have ever signed with a research institute in the south of Europe.

Nº patents / year



Nº licenses
and contract research agreements

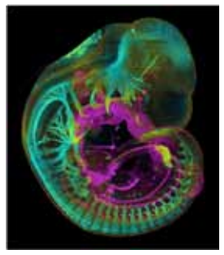


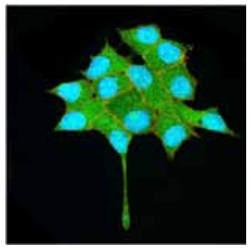
OUTREACH

One of the strategic objectives of the activity of CRG is “to communicate and establish a dialogue with society, educating the public and taking into account their demands and needs.” To this end, during 2011 more than 90 activities were organised, including workshops for primary and secondary schools, summer internships, the open day, the cell model contest (sculptures), science cafés, “Easy Science” conferences and participation in fairs and events for the dissemination of science.

Throughout the year we also worked on the institute’s repository of scientific photographs. In order to expand the photographic resources, in the autumn we organised the first edition of the internal CRG Scientific Photography Competition. Of note is the success of the participation and quality of the photographs submitted. The winning entries were announced at the Christmas party, where the authors received their prizes.

Awarded pictures of the CRG Scientific Photography Competition:
Scientific category (left): Mouse embryo with neurofilament and E-Cadherin staining, by Jim Swoger, Juergen Mayer and Laura Quintana
Artistic category (right): Tree of life, by Julia Burnier





The different findings published in top scientific journals, the activities organised and the self-recognition of the CRG as a reference research institute generated throughout the year a series of articles in print and online media, as well as participation by scientists on radio and television programmes.

Moreover, during 2011 the presence of the CRG in social networks, which are proving to be a very effective additional dissemination tool for all activities, increased. The CRG is also participating in two EU funded science communication projects: EuroStemCell (since 2010) and CommHERE (since 2011), together with leading research institutions across Europe.

2011 was a particularly busy year in terms of seminars, sessions and scientific meetings, all held at the facilities of the institute. In addition, other activities such as Master’s and doctoral classes, training courses, *etc*., have also been organised.

As for the scientific meetings, it is worth highlighting the “Frontiers in Biology” symposium, the “Epigenetics Mechanisms in Health and Disease” workshop, the CRG Annual Symposium entitled “Computational Biology of Molecular Sequences,” and three Core Facilities Technology Symposia.

CRG Outreach figures

Outreach activities: 93
Audience reached: nearly 5.500 people
Press releases: 31
Short pieces of news (web): 32
Written/online media appearances: 587
Radio / TV appearances: 45
Blogs: 38
International meetings: 15
High profile seminars: 118

GENERAL ADMINISTRATION

Management & Research Support

The main goal of the Management & Research Support team is to minimise the administrative tasks of the scientific team so that they focus on developing science of excellence. With this aim, highlights of the past year were:

> Implementation of a new supplier web portal to improve the purchasing and accounting process.

> Improvement of internal accounting control mechanisms, implementation of a new accounting report.

> Improvement of the scientific-technical services internal billing system, which has significantly reduced red-tape. The Scientific-technical services billing has increased by 29%.

> Implementation of improvement actions within the CRG Equal Opportunity Plan, efforts to integrate disabled people at the institute and participation in a specific commission to develop and to implement the Corporate Social Responsibility Plan of the Centre. All these projects are aligned with achieving good HR policies in relation to the excellence of the centre.

> A contingency plan for both IT and General Services has been defined. Their conclusions will be implemented during 2012 and 2013.

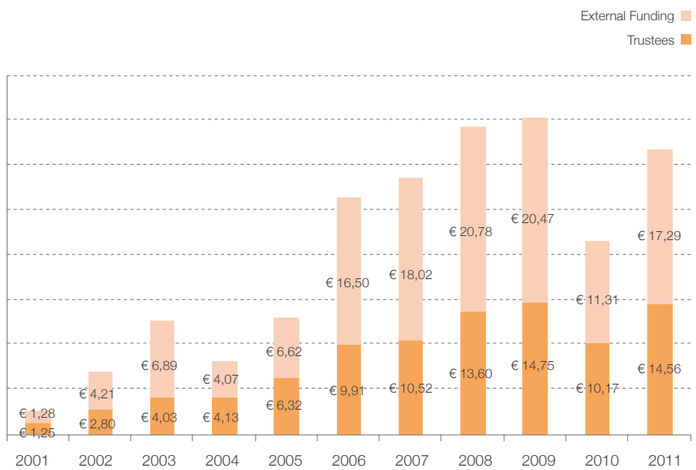
> Increase of the presence of CRG in social networks, especially Twitter, with notable success. Furthermore, the EU approved the project CommHERE, which deals with communication of scientific results in the area of health.

> The Technology Transfer department has obtained more than 500,000€ of additional funds for the assessment of research results and patents.

> The Research Office has continued to actively encourage and individually support the CRG community in applying for grants. This has lead to a very successful year in attracting external funds from highly prestigious funding agencies.

And more importantly, to adapt the team to the growth of the institute and to the new environmental circumstances, and also to increase efficiency, in the last quarter of 2011, the administrative management began working on a reorganisation plan of the Management & Research Support team that will end in 2012.

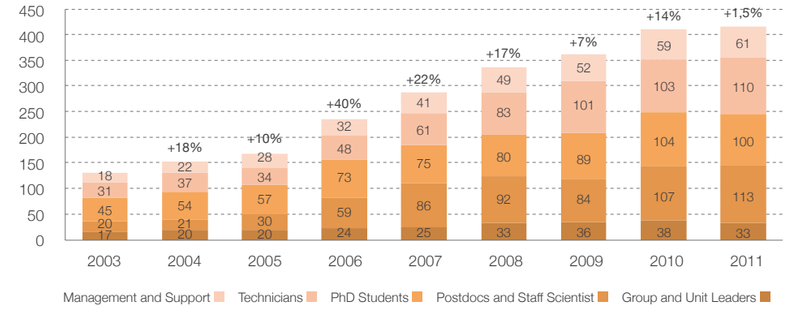
Funding evolution



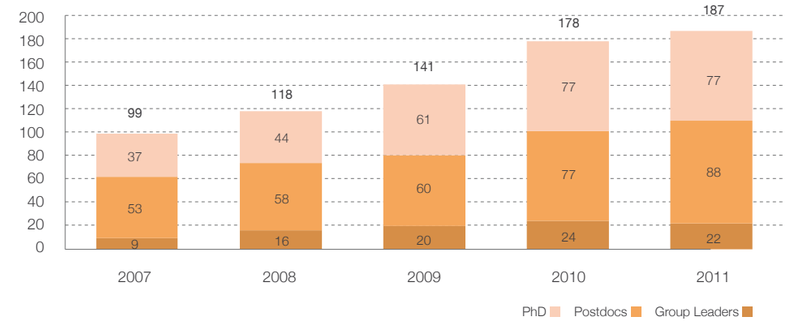
Personnel evolution

On 31 December 2011, 417 people, from 37 countries were employed by CRG.

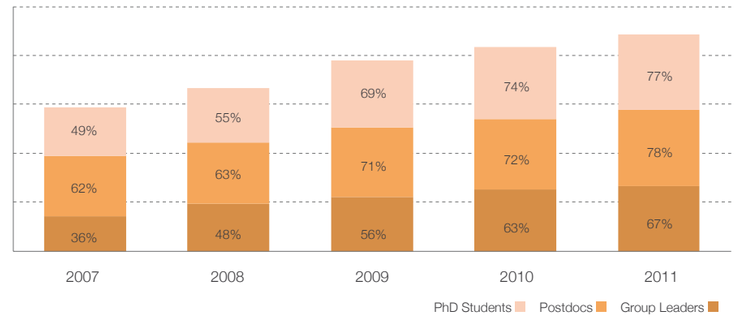
Personnel on 31 December 2011



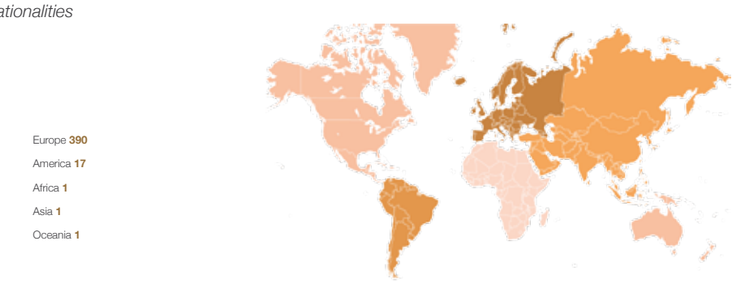
Foreign Researchers



Foreign Scientists Vs. Total

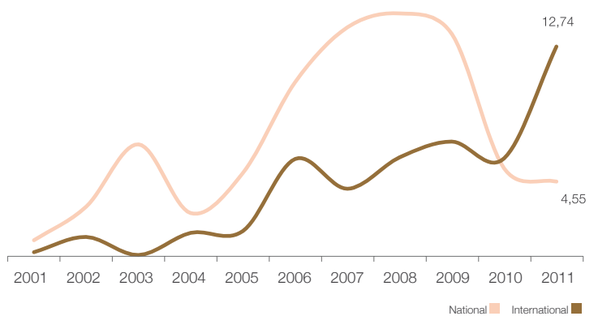


Nationalities



Grants & External Funding

Figure 1. Evolution of External Funding



In 2011 the CRG was very successful in attracting external funding from highly prestigious funding agencies, amounting to a total of 16.9M € (12.3M international, 4.6M € national) (Figure 1). The funds obtained by the European Commission (9.9M €/13 projects, not including projects in negotiation) represent the largest share, accounting for over 58% of the total external grants, an increase of 73% compared to the previous year. The CRG is one of the top research institutes in Spain attracting FP7 funds. With a total over 28M € and 48 projects awarded (including projects in negotiations and ERA-net projects) for the period 2007-2011, the CRG is regularly featured in the top 20, behind only the major universities, large companies and the Spanish research council. In terms of grants from the European Research Council (ERC), an indicator for international excellence, the CRG is currently the top institute in life sciences in Spain with a total of 8 grants (5 Starting Grants: Ben Lehner, Mark Isalan, Hernán López-Schier, Pia Cosma and Manuel Mendoza; and 3 Advanced Grants: Luis Serrano, Vivek Malhotra and Roderic Guigó), according to recent statistics provided by the ERC (Table 1).

Table 1. External Funding

	2011		2010	
Funding Agency	**€000**	**%**	**€000**	**%**
EUROPEAN COMMISSION	9,855.06	58.1%	3.612,31	40,1%
MINISTERIO DE ECONOMIA Y COMPETITIVIDAD	4,161.52	24.5%	4.535,51	50,3%
Howard Hughes Medical Institute	1,088.66	6.4%	0,00	0,0%
European Science Foundation	609.25	3.6%	0,00	0,0%
European Molecular Biology Organization	314.72	1.9%	150,34	1,7%
SWISS NATIONAL SCIENCE FOUNDATION	153.13	0.9%	0,00	0,0%
FUNDACION LILLY	140.00	0.8%	0,00	0,0%
AGAUR - AGENCIA GESTIO D'AJUTS UNIVERSITARIS	137.33	0.8%	76,77	0,9%
FONDS NATIONAL DE LA RECHERCHE LUXEMBOURG	103.24	0.6%	0,00	0,0%
FUNDACIÓN ALICIA KOPLOWITZ	75.00	0.4%	0,00	0,0%
Federation European Biochemical Societies. FEBS	72.00	0.4%	0,00	0,0%
Deutsche Forschungsgemeinschaft (DFG)	54.29	0.3%	0,00	0,0%
NOVARTIS FARMACEUTICA S.A.	42.00	0.2%	0,00	0,0%
MINISTERIO DE EDUCACION Y CIENCIA	50.05	0.3%	0,00	0,0%
HUMAN FRONTIER SCIENCE PROGRAM	41.82	0.2%	207,01	2,3%
FUNDACIO MARATO TV3	0.00	0.0%	125,33	1,4%
FONDATION JEROME LEJEUNE	0.00	0.0%	100,00	1,1%
Other	73.93	0.4%	207,19	2,3%
TOTAL FUNDING	**16,971.98**	**100.0%**	**9.014,46**	**100,0%**
EUROPEAN COMMISSION	2,401.59	n/a	425,00	n/a
MINISTERIO DE ECONOMIA Y COMPETITIVIDAD	0		337,96	
OTHER	516.16		335,35	
TOTAL FUNDING (IN NEGOTIATIONS)	**2,917.75**		**1.098,30**	

The new FP7 projects attracted by CRG researchers in 2011 include one Advanced Grant by the ERC awarded to R. Guigó, head of the Bioinformatics and Genomics Programme, five collaborative projects, one coordination action on science communication and five Marie Curie fellowships and projects[1]. Particularly remarkable are three new collaborative projects coordinated by young CRG PIs: a large collaborative project *4DCellFate* (L.DiCroce), a small or medium-scale collaborative project *BioPreDyn* (J.Jaeger) and an Initial Training Network *FliAct* (M.Louis). This year, the young CRG Group Leaders P.Carvalho and F.Kondrashov have been honoured with "International Early Career Scientists" awards by the Howard Hughes Medical Institute (HHMI) with a total budget of US$1.3 m distributed over 5 years. This is a tremendous achievement for the CRG in the first edition of this programme, considering that 760 researchers from 18 countries applied, and only 28 were selected, 5 of which were from Spanish research institutes (Table 2).

Table 2. The Most Relevant Competitive Grants

PROJECT TITLE	PROJECT TYPE	PI	ROLE	CRG FUNDING (€)
Developing a global understanding of the PRC and NuRD complexes in stem cell differentiation, in health and disease (4D-CellFate)	Collaborative Projects (Large-scale integrating project)	Di Croce, Luciano/ Lehner, Ben	Coordinator/ Partner	2.711,34
Mechanism of Unconventional Protein Secretion (MUPS)	ERC Advanced Grants (AdG)	Malhotra, Vivek	Individual	2.206,96
Chromosome segregation and aneuploidy (Longchrom)	ERC Starting Grants (StG)	Mendoza, Manuel	Individual	1.058,61
A BLUEPRINT of Haematopoietic Epigenomes (BLUEPRINT)	Collaborative Projects (Large-scale integrating project)	Graf, Thomas; Guigó, Roderic; Estivill, Xavier	Partner	1.022,00
Protein interaction machines in oncogenic EGF receptor signalling (PRIMES)	Collaborative Projects (Large-scale integrating project)	Serrano, Luis	Partner	833,44
Systems neuroscience of *Drosophila*: from genes to circuits to behaviour (FLiACT)	Marie Curie Initial Training Networks	Louis, Matthieu	Coordinator	572,56
Evolutionary genomics and genetics	HHMI-New International Competition	Kondrashov, Fyodor	Individual	544,33
Molecular mechanisms underlying ER homeostasis	HHMI-New International Competition	Carvalho, Pedro	Individual	544,33
From Data to Models: New Bioinformatics Methods and Tools for Data-Driven Predictive Dynamic Modelling (BioPreDyn)	Collaborative Projects (Small or medium-scale focused research project)	Jaeger, Johannes	Coordinator	477,99
Studying Physiology and Pathology of Imprinted genes to understand the role of Epigenetic Mutations in Human Disease (INGENIUM)	Marie Curie Initial Training Networks	Cosma, Maria Pia	Partner	238,10
TOTAL FUNDING				**10.209,67**
Uncovering and understanding RNA through Massively Parallel Sequencing (RNA-MAPS)	ERC Advanced Grants (AdG)	Guigó, Roderic	Individual	2.056,64
Effective Technology Transfer in Biotechnology (ETTBio)	INTERREG IVC	Rúbies, Xavier	Partner	184,86
Novel TNF inhibitors as therapeutic agents in autoimmune diseases, osteoporosis and cancer (DIMER-ME)	Assessment Projects (Prova'T)	Serrano, Luis	Individual	157,80
Identification and manipulation of molecular pathways relevant for age-dependent tissue regeneration	AXA Research Projects	Cosma, Maria Pia	Partner	173,50
Characterisation of the role of DNA methylation, hydroxymethylation and TET proteins in progesterone-mediated signalling in breast cancer cells (Hydroxy-Methylation)	Marie Curie Intra-European Fellowships	Verde, Gaetano	Individual	176,05
Synthetic gene regulatory networks for single-stripe gene expression (SynthStripe)	Marie Curie Intra-European Fellowships	Schaerli, Yolanda	Individual	168,90
TOTAL FUNDING (IN NEGOTIATIONS)				**2.917,75**

1. The new projects outlined in the text are reported in the official figures for external funding in 2011, except for the ERC AdG awarded to R. Guigó and two Marie Curie fellowships, which were in negotiation at the end of the year. They are included in the table "Total funding (in negotiations)" and will be reported in the official figures for external funding next year. Similarly, two ERC grants (awarded to V. Malhotra and M. Mendoza) and two Marie Curie fellowships that were awarded in 2010 but officially started in 2011, were in negotiation at the close of 2010, and are, therefore, now reported in the official figures for 2011.

In addition to attracting external grants for excellent basic research and fellowships, the CRG has also succeeded in obtaining additional external funding for strategic CRG activities amounting to 0.3M €, including technology transfer activities, as well as outreach and scientific events (Table 3).

Table 3. Other External Funding	2011		2010	
Entity	**€000**	**%**	**€000**	**%**
Sponsorship	33,14	10,3%	107,81	9,0%
Technology Transfer Activities	256,69	79,8%	1.066,98	89,4%
Other	31,89	9,9%	19,22	1,6%
TOTAL FUNDING	**321,72**	**100,0%**	**1.194,01**	**100,0%**

International & Scientific Affairs

After several years consolidating its research activities of excellence, the CRG is now accepting the challenge of exploring new collaborations on the national, European and International levels. The new collaborations aim to cross disciplines and sectors, as well as borders and cultures to strengthen the international and interdisciplinary perspective and impact of the institute.

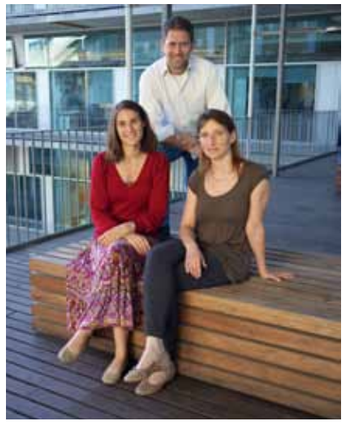
In 2011, we began to set up a plan and strategy to build bridges with the medical and clinical sector in Spain. A first alliance was signed with the Spanish Federation of Rare Diseases (FEDER) for collaboration on raising awareness of rare diseases, and the importance of research in this disease area. Furthermore, workshops with local hospitals have been planned to bring together CRG scientists and clinical researchers to exchange, and eventually integrate, their knowledge and expertise in biomedical research.

At a European level, since 2006 the CRG has been engaged in a partnership with European Molecular Biology Laboratory (EMBL), which supports the EMBL-CRG Systems Biology Programme to advance the understanding of complex biological systems.

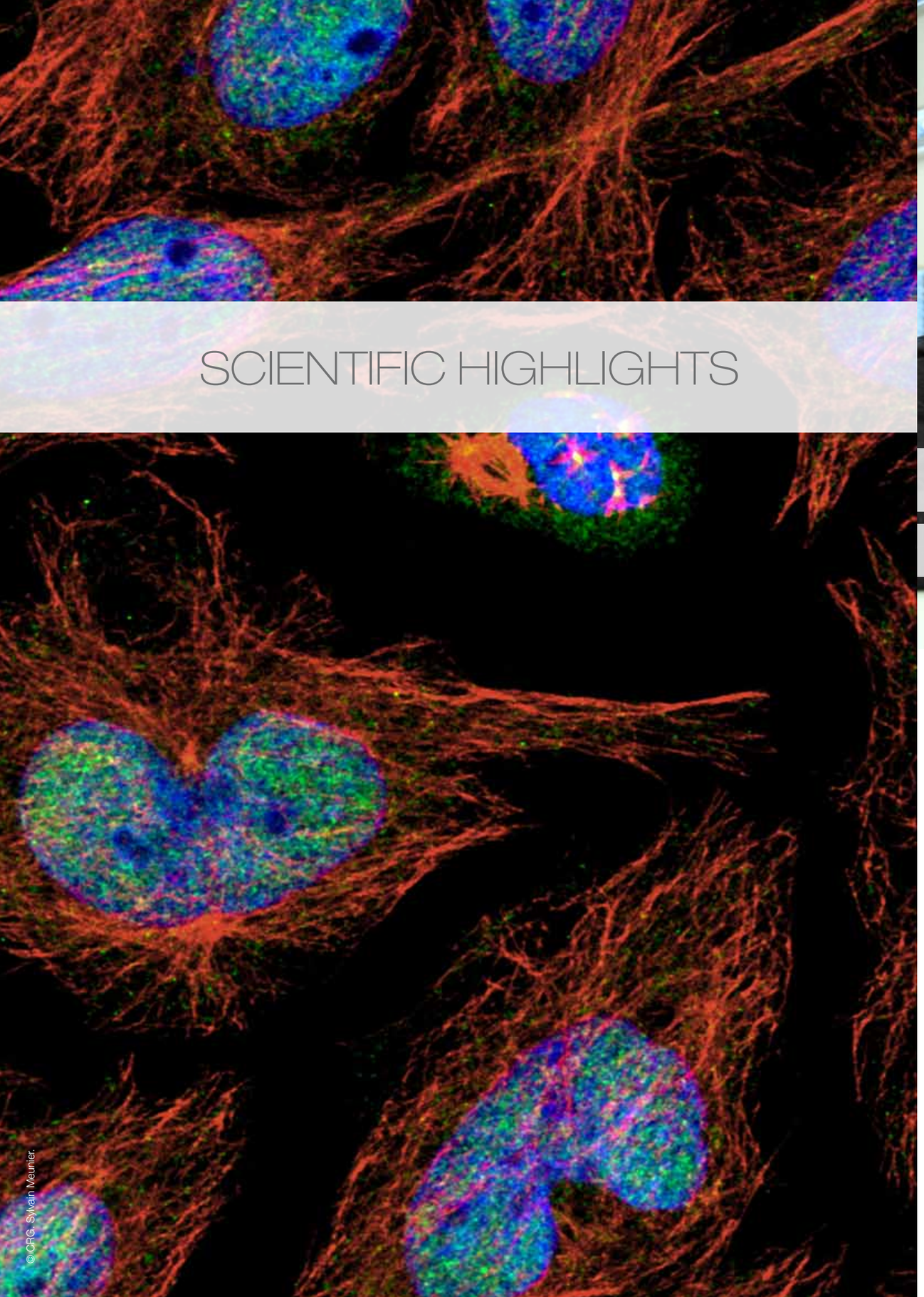
In the international arena, the CRG has initiated a dialogue with the National Center for Biological Sciences in Bangalore and the Genome Institute of Singapore (GIS) to explore common ground for scientific collaboration.

Coordination of collaborative scientific projects

The CRG plays an important role in leading several Spanish and European collaborative projects to advance knowledge and collaboration in several fields of biology and biomedical research, ranging from systems biology, cancer, epigenetics, cellular trafficking, and rare diseases to medical genomics. The current portfolio includes four Spanish Consolider projects and six European projects, three of which were awarded in 2011 (see section on “External funding”).



Members of the International & Scientific Affairs office.



SCIENTIFIC HIGHLIGHTS

SCIENTIFIC HIGHLIGHTS

Gene Regulation

READING A GENE IN A MINUTE

Researchers from the Chromatin and Gene Expression Laboratory at the CRG describe the mechanism that is triggered during the first minute of gene activation, key to developing new strategies to combat cancer.

In the early 60s, the German Peter Karlson identified, for the first time, that steroid hormones in insects act at chromosome level. These hormones regulate the activity of almost all the cells of our bodies. They do so via their ability to activate or repress genes. Despite the time that has elapsed, still little is known about the process. Miguel Beato collaborated with Karlson, and since then has been attempting to unravel the mystery.

Now, researchers from the Chromatin and Gene Expression Laboratory, which Beato directs at the CRG, have thrown a little more light on this question. In a study published in the journal *Genes & Development*, and also featured in *Science*, they have described the mechanism that allows access to the information found in the DNA compacted into chromatin -the ensemble of DNA, histones and other proteins located in the nucleus of eukaryotic cells. The initial steps of the mechanism are triggered during the first minute of gene activation. The results will help in the discovery of new strategies to control the development of hormone-dependent cancers, such as breast or prostate cancer.

The steroid hormone crosses the cell membrane and binds to its corresponding receptor changing its conformation. Immediately afterwards they travel together to the cell nucleus where they look for specific sequences in the DNA, discovered in the 80s by Miguel Beato, with which they interact. These hormone receptor/DNA interactions mark the genes that must be activated or inhibited resulting in the development of hormone-dependent cancers.

“There are four enzyme complexes involved in the first step of the chromatin opening”, explains Beato.

These complexes are known as NURF, ASCOM, KDM5B and CDK2. What they do is displace the linker histone H1 from DNA. Histones are the proteins on which the DNA is wound to stay compact, forming a more complex structure known as chromatin. When the enzyme complexes displace the histone H1, they have access to the information contained in the gene.

Guillermo Vicent, one of the authors, draws acomparison of the scene with a plate of spaghetti. “The plate would be the cell nucleus, the spaghetti, the chromatin. In this context, the fork, which would be the receptor joined to the hormone, stabs at the plate, taking with it a portion of the spaghetti. The fork signals the area within the whole network of chromatin where the enzyme complexes must act.” The function of NURF, ASCOM, KDM5B and CDK2 is to release this network of pasta which is wrapped around the fork in order to read the information it contains, that is, the information from the gene that it has “stabbed”. And all this happens in a single minute.

The study by the Chromatin and Gene Expression laboratory is relevant on a global level. Beato reveals: “For the first time we are seeing the molecular steps in the process of chromatin opening which Karlson observed more than 50 years ago. None of the molecules were previously unknown, but until now nobody knew what role they played in that first minute.” The complexity of the mechanisms required to activate genes was somewhat unexpected.

With these results on the table, we are a little closer to understanding how some tumours develop and, eventually, to designing strategies to help control them. There are a great number of cancers that are regulated by steroid hormones, such as breast cancer, which is, primarily, what this CRG group looks at.

Breast cancer cells are sensitive to progesterone and oestrogen, hormones involved in the menstrual cycle, pregnancy and embryo development, which also control breast cancer cell proliferation. If the receptors are blocked, the tumours stop growing, but after some time they become resistant to the hormones; even so, they still need the receptors. “If we understood how the receptors work”, points out Beato, “we could look for drug combinations that block the growth of the breast cancer cells without harming other cells, something which currently happens in treatments like chemotherapy.”

This, for the director of the Chromatin and Gene Expression laboratory, is the medicine of the future. “It does not entail the treatment of a disease with high doses of a drug, but rather low dose combinations of three or four drugs directed at different steps of the altered control process, as is done, for example, in the treatment of AIDS. This would be personalised medicine, adjusted for different types of cancer.” That is, for certain types of breast cancer one combination of drugs would be used, for others another, and so on for the different types of prostate and lung cancer. In this way, the overall negative consequences for the body, like those of current non-selective chemotherapy, would be avoided.

Reference work:
Vicent GP, Nacht AS, Font-Mateu J, Castellano G, Gaveglia L, Ballaré C, Beato M.
“Four enzymes cooperate to displace histone H1 during the first minute of hormonal gene activation.”
Genes Dev, 25(8):845-62 (2011).

REVEALING THE CELL’S BIOLOGICAL CLOCK

Researchers from the Epithelial Homeostasis and Cancer laboratory at the CRG reveal new discoveries that help us understand how skin stem cells regenerate

We do not even realise, but the truth is that our body loses millions of cells every day. Because they wear out and lose their functionality. Because they age. Because they get damaged. There are many reasons. Normally, this is not a problem because there is an equilibrium between cells that are lost and those that are regenerated (homeostasis). This is the responsibility of the stem cells. But in abnormal conditions, everything gets out of control, cells can become invasive and end up causing a tumour.

In any case, how do know they have to regenerate? How do they communicate with their environment? When do they realise that there is enough energy within the body to perform a function? These are some of the questions being asked in the Epithelial Homeostasis and Cancer lab at the CRG, focused on understanding how skin tissues are maintained throughout the life of a person.

Recently, this team of researchers, led by Salvador Aznar-Benitah, has conducted two studies of global significance that help us to understand some aspects of this process.

The first of these studies, published in the journal *Cell Stem Cell*, identified a protein known as Cbx4 as being essential for the regulation of skin stem cells. When the tissue does not need to regenerate, these cells remain in a quiescent state, that is, they are dormant. At a specific moment, they receive the signal that they have to activate their metabolism and they begin to divide and differentiate into functional cells to regenerate the tissue. Thus, it is not them that specialise; they remain under the skin to provide the necessary cells.

"What we found", explains Aznar-Benitah, ICREA Researcher, "is that the protein Cbx4 is essential to maintain the stem cells in a dormant and unspecialised state. If we remove this protein, the stem cells undergo immediate spontaneous differentiation." Excessive proliferation of these cells, or their early differentiation, is detrimental to the renewal of the tissue, as this causes ageing or the development of pathologies such as cancer.

The second study of the CRG's Epithelial Homeostasis and Cancer laboratory, published in the journal *Nature*, has to do with the role that the circadian rhythm plays in the regenerative capacity of skin stem cells mentioned above.

The skin is exposed to various harmful agents throughout the day, especially ultraviolet light during daylight hours, and pathogens such as bacteria and viruses. Its goal, therefore, is to protect the body from this possible harm. The stem cells need to know when the best time to regenerate the tissue is.

Why? What would happen if a stem cell divided by day, for example? Aznar-Benitah gives us the answer: "The UV light is very mutagenic and is one of the main causes of skin cancer. If the cell divides at this time, there is a high risk of passing on mutations to their daughter cells, and so on successively. By doing this the problem would be perpetuated."

What these researchers have discovered is that there is an internal biological clock that allows the cell to know when it is day and when it is night. It is as if it says that during the day it is better not to divide because the ultraviolet light is going to have an effect, and moreover, the body does not have enough energy for the division and differentiation. As night falls, it informs the cell that it can activate the mechanism.

"Hence, the cell proliferation of a tissue, at least one such as the skin, occurs chiefly in the evening/at night. If we look at proliferation in the morning, we see something, but very little, because it is later in the day when there is less risk of accumulating mutations and when there is more energy."

The experiments that the Epithelial Homeostasis and Cancer laboratory at the CRG have undertaken on mice demonstrate that the accuracy of this biological clock tends to decrease progressively as we age, but also when we change our sleeping habits. This is what happens on a small scale when we cross the Atlantic by plane, for example, and it takes a couple of days to recover because we suffer jetlag. Taken to the extreme, for people who constantly have to change their work shifts, the result is not only ageing of the tissues, but also a greater propensity to develop tumours.

For now, the researchers cannot explain why the biological clock ages. Looking ahead, the goal is to find ways of staying young. And also, according to Aznar-Benitah, restoring a normal circadian rhythm to cancer cells.

Reference works:

Luis NM, Morey L, Mejetta S, Pascual G, Janich P, Kuebler B, Cozutto L, Roma G, Nascimento E, Frye M, Di Croce L, Benitah SA.
"Regulation of human epidermal stem cell proliferation and senescence requires polycomb- dependent and -independent functions of Cbx4."
Cell Stem Cell, 9(3):233-46 (2011). Erratum in: Cell Stem Cell, 9(5):486, Cozutto, Luca [added] (2011).

Janich P, Pascual G, Merlos-Suárez A, Batlle E, Ripperger J, Albrecht U, Cheng HY, Obrietan K, Di Croce L, Benitah SA.
"The circadian molecular clock creates epidermal stem cell heterogeneity."
Nature, 480(7376):209-14 (2011).

IN THE WAITING ROOM FOR PARKINSON'S

It is a little past eight in the morning. Rush hour in Barcelona. Thousands of cars from other towns flood into the Catalan capital while others circulate around its veins and arteries. Controlling and managing this circulatory tangle are the traffic lights, signs and police.

Our brains, to a large extent, resemble the transport network of a large city. The vehicles that circulate via complex neuronal electrical circuits are the genes, which are regulated by different types of signals; some are more active, like the traffic lights and cops; others are more passive, like the signposts; and there are even more ambiguous ones that can end up triggering different results, and even accidents!

In the organism, these signals responsible for coordinating the circulation of the genes are the microRNA, small molecules of RNA which carry out specific functions: "Some are hindrances to the expression of certain genes whilst others are accelerators", explains Xavier Estivill, researcher and coordinator of the Genetic causes of Diseases group at the Centre for Genomic Regulation (CRG) in Barcelona.

As happens with traffic, when one of these regulators is not clear, is misleading or breaks down, imbalance is produced in the brain which may provoke neurodegenerative diseases. This is the case in Parkinson's, the second most frequent neurodegenerative disease, affecting between 1-2% of the population and which usually develops after the age of 60.

Although it is still unknown why, in people with Parkinson's the nerve cells in specific areas of the brain start to die quickly and progressively, such as those of the *substantia nigra* and certain areas of the cortex involved in motor control. The Genes and Disease group of the CRG, led by Xavier Estivill, has been able to identify, for the first time, various microRNA that could act as markers of the early onset of the disease.

Early markers

Over three years the researchers studied the brain of people who, autopsy revealed, were in the early stages of Parkinson's disease. They analysed the regions affected by the disease to check which genes were unregulated, given that they thought they could function as a kind of warning system. "We traced a series of microRNA, which is found in excess in some cases and lacking in others. These small RNA molecules are altered in a prior stage of the disease, so they act as markers of early onset", says Estivill.

The microRNA are responsible for regulating the expression of many other genes, either by sending them signals to degrade and making them disappear, or halting their specific expression. What the research team observed was that when some of these molecules are over or under expressed, this has consequences for other genes, some of which were unknown. "What we found was that it is these changes and alterations that are in the waiting room for Parkinson's", stresses Estivill.

Why these molecules deregulate remains a mystery for the moment. Throughout our life there are many external factors that produce changes in the expression of our genes. For example, the simple fact of being tense before going out to sing in a concert, in the case of musicians, provokes the activation of a series of molecules and neurotransmitters which have an effect on the ecology of the cells of a determined tissue of the body. In the case of Parkinson as well as in Alzheimer's, some factors are known which can be protective and others which can trigger the disease.

Surely, these alterations in the microRNA are also involved in some way in the ageing process of our brains. This study, published in *Human Molecular Genetics*, could lead the way for the early detection of Parkinson's and possible therapies which could act on the expression blocking action of these microRNA.

Reference work:

Miñones-Moyano E, Porta S, Escaramís G, Rabionet R, Iraola S, Kagerbauer B, Espinosa-Parrilla Y, Ferrer I, Estivill X, Martí E.
"MicroRNA profiling of Parkinson's disease brains identifies early downregulation of miR-34b/c which modulate mitochondrial function."
Hum Mol Genet, 20(15):3067-78 (2011).

SCIENTIFIC HIGHLIGHTS

Bioinformatics and Genomics

DECIPHERING THE SECRET LIFE OF NON-CODING RNA

The Gene Function and Evolution group from the Centre for Genomic Regulation presents catRAPID, an application to improve the understanding of gene regulatory networks

It is said that in the eighteenth century, Queen Maria Carolina had one of the most typical homemade dishes from southern Italy banned as she considered it too common. It was a round bread, cooked in the oven and on which other ingredients, such as cheese and tomatoes, were put. King Fernando I adored this dish, so much so that, whenever he could, he disguised himself and ran secretly to a humble neighborhood of Naples where they made it. After a while, the king finally confessed his culinary fascination, which led to the popularisation of pizza throughout Italy.

Today, it is surely one of the most popular dishes around the world. With a thinner or thicker base, softer or crispier, with or without tomato, cheese, salmon, vegetables, meat, there are pizzas for everyone. "If we were to draw an analogy," declares Gian G. Tartaglia, head of the Gene Function and Evolution Group at the Centre for Genomic Regulation, "pizza would symbolize the relationship between RNA and proteins very well."

And he explains: "Human beings are all extremely similar, as are pizza bases. The difference between one and another lies in the "toppings", in the ingredients that we add and which make our pizza, or us ourselves, unique and unrepeatable." But it is not only the ingredients themselves which are important, but also the quantities we add, where we put them, the thickness, how they are cut, their quality and a myriad of small and invisible details that end up making our dish unique.

RNA does the same in our bodies. Traditionally, it has been considered to be the cook responsible for reading and following the recipe written in the DNA. Paradoxically, however, we know few of the functions it performs. It is known that approximately 1% of the RNA contains valuable and useful information for

the synthesis of proteins, the ingredients of the pizza. The rest are, largely, an enormous number of molecules, the long non-coding RNAs, for which the function is unknown. For a long time, it was believed that this 99% was simply trash, waste, evolutionary debris that had remained but which had no function whatsoever.

For Gian Gaetano Tartaglia, doctor of biochemistry and proteomics expert at the head of the Gene Function and Evolution group at the Centre for Genomic Regulation, that idea was it hard to swallow. “It seemed impossible that there were so many molecules, so highly concentrated, so rich in potential information and function, which were simply waste.”

And if we count the number of proteins in humans, mice, worms or a simple fly, we see that the number is more or less the same. These molecules are essential for the survival of an organism, as they are responsible for carrying out a large number of functions. However, despite having similar quantities, human beings are able to perform complex tasks that differentiate us and set us apart from other organisms. “We have evidence that a large part of non-coding RNA interacts with proteins to orchestrate a series of processes within the cell”, notes Tartaglia, who, together with his team, has designed a new method for the prediction of large scale interactions between these non-coding RNA molecules and proteins.

The method is catRAPID, a computational algorithm for making predictions about the interactions between RNA and proteins, based on the physicochemical properties of these molecules as well as identifying areas where they join.

“We started working on this two years ago and it is, at the moment, unique in the world”, says Tartaglia proudly. The researchers have validated their predictions on a large number of protein-protein associations and non-coding RNA molecules *in vitro*. They have, for example, already studied several human genes involved in biological processes such as embryogenesis, cell differentiation in skin and the development of breast cancer.

The purpose of this software, which is a free tool available through the CRG website, is to try to better understand the role of non-coding RNA in the synthesis of proteins involved in gene regulatory networks through experiments and approximations.

If we were to continue with the example of pizza with which we began this article, the CRG Gene Function and Evolution group, says Tartaglia, would be the pizza chef, responsible for assessing whether or not the pizza is good, if the ingredients are right, how much to put on each one, the necessary cooking time, which combinations work and so on. The knowledge of the chef, after years of experience, together with the tests for the creation of new pizzas or evaluation of existing ones, would be the catRAPID method. The researchers have presented the results obtained with this algorithm in the journal *Nature Methods*.

Work reference:
Bellucci M, Agostini F, Masin M, Tartaglia GG.
“Predicting protein associations with long noncoding RNAs.”
Nat Methods, 8(6):444-5 (2011).

SCIENTIFIC HIGHLIGHTS

Cell and Developmental Biology

TAILOR-MADE CHROMOSOMES

The Cytokinesis and Chromosome Segregation group discover how cells measure the length of chromosomes during division

Luc Montagnier recounts that back in the 80’s, when he discovered the AIDS virus together with Françoise Barré-Sinoussi, his laboratory shelves were stuffed full of books. All of them described experiments, but few gave answers. “The history of science is replete with experiments that didn’t work. Once in a while one does work, and it’s that which keeps you motivated”, Manuel Mendoza assures us.

The head of the Cytokinesis and Chromosome Segregation group at the CRG can boast of doing just this. His researchers have described, for the first time, how cells adjust chromosome compaction during cell division.

Think about a car. It is designed in relation to the size of the driver. If the person driving it was a giant, they would not fit. This does not mean that the car does not work on its own, right? Well, the same problem of scale is true in the cell. Living things grow thanks to cell division, the process by which cells duplicate their genetic material (or DNA) and divide it between two daughter cells identical to the original. This genetic material is carried by the chromosomes. In the case of humans, if we could stretch out our chromosomes, each would measure approximately eight centimetres, a disproportionate length with respect to a cell. How then, can they fit inside?

“This is the phenomenon known as mitotic chromosome condensation”, explains Mendoza. When a cell divides, its division apparatus acts like a guillotine, cutting everything in the middle. “The chromosomes have to migrate to the poles of the cell, in equal numbers, before the guillotine falls. And to do this they must be sufficiently compact.” There is a lot of research into the mechanisms of chromosome compaction, but never before had anyone wondered how the cell knows the exact size of the chromosomes and how much they need to be shortened.

Mendoza began to unravel the mystery five years ago, when he was a postdoc at the ETH in Zurich. Along with Yves Barral, he conducted experiments over two years with budding yeast. The aim was to alter the size of the chromosomes to see if the cell realised and adapted or, alternatively, died. To do this, they chose the two longest chromosomes and joined them with a kind of molecular glue. The result was that the cell had one chromosome less, but one which was twice as long; despite this, it divided successfully. What does this mean? That the yeast cell has a mechanism for sensing the length of each chromosome. If one chromosome is longer, it is compacted more during division, and if it is shorter, it is compacted less.

Mendoza moved to the CRG, and the task of finding out how this mechanism works continued in parallel with the laboratory in Zurich. Gabriel Neurohr, from the Cytokinesis and Chromosome Segregation group, performed the majority of the experiments leading to the discovery of the importance of a protein found in the centre of the cell during division: Aurora B.

Aurora B phosphorylates other proteins, *i.e.*, it adds phosphate molecules to regulate their activity, and marks them. And one of the proteins that it phosphorylates is the histone H3, which appears to be responsible for chromosome condensation. “What we did”, explains Mendoza, “was inactivate each one of the proteins (Aurora B and histone H3) with a mutation, to see if that resulted in the chromosomes not being compacted well. The response was positive in both cases.”

To better understand the function of Aurora B, imagine a typical scene in the Japanese subway at rush hour. There are employees whose job is to squash the passengers into the carriage before the train sets off again. If they didn’t, some might get stuck in the door, with some bits of their body outside the train and others inside. In the same way, this protein ensures that the chromosomes are pushed in and packaged before being cut into fragments.

The results of the CRG Cytokinesis and Chromosome Segregation group could explain why some cancer cells are able to proliferate so successfully. Often, the chromosomes of the cancer cells mutate and reorganise, joining together with other chromosomes. The consequence is that they are longer than normal, just as happened when the two yeast chromosomes were stuck together with molecular “glue”.

“You might think that these cells would die because they cannot segregate such a long chromosome. But we know that in cancer, the opposite happens: the cells divide like crazy. The explanation could be that all cells, including cancer cells, have the ability to coordinate the size of the chromosomes during division.” Interrupting this mechanism could be the key to developing new cancer therapies.

Reference work:

Neurohr G, Naegeli A, Titos I, Theler D, Greber B, Díez J, Gabaldón T, Mendoza M, Barral Y.
“A midzone-based ruler adjusts chromosome compaction to anaphase spindle length.”
Science, 332(6028):465-8 (2011).

SCIENTIFIC HIGHLIGHTS

Systems Biology

GENETICS ROLLS THE DICE

More than one coach would give their right arm to sign Messi in his team. Or at least, to have a Messi clone. But would a Messi clone necessarily be as good at football as Messi himself? Even if we cloned the genome of this Argentine genius and, moreover, reproduced the environment in which he grew up?

“Genes and environment are not sufficient to determine the characteristics of an individual. There is a third factor which is rarely considered, but which is likely to contribute to what makes us unique and unrepeatable”, explains Ben Lehner, ICREA Research Professor at the head of the Genetic Systems research group at the CRG, who likes to start his talks and conference presentations with the example of this footballer. The factor that he’s talking about is chance.

Since arriving in the CRG in 2006, Lehner and his team have been trying to understand the principles and evolution of genetic systems, as well as the role stochasticity or chance plays in disease processes. This they do through systems biology, which links experiments with predictions via gene networks. This year, his research and findings, published in *Nature*, *Science*, and *Nature Genetics*, have earned him the “City of Barcelona Award”.

Genetics and chance

One of the main challenges in genetics is to predict phenotypic variations from the genome sequence of an individual and in this way know, for example, if the subject will suffer a certain disease, or react positively to a particular treatment. In order to answer this question, his lab has conducted a study on yeast. “This study is important because we have demonstrated that if we know enough about the function of each gene, then we can make useful predictions about phenotypic differences between individuals using a simple genetic model”, explains Lehner.

For now, it is possible to make these predictions using model organisms. In the case of humans, however, this is very difficult to achieve, as the majority of genes affecting determined phenotypes and disease are not known. This is currently one of the biggest hurdles in the development of predictive and personalised medicine: it is necessary to understand why, in the presence of certain gene mutations, some people develop a disease while others do not.

The CRG researchers carried out a study using the worm *C. elegans* as a model. Given that genome and environment are not sufficient to predict whether a mutation will affect an individual, they developed a method to measure differences in gene expression between individuals, essential for understanding which genes are relevant for each mutation.

"You might think that the difference between the phenotype of two people with the same mutation must be due to environmental differences, or to other mutations present in their genomes. But chance also plays a vital role in explaining why individuals with the same mutation develop a certain disease or not, and why there are mutations that kill some individuals and not others. It is therefore essential to take this into account when making predictions", maintains Lehner.

A question of survival

One of the contributions to differences between individuals is the chaperone content of each individual. These proteins assist the folding of other proteins, a process vital for their correct functioning. The exposure to environmental stress causes proteins to misfold and chaperones are essential for ensuring the survival of individuals. It is also the case that for some mutations, the genetic change generates unstable, badly folded proteins, which can lead to problems in the cell. If there are enough chaperones around, these proteins may remain folded, preventing the mutation from being expressed. A little bit of stress can therefore sometimes protect individuals from the effects of mutations. Moreover, by chance, the levels of chaperones vary from one individual to the next, even if they are genetically identical and are kept under the same environmental conditions. Those "lucky" individuals that happen to express more chaperones are better protected from the harmful effects of mutations.

But if chaperone levels protect us from mutations, stress, and help to guarantee survival, why, then, do organisms not all have high levels of these proteins? The researchers found that in normal, non-stressful conditions, those *C. elegans* worms with more chaperones develop more slowly than the others and reach reproductive maturity later, which is a disadvantage for them. In contrast, those with fewer chaperones breed faster and are therefore more competitive.

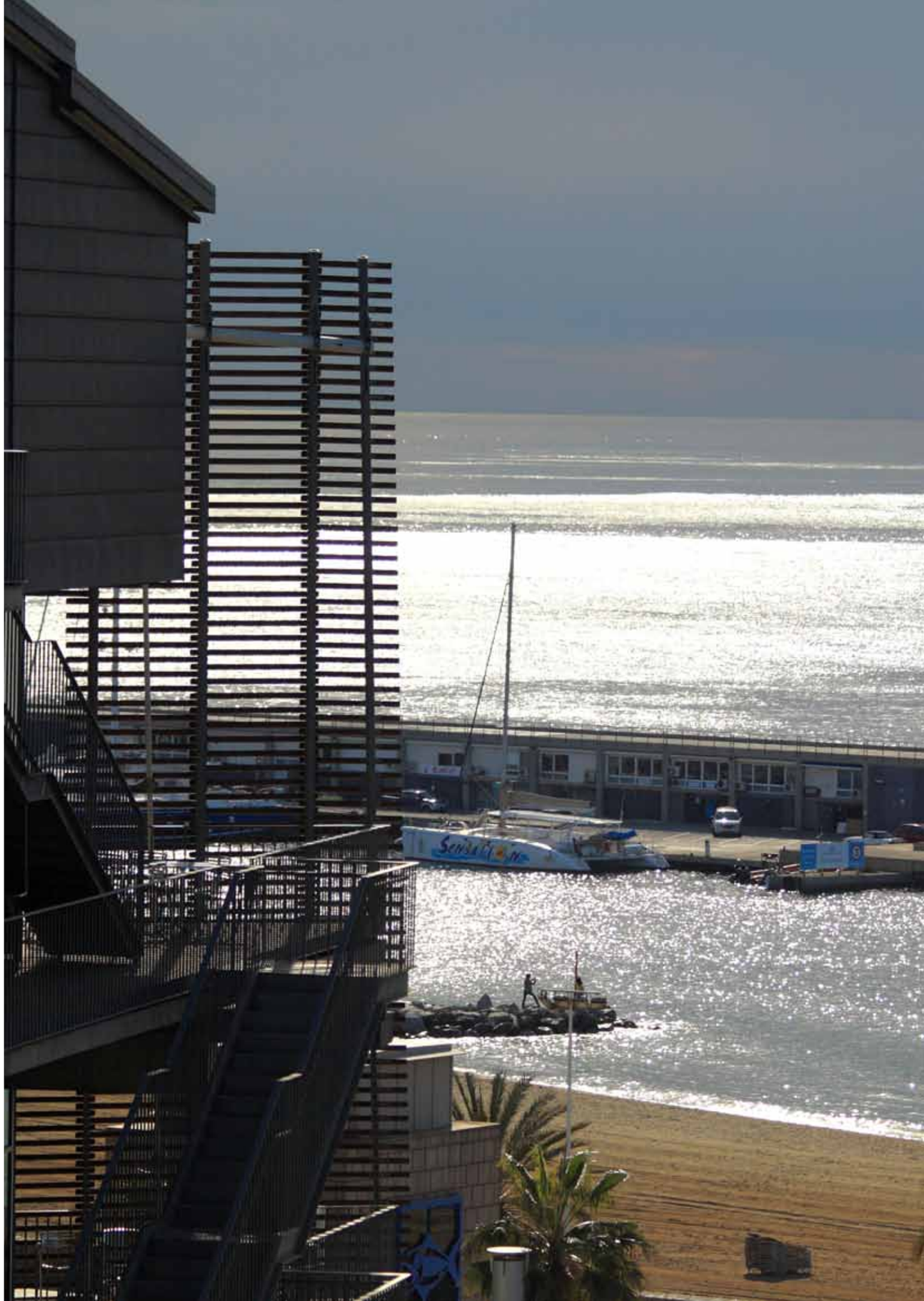
Maybe, think the CRG researchers, this variation in chaperone levels between organisms is an evolutionary tool to create phenotypic diversity from a single genome sequence to guarantee that some of these individuals survive, irrespective of the environmental conditions. Therefore, some individuals having more chaperones than others is a matter of chance. And sometimes genetics, says Lehner, rolls the dice.

Reference works:

Jelier R, Semple JI, Garcia-Verdugo R, Lehner B.
"Predicting phenotypic variation in yeast from individual genome sequences."
Nat Genet, 43(12):1270-4 (2011).

Burga A, Casanueva MO, Lehner B.
"Predicting mutation outcome from early stochastic variation in genetic interaction partners."
Nature, 480(7376):250-3 (2011).

Casanueva MO, Burga A, Lehner B.
"Fitness trade-offs and environmentally induced mutation buffering in isogenic C. elegans."
Science, Epub 2011 Dec 15, doi: 10.1126/science.1213491



SCIENTIFIC REPORT

GENE REGULATION

Group: Chromatin and Gene Expression

Group Structure:

Group Leader: Miguel Beato

Postdoctoral Fellows: Alessandra Ciociola, François Le Dily, Gaetano Verde (since March), Guillermo Vicent (Staff Scientist), Roni Wright, Roser Zaurin (since September)

PhD Students: Laura Gaveglia, Andy Pohl, Diana Reyes, Michael Wierer

Technician/s: Giancarlo Castellano (Consolider), Jofre Font, A. Silvina Nacht

Visitors: Alejandro La Greca, Buenos Aires (untill March)

SUMMARY

The group is interested in understanding how eukaryotic cells respond to external signals, in particular how different signals are transduced to the nucleus to modulate gene expression. Steroid hormones signal to chromatin not only directly via binding of their receptors to DNA, but also indirectly via crosstalk with kinase signalling pathways. Using a model reporter gene and genome wide studies we are trying to unravel how these kinases impinge on chromatin structure and dynamics via functional interactions with chromatin remodeling complexes. In particular, we are studying how the topological information that determines the position of nucleosomes and the general organization of chromatin in the cell nucleus modulates the hormonal response. We also try to use this knowledge to decipher the role of steroid hormones in breast and endometrial cancer cell proliferation and differentiation.

RESEARCH PROJECTS

1. Global Analysis of Chromatin Structure and Dynamics during Hormonal Gene Regulation

L. Gaveglia, R. Wright, F. Le Dily, A. Pohl, G. Castellano

We have performed genome wide nucleosome mapping by massive sequencing of MNase treated chromatin from T47D-MTVL breast cancer cells untreated or treated with hormone for 60 minutes. The results revealed that the nucleosome occupancy changes significantly after hormone treatment, which also generates a large number of new DNase I hypersensitive sites (Gaveglia *et al*, unpublished). In ChIP-seq experiments with an antibody to Progesterone Receptor (PR) that we analyzed with a new software (Althammer *et al*, 2011), we discovered a preferential location of PR binding sites over nucleosomes, which are remodeled after hormone treatment (Ballaré *et al*, unpublished).

We are also analyzing the 3D-structure of the chromosomes and its changes upon hormone treatment using 3C-related techniques and high-resolution microscopy with the aim to establish the spatial relationship between subpopulations of genes regulated by hormones in specific ways (Le Dily *et al*, unpublished).

2. Signalling by Progesterone to Chromatin via Kinase Cascades

A. Ciociola, D. Reyes, M. Wierer, R. Wright

Progesterone controls proliferation and gene expression in breast cancer cells *via* transient activation of kinase signalling pathways, including the Src/Ras/Erk, the PI3K/AKT, CDK2, JAK/STAT, and several other kinases. We have performed gene-profiling studies in breast cancer cell lines to study the response to estrogens and progesterone in the presence of various kinase inhibitors. We find that optimal regulation of over 80% of the hormone target genes requires activation of at least one kinase signalling pathway. We are completing these studies with an analysis of changes in the whole cell phosphoproteome, different times after hormone treatment. These results will be integrated in a dynamic network, which should help identifying relevant nodes connecting various signalling pathways with regulation of different gene cohorts as studied by oligonucleotide arrays.

3. Regulation of MMTV Transcription in the Chromatin Context

G. Vicent, R. Zaurin, F. Le Dily, A. S. Nacht, J. Font

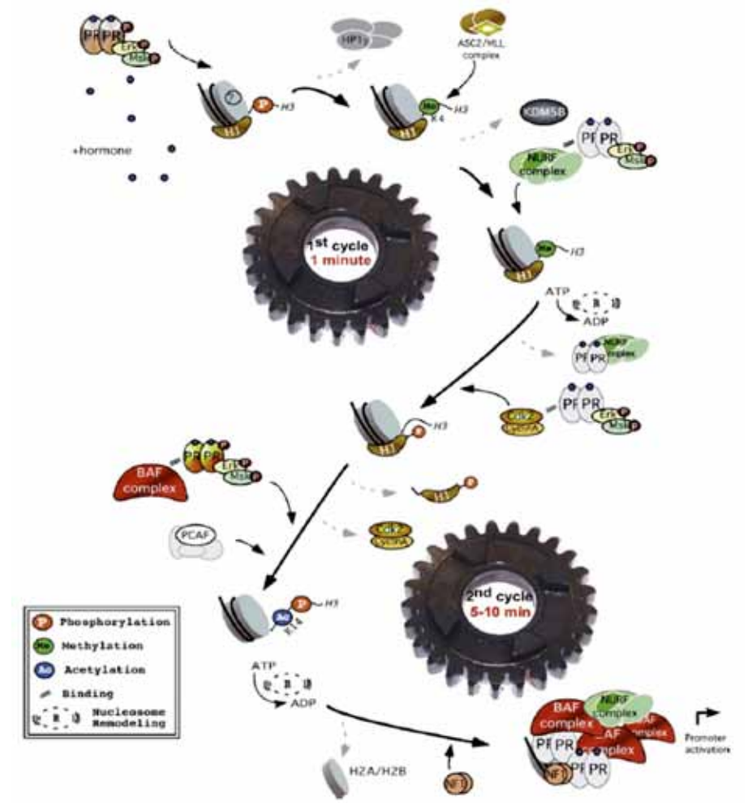
The group studies the kinetics of chromatin structural changes accompanying activation of MMTV promoter (see Model). Already 1 min after progestin treatment the activated Pr-ERK-MSK1 complex recruitsto the promoter the hNURF complex and the ASCOM complex, which are both required for gene activation. The MLL2/3 histone methyltransferases of the ASCOM complex trimethylates histone H3K4 anchoring the NURF complex via an interaction with the BPTF subunit (Vicent *et al*, 2011). Simultaneously the histone demethylase KDM5B/PLU1 is displaced from the promoter. In addition to these three complexes, the CyclinA/CDK2 kinase complex is required for the rapid opening of chromatin by displacement of phosphorylated histone H1 (Vicent *et al*, 2011). ChIP-seq and expression arrays show that H1 displacement is required for hormone induction of most hormone target genes, some of them involved in cell proliferation (Vicent *et al*, 2011). Hormonal gene regulation and histone H1 displacement also require activation of the Poly (ADP-ribose) polymerases PARP1 via phosphorylation by CDK2 (Wright *et al*, submitted).

4. Role of Steroid Hormones in Breast Cancer and Endometrial Physiology

A. Ciociola, D. Reyes, A. La Greca.

We have investigated the mechanism of the mutual inhibitory relationship between BRCA1 and PR function. We found that BRCA1 interacts with PR and leads to its polyubiquitination followed by proteasome-mediated degradation. Moreover BRCA1 is recruited to target promoters and counteracts progesterone induction via monoubiquitination of histone H2A (Calvo & Beato, 2011).

Another line of research is the analysis of the relevance of hormonal crosstalk with various kinase signaling pathways for the proliferative response of breast cancer cells. We found that hormonal activation of MSK1 is critical for activation of cell proliferation in response to hormone but not in response to serum (Reyes *et al*, unpublished). Similar experiments are being performed with endometrial stromal and epithelial cells (La Greca, unpublished).



Model. The two consecutive remodeling cycles during hormone-dependent activation of the MMTV promoter. ***First gear*** *(1 min): activated PR with associated Erk and Msk1 kinases binds to the exposed HRE1 on the surface of the MMTV promoter nucleosome. Histone H3 is phosphorylated at serine 10 by activated Msk1 promoting displacement of an HP1g-containing repressor complex. Along with the activated PR and kinases, ASC2/MLL complex (ASCOM) and NURF are recruited to the promoter chromatin. The combined action of ASCOM recruitment and KDM5B/PLU1 displacement enhances H3K4me3 and anchors NURF at the promoter. NURF remodels the nucleosome and facilitates access of PR with associated Cdk2/CyclinA kinase, which phosphorylates histone H1 and promotes its displacement. Second gear (5–10 min): recruitment of PR-BAF-kinases complexes and PCAF. H3K14 acetylation by PCAF promotes BAF anchoring. BAF mediates ATP-dependent displacement of histones H2A/H2B and thus facilitate binding of NF1. Bound NF1 stabilizes the open conformation of the histones H3/H4 tetramer particle, which exposes the previously hidden HREs allowing synergistic binding of further PR-BAF-kinases complexes and PCAF.*

PUBLICATIONS

Althammer S, González-Vallinas J, Ballaré C, Beato M, Eyras E.
"Pyicos: A versatile toolkit for the anaysis of high-throughput sequencing data"
Bioinformatics, 27(24):3333-40. (2011).

Calvo V, Beato M.
"BRCA1 counteracts progesterone action in breast cancer cells by ubiquitination leading to receptor degradation and epigenetic silencing of target genes"
Cancer Res, 71:3422-3431 (2011).

Vicent GP, Nacht AS, Font-Mateu J, Castellano G, Gaveglia L, Ballare C, Beato M.
"Four enzymes cooperate to displace histone H1 during the first minute of hormonal gene activation."
Genes Dev, 25(8):845-862 (2011).
Editors' Choice Molecular Biology by Beverly A. Purnell *Science*, 332:284 (2011).

Reviews

Beato M, Vicent GP.
"When every minute counts. The enzymatic complexity associated with the activation of hormone-dependent genes"
Cell Cycle, 10(15):2407-9 (2011).

Beato M, Vicent GP.
"Impact of chromatin structure and dynamics on PR signaling. The initial steps in hormonal gene regulation"
Mol Cell Endocrinol, 2011 Sep 12. [Epub ahead of print]

GENE REGULATION

Group: Regulation of Alternative pre-mRNA Splicing during Cell Differentiation, Development and Disease

Juan Valcárcel is an ICREA Research Professor.

Group Structure:

Group Leader: Juan Valcárcel

Staff scientist: Sophie Bonnal

Postdoctoral Fellows: Sergio Barberán, Elias Bechara, Panagiotis Papasaikas , Maria Paola Paronetto (until September 2011), Joao Tavanez

Students: Anna Corrionero (until February 2011), Camilla Ianonne, Juan Ramón Tejedor, Luisa Vigevani (since July 2011)

Technicians: Belén Miñana, Vasiliki Michaki (since June 2011), Anna Ribó, Diana Stork (since November 2011)

SUMMARY

We study molecular mechanisms that control the removal of introns from mRNA precursors (pre-mRNA splicing) and the regulation of alternative splicing. These processes are essential for expression of eukaryotic genes, expand the coding capacity of the genomes of complex organisms and play important roles in the regulation of tumor progression.

RESEARCH PROJECTS

1. A Conformational Change in the Splicing Factor U2AF Measures 3' Splice Site Strength (collaboration with the group of Michael Sattler, Helmholtz Zentrum and Technical University, Munich)

Many cellular functions involve multi-domain proteins, which are composed of structurally independent protein modules connected by flexible linkers. We have studied the recognition of the polypyrimidine (Py) tract associated with 3' splice sites by two RNA recognition motif (RRM) domains of the 65 KDa subunit of the human U2 snRNP auxiliary factor (U2AF65). Binding of U2AF65 to the Py-tract is a key early step in pre-mRNA splicing and also the target of regulation in alternative splice site decisions. We have found that the two RRM domains of U2AF65 adopt two distinct arrangements depending on the length of the Py-tract, which correlates with the efficiency of splice site utilization. Only one of the RRMs contacts short tracts of pyrimidine residues, while the presence of a longer Py-tract stabilizes an open conformation that allows contacts with the RNA that involve both RRMs (Figure 1). As a consequence, the conformational change in U2AF65 effectively *measures* the length of the Py-tract of 3' splice sites and facilitates efficient U2 snRNP recruitment to the intron during spliceosome assembly.

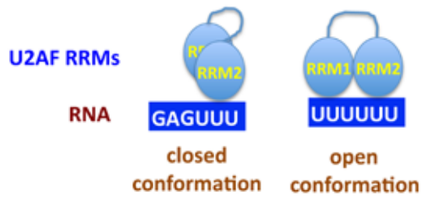


Figure 1.
Model for recognition of polypyrimidine tract length by a conformational change in U2AF65. *The RNA Recognition Motifs (RRMs) of U2AF65 can adopt two alternative conformations. In the absence of RNA, or upon binding to an RNA containing only short uridine stretches, the RRMs are in a closed conformation and only RRM2 contacts the RNA. Upon binding to an RNA containing longer uridine-rich sequences, an open conformation is stabilized and both RRMs interact with RNA, facilitating binding and subsequent events in spliceosome assembly.*

2. Proofreading of 3' Splice Site Recognition by U2AF (collaboration with the group of Michael Sattler, Helmholtz Zentrum and Technical University, Munich)

The Py-tract is an important sequence determinant of 3' splice sites, but pyrimidine-rich stretches outside of 3' splice sites are frequent in mammalian transcriptomes. Therefore the question is what prevents U2AF binding and spurious spliceosome assembly at these sequences? We used *in vitro* and *in vivo* depletion, as well as reconstitution assays using purified components, to identify the RNA binding protein hnRNP A1 as a factor that allows U2AF to discriminate between pyrimidine-rich RNA sequences depending on whether they are followed or not by a 3' splice site AG. Therefore hnRNP A1 acts as a proofreading activity for 3' splice site recognition by U2AF. Biochemical and NMR data indicate that hnRNP A1 forms a ternary complex with the U2AF heterodimer on AG-containing/uridine-rich RNAs, while it displaces U2AF from non-AG-containing/uridine-rich RNAs, an activity that requires the glycine-rich domain of hnRNP A1 (Figure 2). Proofreading assays revealed a role for hnRNP A1 in U2AF-mediated recruitment of U2 snRNP to the pre-mRNA, consistent with the functional relevance of hnRNP A1 for the splicing process.

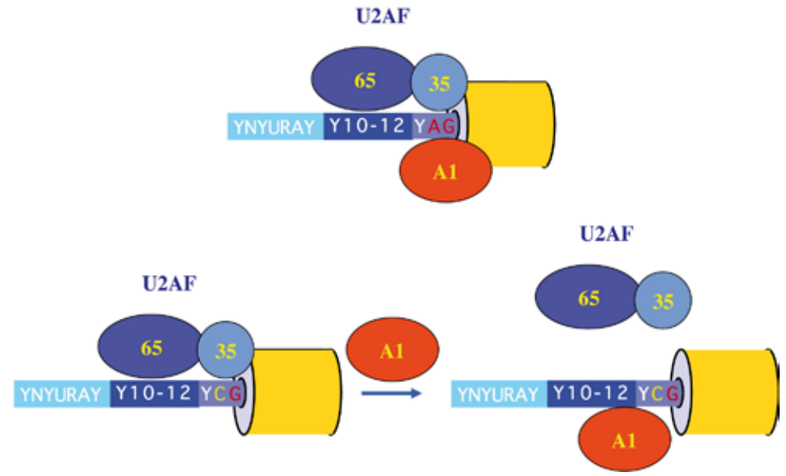
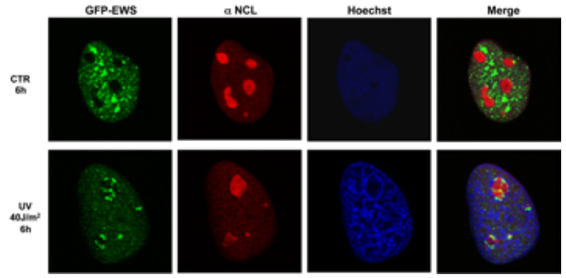


Figure 2.
Model for hnRNP A1-mediated proofreading of 3' splice site recognition by U2AF.
When the U2AF heterodimer binds to a 3' splice site containing a polypyrimidine tract and a 3' splice site AG, hnRNP A1 forms a complex that accommodates all three proteins on the RNA. However, if a bona fide 3' splice site AG is not present, hnRNP A1 displaces U2AF from the polypyrimidine tract, thus proofreading U2AF binding to prevent spurious association of U2AF with uridine-rich sequences outside of 3' splice sites.

3. The Ewing Sarcoma Protein Regulates Alternative Splicing of DNA Damage Control Genes



The Ewing sarcoma (EWS) protein is a member of the TET (TLS/EWS/TAF15) family of RNA- and DNA-binding proteins. Translocations of these genes are characteristic of certain cancers, including Ewing's sarcoma. We have studied whether EWS plays a role in the regulation of alternative splicing. We observed that EWS depletion results in alternative splicing changes affecting genes involved in DNA repair and genotoxic stress. These include the genes ABL1, CHEK2, and MAP4K2, which are important control points for these processes. Chromatin- and RNA crosslinking- immunoprecipitation results indicate that EWS binds to its target RNAs and possibly also to their DNA. We also observed that upon irradiation of cells with ultraviolet (UV) light, EWS is relocalized from the nucleoplasm to nucleoli (Figure 3) and the association of EWS with its target transcripts becomes reduced, leading to changes in alternative splicing similar to those induced by EWS knock down. This includes a change in c-ABL splicing that leads to reduced c-ABL protein expression, which can help to trigger the recognition of DNA lesions by de-repressing damaged DNA binding proteins (DDBs). Consistent with the functional relevance of EWS-mediated alternative splicing regulation, EWS depletion reduces cell viability and proliferation upon UV irradiation, effects that can be reversed by restoring c-ABL expression. Our results uncover a link between EWS and DNA damage, mediated by the regulation of alternative splicing.

Figure 3.
Subnuclear relocalization of EWS to upon irradiation of cells with UV light.
A GFP-EWS fusion is distributed in a speckled pattern in the nucleoplasm of control (CTR) cells. Upon UV light irradiation, GFP-EWS is concentrated in the periphery of nucleoli, as shown by its co-localization with the nucleolar protein nucleolin (NCL). Hoechst indicates DNA staining.

PUBLICATIONS

Corrionero A, Miñana B and Valcárcel J.
"Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug Spliceostatin A."
Genes & Development, 25:445-459 (2011).

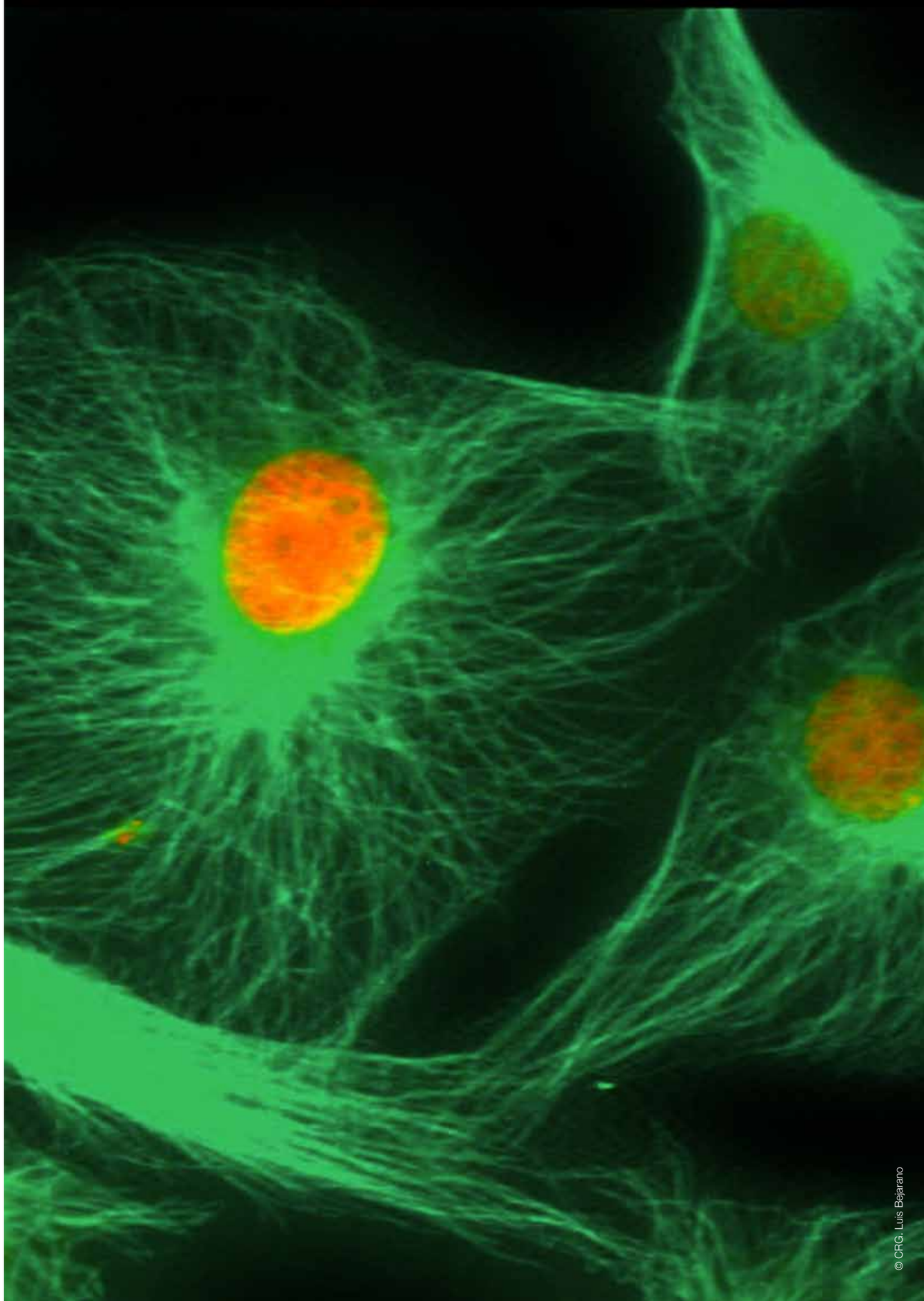
Corrionero A, Raker V, Izquierdo JM and Valcárcel, J.
"Strict 3' splice site sequence requirements for U2 snRNP recruitment after U2AF binding underlie a genetic defect leading to autoimmune disease."
RNA, 17: 401-411 (2011).

Hartmann B, Castelo R, Miñana B, Peden E, Blanchette M, Rio DC, Singh R and Valcárcel J.
"Distinct programs establish widespread sex-specific alternative splicing in Drosophila melanogaster."
RNA, 17:453-468 (2011).

Mackereth CD, Simon B, Madl T, Bonnal S, Zanier K, Gasch A, Rybin V, Valcárcel J and Sattler M.
"Multidomain conformational selection underlies pre-mRNA splicing regulation by U2AF."
Nature, 475: 408-411 (2011).

Paronetto MP, Miñana B and Valcárcel J.
"The Ewing Sarcoma protein (EWS) regulates DNA damage-induced alternative splicing."
Molecular Cell, 43: 353-368 (2011).

Tavanez JP, Madl T, Kooshapur H, Sattler M and Valcárcel J.
"hnRNP A1 proofreads 3' splice site recognition by U2AF."
Molecular Cell, in press.



GENE REGULATION

Group: Regulation of Protein Synthesis in Eukaryotes

Group Structure:

Group Leader: Fátima Gebauer

Lab manager: Olga Coll

Postdoctoral Researchers: Laurence Wurth, Antoine Graindorge

PhD Students: Ana Villalba, Cristina Militti, Emilia Szostak, Marina García

Technician: Anna Ribó

SUMMARY

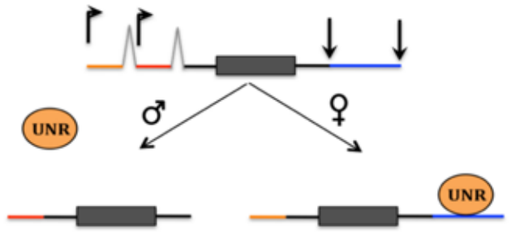
Our group is interested is the regulation of mRNA translation during embryonic development and cell homeostasis. We wish to understand the molecular mechanisms of translational control exerted by RNA-binding proteins (RBPs), the broader roles of translational regulators in RNA biology, and the consequences of altered RBP function in disease. In the past, we have been able to establish extracts from *Drosophila* embryos that recapitulate translational control events important for early development and viability. We have exploited these extracts to study mechanistically different examples of translational control. The results have led to the identification of RBP complexes involved in regulation, and have unravelled a new layer of sex-specific complexity contributed by ubiquitous RBPs recognizing sex-specific alternatively processed transcripts.

RESEARCH PROJECTS

1. Translational Control of Dosage Compensation

X-chromosome dosage compensation is the process that equalizes the expression of X-linked genes in males (XY) and females (XX). Dosage compensation is essential for life and is initiated early during embryonic development. In *Drosophila*, dosage compensation is achieved by hypertranscription of the male X chromosome as a consequence of the binding of the dosage compensation complex (DCC) to hundreds of sites on the X. In females, dosage compensation is repressed via the translational inhibition of the rate-limiting DCC component MSL2. At least two RNA-binding proteins are involved in this repression: the female-specific protein Sex-lethal (SXL) and the ubiquitous protein Upstream of N-ras (UNR) (reviewed in Graindorge *et al*., 2011). Despite high amounts of *msl2* transcript in males, UNR does not bind (and does not repress) *msl2* in this sex, because its binding depends on SXL. Thus, sex-specific binding of RBPs to targets influences the outcome of regulation. To determine the extent of sex-specific binding by UNR, we have identified its targets in both sexes using immunoprecipitation followed by RNA isolation and high-throughput sequencing. The results have revealed a large degree of sex-specific target binding by UNR. This binding is explained not only by sex-specific expression of the targets but, mainly, by alternative processing of target UTRs (Mihailovich *et al*., 2012). These data indicate that alternative promoter usage, alternative splicing and polyadenylation generate sex-specific diversity that is exploited by RBPs to regulate gene expression (Figure 1).

Figure 1.
Alternative UTR processing contributes to sex-specific binding by RBPs.
Alternative promoter usage, alternative splicing and alternative polyadenylation generate transcripts with sex-specific UTRs that are recognized differentially by UNR, a protein present in both males and females.



2. Translational Regulation by Cytoplasmic Polyadenylation

Embryonic axis formation in *Drosophila* depends on the timely translation of localized (*bicoid*) and non localized (*Toll*) transcripts. Translation of these transcripts is activated by cytoplasmic polyadenylation, a process known to require the RNA element CPE and its binding protein CPEB in vertebrates, in addition to the multisubunit complex that binds to the polyadenylation hexanucleotide (Villalba *et al*., 2011). In *Drosophila*, the sequences and factors regulating cytoplasmic polyadenylation are largely unknown. Using a cell-free cytoplasmic polyadenylation system obtained from early embryos we have found that *Toll* mRNA is polyadenylated by a novel, non-canonical mechanism that is independent of the CPE and the hexanucleotide. Our current research focuses in isolating the *Drosophila* machineries responsible for both canonical and non-canonical polyadenylation.

PUBLICATIONS

Mihailovich M, Wurth L, Zambelli F, Abaza I, Militti C, Mancuso FM, Roma G, Pavesi G, Gebauer F.
"Widespread generation of alternative UTRs contributes to sex-specific RNA binding by UNR."
RNA, Epub 2011 Nov 18.

Reviews

Graindorge A, Militti C and Gebauer F.
"Post-transcriptional control of X-chromosome dosage compensation."
WIREs RNA, 2:534-545 (2011).

Villalba A, Coll O, Gebauer F.
"Cytoplasmic polyadenylation and translational control."
Curr Opin Genet Dev, 21:452-457 (2011).

Book chapter

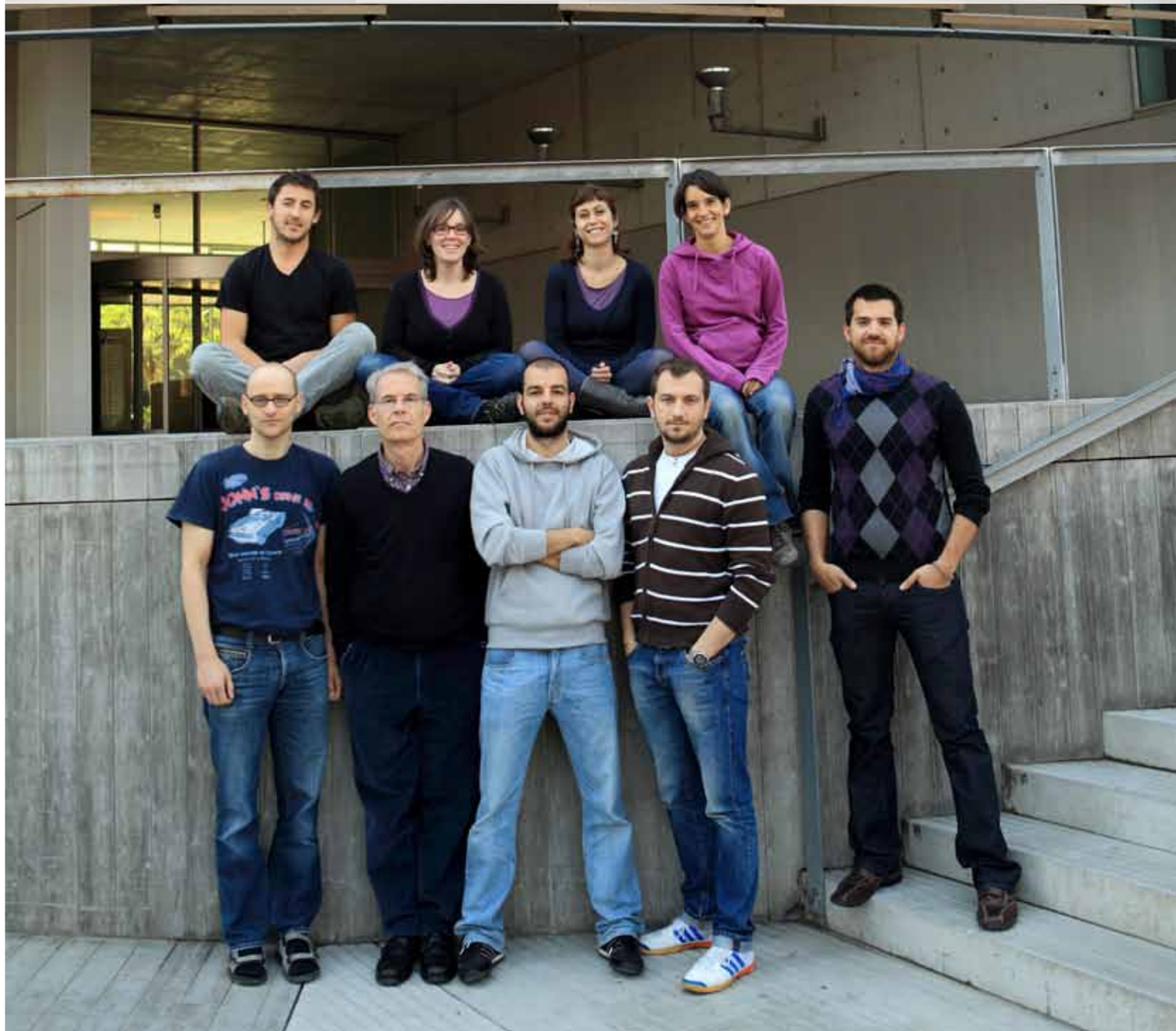
Gebauer F, Preiss T and Hentze MW.
"Translational control: from cis-regulatory elements to complex RNPs and back".
In: *Translational Control*, 4th Edition. Eds: MB Matthews, N. Sonenberg and JWB Hershey, Cold Spring Harbor Laboratory Press, in press.

DIFFERENTIATION AND CANCER

Group: Hematopoietic Differentiation and Stem Cell Biology
Thomas Graf is an ICREA Research Professor.

Group Structure:

Group Leader:	Thomas Graf
Staff scientist:	Chris van Oevelen
Postdoctoral Fellows:	Christos Gekas (EMBO fellow), Eric Kallin (EMBO fellow)
PhD Students:	Luisa Irene de Andrés (CRG), Alessandro DiTullio (CRG), Francesca Rapino ("la Caixa"), Bruno Di Stefano ("la Caixa")
Technicians:	Clara Berenguer (Plan Nacional), Jose Francisco Infante (CRG), Alai Urrutikoetchea (Plan E)



SUMMARY

The laboratory's main interests are mechanisms of transcription factor-induced reprogramming of hematopoietic cells and the function of genes required to establish and maintain the hematopietic stem cell phenotype.

RESEARCH PROJECTS

1. How does C/EBPa Induce both the Activtation and Silencing of Genes during the Transdifferentiation of pre-B Cells into Macrophages?

Using our inducible pre-B cell line system carrying C/EBPaER we have shown that the cells become transgene independent about 24 hours after induction. During this time endogenous C/EBPb, a transcription factor that can also induce transdifferentiation, as well as PU.1, a partner required for C/EBP, become transcriptionally activated. Knockdown of C/EBPb and PU.1 almost completely inhibits the reprogramming, showing that they are required for the establishment of the myeloid fate. In contrast, endogenous C/EBPa only becomes activated after commitment, and knockout experiments showed that it is responsible for the maintenance of the macrophage state. These experiments suggest that C/EBPa induces a stable myeloid transcription factor network consisting of the transcription factor triad C/EBPa, C/EBPb and PU.1. To study the underlying molecular mechanism we have embarked on a major effort to study transcription factor binding and chromatin modifications during the C/EBPa induced transdifferentiation. We have obtained ChIPseq data for uninduced (0h) as well as cells 3h, 12h, 24h and 48h cells, for the binding of the myeloid regulators C/EBPa, C/EBPb and PU.1 to the DNA as well as genome wide changes in the chromatin marks H3K4me1, HeK4me3, H3K27me3, H3K27ac. Our observations so far are that the major changes induced by the myeloid transcription factors occur when C/EBPa binds to enhancers in a sustained manner, resulting in acetylation of H3K27 and activation of myeloid gene expression through enhancer-promoter interactions. Surprisingly, different combinatorial enhancer binding modes of the transcription factor triad result in distinct kinetics of myeloid gene upregulation, reflected in different polymerase II elongation rates at target gene promoters.

2. High C/EBPa Levels Increase the Proportion of pre-B Cells that Transdifferentiate into Macrophages without Cell Division

During C/EBPa induced transdifferentiation pre-B cells divide once and then stop in G0. To investigate whether C/EBPa-induced reprogramming of pre-B cells into macrophages requires cell division we performed BrDU incorporation experiments. We found that 8% of the induced cells did not incorporate BrdU during reprogramming and that 90% of the cells were blocked in transdifferentiation by the cell cycle inhibitor aphidicolin. Surprisingly, however, a comparison of inducible pre-B cells separated into G0/G1 phase and G2/M fractions showed no significant differences in their reprogramming kinetics. In support of this result we showed that knocking down p53 in the inducible pre-B cells does not alter their conversion into macrophages, suggesting that an acceleration of the cell cycle has no effect. To study this apparent discrepancy we performed time-lapse experiments during transdifferentiation, showing that about 8% of the pre-B cells did not divide before acquiring macrophage properties. They also showed that the subpopulation of cells that did not divide transdifferentiated more rapidly than the dividing cells. We found that this cell fraction expresses the highest levels of C/EBPa, suggesting that cells that with lower levels of C/EBPa undergo cell division because they still retain the cycling program of pre-B cells, which are rapidly dividing cells. Together these data show that cell division is not required for the C/EBPa induced transdifferentiation of pre-B cells into macrophages and that high lvels of C/EBPa accelerate the reprogramming process.

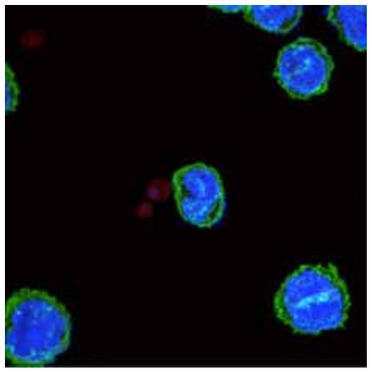
3. Role of the Tet2 Enzyme during the Transdifferentiation of pre-B Cells into Macrophages

To study the epigenetic changes induced during transdifferentiation we asked, in collaboration with Esteban Ballestar (IDIBELL, Barcelona), what happens to the cell's DNA methylation? Using a genome wide promoter-tiling array of methylated DNA binding protein, the reprogrammed macrophages differed very little from the starting pre-B cell line (Rodriguez-Ubreva *et al.*, 2011). This was very surprising since DNA methylation of genes is commonly associated with gene repression yet we observed genes with heavily methylated promoters that became activated during induced transdifferentiation. This turned our attention to the possibility that Tet2, an enzyme that converts methylated cytosine residues into hydroxymethylated cytosines, plays a role in the derepression of myeloid genes. We found that the Tet2 gene becomes rapidly activated during transdifferentiation, mediated by the direct binding of C/EBPa to upstream sequences. We could further show that Tet2 knockdown constructs delayed the C/EBPa induced switch and that the resulting macrophages were impaired in their phagocytic capacity. Gene expression arrays of these cells permitted us to identify about 60 target genes whose activation was inhibited by the Tet2 knockdowns, and of the targets tested, the majority contained Tet2 binding sites in their promoter. Interestingly, these promoters became hydroxymethylated and eventually demethylated. These observations suggest that the hydroxymethylation of specific macrophage promoters by Tet2 accelerated the C/EBPa induced upregulation of the corresponding genes, pointing to a novel mechanism of de-repression. They also suggest that the role of Tet2 in the generation of acute myeloid leukemia consists in the incomplete upregulation of macrophage genes.

4. Transdifferentiation of Human Lymphoma Cells and a Potential Therapeutic Application

In another line of research we asked whether human lymphoma B cells can be reprogrammed into macrophages by C/EBPa. In Burkitt's lymphoma, proliferation of mature B cells is induced by the activation of Myc expression following transocation of the oncogene to the IgH enhancer. We found that several Burkitt's lymphoma lines could be transdifferentiated into functional, quiescent macrophages. Remarkably, the efficiency was nearly 100%, far higher than the frequency of the recently described conversion of human fibroblasts into neurons. We next asked whether transdifferentiation inhibits the formation of human lymphomas in immunodeficient mice. We found (in collaboration with J. A. Martinez Climent, U. de Navarra) using both a Burkitt's lymphoma cells line and an immature B cell leukemia line stably expressing C/EBPaER that tumor formation could be inhibited by their transdifferentiation into macrophages. The formation of lymphomas in the liver could even be delayed by administering the inducer 10 days after inoculation of untreated tumor cells. In addition, all residual tumor cells were found to be lymphoid. These results suggest that it might be possible to find drugs that can induce transdifferentiation as potentially new therapeutics for leukemias and lymphomas.

Uninduced



Induced

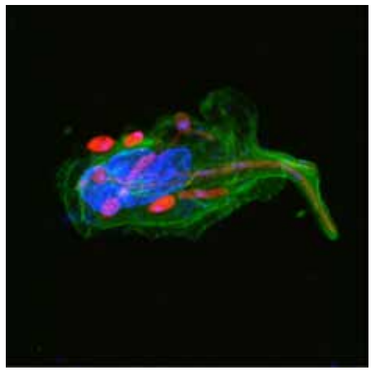


Figure 1.
C/EBPa induced transdifferentiation of immortalized human B cells expressing C/EBPaER. Phagocytic capacity of the cells, showing ingestion of red fluorescent Candida albicans by 4 day induced cells but not by uninduced cells. The cells are GFP positive (green); the nuclei were stained by DAPI (blue). The yeast is in red. Note the differences in cell morphology.

PUBLICATIONS

Graf T.
"Historical origins of transdifferentiation and reprogramming to pluripotency."
Cell Stem Cell, 9:504-516 (2011).

Rodríguez-Ubreva J, Ciudad S, Gómez-Cabrero C, Parra M, Bussmann LH, di Tullio A, Kallin EM, Tegnér J, Graf T and Ballestar E.
"Pre-B Cell to Macrophage Transdifferentiation Without Significant Promoter DNA Methylation Changes"
Nucleic Acid Research, 2011 Nov 15. [Epub ahead of print]

Di Tullio A, Manh TP, Schubert A, Månsson R, Graf T.
"C/EBPalpha-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation."
Proc Natl Acad Sci U S A., 108(41):17016-21 (2011).

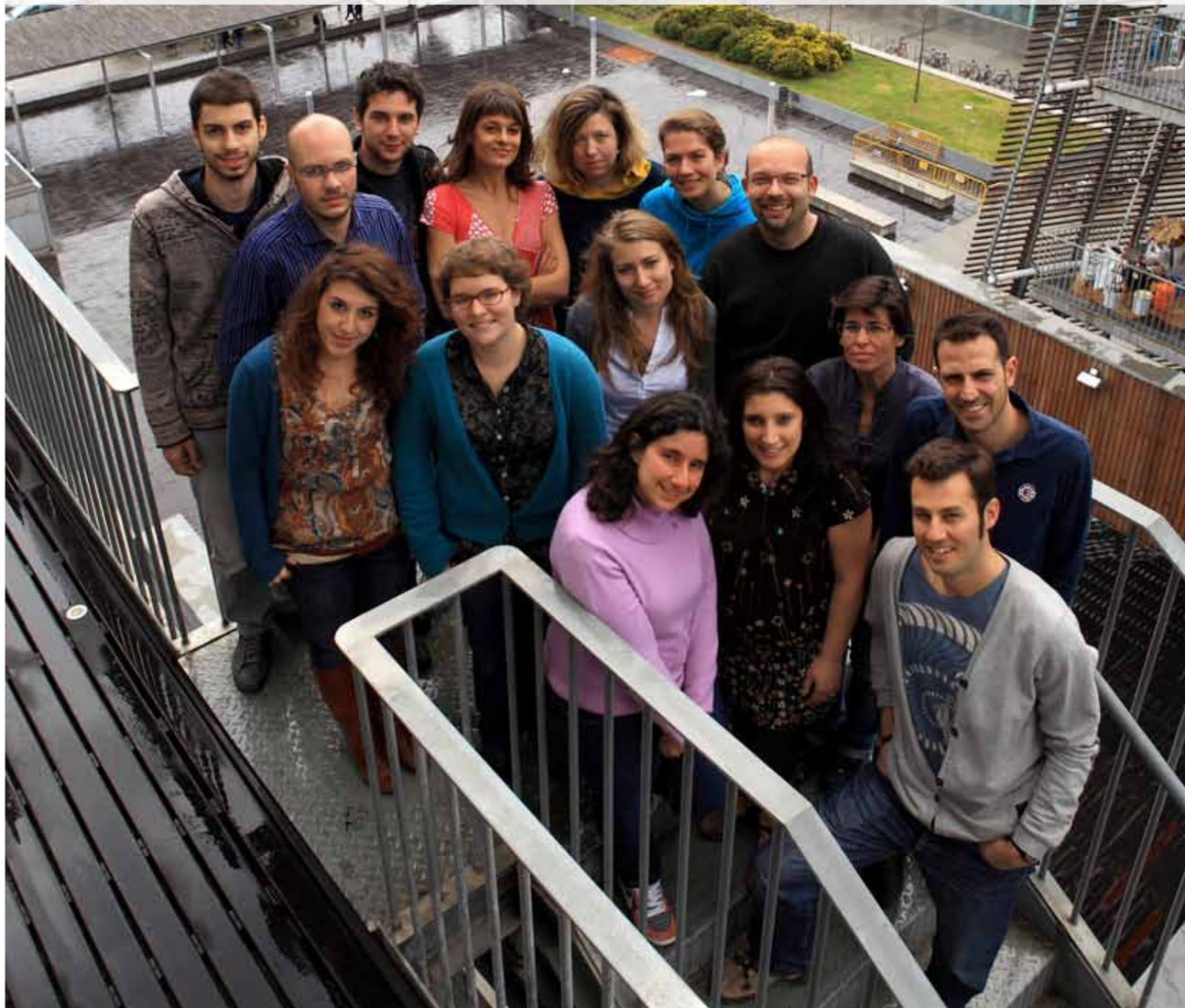
de Andrés-Aguayo L, Varas F, Kallin EM, Infante JF, Wurst W, Floss T, Graf T.
"Musashi 2 is a regulator of the HSC compartment identified by a retroviral insertion screen and knockout mice."
Blood., 118(3):554-64 (2011).

DIFFERENTIATION AND CANCER

Group: Reprogramming and Regeneration
Maria Pia Cosma is an ICREA Research Professor.

Group Structure:

Group Leader:	Maria Pia Cosma
Staff scientist:	Frederic Lluís
Postdoctoral:	Daniela Sanges, Wassim Altarche, Karthik Arumugam, Lucia Marucci, Luigi Ombrato
PhD Students:	Giacom Simonte, Joao Frade, Elisa Pedone, Maria Aurelia Ricci
Master Students:	Francesco Aulicino, Lena Robra
Technicians:	Vanessa Chiganças, Umberto Di Vicino, Maribel Muñoz, Neus Romo



SUMMARY

In our group, we are investigating the mechanisms controlling the reprogramming of somatic cells, and our final goal is to determine if this reprogramming contributes to tissue regeneration in higher vertebrates.

Indeed, whether somatic cell reprogramming can occur *in vivo* in higher vertebrates and what the molecular mechanisms and genes driving reprogramming are, it remains to be defined. We have shown that activation of the Wnt/β-catenin signalling pathway enhances reprogramming of somatic cells after their fusion with stem cells. Remarkably, the activation of this signalling pathway also controls regeneration in response to damage in lower and higher vertebrates; furthermore, cell fusion is one possible mechanism of regeneration in vertebrates. Our main goals are: i.) to dissect the mechanisms of Wnt-mediated somatic cell reprogramming and ii.) to determine whether in mice activation of Wnt/β-catenin signalling controls cell-fusion mediated tissue regeneration.

RESEARCH PROJECTS

1. Identification of β-catenin Targets and Molecular Pathways that Control Cell Reprogramming, and Analysis of their Interactions via Network-Identification Algorithms

Embryonic stem cells (ESCs) express factors that can reprogramme a somatic-cell nucleus. As a result, cell fusions between differentiated cells and embryonic cells produce ESC-like pluripotent reprogrammed hybrids. Our goal is to identify these factors, i.e. the “reprogrammers”, that are targets of β-catenin and that can reprogramme differentiated cells. Reverse engineering and forward algorithms will be used to infer the natural network of interactions surrounding the genes involved in the reprogramming of somatic cells. Furthermore, we will develop mathematical models to dissect out the threshold and timing effects of nuclear factor accumulation that control cell reprogramming. Finally, we will investigate the molecular functions of the identified reprogrammers, and therefore the molecular mechanisms of somatic-cell reprogramming.

2. A study of the Roles of Tcf Factors in Controlling Somatic-Cell Reprogramming

Tcf1 (Tcf7), Lef-1, Tcf3 (Tcf7l1) and Tcf4 (Tcf7l2) form the Tcf family of transcription factors that modulate the transcription of genes by recruiting chromatin remodelling and histone-modifying complexes to their target genes. β-catenin binds to target promoters through its interactions with the Tcf proteins. Tcf3 is the most expressed of the Tcf isoforms in embryonic stem cells, and it is an important player of the pluripotency network. We recently discovered that Tcf3 functions as a repressor of the reprogramming potential of somatic cells. The functions of the other Tcf factors in the reprogramming process are currently under investigation.

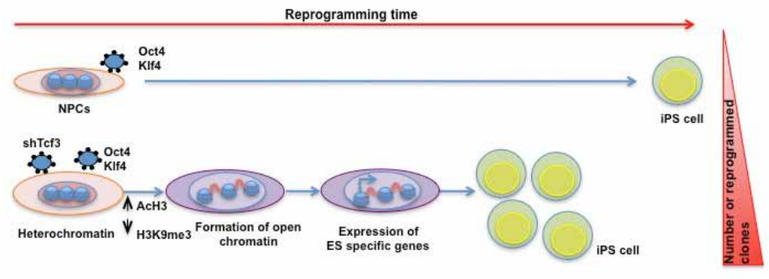


Figure Legend:
Silencing of Tcf3 increases reprogramming efficiency. From Luis, Ombrato et al. PNAS 2011

3. To Determine whether Wnt/β-Catenin-Dependent Reprogramming of Fused Cells is a Mechanism of Regeneration in Higher Vertebrates

We will determine whether Wnt/β-catenin signalling controls *in-vivo* reprogramming of hybrids formed in response to injury. Transplantation of perturbed (Wnt-activated or repressed) adult stem cells into a variety of drug-induced or genetically modified damaged organs will be carried out. Short-term and long-term regeneration will be studied. Genetic approaches will be used to evaluate cell fusion, reprogramming and regeneration in the tissues analyzed.

PUBLICATIONS

Lluis F, Ombrato L, Pedone E, Pepe S, Merrill BJ, Cosma MP.
“T-cell factor 3 (Tcf3) deletion increases somatic cell reprogramming by inducing epigenome modifications.”
PNAS, 108(29):11912-7 (2011).

Reviews

Sanges D, Lluis F and Cosma MP.
“Cell fusion mediated reprogramming: pluripotency or transdifferentiation? Implication in regenerative medicine.”
Adv Exp Med Biol, 713:137-59 (2011).

Sanges D and Cosma MP
“Reprogramming cell fate to pluripotency: the decision-making signalling pathway.”
International Journal of Developmental Biology, 54(11-12):1575-87 (2011).

DIFFERENTIATION AND CANCER

Group: Epigenetics Events in Cancer
Luciano Di Croce is an ICREA Research Professor.

Group Structure:

Group Leader: Luciano Di Croce

Postdoctoral: Luigi Aloia, Lluis Morey, Malte Beringer (since September 2011), Elisabeth Simboeck, Martin Lange (until June 2011)

PhD Students: Santiago Demajo, Sophia Teichmann, Iris Uribesalgo Micás, Joana Ribeiro, Paola Pisano, Jain Payal, Alexandra Santanach (since October 2011)

Technician: Arantxa Gutierrez, Cecilia Ballaré

Visiting Scientists: Livia Caizzi, Sarah Kinkley

SUMMARY

Understanding the genetic basis of cancers has been a topic of intense research, and hundreds of gene mutations have been identified that can cause carcinogenesis. However, in the past few years, increasing evidence has suggested that mutations are not the only genetic changes that lead to cancer. Indeed, perturbations of chromatin structure and of other epigenetic mechanisms can cause inappropriate gene expression and genomic instability, resulting in cellular transformation and malignant outgrowth.

The focus of our laboratory is to understand the basic mechanism of gene regulation and the impact of epigenetic marks on chromatin metabolism, using normal cells, cancer cells, and mES cells as model systems. We will also address some of these questions using several mouse models.

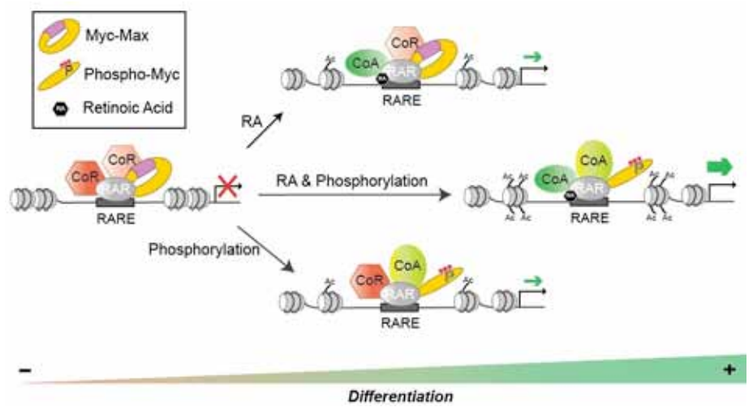
RESEARCH PROJECTS

1. Dual Role of Myc in Leukemia

The transcription factor c-Myc strongly stimulates cell proliferation but also regulates apoptosis, senescence, cell competition, and cell differentiation, and its elevated activity is hallmark for human tumorigenesis. c-Myc induces transcription by forming heterodimers with Max and then directly binding DNA at E box sequences. Conversely, transcription repression depends primarily on the inhibitory interaction of c-Myc/Max with Miz-1 at DNA initiator elements. We recently identified a distinct mechanism of c-Myc gene regulation, in which c-Myc interacts with the retinoic acid receptor alpha (RARα) and is recruited to RAR DNA binding sequences (RAREs). In leukemia cells, this c-Myc/RARα complex functions either as an activator or a repressor of RARα–dependent targets through a phosphorylation switch. Unphosphorylated c-Myc interacts with RARα to repress the expression of RAR targets required for differentiation, thereby aggravating leukemia malignancy. However, if c-Myc is phosphorylated by the kinase Pak2, the c-Myc/RARα complex activates transcription of those same genes to stimulate differentiation, thus reducing tumor burden. We are now inviestigatin the role of c-Myc in balancing proliferation and differentiation, and how modulating this previously unidentified c-Myc activity might provide alternative therapies against leukemia and possibly other types of tumors.

2. ZRF1 and histone H2A ubiquitination

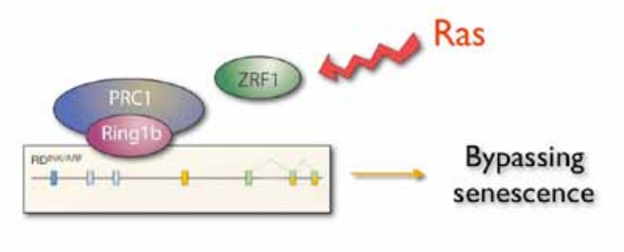
The Polycomb Repressive complex 1 possesses an E3 ligase (Ring1b), the activity of which leads to the monoubiquitination of H2A on lysine 119. H2Aub is the most abundant mono-ubiquitinated protein in the cell, and it was identified more than 20 years ago.



We have recently identified and characterized ZRF1 as the first known protein that specifically binds to H2Aub. Binding of ZRF1 to promoters is necessary not only to displace Polycomb proteins but also to reactivate genes silenced by Polycomb. Indeed, genome-wide analysis showed a significant overlap between ZRF1 and Polycomb target genes.

We will characterize the ZRF1-containing complexes, and we will dissect the molecular steps necessary for promoter reactivation. Our preliminary data suggest that ZRF1 interacts with the mammalian Trithorax complex as well as with a specific histone H2A de-ubiquitinase enzyme.

Furthermore, our recent data also suggest that ZRF1 plays a role in mouse stem cell differentiation. In collaboration with EMBL-Monterotondo, we have started generating the ZRF1 knock-out mice. This will allow us to integrate the transcriptional role of ZRF1 within the context of mouse development. Finally, since ZRF1 is over-expressed in different tumor types, and since ZRF1 regulates the INK/ARF locus, we are studying its role in transcriptional de-regulation in cancer cells.



3. Role of Histone Demethylases in Leukemia

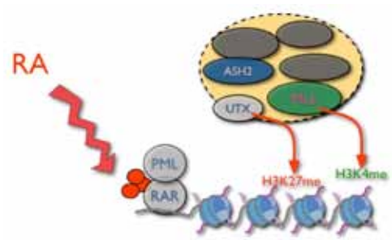
PcG and TrxG proteins were initially described in *Drosophila* as repressors and activators of Hox genes, respectively. More generally speaking, PcG and TrxG proteins play an important role in regulating lineage choices during development and differentiation. Additionally they are implicated in cell proliferation, stem cell identity and cancer, cellular senescence, genomic imprinting, X-inactivation and hematopoiesis.

In collaboration with Dr. R. Shiekhattar (Wistar, USA), we identified UTX as the enzyme responsible for H3K27 demethylation (Lee *et al*, 2007). More recently, UTX have been found mutated in several tumour (van Haaften *et al*, 2009), thus corroborating the important role of epigenetic deregulation in human cancers. Interestingly, UTX is a component of the MLL complex, the mammalian orthologous of Trithorax. Our previous date indicates that after retinoic acid administration, Utx specifically demethylases H3K27 at several Hox genes.

We will investigate:

(i) how Utx is recruited to promoters,
(ii) which are the target genes in the human genome,
(iii) which role has Utx in tumorigenesis.

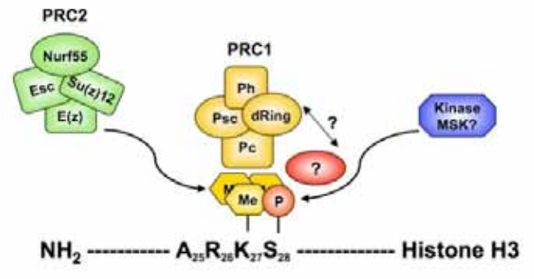
The oncoprotein PML-RARα is one of the most well-studied leukemogenic transcription factors. The PML-RARα fusion protein, responsible for 99% of acute promyelocitic leukemia (APL) cases, arises from a t(15;17) balanced reciprocal chromosomal translocation. It involves the PML gene and the retinoic acid receptor α (RARα) gene. PML-RARα represses target genes through recruitment of DNMTs and Polycomb complex (Villa *et al*, 2007). It is likely that Utx is important for the re-establishment of the proper gene transcription program necessary for cell differentiation. Thus the role of Utx in leukemia is also being investigated.



4. Cross-Talk of Epigenetic Marks in Gene Regulation

Binding of chromatin-associated proteins can be influenced by adjacent additional modification. For instance, binding of HP1 to H3K9me is inhibited by an adjacent phosphorylation on serine 10 (H3S10ph). H3S10ph is associated with actively transcribed chromatin and is recognized by 14-3-3 proteins, whose binding is even favoured by additional lysine 14 acetylation (H3K14ac).

Therefore, we wonder whether additional posttranslational modifications in the proximity of H3K27me, in particular phosphorylation of Serine 28 (H3S28ph), would have any effect on Pc binding and consequently on target gene expression.



Similar to H3S10ph, H3S28ph is an abundant and transient mark during mitosis, while during interphase is also associated with actively transcribed chromatin. However it was shown that H3S28ph targets a different chromatin population than H3S10ph. Proteins, which may specifically bind to H3S28ph, are still unknown.

We have identified several H3S28ph “binder”. We are now characterizing their function with respect to Polycomb binding, cell faith decision, and their role in promoter regulation and chromatin structure.

PUBLICATIONS

Uribesalgo I, Buschbeck M, Gutiérrez A, Teichmann S, Demajo S, Kuebler B, Martín-Caballero J, Roma G, Benitah SA, and Di Croce L.
“E-box independent regulation of transcription and differentiation by c-Myc.”
Nature Cell Biology, 13:1443-49 (2011).

Navarro-Martín L, Viñas J, Gutierrez A, Di Croce L and Piferrer F.
“DNA Methylation of the Gonadal Aromatase (cyp19a) Promoter Is Involved in Temperature-Dependent Sex Ratio Shifts in the European Sea Bass.”
PLOS Genetics, 7:e1002447 (2011).

Janich P, Pascual G, Merlos A, Batlle E, Ripperger J, Albrecht U, Obrietan K, Di Croce L, Benitah SA.
“The circadian molecular clock creates epidermal stem cell heterogeneity.”
Nature, 480:209-14 (2011).

Luis NM, Morey L, Mejetta S, Pascual G, Janich P, Kuebler B, Roma G, Nascimento E, Frye M, Di Croce L, Benitah SA.
“Regulation of human epidermal stem cell proliferation and senescence requires polycomb- dependent and -independent functions of cbx4.”
Cell Stem Cell, 9:233-46 (2011).

Mejetta S, Morey L, Pascual G, Kuebler B, Mysliwiec MR, Lee Y, Shiekhattar R, Di Croce L, and Benitah SA.
“Jarid2 regulates mouse epidermal stem cell activation and differentiation.”
EMBO J, 30:3635-46 (2011).

Reviews

Di Croce L.
"Regulating the shuttling of eukaryotic RNA polymerase II."
Mol Cell Biol, 31:3918-20 (2011).

Simboeck E, Ribeiro JD, Teichmann S, Di Croce L.
"Epigenetics and senescence: Learning from the INK4-ARF locus."
Biochem Pharmacol, 82:1361-70 (2011).

Richly H, Aloia L, Di Croce L.
"Roles of the Polycomb group proteins in stem cells and cancer."
Cell Death Dis, 2:e204 (2011).

Richly H, and Di Croce L.
"The flip side of the coin: Role of ZRF1 and histone H2A ubiquitination in transcriptional activation."
Cell Cycle, 10:745-50 (2011).

Lange M, Demajo S, Jain P, Di Croce L.
"Combinatorial assembly and function of chromatin regulatory complexes."
Epigenomics, 3:567-580 (2011).

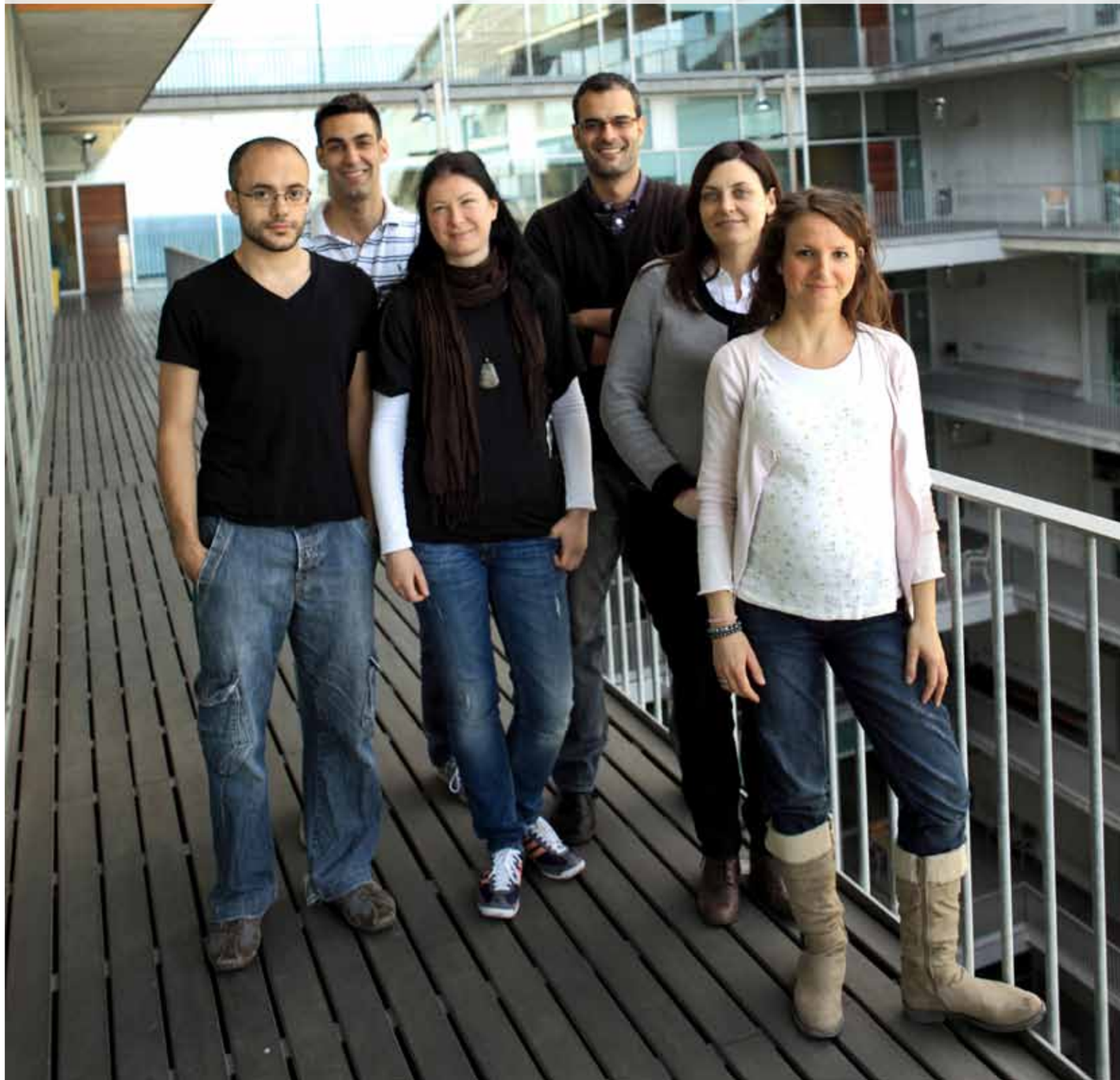
Uribesalgo I, and Di Croce L.
"Dynamics of epigenetic modifications in leukemia."
Briefings in Functional Genomics, 10:18-29 (2011).

DIFFERENTIATION AND CANCER

Group: Epithelial Homeostasis and Cancer
Salvador Aznar Benitah is an ICREA Researcher.

Group Structure:

Group Leader:	Salvador Aznar Benitah
Postdoctoral:	Guiomar Solanas
PhD Students:	Jordi Hernández, Stefania Mejetta
Technicians:	Gloria Pascual
Masters Student:	Lorenzo Rinaldi



SUMMARY

The epidermis, and other stratified epithelia, needs to renew constantly in adults to maintain its function. This process is called homeostasis and relies on a population of **epidermal stem cells (epSCs)** that self renew and can undergo terminal differentiation. EpSCs adhere strongly to specialized niches where they remain relatively quiescent and unspecified. Upon a requirement of tissue replenishment, they become active, proliferate, and egress the niche to contribute to the differentiated compartment. The process is assymmetric, ensuring that the percentage of stem cells is maintained more or less constant after each cycle of activation. The transition between each state (quiescence vs proliferation; adherence vs egression; unspecified vs differentiated) is tightly regulated by the microenvironment and the intrinsic genetic program of the epSCs. Failures in this strict regulation can lead to premature ageing or to the development of tumours. **The aim of our work is to understand the molecular mechanisms that control the behaviour of normal adult stem cells during tissue homeostasis and how their deregulation contributes to carcinogenesis.**

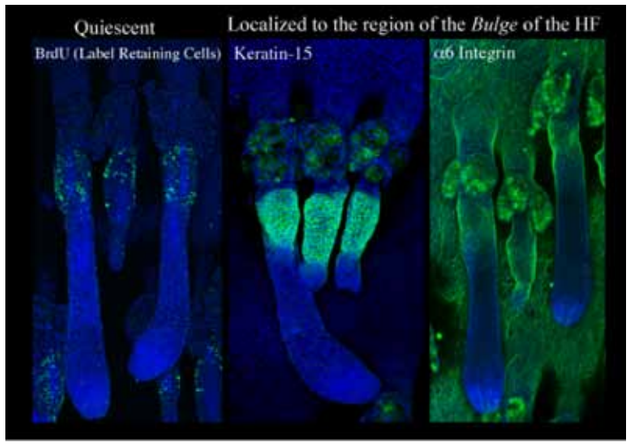


Figure 1:
Epidermal stem cells are located at a region of the hair follicle known as the bulge. They are relatively quiescent and strictly positioned at their niche (Immunostaining and pictures by Peggy Janich).

RESEARCH PROJECTS

Little is still known about the spatiotemporal distribution, and the hierarchy, of the molecular pathways relevant to the transition between the inactive and active states of epidermal stem cells within the stem cell niche. Intriguingly, inactive epSCs express high levels of ¨molecular breaks¨, which make them refractory to activating stimuli. Why then do they respond to such stimuli? In addition, upon stimulation, a very small proportion of epSCs become active, whereas the bulk remains unresponsive. Why don´t all epSCs respond? What is the nature of this stem cell heterogeneity? The consequences of unbalancing this equilibrium must be underscored, since tilting it towards excessive or reduced activation may predispose the tissue to premature aging due to excessive stem cell depletion, carcinogenesis when combined with accumulation of DNA damage, or lack of regenerative potential due to the inability of the stem cells to become active upon tissue damage.

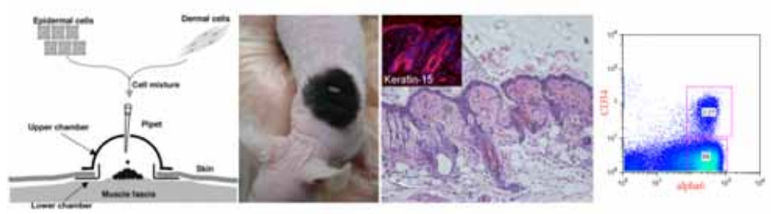
We can summarize our interests in three questions:

(a) **What is the molecular nature of the heterogeneity of epSCs within their niche?** A small percentage of epSCs respond to activating stimuli; why do not all cells respond? How are these restricting mechanisms lost during carcinoma formation?

(b) **Are activating stimuli instructive or permissive for epSC activation?** Inactive epSCs highly express inhibitors of activating stimuli: why do they respond to activating stimuli then? Is response predisposed by an intrinsic genetic program of the epSC?

(c) **What distinguishes the different choice of the two epSCs daughter cells?** i.e. remain at or exit the stem cell niche. How and why is this mechanism lost in carcinomas?

Figure 2:
Epidermal stem cells are the only population that can regenerate the entire epidermal compartment. Epidermal stem cells transplanted onto nude mice can regenerate a homeostatic functional epidermis, hair follicles and sebaceous glands. A niche of epidermal stem cells is re-established once transplanted, which exhibits expression of stem cell markers such as keratin-15, integrin alpha6, and CD34 (Experiment done by Gloria Pascual).



1. Molecular and Genetic Mechanisms Involved in Epidermal Self-Renewal and Differentiation

We have previously identified a signalling axis important for establishing equilibrium between inactive and active epSCs. The small GTPase Rac1 promotes EpSCs quiescence and strong adhesion to the stem cell niche (Watt 2008, Benitah 2005). Conditional epidermal deletion of Rac1 causes an initial burst of proliferation and loss of the quiescent epSC pool, coupled to a massive exit of activated epSCs from their niche. Continuous epSC activation, upon Rac1 deletion, significantly depletes their number, leading to loss of epidermal maintenance and integrity.

Modulation of this pathway has enabled us to induce two states: inactivation (Rac1high/phospho-Myc) and activation (Rac1low/unphospho-Myc) of EpSCs. Based on this, the global comparative transcriptome of human EpSCs in their active versus inactive state has been obtained. Analysis of this data has allowed us to identify key signaling pathways involved in epidermal stem cell behavior. Currently we are analyzing the role of varios selected pathways using cellular and molecular biology tools with primary cultures, as well as with *in vivo* mouse models:

Identification of pathways that establish epidermal stem cell niche heterogeneity: We have identified a molecular clock mechanism that establishes transcriptional oscillations of a large proportion of the genes that constitute the epidermal stem cell signature. This mechanism establishes an equilibrium of two stem cell predisposition states within the niche, one prone to become active, and one prone to remain dormant. Perturbation of this equilibrium, using novel in vivo mouse models, is allowing us to understand how this mechanism affects long term tissue homeostasis, and the predisposition to develop carcinomas. We are further characterizing the molecular signatures of both states.

Identification of chromatin remodelling complexes involved in the stepwise transition from stem cell dormancy, activation and onset of differentiation: We have identified several epigenetic factors whose expression dynamically changes along the axis of dormancy, activation and differentiation. Interestingly, each factor shows a unique pattern of expression and activity. We are validating the role of several of these factors using in vivo models and highthroughput molecular methodologies.

2. Studying Novel Pathways Relevant to Epidermal and Squamous Tumour Onset and Progression

Adult stem cells are potentially the few long term tissue residents that in time may accumulate enough somatic oncogenic mutations which result in the development of neoplasias. Moreover, the behaviour and molecular signature of a small percentage of cancer cells, known as cancer stem cells, recapitulate those of adult stem cells in the normal tissue. Cancer stem cell self-renewal, high potential of invasion and homing into a specific niche, with direct consequences over tumour maintenance and metastasis, are most probably characteristics inherited from normal adult stem cells. However, very little is known about the signalling events and the molecular signature that contribute to the behaviour of cancer stem cells in tumours of epithelial origin.

In collaboration with the Hospital del Mar we are obtaining samples of SCCs (fresh live tissue, and blocks for immunohistological analysis) at different stages of tumour progression. Squamous cell carcinomas are the most diagnosed types of tumours in western countries with poor prognosis when developed in the oral cavitiy. We are analyzing the status of the different pathways studied in the lab with respect to their possible role in squamous cancer stem cells and validating the results using our mouse models.

PUBLICATIONS

Spagnoli FM, Besser D, Benitah SA.
"Where do you come from; where do you go? Pluripotency, differentiation and malfunction of stem cells."
EMBO Rep, 13(1):9-11 (2011).

Blanco S, Kurowski A, Nichols J, Watt FM, Benitah SA, Frye M.
"The RNA-methyltransferase Misu (NSun2) poises epidermal stem cells to differentiate."
PLoS Genet, 7(12):e1002403 (2011).

Nascimento EM, Cox CL, MacArthur S, Hussain S, Trotter M, Blanco S, Suraj M, Nichols J, Kübler B, Benitah SA, Hendrich B, Odom DT, Frye M.
"The opposing transcriptional functions of Sin3a and c-Myc are required to maintain tissue homeostasis."
Nat Cell Biol, 13(12):1395-405 (2011):

Janich P, Pascual G, Merlos-Suárez A, Batlle E, Ripperger J, Albrecht U, Cheng HY, Obrietan K, Di Croce L, Benitah SA.
"The circadian molecular clock creates epidermal stem cell heterogeneity."
Nature, 480(7376):209-14 (2011).

Uribesalgo I, Buschbeck M, Gutiérrez A, Teichmann S, Demajo S, Kuebler B, Nomdedéu JF, Martín-Caballero J, Roma G, Benitah SA*, Di Croce L*.
"E-box-independent regulation of transcription and differentiation by MYC."
Nat Cell Biol, 13(12):1443-9 (2011).
(*co-corresponding authors)

Benitah SA.
"Tumour biology: Skin-cancer stem cells outwitted."
Nature, 478(7369):329-30 (2011).

Luis NM, Morey L, Mejetta S, Pascual G, Janich P, Kuebler B, Cozutto L, Roma G, Nascimento E, Frye M, Di Croce L, Benitah SA.
"Regulation of human epidermal stem cell proliferation and senescence requires polycomb- dependent and -independent functions of Cbx4."
Cell Stem Cell, 9(3):233-46 (2011).
Erratum in: Cell Stem Cell, 9(5):486, Cozutto, Luca [added] (2011).

Mejetta S, Morey L, Pascual G, Kuebler B, Mysliwiec MR, Lee Y, Shiekhattar R, Di Croce L, Benitah SA.
"Jarid2 regulates mouse epidermal stem cell activation and differentiation."
EMBO J, 30(17):3635-46 (2011).

Hunziker L, Benitah SA, Braun KM, Jensen K, McNulty K, Butler C, Potton E, Nye E, Boyd R, Laurent G, Glogauer M, Wright NA, Watt FM, Janes SM.
"Rac1 deletion causes thymic atrophy."
PLoS One, 6(4):e19292 (2011).
Erratum in: PLoS One, 6(6) (2011). doi:10.1371/annotation/5c665e96-0303-438f-afb0-c257c3481692.
Aznar Benitah, Salvador [corrected to Benitah, Salvador Aznar].

DIFFERENTIATION AND CANCER

Group: Mechanisms of Cancer and Aging

Group Structure:

Group Leader:	Bill Keyes
Postdoctoral Researcher:	Jason Doles (EMBO Long Term Fellowship)
PhD Students:	Valeria Di Giacomo ("la Caixa"), Matteo Pecoraro (Plan Nacional / FPI), Mekayla Storer ("la Caixa")
Technician:	Alba Mas (CRG)

SUMMARY

Increasing evidence shows how the processes of cancer and aging are intimately linked, sharing many common molecular and cellular mediators. Importantly, the deregulation of normal stem cell proliferation is emerging as a central event in both processes. Impaired stem cell proliferation is suggested as a primary mediator of the aging process, while deregulated stem cell proliferation is linked with cancer initiation. Similarly, the process of cellular senescence is suggested as a primary cause of the aging process, likely through inhibiting stem cell proliferation, while impaired senescence is a critical component of tumor initiation. ***Our work is interested in understanding how stem cells respond to and cope with oncogene- and age- induced stress, and in uncovering the molecular mechanisms by which a deregulation of processes like stem cell homeostasis and cellular senescence plays a causative role in cancer and aging.***

RESEARCH PROJECTS

1. Investigating the Role of p63 and Aberrant Stem Cell Proliferation in the Pathogenesis of Squamous Cell Carcinoma

Squamous cell carcinoma (SCC) is one of the most frequent solid tumors worldwide representing the second highest cause of skin cancer, one third of lung cancers, and in the case of head and neck SCC, the sixth most common solid tumor type. Although treatable if detected early, SCC presents with a high mortality due to resistance to treatment and tumor recurrence. Recently we identified ΔNp63α as an oncogene that is capable of inducing the development of SCC. Although this isoform of p63 is frequently overexpressed in human SCC, a causative role in tumor development had not been shown. In studies using primary mouse keratinocytes and nude mouse models, we found that ΔNp63α promotes SCC by inhibiting the process of oncogene-induced senescence. Surprisingly however, we found that tumor initiation also involved the aberrant proliferation of epidermal stem cells and the propagation of cells with stem-like properties including, an ability to form self-renewing spheres in 3D-tissue culture, a capacity to differentiate and a resistance to DNA-damaging drugs.

Within many human tumors, including SCC, populations of cells exhibiting properties of stem cells have been identified. These cancer stem cells possess intrinsic growth properties that favor tumor development, resistance to treatment and tumor recurrence after treatment. Such properties include a capacity for self-renewal, an ability to differentiate and an inherent resistance to DNA-damage. However, the origin of these tumorigenic therapy-resistant cells and their mechanisms of proliferation are unknown in many tumor types. Understanding the processes that favour the proliferation of these cells is necessary to design more effective therapies for many cancers. In this project we propose to investigate how ΔNp63α promotes tumor development, focusing on identifying the mechanisms involved and the processes by which aberrant stem cell proliferation may favour tumor development.

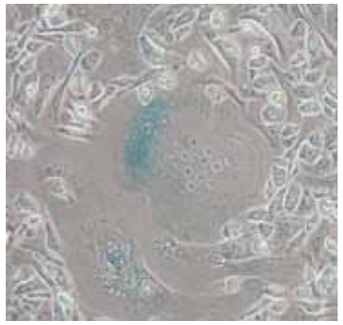


Figure 1.
Senescence-associated β-galactosidase staining identifies senescent cells (blue stain). Note the characteristic morphology, including enlarged flat shape and multiple nuclei. Multiple smaller transformed cells surround.

2. Determining the Function of p63 in Prostate Stem Cells and Prostate Tumor Development

In prostate cancer, the role of p63 is unknown and controversial. Unlike the overexpression of p63 that is seen in SCC, during the development of adenocarcinoma, the most common prostate tumor type, p63 expression is actually lost from the cells that are undergoing malignant conversion. Indeed it has been suggested that it is the p63-positive stem cells that undergo malignant transformation during prostate tumor initiation. However, it is not known if this loss of p63 facilitates, or is required for prostate tumor development, or whether there is a shift in the ratio of expression of p63 isoforms during transformation. To further complicate the situation, loss of expression of p63 has been correlated with the upregulation of genes that are involved in epithelial-to-mesenchymal transformation and metastasis. In prostate tumors, the development of recurrent treatment-resistant tumors that undergo metastasis is the main cause of death from these tumors. By taking a multidisciplinary approach, using functional genetics, *in vivo* animal models and high-throughput genomic screens, it is hoped that this work will identify key genes and mechanisms during tumor development that can be targeted for therapy in future studies.

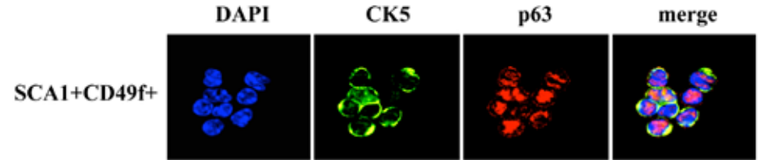


Figure 2.
Stems cells purified from the mouse prostate using antibodies against SCA1 and CD49f express high levels of p63.

3. Novel Pathways Linking Cancer and Aging

Stem cell proliferation must be tightly regulated to allow for proper development and tissue homeostasis, and to respond to stresses such as aging and cancer. However, we are far from having a complete understanding of the mechanisms by which stem cells respond to and compensate for any alteration in the tissue environment, how this may change between different stem cell and tissue types, or how this alters over the course of the lifespan of an organism. We are interested in elucidating the dynamic mechanisms and processes that regulate stem cells in such conditions.

In our ongoing investigations, we have identified the Lsh/Hells gene as a novel p63-target that also links cancer and aging. Lsh is a member of the SNF2-family of chromatin remodelers that is involved in promoting DNA methylation and transcriptional silencing, through recruitment of DNA-methyltransferases or direct interaction with members of the Polycomb-repressive complex-1 (PRC1). Interestingly, mouse models deficient for Lsh exhibit premature aging and enhanced cellular senescence. In our studies, we find that Lsh is overexpressed in some tumors, while a deficiency of Lsh prevents senescence bypass and aberrant tumor-initiating cellular proliferation. We are currently investigating the function of Lsh and other candidates in stem-cell homeostasis, tumor initiation and aging.

PUBLICATIONS

Keyes WM*, Pecoraro M, Aranda V, Vernersson-Lindahl E, Li W, Vogel H, Guo X, Garcia EL, Michurina TV, Enikolopov G, Muthuswamy SK and Mills AA.*
"ΔNp63α is an oncogene that targets chromatin remodeler Lsh to drive skin stem cell proliferation and tumorigenesis."
Cell Stem Cell, 8(2):164-176 (2011).
(*corresponding author)

GENES AND DISEASE

Group: Genetic Causes of Disease

Group Structure:

Group Leader: Xavier Estivill

Staff Scientist: Eulàlia Martí

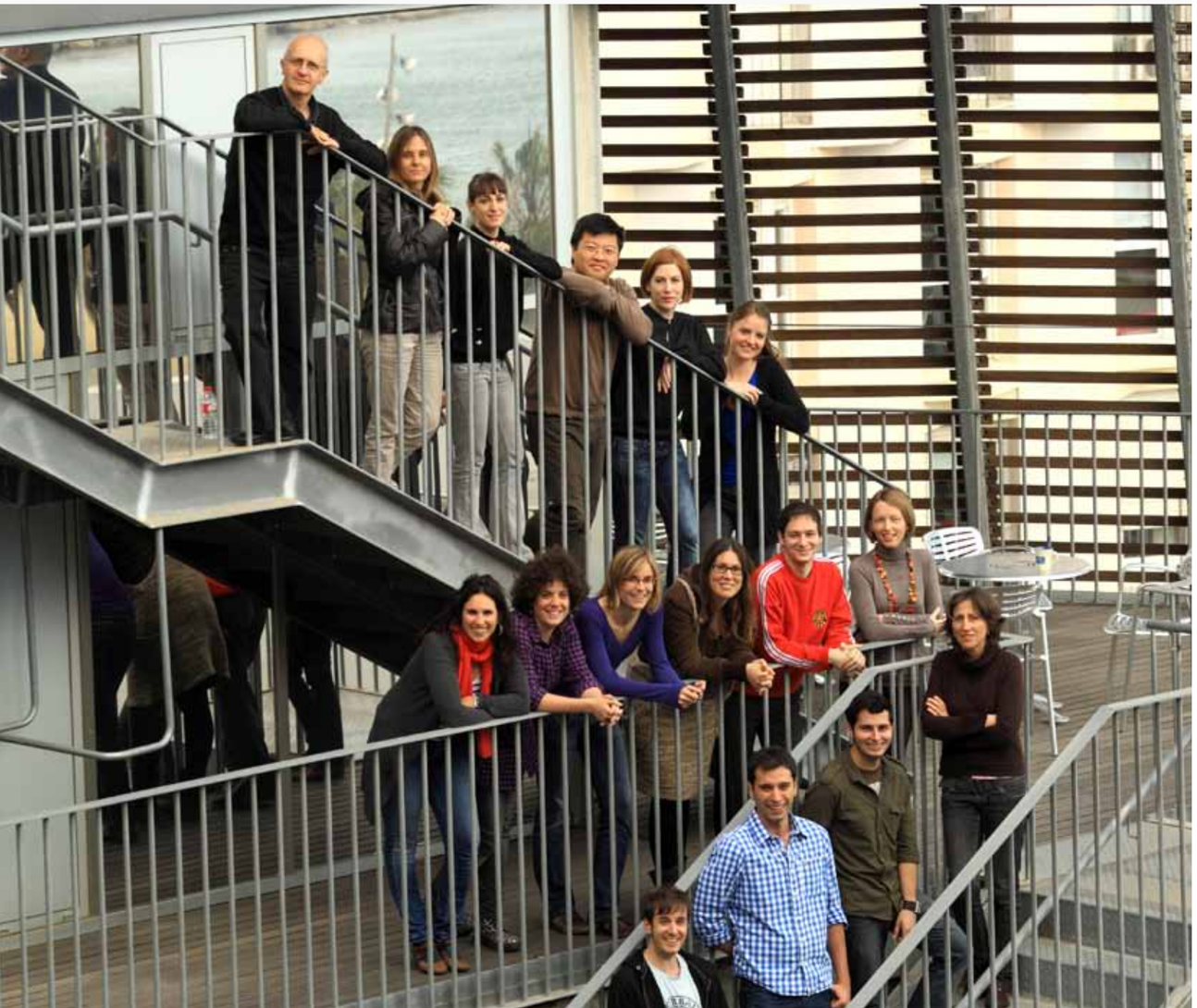
Scientific Officer: Mònica Gratacòs (since March 2011)

Postdoctoral Fellows: Mónica Bañez-Coronel, Mariona Bustamante (CREAL), Georgia Escaramís (CIBERESP), Marc Friedlander, Hyun Hor (since March 2011), Charlotte Hor-Henrichsen (since March 2011), Esther Lizano, Raquel Rabionet (since March 2011), José Manuel C. Tubío

PhD Students: Johanna Aigner, Laia Bassaganyas, Elisa Docampo, Susana Iraola (until December 2011), Elisabet Mateu, Elena Miñones (until August 2011), Lorena Pantano (until May 2011), Daniel Trujillano, Nàdia Vilahur (CREAL)

Technicians: Justo González (since April 2011), Birgit Kagerbauer, Marta Morell, Anna Puig, Cristian Tornador, Sergi Villatoro (until May 2011)

Visiting Scientists: Jean-Jacques Schott (University of Nantes, France), Tim Spector (King's College London, UK)



SUMMARY

We are interested on how different types of genetic variants (single nucleotide polymorphisms, SNPs; structural variations, mainly copy number variants or CNVs, and insertion/deletion variants, including transposable elements) influence human disorders. We are also interested in studying the contribution of epigenetic modifications and non-coding RNA pathways in complex diseases, mainly neuropsychiatric and neurodegenerative disorders. We are approaching these questions through genomic approaches that interrogate SNPs, CNVs and methylation, including high-throughput sequencing for the characterization of the exome and transcriptome. Besides, we are developing functional studies to address the consequences of genetic and epigenetic changes in physiology and pathology. The interaction of genetic factors with different environmental conditions is one of the main areas of research of the group. Finally, we are interested in the role of non-coding RNAs in human disease, in particular in neurodegenerative disorders.

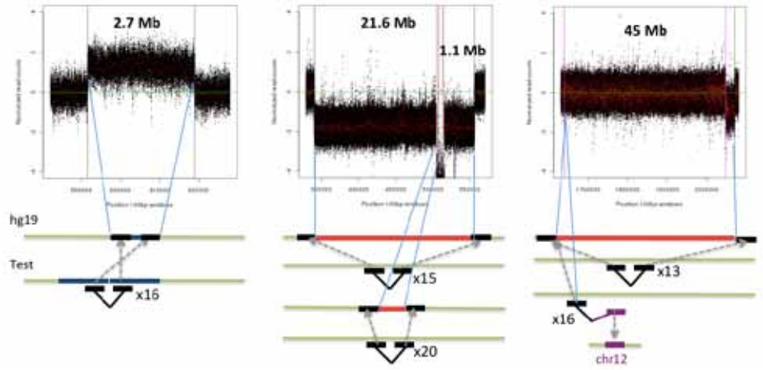
RESEARCH PROJECTS

1. Structural Variations, Environment and Disease

J Aigner, L Bassaganyas, M Bustamante, G Escaramís, J González, C Hor-Henrichsen, R Rabionet, C Tornador, JM C Tubío, N Vilahur, S Villatoro

Structural variations (SV) are genomic regions that contain copy number variants (CNV), inversions, segmental duplications or transposable sequences, and are potentially involved in phenotype variation, including disease. Our group is part of the chronic lymphocytic leukaemia (CLL) International Cancer Genomics Consortium (ICGC), with the aim to fully characterize the genome of CLL cells. We have developed new tools for the analysis of SV and non-coding RNAs and have characterized the genome and transcriptome of several cases of this common type of leukaemia. To identify SV from whole-genome sequence and whole-exome sequence data we have developed the PeSV-Fisher pipeline, which includes paired-end mapping (PEM) and depth of coverage (DOC) strategies to identify structural variations (Escaramis *et al*. unpublished). By identifying read pairs that do not map correctly with respect to each other on the reference genome, we are able to identify potential somatic rearrangements, as well as constitutional genomic changes. High-confidence somatic rearrangements can be confirmed *in silico* by split-reads in predicted breakpoints. The pipeline includes a module FinalCountDown (Escaramis *et al*. unpublished) that allows filtering high-confidence potential events after removing rearrangements with breakpoints involving complex regions of the genome (i.e., segmental duplications, simple tandem repeats, and low-divergent transposable elements). We are applying a new method for CNV detection from multi-sample Exome-seq data, based on comparison of read coverage in enriched regions between several samples. Our method is able to detect heterozygous deletions and duplications as well as higher order genome amplifications, even if more than two alleles exist in the population. We have contributed to the whole-genome sequencing of four CLL cases (Puente *et al*. *Nature* 2011) and whole-exome sequencing of 105 cases (Quesada *et al*. *Nature Genetics* 2011). The analysis of SV in the CLL project will provide gold standard information on the spectrum of SV of the human genome of importance for other human diseases and traits (Figure 1).

Figure 1.



Several other projects evaluate the potential role of structural variations in disease. The group has made specific progress in psoriasis. We have identified a common CNV that involves the deletion of two genes (LCE3B and LCE3C) in patients with psoriasis. The absence of these two genes in psoriatic patients shows that skin barrier alterations play a role in the disease. This association, replicated in several populations, is one of the main genetic variants associated to psoriasis. We have found that patients with other autoimmune disorders, such as rheumatoid arthritis and psoriatic arthritis, also have a higher frequency of LCE3B and LCE3C deletions. We have compared several brain areas (cortex, hippocampus, *substantia nigra*, amygdala and hypothalamus) to the cerebellum of control subjects and of patients affected by Alzheimer's, and has shown several areas that with potential copy number changes. Finally, we have identified a common 45-kb deletion that affects two genes of the same family, BTNL8 and BTNL3, and leads to the formation of a novel fusion transcript and the subsequent down-regulation of the expression level of the neighbouring gene BTNL9. We have found that over-expression of BTNL9 inhibits glucocorticoid-induced apoptosis, suggesting that it has a pro-survival function.

The group collaborates with investigators of the CREAL (Centre for Research in Environmental Epidemiology) to study gene-environment interaction in several phenotypes. The INMA (Infancia y Medio Ambiente) is a network of birth cohorts in Spain with more than 2000 children-mother pairs having DNA available. Approximately 1000 child samples have been genome-wide genotyped. Within the Early Genetics Consortium (EGG) and The EArly Genetics and Lifecourse Epidemiology (EAGLE) consortia, INMA has participated in more than 10 GWAS metaanalysis.

The group is performing whole-exome sequencing of individuals from families that segregate Mendelian diseases. In collaboration with one of the research visitors at our lab (Dr. Jean Jacques Schott) we have identified a gene responsible for hypertension and hyperkalemia (Louis-Dit-Picard *et al*. *Nature Genetics*, in press). Several other disorders (cardiac electric conduction, idiopathic juvenile arthritis, cystic fibrosis, and mitral valve prolapse) are currently under investigation.

2. Genetic Variants Associated to Psychiatric Disorders

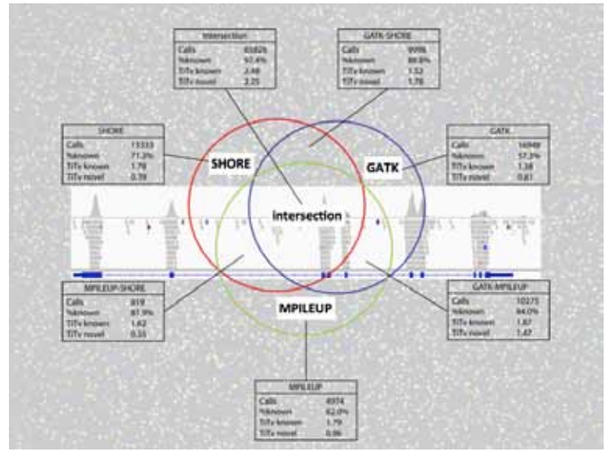
E Docampo, M Gratacòs, H Hor, M Morell, A Puig, D Trujillano

The group has developed activities in the study of substance abuse disorder, anxiety disorders (panic disorder and obsessive-compulsive disorder), eating disorders and affective disorders. In most cases, the methodology involved the selection of variants of SNPs in candidate genes to be involved in the pathophysiology of the disease. We have studied the route of melatonin or genes that regulate circadian rhythm in patients with affective disorders, the endocannabinoid pathway in eating disorders and panic disorder, or galanin, estrogen and glutamate pathways in obsessive compulsive disorder (OCD). We have explored as well the influence of stressing life events and its integration with genetic variants in the development of OCD and its response to pharmacotherapy. In addition, we selected samples of patients with different addictions to perform an association study targeting gene family of neurexins to assess the possible involvement of variants in these genes in addictive processes. We have explored the genetic susceptibility to postpartum depression and the correlation between hormone levels and the occurrence of this disorder. In addition to these lines of research, the group has established international partnerships to replicate the results of other investigations and contribute to studies of GWAS samples.

One of the research lines of the lab has involved the identification of genetic risk factors for fibromyalgia and chronic fatigue syndrome. We have performed a genome-wide association study of 300 fibromyalgia cases and validating findings in a larger cohort of 1000 fibromyalgia cases and 1000 controls. In addition, we are performing whole-exome sequencing of fibromyalgia cases, including some sibpairs, in order to identify additional rare variants potentially involved in disease.

The group has studied the role of miRNAs and CNVs in susceptibility to psychiatric disorders. We have examined, for example, a CNV in obsessive-compulsive disorder, bipolar and anxiety disorders, and the involvement of miRNAs in anxiety disorders and insomnia associated with bipolar disorder. We have carried out functional studies for several miRNAs and have identified interesting regulation pathways for genes expressed in the central nervous system. Finally, in an effort to find both rare and common variants involved in OCD, we have sequenced the exome of 40 patients with a severe and early onset of the disorder, and the candidate variants shared among patients will be compared to the exomes of control individuals (Figure 2).

Figure 2.



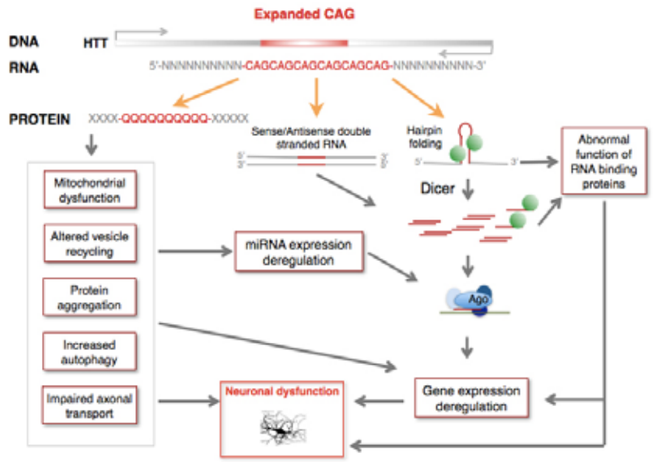
3. Functional Genomics of Neurological Disorders

M Bañez-Coronel, M Friedländer, B Kagerbauer, S Iraola, E Lizano, E Martí, E Mateu, E Miñones, L Pantano

We have characterized non-coding small RNAs, mainly miRNAs, in brain samples from individuals with Huntington disease (HD) and Parkinson disease (PD) and individuals without neuropathology. We evaluated the functional consequences of miR-34b/c downregulation in PD and have found that miR34b/c deficiency induces oxidative stress, mitochondrial dysfunction and loss of ATP cell content. These results suggest that the early downregulation of miR-34b/c contribute to key neuropathological aspects in PD. We have characterized the expression of a novel small-RNA (sRNA) derived from a vault RNA (svRNA2_1a) in brain areas of PD showing a different degree of neuropathological affectation. The present data suggest that svRNA2_1a is a new miRNA-like sRNA, deregulated early in PD that could contribute to early pathology.

The group is also studying the role of small non-coding RNAs in neurodegeneration associated to triplet repeat expansion diseases such as HD and fragile X-associated tremor ataxia syndrome (FXTAS). The group is evaluating the role of CAG and CGG repeated RNA in generating toxic effects in neurons. We have found that expanded CAG in the Huntington gene (HTT) mRNA is neurotoxic though a mechanism involving the biogenesis and activity of short repeated CAG RNAs (sCAG). sCAG may act as miRNAs and use the RNA interference machinery to trigger detrimental effects. Our results suggest a new RNA-dependent pathogenic mechanism in HD that may complement pathogenesis associated with mutant/expanded HTT protein (Figure 3).

Figure 3.



PUBLICATIONS

Articles

Quesada V, Conde L, Villamor N, Ordóñez GR, Jares P, Bassaganyas L, Ramsay AJ, Beà S, Pinyol M, Martínez-Trillos A, López-Guerra M, Colomer D, Navarro A, Baumann T, Aymerich M, Rozman M, Delgado J, Giné E, Hernández JM, González-Díaz M, Puente DA, Velasco G, Freije JM, Tubío JM, Royo R, Gelpí JL, Orozco M, Pisano DG, Zamora J, Vázquez M, Valencia A, Himmelbauer H, Bayés M, Heath S, Gut M, Gut I, Estivill X, López-Guillermo A, Puente XS, Campo E, López-Otín C.
"Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia."
Nat Genet, 44(1):47-52 (2011).

Fullana MA, Alonso P, Gratacòs M, Jaurrieta N, Jiménez-Murcia S, Segalàs C, Real E, Estivill X, Menchón JM.
"Variation in the BDNF Val66Met polymorphism and response to cognitive-behavior therapy in obsessive-compulsive disorder."
Eur Psychiatry, [Epub ahead of print].

Nicchia GP, Ficarella R, Rossi A, Giangreco I, Nicolotti O, Carotti A, Pisani F, Estivill X, Gasparini P, Svelto M, Frigeri A.
"D184E mutation in aquaporin-4 gene impairs water permeability and links to deafness."
Neuroscience,197:80-8 (2011).

Gallego X, Molas S, Amador-Arjona A, Marks MJ, Robles N, Murtra P, Armengol L, Fernández-Montes RD, Gratacòs M, Pumarola M, Cabrera R, Maldonado R, Sabrià J, Estivill X, Dierssen M.
"Overexpression of the CHRNA5/A3/B4 genomic cluster in mice increases the sensitivity to nicotine and modifies its reinforcing effects."
Amino Acids, [Epub ahead of print].

Pantano L, Estivill X, Martí E.
"A non-biased framework for the annotation and classification of the non-miRNA small RNA transcriptome."
Bioinformatics, 27(22):3202-3 (2011).

Rebordosa C, Kogevinas M, Guerra S, Castro-Giner F, Jarvis D, Cazzoletti L, Pin I, Siroux V, Wjst M, Antò JM, de Marco R, Estivill X, Corsico AG, Nielsen R, Janson C.
"ADRB2 Gly16Arg polymorphism, asthma control and lung function decline."
Eur Respir J, 38(5):1029-35 (2011).

Jacquemont S *et al*. (including Gratacos M, Villatoro S, Estivill X).
"Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus."
Nature, 478(7367):97-102 (2011).

Miñones-Moyano E, Porta S, Escaramís G, Rabionet R, Iraola S, Kagerbauer B, Espinosa-Parrilla Y, Ferrer I, Estivill X, Martí E.
"MicroRNA profiling of Parkinson's disease brains identifies early downregulation of miR-34b/c which modulate mitochondrial function."
Hum Mol Genet, 20(15):3067-78 (2011).

Docampo E, Giardina E, Riveira-Muñoz E, de Cid R, Escaramís G, Perricone C, Fernández-Sueiro JL, Maymó J, González-Gay MA, Blanco FJ, Hüffmeier U, Lisbona MP, Martín J, Carracedo A, Reis A, Rabionet R, Novelli G, Estivill X.
"Deletion of LCE3C and LCE3B is a susceptibility factor for psoriatic arthritis: a study in Spanish and Italian populations and meta-analysis."
Arthritis Rheum, 63(7):1860-5 (2011).

Riveira-Munoz E, He SM, Escaramís G, Stuart PE, Hüffmeier U, Lee C, Kirby B, Oka A, Giardina E, Liao W, Bergboer J, Kainu K, de Cid R, Munkhbat B, Zeeuwen PL, Armour JA, Poon A, Mabuchi T, Ozawa A, Zawirska A, Burden AD, Barker JN, Capon F, Traupe H, Sun LD, Cui Y, Yin XY, Chen G, Lim HW, Nair RP, Voorhees JJ, Tejasvi T, Pujol R, Munkhtuvshin N, Fischer J, Kere J, Schalkwijk J, Bowcock A, Kwok PY, Novelli G, Inoko H, Ryan AW, Trembath RC, Reis A, Zhang XJ, Elder JT, Estivill X.
"Meta-analysis confirms the LCE3C_LCE3B deletion as a risk factor for psoriasis in several ethnic groups and finds interaction with HLA-Cw6."
J Invest Dermatol, 131(5):1105-9 (2011).

Alonso P, Gratacòs M, Segalàs C, Escaramís G, Real E, Bayés M, Labad J, Pertusa A, Vallejo J, Estivill X, Menchón JM.
"Variants in estrogen receptor alpha gene are associated with phenotypical expression of obsessive-compulsive disorder."
Psychoneuroendocrinology, 36(4):473-83 (2011).

Muiños-Gimeno M, Espinosa-Parrilla Y, Guidi M, Kagerbauer B, Sipilä T, Maron E, Pettai K, Kananen L, Navinés R, Martín-Santos R, Gratacòs M, Metspalu A, Hovatta I, Estivill X.
"Human microRNAs miR-22, miR-138-2, miR-148a, and miR-488 are associated with panic disorder and regulate several anxiety candidate genes and related pathways."
Biol Psychiatry, 69(6):526-33 (2011).

Morales E, Bustamante M, Gonzalez JR, Guxens M, Torrent M, Mendez M, Garcia-Esteban R, Julvez J, Forns J, Vrijheid M, Molto-Puigmarti C, Lopez-Sabater C, Estivill X, Sunyer J.
"Genetic variants of the FADS gene cluster and ELOVL gene family, colostrums LC-PUFA levels, breastfeeding, and child cognition."
PLoS One, 6(2):e17181 (2011).

Morales E, Guerra S, García-Esteban R, Guxens M, Alvarez-Pedrerol M, Bustamante M, Estivill X, Antó JM, Sunyer J.
"Maternal C-reactive protein levels in pregnancy are associated with wheezing and lower respiratory tract infections in the offspring."
Am J Obstet Gynecol, 204(2):164.e1-9 (2011).

Albacar G, Sans T, Martín-Santos R, García-Esteve L, Guillamat R, Sanjuan J, Cañellas F, Gratacòs M, Cavalle P, Arija V, Gaviria A, Gutiérrez-Zotes A, Vilella E.
"An association between plasma ferritin concentrations measured 48 h after delivery and postpartum depression."
J Affect Disord, 131(1-3):136-42 (2011).

Gilabert-Juan J, Ivorra JL, Tolosa A, Gratacòs M, Costas J, Sanjuán J, Moltó MD.
"Potential involvement of serotonin receptor genes with age of onset and gender in schizophrenia: a preliminary study in a Spanish sample."
Psychiatry Res, 186(1):153-4 (2011).

Tubio JM, Tojo M, Bassaganyas L, Escaramis G, Sharakhov IV, Sharakhova MV, Tornador C, Unger MF, Naveira H, Costas J, Besansky NJ.
"Evolutionary dynamics of the Ty3/gypsy LTR retrotransposons in the genome of Anopheles gambiae."
PLoS One, 6(1):e16328 (2011).

Karwautz AF, Wagner G, Waldherr K, Nader IW, Fernandez-Aranda F, Estivill X, Holliday J, Collier DA, Treasure JL.
"Gene-environment interaction in anorexia nervosa: relevance of non-shared environment and the serotonin transporter gene."
Mol Psychiatry, 16(6):590-2 (2011).

Peñas-Lledó EM, Dorado P, Agüera Z, Gratacós M, Estivill X, Fernández-Aranda F, Llerena A.
"High risk of lifetime history of suicide attempts among CYP2D6 ultrarapid metabolizers with eating disorders."
Mol Psychiatry, 16(7):691-2 (2011).

Puente XS *et al*. (including Escaramis G, Bassaganyas L, Tubío JM, Tornador C, Estivill X).
"Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia."
Nature, 475(7354):101-5 (2011).

Gelabert E, Subirà S, García-Esteve L, Navarro P, Plaza A, Cuyàs E, Navinés R, Gratacòs M, Valdés M, Martín-Santos R.
"Perfectionism dimensions in major postpartum depression."
J Affect Disord, 136(1-2):17-25 (2011).

(…) Estivill X, (…), Gratacòs M, *et al*.
"Fat mass and obesity associated gene (FTO) in eating disorders: association of the rs9939609 obesity risk allele with bulimia nervosa and anorexia nervosa."
BMC Medical Genetics, in press.

Llorens F, Hummel M, Pastor X, Ferrer A, Pluvinet R, Vivancos A, Castillo E, Iraola S, Mosquera AM, González E, Lozano J, Ingham M, Dohm JC, Noguera M, Kofler R, del Río JA, Bayés M, Himmelbauer H, Sumoy L.
"Multiple platform assessment of the EGF dependent transcriptome by microarray and deep tag sequencing analysis."
BMC Genomics, 12:326 (2011).

Vrijheid M, Martinez D, Aguilera I, Bustamante M, Ballester F, Estarlich M, Fernandez-Somoano A, Guxens M, Lertxundi N, Martinez MD, Tardon A, Sunyer J; INMA Project.
"Indoor air pollution from gas cooking and infant neurodevelopment."
Epidemiology, [Epub ahead of print]

Review

Iraola-Guzmán S, Estivill X, Rabionet R.
"DNA methylation in neurodegenerative disorders: a missing link between genome and environment?"
Clin Genet, 80(1):1-14 (2011).

News and Views

Tubio JM, Estivill X.
"Cancer: When catastrophe strikes a cell."
Nature, 470(7335):476-7 (2011).

Book chapter

Ballana E and Estivill X.
"Genética y genómica de las deficiencias auditivas."
In: *Tratado de Audiología*, 2ª Ed., 20:1-11, in press.

GENES AND DISEASE

Group: Gene Therapy

Group Structure:

Group Leader: Cristina Fillat

Postdoctoral Fellows: Maria Victoria Maliandi, Luciano Sobrevals

PhD Students: Anabel Jose, Xavier Bofill, Ana Mato, Eneko Villanueva

Technicians: Núria Andreu (until July)

SUMMARY

Our laboratory focuses on the development of gene therapy strategies to understand and treat human complex genetic diseases. The advances on the basic biological understanding of molecular and cellular events underlying specific disease pathophysiology, facilitates the development of gene therapy in a broad spectrum of human diseases. Efficacy of gene therapy application has already been proven in clinical trials for several rare genetic diseases and in the future it can be foreseen as a medicine to treat patients on an individual basis. Gene therapy has also a great potential as a powerful genetic tool to contribute to the identification of the functional role of specific genes and/or small non-coding elements in certain conditions. Our research interests are in genetic engineering viral vectors that can selectively and efficiently target specific cell types and to study their impact in living animals. Pharmacokinetic and pharmacodynamic studies are being conducted in specific disease mouse models to evaluate the therapeutic response or selected phenotypic rescue.

RESEARCH PROJECTS

1. Pancreatic Cancer

Pancreatic cancer is a neoplasia with a very bad prognosis mainly due to the late diagnosis and inefficient current therapies. We take advantage of the current understanding of the mechanistic pathogenesis and of the comprehensive available data regarding inefficient therapeutic response to design alternative treatments. In the past few years we have been involved in exploring the feasibility of suicide gene-virotherapy. This system is based on the combination of two different principles, based on the attempt to achieve tumor destruction by allowing lytic adenovirus to replicate in tumor cells, multiply and disseminate, and by arming these viruses with the Herpes Simplex Thymidine Kinase gene (TK) that when combined with GCV induces apotosis mediated cell killing. This year we have gained insight in the understanding of the basic mechanisms of tumor cell killing induced by the TK/GCV system. We demonstrated the implication of E-cadherin in the modulation of the gap-junction mediated TK/GCV bystander effect. Interestingly we have also shown the key role of connexin-26 in the bystander effect of gemcitabine toxic metabolites in pancreatic adenocarcinoma. We have shown that TK-expressing oncolytic adenoviruses can be traced by PET imaging providing real time information on the activity of the virus. We have proved to boost the potency of adenoviruses when TK was expressed at the late phase of viral replication and we are actively working on providing tumor selectivity to those viruses. Post-transcriptional, transcriptional and viral capsid modification strategies are under development to confer selective tumor control to adenoviral replication. Moreover we have been able to demonstrate the feasibility of irreversible electroporation as an antitumoral treatment in mouse models of pancreatic cancer.

2. Down Syndrome

Down syndrome (DS) is a complex multi-system disorder, resulting from an extra copy of the human chromosome 21 gene HSA21. Although HSA21 contains more than three hundred genes, it has been hypothesized that the excess of specific target genes (dosage-sensitive) will have a major impact on the DS phenotype. To better understand the contribution of these specific genes, we have engineer a novel approach that takes advantage of the current knowledge from the existing mouse models. This approach is based on the rescue of a phenotype in a mouse model by the viral delivery of shRNA sequences targeting a candidate gene. A proof of principle approach was developed and shown that normalizing the expression of the Dyrk1A dosage sensitive gene in the striatum of adult TgDyrk1A by AAVshDyrk1A delivery rescued motor impairment. This approach highlighted Dyrk1A phenotypic dependence in motor alterations. Mental retardation is the more disabling trait in DS. We have preliminary evidences indicating that Dyrk1A normalization in the hippocampus attenuate specific hippocampal-dependent defects. Studies on the molecular mechanisms by which the normalization of Dyrk1A can reverse such phenotypes are also being explored. This year we have also shown that over-expression of the HSA21 gene RCAN1 in a transgenic mouse results in marked cognitive defects. Despite the contribution triggered by the dosage-sensitive genes to the phenotype the over-representation of functional non-protein-coding elements might be involved in some of the abnormal phenotypes. On going efforts involve the development of viral gene-transfer approaches that help to evaluate the contribution of these non-coding elements to the cognitive phenotype of DS murine models.

PUBLICATIONS

Abate-Daga D, Andreu N, Camacho-Sánchez J, Alemany R, Herance R, Millán O, Fillat C.
"Oncolytic adenoviruses armed with thymidine kinase can be traced by PET imaging and show potent antitumoural effects by ganciclovir dosing."
PLoS One, 6(10):e26142 (2011).

Garcia-Rodríguez L, Pérez-Torras S, Carrió M, Cascante A, García-Ribas I, Mazo A, Fillat C.
"Connexin-26 is a key factor mediating gemcitabine bystander effect."
Mol Cancer Ther, 10(3):505-17 (2011).

Fillat C, Jose A, Bofill-De Ros X, Mato-Berciano A, Maliandi MV, Sobrevals L
"Pancreatic Cancer Gene Therapy: From Molecular Targets to Delivery Systems"
Cancers, 3:368-395 (2011).

Garcia-Rodríguez L, Abate-Daga D, Rojas A, González JR, Fillat C.
"E-cadherin contributes to the bystander effect of TK/GCV suicide therapy and enhances its antitumoral activity in pancreatic cancer models."
Gene Ther,18(1):73-81 (2011).

Dierssen M, Arqué G, McDonald J, Andreu N, Martínez-Cué C, Flórez J, Fillat C.
"Behavioral characterization of a mouse model overexpressing DSCR1/ RCAN1."
PLoS One, 25;6(2):e17010 (2011).

José A, Sobrevals L, Ivorra A, Fillat C.
"Irreversible electroporation shows efficacy against pancreatic carcinoma without systemic toxicity in mouse models."
Cancer Lett, Epub 2011 Nov 10.

Martinez de Lagran M, Benavides-Piccione R, Ballesteros-Yañez I, Calvo M, Morales M, Fillat C, Defelipe J, Ramakers GJ, Dierssen M.
"Dyrk1A Influences Neuronal Morphogenesis Through Regulation of Cytoskeletal Dynamics in Mammalian Cortical Neurons."
Cereb Cortex, [Epub ahead of print]

Book Chapters

Fillat C, Arán JM.
"Tratamiento de la Enfermedades Genéticas y Modelos animales."
In: *Medicina Interna*, Farreras-Rotzman XVI Ed., in press.

GENES AND DISEASE

Group: Gene Function and Murine Models of Disease
Susana de la Luna is an ICREA Research Professor.

Group Structure:

Group Leaders: Mariona Arbonés* /Susana de la Luna

Postdoctoral: Krisztina Arató, María José Barallobre* (CIBERER researcher), Esteban Rozen

PhD Students: Elisa Balducci*, Chiara Di Vona, Sonia Najas *, Julia Rowenstrunk (since October 2011), Andrea Senna (until October 2011)

Technician: Alicia Raya

(*) until October 2011

SUMMARY

The ability of a human being to perform complex tasks relies on the correct formation of neuronal circuits. For this to happen, the many cell types of a mature brain have to be generated at the correct numbers and then differentiate properly during development, a complex process that is exquisitely regulated in time and space. Mutations disturbing the division mode of neuronal precursor cells, cell fate acquisition, differentiation, or natural cell death are likely to cause alterations in brain connectivity resulting in neurological diseases and mental retardation.

Trisomy 21 is the main genetic cause of mental retardation, resulting from neurodevelopmental alterations and also changes in brain homeostasis in the adult. Research in the past few years has revealed that a number of human chromosome 21 (HSA21) genes are overexpressed in Down syndrome (DS) by at least 50% due to gene dosage. Because of the complexity of the DS phenotype, it is very likely that the increased expression leads to perturbations in a great variety of biological pathways. One of the HSA21 genes that is remarkably sensitive to gene dosage is the one encoding the protein kinase DYRK1A. Both its overexpression, as part of the DS critical region, as well as its haploinsufficiency have been linked to neurodevelopmental alterations and mental retardation in humans. In fact, haploinsufficiency of *DYRK1A* is proposed to be considered a distinctive clinical syndrome including mental retardation, primary microcephaly, intra-uterine growth retardation and behavioural problems. Therefore, DYRK1A plays an important role in mammalian brain development that has not been clearly elucidated.

The group has developed a mouse model of *Dyrk1a* haploinsufficiency and it also works with transgenic mouse models in which the *Dyrk1a* gene is present in three copies through a collaboration with Jean Delabar (Universitè Paris). Moreover, since the gene dosage sensitivity might be linked to the ability of DYRK1A to interact with a wide variety of proteins and to participate in several signalling pathways, we are interested in establishing the DYRK1A interactome, as a way to understand the cellular functions of this kinase.

RESEARCH PROJECTS

1. DYRK1A and Nervous System Development

We have previously shown that *Dyrk1a* has an important role in the developing retina by acting as a negative regulator of the intrinsic apoptotic pathway. We thus hypothesized that dysregulation of the apoptotic response in differentiating neurons participates in the neuropathology of diseases that display DYRK1A gene-dosage imbalance effects, such as DS. Therefore, the anti-apoptotic effect of DYRK1A observed in gain-of-function *Dyrk1a* mutant mice should have an impact in the establishment of the retina circuitry in DS. As in the mouse lines overexpressing Dyrk1A, the Ts65Dn mice (a DS trisomic model), present alterations in cellularity, which correlate with reduced developmental cell death. As a consequence, electrical responses in the mature retina are significantly altered. The phenotype is due to the extra copy of *Dyrk1a*, because retinas of trisomic Ts65Dn mice in which *Dyrk1a* dosage have been normalized by genetic means show normal developmental apoptosis and cellularity in the adult.

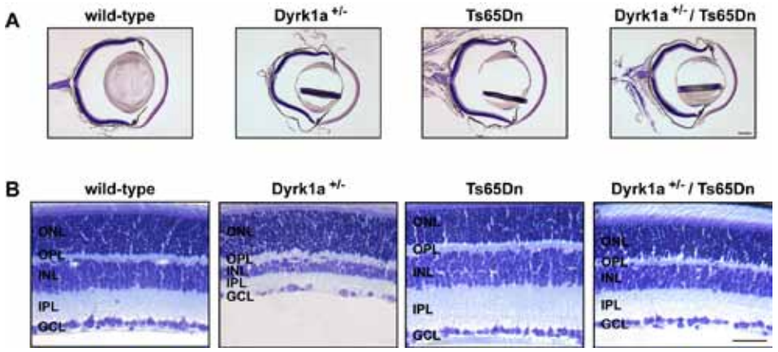


Figure 1.
Normalization of the retina phenotype in the Down syndrome mouse model Ts65Dn by reduction of one copy of Dyrk1a.

2. The DYRK Family of Protein Kinases

One of the approaches we have chosen to elucidate the participation of DYRK1A in different cellular processes and signal transduction pathways is to inferred it from the activity of its interactors and substrates. In this line, several proteomic screens are currently carried out to identify novel DYRK1A partners that can act as regulators of DYRK1A at different levels (stability, subcellular localization, activity) or as substrates.

We have extended our interest to other members of the DYRK family of protein kinases and characterized the less known member of the family, DYRK4. We have found that a combination of promoter usage and alternative splicing gives rise to several protein isoforms that display differential expression patterns and distinct subcellular localization or kinase activity. Therefore, as in other DYRK family members, alternative splicing is used to generate functional diversity. Our results also show that substrate specificity may represent a critical factor that governs biological specificity among the DYRK family of protein kinases.

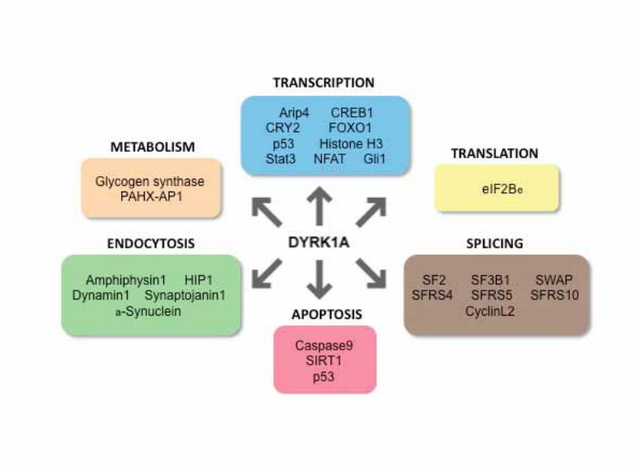


Figure 2.
DYRK1A may affect several cellular processes through phosphorylation of different substrates.

PUBLICATIONS

Aranda S, Laguna A, de la Luna S.
"DYRK family of protein kinases: evolutionary relationships, biochemical properties, and functional roles."
FASEB J, 25(2):449-62 (2011).

Calvo CF, Fontaine RH, Soueid J, Tammela T, Makinen T, Alfaro-Cervello C, Bonnaud F, Miguez A, Benhaim L, Xu Y, Barallobre MJ, Moutkine I, Lyytikka J, Tatlisumak T, Pytowski B, Zalc B, Richardson W, Kessaris N, Garcia-Verdugo JM, Alitalo K, Eichmann A, Thomas JL.
"Vascular endothelial growth factor receptor 3 directly regulates murine neurogenesis."
Genes Dev, 25(8):831-44 (2011).

Oeschger FM, Wang WZ, Lee S, García-Moreno F, Goffinet AM, Arbonés ML, Rakic S, Molnár Z.
"Gene expression analysis of the embryonic subplate."
Cereb Cortex, [Epub ahead of print]

Papadopoulos C, Arato K, Lilienthal E, Zerweck J, Schutkowski M, Chatain N, Müller-Newen G, Becker W, de la Luna S.
"Splice variants of the dual specificity tyrosine phosphorylation-regulated kinase 4 (DYRK4) differ in their subcellular localization and catalytic activity."
J Biol Chem, 286(7):5494-505 (2011).

GENES AND DISEASE

Group: Neurobehavioral Phenotyping of Mouse Models of Disease

Group Structure:

Group Leader: Mara Dierssen

Predoctoral Students: Carla Obradors (until January 2011), Susanna Molas, Davide D'Amico, Meritxell Pons, Debora Masini, Sivina Catuara

Postdoctoral Fellows: Monica Joana Pinto Dos Santos, Maria del Carmen de Agustín Pavón, Thomas Gener (since February 2011)

Technician: María Martínez de Lagrán Cabredo, Tiziana Pederzani

Mouse phenotyping: Ignasi Sahún Abizanda

Master/Graduate Students: Jofre Güell Bosch (since December 2011), Jacobo Chamorro López (since March 2011), Sara Padrell (since December 2011)

SUMMARY

The overall goal of our research is to understand the role of putative candidate genes for human complex genetic diseases that affect cognitive systems, using genetically modified mouse models as our main experimental tool. The characterization of these models allows obtain better knowledge of the genetic substrates regulating the expression of complex behavioral traits and the pathogenesis of neuropsychiatric and neurological disorders. Understanding the genetic and neural circuits disturbed in mental retardation and neuropsychiatric disorders is one of the significant challenges in ultimately treating it. Answers may emerge from systems neuroscience approaches that combine cognitive, imaging, and genetic analyses with the results from animal and cellular models. Our results have already led to one patent and a clinical assay (phase I).

RESEARCH PROJECTS

During the last years our group has contributed significantly to research into cognitive processes and their relationship to brain function in neuropsychiatric disorders and mental retardation. The work has been relevant to the translation of research from basic behavioral science and integrative neuroscience, to clinical issues. By taking validated mouse models of human developmental cognitive disorders such as Down syndrome we have devised fundamental problems related to synaptic plasticity in these models and have devised new therapeutic strategies to rescue learning and memory, and delineate the cellular and molecular correlates of effective therapies and their mechanistic interrelations. The true value of our studies lies in the systematic and wide-angled approach that has lead to an integrated overview of the mechanisms underlying (different forms) of memory. From the translational angle, the added value of her project relies in the fact that a pilot study in humans has been already performed based on our basic research.

Concretely the group has contributed to characterize the functional and structural organization of the cognitive networks in Down syndrome (DS) models and the mechanistic aspects underlying the neuropathology. There are a number of mouse models of DS that replicate some of the cognitive and behavioral alterations of DS humans (for a review see Lott and Dierssen 2010). However, the link between the genetic and neuronal basis of the alteration and the corresponding behavioral and cognitive phenotype is still missing.

a) DS children received special care programs as a treatment to improve social and cognition capacities, however the positive effects of this treatment are limited and temporary. We are interested in dissection the molecular and cellular mechanisms underlying this process and for that we used environmental enrichment paradigms in trisomic murine models (Baamonde *et al* 2011, Pons-Espinal in preparation).

b) We address the genetic mechanisms involved in the dendritic structural abnormalities observed in mental retardation disorders (Martínez de Lagrán *et al* 2012) and the possible involvement of Rho signaling and actin cytoskeleton (EU project granted; M Dierssen, coordinator).

c) We have performed a proteomic analysis for searching protein dysregulation in fetal DS brain before neurodegenerative changes are observed (Sun *et al*, 2011).

d) We have performed a large-scale analysis of gene interactions involved in cognitive phenotype using genetically engineered aneuploid mice for overlapping mouse chromosomes homologous to HSA21 (Sahún *et al* in preparation, coll. Ionas Erb/Cedric Notredame).

e) We have also continued the characterization of single gene transgenic models focusing on the fine dissection of different forms of learning and memory (Dierssen *et al* 2011).

f) Epigenetic contributions to the functional and structural anomalies in DS. We have analyzed histone deacetylation patterns and DNA methylation patterns (coll. M Esteller).

g) Several physiological mechanisms of cortical neurons and synapses play a key role in constraining the extent and strength of network activation, among them are the balance between excitation and inhibition in the cortex (coll. MV Sánchez-Vives; IDIBAPS).

1. Neuropsychiatric Disorders

a) **Panic disorder**
We use a mouse model that overexpresses the human *NTRK3* gene, which encodes the high affinity receptor of neurotrophin 3 - TrkC, to (1) address the contribution of this receptor to panic disorder and (2) study the role of TrkC in hippocampal function. We have found that the hippocampal cognitive phenotype of the TgNTRK3 mice is altered as compared to wt mice and that the three main brain structures that regulate fear (amygdala, hippocampus and prefrontal cortex) present an altered activation pattern both at naïve state and after fear conditioning. Moreover we are also addressing the exact molecular mechanism through which overexpression of TrkC leads to an altered hippocampal function. Both the MAPK signalling cascade, which is downstream of TrkC and the role of TrkC as a synaptogenic molecule are good candidates that we are exploring. We address the pathogenetic mechanisms that underlie the inability of persons with anxiety disorders to correctly identify the fear-related information using genetically modified mouse models (D'Amico, PhD thesis; Santos M., D'Amico D., *et al*, in preparation).

b) **Nicotine dependence**
The second line is focused on the genetic-susceptibility to nicotine dependence (X. Gallego *et al*,; S. Molas PhD project)

c) **Obsessive-compulsive behavioral components of obesity**
We focus on the development of a behavioral model of the obsessive-compulsive elements involved in eating disorders (J. MacDonald *et al* in preparation; Heyne *et al* 2009; Mercader *et al* submitted) and in the bioinformatics analysis of the behavioral structure.

d) **Schizophrenia**
We are collaborating in the behavioural characterization of murine models of schizophrenia (coll O Marin and Texeira *et al* 2011).

PUBLICATIONS

Gallego X, Ruiz J, Valverde O, Molas S, Robles N, Sabriá J, C Crabbe J, Dierssen M
"Transgenic Over Expression of Nicotinic Receptor Alpha 5, Alpha 3, and Beta 4 Subunit Genes Reduces Ethanol Intake in Mice"
Alcohol, [Epub ahead of print].

Teixeira CM, Martín ED, Sahún I, Masachs N, Pujadas L, Corvelo A, Bosch C, Rossi D, Martinez A, Maldonado R, Dierssen M, Soriano E.
"Overexpression of Reelin prevents the manifestation of behavioral phenotypes related to schizophrenia and bipolar disorder."
Neuropsychopharmacology, 36(12):2395-405 (2011).

Viñals X, Molas S, Gallego X, Fernández-Montes RD, Robledo P, Dierssen M, Maldonado R.
"Overexpression of α3/α5/β4 nicotinic receptor subunits modifies impulsive-like behavior."
Drug Alcohol Depend., [Epub ahead of print] (2011).

Baamonde C, Martínez-Cué C, Flórez J, Dierssen M.
"G-protein-associated signal transduction processes are restored after postweaning environmental enrichment in Ts65Dn, a Down syndrome mouse model."
Dev Neurosci., 33(5):442-50 (2011).

Mocholí E, Ballester-Lurbe B, Arqué G, Poch E, Peris B, Guerri C, Dierssen M, Guasch RM, Terrado J, Pérez-Roger I.
"RhoE deficiency produces postnatal lethality, profound motor deficits and neurodevelopmental delay in mice."
PLoS One, 6(4):e19236 (2011).

Sun Y, Dierssen M, Toran N, Pollak DD, Chen WQ, Lubec G.
"A gel-based proteomic method reveals several protein pathway abnormalities in fetal Down syndrome brain."
J Proteomics., 74(4):547-57 (2011).

Gallego X, Molas S, Amador-Arjona A, Marks MJ, Robles N, Murtra P, Armengol L, Fernández-Montes RD, Gratacòs M, Pumarola M, Cabrera R, Maldonado R, Sabrià J, Estivill X, Dierssen M.
"Overexpression of the CHRNA5/A3/B4 genomic cluster in mice increases the sensitivity to nicotine and modifies its reinforcing effects."
Amino Acids, [Epub ahead of print].

Martinez de Lagran M, Benavides-Piccione R, Ballesteros-Yañez I, Calvo M, Morales M, Fillat C, Defelipe J, Ramakers GJ, Dierssen M.
"Dyrk1A Influences Neuronal Morphogenesis Through Regulation of Cytoskeletal Dynamics in Mammalian Cortical Neurons."
Cereb Cortex, [Epub ahead of print].

Dierssen M, Arqué G, McDonald J, Andreu N, Martínez-Cué C, Flórez J, Fillat C.
"Behavioral characterization of a mouse model overexpressing DSCR1/ RCAN1."
PLoS One., 6(2):e17010 (2011).

PATENTS

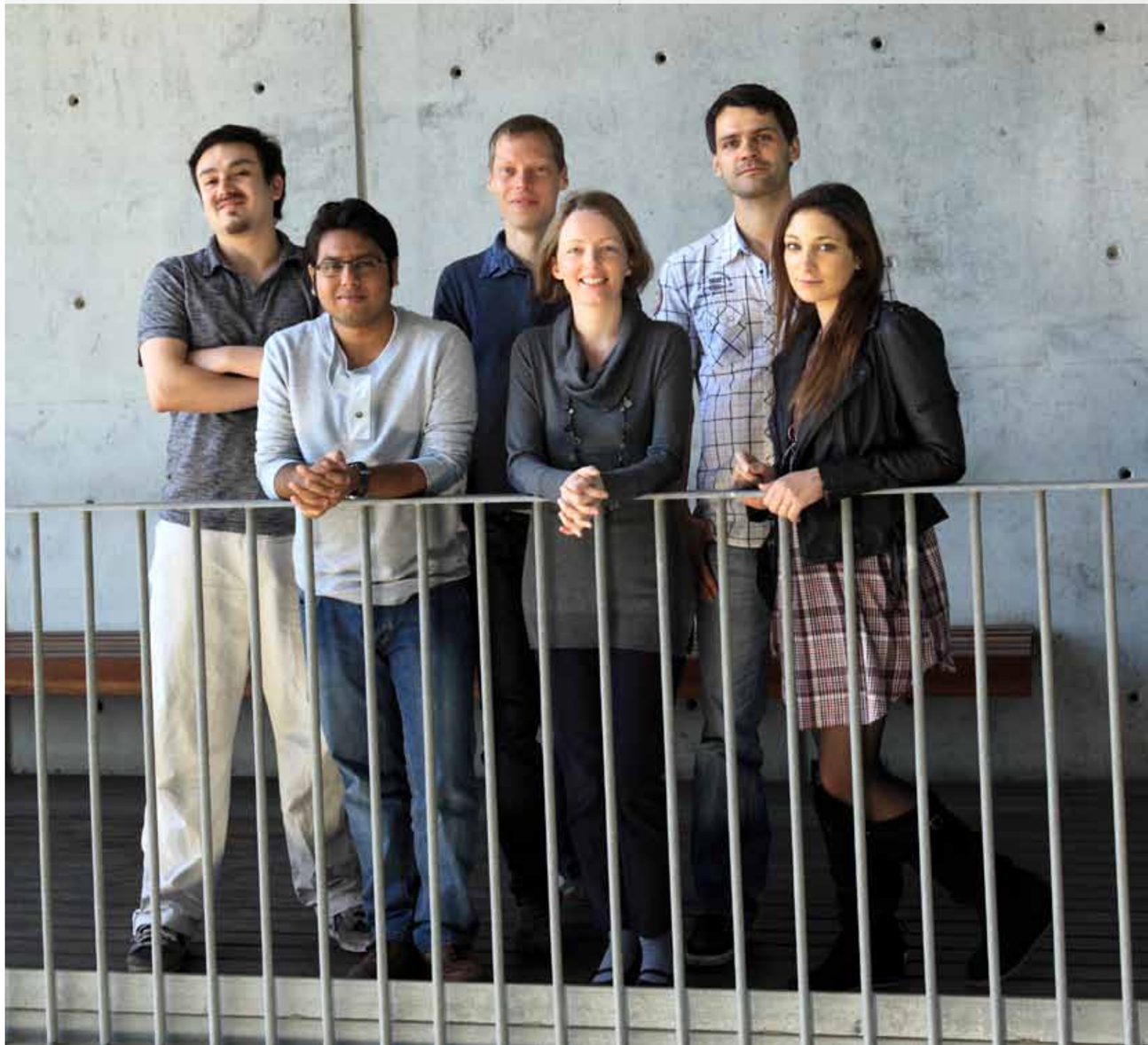
CNB-CIBERNED-CRG
"Compuestos para el tratamiento de enfermedades neurodegenerativas."
P6529ES00.

GENES AND DISEASE

Group: Genomic and Epigenomic Variation in Disease

Group Structure:

Group Leader:	Stephan Ossowski
Postdoctoral Fellows:	Oliver Drechsel (since September 2011), Charlotte Hor (since March 2011)
PhD Students:	Luis Zapata (since June 2011)
Technicians:	Daniela Bezdan (since March 2011)



SUMMARY

Next Generation Sequencing (NGS) techniques enable analysis of genome, transcriptome and epigenome of an individual in a stage, tissue or even cell specific manner at single nucleotide resolution. Hundreds of samples can be analyzed at different molecular or regulatory levels using the same technology platform. This allows for identification of disease specific alterations at the molecular level and will likely result in optimized individual treatment of patients. Furthermore NGS techniques are used for metagenomic analysis of pathogenic and non-pathogenic bacteria that will provide a better understanding of pathogenicity, host defense, antibiotic resistance and the impact of drugs on the human microbiome.

The focus of our group is to integrate several NGS based analysis methods including whole genome and exome sequencing, ChIP-seq, RNA-seq and DNasel-seq in order to detect genomic, genic and epigenomic variation related to disease or intolerance to specific treatments. We seek to develop and improve computational analysis algorithms for each sequencing method and to integrate and correlate resulting marks and measures, e.g. genomic variants, gene expression, histone modifications, DNA methylation and chromatin conformation. We envision studying multiple snapshots of the genomic and epigenomic landscape of tissues during the development of a disease, i.e. the personal OMICs profile of a patient, to better understand the impact of genetic predisposition, epigenomic and regulatory variability, viral and bacterial infections and other environmental effects on the development of complex diseases.

RESEARCH PROJECTS

1. Human Brain Epigenome Atlas: Tissue and Allele Specific Epigenomic Variation in Neurological Disorders

Complex disorders, including neurodegenerative diseases, are thought to arise due to the environment acting on genetically susceptible individuals. Despite numerous genetic and epidemiological studies, the causative genetic variants and environmental factors, as well as the mechanisms by which they lead to disease, remain elusive. Epigenetic marks, defined as the modifications to the DNA molecule and chromatin structure not affecting the nucleotide sequence itself, are simultaneously a function of the underlying sequence, and a reflection of the impact of the environment on the cell, tissue, or whole organism. Thus, the epigenome may represent the long sought after interface between genetic background and environment. While an increasing number of genetic mutations have been linked to human complex diseases, aberrant DNA methylation and chromatin states have been observed in cancer, as well as neurological and autoimmune disorders.

The goal of this project is to establish a genome wide map of epigenetic markers in human brain samples from control individuals and subjects with neurodegenerative disease, at nucleotide resolution and encompassing allele-specific effects, in order to determine epigenomic patterns in disease and their correlation to genomic variants and differential gene expression. Disease specific epigenomic patterns could be interesting biomarkers for determination of causal genomic loci, development of new treatment methods, and could help direct the search for peripheral biomarkers for early diagnosis and prognosis. Comparison of chromatin state and expression patterns between healthy and diseased tissues might also shed light on the functional implications of intergenic variants linked with complex diseases by e.g. aQTL analysis or GWAS. We aim to establish multiple next generation sequencing based analysis methods to map diverse epigenetic marks, i.e. DNA methylation, histone methylation and acetylation, DNAseI sensitive sites (open chromatin), chromatin conformation as well as occupancy of the chromatin by key proteins as RNA polymerase II and CTCF. Sequencing methods include whole genome and exome sequencing, whole genome bisulfite sequencing, ChIP-seq, DNAseI-seq and ChIA-PET. Adapting algorithms used by Ernst et al. for the Encode project we will subsequently generate tissue and disease specific maps of chromatin states. The information generated in this project will be crucial for the detection of epigenomic patterns in disease and their correlation to genomic variants and differential gene expression.

2. Exome Sequencing: A Key Approach to Study Human Disease and Cancer

Novel or inherited genetic variations can lead to drastic phenotypes including rare and common diseases. Human exome analysis using next generation sequencing (Exome-seq) has recently been established as a key approach to identify genetic variations in protein coding genes. We have developed computational methods for the identification and functional analysis of causal and disease associated mutations in Exome-seq studies of rare and common, mendelian and complex diseases. Our studies cover various types of diseases including cardio-vascular (e.g. familial hyperkalemic hypertension, varicose veins), neurological (e.g. Parkinson's disease and congenital myopathies) and rheumatologic diseases (e.g. fibromyalgia) as well as cancer (chronic lymphocytic leukemia). To facilitate the routine application of Exome-seq for diagnosis and improved personalized treatment in hospital environments we have implemented Exome-CRG. This pipeline performs read alignment, SNP and indel prediction, CNV identification, functional annotation of coding variants and adds OMICs information from e.g. dbSNP, 1000genomes and OMIM. To improve data accessibility and to facilitate comparison between studies we develop BioMart and browser based retrieval tools. These incorporate predicted variants together with relevant OMICs data on diseases as well as common and rare variants in the population, focusing on samples from the Iberian Peninsula. Applying our approach we have recently identified *KLHL3* as a gene responsible for familial hyperkalemic hypertension (FHHt). A novel damaging missense mutation in a family with three affected members was identified by the analysis of Exome-seq data.

3. Understanding Cancer Genomes: Computational Analysis of Structural Variants and Correlated Transcriptional and Epigenomic Variation in Leukemia

In this project we will develop novel methods for detection of structural variants (SV), copy number variants (CNV) and movement of transposable elements using Next Generation Sequencing (NGS) data from chronic lymphocytic leukemia tumors (CLL) sequenced as part of the International Cancer Genome Consortium. Further we will combine information across multiple tumor and normal samples from hundreds of CLL patients sequenced using multiple strategies (genome or exome sequencing, RNA-seq) to study driver and passenger mutations of the tumor, mutation heterogeneity as well as mutations influencing metastasis. The group is participating in the International Cancer Genomics Consortium (ICGC) with the aim to fully characterize the genome of chronic lymphocytic leukemia (CLL). The role of our group is to optimize the computational analysis of NGS data in order to obtain high quality micro-indel and structural variant predictions in normal and leukemic cells from patients with CLL. The Spanish contribution to the ICGC is a collaborative effort between several centers in Spain (Clinic Hospital, University of Oviedo and CNIO, among others). We have made substantial progress in implementing a new algorithm to detect small to medium sized indels (1 to 200bp), which are hard to detect using paired-end mapping (PEM) based strategies. We have further set up a standardized analysis pipeline including SNP, micro-indel, SV and CNV prediction, which is currently applied to whole genome data of 50 patients and Exome data of 120 patients.

4. Determining Genetic Factors and Metagenomic Alterations Related to increased virulence and multi-antibiotic resistances in *S. aureus* and *P. aeruginosa*

In this project we will study genetic factors influencing *Staphylococcus aureus* and *Pseudomonas aeruginosa* infections responsible for significant morbidity and mortality in community and health care settings due to increasing frequency of antibiotic resistance. Next generation sequencing technology allows for a detailed analysis of the genes and variants correlated to e.g. pathogenicity and antibiotic resistance. New algorithms are developed for comparison of hundreds of sequenced strains. Further we develop approaches for candidate gene screening using T-DNA directed insertion-site sequencing of *S. aureus* mutants and metagenomic sequencing of the nares. We have already sequenced 12 strains of *S. aureus* including pathogenic and non-pathogenic strains as well as 23 strains of *P. aeruginosa*, which are currently being assembled and compared in order to detect causal variants.

PUBLICATIONS

Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, Wang X, Ott F, Müller J, Alonso-Blanco C, Borgwardt K, Schmid KJ, Weigel D.
"Whole-genome sequencing of multiple Arabidopsis thaliana populations."
Nat Genet., 43(10):956-63 (2011).

Schneeberger K[1], Ossowski S[1], Ott F, Klein JD, Wang X, Lanz C, Smith LM, Cao J, Fitz J, Warthmann N, Henz SR, Huson DH, Weigel D. ([1]contributed equally)
"Reference-guided assembly of four diverse Arabidopsis thaliana genomes."
Proc Natl Acad Sci U S A., 108(25):10249-54 (2011).

Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, Haberer G, Hollister JD, Ossowski S, Ottilar RP, Salamov AA, Schneeberger K, Spannagl M, Wang X, Yang L, Nasrallah ME, Bergelson J, Carrington JC, Gaut BS, Schmutz J, Mayer KF, Van de Peer Y, Grigoriev IV, Nordborg M, Weigel D, Guo YL.
"The Arabidopsis lyrata genome sequence and the basis of rapid genome size change."
Nat Genet., 43(5):476-81 (2011). (*)

Guo YL, Fitz J, Schneeberger K, Ossowski S, Cao J, Weigel D.
"Genome-wide comparison of nucleotide-binding site-leucine-rich repeat-encoding genes in Arabidopsis."
Plant Physiol.,157(2):757-69 (2011). (*)

Klein JD, Ossowski S, Schneeberger K, Weigel D, Huson DH.
"LOCAS--a low coverage assembly tool for resequencing projects."
PLoS One.;6(8):e23455 (2011).

(*) These publications result from the work of Dr. Stephan Ossowski at the Max Planck Institute for Developmental Biology, Tübingen, Germany

BIOINFORMATICS AND GENOMICS

Group: Bioinformatics and Genomics

Group Structure:

Group Leader:	Roderic Guigó
Staff Scientist:	Rory Johnson
Postdoctoral Fellows:	Sarah Djebali, David Martin, Pedro Gabriel Dias Ferreira, David González, Angelika Merkel, Andrea Tanzer, Dmitry Pervouchine, Marta Melé
PhD Students:	Hagen Tilgner, Anna Kedzierska, Marco Mariotti, Joao Curado, Didac Santesmases, Jorgen Skancke (visiting student)
Technicians:	Julien Lagarde, Francisco Câmara, Maik Röder, Carmen Arnán, Alessandra Breschi
Students:	Cinzia de Benedictis, Marcel Costa, Joan Pallares, Daniel Soronellas, Lukas Kuderna, Chiara Medoro, Jean Monlong

SUMMARY

Research in our group focuses on the investigation of the signals involved in gene specification in genomic sequences (promoter elements, splice sites, translation initiation sites, etc…). We are interested both in the mechanism of their recognition and processing, and in their evolution. In addition, but related to this basic component of our research, our group is also involved in the development of software for gene prediction and annotation in genomic sequences. Finally, our group has actively participated in the analysis of many eukaryotic genomes and it in involved in the NIH funded ENCODE project.

RESEARCH PROJECTS

1. Gene Prediction/Genome Annotation

We have continued working on the development of geneid, an *ab initio* gene prediction program, and SGP2 a comparative gene finder. As part of this effort we also created a semi-automated software pipeline tool (geneidTRAINer) that which simplifies the training of the program geneid. Furthermore, in 2011 we have continued our collaboration with the Baylor College of Medicine (Houston, USA) in the annotation of several hymenoptera species (*B.terrestris/B.impatiens/A.mellifera*) by generating gene predictions that are being incorporated into the annotation pipelines for these organisms. We have also continued our collaboration with the Broad institute (MIT, Boston) in the annotation of fungi and protozoan genomes. We hope to make this tool available to the general public in the course of 2012. We have now generated fifty parameter files for gene prediction in a wide range of species spanning all Kingdoms of life (*http://genome.crg.es/software/geneid/index.html#parameters*). Moreover, we have also continued our participation in three large genome sequencing/annotation consortia projects on: 1) the annotation of the *Cucumis metuliferus* (melon) genome 2) the sequencing and annotation of the common bean (*Phaseolus vulgaris*) genome and 3) the assembly, annotation and comparative analysis of Iberian and European lynx genomes.

We have also proceeded with our effort in developing improved methods to predict and identify selenoprotein genes. In collaboration with Gladyshev's lab at Harvard, we developed a new method to scan nucleotide sequences for eukaryotic Selenocysteine Insertion Sequences (SECIS), which proved to be superior to existing methods. We then built a pipeline (Seblastian) to predict selenoprotein genes in nucleotide sequences, that uses such SECIS predictions and searches upstream of them for potential selenoprotein coding genes. We ran the pipeline on a number of eukaryotic genomes, predicting several new selenoproteins. We also built a web service for these two new programs, which will be made public soon.

Also in collaboration with the same lab, we worked on the computational characterization of the vertebrate and mammalian selenoproteins. This resulted mainly in the redefinition of the ancestral vertebrate selenoproteome, and in the description of how this evolved along the various lineages through gene duplications, losses, and replacements of selenocysteine with cysteine.

2. Characterization of Long Non-Coding (lnc)RNAs

Long Noncoding RNAs (lncRNA) constitute a large portion of the mammalian transcriptome even though their biological functions remained unclear. Using RNAseq data we have recently shown that lncRNAs display differential tissue expression, which is closely correlated with the presence of active or repressive chromatin signatures. Importantly, we found increased expression of a number of lncRNAs following differentiation of adult human stem cells. Functional analysis of these lncRNAs in multiple cell lines revealed their positive regulation of neighboring protein-coding transcripts. This work was published in the journal Cell in 2010 (Orom *et al*., 2010) and reviewed recently (Derrien *et al*., 2012). We subsequently have begun to profile human lncRNA expression in multiple human tissues, disease cohorts and differentiation models. In order to facilitate such a task during 2010-11 we designed a custom microarray platform containing ~10,000 lncRNAs, based on the GENCODE annotation of human genes. In these arrays each lncRNA is interrogated by multiple distinct microarray probes, allowing sensitive and accurate detection. We are using these arrays with various collaborators to study lncRNA involvement in a number of biological processes. In parallel, we are developing tools to study candidate lncRNAs in loss-of-function assays. We have adapted retroviral delivery systems and short hairpin technology to target lncRNAs in a high-throughput manner, and are currently testing this as a method for functional lncRNA discovery. Within the ENCODE project we have carried an exhaustive characterization of human lncRNAs. We have also been investigating into the function of lncRNAs by statistical analysis of high-throughput RNA-seq data, particularly focusing on the possible association between lncRNA expression and splicing. On a slightly different focus, and based on the existence of increasing evidence that lncRNA may also play a relevant role in the regulation of several cellular processes, we started employing comparative genomics approaches to study conservation of lncRNAs in Great Apes. Previous studies at the patterns of conservation comparing vertebrate genomes have found that lncRNAS are indeed under purifying selection. However, except for a few well-known cases such as Xist or HotAir, most of the annotated lncRNA have an unclear function. We believe that studying how natural selection has targeted lncRNAs along the Great Ape lineage will allow shedding some light in our understanding of the relevance and functionality of these emerging elements of the genome.

3. Methods for Transcriptome Analysis and the ENCODE Project

The field of transcriptomics has recently been given a huge boost from the use of so-called "second" generation high throughput sequencing technologies to sequence RNA samples. Second-generation sequencing technologies provide an unprecedented capacity for surveying the nucleic acid content of cells. Specially since these techniques started to be applied to transcriptome sequencing we have become increasingly aware of the large number of genes that show alternative splice forms in human as well as the large variety of splice forms that these genes can have, that may range from just two splice variants to hundreds. On the other hand, the accelerating rate of data production with these new technologies is moving the bottleneck in many studies from the data generation to the actual analysis of these data. Because of this it is important to design methods with which we can analyze them in a fast and efficient manner. Our aim is to use the data from these experiments in order to determine the exact transcript abundances within the cell. Not only as a list of the transcripts that are expressed at the qualitative level, but also the exact expression level of each transcript and alternative variant within the cell, while at the same time developing a highly automated method that will allow us to take advantage of the huge amounts of data available. Following up on this strategy and as part of the ENCODE projects lead by Tom Gingeras (Transcriptome) and Tim Hubbard (GENCODE), our group has been working towards the development of a number of tools for RNASeq processing. These include the GEM read aligner (*http://big.crg.cat/services/gem_genome_multi_tool_library*), the Flux Capacitor (*http://big.crg.cat/services/flux_capacitor*) for transcript quantification, and NextGeneid for "de novo" transcript modeling and discovery. In 2011 we incorporated these tools, as well as tools developed elsewhere, into GRAPE (General RNASeq Analysis Pipeline Environment), a robust, efficient and scalable software system for the storage, organization, access, and analysis of RNASeq data. The system has three main components: a structured repository hosting the raw and processed data, an RNASeq pipeline to transparently produce transcript models and quantifications from sequence reads, and a common interface to both, the data and the analysis results (*http://rnaseq.crg.cat/*). GRAPE has been used to process the RNASeq data produced by the ENCODE project and elsewhere. GRAPE is still under development, but a preliminary version can be accessed at *http://big.crg.cat/services/grape*. The GRAPE dashboard, under development, offers a point of entry to the data and results through a schematics of the project's experimental design (Figure 1).

Figure 1 - *Interface to the results of the RNASeq pipeline.* ***Left****. Entry page to an experiment, with an overall summary of the processing results.* ***Right****. Snapshots from some of the analysis pages.* ***Top****. Distribution of read across transcripts for different transcripts lengths.* ***Bottom****. Lane to lane correlation between transcript quantifications.*

We have also recently developed statistical methods to assess variability of splicing ratios, to identify genes with condition specific splicing patterns, and to de-convolute the contribution of expression variation vs splicing variation to the variation in the abundance of alternative splice forms (Figure 2, Gonzalez-Porta *et al*., 2011).

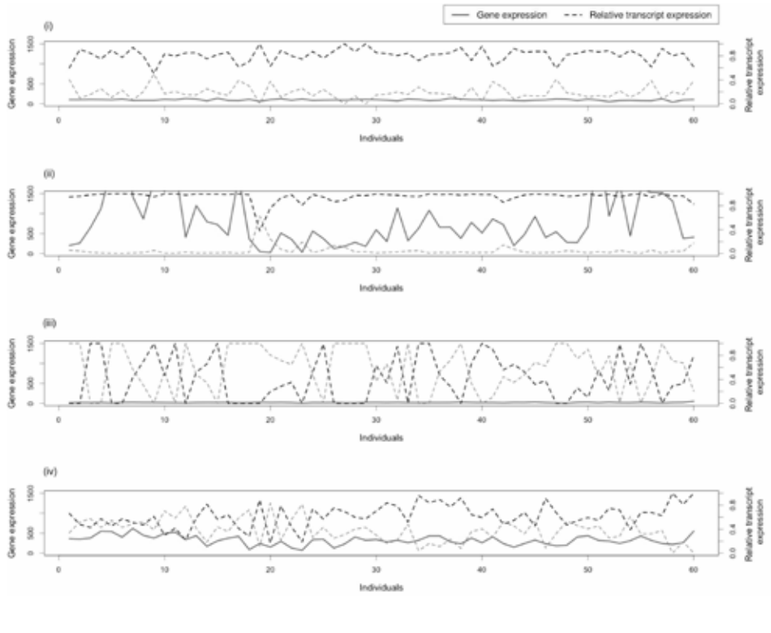


Figure 2 - *Variability in gene expression versus variability in splicing ratios. Behavior of four different genes regarding expression and splicing variability in human populations. Four genes with two splice forms each have been selected to illustrate possible extreme cases: (i) Low variability in both gene expression and splicing (as exhibited by the Vacuolar protein sorting-associated protein 28 homolog gene, VPS28); (ii) variability in gene expression, but quite constant splicing ratios (as exhibited by Prothymosin alpha, PTMA); (iii) constant gene expression, but variability in alternative splicing ratios (as exhibited by the Coiled-coil domain containing 43 gene, CCDC43); and (iv) variability in both gene expression and alternative splicing ratios (as exhibited by the Heterogeneous nuclear ribonucleoprotein M, HNRNPM). The x-axis denotes the 60 individuals in which the values have been profiled and the y-axis both absolute gene expression (measured in RPKMs) and relative splicing ratios. (from Gonzalez-Porta et al., 2011)*

Finally, as part of the GENCODE project, we have developed a pipeline for experimental verification of transcript annotations. The pipeline, which we termed RT-PCRSeq, is based on an efficient multiplexing of RT-PCR reactions using high throughput sequencing.

We are also continuing with our efforts to understand RNA processing, and in particular pre-mRNA splicing. In 2011 one of our tools (AStalavista; *http://big.crg.cat/services/astalavista*) which retrieves all alternative splicing events from generic transcript annotations was employed on an important focusing study on the origin and evolution of alternative splicing (Mudge *et al*., 2011).

Finally, we have a long-standing interest in the assessment of methods to delineate transcript structures and transcript abundances, and within GENCODE, we have organized RGASP 1 and 2 community benchmark experiments (*http://www.gencodegenes.org/rgasp/*).

4. Modeling Splicing from Chromatin

The search of sequence prerequisites for nucleosome positioning has been a long-standing problem at the intersection of chromatin structure and gene regulation. With the use of structural information directly related to the primary DNA sequence we propose a new property of natural nucleosome forming sequences, which is inherently related to their intrinsic curvature and its symmetry. A measure of this property has been introduced and a corresponding method validated against novel high quality datasets of human nucleosomes, obtained through close collaboration with CRG's Chromatin and Gene Expression Group (Miguel Beato). We are also continuing our close collaboration with the group of Juan Valcárcel, from the CRG's Gene Regulation programs focusing on the investigation of the mechanisms through which splice signals are recognized and processed. We are developing new methods to infer sequences that may play a role in the regulation of alternative splicing, and have continued investigating the dynamics of the evolution of U12 introns.

5. Development of Methods for Evolutionary Genomics Studies

In 2011 we published an article (Kedzierska *et al*., 2011) that describes the development of a novel method for phylogenetic reconstruction (SPIn). This method is based on linear INvariants (SPIn), which uses recent insights on linear invariants to characterize a model of nucleotide evolution for phylogenetic mixtures on any number of components. SPIn is unique in which it does not require an input tree and is designed to deal with non-homogeneous phylogenetic data consisting of multiple sequence alignments showing different patterns of evolution, *i.e*. concatenated genes, exons and/or introns.

We also concluded a collaborative effort focusing on the CTCF vertebrate insulator protein. We have found that constitutive sites form mandatory boundaries within vertebrate genomes are preferably located close to transcription factor-encoding genes involved in development (Martin *et al*., 2011).

PUBLICATIONS

Gonzalez-Porta, M, Calvo M, Sammeth M, Guigo R.
"Estimation of alternative splicing variability in human populations".
Genome Research, 22(3):528-3 (2011).

Kedzierska AM, Drton M, Guigó R and Casanellas M.
"SPIn: model selection for phylogenetic mixtures via linear invariants".
Mol Biol Evol, 29(3):929-37 (2011).

Martin *et al*. (incl. R. Guigo).
"Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes".
Nat Struct Mol Biol, 18(6):708-14. Erratum in: Nat Struct Mol Biol, 2011, 18(9):1084 (2011).

Derrien T, Guigó R.
[Long non-coding RNAs with enhancer-like function in human cells].
Med Sci (Paris), 27(4):359-61 (2011).

Maxwell *et al*. (incl. J. Curado, H. Tilgner and R. Guigo).
"Interplay between BRCA1 and RHAMM Regulates Epithelial Apicobasal Polarization and May Influence Risk of Breast Cancer".
PLoS Biol, 9(11):e1001199 (2011).

Mudge *et al*. (incl. T. Alioto, T. Derrien and R. Guigo)
"The origins, evolution and functional potential of alternative splicing in vertebrates".
Mol Biol Evol, 28(10):2949-59 (2011).

Puente *et al*. (incl. R. Guigo).
"Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia".
Nature, 475(7354):101-5 (2011).

Orom UA, Derrien T, Guigo S, Shiekhattar R.
"Long noncoding RNAs as enhancers of gene expression."
Cold Spring Harb Symp Quant Biol, 75:325-31 (2011).

The "ENCODE Project Consortium".
"A user's guide to the encyclopedia of DNA elements (ENCODE)".
PLoS Biol, 9(4):e1001046 (2011).

Guigó R.
"Book review of CAP-ANALYSIS GENE EXPRESSION (CAGE)"
Bioessays, 2011 Dec 9.

BIOINFORMATICS AND GENOMICS

Group: Comparative Bioinformatics

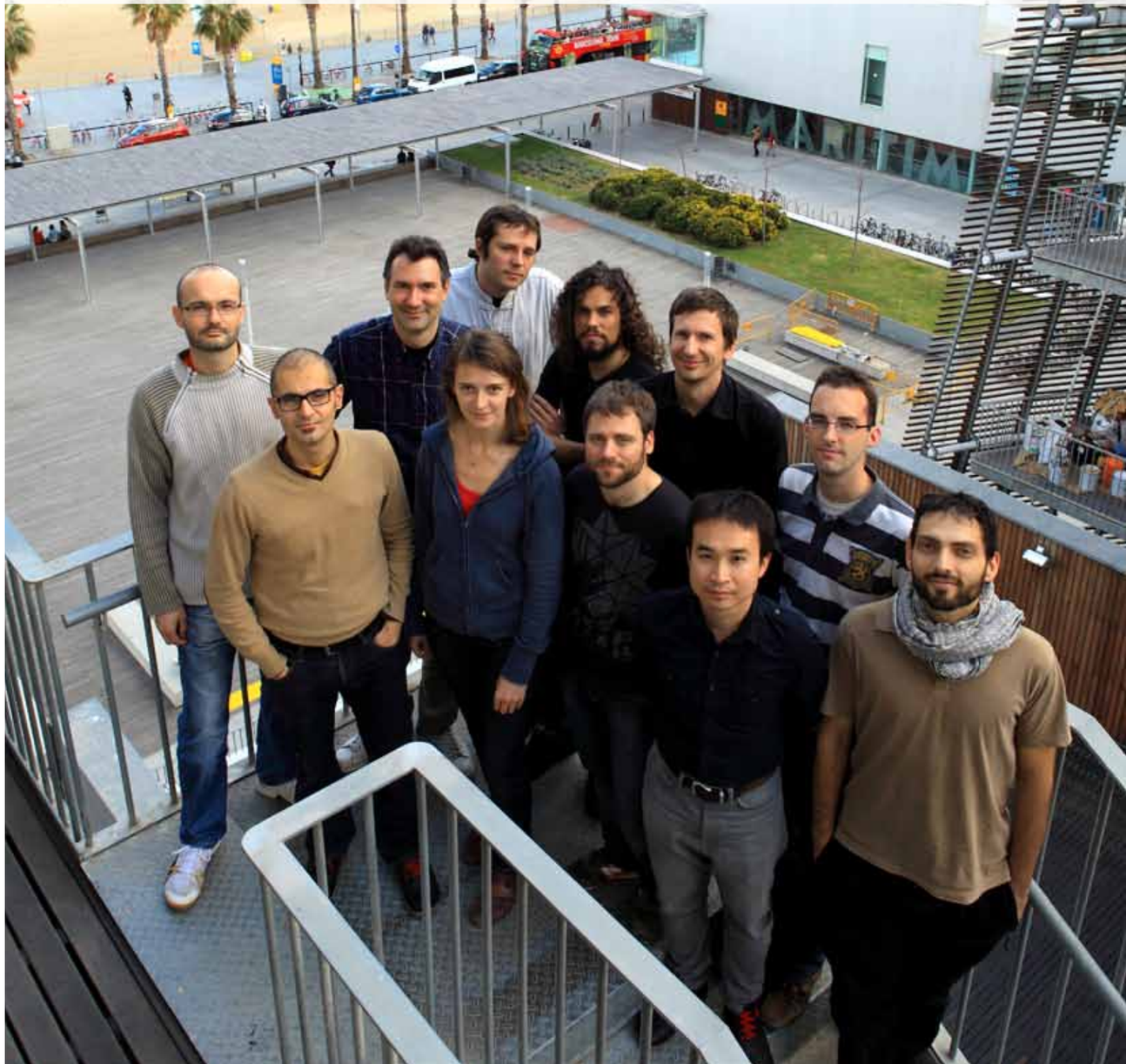
Group Structure:

Group Leader: Cédric Notredame

Postdoctoral Fellows: Ionas Erb, Jean-François Taly, Cedrik Magis, Darek Kedra

Students: Jia-Ming Chang, Giovanni Bussoti, Carten Kemena, Jose Espinosa, Pablo Prieto

Technicians: Paolo Di Tomasso



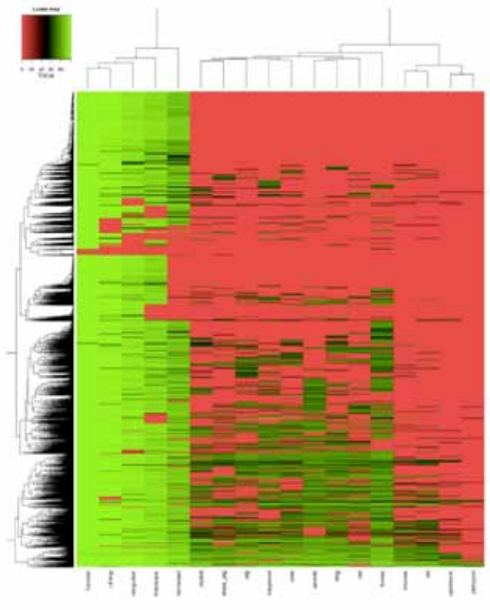
SUMMARY

The main focus of the group is the development of novel algorithms for the comparison of multiple biological sequences. Multiple comparisons have the advantage of precisely revealing evolutionary traces, thus allowing the identification of functional constraints imposed on the evolution of biological entities. Most comparisons are currently carried out on the basis of sequence similarity. Our goal is to extend this scope by allowing comparisons based on any relevant biological signal such as sequence homology, structural similarity, genomic structure, functional similarity and more generally any signal that may be identified within biological sequences. Using such heterogeneous signals serves two complementary purposes: (i) producing better models that take advantage of the evolutionary resilience, (ii) improving our understanding of the evolutionary processes that leads to the diversification of biological features. We develop these novel methods in close collaboration with experimental groups and make them available through an international network of web severs that can be accessed via *www.tcoffee.org*. In addition to the CRG and the Catalan government, the group is supported by the Plan Nacional, "la Caixa" Foundation and by two international FP7 consortiums: Quantomics (dedicated to the survey of genomic variation in farm animals) and Leishdrug, a project dedicated to the development of a new class of drugs targeting kinases in Leishmania Major.[1-10]

RESEARCH PROJECTS

1. Homology Modelling of Non Coding RNA

Aligning non-coding sequences, so as to reveal the constraints that may have shaped their diversity remains a major challenge. Non-coding nucleotide sequences are, indeed, fast evolving entities. This property, combined with the limited alphabet size, can hamper the use of the most commonly used statistical tools thus making it hard to assemble biologically meaningful alignments. As hard to get as they may seem, these models have never been in such dire need, with the number of functionally relevant non-coding sequences growing faster every year. Indeed more than 10.000 novel non-coding transcripts have been reported since 2009, and the systematic usage of Next Generation Sequencing data is gradually revealing a transcriptomic landscape of increasing complexity. With so much novelty at hand, assembling trustworthy models allowing the comparison of functional features is a compulsory step. Designing suitable tools, so that biologists can tap into this wealth of knowledge, is one of our prime goals.



We are therefore actively involved in the development of RNA dedicated tools. For instance, R-Coffee, the RNA flavour of the T-Coffee multiple sequence alignment package, makes it possible to rapidly align RNA sequences while taking into account their secondary structure. R-Coffee also allows the combination of slow but accurate RNA alignment methods like CONSAN and we are now investigating the possibility of combining sequence and structural RNA information in order to produce even more accurate models. We have also started exploring the issue of RNA detection, with a novel method named BlastR[1], that uses di-nucleotide information in order to increase the sensitivity of database search. We are routinely using this method in order to estimate the evolutionary trace associated with known transcripts[3].

On this line of research, our future goal will be to improve the alignment methods so that they can combine expression and evolutionary evidences in order to estimate accurate gene models.

2. Large Scale Protein Sequence Alignments

With new genomes being reported at the pace proteins sequences were published a decade ago, the size of samples available for phylogenetic analysis, and, in general, any kind of study focused on protein variability, has increased at an unprecedented level[6]. If we want the community to make the best out of this information, existing tools will need to be adapted so that they become efficient enough to deal with datasets several orders of magnitude larger than the ones currently available.

Our group is addressing this problem from several angles. First of all, we keep developing the T-Coffee algorithm, in an attempt to improve both its scalability and its accuracy[5]. We have done so in collaboration with the Sinica Academia in Taiwan. When scaling up aligners, the loss of accuracy is a major issue. We are therefore developing a novel methodology that makes it possible to estimate the accuracy of a multiple alignment using a limited number of structures (as few as one)[4]. This methodology, named STRIKE, makes it relatively straightforward to rank alternative alignments of the same sequences, a problem that has been shown to be critical when building phylogenies.

We take great care at insuring that all the methods developed in the group receive optimal visibility, either through web servers that insure access by non-specialist (*www.tcoffee.org*) or via dedicated protocols.

3. Multiple Genome Alignments

Multiple genome alignment (MGA) is a complex and important problem. MGAs can be very useful in order to identify functional elements in a genome. This identification usually occurs through the identification of sites evolving under positive or purifying selection. In this context, our goal is to address the problem using as much as possible available functional information. We have started approaching MGAs from the angle of multiple promoter region alignments. Our methodology, named Pro-Coffee relies on the use of di-nucleotide substitution patterns. Its main originality, however, is not so much the novelty of the algorithm, but rather to propose a novel approach for the fine tuning of such aligners, an approach that depends on the combined used of experimental Chlp-Seq datasets (done on transcription factors) and more traditional comparative genomics processes.

PUBLICATIONS

Bussotti G, Raineri E, Erb I, Zytnicki M, Wilm A, Beaudoing E, Bucher P and Notredame C.
"BlastR--fast and accurate database searches for non-coding RNAs."
Nucleic Acids Res, 39(16):6886-95 (2011).

Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang JM, Taly JF and Notredame C.
"T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension."
Nucleic Acids Res, 39(Web Server issue):W13-7 (2011).

Esteve-Codina A, Kofler R, Palmieri N, Bussotti G, Notredame C and Perez-Enciso M.
"Exploring the gonad transcriptome of two extreme male pigs with RNA-seq."
BMC Genomics, 12:552 (2011).

Kemena C, Taly JF, Kleinjung J and Notredame C.
"STRIKE: evaluation of protein MSAs using a single 3D structure."
Bioinformatics, 27(24):3385-91 (2011).

Lin HN, Notredame C, Chang JM, Sung TY and Hsu WL.
"Improving the alignment quality of consistency based aligners with an evaluation function using synonymous protein words."
PLoS One, 6(12):e27872 (2011).

Merhej V, Notredame C, Royer-Carenzi M, Pontarotti P and Raoult D.
"The rhizome of life: the sympatric Rickettsia felis paradigm demonstrates the random transfer of DNA sequences."
Mol Biol Evol, 28(11):3213-23 (2011).

Rius J, Cores F, Solsona F, van Hemert JI, Koetsier J and Notredame C.
"A user-friendly web portal for T-Coffee on supercomputers."
BMC Bioinformatics, 12:150 (2011).

Taly JF, Magis C, Bussotti G, Chang JM, Di Tommaso P, Erb I, Espinosa-Carrasco J, Kemena C and Notredame C.
"Using the T-Coffee package to build multiple sequence alignments of protein, RNA, DNA sequences and 3D structures."
Nat Protoc, 6(11):1669-82 (2011).

Torrent M, Di Tommaso P, Pulido D, Nogues MV, Notredame C, Boix E and Andreu D.
"AMPA: an automated web server for prediction of protein antimicrobial regions."
Bioinformatics, 28(1):130-1 (2011).

Erb I, Gonzalez-Vallinas JR, Bussotti G, Blanco E, Eyras E and Notredame C.
"Use of ChIP-Seq data for the design of a multiple promoter-alignment method."
Nucleic Acids Res, in press.

BIOINFORMATICS AND GENOMICS

Group: Comparative Genomics

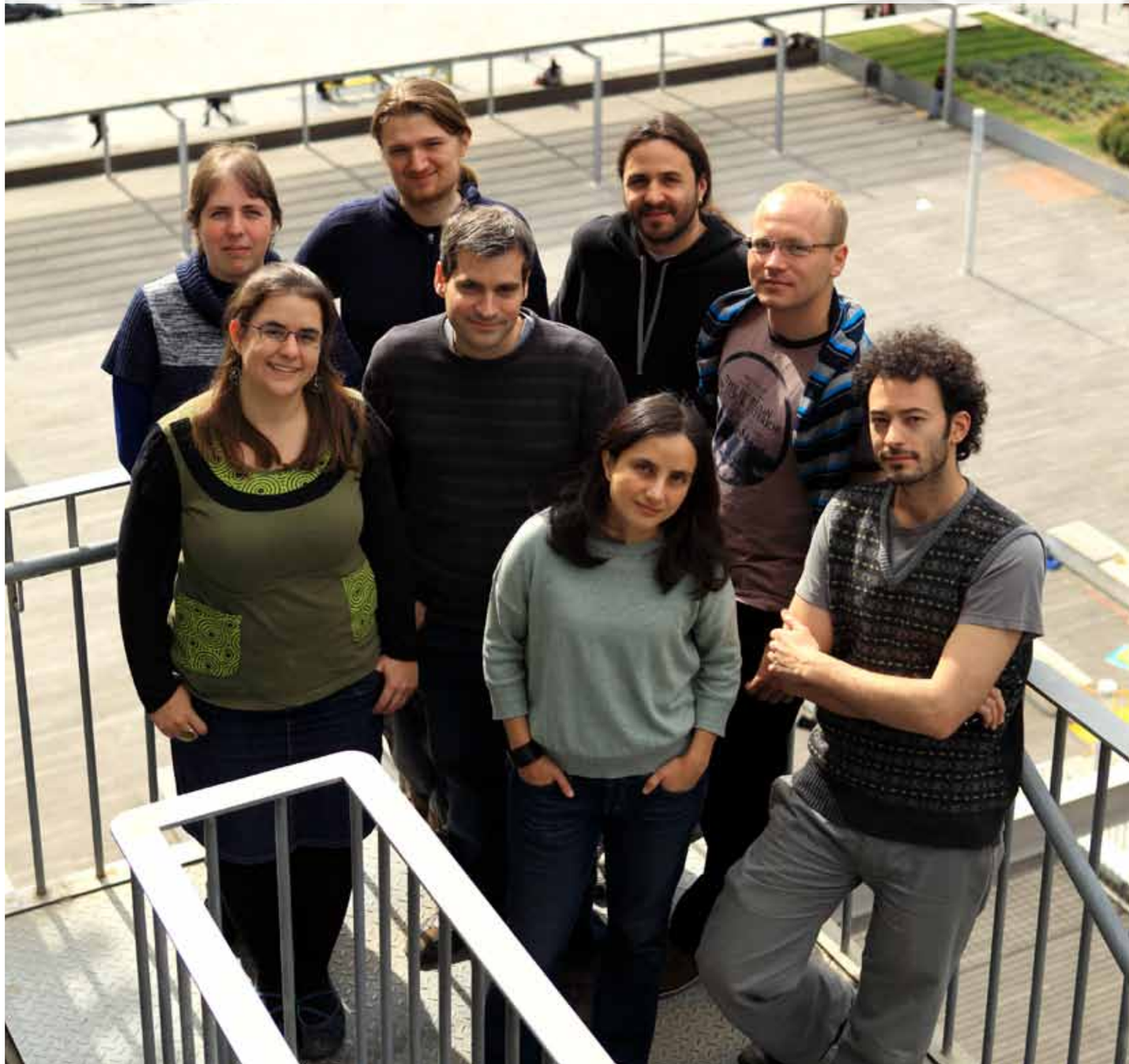
Group Structure:

Group Leader: Toni Gabaldón

Postdoctoral Fellows: Jaime Huerta-Cepas, Gabriela Aguileta, Fran Supek, Damien de Vienne, Ester Saus

Students: Salvador Capella-Gutiérrez, Leszek Pryscsz, Alexandros Pittis

Technician: Marina Marcet-Houben



SUMMARY

Our research interests are focused around the use of comparative genomics and phylogenomics to study the origin, evolution and function of complex biological systems. This includes understanding how specific biochemical pathways, protein complexes or cellular organelles emerged and evolved as well as using this evolutionary information to gain insight into their function. Through collaborations with experimental groups we apply comparative genomics to discover new mechanisms and genes involved in interesting processes, especially those of clinical relevance (see lines of research). On the technical side, our work often involves the development of new bioinformatics tools and algorithms that we make available to the community. You can access more info at *http://gabaldonlab.crg.es*

RESEARCH PROJECTS

1. Discovery of New Genes Involved in Mitochondrial Disease

Mitochondria play a central role in the cellular metabolism and the impairment of many mitochondrial proteins leads to disease. The list of such diseases is continuously growing and includes Parkinson, Alzheimer and Huntington diseases. Despite recent advances, the molecular basis of many mitochondrial diseases is yet to be understood and there is growing need to identify disease-causing genes and to unravel their functions. Recently, proteomics analyses have identified a large set of proteins that function inside the mitochondrion, the so-called mitochondrial proteome. Most of these proteins are not functionally characterized and it is expected that many of them may be involved in mitochondrial diseases. Moreover, other mitochondrial proteomic sets are being characterized in different species facilitating an evolutionary analysis of the mitochondrial system. The aim of this research line is to integrate different types of data and automatically combine them in order to facilitate the identification and functional characterization of mitochondrial disease-related genes.

2. Comparative Genomics of Fungal Pathogens

Fungal infections constitute an ever-growing and significant medical problem. Diseases caused by such pathogens range from simple toe nail infections, to life-threatening systemic mycoses in patients with impaired immune systems. The molecular mechanisms driving invasion of mammalian hosts by fungal pathogens poses many scientifically challenging problems, which are as yet little understood. The ability to infect humans has emerged in several lineages throughout the fungal tree of life. Therefore, the problem of elucidating the mechanism for pathogenesis of fungi, as proposed here, can be approached with an evolutionary perspective by detecting specific adaptations in pathogenic lineages. This year we have investigated the genomic changes that underlie the emergence of pathogenesis in the *Candida glabrata* and *Candida parapsilosis* clades by sequencing genomes and transcriptomes of related species and strains that differ in their virulence traits.

3. Phylogenomics and Genome Evolution

In the genomic era it has been possible to move from the evolutionary analysis of single protein families (phylogenetics) to that of complete genomes and proteomes (phylogenomics). To achieve this transition new tools have been developed that allow the large-scale reconstruction of thousands of phylogenetic trees in an automatic way. This computerization of the whole process of tree construction often involves the use of standard parameters and conditions for all tree families, inevitably resulting in poor or incorrect phylogenies in many cases. Moreover, interpreting such type of complex data poses many difficulties and does require the development of novel algorithms, tools, forms of representing the data and even new semantics and concepts. We combine the development of original algorithms to treat phylogenomic data with its application to gain knowledge on problems of biological relevance (see Figure 1). In particular we are interested in developing post-processing methods to interpret sequence alignments and phylogenetic trees in a large-scale and to mine such data to find evidence for functional interactions between proteins.

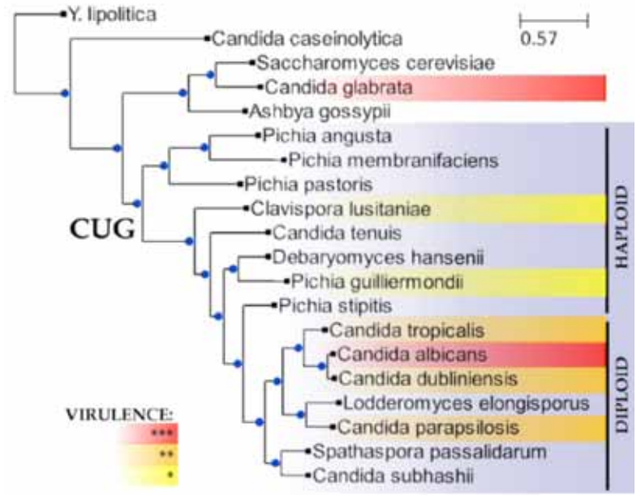


Figure 1.
Phylogenetic relationships of currently sequenced Candida species differing in their degree of virulence. The implied scenario involves parallel independent origins of the virulence towards humans.

4. Evolution of the Eukaryotic Cell

Every eukaryotic organism shows a high level of sub-cellular compartmentalization that is significantly more intricate than the most complex prokaryotic cell. How such degree of complexity came to be is still not fully understood. In this context, endo-symbiotic events with bacterial organisms have been proposed to be the source of a number of organelles including mitochondria, chloroplasts and peroxisomes. Only recently, it has been possible to contrast these hypotheses with the growing availability of completely sequenced genomes and organellar proteomic data. We use large-scale evolutionary analyses to investigate the origin and evolution two most widespread organelles for which an endosymbiotic origin has been proposed: mitochondria and peroxisomes.

PUBLICATIONS

Pryszcz LP, Huerta-Cepas J, Gabaldón T.
"MetaPhOrs: orthology and paralogy predictions from multiple phylogenetic evidence using a consistency-based confidence score."
Nucleic Acid Res, 39(5):e32 (2011).

Huerta-Cepas J, Capella-Gutiérrez S, Pryszcz LP, Denisov I, Kormes D, Marcet-Houben M, Gabaldón T.
"PhylomeDB v3.0: An expanding repository of genome-wide collections of phylogenies, alignments and orthology and paralogy predictions."
Nucleic Acid Res, 39(Database issue):D556-60 (2011).
Featured article by NAR's editors.

Huerta-Cepas J, Gabaldón T.
"Assigning duplications to relative temporal scales in genome-wide studies."
Bioinformatics, 27(1):38-45 (2011).

Roetzer A, Gabaldón T, Schüller C.
"From Saccharomyces cerevisiae to Candida glabrata in a few easy steps: important adaptations for an opportunistic pathogen."
FEMS Microbiol Lett, 314(1):1-9 (2011).

Roetzer A, Klopf E, Gratz N, Marcet-Houben M, Hiller E, Rupp S, Gabaldón T, Kovarik P, Schüller C.
"Regulation of Candida glabrata oxidative stress resistance is adapted to host environment."
FEBS Lett, 585(2):319-27 (2011).

Marcet-Houben M, Gabaldón T.
"TreeKO: A duplication-aware algorithm for the comparison of phylogenetic trees."
Nucleic Acid Res, 39(10):e66 (2011).

Neuhor G, Naegeli A, Titos I, Theler D, Greber B, Diez J, Gabaldón T, Mendoza M, Barral Y.
"A Midzone-Based Ruler Adjusts Chromosome Compaction to Anaphase Spindle Length."
Science, 332(6028):465-8 (2011).

Vellozo AF, Véron AS, Baa-Puyoulet P, Huerta-Cepas J, Cottret L, Febvay G, Calevro F, Rahbé Y, Douglas AE, Gabaldón T, Sagot MF, Charles H, Colella S.
"CycADS: an annotation database system to ease the development and update of BioCyc databases."
Database, 2011:bar008 (2011).

Huerta-Cepas J, Dopazo J, Huynen MA, Gabaldón T.
"Evidence for short-time divergence and long-time conservation of tissue-specific expression after gene duplication."
Brief Bioinform, 12(5)442-8 (2011).

Sánchez R, Serra F, Tárraga J, Medina I, Carbonell J, Pulido L, de María A, Capella-Gutíerrez S, Huerta-Cepas J, Gabaldón T, Dopazo J, Dopazo H.
"Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypothesis testing."
Nucleic Acids Res, 39(Webserver Issue):W470-4 (2011).

Faria R, Pinheiro A, Gabaldón T, Weiss S, Alexandrino P.
"Molecular tools for species discrimination and detection of hybridization between two closely related Clupeid fishes Alosa alosa and A. fallax."
J Appl Ichtyil, 27(supl.3):16-20 (2011).

Lopes-Silva L, Marcet-Houben M, Zerlotini A, Gabaldón T, Oliveira G, Alves-Nahum L.
"Evolutionary histories of expanded peptidases families in Schistosoma mansoni."
Mem Inst Oswaldo Cruz, 106(7):846-77 (2011).

Book chapters

Marcet-Houben M and Gabaldón, T.
"Evolution of fungi and their respiratory metabolism."
In: *Evolutionary Biology: Concepts, Biodiversity, Macroevolution and Genome Evolution*, p. 257-272, Edited by Pontarotti P, Springer-Verlag, Berlin (2011).

Peña A, Teeling H, Huerta-Cepas J, Santos F, Yarza P, Brito-Echeverría J, Lucio M, Schmitt-Kopplin P, Meseguer I, Schenowitz C, Dossat C, Barbe V, Dopazo J, Rosselló-Mora R, Schüler M, Glöckner FO, Amann R, Gabaldón T, Antón J.
"From Genomics to Microevolution and Ecology: The Case of Salinibacter ruber."
In: *Halophiles and Hypersaline Environments: Current Research and Future Trends*, p. 109-122, Edited by Ventosa A, Oren A, Ma Y, Springer-Verlag, Berlin (2011).

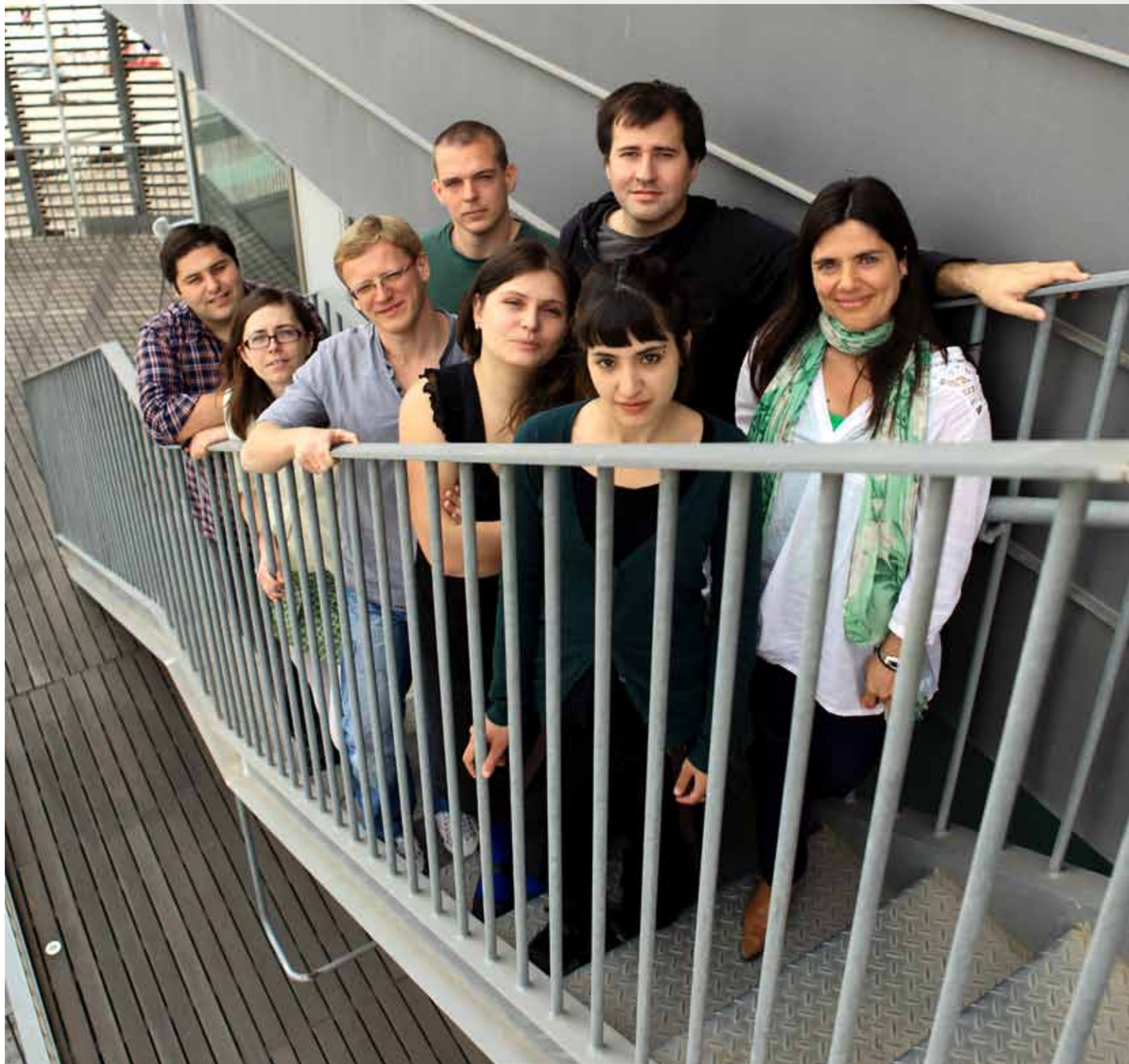
Gabaldón T.
"Mitochondrial origins."
In: *Organelle Genetics: Evolution of organelle genomes and gene expression*, p. 3-18, Ed. Charles Bullerwell, Springer-Verlag, Berlin (2011).

BIOINFORMATICS AND GENOMICS

Group: Evolutionary Genomics

Group Structure:

Group Leader:	Fyodor Kondrashov
Graduate Students:	Margarita Meer, Inna Povolotskaya, Onuralp Soylemez
Postdoctoral Fellows:	Peter Vlasov, Romain Derelle, Tobias Warnecke
Research Assistants:	Michael Breen, Maria Plyuscheva



SUMMARY

Our laboratory is engaged in the study of evolutionary biology in the broader sense, without limiting our queries to any specific organisms of mechanisms of action. In principle, there are just three fundamental mechanisms that together encompass the evolutionary process: the emergence of new variability through the mutational process, and the action and genetic drift and selection on this variability. The interplay and action of these three factors leads to the diversity all evolutionary phenomena that we observe and study. In particular, since selection acts on the level of the phenotype the question of how genotypic changes manifest themselves on the phenotype becomes a crucial question from the perspective of how selection shapes the direction and tempo of the evolutionary process.

After the success of the past year we have taken some time to explore new areas of research. In particular, several papers dealing with the issues of epistasis and protein evolution are in review, along with a paper on stop codon evolution. However, most of the work accomplished in 2011 will be published, hopefully, in 2012 and be the subject of the next year's review. The published work of the past year has focused on the questions of the evolutionary roots of eukaryote species, the geographic distribution of a couple of avian species and a conceptual review of the evolutionary importance of transcriptional error rates.

RESEARCH PROJECTS

1. The Consequences of Error-Prone Gene Expression

The production of functional proteins is a delicate task, involving a number of complex molecular machines. None of these machines, from the RNA polymerase, which generates the initial transcript, to the ribosome, which translates processed mRNA into protein, are perfect. In fact, errors during gene expression are commonplace. We are trying to understand how genetic systems have evolved to reduce the number of errors being made and/or minimize their impact on organismal fitness. We are also interested to characterize the trade-offs that come with the need for error proofing. For example, when does reducing the incidence of errors even further become undesirable because it affects the speed of gene expression?

In some bacteria, population sizes are very small so that selection cannot efficiently eliminate slightly deleterious variants from the population. Often, these variants are slightly deleterious because they increase the chance of errors during expression, so a larger fraction of transcripts is mis-spliced or more proteins fail to fold properly. We recently showed that across pathogenic bacteria, there is a predictable knock-on effect of reduced population size: In several lineage where population sizes have shrunk and more deleterious mutations accumulated, chaperone genes evolve faster, probably to compensate for the increased error burden.

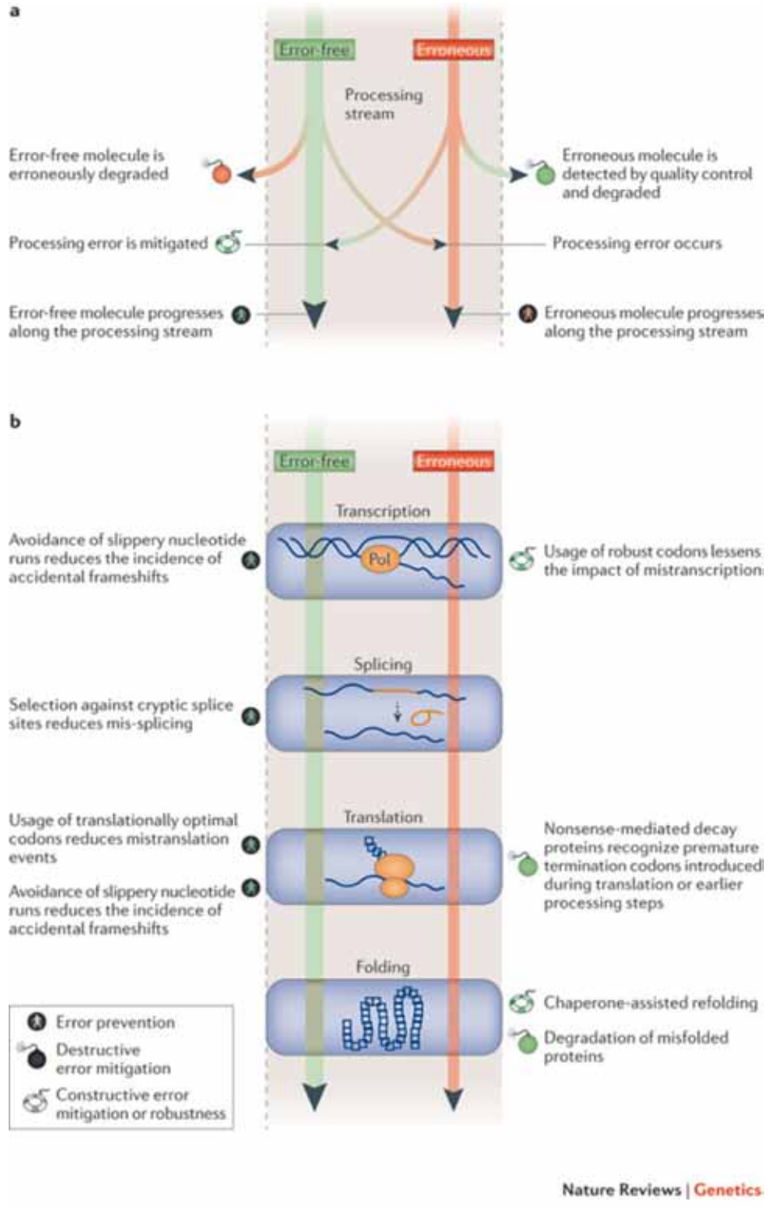


Figure 1:
a) At any point during gene expression, the relevant expression product is either error-free or has accumulated one or more errors. Both error-free and erroneous intermediates can progress further along the same processing stream or be degraded. In addition, new errors can be acquired and previous errors mitigated. For example, this can occur when chaperones unfold or disaggregate proteins that had initially failed to fold correctly, so that error-free gene products can become erroneous and vice versa. These alternative processing fates have different consequences for fitness, with presumably beneficial and detrimental fates shown by green and red symbols, respectively. Features of the gene that promote error-free processing constitute adaptations for error prevention. Conversely, we can speak of error mitigation when an error has already occurred, but that error is either corrected outright (termed constructive mitigation), as in the chaperone example above, or its impact is reduced, such as through targeting an erroneous transcript for degradation (termed destructive mitigation). b) Some key steps in the expression of protein-coding genes, showing examples of error prevention, as well as constructive and destructive mitigation. Pol, RNA polymerase.

2. Early Origins of the Eukaryotes

The origin and early evolution of eukaryotes is a fundamental question that remains hard to study. Until now this question has avoided a solution because of the rapid diversification of early eukaryotes relative to the phylogenetic distance to the nearest outgroups in Archaea. In a publication that appeared this year we utilized an elegant approach to sidestep the issue of distant outgroups by taking advantage of the symbiotic origin of the eukaryote. Prior to the Last Common Eukaryote Ancestor (LCEA) the eukaryotic cell precursor established a symbiosis with a bacterial cell, which has evolved into the mitochondria. Thus, many of the genes encoded in all of the modern eukaryota have a more recent common origin through the symbiotic horizontal transfer compared with the vertically transferred genes originally encoded in the eukaryotic cell precursor. We take advantage of this symbiotic event and reconstruct the eukaryotic phylogeny using protein sequences acquired through the massive lateral gene transfer event from the mitochondrial precursor and their orthologoues from Proteobacteria. This trick provides outgroup sequences that were close enough that our analysis does not suffer from the long branch attraction artifacts that have plagued all previous attempts and reliably reconstructs the root of the eukaryotic tree. Besides the purely fundamental value in understanding the early evolution of the most complex superkingdom of life, the present analysis finally provides the proper phylogenetic framework to all evolutionary studies involving the genomes of 'primitive' eukaryotes.

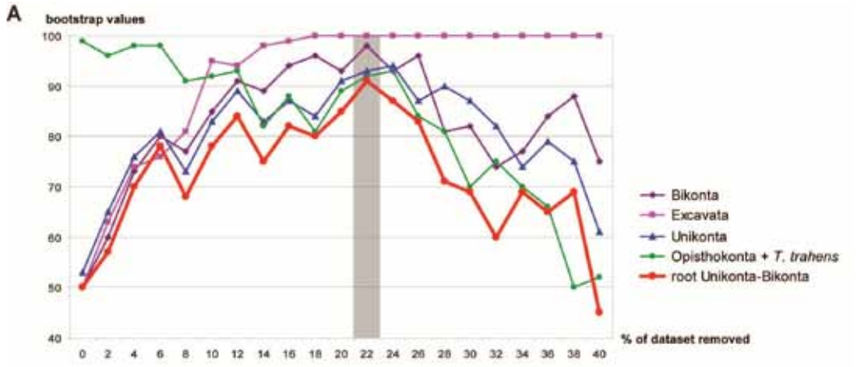


Figure 2

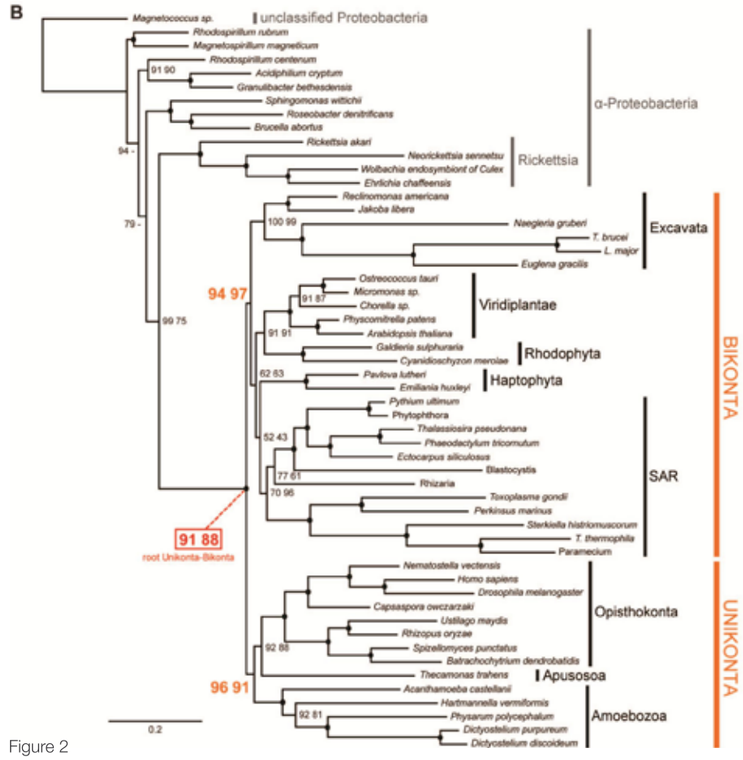


Figure 2

PUBLICATIONS

Koblik EA, Red'kin YA, Meer MS, Derelle R, Golenkina SA, Kondrashov FA, Arkhipov VY.
"Acrocephalus orinus: a case of mistaken identity."
PLoS One, 6(4):e17716 (2011).

Torruella G, Derelle R, Paps J, Lang BF, Roger AJ, Shalchian-Tabrizi K, Ruiz-Trillo I.
"Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains."
Mol Biol Evol, Epub 2011 Jul 18. (*)

Warnecke T, Rocha EP.
"Function-specific accelerations in rates of sequence evolution suggest predictable epistatic responses to reduced effective population size."
Mol Biol Evol, 28(8):2339-49 (2011). (**)

Warnecke T, Hurst LD.
"Error prevention and mitigation as forces in the evolution of genes and genomes."
Nat Rev Genet, 12(12):875-81 (2011).

Derelle R, Lang BF.
"Rooting the eukaryotic tree with mitochondrial and bacterial proteins."
Mol Biol Evol, Epub 2011 Dec 1.

(*) This publication results from the work of Dr. Romain Derelle at the Departament de Genètica and Institut de Recerca en Biodiversitat (Irbio), Universitat de Barcelona, Barcelona, Spain

(**) This publication results from the work of Dr. Tobias Warnecke at the Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom

BIOINFORMATICS AND GENOMICS

Group: Gene Function and Evolution

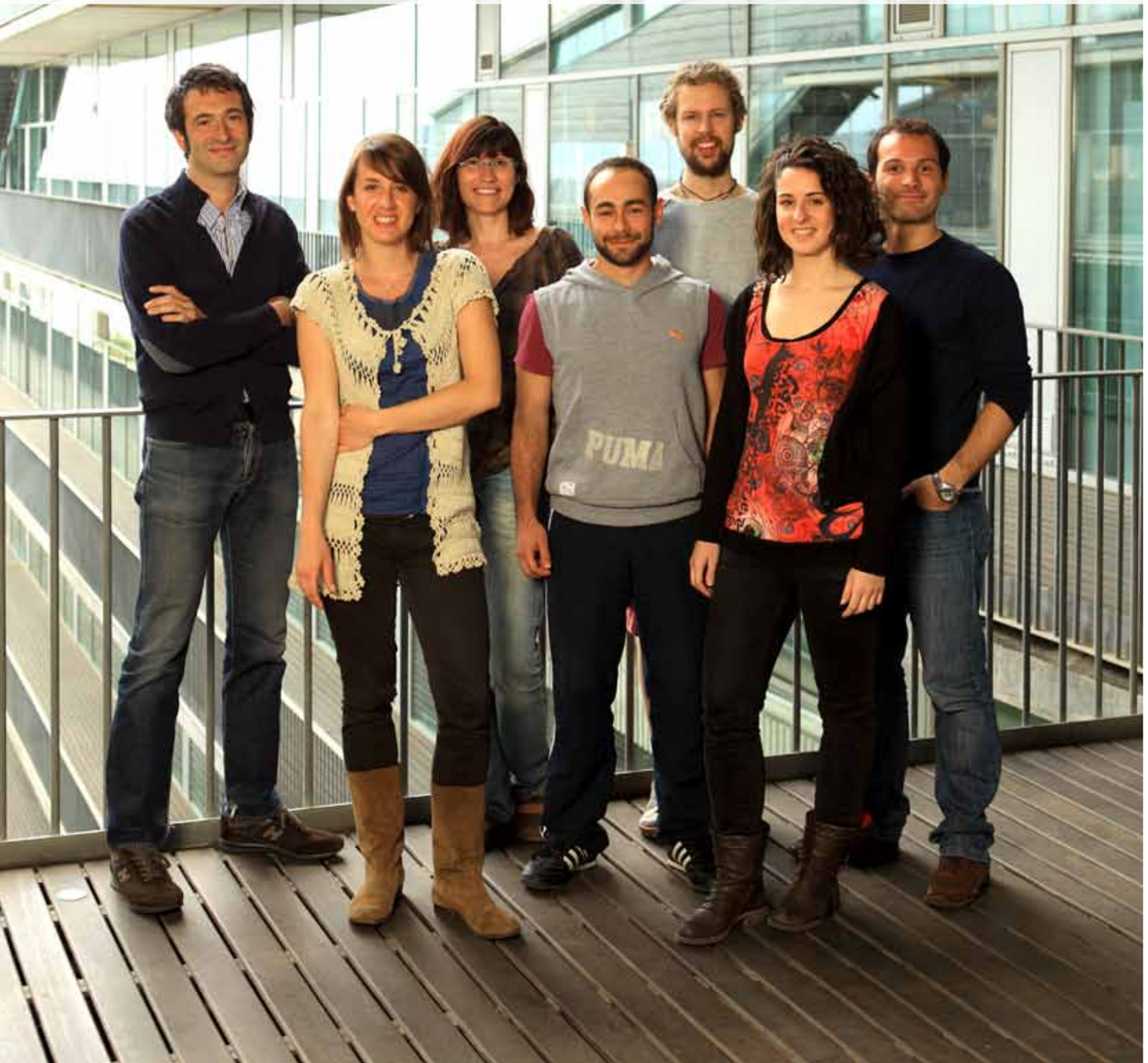
Group Structure:

Group Leader: Gian Gaetano Tartaglia

Postdoctoral Fellows: Marianela Masin (co-supervised with Dr. Salvatella, IRB), Matteo Bellucci (until June 2011), Benedetta Bolognesi (from November 2011)

Technicians: Silvia Rodriguez (from November 2011)

Doctoral Students: Federico Agostini, Davide Cirillo, Petr Klus



SUMMARY

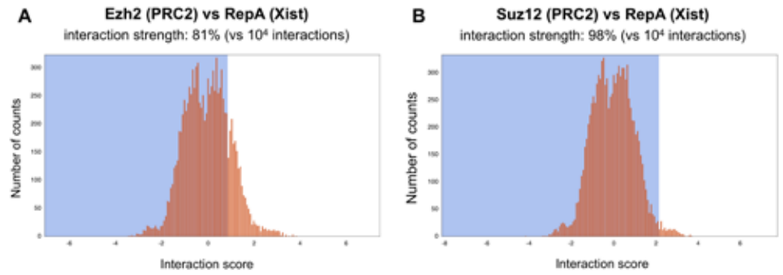
Long non-coding RNA have emerged as key players involved in the control of transcriptional and post-transcriptional gene regulatory pathways. Only a limited number of functional lncRNAs has been identified so far, but great regulatory potential is envisaged. We aim to identify novel protein-RNA interactions using our computational approach, catRAPID[1] and a number of *in vitro* experiments. We are also interested in the prediction of protein solubility[2] and interactomes of amyloid fibrils in the cellular context[3]. We are currently developing models for calculating the interaction potential of proteins with molecular chaperones.

RESEARCH PROJECTS

1. Large-Scale Predictions of Protein-RNA Associations

In 2011, we introduced a theoretical framework to study interactions between protein and RNA molecules. Our algorithm, catRAPID, is the first computational method able to perform large-scale predictions of protein-RNA associations[1]. We trained catRAPID on a large set of protein-RNA pairs available in the Protein Data Bank to discriminate interacting and non-interacting molecules using the only information contained in primary structures. In our method, secondary structure propensities account for 72% of the ability to predict protein-RNA associations, followed by hydrogen bonding (58%) and van der Waals (26%) contributions. catRAPID was greatly appreciated by many experts in the field - suffice to say that more than 2500 different users ran our web-server in the last year -. Collaborations with experimental scientists confirmed the quality of our predictions and a number of articles are currently in preparation. Our scientific strategy relies on the interaction with experimental groups to interpret existing data and propose validation of our predictions. At the same time, we exploit our computational methods to understand basic principles that underlie evolution of genomes. In 2011, we started to investigate the role played by non-coding transcripts in protein networks. The application of our prediction method represents a crucial point to advance in the comprehension of RNA-mediated regulation.

catRAPID webserver *http://tartaglialab.crg.cat/catrapid.html*



Prediction of protein-RNA associations: *In agreement with experimental evidence, we predict that A) Enhancer of zeste homolog 2 (Ezh2) interacts with Xist Repeat A region (RepA); B) Suppressor of zeste 12 protein homolog (Suz12) interacts with RepA. The interaction strength is the propensity to bind with respect to a control set (10^4 protein-RNA associations) and is indicated with blue areas.*

2. Mining Proteostasis by Playing with Protein Concentration, Aggregation and Disorder

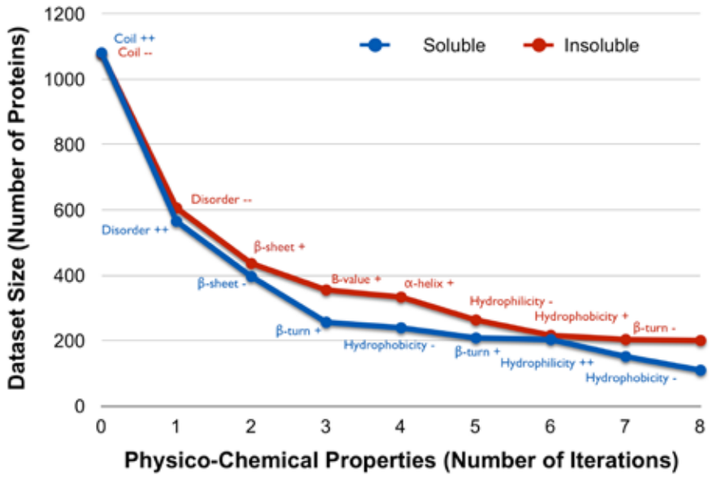
Protein aggregation and structural disorder are directly linked to a number of pathologies, ranging from cancer to neurodegenerative diseases. At high concentrations proteins are more prone to aggregate and establish aberrant interactions with disordered proteins[3]. Protein concentration and aggregation are connected[4,5], but a crucial relation also exists between concentration and the mechanism by which disordered proteins may become harmful for cell viability[6]. Indeed, over-expression of disordered proteins results in severe cytotoxicity, as at high concentrations disordered regions can establish promiscuous interactions, which alter proteostasis and cause impaired cell viability[6]. In this project started in 2011 with Dr. Lehner, we are using *S. cerevisiae* as a model to characterize the interplay among concentration,

aggregation and disorder. In order to select appropriate candidates, we employ a series of predictions using algorithms to estimate the aggregation as well as the disorder propensities of specific amino acid sequences. By measuring the growth rate of the strains we plan to understand the concentration levels at which disordered proteins or aggregation prone ones become toxic. Crucial insight arises from a comparison between proteins that do not cause toxicity when over-expressed. In the near future, we aim to investigate the interactomes of disordered or aggregation prone proteins3 and how protein concentration may affect the type of interactors recruited, especially at those levels in which toxicity arises.

3. Large-Scale Predictions of Protein Solubility

In order to investigate the relationship between the thermodynamics and kinetics of protein aggregation, we compared the solubility of proteins with their aggregation rates. We found a significant correlation between these two quantities by considering a database of protein solubility values measured using an *in vitro* reconstituted translation system containing about 70% of *E. coli* proteins. The existence of such correlation suggests that the thermodynamic stability of the native states of proteins relative to the aggregate states is closely linked with the kinetic barriers that separate them. In order to create the possibility of conducting computational studies at the proteome level to investigate further this concept, we developed a method of predicting the solubility of proteins (ccSOL) based on their physicochemical properties.

ccSOL webserver *http://tartaglialab.crg.cat/ccsol.html*



Protein solubility: *An iterative procedure was employed to characterize properties of soluble and insoluble proteins in* ***E. coli****. The signs +/- indicate accruement/depletion of physico-chemical properties in the soluble and insoluble datasets [(+) positive selection, (+ +) positive selection and steep gradient, (–) negative selection, (– –) negative selection and steep gradient].*

4. The Clever Machine

In 2011, we started to develop a tool for feature discovery in protein datasets. The ambitious name "the clever machine" refers to the program's capability of deriving the result without any parameterization or additional information from the user, who simply provides the data sets. The proposed output of the tool is a set of graphs and textual description highlighting features of individual data sets. Furthermore, the tool will be able to detect datasets without any significant features and report such finding to the user (for example, randomly generated datasets or a cell lysate containing a mixture of multiple protein types). The program relies on a large number of experimental scales, each of which is translated into a protein profile for all of the individual proteins. Statistics about those profiles are then combined and compared between the datasets with a complex set of rules guiding the decision process.

5. Protein-Protein Interactions and α-Synuclein Degradation in Parkinson's Disease

Aberrant α-synuclein degradation is implicated in Parkinson's disease (PD) pathogenesis because the protein accumulates in the Lewy inclusion bodies associated with the disease. Little is known, however, about the degradation pathways of α-synuclein and its pathogenic mutations A30P and A53T. It has been observed that wild-type a-synuclein is selectively translocated into lysosomes for degradation by the chaperone-mediated autophagy pathway. Together with the laboratories of Evolutionary Genomics headed by Dr. Fyodor Kondrashov and Genetic Causes of Disease headed by Prof. Xavier Estivill, we hypothesized the α-synuclein and Lamp2a are subject to co-evolution, and that alteration of their binding ability is a crucial factor that could have led the PD in humans. In order to validate our hypothesis and to address this fundamental question we carried out experiments of co-immunoprecipitation probing both wild type and A53T α-synuclein against different substitutions in Lamp2a sequences. We are now planning to test the effect of the different interactions in isolated intact lysosomes to evaluate the pathologic accumulation of a-synuclein in neuronal cells.

1. Bellucci, M., Agostini, F., Masin, M. & Tartaglia, G. G. Predicting protein associations with long noncoding RNAs. ***Nat. Methods*** *8, 444–445 (2011).*
2. Agostini, F., Vendruscolo, M. & Tartaglia, G. G. Sequence-Based Prediction of Protein Solubility. ***Journal of Molecular Biology*** *(2011). doi:10.1016/j.jmb.2011.12.005*
3. Olzscha, H. ***et al.*** *Amyloid-like Aggregates Sequester Numerous Metastable Proteins with Essential Cellular Functions.* ***Cell*** *144, 67–78 (2011).*
4. Tartaglia, G. G., Pechmann, S., Dobson, C. M. & Vendruscolo, M. Life on the edge: a link between gene expression levels and aggregation rates of human proteins. ***Trends Biochem*** *Sci 32, 204–6 (2007).*
5. Baldwin, A. J. et al. Metastability of native proteins and the phenomenon of amyloid formation. ***J. Am. Chem. Soc.*** *133, 14160–14163 (2011).*
6. Vavouri, T., Semple, J. I., Garcia-Verdugo, R. & Lehner, B. Intrinsic Protein Disorder and Interaction Promiscuity Are Widely Associated with Dosage Sensitivity. ***Cell*** *138, 198–208 (2009).*

PUBLICATIONS

Agostini F, Vendruscolo M, Tartaglia GG.
"Sequence-Based Prediction of Protein Solubility."
J Mol Biol, 2011 Dec 9 [Epub ahead of print].

Murakami T, Yang SP, Xie L, Kawano T, Fu D, Mukai A, Bohm C, Chen F, Robertson J, Suzuki H, Tartaglia GG, Vendruscolo M, Kaminski Schierle GS, Chan FT, Moloney A, Crowther D, Kaminski CF, Zhen M, St George-Hyslop P.
"ALS mutations in FUS cause neuronal dysfunction and death in Caenorhabditis elegans by a dominant gain-of-function mechanism."
Hum Mol Genet, 21(1):1-9 (2011). (*)

Mossuto MF, Bolognesi B, Guixer B, Dhulesia A, Agostini F, Kumita JR, Tartaglia GG, Dumoulin M, Dobson CM, Salvatella X.
"Disulfide bonds reduce the toxicity of the amyloid fibrils formed by an extracellular protein."
Angew Chem Int Ed Engl, 50(31):7048-51 (2011).

Baldwin AJ, Knowles TP, Tartaglia GG, Fitzpatrick AW, Devlin GL, Shammas SL, Waudby CA, Mossuto MF, Meehan S, Gras SL, Christodoulou J, Anthony-Cahill SJ, Barker PD, Vendruscolo M, Dobson CM.
"Metastability of native proteins and the phenomenon of amyloid formation."
J Am Chem Soc, 133(36):14160-3 (2011). (*)

Bellucci M, Agostini F, Masin M and Tartaglia GG.
"Predicting protein associations with long noncoding RNAs."
Nat. Methods, 8:444–445 (2011).

Jahn TR, Kohlhoff KJ, Scott M, Tartaglia GG, Lomas DA, Dobson CM, Vendruscolo M, Crowther DC.
"Detection of early locomotor abnormalities in a Drosophila model of Alzheimer's disease."
J Neurosci Methods, 197(1):186-9 (2011).

Raimondi S, Guglielmi F, Giorgetti S, Di Gaetano S, Arciello A, Monti DM, Relini A, Nichino D, Doglia SM, Natalello A, Pucci P, Mangione P, Obici L, Merlini G, Stoppini M, Robustelli P, Tartaglia GG, Vendruscolo M, Dobson CM, Piccoli R, Bellotti V.
"Effects of the known pathogenic mutations on the aggregation pathway of the amyloidogenic peptide of apolipoprotein A-I."
J Mol Biol, 407(3):465-76 (2011).

Szczepankiewicz O, Cabaleiro-Lago C, Tartaglia GG, Vendruscolo M, Hunter T, Hunter GJ, Nilsson H, Thulin E, Linse S.
"Interactions in the native state of monellin, which play a protective role against aggregation."
Mol Biosyst, 7(2):521-32 (2011). (*)

Olzscha H, Schermann SM, Woerner AC, Pinkert S, Hecht MH, Tartaglia GG, Vendruscolo M, Hayer-Hartl M, Hartl FU, Vabulas RM.
"Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions."
Cell, 144(1):67-78 (2011).

(*) All these publications result from the work of Dr. Gian Gaetano Tartaglia at the Dept. of Chemistry, University of Cambridge, United Kingdom.

CELL AND DEVELOPMENTAL BIOLOGY

Group: Intracellular Compartmentation
Vivek Malhotra is an ICREA Research Professor and the coordinator of the programme.

Group Structure:

Group Leader: Vivek Malhotra

Postdoctoral Fellows: Felix Campelo, David Cruz, Amy Curwin, Juan Duran, Patrik Erlmann, Sandra Mitrovic, Cristina Nogueira, Josse van Galen, Julien Villeneuve , Yuichi Wakana

PhD Students: Caroline Bruns, Ana Catarina De Olivera

Technician: Anne-Marie Alleaume, Maria Ortega, Jean-François Popoff, Margherita Scarpa

SUMMARY

We are interested in the mechanisms of membrane fission involved in the biogenesis of transport carriers at the TGN. Our lab also studies the overall process of cargo sorting and vesicle trafficking in both conventional and unconventional protein secretion. We are also interested in the mechanism of regulated secretion of mucins from the human Goblet cells with the intend to get a better understanding of chronic obstructive pulmonary disorders (COPD).

RESEARCH PROJECTS

1. Membrane Fission

Yuichi Wakana, Felix Campelo, David Cruz, Josse van Galen, Margherita Scarpa

The Golgi apparatus is composed of cisternae (flat membranes) that are stacked and kept near the centrioles in mammalian cells. Why such unique organization and spatial location? How is this organization regulated during protein transport? The secret here is membrane fission (cutting), an essential process for the generation of transport carriers. Membrane fission must be regulated to generate transport carriers commensurate with cargo size, prevent the formation of empty carriers or the complete conversion of Golgi into small vesicles during protein secretion.

A compound called illimaquinone (IQ) was identified based on its property to specifically vesiculate Golgi membranes into small 60-90 nm size vesicles (Takizawa *et al*., *Cell* 1993). IQ mediated Golgi vesiculation was reconstituted *in vitro*, which revealed the involvement of trimeric G protein subunits βγ (Jamora *et al*., *Cell* 1997). Soon thereafter we found that Gβγ activated a serine/threonine protein kinase called PKD (Jamora *et al*., *Cell* 1999). The next obvious step was to test whether PKD is required for formation of transport carriers from Golgi membranes during protein secretion. This was indeed the case and we found that inhibition of PKD blocked secretion of proteins, interestingly and specifically, from Golgi to the cell surface (Liljedahl *et al*., *Cell* 2001). How is PKD recruited to the Golgi membranes? We found that diacylglycerol (DAG) was required for the recruitment of PKD to the Golgi membranes (Baron and Malhotra, *Science* 2001). In collaboration with Dr. Thomas Suefferlein, we have recently found that PKD also binds to ARF1 at the Golgi. There are numerous reports now on the involvement of PKD in protein secretion but the challenge is to understand the molecular mechanism of downstream events. PKD has been shown to activate a lipid kinase called PI4KIIIb, which converts phophatidylinositol (PI) into phophatidylinositol 4-phosphate (PI4P), a lipid required for Golgi to cell surface transport. PKD also phosphorylates a ceramide transfer protein called CERT. The role of ceramide at the TGN has not been elucidated yet (Bard and Malhotra, *Ann Rev Cell and Dev Biol* 2006). Clearly, a number of components involved in PKD dependent TGN to cell surface transport have been identified and in order to gain mechanistic insights, we reconstituted this process in permeabilized mammalian cells. This allowed us to purify a class of PKD dependent transport carriers that were analyzed by mass spectrometry, which led to the characterization of a number of new proteins. This assay is being used to reveal the role of proteins and lipids that assemble at the TGN for the generation of TGN to cell surface transport carriers (Wakana *et al*., in revision).

2. New Transport Components

Amy Curwin, Julien Villeneuve, Patrik Erlmann, Ana Catarina Vidinhas De Olivera, Maria Ortega, David Cruz, Felix Campelo, Anne-Marie Alleaume

A genome wide screen was carried out to identify new components required for protein secretion. From the 22,000 genes tested, 110 gene products were identified to be essential for secretion. This contained 22 previously known components. The rest were cloned, tagged and expressed to identify the intracellular location of the cognate proteins. 20 genes products were further selected, based on their localization to compartments of the secretory pathway (Bard *et al*., *Nature* 2006). The new genes of interest are called TANGO for Transport And Golgi Organization. TANGO1 and a previously identified gene called twinstar/cofilin are being further characterized for their roles in protein transport and Golgi organization.

TANGO1 is required for loading cargo into transport carriers at the Endoplasmic Reticulum (ER) exit sites (Saito *et al*., *Cell* 2009). Cofilin is required for the sorting of secretory cargo at the TGN (von Blume *et al*., *JCB* 2009; von Blume *et al*., Dev. *Cell* 2011).

Our current aims are to understand the mechanism by which TANGO1 loads cargo without entering the transport carrier, and the process by which cofilin dependent actin remodeling helps in cargo sorting.

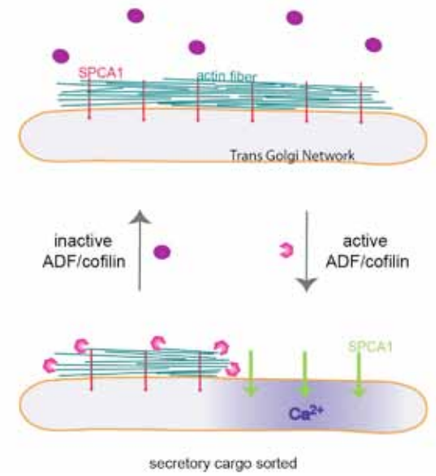


Figure 1.
Cofilin dependent actin depolymerization at the TGN is required for SPCA1 mediated Ca^{2+} pumping into the lumen TGN, regulating the sorting of a subset of secretory cargo.
.

3. Unconventional Secretion

Juan Duran, Caroline Bruns, Josse van Galen and Amy Curwin

Some of the interleukins, fibroblast growth factor-2 (FGF2), inhibitor of macrophage migration (MIF), Galectins etc., are secreted from cells without entering the ER-Golgi pathway. These secreted proteins are key players in the immune response, cell growth, angiogenesis, but the mechanism of their release from cells remains mysterious. We have found that the Golgi associated protein called GRASP in Dictyostelium is required for secretion of a protein called AcbA. AcbA, like the proteins mentioned above, lacks a signal sequence necessary for targeting to the ER, and is secreted unconventionally (Kinseth *et al*., *Cell* 2007). Unconventional secretion of AcbA ortholog Acb1 in yeast requires GRASP, genes involved in autophagosome formation, transport to the early endosome, multivesicular body formation and the cell surface specific T-SNARE (Duran *et al*., *JCB* 2010). We are using yeast and mammalian cells to reveal the mechanism of unconventional secretion.

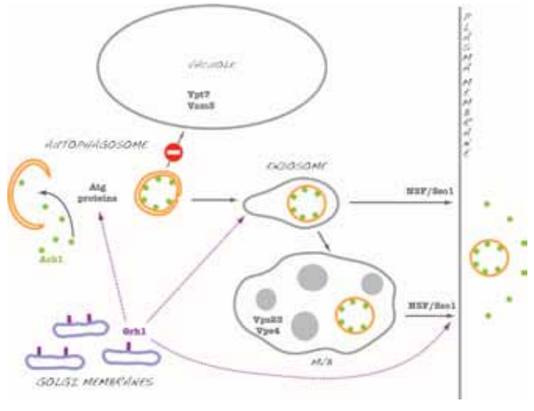


Figure 2.
The Acb1 secretion pathway. Cytosolic Acb1 is packaged into autophagosomes, which fuse with early endosomes. The early endosomes containing Acb1 either fuse directly with the cells surface, or more likely mature into a multivesicular body (MVBs). The MVB fuses with the cell surface to release exosomes containing Acb1. Grh1 is required for the secretion of Acb1 and has to be membrane associated for its role in this pathway. However, the exact site of action for Grh1 in unconventional secretion remains unknown.

4. Mechanism of Mucin Secretion

Sandra Mitrovic, Cristina Nogueira, Jean-François Popoff

Mucins are heavily glycosylated secretory proteins, however, the mechanism by which they are sorted at the TGN, packed into secretory storage granules and then secreted in a regulated manner is not known. We have reconstituted mucin secretion in human goblet cells and screened the genome to identify components involved in mucin synthesis and secretion. A large number of gene products have been identified and our aim is to understand their mechanism of action in mucin secretion. We are specifically interested in cell surface proteins that can be targeted to inhibit mucin secretion. This has the potential of developing valuable therapeutics for Asthma and chronic obstructive pulmonary disorders (COPD).

5. A Golgi Organization Specific Cell Cycle Checkpoint

Julien Villeneuve

Inhibiting changes in Golgi organization prevents entry of cells into mitosis (Sutterlin *et al*., Cell 2004). Thus a mechanism exists to monitor organizational changes in Golgi and if there is any defect, entry of cells into mitosis is blocked. What is the molecular mechanism of this Golgi specific checkpoint? How is this event coordinated with other check points that are activated upon DNA damage and defective spindle dynamics? A combination of *in vitro* approaches and system wide siRNA is being used to identify the Golgi specific cell cycle checkpoint.

PUBLICATIONS

Malhotra V and Campelo F.
"PKD Regulates Membrane Fission to Generate TGN to Cell Surface Transport Carriers."
Cold Spring Harb Perspect Biol, 3(2). pii: a005280 (2011).

Von Blume J, Alleaume AM, Cantero-Recasens G, Curwin A, Carreras-Sureda A, Zimmermann T, van Galen J, Wakana Y, Valverde MA and Malhotra V.
"ADF/Cofilin regulates secretory cargo sorting at the TGN via the Ca^{2+} ATPase SPCA1."
Dev Cell, 20:652-662 (2011).

Saito K, Yamashiro K, Ichikawa Y, Erlmann P, Kontani K, Malhotra V and Katada T.
"cTAGE mediates collagen secretion through interaction with TANGO1 at Endoplasmic reticulum exit sites."
Mol Biol Cell, 22:2301-2308 (2011).

Malhotra V and Erlmann P.
"Protein export at the ER: TANGO1 helps COPII carriers grow in size."
EMBO J, 30:3475-3480 (2011).

Bruns C, McCaffery JM, Curwin AJ, Duran JM and Malhotra V.
"Biogenesis of a novel compartment for unconventional protein secretion."
J Cell Biol, 195:979-992 (2011).

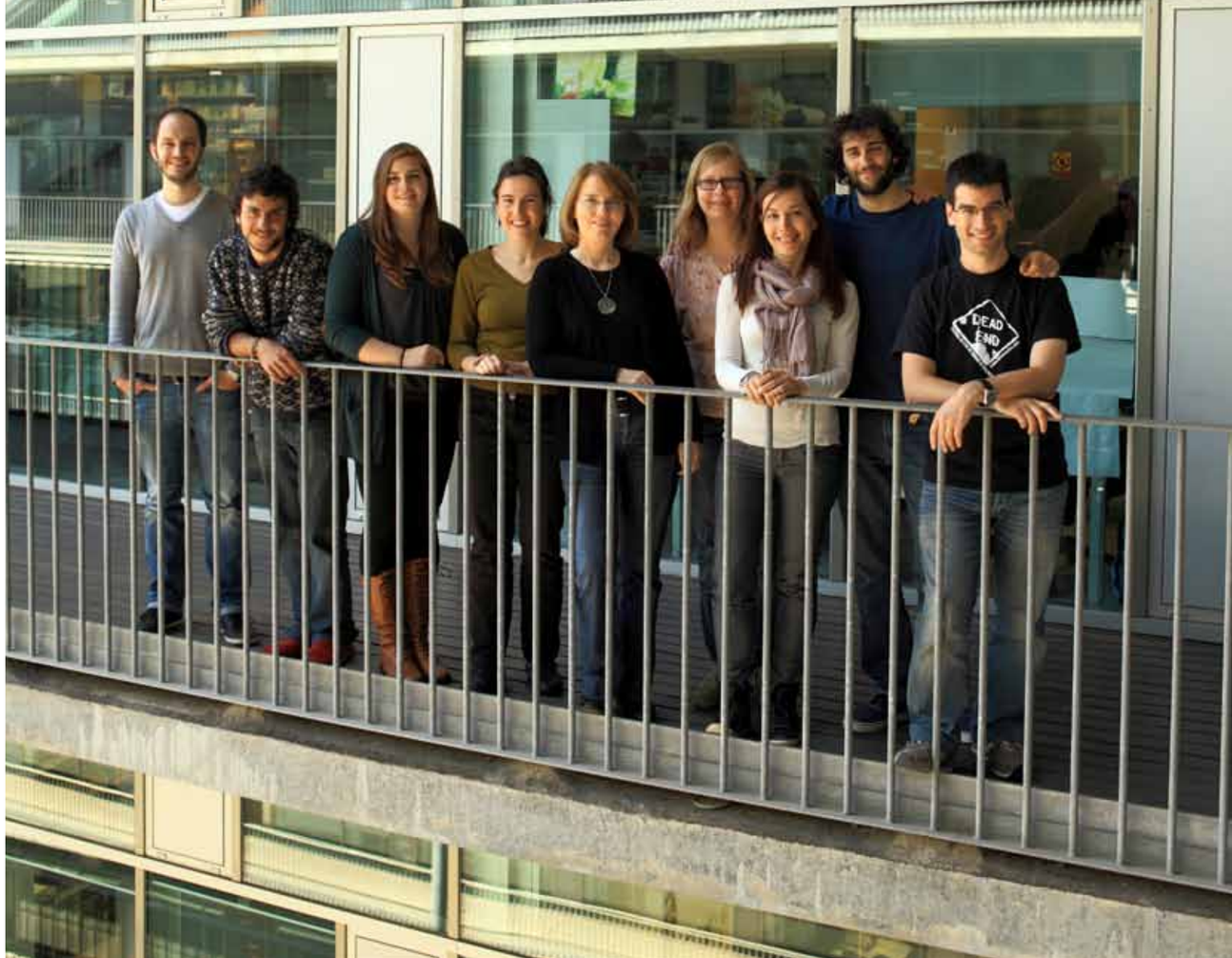
Campelo F and Malhotra V.
"Membrane Fission: the biogenesis of transport carriers."
Ann Rev Biochem, in press.

CELL AND DEVELOPMENTAL BIOLOGY

Group: Microtubule Function and Cell Division
Isabelle Vernos is an ICREA Research Professor and a member of the ERC Scientific Council

Group Structure:

Group Leader:	Isabelle Vernos
Postdoctoral Fellows:	Teresa Sardon (until June 2011), Sylvain Meunier, Roser Pinyol, Marti Badal, Nathalie Brouwers (from October 2011), Vanessa Ferreira (from July to December 2011)
Students:	Martin Schütz, Antonios Lioutas, Jacopo Scrofani, Tommaso Cavazza
Trainee:	Rebecca Adikes (from September 2011)
Technicians:	Leonor Avila, Nuria Mallol



SUMMARY

Research in my lab is directed at understanding the role of the microtubule network in cell organization and function. To address this question we study various microtubule-associated proteins (molecular motors and MAPs) and their regulators (kinases, phosphatases and the small GTPase Ran during M-phase). One major goal is to unravel how the self-organization of cellular components results in the morphogenesis of dynamic molecular machines.

Our favourite experimental system is the Xenopus egg extract for studies on cell cycle progression and regulation, microtubule dynamics, spindle assembly and chromosome behaviour (Karsenti and Vernos, 2001). We also perform experiments in human tissue culture cells and *in vitro* with purified components. Altogether these experimental approaches allow us to gain a better insight on the function and regulation of novel proteins involved in spindle assembly and dynamics.

RESEARCH PROJECTS

Cell division is characterized by the dramatic reorganization of the microtubule network into a spindle shaped apparatus that segregates the chromosomes into the two daughter cells. Spindle assembly and function rely on complex protein interaction networks that are finely regulated in time and in space. In addition to phosphorylation-dephosphorylation reactions, recent work has shown that the small GTPase Ran in its GTP bound form plays an important role in the spatial regulation of spindle assembly (Gruss and Vernos, 2004). To understand the molecular mechanism underlying cell division we study the process of microtubule nucleation and stabilization during M-phase and the role of molecular motors in bipolar spindle assembly and chromosome movements.

1. The Centrosome: a Cellular Regulatory Hub

The centrosome is the major site for microtubule nucleation in animal cells and its activity is finely regulated during the cell cycle. In late G2 and prophase, the pericentriolar material expands by recruiting additional components, such as the γ-tubulin-ring complex and as a result the MT nucleation activity of the centrosome increases. Several kinases from different families are recruited to the centrosome in G2/M phase and become specifically activated as the cell enters mitosis. We are studying how these kinases participate to centrosome maturation promoting microtubule nucleation and stabilization during cell division.

The centrosomal Aurora A kinase has been implicated in several important processes including centrosome maturation during G2, mitotic entry, centrosome separation and bipolar spindle assembly. We have previously shown that Aurora A interacts with TPX2 in a RanGTP dependent manner after nuclear envelope breakdown resulting in kinase activation. Using the Xenopus egg extract system we showed that Aurora A works though different mechanisms to regulate MT assembly during mitosis, ensuring bipolar spindle formation (Sardon *et al*, 2008). At the centrosome active Aurora A is required for efficient MT nucleation. In addition, it promotes MT growth by recruiting TACC3 that works in concert with XMAP215 to oppose the destabilizing activity of XKCM1 (Peset *et al*, 2005). We are currently studying the role of TACC3 in spindle assembly.

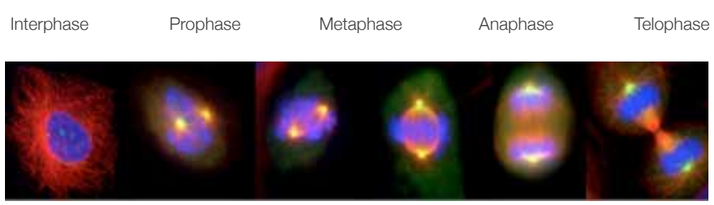


Figure 1
Immunofluorescence analysis of HeLa cells shows the specific localization of the kinase Aurora A to the centrosome and the spindle poles during mitosis. MTs are labelled in red, DNA in blue and Aurora A in green.

2. The RanGTP Pathway for MT Nucleation and Stabilization in M-Phase

After nuclear envelope breakdown, a centrosome independent pathway relying on a RanGTP gradient triggers MT nucleation and promotes MT stabilization in the vicinity of the condensed chromatin. We want to understand the molecular mechanism that triggers MT nucleation and stabilization through this acentrosomal pathway.

We previously identified TPX2 as being essential for this pathway and for spindle assembly both in mitosis and meiosis (Gruss and Vernos, 2004; Gruss *et al*., 2002; Wittmann *et al*., 2000).

To understand how the RanGTP dependent MT assembly pathway works we aim at identifying the proteins that are directly regulated by RanGTP during cell division. We recently identify the protein MCRS1 as one of them. MCRS1 is nuclear during interphase and during mitosis it localizes specifically to the minus–ends of the chromosomal MTs and the K-fibres protecting them from the MT depolymerase MCAK. MCRS1 unique localization reveals the existence of a specific mechanism controlling the stability and/or dynamic of this class of MTs during mitosis. Our results suggests that specific proteins may be 'loaded' through to the RanGTP pathway, on the K-fibres to confer them their properties in terms of organization and dynamics.

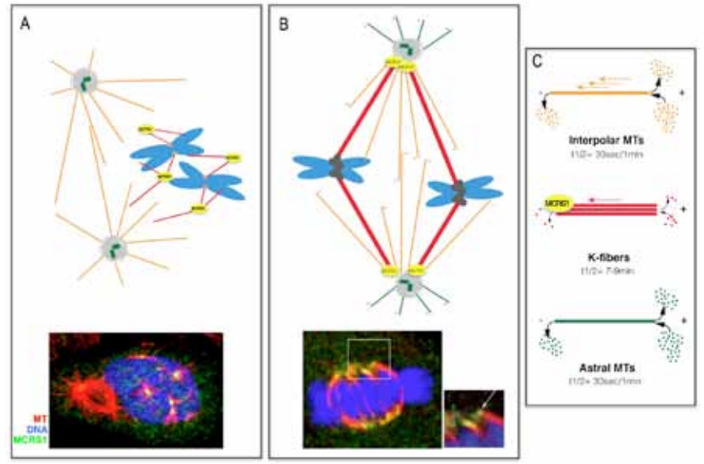


Figure 2
Schematic representation of the different classes of MTs that constitute the mitotic spindle and their dynamic properties. Centrosomes nucleate astral MTs. Some of them interact in an anti-parallel fashion (interpolar MTs). Other MTs are nucleated in an acentrosomal way close to the chromatin through a pathway that involves RanGTP. These MTs may be largely responsible for the formation of the K-fibres although this has not been demonstrated yet. The immunofluorescence images show the specific localization of MCRS1 (yellow) to the center of the chromosomal MT asters (A) and later on to the minus-ends of the K-fibres (B) (MTs are shown in red, DNA in blue). For more details see Meunier and Vernos (2011).

3. Role of Molecular Motors in Spindle Assembly and Chromosome Movements

Spindle bipolarity is essential for correct chromosome segregation but the mechanism underlying its establishment is still not completely understood. Furthermore very little is known about how metaphase spindles maintain a stable bipolar configuration before anaphase. Although it has been established that a balance of forces, generated by plus and minus-end directed motors, mostly Eg5 and dynein, is required for bipolar spindle assembly, whether these forces still play a role in metaphase is unknown.

We recently found that another motor, Hklp2, the human homologue of Xklp2 (Boleti *et al*, 1996; Wittmann *et al*, 1998) plays a role in bipolar spindle assembly and stability. Hklp2 localizes to the spindle microtubules and the chromosomes in metaphase. It is therefore a novel chromokinesin (Vanneste *et al*, 2011). Hklp2 steady state distribution is essential for its role in promoting the switch from the monopolar to the bipolar configuration and in stabilizing spindle bipolarity in metaphase (Vanneste *et al*, 2009). We are currently studying how Hklp2 functions and how it is regulated during the cell cycle.

PUBLICATIONS

Meunier S and Vernos I.
"K-fibre minus ends are stabilized by a RanGTP-dependent mechanism essential for functional spindle assembly."
Nat Cell Biol, 13(12):1406-14 (2011).

Article commented in:
New and Views: Petry S and Vale RD. *"A new cap for kinetochore fibre minus ends."* Nature Cell Biology 13:1389–1391(2011).
Research Highlights: *"Cytoskeleton: A stabilizing influence for K-fibres."* Nature Reviews Molecular Cell Biology, 13(1):2-3 (2011).

Vanneste D, Ferreira V and Vernos I.
"Chromokinesins: localization-dependent functions and regulation during cell division."
Biochem Soc Trans, 39(5):1154-60 (2011).

CELL AND DEVELOPMENTAL BIOLOGY

Group: Sensory Cell Biology and Organogenesis

Group Structure:

Group Leader: Hernán López-Schier

Postdoctoral Fellows: Adele Faucherre, Jacobo Cela, Rachele Allena (shared with James Sharpe lab), Sabrina Desbordes

Graduate Students: Indra Wibowo (until March 2011), Filipe Pinto Teixeira (until September 2011), Jesús Pujol Martí, Alessandro Mineo, Andrea Zecca, Oriol Viader

Technician: Andrea Durán (until December 2011)

SUMMARY

Research in my laboratory focuses on understanding the fundamental principles that govern the development of tissues and organs in vertebrates. We attempt to define the cellular and molecular bases of the acquisition and maintenance of tissue architecture and neural circuits, and their relationship to the function of sensory organs.

RESEARCH PROJECTS

1. Cellular Responses to Polarity Signals, in particular Planar Cell Polarity

We use the mechanosensory lateral line of the zebrafish (*Danio rerio*) as a model system to study cell-fate specification, and the formation and remodelling of epithelial architecture during organ development and regeneration. For our studies, we employ cellular, genetic and molecular approaches and state-of-the-art optical imaging techniques, and are also developing methods to analyse *in vivo* the reinnervation of sensory cells during regeneration. In the long term, our studies should provide insight into how sensory organs develop and regenerate, and how their cellular organisation and function are maintained throughout life.

The coordinated orientation of polarised cells within the plane of an epithelium is termed planar cell polarity. The orientation of hair cells within the neuroepithelium of the inner ear represents a striking example of planar cell polarity in vertebrates. Directional deflections of apical mechanosensitive organelles (stereocilia), respectively open or close transduction channels to depolarise or hyperpolarise the hair cell's plasma membrane. The axis of morphological polarity of the stereocilia therefore corresponds to the direction of excitability of the hair cell, and bestows the organ with maximal sensitivity to mechanical stimuli. The senses of hearing and equilibrium thus rely on the exquisite precision with which hair cells are oriented across the sensory epithelium. In spite of its importance, we only have a very superficial knowledge of the mechanisms that control the planar polarisation of hair cells.

Some aquatic vertebrates sense directional water movements with the lateral-line system, a sensory organ closely related to the inner ear. This system comprises a stereotyped array of sensory clusters called neuromasts, each with a very simple organisation. A neuromast contains two types of peripheral supporting cells and a few centrally located hair cells innervated by afferent and efferent axons (Figure 2 left). Hair cells in neuromasts are polarised within the plane of the epithelium in a way comparable to that of the inner ear (Figure 2 right). The lateral-line organ of the zebrafish is thus ideally suited to investigate the mechanisms that control hair-cell planar polarisation.

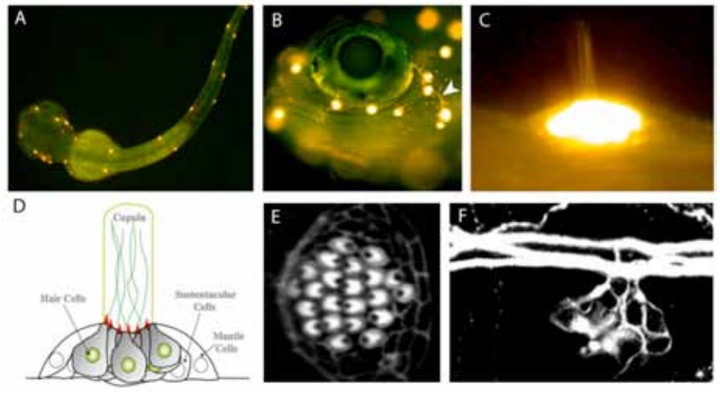


Figure 1
Low magnification view of a living zebrafish larva whose hair cells in the lateral line were labelled with the fluorescent vital dye DiAsp (bright orange). It shows the superficial and systematic distribution of neuromasts along the anterior (head) and posterior (trunk and tail) lateral-line systems. ***B****: Higher magnification of head neuromasts revealing the transynaptic transport of DiAsp from hair cells to afferent nerves (white arrowhead).* ***C****: The highest magnification of a neuromast labeled with DiAsp shows a core of hair cells and their apical-projecting hair bundles.* ***D****: Scheme of a neuromast with its constituent cells. Neurons are not depicted.* ***E****: The orientation (planar polarity) of hair cells is evident in this neuromast whose actin-rich hair bundles were stained with phalloidin. The kinocilium (not stained) appears as a black hole giving the hair bundle its horseshoe appearance.* ***F****: Axonal arbor of two afferent neurons branching off the lateralis nerve below a neuromast.*

We have now identified a series of mutations that disrupt the establishment and maintenance of planar cell polarity in neuromasts. The combination of these strains with several of our multicolour transgenic animals will permit us to generate three- and four-dimensional images of living wild type and mutant specimens with great precision, and to track protein localisation patterns within seconds, or cellular behaviours over days. The combination of the genetic approaches afforded by the zebrafish with live imaging shall allow us to understand sensory-organ development, regeneration, and function in whole animals and at the single-cell level.

2. Sensory Organ Innervation

Historically, planar cell polarity has been studied in invertebrates on tissues that undergo polarisation during a very brief period, to eventually become fixed with negligible or non-existent plasticity, including lack of cellular proliferation, tissue remodelling or cell migration. Such tissues, consequently, will not undergo repair or regeneration after cell death or mechanical damage. Genetic and molecular studies in *Drosophila* have shown that the establishment of planar polarity relies on the concerted activity of many proteins. The cellular responses to polarity cues, especially in remodelling tissues, are not understood. We are defining the cellular and molecular bases underlying the acquisition and maintenance of planar cell polarity, epithelial architecture and innervation in a vertebrate, and its relationship to organ function.

3. Sensory-Organ Growth and Regeneration, with an Emphasis on Epithelial Remodelling

Sensory perception is a complex process that allows organisms to sample the environment and to react appropriately. Sensory dysfunction can thus be a major handicap that dramatically decreases the quality of life of the affected individual. All sensory modalities are liable to deteriorate during one's lifetime. Hearing deficits, for example, afflict more than 10% of the population in industrialized countries, including 0.1% of newborn children and 50% of those aged 80 years or over. Some sensory organs have an impressive capacity to recover after environmental insult, while others can lose function permanently. The inner ear is among the later: hearing loss owing to the degeneration or denervation of the mechanosensory hair cells is irreversible.

Although the search for a hair-cell progenitor resident in sensory epithelia has been pursued for over twenty years, to date there are no reports demonstrating the identification, or even the existence of such cell type. Our recent work has identified a hair-cell progenitor in neuromasts, which allows us the analysis of hair-cell development from its very outset. It also suggests the existence of a stem-cell population, and pinpoints its location within the neuromast. Within the context of this research, we are also attempting to devise methods to follow every cells and complex tissue movements to reconstruct a digital organ *in vivo* (Figure 2).

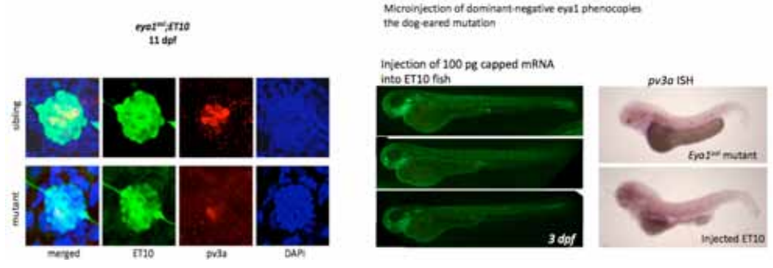


Figure 2
High magnification view of wild type and eya1/dog eared zebrafish larvae whose hair cells in the lateral line were labelled with PV3a (red). Fish are transgenics for ET10, which expresses the GFP in supporting cells (green). Blue is DAPI staining. It shows that hair cells degenerate in eya1/dog eared. ***B****: Low magnification of zebrafish of the ET10 line expressing Eya1^{DN}. It phenocopies the eya1/dog eared mutation.*

4. Tissue Mechanics

We have recently begun to investigate the physical forces that shape and maintain the architecture of epithelia in the zebrafish. We plan to use genetic, microscopic, and biophysical approaches to address this biological problem. We are also developing mathematical tools to analyse the quantitative data generated through the above-mentioned approaches, in the hope to model organ formation *in silico*.

These studies shall provide insight into how organs develop and function throughout life, and also how they regenerate and re-innervate to recover function after damage. This not only represents a very interesting biological problem, but also is relevant to the successful application of therapies aimed to restore sensory function in humans, for aberrant repair would prevent the organ from performing properly.

PUBLICATIONS

Wibowo I, Pinto-Teixeira F, Satou C, Higashijima S, López-Schier H.
"Compartmentalized Notch signaling sustains epithelial mirror symmetry."
Development, 138(6):1143-52 (2011).

Faucherre A, López-Schier H.
"Delaying Gal4-driven gene expression in the zebrafish with morpholinos and Gal80."
PLoS One, 61):e16587 (2011).

Swoger J, Muzzopappa M, López-Schier H, Sharpe J.
"4D retrospective lineage tracing using SPIM for zebrafish organogenesis studies."
J Biophotonics, 4(1-2):122-34 (2011).

Book chapter

Duran A, López-Schier H.
"El peix zebra com a sistema model per a l'estudi de malalties humanes".
In: *Organismes model en biologia*, Montserrat Corominas & Marc Valls, ed., Treballs de la SCB, 62:141-153 (2011).

CELL AND DEVELOPMENTAL BIOLOGY

Group: Cytokinesis and Chromosome Segregation

Group Structure:

Group Leader:	Manuel Mendoza
Postdoctoral Fellows:	Alexandre Vendrell, Petra Stockinger (shared with the Solon lab; since June 2011)
PhD Students:	Gabriel Neurohr, Aina Masgrau, Iris Titos, Nuno Amaral (since September 2011)
Technician:	Trinidad Sanmartin
Visiting Student:	Francesca DiGiovanni (since November 2011)

SUMMARY

Living cells have a fascinating ability to generate complex and dynamic internal structures. Nowhere is this property more evident than during mitosis and cytokinesis: in a very short time (often of the order of a few minutes) cells alter their shape, duplicate and partition their internal components, and divide into two apparently identical halves. To ensure genomic stability, cytokinesis must take place exclusively after the last pair of sister chromatids have been pulled out of the cleavage plane. We are interested in mechanisms that enforce this coordination in the yeast *Saccharomyces cerevisiae*. Regulatory systems identified in yeast are then validated in animal cells (such as *Drosophila*), to ensure that our key findings are relevant for the fidelity of mitosis and genetic stability in multicellular organisms.

RESEARCH PROJECTS

1. Regulation of Chromosome Compaction and Spindle Length

Gabriel Neurohr, Iris Titos

During anaphase, chromosome partitioning requires mitotic chromosomes to be compact enough to allow their segregation; conversely, the spindle must elongate enough to segregate the longest chromosome. But what determines the size of mitotic chromosomes, and the length of the anaphase spindles? To address these questions, we have generated budding yeast cells in which chromosome arm length is progressively increased through chromosome fusions. The resulting compound long chromosomes (LC) do not affect cell viability or growth rate, in spite of their increased length (more than 50% longer than the longest wild type chromosome). Instead, cells successfully segregate these LCs by mutually adjusting their level of condensation and the size of the anaphase spindle during anaphase. We are starting to understand the molecular mechanisms responsible for the mutual regulation of chromosome compaction and spindle size.

We have found that adaptive hyper-condensation of long chromosome arms depends on the Aurora-B kinase at the spindle midzone, Ser10 of histone H3, and condensin activity. Perturbation of this regulatory system also results in decompaction of normal chromosomes during anaphase. Thus, the spindle midzone functions as a ruler that adapts the condensation of long chromosome arms to spindle length to promote their faithful segregation during anaphase, regardless of variations in chromosome or spindle length (Neurohr *et al*., 2011).

2. Coordination of Chromosome Condensation with Cell Size in Metazoans

Petra Stockinger

Our data indicate that spindle length and the level of mitotic chromosome condensation are not predetermined, but mutually coordinated through feedback regulatory loops. These regulatory systems could play important roles upon cell size changes during development in metazoans. An ideal system to study whether such coordination exists, and then to assess its relevance, is the division of neural stem cells in the fruit fly *Drosophila melanogaster*. In this well-characterized system, a large neuroblast (NB) stem cell divides asymmetrically to give origin to a small ganglion mother cell (GMC), which then divides symetrically. Intriguingly, chromosomes are much more compact, and spindles are shorter, in small GMCs than in large NBs. We are testing whether the same process that mediates hyper-condensation of the long chromosomes in yeast also regulates the coordination of chromosome and spindle length during asymmetric divisions in this *Drosophila* lineage.

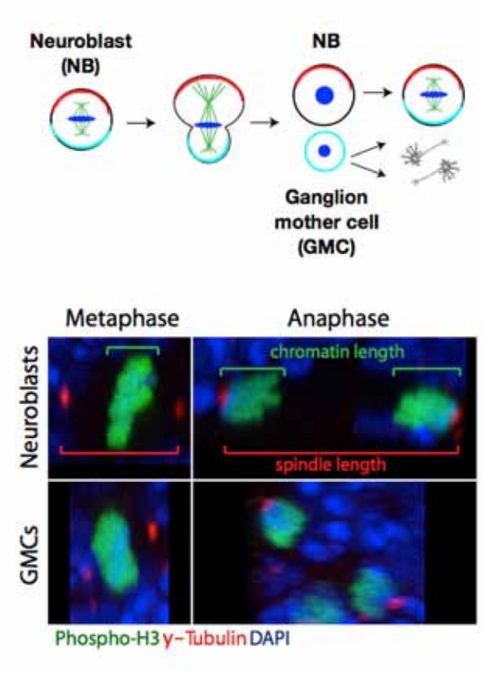


Fig. 1.
Characterization of chromosome compaction and spindle length, in neuroblasts and GMCs in ***Drosophila*** *larvae. Brains were stained with antibodies to visualize mitotic chromosomes (phospho-histone H3), spindle poles (gamma tubulin) and DNA (DAPI). Image by Petra Stockinger.*

3. How Cells Respond to Errors in Chromosome Segregation

Alex Vendrell, Nuno Amaral

Budding yeast cells in which chromosome segregation is impaired through a variety of mutations delay the last step of cytokinesis (abscission). Cells in which the Aurora kinase Ipl1 is inactivated or which lack the cortical proteins Boi1 and Boi2, are unable to inhibit cytokinesis in the presence of chromosome segregation defects. This indicates that an Ipl1- and Boi1/2-dependent checkpoint, which we call NoCut, inhibits abscission upon chromosome segregation defects.

To better understand how errors in chromosome segregation affect cytokinesis, we are characterizing cell cycle progression in a condensin mutant that is unable to complete abscission. By live cell microscopy, the earliest defect of these cells is a delay in the segregation of chromosome arms away from the cleavage plane, and inhibition of cytokinesis. In parallel, to directly study the effects of lagging chromosomes during cytokinesis in the absence of condensation defects, we are characterizing the effect of dicentric chromosomes in cell cycle progression. The cellular response to chromatin bridges generating through these different methods will prove instrumental in dissecting the molecular pathways coordinating chromosome segregation and cell division.

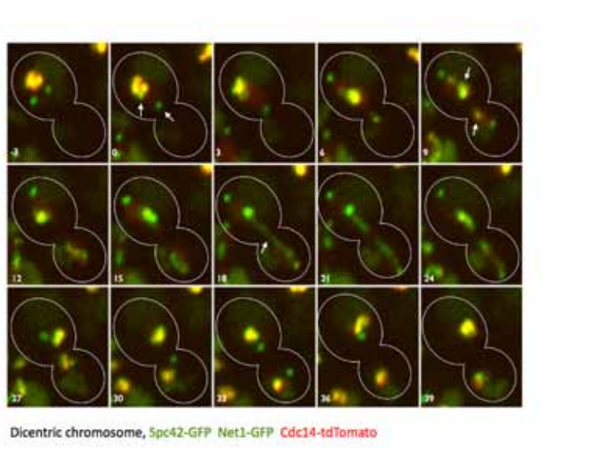


Fig. 2.
Segregation of a dicentric chromosome in a budding yeast cell, imaged thanks to Net1-GFP labeling of the rDNA locus and Spc42-GFP labeling of spindle poles. Cell cycle progression is visualized thanks to Cdc14-tomato, a protein that is released from the nucleolus during anaphase. Image by Nuno Amaral.

4. The Role of Boi1 and Boi2 during Abscission and Polarized Growth

Aina Masgrau

The Boi1 protein and its functionally redundant homolog Boi2p are implicated in both the NoCut checkpoint and in actin cytoskeleton reorganization during polarized growth. We have found that cells lacking both Boi1 and Boi2 are viable, show mild cell polarity defects, and are unable to inhibit abscission (the last step of cytokinesis) upon chromosome segregation defects. In other genetic backgrounds however, *boi1Δ boi2Δ* mutants are unviable and arrest as large depolarized cells. *BOI1* and *BOI2* interact physically or genetically with the Rho like GTPases Cdc42, Rho3 and Rho4, but their mechanism of action in either polarized growth or cytokinesis is not known. To understand the function of Boi proteins in cell polarity and cytokinesis, we are characterizing the phenotype of conditional *boi1 boi2* double mutants in a defined genetic background. In addition, we are investigating the genetic basis of viability in surviving *boi* mutant strains by whole genome sequencing of surviving *boi1 boi2* mutants, in collaboration with the lab of Toni Gabaldón.

PUBLICATIONS

Neurohr G, Naegeli A, Titos I, Theler D, Greber B, Díez J, Gabaldón T, Mendoza M and Barral Y.
"A midzone-based ruler adjusts chromosome compaction to anaphase spindle length."
Science, 332(6028), 465–468 (2011).

CELL AND DEVELOPMENTAL BIOLOGY

Group: Biomechanics of Morphogenesis

Group Structure:

Group Leader: Jérôme Solon

Postdoctoral Fellow: Laure Saias, Kai Dierkes, Petra Stockinger (shared with the Mendoza lab; since June 2011)

Graduate Students: Natalia Czerniak, David Peran Hayes, Anghugali Sumi

Technician: Arturo D'Angelo

SUMMARY

A developing organism undergoes dramatic tissue reshaping and rearrangements. These tissue movements require the precise coordination in space and time of hundreds of cells. This coordination is achieved by the strong interplay between expression of regulatory genes and mechanical forces exerted by the cells. Our goal is to reveal the mechanisms driving tissue rearrangements during morphogenesis.

We are using *Drosophila* late embryogenesis as a model system. At that stage of development, we can observe the coordinated formation of several organs, such as the central nervous system, trachea, heart, gut and muscles, together with the completion of late morphogenetic processes, dorsal closure and head involution. Our group aims to understand the mechanisms underlying these late morphogenetic processes and organs formation during embryogenesis.

RESEARCH PROJECTS

1. Mechanisms Driving Tissue Constriction and Fusion during *Drosophila* Development at the Cellular Scale

We are focusing on tissue constriction, a major morphogenetic process occurring several times during the development of an organism. This consists in the apical constriction of an acto-myosin meshwork. During Drosophila embryogenesis, tissue constriction leads to the fusion and sealing of the embryo's epidermis in a process called dorsal closure (DC). Intriguingly, dorsal closure presents many similarities with wound healing processes in humans. My group is interested in revealing the mechanisms driving DC.

DC consists in the closing of a gap in the epidermis on the dorsal side of the embryo (Fig 1 A). The process takes place after the germ band retraction. It starts with the combined contraction of a monolayer of cells covering the gap, the amnioserosa tissue, with the reinforcement of an actin cable surrounding the contracting tissue. Eventually, at the end of DC, once the two epidermal layers are close enough, they will fuse with a zippering occurring at the two canthis of the opening. The interplay between these three forces, amnioserosa constriction, actin cable reinforcement and filopodia zippering, and their regulation are still poorly understood.

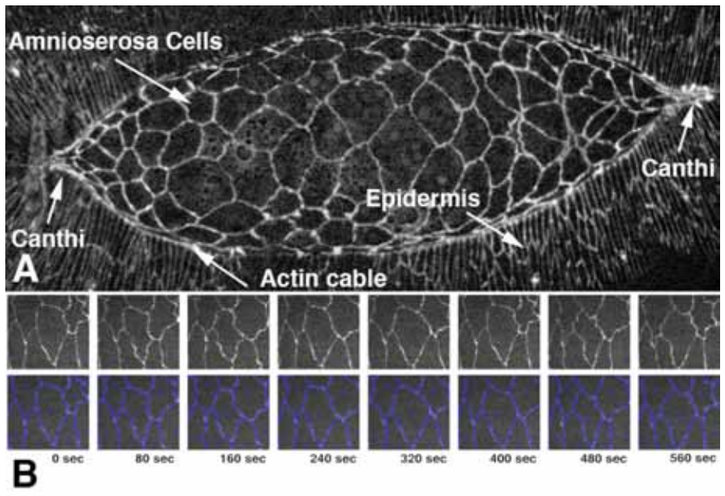


Figure 1 ***(A) Armadillo-GFP highlighting cell membranes during Drosophila dorsal closure***. *The epidermal tissues converge from the lateral part (top and bottom of the picture) to the dorsal part (in the center of the picture) of the embryo. The dorsal part is covered with amnioserosa cells, which provide the initial force for tissue movement.* ***(B) Typical apical surface area pulsations of an AS cell in a GFP-Arm expressing embryo***. *The upper panel shows raw data, the lower panel shows the superimposed segmented image (extracted from Solon et al, 2009).*

Recently, we found that the progression of the epidermis toward the dorsal part of the embryo is due to complex pulsed contractions of the amnioserosa cells coupled with the reinforcement of the actin cable, stabilizing the whole structure in a ratchet-like manner (Figure 1 B).

We want to investigate the molecular regulation of the amnioserosa contractions, what are the mechanisms triggering and regulating the amnioserosa cells, and what molecular mechanisms translate these contractions into tissue movement?

2. Three-dimensional Tissue Remodeling during *Drosophila* Dorsal Closure

We are also interested in revealing the mechanisms underlying tissue remodelling during tissue constriction. Single Plane Imaging Microscopy (SPIM, in collaboration with James Sharpe group) allow us to image a slice in the dorsal-ventral axis of the embryo and to get a high-resolution image of the apical and basal activity of the cells (Figure 2-A). We are combining this technique with regular confocal imaging to extract *in vivo* the 3D dynamics of the amnioserosa contractions (Figure 2-B).

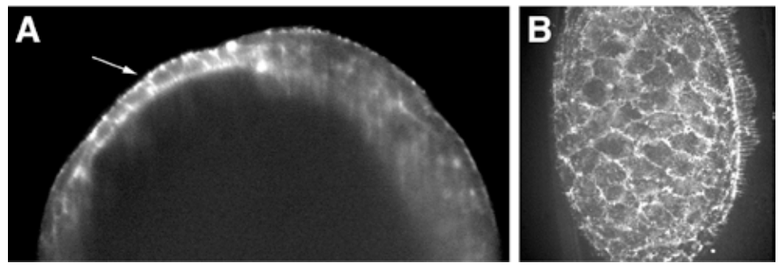


Figure 2. ***Three-dimensional imaging of DC. (A)*** *Live SPIM imaging of a* ***Drosophila*** *embryo in the dorsal ventral axis. The arrow indicates the apical surface the amnioserosa tissue covering the dorsal part of the embryo.* ***(B)*** *Live imaging of the apical surface of the amnioserosa tissue covering the dorsal part of the embryo during DC with a confocal microscope. Images in (A) and (B) are two orthogonal orientations.*

We have found that amnioserosa tissue is able to modulate its volume in order to proceed an efficient closure. During late DC, individual cells undergo a pressure due to the constriction inducing a decrease in volume of each cell. Additionally, the amnioserosa cells are also delaminating during late DC, reducing the number of cells within the tissue. We want to understand how and why cells are adjusting their volume and what trigger the delamination of some of them. We are also investigating how cells could eventually elongate in the apico basal axis and what are the molecular mechanisms underlying this elongation.

3. Mechanisms Driving Head Involution and its Coordination with Dorsal Closure

Simultaneously to DC, the process of head involution (HI) is occurring. Both are highly coordinated in time and are known to share similar genetic regulation. If DC has been intensively studied in the recent years, very few is known about HI. The process of HI consists in the internalization of the central nervous system. The head tissues are first progressing towards in the anterior posterior direction in an inward movement. During this process, massive apoptosis is occurring. The regulation of this apoptosis is necessary for the completion of HI. The inward movement is followed afterwards by the outward progression of the epidermis towards the anterior pole of the embryo to finish covering all the head. We are interested in understanding the mechanisms at the origin of the tissue remodelling occurring during HI. What are the forces generated and how are they regulated in space and time?

We are also interested in revealing the connections between HI and DC at the origin of their spatio-temporal coordination.

We are addressing all these questions by combining *Drosophila* genetics and state-of-the-art imaging techniques available at the ALMU (spinning disk, confocal and two-photon microscopy) with automated image analysis and physical modeling.

PUBLICATIONS

Sironi L, Solon J, Conrad C, Mayer TU, Brunner D, Ellenberg J.
"Automatic quantification of microtubule dynamics enables RNAi-screening of new mitotic spindle regulators."
Cytoskeleton (Hoboken), 68(5):266-78 (2011).

Azevedo D, Antunes M, Prag S, Ma X, Hacker U, Brodland GW, Hutson MS, Solon J, Jacinto A.
"DRhoGEF2 regulates cellular tension and cell pulsations in the Amnioserosa during Drosophila dorsal closure."
PLoS One, 6(9):e23964 (2011).

CELL AND DEVELOPMENTAL BIOLOGY

Group: Organelle Biogenesis and Homeostasis

Group Structure:

Group Leader: Pedro Carvalho

Postdoctoral fellow: Ombretta Foresti (since May 2011)

PhD Students: Alexandra Grippa, Annamaria Ruggiano

Technician: Josep Pareja

SUMMARY

Our lab is interested in the mechanisms of protein and lipid homeostasis at the endoplasmic reticulum (ER).

RESEARCH PROJECTS

1. Mechanisms to Detect and Eliminate Misfolded ER Proteins

a- **ER- associated protein degradation (ERAD)**

The first defense line to eliminate misfolded proteins from the ER is mediated by a pathway called ER-associated protein degradation or ERAD. This pathway consists of different steps that initiates with the recognition of a protein as being misfolded. Substrates are subsequently moved across the ER membrane back into the cytoplasm, a process known as retrotranslocation. On the cytosolic side of the ER, substrates are ubiquitinated by specific, membrane-bound ubiquitin-ligases. An ATPase complex facilitates the release of ubiquitinated substrates from the ER membrane into the cytosol where they are eventually degraded by the proteasome.

A few years ago, as a first step towards a mechanistic understanding of the pathway, we used S. *cerevisiae* to systematically map the interactions between ERAD components. This work showed that Doa10p and Hrd1p, the yeast ubiquitin ligases previously implicated in ERAD by genetic studies, are core components of two distinct protein complexes required for the degradation of different classes of substrates: ER proteins with a misfolded domain in the cytosol (ERAD-C substrates) are targeted to Doa10 complex while proteins with a misfolded luminal domain (ERAD-L substrates) are targeted to Hrd1 complex. A subset of the components of the Hrd1 complex is required for the degradation of proteins with misfolded intramembrane domains (ERAD-M substrates). Interestingly, each one of the ubiquitin-ligase complexes contained components involved in substrate recognition, ubiquitination and membrane extraction, a solution that might facilitate the coordination of the complex series of events occurring on opposite sides of the ER membrane. This finding offered a novel framework for the organization and cooperation among ERAD components that appears to be conserved among eukaryotes.

How are ERAD substrates moved across the ER membrane into the cytosol where they will eventually be degraded by the proteasome? The translocation of newly synthesized proteins into the ER is very well characterized and is known to occur through a protein conducting channel, the translocon. Therefore, it has been postulated that the retrotranslocation step in ERAD is also mediated by a protein conducting channel however the identity of this channel is not known. To start addressing this issue we developed a protein crosslinking strategy to identify the binding partners of an ERAD-L substrate (luminal misfolded protein) that is undergoing retrotranslocation. This work identified the ubiquitin ligase Hrd1p as the main membrane component in the retrotranslocation of this model substrate. Future studies are now required to understand the mechanism by which Hrd1p and probably other membrane bound ubiquitin ligases involved in ERAD, like Doa10p, promote substrate retrotranslocation.

Most of ERAD components have been identified and a general idea for the organization of the pathway is now available, therefore the current challenge is to reveal the mechanisms of ERAD. Towards this goal, we are developing biochemical tools to recapitulate individual steps of the ERAD pathway *in vitro*. These will allow us to address mechanistically critical issues, like for example how substrates are recognized by ERAD components or dissecting the driving force for ERAD substrate retrotranslocation.

b- **ER- phagy**

Certain proteins fail to engage in ERAD, like for example misfolded proteins that have the propensity to aggregate. These are predominantly targeted by ER-phagy, a form of selective autophagy in which certain protein aggregates in the ER are engulfed in autophagosomes and delivered to the lysosome for degradation. It is known that ER-phagy is stimulated when the levels of misfolded proteins in the ER are abnormally high and the folding capacity of this organelle is exceeded, a condition known as ER stress and that is common in many diseases. Other aspects of ER-phagy are completely mysterious. For example, how accumulation of misfolded proteins in the ER is communicated to the autophagy machinery in the cytosol? Does the ER-containing autophagosomes form at specific ER subdomains that contain the misfolded proteins or do they form at random ER sites? We are currently taking a genetic approach to identify components of the ER-phagy pathway in S. *cerevisiae*. This will provide a molecular handle for subsequent mechanistic studies of ER-phagy.

c- **Partition of ER proteins during cell division**

Asymmetric partition of cytosolic misfolded proteins during cell division is a well documented phenomenon both in eukaryotes and prokaryotes. Interestingly, this process is invariably linked to the generation and maintenance of long-lived cellular lineages that need to be damage-free, like for example stem cells. There are some recent indications that misfolded ER proteins might also partition asymmetrically during cell division, however the biological significance of this observation is not yet known. Some of the questions we are interested in understanding are, for example: do certain physiological conditions (like stress) affect the partition of ER proteins? Is asymmetric partition of ER proteins important to establish different cell fates? Do all misfolded ER proteins segregate asymmetrically or only certain classes (aggregated Vs non aggregated)? We are using a strategy to survey the segregation of proteins during mitosis that will allow us to address these and other related issues.

2. Mechanisms of Lipid Droplet Formation and Dynamics

In most cells energy is stored as neutral lipids, mostly triglycerides and sterol esters, in a dedicated cellular compartment, the cytoplasmic lipid droplets (LDs). LDs are found in virtually every eukaryotic cell and play a central role in cellular lipid and energy metabolism. Despite their ubiquitous presence and importance, a large number of questions regarding the physiology of LDs are poorly understood. LDs are composed of a single lipid layer and therefore distinct from all other cellular compartments. How do LDs originate at the endoplasmic reticulum (ER) and what is the machinery involved? How is the size, number and the storage capacity of the LDs regulated? How are specific proteins and lipids targeted to LDs?

To address these questions we are 1) characterizing the ER protein complexes required for LD formation and regulation; 2) developing a cell-free system that recapitulates the biogenesis of LDs *in vitro* that will ultimately reveal the molecular mechanisms of lipid droplet biogenesis.

PUBLICATIONS

Gauss R, Kanehara K, Carvalho P, Ng DT, Aebi M.
"A complex of pdi1p and the mannosidase htm1p initiates clearance of unfolded glycoproteins from the endoplasmic reticulum."
Mol Cell, 42(6): 782-93 (2011).

Stanley AM*, Carvalho P*, Rapoport TA.
"Recognition of an ERAD-L substrate analyzed by site-specific in vivo photocrosslinking."
FEBS Lett, 585(9):1281-6 (2011).
*Equal contribution

SYSTEMS BIOLOGY

Group: Systems Analysis of Development
James Sharpe is an ICREA Research Professor.

Group Structure:

Group Leader: James Sharpe

Staff Scientist: Jim Swoger

Postdoctoral Fellows: Jean-Francois Colas, Henrik Westerberg, James Cotterell, Niamh Nowlan, Rachele Allena, Andreea Munteanu, Marco Musy

Technician: Laura Quintana, Martina Niksic

PhD Students: Bernd Boehm, Luciano Marcon, Gaja Lesnicar-Pucko, Jelena Raspopovic, Juergen Mayer, Alba Jiménez, Manu Uzkudun

SUMMARY

The Sharpe lab has 2 primary goals:

(1) To further our understanding of *organogenesis as a complex system*, by bringing together a diverse range of techniques from biology, physics, imaging and computer science. The larger part of the lab focuses on a well-characterised standard model of development – the vertebrate limb (using both mouse and chick). For this project we constructing a collection of computer simulations (Boehm *et al*, 2010, Marcon *et al*., 2011) which are based on high-quality quantitative empirical data generated by our own new 3D and 4D imaging technologies. A smaller but equally important project within this topic is a more abstract exploration of the patterning potential of gene network motifs.

(2) Building on the success of the 3D imaging technique developed within the lab called Optical Projection Tomography (OPT – *Science* 296:541, 2002), the other major goal of the lab is to continue developing and improving 3D and 4D imaging technology. Recent success in this direction includes the development of time-lapse OPT imaging of mouse limb development *in-vitro* (*Nature Methods* 5:609-12, 2008).

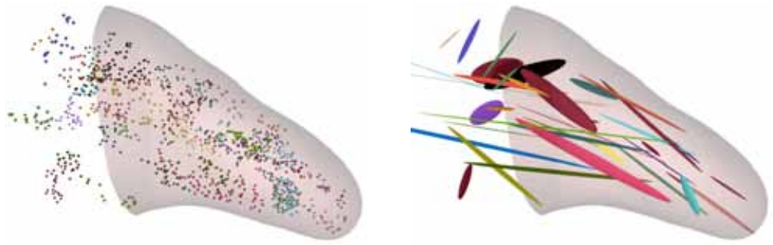
RESEARCH PROJECTS

1. Improvements to OPT and SPIM Imaging

An important focus of the lab is the development of new mesoscopic imaging technology, in particular OPT and SPIM. These technologies are used in a variety of projects including limb bud imaging (a central project for the Sharpe lab), adult mouse pancreas imaging to quantify beta-cell mass in models of diabetes (EU VIBRANT project), and pinpointing the 3D distributions of labelled T-cells within intact mouse lymph nodes. During the last year we have made numerous improvements to the system, including the ability to perform high-resolution images of larger regions of tissue by 3D tiling and reducing the shadow artefact of SPIM by the inclusion of a resonant scan mirror. These modifications have allowed us to achieve unprecedented detail in our analysis of the mouse pancreas and lymph node, as well as enabling cell-resolution analysis of limb bud morphogenesis (described in next section).

2. 3D Cell-Resolution Clonal Analysis of Limb Bud Morphogenesis

In addition to the 2D map, we are pursuing data-capture for the larger project to create a full 3D description of these movements. We are using a double transgenic mouse line which expresses tamoxifen-inducible Cre, and also a flox-stopped EYFP reporter. Both genes are expressed ubiquitously, but by titrating the concentration of injected tamoxifen we can control the frequency of recombination events to the point of obtaining single clones in a limb bud. Although the location of clone induction cannot be controlled, the time of induction can. In this way embryos can be harvested at known time-intervals which contain clones of labelled cells which have grown and expanded for a fixed number of hours. This large collection of clones, which are all scanned in 3D using SPIM (see point 1 above), is compiled into a database which covers all the timepoints important for limb development. The SPIM images display enough spatial resolution to allow the individual cells to be located within each clone. We can therefore perform statistical analysis on these distributions (principle component analysis) to characterise the degree of anisotropic growth represented by each clone. We have also devised a new theoretical approach to incorporate these data into a finite-element model. In this way we are managing to reconstruct *in-silico* the realistic 3D tissue movement maps for this example of mammalian organogenesis.



Clonal analysis of the developing mouse limb bud. Analysis of many genetically-encoded clones, SPIM imaged from many different experiments, which have been mapped into a single reference limb bud (left). The primary orientations of their anisotropic growth has then been calculated and represented as 3D ellipsoids (right).

3. In-Ovo Time-Lapse Multiphoton to Reveal Cellular Activities during Limb Elongation

In addition to tissue-level imaging, we have also performed time-lapse multiphoton microscopy as a means to observe the dynamics of individual cells during limb bud morphogenesis. For this approach we use the chick embryo as a model system, as normal limb development can be maintained during in-ovo imaging, while this is not reliable for mouse *in-vitro* culture systems. A membrane-bound GFP construct is electroporated into the flank mesenhcyme which will give rise to the limb bud. Using this approach we are documenting the orientation of the cells with respect to the major axes of the limb (in different positions in time and space), as well as their dynamic behaviours, such as intercalation and migration.

4. Control of Pattern Formation by Gene Regulatory Networks

The tissue movement maps being created (both in 2D and 3D) will serve as the framework within which to simulate and explore the gene network designs responsible for controlling limb development. Over the last year we have been modelling how self-organising reaction-diffusion networks may be able to control formation of the skeletal pattern during limb bud development. In particular, we are exploring how an accurate pattern may depend on a close integration of a Turing-type patterning mechanism with positional information gradients controlled from the boundaries of the limb bud.

PUBLICATIONS

Cheddad A, Svensson C, Sharpe J, Georgsson F, Ahlgren U.
"Image Processing Assisted Algorithms for Optical Projection Tomography."
IEEE Trans Med Imaging, 31(1):1-15 (2011).

Quintana L and Sharpe J.
"Optical projection tomography of vertebrate embryo development."
Cold Spring Harb Protoc, (6):586-94 (2011).

Quintana L, Sharpe J.
"Preparation of mouse embryos for optical projection tomography imaging."
Cold Spring Harb Protoc, 2011(6):664-9 (2011).

Domínguez-Frutos E, López-Hernández I, Vendrell V, Neves J, Gallozzi M, Gutsche K *et al*.
"N-myc controls proliferation, morphogenesis, and patterning of the inner ear."
J Neurosci, 31(19):7178-89 (2011).

Capellini TD, Handschuh K, Quintana L, Ferretti E, Di Giacomo G, Fantini S, Vaccari G, Clarke SL *et al*.
"Control of pelvic girdle development by genes of the Pbx family and Emx2"
Dev Dyn, 240(5):1173-89 (2011).

Boehm B, Rautschka M, Quintana L, Raspopovic J, Jan Z, Sharpe J.
"A landmark-free morphometric staging system for the mouse limb bud."
Development, 138(6):1227-34 (2011).

Marcon L, Arqués CG, Torres MS, Sharpe J.
"A computational clonal analysis of the developing mouse limb bud."
PLoS Comput Biol, 7(2):e1001071 (2011).

Swoger J, Muzzopappa M, López-Schier H, Sharpe J.
"4D retrospective lineage tracing using SPIM for zebrafish organogenesis studies."
J Biophotonics, 4(1-2):122-34. Epub 2010 Oct 5 (2011).

Herrmann F, Garriga-Canut M, Baumstark R, Fajardo-Sanchez E, Cotterell J, Minoche A, Himmelbauer H, Isalan M.
"p53 Gene repair with zinc finger nucleases optimised by yeast 1-hybrid and validated by Solexa sequencing."
PLoS One, 6(6):e20913 (2011).

Reviews

Sharpe J.
"Two ways to use imaging: focusing directly on mechanism, or indirectly via behaviour?"
Curr Opin Genet Dev, 21(5):523-9 (2011).

Hopyan S, Sharpe J, Yang Y.
"Budding behaviors: Growth of the limb as a model of morphogenesis."
Dev Dyn, 240(5):1054-62 (2011).

Book

Sharpe J and Wong RO.
"Imaging in Developmental Biology: A Laboratory Manual."
Cold Spring Harbor Laboratory Press, (2011).

SYSTEMS BIOLOGY

Group: Design of Biological Systems
Luis Serrano is an ICREA Research Professor.

Group Structure:

Group Leader: Luis Serrano

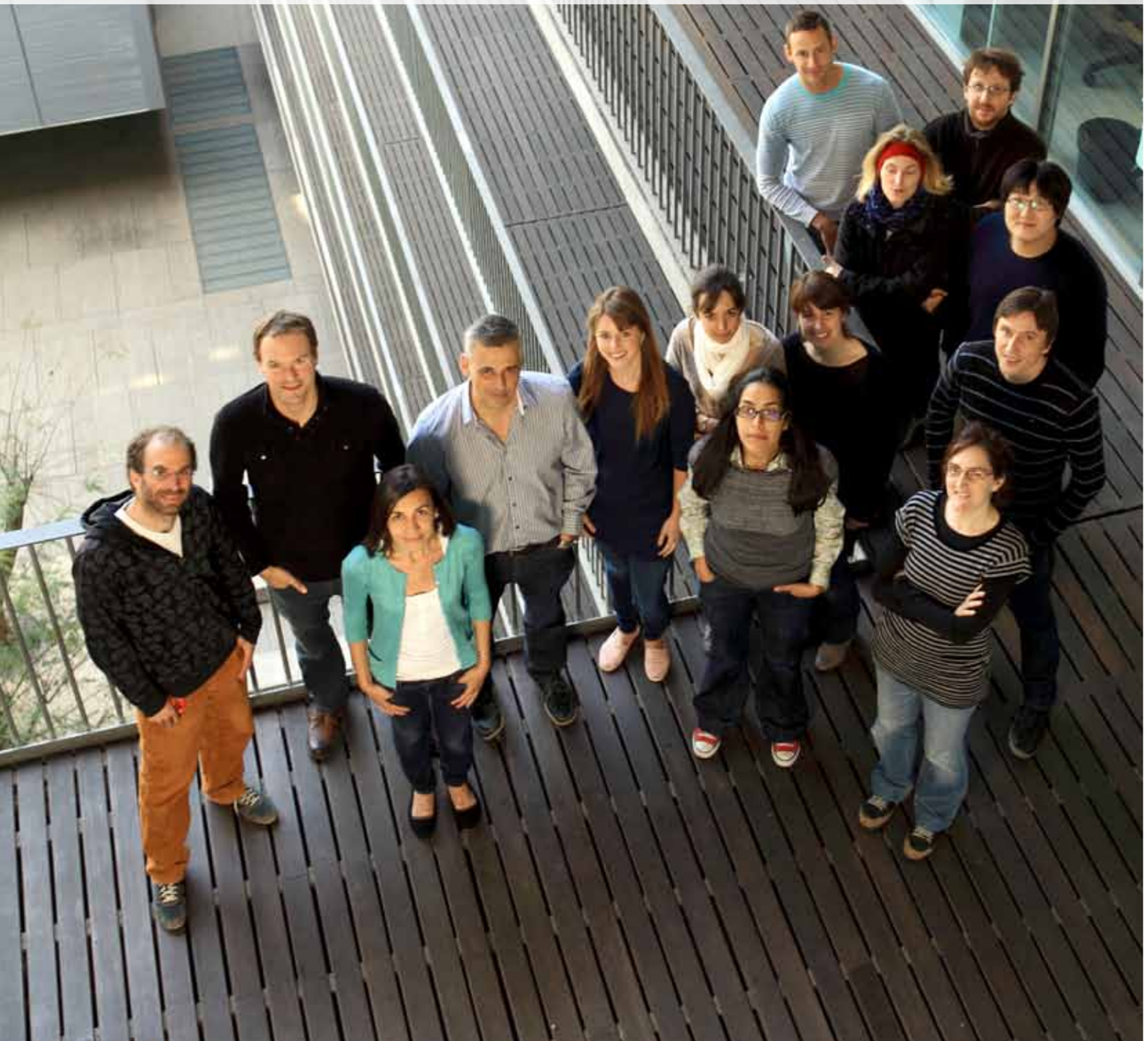
Staff Scientist: Christina Kiel

Postdoctoral Fellows: Tobias Maier, Eva Yus, Almer Van der Sloot, Maria Lluch, Julia Burnier, Javier Delgado, Besray Unal (INTERPOD program, shared with Yogi Jaeger)

Students: Anne Campagna, Judith Wodke, Bernhard Paetzold, Kiana Toufighi (shared with Ben Lehner), Erik Verschueren, Marie Jeanne Trussart

Collaborator: Peter Vanhee

Technician: Sira Martinez, Tony Ferrar, Violeta Beltran



SUMMARY

Our group is interested in the quantitative understanding and rational engineering of living systems (ranging from gene networks to organisms). For this purpose we use a combination of tools that involve software for protein design and simulations of networks and experimental approaches. Our approach is based on first understanding a system and then engineering it to obtain the properties we want. Our philosophy is also whenever possible identifying the possible practical applications for human health and biotechnology of our work.

RESEARCH PROJECT

1. Quantitative Understanding of *M. pneumoniae*

The idea of harnessing living organisms for treating human diseases is not new and has been in fiction books since a long time ago. So far the majority of the living vectors used in human therapy are viruses, which have the disadvantage of the limited number of genes and networks that can contain. Bacteria have the advantage of allowing the cloning of complex networks and the possibility of making a large plethora of compounds either naturally or through careful re-design. One of the main limitations for the use of bacteria to treat human diseases is their complexity, the existence of a cell wall that difficult the communication with the target cells, the lack of control over its growth and the immune response that will elicit on its target. Ideally one would like to have a very small bacterium (of the size of a mitochondria), with no cell wall, which could be grown *in vitro*, could be genetically manipulated, for which we will have enough data to allow a complete understanding of its behavior and which could live as a human cell parasite. Such a microorganism could in principle be used as a living vector in which genes of interests, or networks producing organic molecules of medical relevance, could be introduced under *in vitro* conditions and then inoculated either on extracted human cells or in the organism, and then become a new organelle in the host. Once the living vector enters inside the host cells it could then produce and secrete into the host proteins which will be needed to correct a genetic disease, or drugs needed by the patient. Putting it into engineering terms, the living vector will be alike a processor which will have a complicated set of instructions and circuits but will only communicate with the host through input and output outlets. Thus the processor could be reprogrammed but the interface with the hosting cell will remain the same. For some particular applications it will not be needed to integrate the bacteria as an organelle, but rather have it inside the cell for a limited amount of time to achieve its goal and then eliminate it by antibiotic treatment for example.

In order to achieve the above goals we need to understand in excruciating detail the Biology of the target bacterium as well as how to interface with the host cell cycle (**Systems Biology aspect**). Then we need to have the engineering tools (network design, protein design, simulations …) in order to modify the target bacterium to behave like an organelle once inside the cell (**Synthetic Biology aspect**). Thus this project has two objectives:

a) Obtain a complete quantitative understanding of a free-living organism (a bacterium in this case).

b) Engineer the bacterium to enter into a mammalian cell line, adapt to the host so as to keep a fixed number of bacteria per host, respond to the host environment and secrete into the host proteins or organic molecules that will provide missing functionalities.

In 2011 we have quantified at copy level the proteome of *M. Pneumonia* and related it to RNA abundance and protein half life (Figure 1). Our analysis shows as reported in other organisms that the correlation between RNA and protein is poor and therefore that both are needed to understand biological processes (Maier *et al.*, 2011).

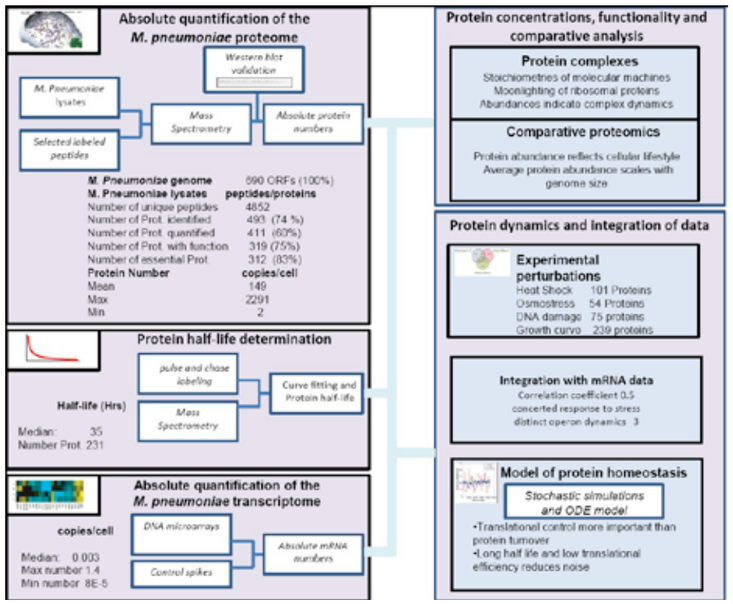


Figure 1.
Workflow of the work done to determine absolute protein and RNA concentrations in M. Pneumonia, as well as protein half lifes, and the integration of this data with computer simulations of protein homeostasis (Maier et al.,2011).

2. Signal Transduction and Disease

Understanding signal transduction pathways is capital for human health. Different cell types share many of their signaling molecules yet can respond specifically to the same stimuli, through ways not fully understood. With increasing information available from large-scale '–omics' experiments in recent years, the representation of signaling systems has changed from the traditional depiction of linear pathways to complex network maps ("everything does everything to everything", Dumont *et al*., 2002; *Am. J. Physiol. Cell. Physiol*. 283, C2-C28). Therefore, it is difficult to elucidate when knocking out a protein or blocking an activity with a drug what is the relationship between phenotype and the interaction affected. On the other hand cells respond specifically to a large number of different external signaling molecules and stimuli although they frequently sharing downstream signaling modules. For example, stimulation of a similar MAPK module/ pathway results in either transient or sustained ERK activation, depending on whether cells were stimulated with epidermal growth factor (EGF) or nerve growth factor (NGF), and the two factors induce distinct cell fates (Marshall, 1995; *Cell* 80, 179-185). Furthermore, even the same growth factor, can elicit different downstream signaling events, depending on the cell type (Kiel and Serrano, 2009; *Sci. Signal*. 2, ra38).

In principle, one simple way to explain the different responses in different cell lines to the same signal could be that some critical proteins in the network are differentially expressed. It is clear that a central 'hub'protein in the network cannot interact with all of its partner proteins simultaneously and that some interactions are mutually exclusive (Kiel *et al*., 2011; *Mol. Syst. Biol*. 7, 551). One could envisage that differences in protein concentrations between cell lines could change cellular output if there is competition at a critical branching point of the network, e.g., if an upstream hub is expressed at a low concentration, then the concentrations (and affinities) of competing binding partners could determine the signaling pathway taken.

In our group we want to understand signal transduction in a quantitative way to the point that we can model accurately the response of different cells to drugs or mutations. As a scientific target we have selected signal transduction in vision and the MAPK pathway and our final goal is to obtain a global quantitative understanding with the idea of designing better therapies for diseases involving its deregulation. To understand that pathway in a quantitative predictive way we are building the interaction network at structural level (Figure 1), mapping all known diseases mutations, determining the concentrations of all proteins in different cell lines and applying different perturbations to understand how the network is regulated. The results of these analyses plus data from the literature regarding Kds, kons, koffs is used to model mathematically the network.

In 2011, by integrating three proteomic datasets, literature mining, computational analyses, and structural information, we have generated a multiscale signal transduction network linked to the visual G protein-coupled receptor (GPCR) rhodopsin, the major protein component of rod outer segments (Figure 2) (Kiel *et al* 2011). This reconstruction demonstrates a specific disease susceptibility of the core visual pathway due to the uniqueness of its components present mainly in the eye.

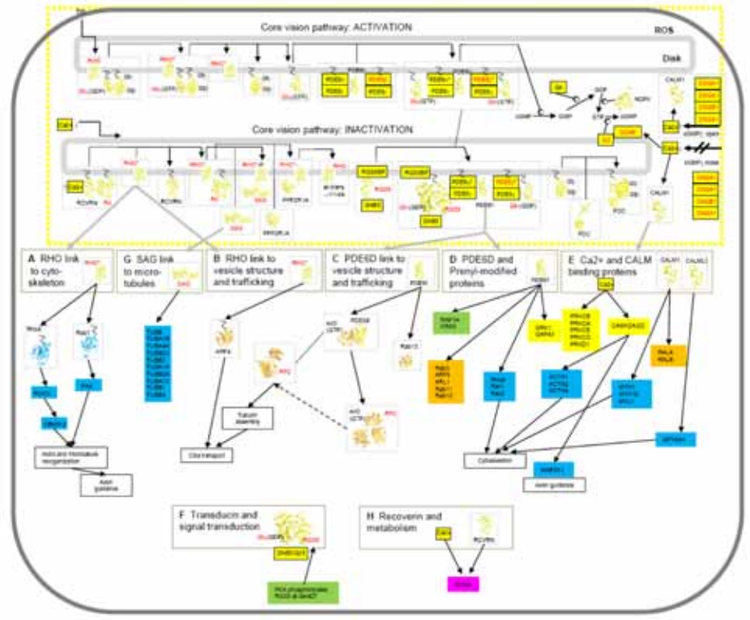


Figure 2.
Structural coverage of the core vision pathway and its links to other functional modules (Kiel ***et al.****, 2011). The published core pathway (Dell'Orco et al, 2009;* ***Mol Biosyst 5****: 1232-1246) was extended using evidence from our high-confidence network. Outputs to different functional cellular processes emanating from the proteins in the pathway are indicated, and structures are displayed if available by ribbon representation.*

PUBLICATIONS

De Masi F, Grove CA, Vedenko A, Alibes A, Gisselbrecht SS, Serrano L, Bulyk ML, Walhout AJ
"Using a structural and logics systems approach to infer bHLH-DNA binding specificity determinants."
Nucleic Acids Res, 39: 4553-4563 (2011).

Gfeller D, Butty F, Wierzbicka M, Verschueren E, Vanhee P, Huang H, Ernst A, Dar N, Stagljar I, Serrano L, Sidhu SS, Bader GD, Kim PM
"The multiple-specificity landscape of modular peptide recognition domains."
Mol Syst Biol, 7: 484 (2011).

Guell M, Yus E, Lluch-Senar M, Serrano L
"Bacterial transcriptomics: what is beyond the RNA horiz-ome?"
Nat Rev Microbiol, 9: 658-669 (2011).

Kiel C, Serrano L
"Challenges ahead in signal transduction: MAPK as an example."
Curr Opin Biotechnol. 2011 Oct 28. [Epub ahead of print]

Kiel C, Vogt A, Campagna A, Chatr-aryamontri A, Swiatek-de Lange M, Beer M, Bolz S, Mack AF, Kinkl N, Cesareni G, Serrano L, Ueffing M
"Structural and functional protein network analyses predict novel signaling functions for rhodopsin."
Mol Syst Biol, 7: 551 (2011).

Maier T, Schmidt A, Guell M, Kuhner S, Gavin AC, Aebersold R, Serrano L
"Quantification of mRNA and protein and integration with protein turnover in a bacterium."
Mol Syst Biol, 7: 511 (2011).

Munoz IG, Prieto J, Subramanian S, Coloma J, Redondo P, Villate M, Merino N, Marenchino M, D'Abramo M, Gervasio FL, Grizot S, Daboussi F, Smith J, Chion-Sotinel I, Paques F, Duchateau P, Alibes A, Stricher F, Serrano L, Blanco FJ, Montoya G
"Molecular basis of engineered meganuclease targeting of the endogenous human RAG1 locus."
Nucleic Acids Res, 39:729-743 (2011).

Nadra AD, Serrano L, Alibes A
"DNA-binding specificity prediction with FoldX."
Methods Enzymol, 498:3-18 (2011).

Rakoczy EP, Kiel C, McKeone R, Stricher F, Serrano L
"Analysis of disease-linked rhodopsin mutations based on structure, function, and protein stability calculations."
J Mol Biol, 405:584-606 (2011).

Sorek R, Serrano L.
"Bacterial genomes: from regulatory complexity to engineering."
Curr Opin Microbiol, 14:577-578 (2011).

Szegezdi E, Reis CR, van der Sloot AM, Natoni A, O'Reilly A, Reeve J, Cool RH, O'Dwyer M, Knapper S, Serrano L, Quax WJ, Samali A.
"Targeting AML through DR4 with a novel variant of rhTRAIL."
J Cell Mol Med, 15: 2216-2231 (2011).

Van Durme J, Delgado J, Stricher F, Serrano L, Schymkowitz J, Rousseau F.
"A graphical interface for the FoldX forcefield."
Bioinformatics, 27:1711-1712 (2011).

Vanhee P, van der Sloot AM, Verschueren E, Serrano L, Rousseau F, Schymkowitz J.
"Computational design of peptide ligands."
Trends Biotechnol, 29:231-239 (2011).

Vanhee P, Verschueren E, Baeten L, Stricher F, Serrano L, Rousseau F, Schymkowitz J.
"BriX: a database of protein building blocks for structural analysis, modeling and design."
Nucleic Acids Res, 39:D435-442 (2011).

Verschueren E, Vanhee P, van der Sloot AM, Serrano L, Rousseau F, Schymkowitz J.
"Protein design with fragment databases."
Curr Opin Struct Biol, 21: 452-459 (2011).

SYSTEMS BIOLOGY

Group: Gene Network Engineering
This group is part of the EMBL/CRG Research Unit in Systems Biology

Group Structure:

Group Leader: Mark Isalan

Postdoctoral Fellows: Mireia Garriga, Phil Sanders, Yolanda Schaerli, Manjunatha Kogenaru

Students: Andreia Carvalho (graduated Mar 2011), Marco Constante (graduated Dec 2011), Diego Barcena (started Sept 2011), Vivek Senthivel (started Oct 2011), Jose Aguilar-Rodriguez (visiting)

Technicians: Rebecca Baumstark, Carmen Agustin Pavon

SUMMARY

The primary aim of the group is to study the interactions between biological components with a view to having a predictive understanding of biological networks. Ultimately, we would like to be able to build functional gene networks that actually behave in the way we would anticipate. To achieve this, the group has been structured to combine: (1) studying large-scale networks from the top-down (bacterial shuffle networks), (2) reconstructing small networks from the bottom-up (synthetic patterning networks), and (3) protein engineering of zinc fingers for gene repair and network engineering.

RESEARCH PROJECTS

1. Shuffling *E. coli* Transcription

The major work of our lab is based around a study on rewiring large-scale bacterial gene networks. We recently shuffled the transcription network of *E. coli* to ask the question, "what happens if you add lots of new links on top of an existing biological network?"

Genomes evolve by duplicating genes, which then acquire new controlling connections. We speeded-up this process by taking some of the most highly-connected transcription factor (TF) coding sequences and rewiring them to different promoter control regions (by adding promoter-ORF fusions on plasmids). In this way, the multiple inputs to a given promoter regulatory region were connected to the multiple targets of a TF (outputs). At least two outcomes were possible (see Figure 1): (i) Pathways that never communicated before could be linked together in a new synthetic crosstalk. (ii) Regulatory sites downstream of a TF cascade could be connected back to the TF, creating multiple positive and/or negative feedback loops. Thus, very complicated new connections, as well as other motifs like feedforward loops, could potentially be created by single promoter-ORF fusions.

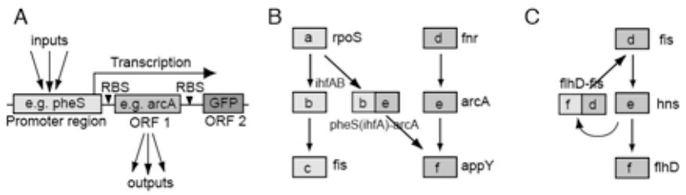


Figure 1. *Rewiring the network by adding shuffled promoter-ORF combinations. (A) Plasmid constructs containing a promoter regulatory region, linked to a different ORF (transcription regulator or sigma factor), create new network links. A GFP ORF with an independent ribosome binding site (RBS) acts as a reporter. (B) Parallel pathways ("a-b-c" and "d-e-f") can be linked together with a rewiring construct, b-e. The inputs to the promoter "b" now output through the ORF "e". Examples of actual gene regions used in this study are written next to the gene boxes. (C) Promoter regions downstream of their new partner ORF can create new feedback loops. The linear "d-e-f" pathway is converted to a feedback loop, by adding the "f-d" construct, such that the inputs into "f" now output through "d".*

For the rewiring experiments, we chose transcription network genes with various levels of connectivity within the tree-like hierarchy of the *E. coli* transcription network; some were connected to 10s, 100s or even 1000s of other genes. The surprising result was that even highly-connected genes could be "rewired" and were generally very well-tolerated by the bacteria. Some new networks even conferred new specific fitness advantages on the bacteria, such as improved heat resistance. We built ~600 rewired networks and showed that very complex new feedback loops and rewirings were viable and could act as a substrate for evolution. Therefore the existing network is highly evolvable through such gene shuffling, even at network hubs.

Reference: Isalan *et al*. *"Evolvability and hierarchy in rewired bacterial gene networks"*. Nature, 452(7189):840-5 (2008).

The propagation of perturbations in shuffled networks

The Shuffle project has been extended in the past year by choosing ~100 network constructs for Affymetrix microarrays, under highly-standardised conditions, in biological triplicates. We already have very promising results for ~60 of the chosen networks, which is why we have extended the project to 100 networks. The microarrays clearly show that the different networks make systematic perturbations from 10s to 1000s of genes (the most perturbed clone has 2795/4070 genes perturbed - ~70% of the transcriptome - and is still viable). By clustering the genes and networks into patterns of similar gene expression, we can visualise the extent to which perturbations propagate across the network and can find some common 'states' which the bacteria tend towards. We believe these states to represent 'attractors' in the system, which we want to understand. We are working towards converting the array data into a predictive model that will describe which network perturbations will induce which gene expression states.

2. Patterning Gene Networks

We are carrying out several spatial and temporal pattern-engineering projects in parallel, corresponding to different scales of cellular organisation. These start from engineering localised gradients and patterns in single cells and move towards designing or selecting gene networks using diffusible factors that operate over fields of cells. The various projects use both bacterial and mammalian cells.

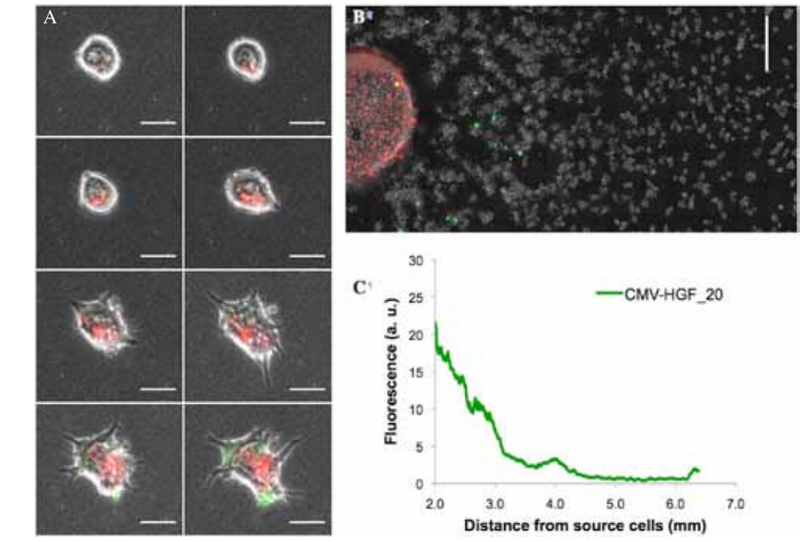


Figure 2. ***Synthetic 3D 'morphogen' patterns in MDCK cysts. A**. Localized HGF transfection (red), followed by HGF secretion, distal induction of GFP expression and tubule formation, in an MDCK cyst containing a reporter gene network. **B**. Spatial expression of GFP over fields of cysts, in response to a localized source of HGF (secreting cells; expressing a red fluorescent protein). 24hrs after seeding the HGF-secreting cells, the cysts surrounding the sender region express GFP and undergo tubule formation. Cysts further way from the sender cells do not express GFP and do not tubulate or scatter. **C**. Quantification of the expression of GFP using imageJ (Radial profile plugin) which measures average GFP intensity radially, from a source. Bars, 1 mm.*

3. Synthesising Zinc Fingers for Genome Engineering and Gene Repair

As part of an EU-funded ERC project (FP7 ERC Zinc-Hubs) we are building a number of artificial sequence-specific DNA-binding proteins using our established protocol Isalan M, Klug A and Choo Y. *Nature Biotechnology*, 19, 656-60, 2001). We aim to develop upon the recent reports of endogenous gene repair using zinc finger nucleases (Bibikova *et al*. *Science*, 300, 764, 2003; Urnov et al., *Nature,* 435, 646-51, 2005). Ultimately, we wish to probe the effects of mutating or rewiring hub genes, within the context of the entire mammalian transcription network.

PUBLICATIONS

Isalan M.
"Zinc-finger nucleases: how to play two good hands."
Nat Methods, 9(1):32-4. (2011).

Constante M, Grünberg R, Isalan M.
"A biobrick library for cloning custom eukaryotic plasmids."
PLoS One, 6(8):e23685 (2011).

Herrmann F, Garriga-Canut M, Baumstark R, Fajardo-Sanchez E, Cotterell J, Minoche A, Himmelbauer H, Isalan M.
"p53 Gene repair with zinc finger nucleases optimised by yeast 1-hybrid and validated by Solexa sequencing."
PLoS One, 6(6):e20913 (2011).

SYSTEMS BIOLOGY

Group: Genetic Systems
This group is part of the EMBL/CRG Research Unit in Systems Biology.
Ben Lehner is an ICREA Research Professor.

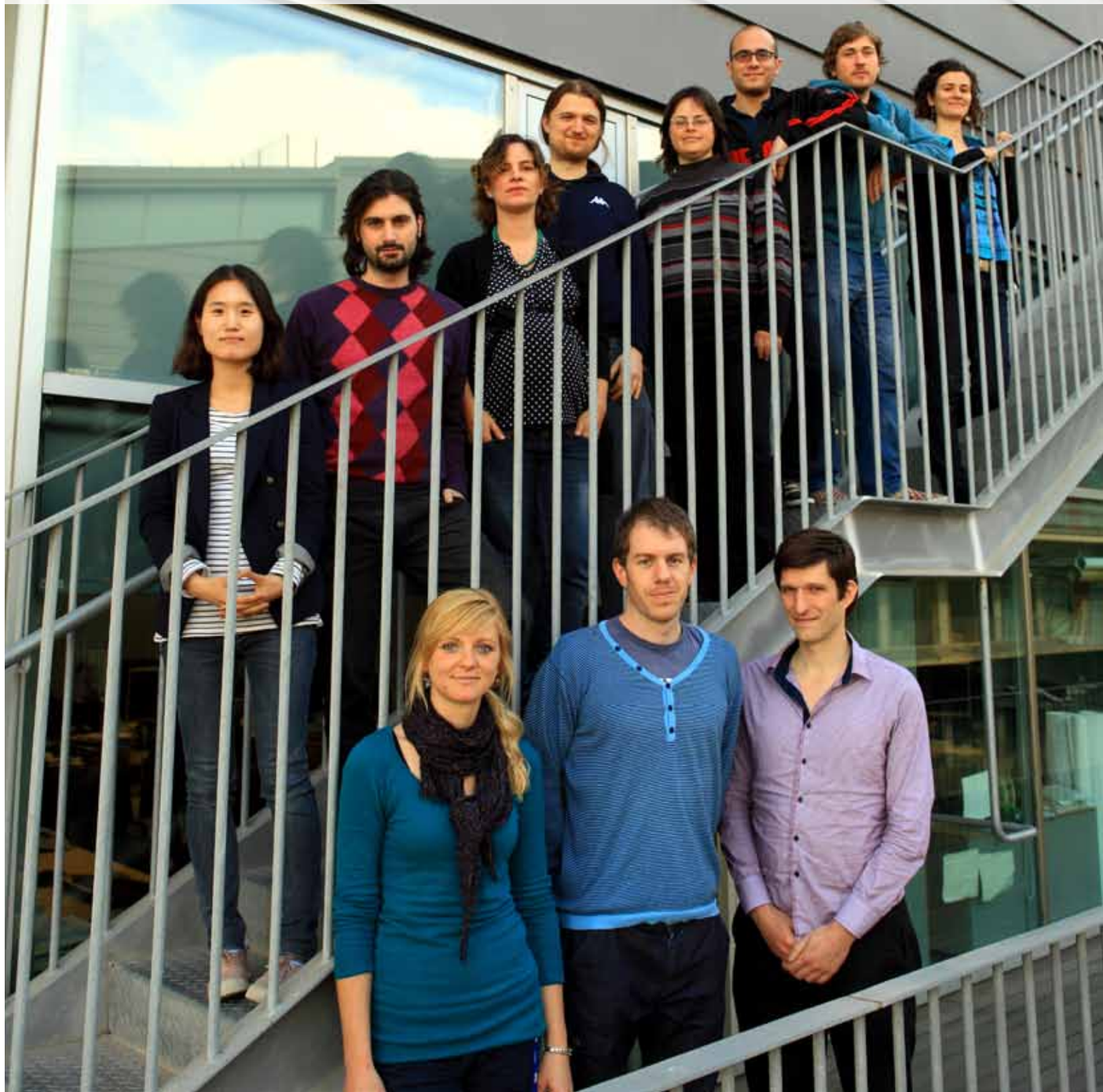
Group Structure:

Group Leader: Ben Lehner

Postdoctoral Researchers: Jennifer Semple, Rob Jelier, Olivia Casanueva, Benjamin Schuster-Böckler, Mirko Francesconi

PhD Students: Alejandro Burga Ramos, Angela Krüger, Adam Klosin

Technicians: Laura Biondini, Kadri Reis



SUMMARY

Most mutations, for example disease causing mutations in humans, are not harmful in all of the individuals who carry them. When do genetic changes result in phenotypic change? When do they not? Why is this? And how can this be predicted? These are the main questions that drive our research, and we use both experimental and computational approaches to address them. Most of our work is hypothesis driven, but we have also used systematic data collection and integration. Our favourite model organisms are *C. elegans* and budding yeast because in these systems we can perform both large-scale and highly quantitative genetic and phenotypic analyses. A central goal is to understand and predict when genetic variation results in phenotypic variation, both at the level of the typical outcome in a population and also in each particular individual. Moreover, to use this information to better understand both sequence and phenotypic evolution.

Work that we have published in the past year has addressed a number of basic questions in genetics, including:

(i) how well can we predict phenotypic variation from the complete genome sequences of individuals?
(ii) why do mutations have different outcomes in different individuals, even if they are genetically identical and share a common environment?
(iii) can we predict when mutations in regulatory regions are detrimental?
(iv) what is the functional consequence of base composition variation in the human genome?

RESEARCH PROJECTS

1. Whole Genome Reverse Genetics: Predicting Phenotypic Variation from the Complete Genome Sequences of Individuals

A central challenge in genetics is to predict phenotypic variation from individual genome sequences. We attempted this for the first time, constructing and evaluating phenotypic predictions for 19 strains of *Saccharomyces cerevisiae*. We used conservation-based methods to predict the impact of protein-coding variation within genes on protein function. We then ranked strains using a prediction score that measures the total sum of function-altering changes in different sets of genes reported to influence over 100 phenotypes in genome-wide loss-of-function screens. We evaluated our predictions using large-scale quantitative experiments. When the set of genes linked to a trait was evaluated as reliable using a gene network, we were able to make reasonable predictions about the phenotypes of individuals from their genome sequences. This approach – making and evaluating whole genome reverse genetic predictions in model organisms – provides a general strategy to evaluate and improve computational methods for personalised genomics, and shows the plausibility of making accurate phenotypic predictions for diverse phenotypes from individual genomes.

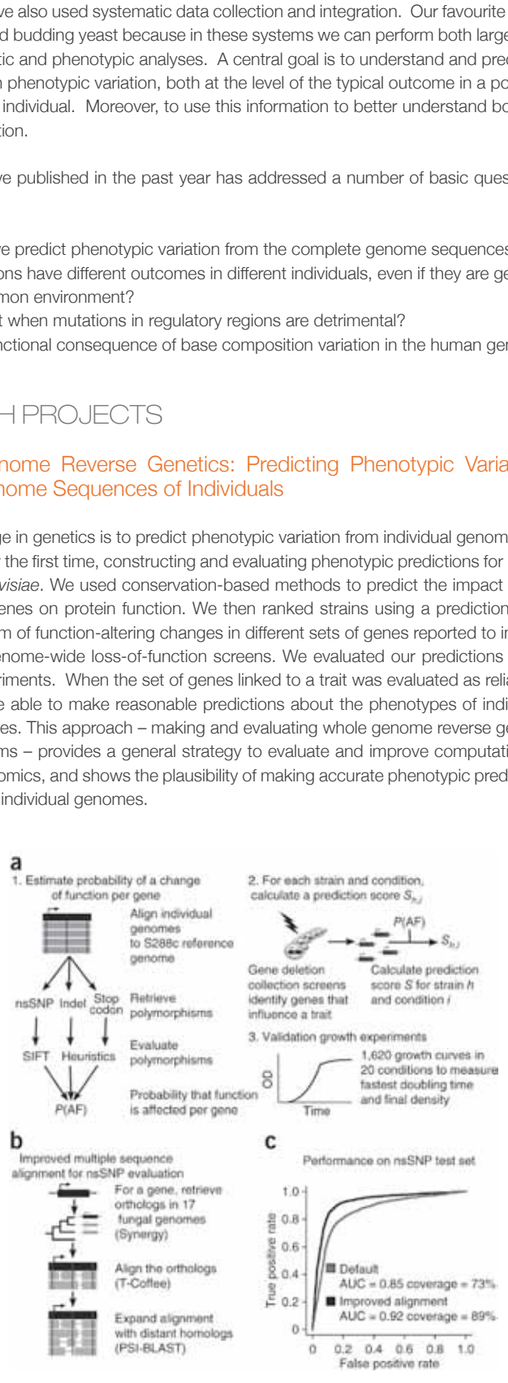


Figure 1.
Making and evaluating whole genome reverse genetic predictions in budding yeast.

2. Inter-Individual Variation in the Outcome of Mutations: Early Stochastic Variation in Gene Expression Predicts the Outcome of Mutations in Individuals

Many mutations, including those that cause disease, only have a detrimental effect in a subset of individuals. The reasons for this are usually unknown, but may include additional genetic variation and environmental risk factors. However, phenotypic discordance remains even in the absence of genetic variation, for example between monozygotic twins, and incomplete penetrance of mutations is frequent in isogenic model organisms in homogeneous environments. We proposed a model for incomplete penetrance based on genetic interaction networks. Using *Caenorhabditis* elegans as a model system, we identify two compensation mechanisms that vary among individuals and influence mutation outcome. First, feedback induction of an ancestral gene duplicate differs across individuals, with high expression masking the effects of a mutation. This supports the hypothesis that redundancy is maintained in genomes to buffer stochastic developmental failure. Second, during normal embryonic development we find that there is substantial variation in the induction of molecular chaperones such as Hsp90 (DAF-21). Chaperones act as promiscuous buffers of genetic variation, and embryos with stronger induction of Hsp90 are less likely to be affected by an inherited mutation. Simultaneously quantifying the variation in these two independent responses allows the phenotypic outcome of a mutation to be more accurately predicted in individuals. Our model and methodology provide a framework for dissecting the causes of incomplete penetrance. Further, the results establish that inter-individual variation in both specific and more general buffering systems combine to determine the outcome inherited mutations in each individual.

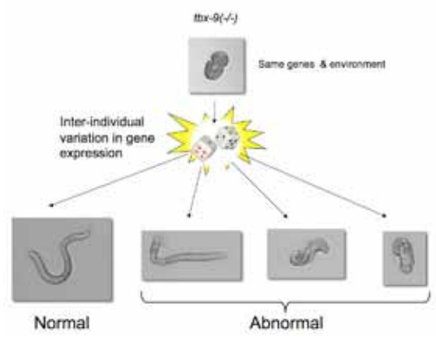


Figure 2.
Variation in gene expression during development results in some individuals being affected by an inherited mutation whilst others survive.

3. Fitness Trade-Offs, Variable Signalling and Environmentally Induced Mutation Buffering in Isogenic Organisms

The consequences of many inherited mutations are enhanced when chaperone activity is reduced. We hypothesised that the stimulation of a stress response might reduce the effects of diverse mutations. We tested this hypothesis in *C. elegans*, and found moreover that this induced mutation buffering varies across isogenic individuals because of non-genetic inter-individual differences in stress signalling. We also demonstrated that this variation in signal transduction has important consequences in wild-type animals, producing some individuals with higher stress resistance but lower reproductive fitness and other individuals with lower stress resistance and higher reproductive fitness. This may be beneficial in an unpredictable environment, acting as a "bet-hedging" strategy to diversify risk. These results illustrate how transient environmental stimuli can induce protection against mutations, how environmental responses can underlie variable mutation buffering, and how a fitness trade-off may make variation in signal transduction advantageous.

4. Chromatin Organisation in Sperm is a Major Functional Consequence of Base Composition Variation in the Human Genome

We are interested in the chromatin organization of germ cells because of the potential for the transmission of epigenetic information from one generation to the next. In most cells nucleosomes package DNA.

In sperm, however, nucleosomes are only retained at a small fraction of the genome. At a fine scale, the human genome varies extensively in the content of GC versus AT base pairs, and we found that in both genic and non-genic regions this predicts very well where nucleosomes are retained in mature sperm. These regions include transcription start sites, especially for genes that are expressed in all cells and for genes that regulate development. We also showed that regions that retain nucleosomes in sperm are likely to be protected from DNA methylation in the early embryo, suggesting a further connection between the presence of nucleosomes on the paternal genome and the establishment of gene regulation in the embryo.

5. Genome-Scale Prediction of Detrimental Mutations in Transcription Networks

The consequences of gene inhibition have been systematically studied and can be predicted reasonably well across a genome. However, many sequence variants important for disease and evolution may alter gene regulation rather than gene function. The consequences of altering a regulatory interaction (or "edge") rather than a gene (or "node") in a network have not been as extensively studied. We used an integrative analysis and evolutionary conservation to identify features that predict when the loss of a regulatory interaction is detrimental in the extensively mapped transcription network of budding yeast. Properties such as the strength of an interaction, location and context in a promoter, regulator and target gene importance, and the potential for compensation (redundancy) associate to some extent with interaction importance. Combined, however, these features predict quite well whether the loss of a regulatory interaction is detrimental across many promoters and for many different transcription factors. Thus, despite the potential for regulatory diversity, common principles can be used to understand and predict when changes in regulation are most harmful to an organism.

PUBLICATIONS

Burga A, Casanueva MO, Lehner B.
"Predicting mutation outcome from early stochastic variation in genetic interaction partners."
Nature, 480:250-254 (2011).

Casanueva MO, Burga A, Lehner B.
"Fitness trade-offs and environmentally-induced mutation buffering in isogenic C. elegans."
Science, Epub 2011 Dec 15.

Jelier R, Semple JI, Garcia-Verdugo R, Lehner B.
"Predicting phenotypic variation in yeast from individual genome sequences."
Nature Genetics, 43(12):1270-4 (2011).

Vavouri T, Lehner B.
"Chromatin organization in sperm may be the major functional consequence of base composition variation in the human genome."
PLoS Genetics, 7(4):e1002036 (2011).

Francesconi M, Jelier R, Lehner B.
"Integrated genome-scale prediction of detrimental mutations in transcription networks."
PLoS Genetics, 7(5):e1002077 (2011).

Reviews

Lehner B.
"Molecular mechanisms of epistasis within and between genes."
Trends in Genetics, 27(8):323-3 (2011).

Lehner B, Kaneko K.
"Fluctuation and response in biology."
Cell and Molecular Life Sciences, 68(6):1005-10 (2011).

SYSTEMS BIOLOGY

Group: Sensory Systems and Behaviour
This group is part of the EMBL/CRG Research Unit in Systems Biology

Group Structure:

Group Leader:	Matthieu Louis
Postdoctoral Fellows:	Alex Gomez-Marin, Sam Reid
PhD Students:	Julia Riedl, Aljoscha Schulze, Andreas Braun
Visiting Scientist:	Balaji Iyengar
Technicians:	Mariana Lopez-Matas, Moraea Philipps, Vani Rajendran

SUMMARY

The aim of our group is to unravel structure-function relationships between sensory coding, neural circuits and orientation behaviour. We are interested in understanding how orientation decisions come about in terms of neural circuit computation. This problem is tackled in the olfactory system of the fruit fly *Drosophila melanogaster* larva. Our research combines a variety of experimental and computational approaches to define the sensorimotor integration that converts sensory inputs into coordinated motor outputs. To this end, we are developing tools to track behaviour in real time and to monitor the activity of targeted neurons. Using optogenetics, we create artificial patterns of neural activity to test mechanistic hypotheses about the spatio-temporal integration of virtual sensory signals. Finally, we seek to identify the neural circuits involved in the processing of olfactory information and in the making of orientation decisions during chemotaxis. Overall, our work clarifies how a simple brain exploits sensory information to control behaviour.

RESEARCH PROJECTS

1. Peripheral Representation of Odours

Mariana Lopez-Matas, Moraea Phillips, Julia Riedl, Aljoscha Schulze

The larval 'nose' is composed of 21 olfactory sensory neurons (OSNs) expressing typically one type of odorant receptors along with a ubiquitously expressed co-receptor (Orco). Individual odorant receptors have overlapping, yet distinct, ligand tuning properties. Each OSN can be viewed as distinct information channel to the olfactory system. Whether and how one type of odorant receptor alone is capable of encoding the quality of an odour remains an open question. We have undertaken to disentangle the contribution of single OSNs to the representation of odour stimuli. In classical conditioning experiments, we found evidence that a single functional OSN is sufficient to mediate odour quality discrimination.

We developed a protocol to perform *in vivo* single-unit extracellular recordings from the larval olfactory system. Using this technique, we can monitor the activity of identified OSNs in the background of non-olfactory sensory neurons (mechanosensors and thermosensors). To separate odour-driven signals from irrelevant background, we exploit optogenetics to identify spikes of interest throughout the recording. We have begun to compare the activity elicited in identified OSNs in response to different odours. By taking advantage of a new odour delivery system that allows us to generate temporal changes in odour concentration during electrophysiological recordings, we are studying the encoding of naturalistic stimuli in the peripheral olfactory system. By creating a virtual odour reality we are able to reconstruct sensory dynamics comparable to those that a larva experiences during chemotaxis. With this approach we are aiming to explain the ability (or inability) of larvae with a single functional OSN to discriminate between distinct odours. In addition, we have started characterizing the response of single OSNs to concentration time courses corresponding to episodes of behaviour in odour gradients. We find that the activity pattern of a single OSN captures features associated with temporal changes in the stimulus intensity.

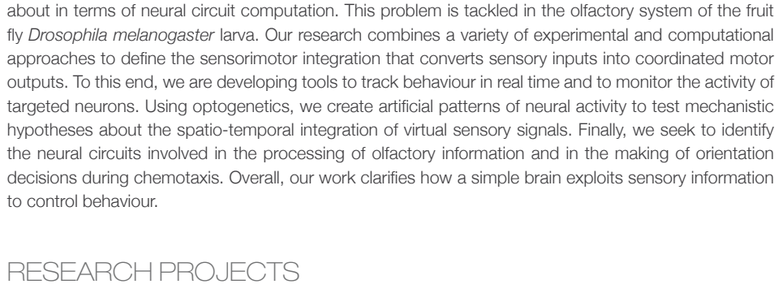


Figure 1: *A novel method for the recording of single identified olfactory sensory neurons. Sorting of the signal is achieved by defining the waveform of spikes arising from a given OSN upon excitation by channelrhodopsin (ChR2). This template-based spike sorting strategy was devised in collaboration with Vivek Jayaraman and Parvez Ahammad (HHMI-Janelia Farm).*

2. Sensorimotor Algorithm Controlling Larval Chemotaxis

Alex Gomez-Marin

Bacteria chemotax according to *indirect* orientation mechanism that consists in an improved biased random walk. In contrast, *Drosophila* larvae employ a *direct* orientation mechanism where motion is directed toward the local odour gradient. Having ruled out a mechanism purely based on stereo-olfaction (comparisons between bilateral odorant receptors in the larval head), we hypothesized that larvae orient based on active sampling where lateral head movements ("head casts") sequentially sample odour intensities in space. This essential feature of larval chemotaxis entails comparisons between stimuli measured at different times. To test this hypothesis, we have developed a computer-vision algorithm for high-resolution tracking. This tool enables us to extract morphological and kinetic features describing the behavioural state of a larva.

Larvae chemotax by punctuating bouts of straight runs with turns. Both the timing and direction of individual turns are controlled in a stimulus-dependent manner (Figure 2). In response to stereotyped decreases in concentration, larvae switch from running to turning. The initiation of a turn is preceded by a phase of head casting during which concentrations differences are measured on either side of the body axis (Figure 2). In more than 75% of the cases, a turn is implemented toward the side of higher concentrations. These results demonstrate that orientation results from an active sampling mechanism. Larvae perform a spatio-temporal comparison of differences in concentration measured during lateral head movements (head casting). These findings suggest the participation of a spatial short-term memory to the decision-making process guiding larval chemotaxis.

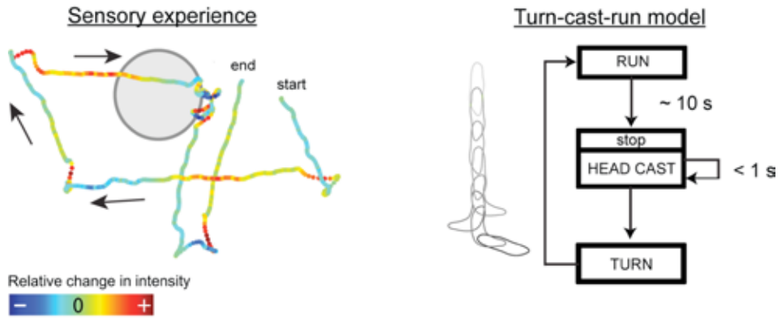


Figure 2: ***(Left)*** *Trajectory of a larva evolving in the vicinity of an odour source (gray circle). The time course of the olfactory stimulus experienced by the larva is colour-coded according to the legend at the bottom.* ***(Right)*** *Model of the sensorimotor integration controlling larval chemotaxis.*

3. Remote Control of Orientation Behaviour in Virtual Reality Experiments

Alex Gomez-Marin, Vani Rajendran and Aljoscha Schulze in collaboration with Vivek Jayaraman

Motor outputs and sensory inputs are coupled in a feedback loop that can be challenging to investigate experimentally. To address this problem, we combine real-time tracking, electrophysiology, optogenetics, quantitative analysis of behaviour and modeling. We have built a fully automated tracking system that can resolve the posture and kinetic properties of a larva at a rate of 30 Hz. The tracking system is outfitted with a set of LEDs that can be activated according to the behavioral history of the larva (head casting, running or turning, etc). We express channelrhodopsin-2 in targeted OSNs to manipulate the olfactory inputs experienced by freely moving larvae (Figure 1). We are now in position to systematically investigate the computational bases of spatiotemporal integration during orientation behaviour by reverse engineering odour evoked activity patterns with light. This approach allows us to precisely control the temporal, spatial and quantitative aspects of a stimulus, either preset or contingent upon the animal decisions (Figure 2). To date, we have managed to initiate (delay) the onset of turns by associating runs with decreases (increases) in light intensity that mimic down-gradient (up-gradient) experiences. Our goal is to elaborate a model that accurately predicts behavioral decisions from sensory signals.

4. Mapping New Neuronal Centres Participating in Orientation Behaviour

Sam Reid and Julia Riedl

To map the circuits participating in larval chemotaxis, we have initiated a behavioral screen. Whereas the first- and second-order neurons of the fly olfactory circuit are well characterized, our understanding of the downstream neurons integrating dynamic olfactory information remains poor. The Gal4-UAS binary expression system was applied to genetically silence targeted neural populations of the larval brain. More than 1000 GAL4 enhancer trap driver lines were crossed with a UAS reporter expressing tetanus toxin (TNT), a protein that efficiently silences neuronal activity by impairing synaptic transmission.

Based on high-resolution behavioral quantification and detailed assessment of expression patterns, we focused on two Gal4-lines covering a distinct population of neurons with a deficit specific to chemotaxis. One line shows altered integration of odour information during forward locomotion, failing to initiate reorientation at the correct time point. The second is severely impaired in the execution of active sampling, therefore showing strongly prolonged run episodes while using compensatory behaviours such as stalls and reverse locomotion which are rarely seen in wild type larvae. By means of controlled gain-of-function experiments, we confirmed the role of specific neurons in the modulation of run persistency and the execution of head casting. We are now conducting functional calcium imaging experiments to characterize the response properties of these neurons to dynamic odour stimuli.

5. Neural Computation Underpinning Multisensory Integration in the Larva

Andreas Braun

To adapt to complex environments, every animal constructs a model of its surroundings by combining information conveyed by different sensory modalities. This process, called multisensory integration, is essential to form a coherent representation of the external world. We investigate the potential of *Drosophila* larvae to perform multisensory integration. During their navigation toward an odour source, larvae are capable of exploiting other sensory modalities to increase their orientation performance. Like humans, larvae make decisions in conditions of incomplete evidence corrupted by noise. Our goal is to clarify the mechanisms underlying the encoding of fluctuating signals and the neural computation performing multisensory integration.

PUBLICATIONS

Gomez-Marin A and Louis M.
"Active sensation during orientation behavior in the Drosophila larva: more sense than luck."
Curr Opin Neurobiol, 2011 Dec 12. [Epub ahead of print]

Gomez-Marin A, Stephens GJ and Louis M.
"Active sampling and decision making in Drosophila chemotaxis."
Nat Commun, 2:441 (2011).

SYSTEMS BIOLOGY

Group: Comparative Analysis of Developmental Systems
This group is part of the EMBL/CRG Research Unit in Systems Biology

Group Structure:

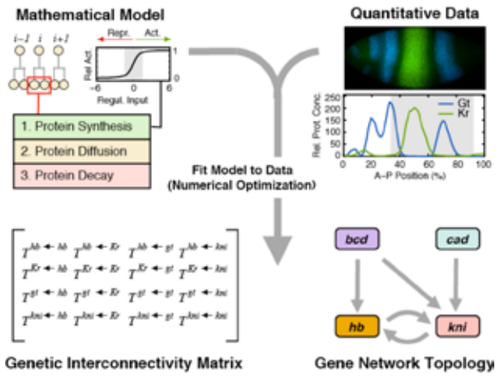
Group Leader:	Johannes Jaeger
Technician/Lab Manager:	Hilde Janssens
Technician/Programmer:	Damjan Cicin-Sain
Postdoctoral Fellows:	Anton Crombach, Eva Jiménez Guri, Bárbara Negre, Karl Wotton
PhD Students:	Astrid Hörmann, Berta Verd

SUMMARY

Natural selection acts on phenotypic variability. In particular, positive selection acts on beneficial mutations which increase fitness of an individual. The nature and availability of such beneficial mutations is a long-standing controversy in evolutionary theory. We still do not understand the kinds of phenotypes that are more (or less) likely to arise during development, and how their distribution affects evolution at the level of the genotype. As long as these issues remain unresolved, we are unable to explain the emergence of novelty and complexity in evolution.

To tackle this problem we must understand the mapping from genotype to phenotype. This relationship is complex, involving many relevant factors and their non-linear inter-actions. Therefore, we need a systems-biology approach. Our work is based on a computational, reverse-engineering method, which combines quantification of gene expression with mathematical modelling using gene circuits (Figure 1). Gene circuits are computational tools to extract the regulatory structure of a network from quantitative, spatial gene expression data. We use gene circuit models to study the developmental and evolutionary dynamics of a number of three particularly well studied gene regulatory networks (Figure 2): the gap gene system involved in segment determination, the gene regulatory network involved in muscle and heart formation, and the thoracic patterning network involved in positioning sensory bristles during insect development. It is our aim to compare these gene networks across four different dipteran species (flies, midges, and mosquitoes) in order to predict which regulatory interactions in each system are conserved, and which have diverged between species. We will test these predictions experimentally, using RNA interference (RNAi), and trace changes in regulatory interactions to molecular changes in regulatory sequences. In addition, we will use an *in silico* evolution approach to study evolutionary transitions between species. This is achieved by simulating populations of gene networks under selection to evolve from primitive to derived modes of development.

Figure 1. The gene circuit method (Jaeger & Reinitz, 2006).



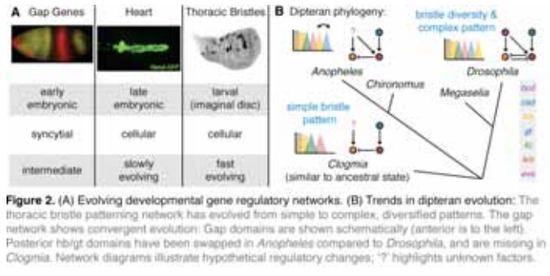


Figure 2. (A) Evolving developmental gene regulatory networks. (B) Trends in dipteran evolution: The thoracic bristle patterning network has evolved from simple to complex, diversified patterns. The gap network shows convergent evolution: Gap domains are shown schematically (anterior is to the left). Posterior hb/gt domains have been swapped in *Anopheles* compared to *Drosophila*, and are missing in *Clogmia*. Network diagrams illustrate hypothetical regulatory changes; '?' highlights unknown factors.

In contrast to traditional experimental approaches, it provides an integrative view of network evolution across multiple levels, from the molecular to the phenotypic. To our knowledge, this has not yet been achieved for any developmental system. If successful, our project will produce insights into the nature of phenotypic variability which have the potential to radically transform the methodology and conceptual framework of evolutionary developmental biology and evolutionary theory in general.

RESEARCH PROJECTS

1. A Systems-Level Analysis of Giant (*gt*) Regulation in *Drosophila melanogaster*

Astrid Hörmann, Hilde Janssens

Aim. We still do not have a precise mechanistic understanding of eukaryotic gene regulation. For this reason, our current gene network models do not include molecular details, such as specific transcription factor binding sites and *cis*-regulatory elements. To resolve this issue, we are looking into the regulation of a particular gap gene (*gt*) using our reverse-engineering approach with a detailed model of transcriptional regulation. This enables us to identify and analyze contributions of particular binding sites to the expression pattern of *gt*. In particular, we want to address important open questions such as how individual binding sites constitute a *cis*-regulatory element or how such elements interact to result in the expression of a whole, endogenous gene.

Results. A preliminary analysis of *gt cis*-regulatory elements has revealed that distinct elements contribute to different *gt* expression domains in a non-additive way, and we have identified elements responsible for early vs. late regulation. We have created strains of *Drosophila* carrying reporter constructs for these elements using site-specific transgenesis. We have quantified gene expression in two of these strains, and are using the resulting data sets to fit a model of transcriptional regulation. Preliminary analyses of the model indicate that early *gt* expression depends on maternal gradients, while later expression depends on auto-regulation. We plan to verify these predictions and extend our analysis using additional reporter strains generated by BAC recombi-neering and staining using quantitative multiplex in situ hybridization.

2. A Quantitative Study of Gap Gene Mutants in *Drosophila melanogaster*

Hilde Janssens

Aim. If we are to study evolutionary transitions using mathematical models of gene networks, we need a modeling formalism that captures the variational properties of the network. However, it remains unclear whether our current gene circuit models are able to correctly reproduce expression in mutant embryos. We will investigate these issues systematically by generating quantitative gene expression data for *Drosophila* gap gene mutants (with J. Reinitz, Stony Brook, USA), and by testing various modeling formalisms with regard to their ability to reproduce wild-type/mutant patterns correctly.

Results. Data sets for mutants of *Kr* and *kni* (Reinitz) are available, and we have created a high-quality data set for mutants of the terminal gap gene *tll*. It reveals extensive embryo-to-embryo variation as some individuals have six, some seven stripes of the pair-rule gene *even-skipped (eve)*. We have analyzed these data in detail, quantifying levels of expression and variability. This work will be submitted to *Developmental Biology*. We are now fitting models to these data to uncover the source of the variability and to systematically test various network modeling formalisms.

3. A Quantitative, Comparative Study of Gap Gene Regulation in Dipterans

Karl Wotton, Eva Jiménez-Guri, Anton Crombach, Berta Verd, Damjan Cicin-Sain

Aim. We aim to create gene circuit models of the gap gene network in three dipteran species: *Drosophila melanogaster, Megaselia abdita and Clogmia albipunctata*. To achieve this, we require detailed characterization of the early stages of development, as well as quantitative spatial gene expression patterns for all three species. Such data sets can be acquired using immunofluorescence (for protein) or *in situ* hybridization (for mRNA expression patterns; Figure 3). Gene circuits are then obtained by fitting models to data (Figure 1). The resulting gene network topologies will be used to predict, which aspects of the network have been conserved, and which ones diverged during evolution. These predictions will be verified using RNAi and reporter assays.

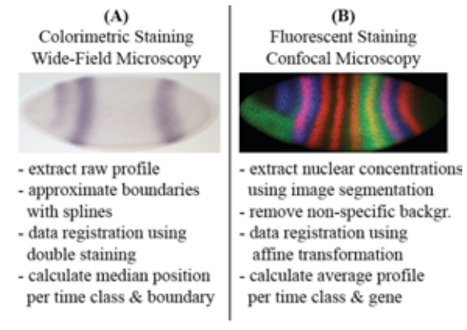


Figure 3. Comparison of data quantification methods for colorimetric (mRNA) and fluorescent (protein) staining. *Drosophila* embryos stained for *gt* (A), and Gt (blue), Kni (green), and Eve (red) (B) are shown.

Results. We have characterized the early development of all three species by DIC time-lapse microscopy. We have established RNAi protocols in *Megaselia* (Figure 4) and are doing the same for *Clogmia*. We have created early embryonic transcriptomes and have sequenced the genomes of both *Megaselia* and *Clogmia* (using 454 and HiSeq technologies). We have cloned the complete set of gap genes from *Megaselia*. We have raised antibodies against maternal proteins in *Megaselia* and *Clogmia*. We have systematically characterized gap gene expression at the mRNA level in both *Drosophila* and *Megaselia* using colorimetric *in situ* hybridization (Figure 3; similar data for *Clogmia* were acquired previously by Mónica García Solache in Cambridge, UK). We have developed a novel data quantification pipeline for our mRNA data, which was used to create expression databases and integrated expression patterns for all three species (Figure 5).

Figure 4. *bcd* RNAi knock-down in *Megaselia* embryos. *kni* expression is shown.

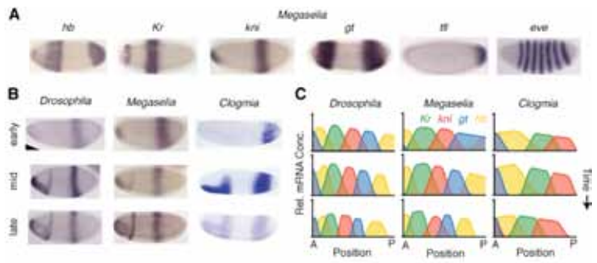


Figure 5. Gap gene datasets. (A) Gap gene expression patterns in *Megaselia* at mid blastoderm. (B) Comparison of *kni* expression over time in all three species. (C) Integrated datasets for *Drosophila*, *Megaselia* and *Clogmia*. Embryos stained by *in situ* hybridisation. A = anterior, P = posterior.

We have created an online database for these data (*superfly.crg.eu*). These data are now being used to fit gene circuit models, and will serve as a basis for our *in silico* evolution study (see below). We have shown in *Drosophila* that we can obtain consistent network structures for gap genes using both mRNA and protein data (Figure 6; work submitted to *PLoS Computational Biology*). In the same study, we have defined minimal data requirements for the gene circuit method. We are now obtaining models for the other two species. We are extending quantification of maternal protein gradients for this purpose (Figure 7). Besides, we are continually improving numerical algorithms required for model fitting, solution and analysis.

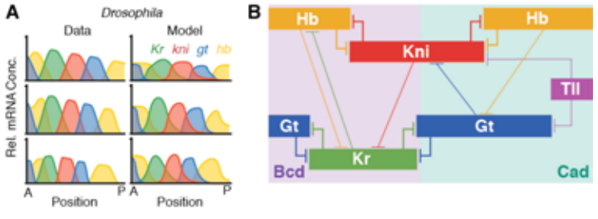


Figure 6. (A) *Drosophila* gap gene circuit models vs. mRNA data. (B) Gap gene network recovered by both models fit to protein or mRNA data. Background indicates activating inputs by maternal factors, T-bars represent repression. Boxes represent position of gap domains along the A–P axis (anterior is to the left).

4. Modeling the Evolutionary Dynamics of the Gap Gene Network

Anton Crombach

Aim. We want to explore the possibility that intermediate stages of evolution could be predicted (or reconstructed) using *in silico* evolution. In particular, we are interested whether these evolutionary transitions are constrained by network topology, or whether they depend on selection alone. Predictions from such an analysis will be tested against the data sets described above—or against qualitative expression data obtained from other suitable dipteran species.

Results. We are currently adapting evolutionary simulation code for use with our gene circuit models. We have obtained access to the Mare Nostrum computer at the Barcelona Supercomputing Center (BSC), where calculations will be carried out.

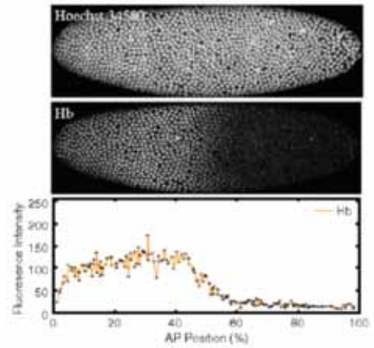


Figure 7. Hb protein expression in *Clogmia* (embryo images courtesy of Ken Siggens).

5. Modeling Development and Evolution of the Fly Heart

Eva Jiménez-Guri, Damjan Cicin-Sain

Aim. Apart from our work on the gap gene system, we are also applying our reverse-engineering method to other gene networks. One example is the network underlying muscle and heart development. This network is interesting since it is active at a later stage of development (cellularized tissues, tissue movements) and evolves extremely slowly, its core components being conserved between vertebrates and flies.

Results. We have cloned muscle- and heart-determining genes from *Megaselia* and *Clogmia* and are currently characterizing their expression and function. We are developing data quantification methods for spatial gene expression at late stages of *Drosophila* development. These will be used to create datasets for model fitting.

6. Inferring the Thoracic Bristle Patterning Network of *Drosophila*

Bárbara Negre

Aim. Another interesting example of an evolving network is the thoracic bristle patterning network. In contrast to both gap and heart networks, it is evolving extremely fast. We ultimately aim to apply our reverse-engineering approach to study the development and evolution of this network. Before this is possible, however, we will need to gain a better understanding of how it functions in *Drosophila*. For this reason, we are performing an expression screen that aims to identify novel regulators of bristle patterning. It combines expression studies with a qualitative modeling approach based on graphical primitives, to identify and predict new regulatory interactions.

Results. We have obtained probes and stained more than 100 candidate genes for regulation of bristle positioning. We are currently imaging and quantifying these expression patterns, which will be used in our expression screen.

PUBLICATIONS

Review

Jaeger J.
"The Gap Gene Network."
Cell Mol Life Sci, 68(2):243-74. Epub 2010 Oct 8 (2011).

CORE FACILITIES

Genomics Unit: Ultrasequencing Unit

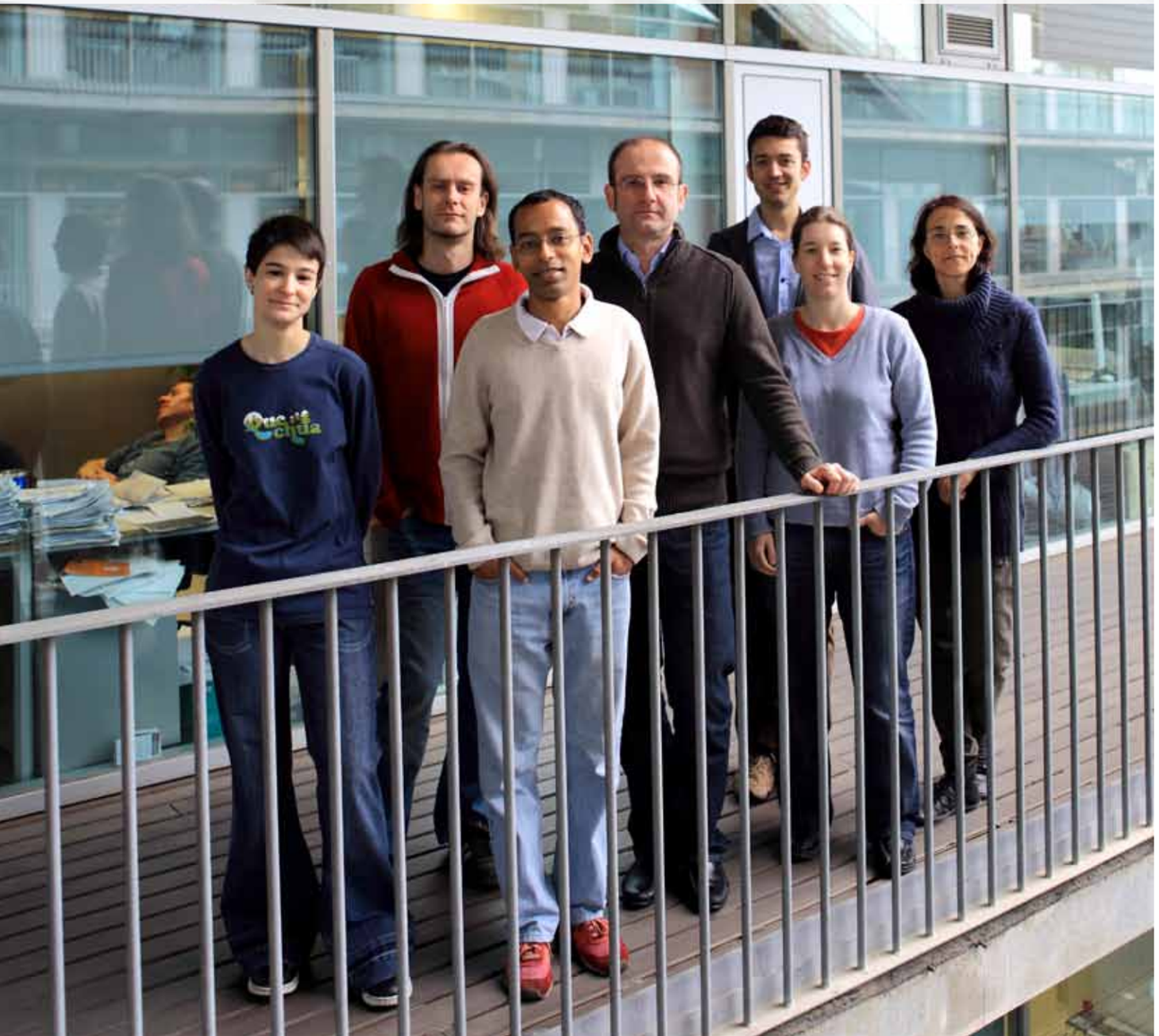
Unit Structure:

Head of the Unit: Heinz Himmelbauer

Laboratory Technicians: Anna Menoyo, Rebecca Curley, Maik Zehnsdorf, Miquel Àngel Adrover (until June 2011), Kadri Reis (until September 2011), Hui Kang (MPI Berlin, until August 2011), Núria Andreu (from July 2011), Irene González (from November 2011), Ester Cuenca (from December 2011)

Bioinformaticians: Debayan Datta, Juliane Dohm (MPI Berlin)

PhD Student: André Minoche (MPI Berlin)



SUMMARY

The Ultrasequencing Unit utilizes different high-throughput sequencing platforms (Illumina Genome Analyzer, Illumina HiSeq2000, Roche FLX). State-of-the-art sequencing services are provided to the CRG and to other users from academia or from the private sector. Additionally, the Unit performs its own research in the genomics field.

SERVICES

Second generation sequencing allows addressing a large number of different questions in biology and in biomedical sciences. The Unit is equipped with two Illumina Genome Analyzer IIx sequencers, one Illumina HiSeq2000 instrument and one FLX sequencer from 454/Roche. No additional sequencing instruments were acquired during 2011. However, company representatives were invited to present new technologies (Ion Torrent, Pacific Biosciences, Illumina MySeq) in order to evaluate the prospects of these new technologies for the Unit. During 2011, the Unit performed 42 Illumina GA sequencing runs. On the Hiseq2000, 57 runs were performed. Since installation of the Hiseq sequencer, runs are distributed between the two Illumina platforms so that short, single-read runs are performed on the GA, while long-read and paired-end runs are conducted on the Hiseq. 40 sequencing runs were performed on the Roche/FLX sequencer.

Regarding applications that we provide sequencing service for, the next generation sequencing platform at the CRG is certainly unique in offering a wide range of state-of-the-art protocols, which include genome sequencing and resequencing, exome selection, amplicon sequencing, ChIP-Seq, and different variations of transcriptome characterization at the level of total RNA, small RNA, or mRNA, including directional sequencing protocols.

During 2011, new protocols were established for both Illumina and 454 sequencing, including the use of kits which allow transcriptome sequencing from small sample amounts. Also, protocols for the preparation of matepair libraries from partially degraded DNA were developed, and different options for the sequencing of low-complexity PCR product pools were tested. Chip-Seq library preparation protocols were streamlined so that they encompassed additional quality control steps in order to ensure good data quality. On the Illumina Hiseq2000, we adopted version 3 sequencing chemistry in 2011 which currently allows the generation of over one billion reads per flow cell.

The services offered to users include the preparation of samples ready for sequencing on the GA, HiSeq2000 or on the 454-FLX instruments, and performing the sequencing runs. After basecalling, the read sequences are quality-checked and mapped against a reference genome (if applicable) and are made available to the users as fastq sequence data files.

RESEARCH PROJECTS

The research of the Ultrasequencing Unit focuses on the development of new procedures, both in the lab, and for data analysis, such as the characterization of errors and biases encountered in HiSeq2000 sequencing (Figure 1). With funding provided by the German Federal Ministry of Education and Research (BMBF), we have generated a high-quality reference genome sequence for sugar beet (*Beta vulgaris*). After misassembly correction, the reference assembly RefBeet-1.0 encompassed 596 Mbp with an N50 length of 1.5 Mbp. Sixty-six percent of RefBeet-1.0 was chromosomally or genetically anchored: this proportion of the draft was contained within 222 scaffolds (N50 length of 3 Mbp). The annotation of coding regions within the genome draft was supported by almost 600 million mRNA-Seq reads. So far, 31267 genes have been discovered with full or partial experimental support. To analyse intraspecific variation we sequenced several additional lines and compared them to the reference genome (Figure 2).

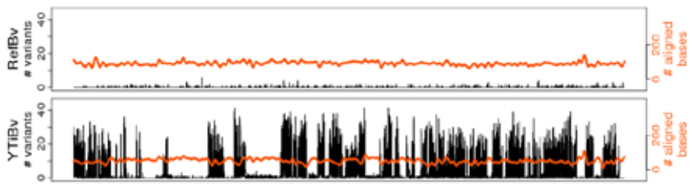


Figure 1: *Biased distribution of low quality bases in PhiX read data (HiSeq2000 sequencing). Reads from forward and from reverse strands were analyzed separately (A) Number of bases within B-tails (consecutive bases of Q-score=2 at 3' end of a read) per position; (B) average Q-score of bases in untrimmed reads; (C) average Q-score of bases in B-tail trimmed reads; (D) observed per-base substitution error rate. Accumulation of low quality values is observed even after B-tail removal. Peaks of observed error rates occur at positions where increased low quality counts are detected, and in most cases peaks are seen only on one strand (Minoche et al., 2011).*

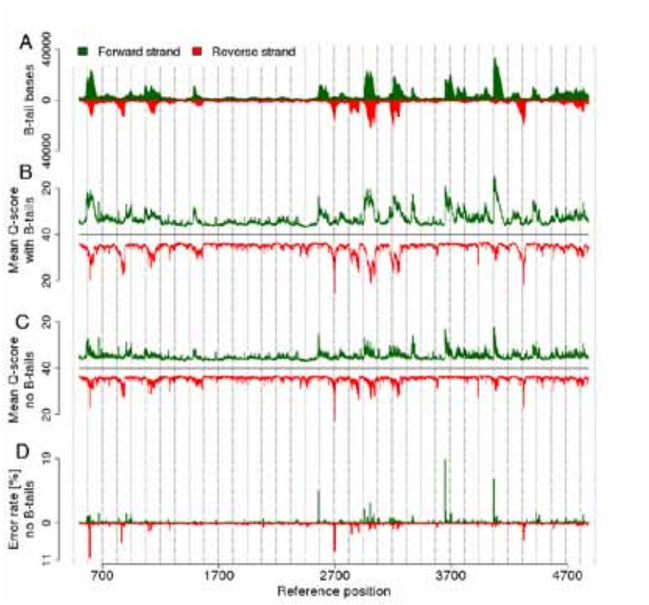


Figure 2: *Comparison of sugar beet chromosome 7 (50 Mb) between a reference accession and another line. Columns in black indicate SNV frequency per window of 2 kbp. The red line indicates read coverage. The top panel shows a comparison between the reference genome to Illumina reads generated from the reference for background determination. The lower panel shows a comparison between the reference and another line. Conspicuous is the mosaic of regions exhibiting low or high sequence diversity.*

PUBLICATIONS

Minoche AE, Dohm JC, Himmelbauer H.
"Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems."
Genome Biology, 12:R112 (2011).

Quesada V, [..], Himmelbauer H, [..] López-Otín C.
"Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia."
Nature Genetics, 44:47-52 (2011).

Dobbernack G, Meinl W, Schade N, Florian S, Wend K, Voigt I, Himmelbauer H, Gross M, Liehr T, Glatt HR.
"Altered tissue distribution of 2-amino-1methyl-6-phenylimidazol[4,5-b]pyridine-DNA adducts in mice transgenic for human sulfotransferases 1A1 and 1A2."
Carcinogenesis, 32:1734-1740 (2011).

Schulz N, Himmelbauer H, Rath M, van Weeghel M, Houten S, Suhre K, Scherneck S, Vogel H, Kluge R, Wiedmer P, Joost H-G, Schürmann A.
"Role of medium and short-chain L-3-hydroxyacyl-CoA dehydrogenase in the regulation of body weight and thermogenesis."
Endocrinology, 152:4641-4651 (2011).

Schroeder IS, Sulzbacher S, Nolden T, Fuchs J, Czarnota J, Meisterfeld R, Himmelbauer H, Wobus AM.
"Induction and selection of Sox17 expressing endoderm cells generated from murine embryonic stem cells."
Cells Tissues Organs, in press.

Zakrzewski F, Weisshaar B, Fuchs J, Bannack E, Minoche AE, Dohm JC, Himmelbauer H, Schmidt T.
"Epigenetic profiling of heterochromatic satellite DNA."
Chromosoma, 120:409-422 (2011).

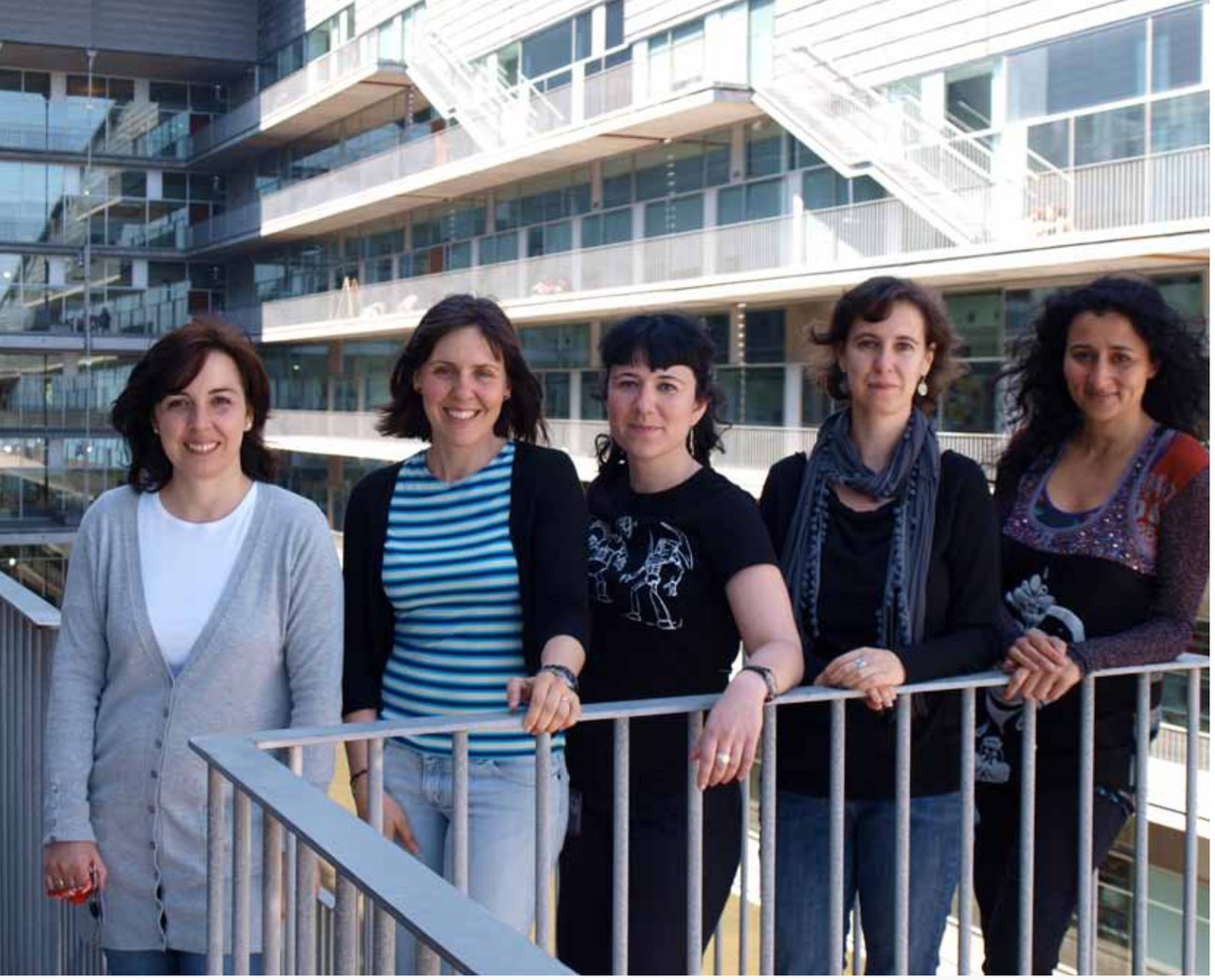
Llorens F, Hummel M, Pastor X, Ferrer A, Pluvinet R, Vivancos A, Castillo E, Iraola S, Mosquera AM, Gonzalez E, Lozano J, Ingham M, Dohm J., Noguera M, Kofler R, Del Rio JA, Bayés M, Himmelbauer H, Sumoy L.
"Multiple platform assessment of the EGF dependent transcriptome by microarray and deep tag sequencing analysis."
BMC Genomics, 12:326 (2011).

Herrmann F, Garriga-Canut M, Baumstark R, Fajardo-Sanchez E, Cotterell J, Minoche A, Himmelbauer H, Isalan M.
"p53 Gene Repair with Zinc Finger Nucleases Optimised by Yeast 1-Hybrid and Validated by Solexa Sequencing."
PLoS One, 6:e20913 (2011).

Puente XS, [..] Castillo E, Dohm JC, Himmelbauer H, [..] Campo E.
"Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia."
Nature, 475:101-105 (2011).

Esteve A, Kofler R, Himmelbauer H, Ferretti L, Vivancos AP, Groenen MAM, Folch JM, Rodríguez MC, Pérez-Enciso M.
"Partial short-read sequencing of a highly inbred Iberian pig and genomics inference thereof."
Heredity, 107:256-264 (2011).

Meunier D, Smits R, Hägebarth A, Lüttges A, Jaussi R, Quintanilla-Fend L, Graff JM, van de Wetering M, Muncan V, Clevers H, Himmelbauer H, Fodde R, Fundele R.
"Expression analysis of Proline Rich 15 (Prr15) in mouse and human gastrointestinal tumours."
Molecular Carcinogenesis, 50:8-15 (2011).



CORE FACILITIES

Genomics Unit: Genotyping Unit

Unit Structure:

Manager of the Unit: Magda Montfort

Technicians: Pilar Herruzo, Cinta Pegueroles (until March 2011), Carles Arribas (until May 2011), Ester Cuenca (from April to November 2011), Sheila Santín (from April 2011), Susana Coca (from May 2011)

SUMMARY

Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation in the human genome. A fraction of this genetic variation is likely to explain the majority of the differences between individuals, including their predisposition to develop common human disorders, such as cardiovascular disease, diabetes, asthma and cancer, and their differences in response to drugs. Other goals of SNP research include population genetics and trait selection for agricultural, cattle farming or aquaculture applications. Some of the genotyping technologies also enable to accurately characterize copy number variants (CNVs), loss of heterozygousity (LOH) and DNA methylation status.

SERVICES

The Genotyping Unit is one of the nodes of the Plataforma en Red de Genotipado Carlos III (CeGen-ISCIII, *www.cegen.org*). It provides support to PRBB users and external users from public or private institutions for genotyping projects in every aspect of research, from experiment design, DNA extraction, genotyping, methylation profiling, and data interpretation.

At the CeGen-ISCIII CRG Barcelona Node several medium and high throughput genotyping and related services were available in 2011:

1. Automated DNA extraction from blood or other tissues (Chemagen)
2. DNA quantification using Picogreen (Molecular Probes)
3. Whole Genome Amplification using GenomiPhi (Amersham)
4. Custom Genotyping by VeraCode GoldenGate (Illumina): genotyping of 48-384 SNPs selected by customer.
5. Custom Genotyping by GoldenGate technology (Illumina): genotyping of 96-3,072 SNPs selected by customer.
6. Focused-content SNP Genotyping with GoldenGate technology on all catalogue Illumina products: African American admixture panel, DNA test panel, Cancer SNP panel and MHC panel set, Human Linkage V panel, Mouse LD and MD Linkage panels and VeraCode ADME Core Panel.
7. Focused-content SNP Genotyping with Infinium technology on all catalogue Illumina products: HumanOmni5-quad, HumanOmni2.5-quad, HumanOmni1-quad, HumanOmni-Express, Human660W-quad, HumanCytoSNP-12, HumanLinkage-24, BovineSNP50, CanineSNP20, EquineSNP50, others.
8. Custom DNA methylation profiling using Veracode GoldenGate (Illumina): 96 or 384 sites selected by costumer.
9. Focused-content DNA methylation profiling using Infinium technology (Illumina): HumanMethylation27 (Figure 1) and HumanMethylation450.

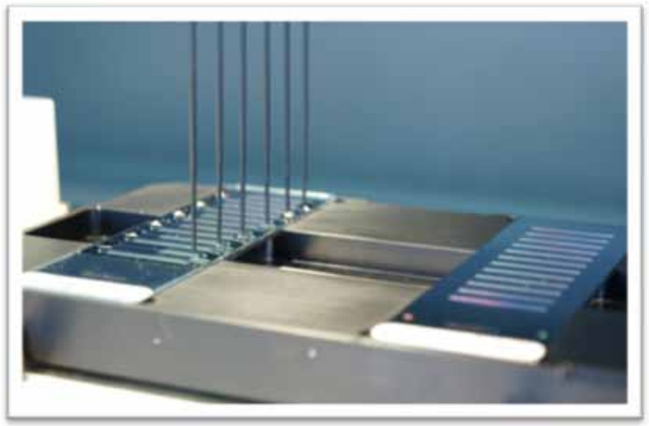


Figure 1.
Loading DNA samples on Illumina HumanMethylation27 BeadChips using liquid handling robots.

All services are integrated with robust software tools for experimental design, management of data and analyses. Extensive quality control measures let us further refine the quality of data.

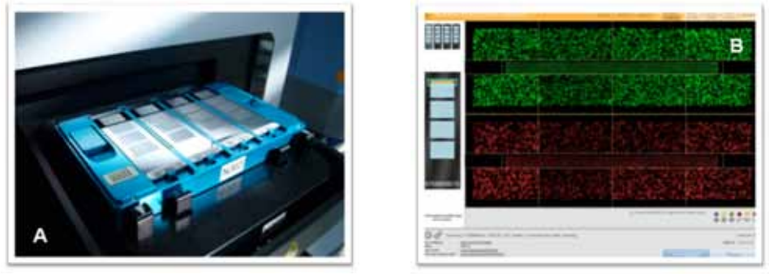


Fig 2.
A. The facility is equipped with state-of-the art instruments including high resolution scanners. B. Section derived from a HumanOmni1-Quad Beadchip under standard scanning conditions using Illumina iScan system.

During this period, Genotyping Unit has finished 60 genotyping projects, produced more than 3.6 million genotypes for GoldenGate custom designs and processed near 500 beadchips (doubling last year's beadchips) through Infinium technology. It is worth to notice that the genotyping platforms were used at around 80% capacity during this period. The number of processed samples and the number of work orders for each service are detailed in the following table.

Service	Samples	Work Orders
Custom Genotyping	15664	29
Infinium Genotyping	2020	18
Infinium Methylation	696	13
DNA Extractions	1596	8
DNA Quantifications	20639	48
DNA WGA	1437	4

PUBLICATIONS

Mancuso FM, Montfort M, Carreras A, Alibes A, Roma G.
"HumMeth27QCReport: an R package for quality control and primary analysis of Illumina Infinium methylation data."
BMC Res Notes, 4(1):546 (2011).

PUBLICATIONS BY EXTERNAL AND CRG GENOTYPING UNIT USERS

Abulí A, Fernández-Rozadilla C, Alonso-Espinaco V, Muñoz J, Gonzalo V, Bessa X, González D, Clofent J, Cubiella J, Morillas JD, Rigau J, Latorre M, Fernández-Bañares F, Peña E, Riestra S, Payá A, Jover R, Xicola RM, Llor X, Carvajal-Carmona L, Villanueva CM, Moreno V, Piqué JM, Carracedo A, Castells A, Andreu M, Ruiz-Ponte C, Castellví-Bel S; Gastrointestinal Oncology Group of the Spanish Gastroenterological Association.
"Case-control study for colorectal cancer genetic susceptibility in EPICOLON: previously identified variants and mucins."
BMC Cancer, 11:339 (2011).

Abulí A, Fernández-Rozadilla C, Giráldez MD, Muñoz J, Gonzalo V, Bessa X, Bujanda L, Reñé JM, Lanas A, García AM, Saló J, Argüello L, Vilella A, Carreño R, Jover R, Xicola RM, Llor X, Carvajal-Carmona L, Tomlinson IP, Kerr DJ, Houlston RS, Piqué JM, Carracedo A, Castells A, Andreu M, Ruiz-Ponte C, Castellví-Bel S; Gastrointestinal Oncology Group of the Spanish Gastroenterological Association.
"A two-phase case-control study for colorectal cancer genetic susceptibility: candidate genes from chromosomal regions 9q22 and 3q22."
Br J Cancer, 105(6):870-5 (2011).

Alonso P, Gratacòs M, Segalàs C, Escaramís G, Real E, Bayés M, Labad J, Pertusa A, Vallejo J, Estivill X, Menchón JM.
"Variants in estrogen receptor alpha gene are associated with phenotypical expression of obsessive-compulsive disorder."
Psychoneuroendocrinology, 36(4):473-83 (2011).

Bertinetto FE, Calafell F, Roggero S, Chidichimo R, Garino E, Marcuccio C, Coppo R, Scolari F, Frascá G, Savoldi S, Schena FP, Amoroso A
"Search for genetic association between IgA nephropathy and candidate genes selected by function or by gene mapping at loci IGAN2 and IGAN3."
Nephrol Dial Transplant, 2011 Nov 29. [Epub ahead of print]

Cabezas JA, Ibáñez J, Lijavetzky D, Vélez MD, Bravo G, Rodríguez V, Carreño I, Jermakow AM, Carreño J, Ruiz-García L, Thomas MR, Martínez-Zapater JM.
"A 48 SNP set for grapevine cultivar identification."
BMC Plant Biol, 11:153 (2011).

Calvanese V, Fernández AF, Urdinguio RG, Suárez-Alvarez B, Mangas C, Pérez-García V, Bueno C, Montes R, Ramos-Mejía V, Martínez-Camblor P, Ferrero C, Assenov Y, Bock C, Menendez P, Carrera AC, Lopez-Larrea C, Fraga MF.
"A promoter DNA demethylation landscape of human hematopoietic differentiation."
Nucleic Acids Res, 2011 Sep 12. [Epub ahead of print]

Garcia-Linares C, Fernandez-Rodriguez J, Terribas E, Mercade J, Pros E, Benito L, Benavente Y, Capella G, Ravella A, Blanco I, Kehrer-Sawatzki H, Lazaro C, Serra E.
"Dissecting Loss of Heterozygosity (LOH) in Neurofibromatosis Type 1-Associated Neurofibromas: Importance of Copy Neutral LOH."
Hum Mutat, 32(1):78-90 (2011).

Carreño O, Corominas R, Fernández-Morales J, Camiña M, Sobrido MJ, Fernández-Fernández JM, Pozo-Rosich P, Cormand B, Macaya A.
"SNP variants within the vanilloid TRPV1 and TRPV3 receptor genes are associated with migraine in the Spanish population."
Am J Med Genet B Neuropsychiatr Genet, 2011 Dec 7. [Epub ahead of print].

Toma C, Hervás A, Balmaña N, Salgado M, Maristany M, Vilella E, Aguilera F, Orejuela C, Cuscó I, Gallastegui F, Pérez-Jurado LA, Caballero-Andaluz R, de Diego-Otero Y, Guzmán-Alvarez G, Ramos-Quiroga JA, Ribasés M, Bayés M, Cormand B.
"Neurotransmitter systems and neurotrophic factors in autism: association study of 37 genes suggests involvement of DDC."
The World Journal of Biological Psychiatry, in press.

Domingues-Montanari S, Hernandez-Guillamon M, Fernandez-Cadenas I, Mendioroz M, Boada M, Munuera J, Rovira A, Maisterra O, Parés M, Gutierrez M, Alvarez-Sabin J, Chacón P, Delgado P, Montaner J; Stroke Project Cerebrovascular Diseases Study Group, Spanish Society of Neurology.
"ACE variants and risk of intracerebral hemorrhage recurrence in amyloid angiopathy."
Neurobiol Aging, 32(3):551 (2011).

Duell EJ, Sala N, Travier N, Muñoz X, Boutron-Ruault MC, Clavel-Chapelon F, Barricarte A, Arriola L, Navarro C, Sánchez-Cantalejo E, Quirós JR, Krogh V, Vineis P, Mattiello A, Tumino R, Khaw KT, Wareham N, Allen NE, Peeters PH, Numans ME, Bueno-de-Mesquita HB, van Oijen MG, Bamia C, Benetou V, Trichopoulos D, Canzian F, Kaaks R, Boeing H, Bergmann MM, Lund E, Ehrnström R, Johansen D, Hallmans G, Stenling R, Tjønneland A, Overvad K, Ostergaard JN, Ferrari P, Fedirko V, Jenab M, Nesi G, Riboli E, González CA.
"Genetic variation in alcohol dehydrogenase (ADH1A, ADH1B, ADH1C, ADH7) and aldehyde dehydrogenase (ALDH2), alcohol consumption, and gastric cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort."
Carcinogenesis, 2011 Dec 5. [Epub ahead of print].

Fernandez-Cadenas I, Mendioroz M, Domingues-Montanari S, Del Rio-Espinola A, Delgado P, Ruiz A, Hernandez-Guillamon M, Giralt D, Chacon P, Navarro-Sobrino M, Ribo M, Molina CA, Alvarez-Sabin J, Rosell A, Montaner J.
"Leukoaraiosis is associated with genes regulating blood-brain barrier homeostasis in ischaemic stroke patients."
Eur J Neurol, 18(6):826-35 (2011).

Fernàndez-Castillo N, Cormand B, Roncero C, Sánchez-Mora C, Grau-Lopez L, Gonzalvo B, Miquel L, Corominas R, Ramos-Quiroga JA, Casas M, Ribasés M.
"Candidate pathway association study in cocaine dependence: The control of neurotransmitter release."
World J Biol Psychiatry, 2011 Mar 23. [Epub ahead of print].

Fullana MA, Alonso P, Gratacòs M, Jaurrieta N, Jiménez-Murcia S, Segalàs C, Real E, Estivill X, Menchón JM.
"Variation in the BDNF Val66Met polymorphism and response to cognitive-behavior therapy in obsessive-compulsive disorder."
Eur Psychiatry. 2011 Dec 6. [Epub ahead of print].

Gomaa NH, Montesinos-Navarro A, Alonso-Blanco C, Picó FX.
"Temporal variation in genetic diversity and effective population size of Mediterranean and subalpine Arabidopsis thaliana populations."
Molecular Ecology, 20:3540-3554 (2011).

Javed A, Melé M, Pybus M, Zalloua P, Haber M, Comas D, Netea MG, Balanovsky O, Balanovska E, Jin L, Yang Y, Arunkumar G, Pitchappan R, Bertranpetit J, Calafell F, Parida L; The Genographic Consortium.
"Recombination networks as genetic markers in a human variation study of the Old World."
Hum Genet, 2011 Oct 18. [Epub ahead of print].

Melé M, Javed A, Pybus M, Zalloua P, Haber M, Comas D, Netea MG, Balanovsky O, Balanovska E, Jin L, Yang Y, Pitchappan RM, Arunkumar G, Parida L, Calafell F, Bertranpetit J; The Genographic Consortium.
"Recombination Gives a New Insight in the Effective Population Size and the History of the Old World Human Populations"
Mol Biol Evol, Epub 2011 Sep 1.

Méndez-Vigo B, Picó FX, Ramiro M, Martínez-Zapater JM, Alonso-Blanco C.
"Altitudinal and Climatic Adaptation Is Mediated by Flowering Traits and FRI, FLC, and PHYC Genes in Arabidopsis."
Plant Physiol, 157(4):1942-55 (2011).

Morales E, Bustamante M, González JR, Guxens M, Torrent M, Méndez MA, García-Esteban R, Julvez J, Forns J, Vrijheid M, Molto-Puigmarti C, López-Sabater MC, Estivill X, Sunyer J.
"Genetic Variants of the FADS Gene Cluster and ELOVL Gene Family, Colostrums LC-PUFA Levels, Breastfeeding, and Child Cognition."
PLoS ONE, 6(2): e17181 (2011).

Muiños-Gimeno M, Espinosa-Parrilla Y, Guidi M, Kagerbauer B, Sipilä T, Maron E, Pettai K, Kananen L, Navinés R, Martín-Santos R, Gratacòs M, Metspalu A, Hovatta I, Estivill X.
"Human microRNAs miR-22, miR-138-2, miR-148a, and miR-488 are associated with panic disorder and regulate several anxiety candidate genes and related pathways."
Biol Psychiatry, 69(6):526-33 (2011).

Ribasés M, Ramos-Quiroga JA, Sánchez-Mora C, Bosch R, Richarte V, Palomar G, Gastaminza X, Bielsa A, Arcos-Burgos M, Muenke M, Castellanos FX, Cormand B, Bayés M, Casas M.
"Contribution of LPHN3 to the genetic susceptibility to ADHD in adulthood: a replication study."
Genes Brain Behav, 10(2):149-57 (2011).

Sala N, Muñoz X, Travier N, Agudo A, Duell EJ, Moreno V, Overvad K, Tjonneland A, Boutron-Ruault MC, Clavel-Chapelon F, Canzian F, Kaaks R, Boeing H, Meidtner K, Trichopoulos A, Tsiotas K, Zylis D, Vineis P, Panico S, Palli D, Krogh V, Tumino R, Lund E, Bueno-de-Mesquita HB, Numans ME, Peeters PH, Quirós JR, Sánchez MJ, Navarro C, Ardanaz E, Dorronsoro M, Hallmans G, Stenling R, Manjer J, Allen NE, Travis RC, Khaw KT, Jenab M, Offerhaus GJ, Riboli E, González CA.
"Prostate stem-cell antigen gene is associated with diffuse and intestinal gastric cancer in Caucasians: Results from the EPIC-EURGAST study."
Int J Cancer, 2011 Jun 16. [Epub ahead of print]

Sikora M, Laayouni H, Calafell F, Comas D, Bertranpetit J.
"A genomic analysis identifies a novel component in the genetic structure of sub-Saharan African populations."
European Journal of Human Genetics, 19(1):84-88 (2011).

Toma C, Hervás A, Balmaña N, Vilella E, Aguilera F, Cuscó I, del Campo M, Caballero R, De Diego-Otero Y, Ribasés M, Cormand B, Bayés M.
"Association study of six candidate genes asymmetrically expressed in the two cerebral hemispheres suggests the involvement of BAIAP2 in autism."
J Psychiatr Res, 45(2):280-2 (2011).

Zinelabidine LH, Haddioui A, Rodríguez V, Cabello F, Eiras-Dias JE, Cabello F, Martínez-Zapater JM, Ibáñez J.
"Identification by SNP analysis of a major role for Cayetana Blanca in the genetic network of Iberian Peninsula grapevine varieties."
Am J Enol Vitic 63, Published ahead of print November 11, 2011.

CORE FACILITIES

Genomics Unit: Microarrays Unit

Unit Structure:

Manager of the Unit: Anna Ferrer

Technicians: Maria Aguilar, Heidi Mattlin (since February 2011), Inês Guimãraes (from May 2011)

Bioinformaticians: Sarah Bonnin, Manuela Hummel

SUMMARY

Microarrays allow us to analyze expression profiles (mRNA and microRNA levels) and structural variation (DNA copy number) on a genome-wide level. The Microarrays Unit is equipped with all the instruments to carry out microarray analysis with the latest version of Agilent, Illumina and Roche-NimbleGen technologies. For all technologies, the Microarrays Unit provides a full service, including technical advice in experimental design, quality control of starting material, sample and array processing, and data analysis.

SERVICES

The laboratory provides microarray methodologies as a service to scientists from the CRG, other PRBB institutions and external public and private institutions. Services offered include: sample quality control, RNA and DNA sample labelling, hybridization of microarrays and data processing and analysis, as well as array design and fabrication by spotting.

The Microarrays Unit offers a range of options suitable for a variety of applications including:

- mRNA expression profiling on Agilent microarrays using the 2-color and 1-color protocols, and Illumina BeadArrays using 1-color protocol.
- microRNA expression profiling on Agilent microarrays.
- array based comparative genomic hybridization (aCGH) on spotted BAC, Agilent and Roche-NimbleGen microarrays
- chromatin immunoprecipitation on array (ChIP-on-chip) on promoter and CpG island microarrays from Agilent.

During 2011 the unit has finished 52 projects. From those, 45 consisted in mRNA expression profiling, the most common application at the Microarrays Unit. In total, we have processed 706 samples for mRNA expression profiling, 24 samples for microRNA profiling and 24 for aCGH. In addition, we have fabricated 160 arrays using the Versarray spotter. The unit also provides access and support to real-time qPCR instruments and offers a service for DNA shearing using the Covaris system.

DEVELOPMENT AND NEW SERVICES

There's an increasing need at the scientific community to use lower amount of sample. For that microarrays unit has made an effort to implement new protocols and a workflow, starting from the RNA extraction from low number of cells using the Agencourt RNA Clean XP beads (Beckman Coulter) (Figure 1), continuing with the Whole Transcriptome Amplification using the TransPlex® Complete Whole Transcriptome Amplification Kit (Sigma) and ending with the hybridization of the amplified material to the Agilent gene expression microarrays.

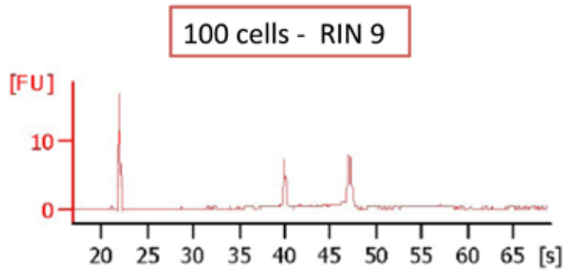


Figure 1:
Quality control of RNA obtained by using the Agencourt RNA Clean XP beads from 100 cells (QC done by the Bioanalyzer).

This methodology has already been used for 7 projects at the unit, 3 of those, doing the complete workflow, from the RNA extraction till the hybridization, and 4 using an already amplified sample to hybridize to the gene expression microarray.

As every year, we have carried out an evaluation of the microarray technologies the unit has, as our annual internal control. We used the MAQC samples (*Universal Human Reference RNA*, and *Brain RNA*). This year we did not include in the test the Illumina platform and the Agilent 2-colors protocol, since they are not used that much at the unit. On the other hand, the platform comparison comprised our first test of the new Agilent Exon 4x180K arrays.

The 4x180K Exon array includes less genes than the 8x60K gene expression format, but each gene is represented by many different probes, corresponding to different exons (Figure 2A). The new array type is especially suited to studying alternative splicing, whereas for classical gene expression studies we still recommend the standard Agilent gene expression formats.

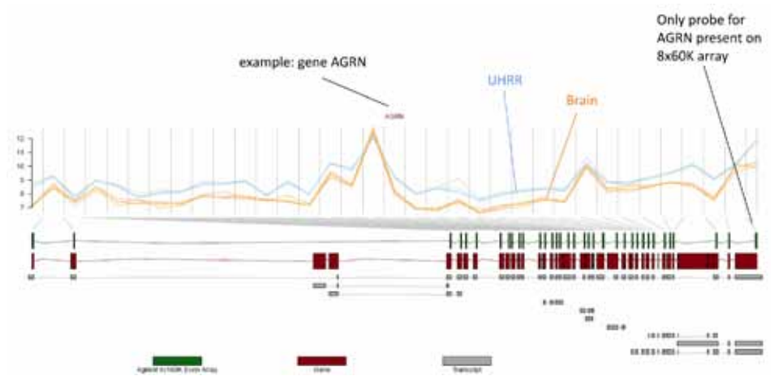


Figure 2A:
Illustration of the genomic positions of exons and exon array probes, using the example of gene AGRN, along with corresponding expression values obtained for replicated UHRR and Brain samples in the test experiment.

Both Agilent arrays show good results in our test experiments (Figure 2B). The performance in terms of reproducibility, sensitivity, specificity, accuracy, and inter-platform comparability is very much in line with our results from the last years.

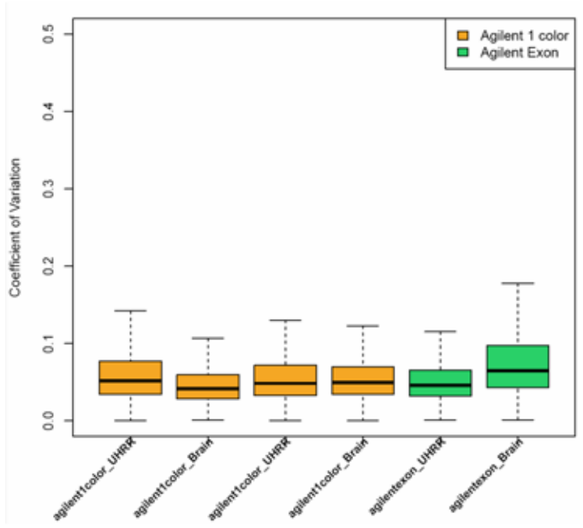


Figure 2B:
Variability across sample replicates, measured by the coefficient of variation, for each sample type within each platform/protocol. The lower the coefficient of variation, the better is the technical reproducibility.

In June 2011 microarrays unit acquired the Roche-NimbleGen platform for microarrays experiments. With the Roche-NimbleGen platform it is possible to perform aCGH/CNV, DNA methylation, chromatin immunoprecipitation on a chip, and gene expression experiments. With the acquisition it's possible to offer more microarray options to the users. For the moment, the aCGH protocol is the only one that has been established and already used for a custom aCGH 3x720k array project.

PUBLICATIONS

Hummel M, Bonnin S, Lowy E, Roma G.
"TEQC: an R-package for quality control in target capture experiments."
Bioinformatics, 27(9):1316-7 (2011).

Llorens F, Hummel M, Pastor X, Ferrer A, Pluvinet R, Vivancos A, Castillo E, Iraola S, Mosquera AM, Gonzalez E, Lozano J, Ingham M, Dohm JC, Noguera M, Kofler R, del Rio JA, Bayes M, Himmelbauer H, Sumoy L.
"Multiple platform assessment of the EGF dependent transcriptome by microarray and deep tag sequencing analysis."
BMC Genomics, 12:326 (2011).

CORE FACILITIES

CRG/UPF Proteomics Unit

Unit Structure:

Head of the Unit: Henrik Molina (until June 2011)

Acting Unit Manager: Cristina Chiva (from July to December 2011)

Scientists: Cristina Chiva (UPF), Eva Borras (UPF), Susagna Poses (ProteoRed), Guadalupe Espadas-Garcia (CRG), Francesco Mancuso (CRG)

SUMMARY

The word proteome refers to the set of proteins expressed by the genome in a given cell type, tissue or organism at a certain time and under defined conditions. By definition the proteome is a dynamic entity that changes over time, between different cellular compartments, and in response to cell stimulation. Therefore, beyond the identification of the expressed proteins, protein quantification has become essential to understand differences in cellular physiology and cell signalling pathways. Indeed the field of proteomics has turned quantitative in the last few years and several strategies for protein quantification have been developed including both labelled and label-free strategies. Moreover, a myriad of post-translational modifications such as phosphorylation, acetylation, and methylation among many others, modify protein sequences and alter their activity and influence their interaction with other proteins. Therefore, in addition to protein quantification, the identification and the localization of certain post-translational modifications is a challenge of great importance in the proteomics field.

Mass spectrometry has proven a valuable technology for studying the proteomes and it is safe to say that mass spectrometry has been the driving force of proteomics. Today, close to half of all proteomics studies rely on mass spectrometry, which is also the case for the CRG/UPF Joint Proteomics Unit. The main issues in the field address where and when specific proteins are expressed; how proteins interact to each other; how proteins are modified by post-translational modifications; how proteins respond with respect to stimuli or other cellular states; and how exons are combined to produce alternative gene products or protein splicing variants.

The Proteomics Unit works in close collaboration with many researchers at the Centre de Regulació Genòmica (CRG) and Universitat Pompeu Fabra (UPF), as well as several other institutions hosted at the Parc de Recerca Biomèdica de Barcelona (PRBB). Provided services include the identification of proteins by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS), determination of molecular mass; identification of post-translational modification, and relative and absolute quantitation of proteins based on stable isotopes. The unit operates two modern Orbitrap mass spectrometers both equipped with nano-HPLC. Mid 2011 the unit bought an additional mass spectrometer (Q-Trap 5500 AB Sciex) that is dedicated to targeted proteomics experiments, a technique in which a preselected list of proteins can be consistently quantified in different experimental conditions or system perturbations.

FINANCE

In 2011 the Proteomics Unit implemented a new cost model for services, in which several proteomics workflows were split in blocks representing an independent step of a protocol. It is now possible to combine different experimental blocks to offer customized solutions to our users without charging for blocks that are not used.

AFFILIATIONS

The Joint UPF/CRG Proteomics Core Facility is a member of ProteoRed, the Spanish Proteomics Network, and Prime-XS, the European project to promote, develop and grant access to proteomics core facilities.

PUBLICATIONS

Hyland EM, Molina H, Poorey K, Jie C, Xie Z, Dai J, Qian J, Bekiranov S, Auble DT, Pandey A, Boeke JD.
"An evolutionarily 'young' lysine residue in histone H3 attenuates transcriptional output in Saccharomyces cerevisiae."
Genes Dev, 25(12):1306-19 (2011).

Mancuso FM, Montfort M, Carreras A, Alibes A, Roma G.
"HumMeth27QCReport: an R package for quality control and primary analysis of Illumina Infinium methylation data."
BMC Res Notes, 4(1):546 (2011).

PUBLICATIONS BY CRG PROTEOMICS UNIT USERS

García-Santamarina S, Boronat S, Espadas G, Ayté J, Molina H, Hidalgo E.
"The oxidized thiol proteome in fission yeast--optimization of an ICAT-based method to identify H2O2-oxidized proteins."
J Proteomics, 74(11):2476-86 (2011).

Moriyama M, Fukuhara T, Britschgi M, He Y, Narasimhan R, Villeda S, Molina H, Huber BT, Holers M, Wyss-Coray T.
"Complement receptor 2 is expressed in neural progenitor cells and regulates adult hippocampal neurogenesis."
J Neurosci, 31(11):3981-9 (2011).

Zachara NE, Molina H, Wong KY, Pandey A, Hart GW.
"The dynamic stress-induced "O-GlcNAc-ome" highlights functions for O-GlcNAc in regulating DNA damage/repair and other cellular pathways."
Amino Acids, 40(3):793-808 (2011).

CORE FACILITIES

CRG/UPF Advanced Light Microscopy Unit

Unit structure:
Head of the Unit: Timo Zimmermann

Microscopy Specialists: Raquel Garcia Olivas, Arrate Mallabiabarrena, Xavier Sanjuan (UPF)

SUMMARY

The Advanced Light Microscopy Unit (ALMU) of the CRG and UPF serves as a core facility for high-end light microscopy for PRBB researchers. A range of instruments with unique capabilities fully covers the spectrum of advanced imaging applications from thick tissue reconstruction to fast *in-vivo* imaging to the sensitive detection of very faint signals of single molecules. The staff of the facility provides advice in the initial experiment planning, training of the researchers on the instruments and assistance with the subsequent data analysis. It is the aim of the facility to provide a link for the biological questions of researchers to the full capabilities of advanced light microscopy at the organismic, cellular and molecular level. Methods available in the facility include super-resolution microscopy by stimulated emission depletion (STED), optical sectioning (single photon and multi-photon microscopy), spectral imaging, *in-vivo* timelapse imaging, Total Internal Reflection Fluorescence (TIRF) Microscopy and methods for the study of molecular properties and interactions like Fluorescence Correlation Spectroscopy (FCS), Fluorescence Lifetime Imaging Microscopy (FLIM), Fluorescence Resonance Energy Transfer (FRET) detection and Fluorescence Recovery after Photobleaching (FRAP). Additionally, dedicated software packages for data visualization and analysis are available for 3D rendering, particle tracking and image analysis.

The unit is used regularly by researchers from CRG and UPF and additionally by researchers from other PRBB institutes. Applications range from immunofluorescence imaging of fixed samples to timelapse observations spanning several days.

FACILITY OVERVIEW

As in the years before, the Advanced Light Microscopy Unit continued in 2011 to provide instrumentation at the forefront of imaging technology through the upgrade of the STED super-resolution microscope with two Hybrid detectors that represent the latest in sensitive detection technology and a custom incubation solution for *in-vivo* imaging that makes the system unique in Spain and one of only few such systems worldwide.

The unit also extended the functionality of its confocal High Content Screening software HCS A from Leica by adding a module for computer-aided microscopy (CAM) to perform automated real-time decision making by the computer during the microscope experiment. The module was acquired for SysteMTB, a FP7 consortium of research groups working on Tuberculosis which is coordinated by Luis Serrano and will be used for the selection of suitable specimens for the analysis of the intracellular localization of all mycobacterial proteins.

Among the nine available imaging systems of the unit, no two are identical in their features. Because of this, a wide range of microscopy applications can be covered. However, most applications can be performed on multiple systems. This redundancy ensures that experiment planning is not impaired by the limited availability of a single system. Reflecting the variety in available instrumentation, the experiments performed in the year ranged from *in-vivo* timelapse experiments spanning several days to the high-resolution 3D imaging of multiple intranuclear components.

The total booked microscope usage time of the joint unit in 2011 reached 17300 hours corresponding to seven hours of daily usage on the seven bookable microscope systems plus many additional hours on equipment without mandatory booking. The usage has stabilized at the levels reached in 2010, reflecting the continuing high need for light microscopy by CRG, UPF and PRBB researchers. During the year, 130 users from 24 CRG research groups and 39 users from 12 UPF groups have used the unit. Additionally the unit was used by 26 users from the other PRBB institutes IMIM and IBE. On average, 84 investigators used the unit every month.

During the year, the ALMU staff has participated in teaching masters courses of the Universitat Pompeu Fabra (UPF), as well as in microscopy and image processing courses in other Barcelona universities. They have continued to participate as speakers and instructors in courses, workshops and seminars at institutes in Spain and internationally.

In June the ALMU co-hosted with Leica a super-resolution microscopy two-day workshop both on STED as well as on the latest released super-resolution method, Ground State Depletion (GSD). For this purpose, a GSD microscope was for the first time ever set up in Spain and was available for testing by CRG users for a week. Together with Doris Meder and Renza Roncarati, Timo Zimmermann, co-organised in November a technology symposium on Advanced Light Microscopy and Screening topics that brought together three international high-level speakers from that field from the Netherlands and Germany.

The ALMU is participating in EuroBioImaging, an initiative of the biomedical imaging field that is currently in the preparatory phase of the ESFRI roadmap. Timo Zimmermann is currently the national contact person for Spain for biological imaging in this initiative and involved in the workgroups for general and special access. He is very actively involved in the formation of the Spanish Advanced Light Microscopy Network (Red Española de Microscopía Óptica Avanzada, REMOA).

Together with the Institute for Photonic Sciences (ICFO) in Castelldefels, CRG has formed an alliance for super-resolution light nanoscopy that will provide access for Spanish researchers to this new technology. Together, the two institutes provide the only two currently available STED systems in Spain, as well as a range of other super-resolution microscopy methods that are currently unmatched in Spain and are in the process of being even further extended. This initiative is supported by MICINN European Infrastructure grants AIC10-A-000513 and AIC-A-2011-0827.

FIGURES

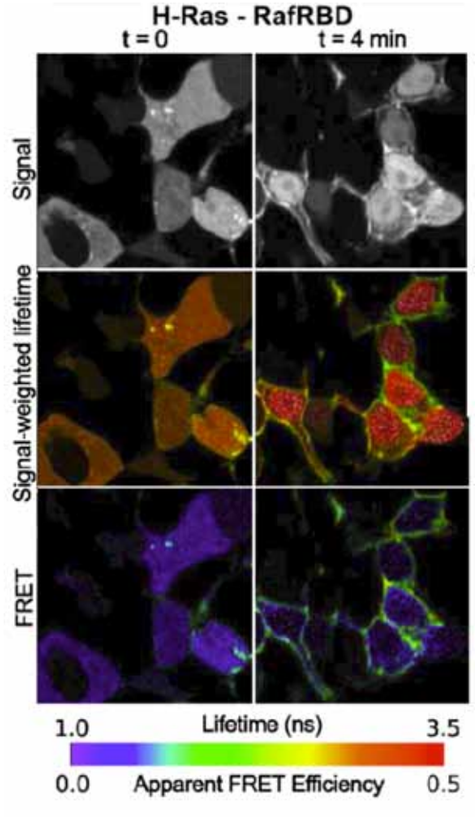


Figure 1: *Time-correlated single photon counting (TCSPC) Fluorescence Lifetime measurement of living cells expressing Citrine and mCherry tagged Ras and Raf domains before and after stimulation with EGF. Top row: Intensity image of Citrine-tagged H-Ras. Middle row: The fluorescence lifetime of Citrine H-Ras at the membrane changes significantly upon stimulation with EGF. Bottom row: The change in lifetime at the membrane is due to localized interactions of the Citrine H-Ras with an mCherry-tagged Raf domain which leads to Fluorescence Resonance Energy Transfer between the two fluorescent tags. Data by Raik Gruenberg and Julia Burnier*

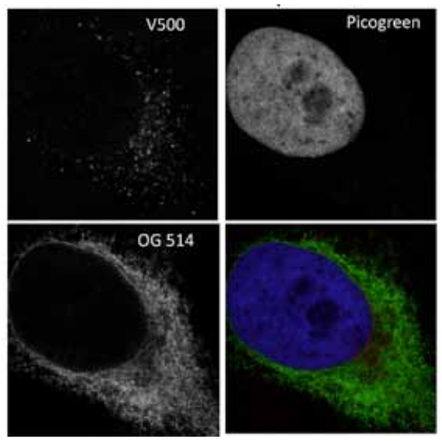


Figure 2: *A three fluorophore super-resolution image of Sec 31 (labeled with Horizon BD V500), DNA (labeled with Picogreen) and Calreticulin (labeled with Oregon Green 514) taken with Continuous Wave Stimulated Emission Depletion (CW-STED). All three selected fluorophores can be emission depleted using the same strong orange laser and thus provide details below the optical resolution limit. Due to the single depletion line requirement, the three emission spectra have significant overlap, but are distinct enough to be clearly separated after post-processing.*

PUBLICATIONS

von Blume J, Alleaume AM, Cantero-Recasens G, Curwin A, Carreras-Sureda A, Zimmermann T, van Galen J, Wakana Y, Valverde MA, Malhotra V.
"ADF/cofilin regulates secretory cargo sorting at the TGN via the Ca2+ ATPase SPCA1."
Dev Cell, 20(5):652-62 (2011).

CORE FACILITIES

Biomolecular Screening and Protein Technologies Unit

Unit structure:
Head of the Unit: Renza Roncarati

Technicians: Miriam Alloza, Carlo Carolis, Raul Gomez, Silvia Speroni

SUMMARY

The scope of the Biomolecular Screening & Protein Technologies unit (BMS&PT) is to provide researchers at CRG/PRBB and external institutions with a state-of-the-art technology platform and broad knowledge and expertise to perform medium to high-throughput RNA interference assays, chemical screenings, protein engineering and biophysical characterization of proteins and nucleic acids.

The unit performs biochemical, cellular and image-based high content screening assays. Part of the available equipment is endowed with various degrees of automation and utilized for a variety of applications comprising the preparation of expression vectors, screening of diverse expression conditions and protein purification on a milligram scale.

MAJOR PROJECTS AND ACHIEVEMENTS

In 2011 the High Throughput Screening Unit was merged with the Protein Services Unit, and named “Biomolecular Screening & Protein Technologies Unit”. This fusion has strengthened an already existing collaboration, and provides synergies and the basis for new technology developments and for better serving many different end-users.

In 2011 the Unit has been working on two screening projects in collaboration with CRG laboratories.

In the first project, several automated liquid handling steps comprising media and bacteria manipulations were optimized to perform a screening aimed at identifying genes involved in lineage reprogramming in *C. elegans* (laboratory of Ben Lehner). The primary screening was completed and the automated protocols are now being used routinely for result validation and other experiments.

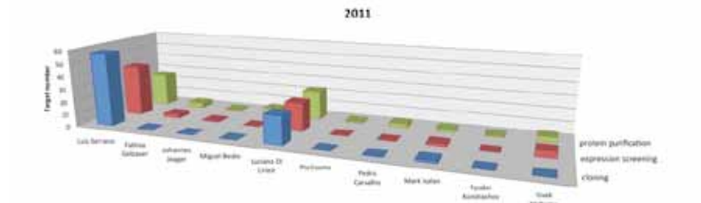
The aim of the second screening project was the identification of small molecule compounds able to induce human leukemic B cell transdifferentiation into macrophages (laboratory of Thomas Graf). The developed assay system employs cells expressing a differentiation-driven reporter gene in conjunction with live-cell imaging. Assay optimization was initiated to define the optimal assay conditions in terms of kinetics of gene expression, most suitable image analysis mode and assay reproducibility. Currently, final assay validation is ongoing.

In order to establish a compound library for this and other screening projects in the future, small focused libraries and diversity collections of tool compounds with described function were purchased comprising the Tocris kinase inhibitors toolbox and the Prestwick Chemical Library®. The latter contains 100% FDA approved drugs that were selected for their high chemical and pharmacological diversity as well as for their known bioavailability and safety in humans. This collection will be expanded by addition of further compounds selected through a ligand–based *in silico* screening approach performed by Chemotargets S.L. (Jordi Mestres) around pathways of interest. The current compound collection was established with financial contribution from the CRG Technology Transfer Department within the framework of a valorization program. As of today, 2039 compounds are available for screening.

Small organic molecule libraries	N cmpds
#3514 Tocris kinase inhibitors toolbox	80
The Prestwick Chemical Library®	1200
Prestwick Phytochemical Library®	320
“Chemotargets” S.L. virtual-screening focused library	1083

The BMS&PT Unit has developed and implemented a high throughput pipeline for the production of proteins in *E. coli* which encompasses: (i) multi-construct design, (ii) ligation-free cloning into customized expression vectors, (iii) semi-automated small-scale screening for soluble expression, (iv) scale-up and semi-automated purification. Application of this process significantly reduces the lead-time for the identification of soluble proteins and enables the parallel processing of multiple protein targets.

In 2011, the BMS&PT Unit has performed more than 20 projects in collaboration with 10 different CRG laboratories. Most of the experiments aimed at determining optimal expression conditions and purifying target proteins for functional assays and structural analysis and/or biophysical characterization as well as for antibody production or antibody purification.



	Luis Serrano	Fatima Gebauer	Johannes Jaeger	Miguel Beato	Luciano Di Croce	Pia Cosma	Pedro Carvalho	Mark Isalan	Fyodor Kondrashov	Vivek Malhotra
cloning	57	0	0	0	22	0	0	2	0	0
expression screening	40	3	1	0	22	0	0	2	0	5
protein purification	26	4	1	5	23	1	3	2	1	4

One of these projects is part of the European FP7 project PROSPECTS (Proteomics in Time and Space), in which the Serrano lab is partner. The overall aim is to quantitatively characterize the Wnt signal transduction pathway. Currently, 44 proteins have been shown to be involved in the Wnt pathway. The BMS&PT Unit has cloned the corresponding genes. All proteins successfully expressed will be purified and used by the Serrano team to determine absolute protein concentrations in different cell lines.

A second European FP7 project in which the Unit is involved is SysteMTb (Systems Biology of *Mycobacterium tuberculosis*). Here, the main work of the facility has been to develop a high-throughput cloning method for generating the fluorescent-tagged Mtb-ORFeome library for subsequent systematic protein localization studies in Bacillus Calmette-Guérin.

In 2011 the Unit has also set up and optimized a series of automated protocols that are provided as regular services, including immunofluorescence staining, plasmid miniprep preparation, “sample normalization” for normalizing raw nucleic acid and protein samples to a specified concentration, and a protocol for dispensing, transfer and reformatting EasyClone library plates.

CORE FACILITIES

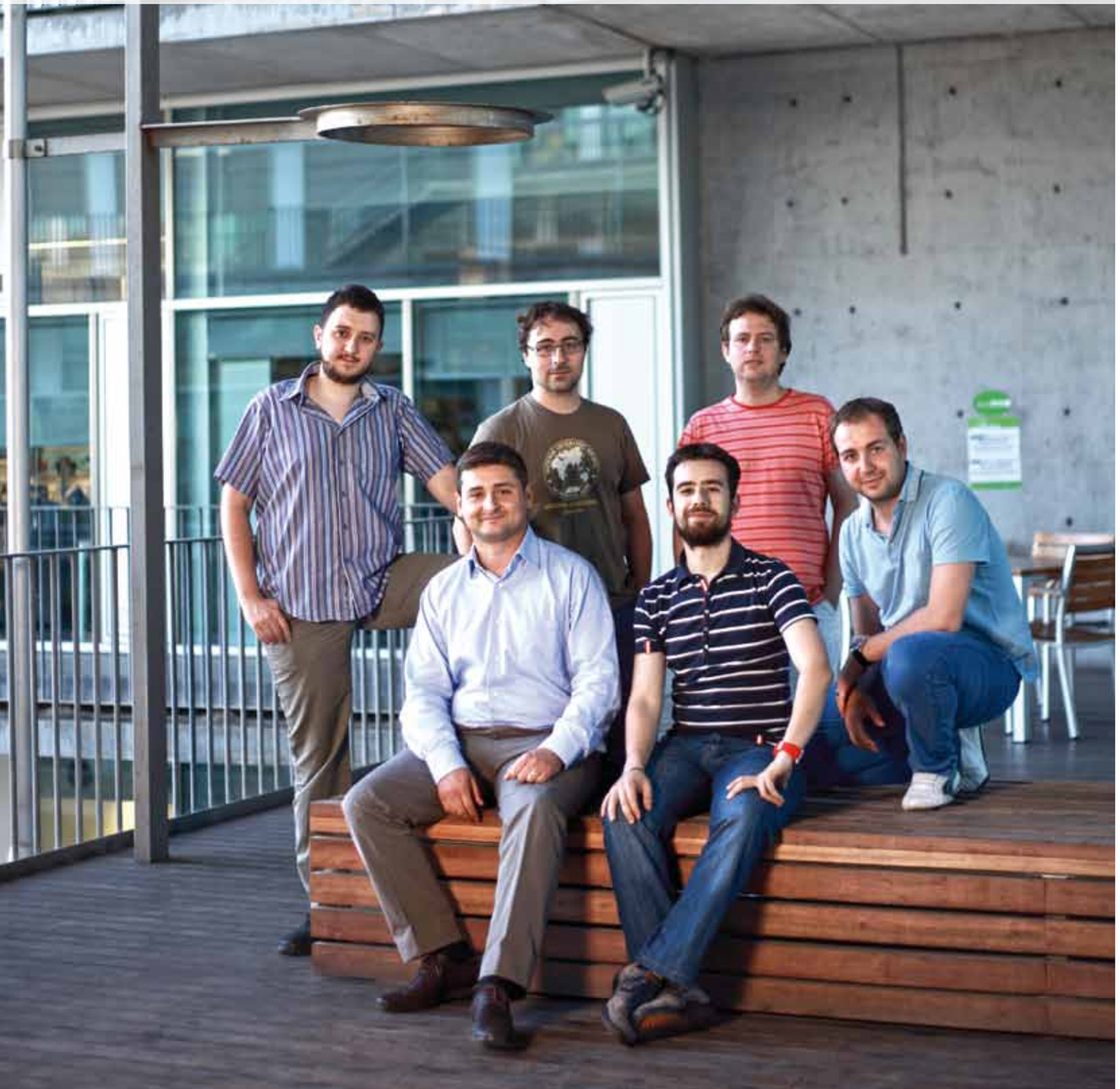
Bioinformatics Unit

Unit Structure:

Head of the Unit: Guglielmo Roma (until July 2011)

Acting Manager of the Unit: Ernesto Lowy (since July 2011)

Bioinformaticians: Ernesto Lowy, Francesco Mancuso, Luca Cozzuto



SUMMARY

The Bioinformatics core provides expertise in bioinformatics, statistics, data analysis, and scientific software development to support CRG, PRBB, and external research groups. The team periodically organizes courses to train biologists on the use of bioinformatics resources.

FACILITY OVERVIEW AND SERVICES

After its creation in September 2009 and consolidation in 2010, 2011 has been a year with some changes in the unit. The former head of the unit (Guglielmo Roma) left us in July 2011 and since then Ernesto Lowy has been managing the unit. Furthermore, another member of the facility (Andreu Alibes) left us as well. A bionformatician position has been announced and will be filled during April 2012.

During this year, the team has contributed to several research projects by providing support and expertise in programming and advanced data analysis, focusing primarily on high-throughput genomics (including microarrays, genotyping, and next-generation sequencing) and proteomics technologies. All methodologies developed and implemented in collaboration with the colleagues from other core facilities and research groups are therefore offered as data analysis services to new users.

This last year we have continued getting requests for advanced data analysis of Microarrays data coming from different platforms such as Affymetrix, Agilent, and Illumina. Statistical treatment of the data was done with R and various packages from Bioconductor. The analysis workflow consists of data preprocessing, normalization, and statistical inference. Then, identified gene lists can be further investigated by applying various downstream methods to interpret the biological significance, including gene set enrichment, over-representation of biological functions using ontology terms, and pathways analysis. Pattern discovery tools (*de novo* and not) are used to look for sequence motifs shared in upstream regions of co-regulated genes. Apart from the previously described type of analyses, the unit has gained expertise in the analysis of SNP and CGH array data. Both technologies have been used to genotype and study the structural variation in different individuals and required from the unit to develop a set of R scripts to detect Duplications/Deletions and regions of LOH (loss of heterozygosity). Additionally, during 2011 our unit had the opportunity of performing a reverse engineering study in *E. coli*. This study analyzed the effect of creating different combinations of promoters and genes on the biological networks of *E. coli* and required us to statistically treat the data in a more sophisticated way using the holistic approach characterizing the Systems Biology field.

Other very popular request among users of the facility is the analysis of Next-Gen sequencing data. Along 2011, we have analyzed a wide range of data, including ChIP-seq, RNA-seq, resequencing for polymorphism detection and De-novo genome and transcriptome assembly data. In order to provide a better service we have improved the existing pipelines and created new pipelines for some of the analyses. For example, the ChIP-seq pipeline has been modified to include new methods of visualizing the enrichment around the TSS (transcription start site), new algorithms to detect broader peaks of enrichment that are characteristic of Histones modifications and RNApol occupancy studies and finally, new ways of integrating the results produced across different immunoprecipitation experiments. Furthermore, we have developed two different pipelines to generate De-novo genome and transcriptome assemblies. Both of them include a quite sophisticated filtering procedure to remove sequencing artifacts and possible contaminants, this has improved greatly the quality of the assemblies produced by the pipeline. The pipeline for the De-novo transcriptome assembly has been successfully applied, in collaboration with the Bioinformatics & Genomics program, to analyze the RNA-seq data generated within the framework of the Spanish collaborative effort to produce a draft assembly and annotation of the *Lynx pardinus* species (Iberian Lynx). Apart from the previously mentioned pipelines, we have put a great effort to establish a standard analysis procedure to functionally annotate the protein coding genes coming from projects dedicated to sequence new genomes. This pipeline is specially designed to identify disease resistance genes in plants but is also capable of correctly annotating a certain protein with Gene Ontology (GO) information, metabolic pathway, reaction information and it is also capable of identifying the different domains at the protein level. This pipeline has been successfully used in the Melonomics project, which is a project funded by Genoma España, several private companies and five regional governments whose main focus is producing a draft sequence of the melon genome and its annotation. Additionally, the unit has also contributed to the Melonomics project by providing a comprehensive annotation of the non-coding RNAs present in the Melon genome.

Additionally, the unit has continued collaborating with the proteomics unit to investigate different extraction tools, giving important results that are relevant for the proteomics field. The main goal of this study is to highlight that extraction tools do differ and some of the quality metrics used in the project exemplify why it is important that this first step of any proteomics data analysis would be considered.

Since 2010 and after the evaluation of several commercially available laboratory management tools, the unit started to work on the adoption and deployment of easy-to-use Wikipedia-like systems intended for an accurate tracking and management of the different laboratory processes in the core facilities and other internal services. During 2011 this project has continued with the incorporation of the CRG Microarray core facility, which has joined the project and is working in close collaboration with us. Additionally, we completed the Biomolecular Screening & Protein Technologies Unit data management wiki-based system, this tool has become an important asset for the facility and it is being routinely used to control the different laboratory workflows. Finally, the bioinformatics wiki-based system was also improved in the last year, incorporating a useful feature to track the status of the different user requests.

Another important activity of the unit during the last year has consisted on the development of customized software solutions for the different research groups at the CRG. We have been working in the design and implementation of scientific databases, apart from developing different methods to display the data present in these databases. As an example, the unit is involved in the implementation of a web front-end and a relational database back-end system that will contain more than 15,000 different fly embryo images. This system will enable researchers to manage, display and integrate the experimental data present in the database.

Finally, it does worth to mention that during 2011 the unit has been heavily involved in different training activities. We have organized several technical seminars to train researchers on the use of different bioinformatics tools and analysis. Besides, in 2011 the unit organized a multisession hands-on Perl programming course.

PUBLICATIONS

Mancuso FM, Montfort M, Carreras A, Alibes A, Roma G.
"HumMeth27QCReport: an R package for quality control and primary analysis of Illumina Infinium methylation data."
BMC Research Notes, 4:546 (2011).

Mihailovich M, Wurth L, Zambelli F, Abaza I, Militti C, Mancuso FM, Roma G, Pavesi G, Gebauer F.
"Widespread generation of alternative UTRs contributes to sex-specific RNA binding by UNR."
RNA, Epub 2011 Nov 18.

Uribesalgo I, Buschbeck M, Gutiérrez A, Teichmann S, Demajo S, Kuebler B, Nomdedéu JF, Martín-Caballero J, Roma G, Benitah SA, Di Croce L.
"E-box-independent regulation of transcription and differentiation by MYC."
Nat Cell Biol, 13(12):1443-9 (2011).

Luis NM, Morey L, Mejetta S, Pascual G, Janich P, Kuebler B, Cozzuto L, Roma G, Nascimento E, Frye M, Di Croce L, Benitah SA.
"Regulation of Human Epidermal Stem Cell Proliferation and Senescence Requires Polycomb- Dependent and -Independent Functions of Cbx4."
Cell Stem Cell, 9(3):233-46 (2001).
(Erratum in Cell Stem Cell, Volume 9, Issue 5, 486, 4 November 2011).

Hummel M, Bonnin S, Lowy E, Roma G.
"TEQC: an R-package for quality control in target capture experiments."
Bioinformatics, 27(9):1316-7 (2011).

CORE FACILITIES

CRG/UPF Flow Cytometry (FACS) Unit

Unit Structure:
Head of the Unit: Òscar Fornas (UPF)

Operators: Erika Ramírez (CRG), Sabrina Bascones (IMIM)

SUMMARY

Flow Cytometry is a technology that day to day offers more possibilities in cell studies and is becoming an essential technology for biomedical research. The Joint CRG/UPF Flow Cytometry Facility offers state of the art instrumentation that enables researchers to perform a wide range of applications, from simple to multicolour flow analyses.

The facility was established by Universitat Pompeu Fabra (UPF) in 2001, and operates as a Joint CRG/UPF Facility since September 2009. Until September 2009 the UPF facility had 4 instruments and 1 operator/Unit Head. With the CRG/UPF agreement 1 additional operator and 3 additional instruments were incorporated into the facility, drastically increasing the service capacity.

The following instruments are available in the facility:

Analyzers:

FACScan (UPF)	(1 laser and 3 fluorescences)
FACScalibur (UPF)	(2 lasers and 3+2 fluorescences)
LSR I (UPF)	(3 lasers and 3+2+2 fluorescences)
FACScanto (CRG)	(2 lasers and 4+2 fluorescences)
LSR II (CRG)	(4 lasers and 4+2+2+2 fluorescences)

Cell sorters:

FACSvantage (UPF)	(3 lasers and 3+2+2 fluorescences)
FACSAria (CRG)	(5 lasers and 5+5+3+3+2 fluorescences)

In January 2011 a third operator, supported by IMIM, was incorporated. This has been a crucial improvement of the user support. Since the two operators work in two overlapping shifts, the facility could extend the service hours into the late evening. The facility now offers assistance with experimental design, training on the analysers, and cell sorting services during 12 hours per day, from 8:30 to 20:30. Consequently, the usage hours have gone up by 18%, from 4427 in 2010 to 5242 in 2011. In total, the facility had 160 users in 2011.

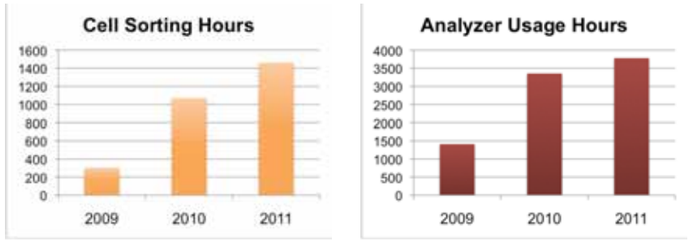
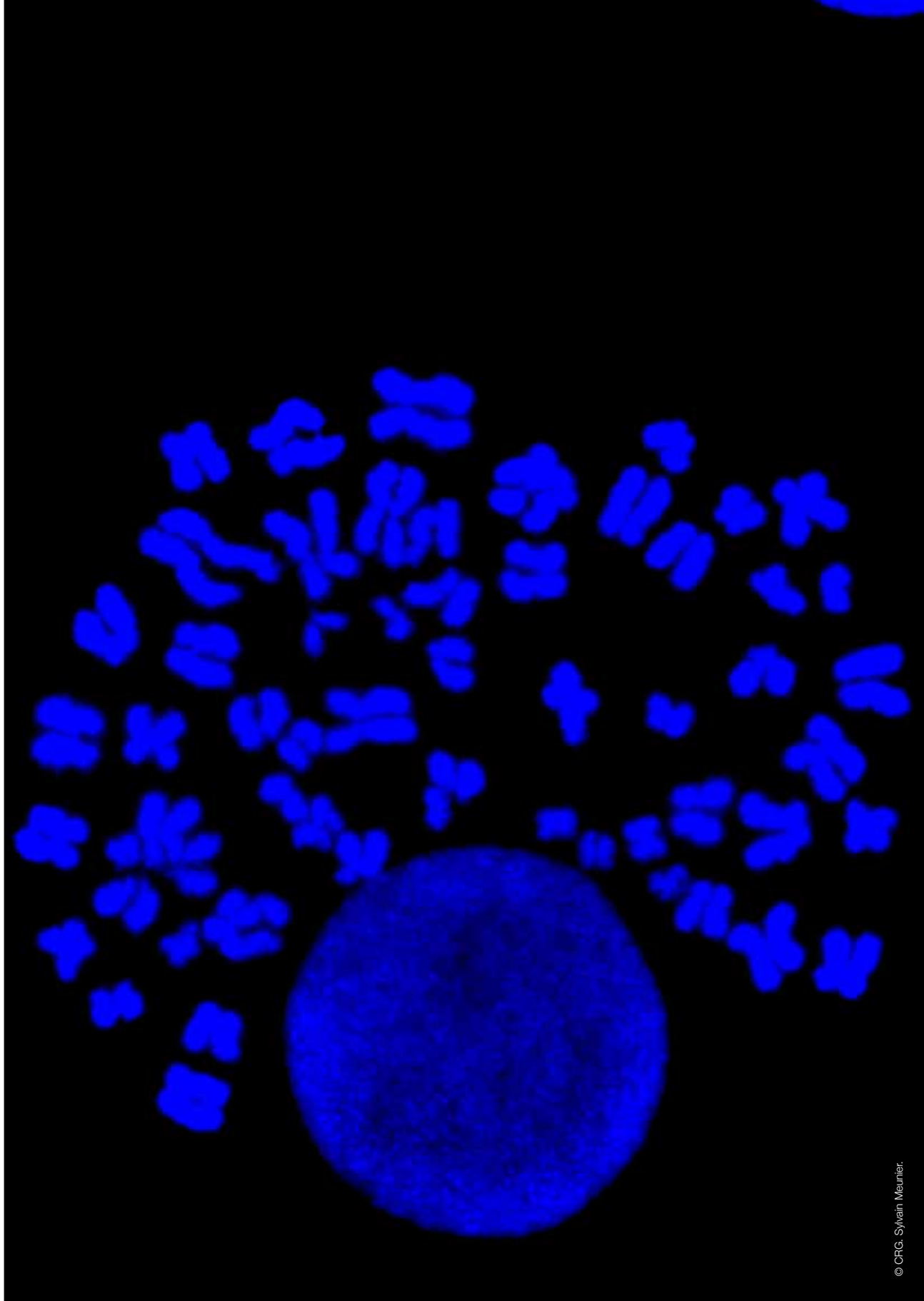


Figure 1:
Usage statistics of the Joint CRG/UPF Flow Cytometry Unit

During 2011 the head of the facility has participated in master teaching courses of the UPF as well as in flow cytometry courses at other institutions in Barcelona. He also participated as a tribunal member of 3 "*Tesinas*" in the Master of *Tropical Medicine and International Health* at the Universitat Autònoma de Barcelona (UAB).



CORE FACILITIES

In-House Services: Histology Service

Service structure:

Service Manager: Salvador Aznar Benitah PhD, ICREA Researcher, PI of the Epithelial and Homeostasis Laboratory, Dept. Differentiation and Cancer (CRG)

Staff Scientists: Alexis Rafols Mitjans, Senior Technician, FP (CRG), Marina Nuñez Alfonso, FP (CRG)

Associates: Juan Martin Caballero PhD, Veterinarian, Head and of the Animal Facility of the PRBB
Juana M. Flores, Professor of Pathological Anatomy and Director of the Department of Medicine and Surgery, Faculty of Veterinary, Madrid's Complutense University

SUMMARY

The Histology Unit was established to provide the CRG with the histopathological analysis required when analyzing tissues, and research based on *in vivo* models. Studies involving developmental biology, mechanisms of homeostasis, and pathology, require an in depth analysis of the morphological and molecular characteristics of the tissue under study. A common tool for these studies is to generate animal models that may recapitulate a specific process or a particular disease. The histological and immunohistological analysis are required to establish changes in the citoarchitecture, morphology of the tissue, and the specific molecular modifications that cause and accompany these. Immunohistochemistry is an essential tool to determine the localization of proteins and RNAs at the tissue and cellular level. **To this respect, the main goal of the Histology Unit is to perform and optimize histological processing and analysis of the tissue, from experimental animal models and human origin**. The unit also provides mentoring and training in common histological techniques, in close association with the different groups at the CRG.

SERVICES

The aim of this unit is to provide researchers with assistance in the histological processing and analyzing of samples derived from *in vivo* models. The unit centralizes and performs all the histological analysis of the CRG and works in close association to the Animal Unit of the PRBB (headed by Dr. Juan Martin Caballero). It provides the following services:

> Preparation of paraffin embedded samples.
> Sectioning of paraffin embedded, fresh and frozen samples.
> Common histological stainings.
> Full post-mortem histopathological analysis of the mouse models.

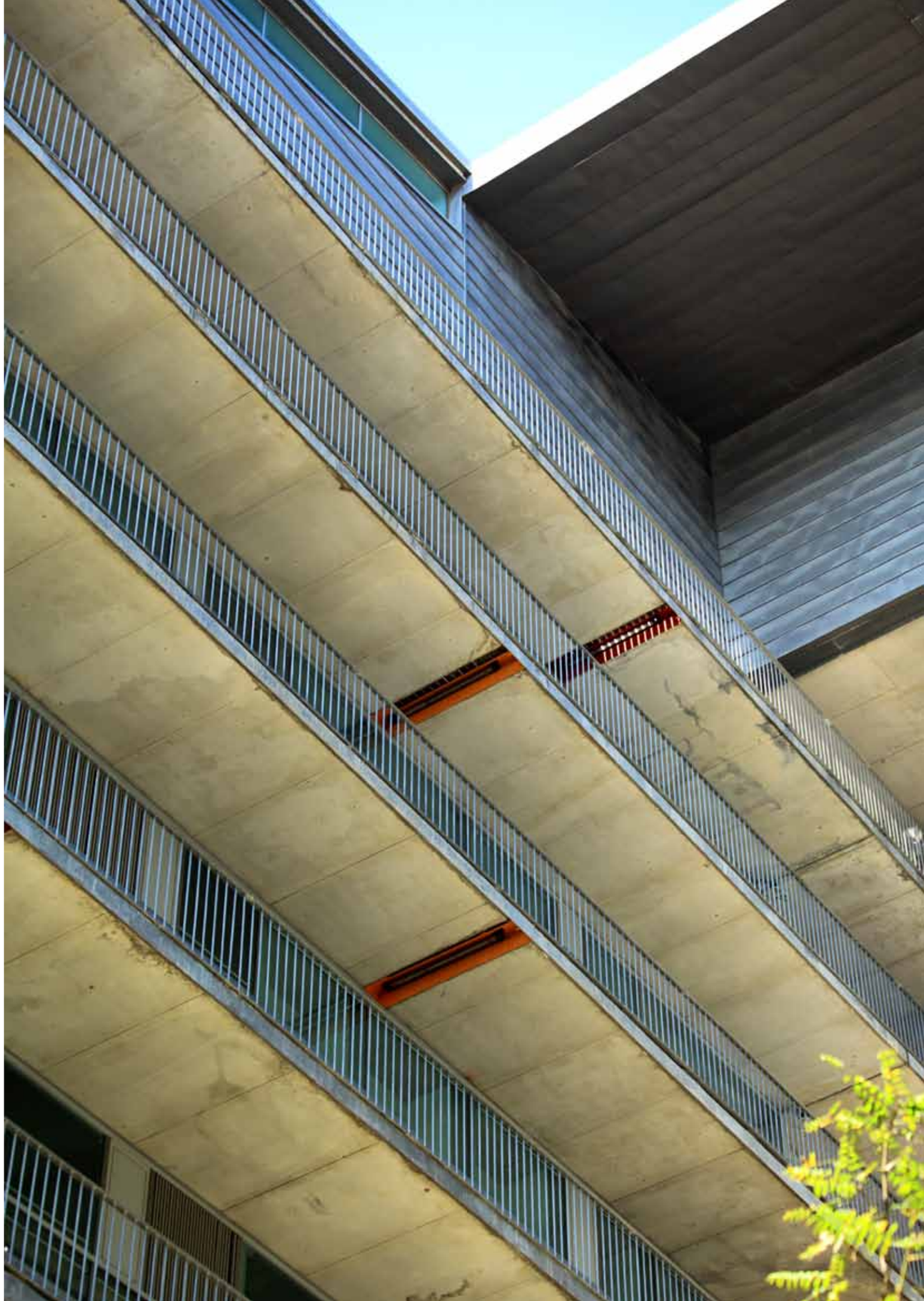
The unit also provides researchers with the necessary equiment and material to perform their own histological preparations, processing and analysis, as well as with experimental planning and training. When required, custom analysis routines will be designed.

The full Postmortem Histopathological analysis of mouse models is provided externally by Dr. Juana M. Flores (Professor of Pathological Anatomy and Director of the Department of Medicine and Surgery, Faculty of Veterinary, Madrid's Complutense University), in association with the CRG.

EQUIPMENT

> Two cryostats: Sectioning of frozen tissue blocks.
> Two microtomes: Sectioning of paraffin blocks.
> Two vibratomes: Sectioning of fresh tissue.
> Two histological water baths: For paraffin embedding and sectioning.
> Paraffin dispenser, hot and cold plate: Paraffin embedding.
> Autostainer (Leica): Performs histological staining of frozen and paraffin sections. Can perform over 250 stainings per day in an automated manner.
> Tissue Processor (Leica): Automated paraffin embedding. Provides the unit with the potential to embed 100 blocks of paraffin a day.
> Shaker and Precision balance.
> One Olympus BX51 microscope and an Olympus DP70 digital camera. The microscope incorporates two softwares, Neurolucida-mbf bioscience, MicroBrightField, Inc.; and CAST-Olympus which allows to study cellular morphology and 3D reconstruction of tissue by sterological analysis.
> Movable Fumehood.

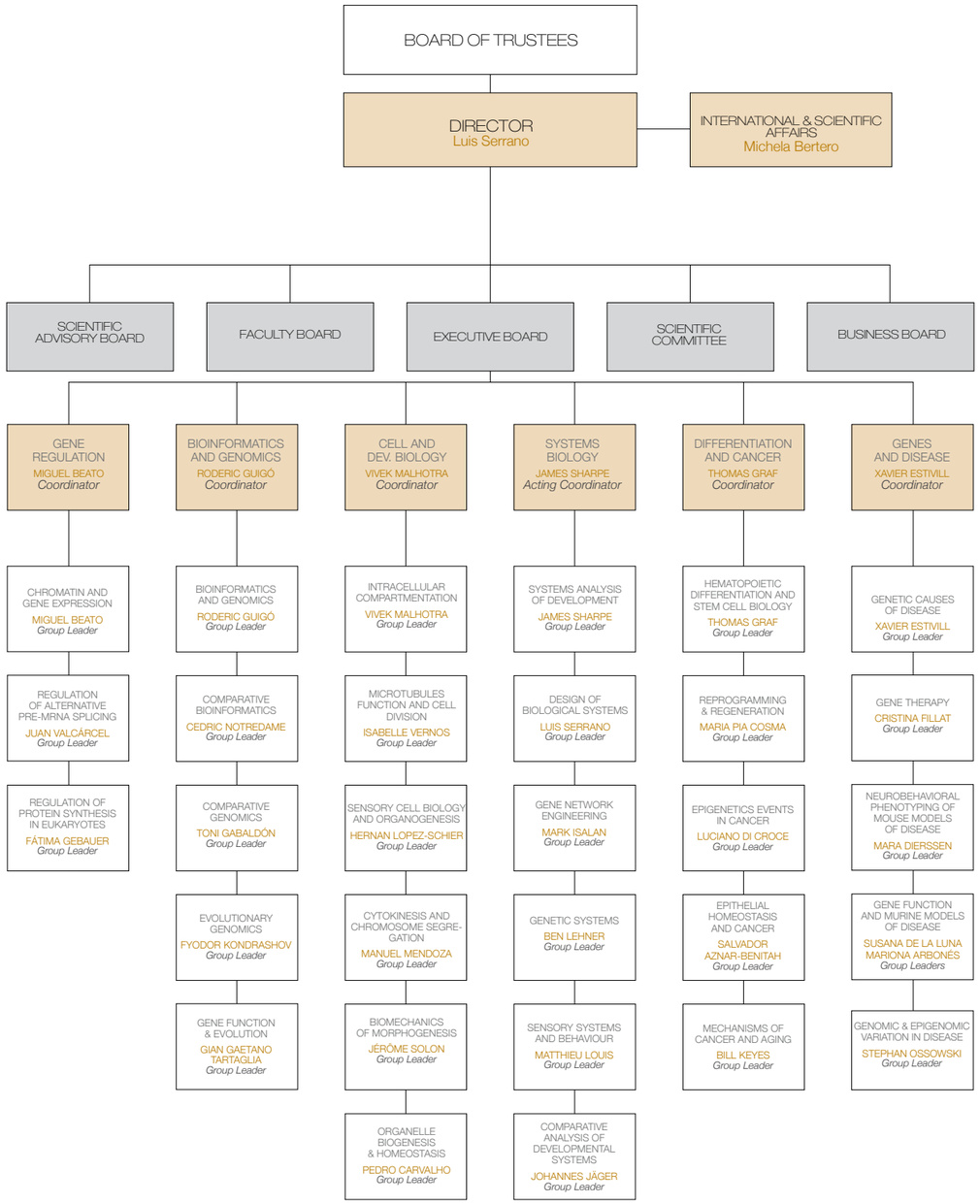
Most of the tissue processing equipment is duplicated to enable both technicians to work simultaneously. This has allowed us to greatly speed up the time required for finishing a requested service, thereby reducing waiting times and waiting list. Currently, the average time required for a medium sized service does not exceed four days.



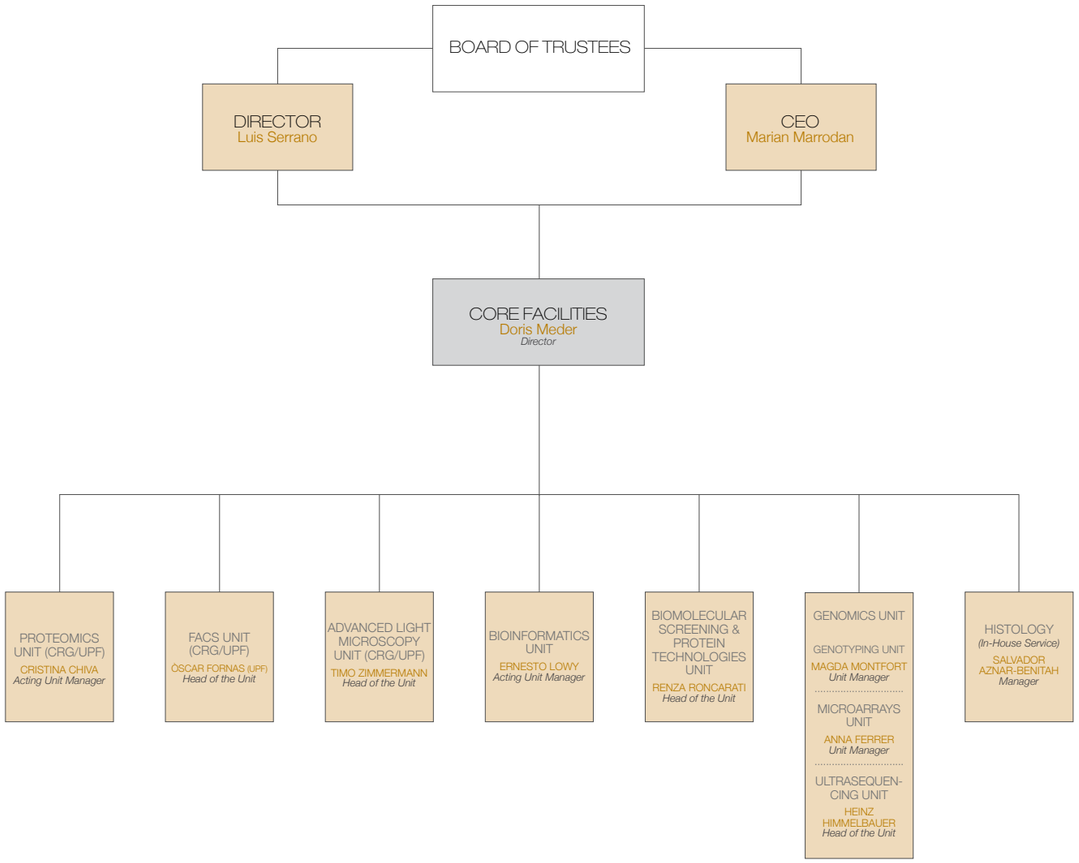
Appendix 1

CRG STRUCTURE

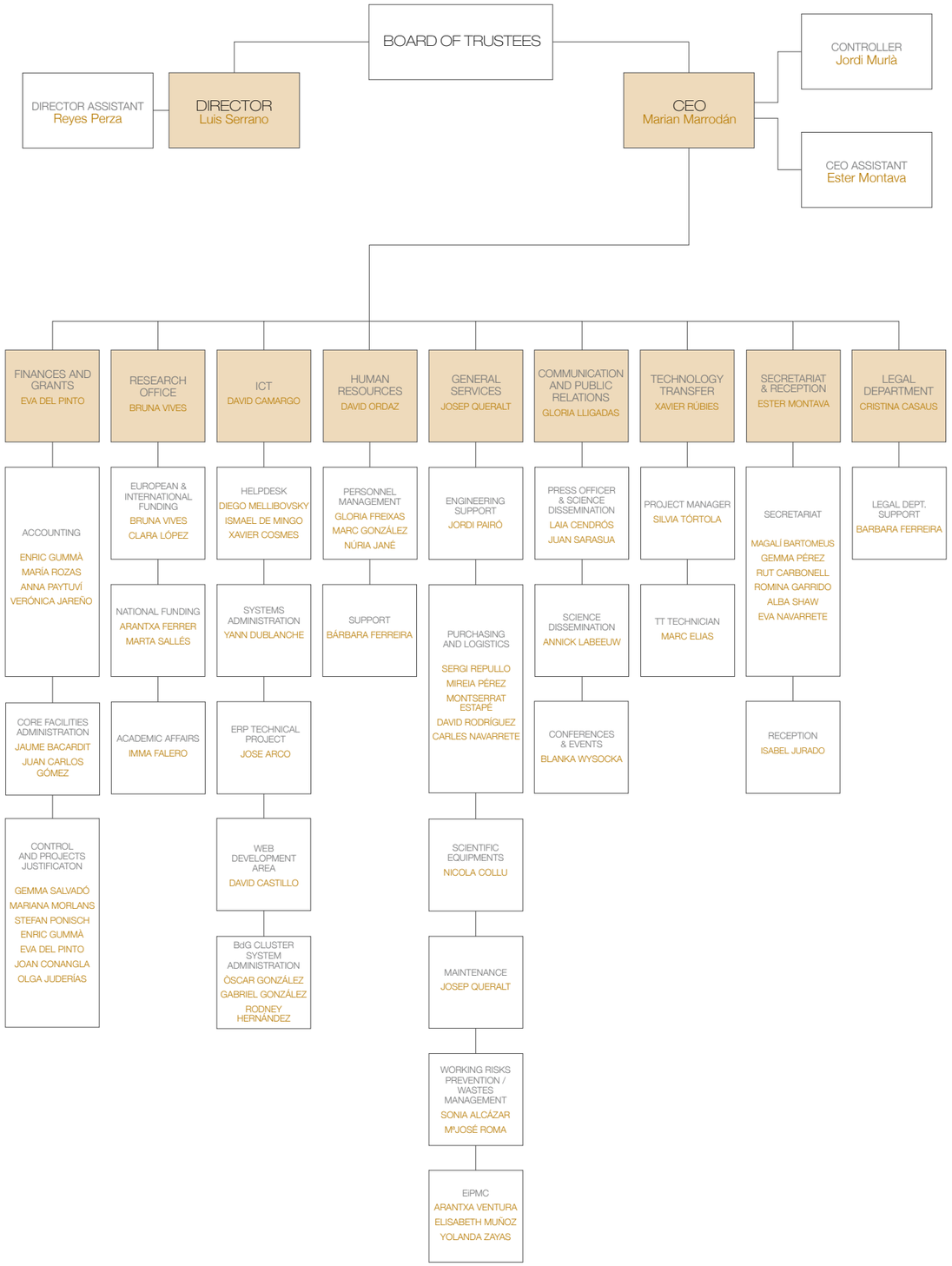
SCIENTIFIC STRUCTURE



CORE FACILITIES STRUCTURE



MANAGEMENT & RESEARCH SUPPORT STRUCTURE



SCIENTIFIC ADVISORY BOARD (SAB)

Chairman

Prof. Tariq Enver
UCL Cancer Institute,
London, United Kingdom

Members

Prof. Peer Bork
Structural & Computational Biology Unit,
European Molecular Biology Laboratory (EMBL),
Heidelberg, Germany

Prof. Scott D. Emr
Weill Institute for Cell and Molecular Biology,
Cornell University,
Ithaca, USA

Prof. Tim Hunt
Cancer Research UK,
Hearts, United Kingdom

Prof. Walter Kolch
Systems Biology Ireland & Conway Institute,
University College,
Dublin, Ireland

Prof. Tony Kouzarides
The Wellcome Trust/Cancer Research
UK Gurdon Institute,
Cambridge, United Kingdom

Prof. Iain Mattaj
European Molecular Biology Laboratory (EMBL),
Heidelberg, Germany

Prof. Angela Nieto
Institute for Neurosciences, CSIC-UMH,
Alicante, Spain

Prof. Laura P.W. Ranum
Center for NeuroGenetics,
University of Florida
Gainesville, USA

Prof. Wolf Reik
Babraham Institute Cambridge,
Department of Physiology,
Development & Neuroscience,
University of Cambridge, United Kingdom

Prof. Anna Tramontano
Department of Physics,
Sapienza University of Rome,
Rome, Italy

Prof. Veronica van Heyningen
Medical Research Council,
Human Genetics Unit,
Western General Hospital,
Edinburgh, United Kingdom

Prof. Marc Vidal
Dana Farber Cancer Institute,
Boston, USA

BUSINESS BOARD

President

Dr. Josep M. Taboada
Medical Director,
Sanofi-Aventis

Vice-president

Dr. Antonio Parente
President,
Grup Lipotec

Members

Mr. Ramon Aymerich
Economics Editor in Chief,
La Vanguardia

Mr. Josep Lluís Sanfeliu
Co-founder partner,
Ysios Capital Partners

Dr. Pere Berga
R+D Management Director,
Almirall Prodesfarma, S.A.

Sr. Fernando Turró
General Manager,
Contratas y Obras

Dr. Clara Campàs
Executive Vice-President,
Advancell Therapeutics

Dr. Luis Vives
Strategy Professor,
ESADE Business School

Dr. Antoni Esteve
President,
Esteve

Dr. Josep Prous, Jr.
Executive Vice President,
Prous Science

Sr. Jordi Ramentol
General Manager,
Grupo Ferrer Internacional, S.A.

Appendix 2

PRBB-CRG SESSIONS (2011)

16-12-11 Stavros J. Hamodrakas
Dept. of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Panepistimiopolis, Athens, Greece
"Revealing the code of a protective amyloid"

02-12-11 Francis Stewart
Epigenetic regulation and genomic engineering, Biotechnology Center TU Dresden (Biotec), Dresden, Germany
"H3K4 methylation in mouse development"

22-11-11 Edward Marcotte
Center for Systems and Synthetic Biology, Dept of Chemistry and Biochemistry, Inst. for Cellular and Molecular Biology, Univ. of Texas at Austin, USA
"Insights from proteomics into cellular evolution and surprising disease models"

18-11-11 Alfonso Martinez-Arias
Department of Genetics, University of Cambridge, United Kingdom
"Transition states and ß-catenin signalling in self renewal and differentiation of Stem Cell populations"

16-11-11 Dinshaw Patel
Structural Biology, Memorial Sloan-Kettering Cancer Center, New York, USA
"Structural Biology of Methylation-mediated Epigenetic Regulation"

04-11-11 Pavel Tomancak
Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany
"Patterns of gene expression in development and evolution"

21-10-11 Cedric Blanpain
Interdisciplinary Research Institute (IRIBHM), Université Libre de Bruxelles, Brussels, Belgium
"Studying stem cells during cancer initiation and growth"

14-10-11 Alex Stark
Research Institute of Molecular Pathology (IMP), Vienna, Austria
"Regulatory Genomics in Drosophila"

07-10-11 Tarun Kapoor
Laboratory of Chemistry and Cell Biology, The Rockefeller University, New York, USA
"Examining mechanisms required for error-free cell division"

16-09-11 Freda Miller
Developmental & Stem Cell Biology, Hospital for Sick Children Research Institute and University of Toronto, Canada
"Neural stem cells: from development to repair"

09-09-11 Rui Costa
Neurobiology of Action, Champalimaud Neuroscience Program, Instituto Gulbenkian de Ciencia, Oeiras, Portugal
"Learning and organizing new action repertoires"

01-07-11 Jean-Michel Claverie
Structural and Genomic Information, CNRS;
Institut de Microbiologie de la Méditerranée, Aix-Marseille Université, Marseille, France
"Mimivirus and other Megaviridae: freaks of nature or key players in Evolution?"

17-06-11 Stefan Dimitrov
Chromatin and Epigenetics, Department of Cell Differentiation and Transformation, Institut Albert Bonniot, Grenoble, France
"The smart chromatin nanomachines: how do they work?"

03-06-11 Yosef Shiloh
Department of Human Molecular Genetics & Biochemistry, Sackler School of Medicine, Tel Aviv University, Israel
"The ATM-mediated DNA damage response: from pathways to systems - and back"

27-05-11 Wen-Hsiung Li
Director, Biodiversity Res. Center, Academia Sinica, Taiwan and James Watson Prof., Ecology & Evolution, Univ. of Chicago
"Chromatin Structure, Gene Clustering Pattern and Evolution of Gene Regulation"

06-05-11 Charlie Boone
The Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Canada
"The Genetic Landscape of the Cell"

15-04-11 Magdalena Götz
Institute for Stem Cell Research, Helmholtz Center Munich, and Physiological Genomics, University of Munich, Germany
"Neurogenesis from glial cells – cellular and molecular mechanisms"

08-04-11 Ricard Solé
ICREA Research Professor, Complex Systems Lab (UPF), Barcelona, Spain
"How to make a biological computer"

01-04-11 Jim Rothman
Department of Chemistry, Yale University, New Haven, USA
"The structural and thermodynamic basis of synchronous neurotransmitter release"

18-03-11 Daniel Hartl
Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, USA
"Evolutionary Pathways of Drug Resistance"

11-03-11 Conly Rieder
Division of Translational Medicine, Wadsworth Center, and Department of Biomedical Sciences, State University of New York, Albany, USA
"Spindle Poisons, the Mitotic Checkpoint and Mitotic Slippage"

04-03-11 Brian Link
Dept. Cell Biology, Neurobiology, Anatomy, Medical College of Wisconsin, Milwaukee, USA
"A cell biological perspective on retinal neurogenesis"

25-02-11 Patrick Phillips
Ecology and Evolutionary Department, University of Oregon, USA
"Sex, death and redemption: Using experimental evolution to test evolutionary hypotheses and to explore the structure of genetic systems"

18-02-11 Noam Sobel
Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel
"Predicting odor perception and neural activity from odorant structure"

11-02-11 Cayetano González
Cell Division Group, Institute for Research in Biomedicine Barcelona, Spain
"Modeling cancer in flies"

04-02-11 Tony Kouzarides
Gurdon Institute, Cambridge University, UK
"Chromatin modifications: their function and role in cancer" and *"Vencer el Cancer: a new Spanish foundation against cancer"*

28-01-11 Bill Bialek
Joseph Henry Lab. of Physics, Lewis-Sigler Inst. for Integrative Genomics, and Princeton Center for Theoretical Science, Princeton University, USA
"Life at its best: Optimization and the physical limits to biological function"

14-01-11 Alfonso Valencia
Programa de Biología Estructural y Biocomputación, Centro Nacional de Investigaciones Oncológicas, Madrid, Spain
"The Secret Life of Proteins, from Splicing to Cancer Networks"

PROGRAMME & CORE FACILITIES SEMINARS (2011)

SYSTEMS BIOLOGY PROGRAMME

19-12-11 Dominic GRUEN
Nikolaus Rajewsky lab, Max-Delbrück-Center, Berlin, Germany
"Evolution of gene expression during early nematode development"

13-10-11 Julio SAEX
Systems Biomedicine, European Bioinformatics Institute (EMBL-EBI), Hinxton UK & Genome Biology Unit, EMBL, Heidelberg, Germany
"Network models to understand and target signal deregulation in cancer"

05-10-11 Hernán G. GARCIA
Thomas Gregor's lab, Princeton University, Princeton, USA
"Quantitative Dissection of the Simple Repression Input-Output Function"

21-09-11 Amit ZEISEL
Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot, Israel
"Complex dynamics of transcriptional cellular response"

08-07-11 Verônica GRIENEISEN
John Innes Centre, Norwich, United Kingdom
"Feedbacks between Cell Shape and Polarity"

07-07-11 Theodore J. PERKINS
Ottawa Hospital Research Institute and Ottawa University, Ottawa, Canada
"State Sequence Inference for Jump Markov Processes: HIV Evolution, Ion Channel Dynamics, and Other Examples"

06-07-11 Nicolas BUCHLER
Institute for Genome Sciences and Policy GSP, Duke University, Durham, USA
"Building a genetic transistor in yeast: How protein sequestration generates a tunable threshold response"

05-07-11 Jörg STELLING
ETH Zurich, Zurich, Switzerland
"Biological systems analysis under uncertainty"

04-07-11 Nicolas LE NOVÈRE
EMBL-EBI, Hinxton, United Kingdom
"Decoding calcium signals involved in synaptic plasticity"

06-06-11 Christen MIRTH
Development, Evolution and the Environment, Instituto Gulbenkian de Ciência, Oeiras, Portugal
"The evolution of foraging behaviour in larvae of the genus Drosophila"

02-06-11 Jordi CASANOVA
Development and Morphogenesis in *Drosophila*, Cell and Developmental Biology, IRB Barcelona, Spain
"From genes to shape: notions from the study of the Drosophila tracheal system"

01-06-11 Olena RIABININA
The Ear Institute, University College London, London, United Kingdom
"Hearing in Drosophila: courtship songs and species-specific ear tuning"

24-05-11 Ulrich BRAUNSCHWEIG
Division of Gene Regulation (B4), Netherlands Cancer Institute, The Netherlands
"Chromatin protein targeting and principal chromatin types in Drosophila cells"

20-05-11 Angela BRAND
Institute for Public Health Genomics (IPHG), Maastricht University, Maastricht, The Netherlands
"Systems Biology for Public Health"

06-05-11 Peter PETERS
Netherlands Cancer Institute (NKI), Amsterdam, The Netherlands
"Type VII Secretion System of mycobacteria; toward visualization of nanomachines in their native cellular environment using electron cryo-microscopy"

13-04-11 Tzachi PILPEL
Computational Functional Genomics, Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel
"Genome-wide post-transcription regulation"

18-02-11 Noam SOBEL
Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel
"Sobel goes informal - The human nose: from scent tracking to chemical communication and electric wheelchair driving"

28-01-11 Bill BIALEK
Joseph Henry Lab. of Physics, Lewis-Sigler Inst. for Integrative Genomics, and Princeton Center for Theoretical Science, Princeton University, Princeton, USA
"Optimizing information transmission"

26-01-11 Bill BIALEK
Joseph Henry Lab. of Physics, Lewis-Sigler Inst. for Integrative Genomics, and Princeton Center for Theoretical Science, Princeton University, Princeton, USA
"Maximum entropy models for biological networks"

BIOINFORMATICS & GENOMICS PROGRAMME

05-12-11 Zoe DONALDSON
Robert Wood Johnson Health & Society Scholar, Columbia University, New York, USA
"The neurogenetics of sociality: examining a functional role for V1a receptor genetic diversity"

09-11-11 Alessandra VIGILANTE
European Bioinformatics Institute (EBI-EMBL), Hinxton, Cambridge, United Kingdom
"Gene duplication and Alternative splicing: genome scale analyses of two complex mechanisms"

17-10-11 Mitch GUTTMAN
Massachusetts Institute of Technology, Broad Institute of MIT and Harvard, Cambridge, USA
"Functional integration of lincRNAs in the molecular circuitry of the cell"

13-10-11 Rahul SIDDHARTHAN
The Institute of Mathematical Sciences, Chennai, India
"Evolution of centromeres in budding yeasts"

06-10-11 Aled EDWARDS
Banting and Best Department of Medical Research, The University of Toronto, Toronto, Canada
"Open access chemical biology of epigenetic signalling"

26-09-11 Meromit SINGER
Computer Science Division, University of California, Berkeley, USA
"A genomewide study of DNA methylation in primates via evolution and high-throughput sequencing"

19-09-11 Ben BROWN
Statistics, University of California, Berkeley, USA
"Nonparametric inference in functional and translational genomics"

29-08-11 Vincent LACROIX
Université de Lyon, Biométrie et Biologie Évolutive, CNRS, Université Lyon 1, Equipe BAMBOO, INRIA Grenoble Rhône-Alpes, France
"Identification of polymorphism in NGS data without a reference genome"

30-06-11 Toni REVERTER
Computational and Systems Biology, CSIRO Livestock Industries, Sta Lucia, Australia
"Gene network inference applied to two disparate scenarios: Cow puberty and colon cancer"

23-06-11 Guy COCHRANE
European Nucleotide Archive Team, European Bioinformatics Institute (EBI-EMBL), Hinxton, Cambridge, United Kingdom
"European Nucleotide Archive: Life on the log scale"

14-06-11 Pierre BALDI
School of Information and Computer Sciences, University of California, Irvine, USA
"Integrative Systems Biology Approaches to Gene Regulation and Beyond"

23-05-11 Brian DALRYMPLE
Molecular Technologies/Matching the Animal to the Market, CSIRO Livestock Industries, St Lucia, Australia
"Comparative systems biology: a livestock perspective"

14-04-11 Larry STANTON
Deputy Director, Genome Institute of Singapore, Singapore
"Applying genomic approaches to elucidate transcriptional networks that regulate cell fate decisions"

14-04-11 Dmitri PERVOUCHINI
Department of Bioengineering and Bioinformatics, Moscow State University, Institute for Information Transmission Problems, Russian Academy of Science, Moscow, Russia
"The role of secondary structure and transcription kinetics in pre-mRNA splicing"

17-02-11 Eduard PORTA-PARDO
Computational Cell Biology Group, Institut de Medicina Predictiva i Personalitzada del càncer (IMPPC), Badalona, Spain
"Using ontologies to analyze the human mutational spectrum"

11-02-11 Yijun RUAN
Genome Institute of Singapore, Singapore
"3D chromatin architecture and transcription regulation"

CELL & DEVELOPMENTAL BIOLOGY PROGRAMME

12-12-11 Nicole GORFINKIEL
Centro de Biología Molecular Severo Ochoa, Madrid, Spain
"Contractile activity of amnioserosa cells during dorsal closure in Drosophila"

09-11-11 Christer EJSING
Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark
"Lipidomics, new tools and applications"

08-11-11 Boris CHAGNAUD
Department of Neurobiology, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany
"Temporal patterning of vocalizations and its consequences for the prevention of acoustic reafference– a lesson from vocalizing fish"

28-10-11 Maria José MENDIBURO
Laboratory of Patrick Heun, Max Planck Institute, Freiburg, Germany
"Drosophila CENH3 is sufficient for centromere formation"

06-10-11 Guillaume SALBREUX
Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany
"Physics of the actin cortex in polar oscillations of dividing cells"

19-05-11 Raghu PADINJAT
The National Centre for Biological Sciences (NCBS), Bangalore, India
"Regulation of the phosphatidylinositol signalling cascade"

18-04-11 Petra STOCKINGER
Laboratory of Dr. Carl-Philipp Heisenberg, Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG), Dresden, Germany
"Regulation of motoneuron migration in the zebrafish hindbrain"

14-04-11 Snezhka OLIFERENKO
Temasek Life Sciences Laboratory, Singapore, Singapore
"Divergent strategies of the nuclear membrane management during mitosis"

31-03-11 Ombretta FORESTI
Institute of Molecular and Cellular Biology, University of Leeds, Leeds, United Kingdom
"Recycling of the plant vacuolar sorting receptor, a tail's tale"

28-03-11 Miquel TUSON
Kathryn Anderson lab, Developmental Biology Program, Sloan-Kettering Institute, New York, USA
"Protein kinase A acts at the basal body of the primary cilium to promote Gli3 processing and inhibit Gli2 activation"

18-03-11 Georgina SORROSAL DE LUNA
Marco Milan's lab, IRB Barcelona, Spain
"Scarface, a secreted serine protease-like protein, regulates polarized localization of laminin A at the basement membrane of the Drosophila embryo"

17-03-11 Marta BATLLE
Ferran Azorin's Lab, IRB Barcelona, Spain
"Analysis of the subcellular localisation and functional properties of DDP1, Drosophila's vigilin"

15-03-11 Alexandre DALET
Ludwig Institute for Cancer Research, de Duve Institute, Université Catholique de Louvain, Brussels, Belgium
"Splicing of antigenic peptides by the proteasome"

GENES & DISEASE PROGRAMME

22-12-11 Roger PIQUE-REGI
Department of Human Genetics at the University of Chicago, USA
"Variation in chromatin and heritable variation in gene expression"

21-07-11 Anna VAN TETERING-HOUBEN
Cell Biology Department, Netherlands Cancer Institute, The Netherlands
"Biochemical and functional characterization of Autotaxin in cancer"

13-07-11 Tim SPECTOR
Kings College London, London, United Kingdom
"Use of Twins in the Omics era"

11-02-11 Oliver DRECHSEL
Max-Planck-Institute for Molecular Plant Physiology, Postdam-Golm, Germany
"Investigating structure and function of the chloroplast genome"

DIFFERENTIATION & CANCER PROGRAMME

13-12-11 Melike LAKADAMYALI
Institut de Ciències Fotòniques (ICFO), Castelldefels, Spain
"Imaging cellular and sub-cellular structure and dynamics at high spatial resolution"

04-11-11 Jasna MEDVEDOVIC
Group of Meinrad Busslinger, Institute of Molecular Pathology, Vienna, Austria
"Molecular mechanisms in Pax5-mediated gene regulation and Igh locus contraction"

06-09-11 Magnus HANSSON
Department of Molecular Toxicology, Institute of Environmental Medicin, Karolinska Institute, Stockholm, Sweden
"Molecular mechanisms of the transcriptional coactivator Mastermind-like 1"

29-07-11 Scott I. SIMON
Department of Biomedical Engineering, University of California, Davis, USA
"Engineering neutrophil response for improved resolution of Staph-aureus wounds"

18-07-11 Anindya DUTTA
Dept. of Biochemistry & Molecular Genetics, Mellon Prostate Cancer Research Institute, University of Virginia, Charlottesville, USA
"Small RNAs in control of cell proliferation and muscle differentiation"

14-07-11 Ligia CARINHA GOMES
Dulbecco-Telethon Institute at the Venetian Institute of Molecular Medicine, Padova, Italy
"Autophagy and mitochondrial elongation: sustaining cell viability under difficult conditions"

28-06-11 Elena CATTANEO
Director of the Centre for Stem Cell Research (UniStem), University of Milan, Milan, Italy
"Huntingtin and stem cells in Huntington's Disease"

27-06-11 Jyotsna DHAWAN
Institute for Stem Cell Biology and Regenerative Medicine (inStem), National Center for Biological Sciences, Bangalore, India
"Quiescence and Stem Cell Function"

25-05-11 Mathias PAWLAK
Jaenisch Group, Whitehead Institute, Cambridge, USA
"The role of de novo DNA methyltransferases"

28-04-11 Christian SCHRÖTER
EMBO LT candidate, MPI-CBG Dresden, Germany
"Dynamics of Gene Regulatory Networks in Somitogenesis and Early Lineage Specification".

29-03-11 Harald KRANZ
Gene Bridges GmbH, Heidelberg, Germany
"Red/ET Recombination technology: Recombineering – a powerful tool to modify the mouse genome"

16-03-11 Yen-Sin ANG
Laboratory of Ihor Lemischka, New York Stem Cell Center, New York, USA
"Integrated Epigenetic & Transcriptional control of pluripotency"

GENE REGULATION PROGRAMME

14-11-11 Natalia PINZÓN
Laboratoire de Biologie Moléculaire Eucaryote, Université Paul Sabatier, Toulouse, France
"The complex universe of small noncoding RNAs".

14-11-11 Nicola CAVALLARI
Dept. of Biochemistry and Molecular Biology, University of Ferrara, Ferrara, Italy
"Engineered U1-snRNAs in the treatment of severe coagulopathies".

04-11-11 Priyanka SHARMA
German Cancer Research Center (DKFZ), Heidelberg, Germany
"Citrullination of Histone H3 at position 8 and its effect on transcription regulation"

27-10-11 Lidia LÓPEZ-SERRA
Chromosome Segregation Laboratory, Cancer Reserach UK- London Research Institute, London, United Kingdom
"Chromatin architecture and the Scc2/Scc4 Cohesin and Condensin loading complex".

17-10-11 Tim HUGHES
Donnelly Centre and Banting and Best Department of Medical Research, University of Toronto, Toronto, Canada
"Exploring the eukaryotic protein-nucleic acid interactome"

06-05-11 Adrien DECORSIERE
INSERM U563, Institut Claudius Regaud, Toulouse, France
"Essential role for the interaction between hnRNP H/F and a G-quadruplex in maintaining p53 3' end processing and function during DNA damage"

02-05-11 Yves ROIGNANT
Developmental Genetics Program, Skirball Institute of Biomolecular Medicine, New York, USA
"A Novel Role for the Exon Junction Complex in Splicing Large Heterochromatic introns"

02-05-11 Reini FERNÁNDEZ DI LUCO
National Cancer Institute, NIH, Bethesda, USA
"Epigenetics in alternative splicing"

02-05-11 Miguel A. ZARATIEGUI
Cold Spring Harbor Laboratory, New York, USA
"Dissecting Heterochromatin Inheritance"

02-05-11 Luisa DI STEFANO
Massachussetts General Hospital, Boston, USA
"Functional antagonism between histone H3K4 demethylases in vivo"

02-05-11 Kiran PADMANABHAN
Dept. of Neurobiology, Harvard Medical School, Boston, USA
"Circadian Rythms in mice: Transcriptional regulation by PERIOD protein complexes"

02-05-11 Guillaime FILION
The Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands
"Unraveling the drosophila chromatin by protein-genome profiling"

02-05-11 Emmanuel ROSONINA
Dept. of Biological Sciences, Columbia University, New York, USA
"Beyond the 1st round of transcription: SUMO clears promoters for reactivation"

01-04-11 William Lee KRAUS
Cecil H. and Ida Green Center for Reproductive Biology Sciences, The University of Texas, Southwestern Medical Center, Dallas, USA
"Regulation of Chromatin Structure and Gene Expression by PARP-1"

21-02-11 Marta PABIS
Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany
"The cap-binding complex promotes co-transcriptional splicing in vivo"

CORE FACILITIES

29-11-11 Leila SMITH
Fluidigm Europe B.V., Amsterdam, The Netherlands
"The BioMark™ HD System: recent advancements in next generation microfluidic qPCR platform enabling single cell gene expression, microRNA profiling."

22-11-11 Jim WHITE
NanoString Technologies, Seattle, USA
"The nCounter Analysis System: multiplex gene analysis made easy"

26-09-11 Jennifer MOLETTE
CLC Bio, Aarhus, Denmark
"CLC bio platform for NGS data analysis"

21-09-11 Veronique BERCHER
Field Application Specialist, Reagents and Multimode Plate Readers, PerkinElmer, Waltham, USA
"An overview of biochemical and cellular assays with label-free technologies"

14-09-11 Jonas KORLACH
Pacific Biosciences, Menlo Park, California, USA
"Single-Molecule, Real-Time (SMRTT) Monitoring of Biomolecules: DNA Sequencing and Beyond"

28-06-11 Deepak SINGH
Senior Director, Sales, Pacific Biosciences Europe, Menlo Park, California, USA
"Revealing SMRT Biology. Overview of single molecule sequencing on the PacBio RS platform"

23-03-11 Patrick HOLLENDER
Life Technologies (SAS), France
"Ion Torrent: Semiconductor Sequencing for LifeT"

15-02-11 Christian ZINSER & Nancy BRETSCHNEIDER
Genomatix Software GmbH, Munich, Germany
"Training and workshop on comprehensive methods for Next-Generation Sequencing data analysis"

07-02-11 Robert KUHN
UCSC Genome Browser, University of California, Santa Cruz, USA
" UCSC Genome Browser"

TECHNOLOGY TRANSFER DEPARTMENT

20-06-11 Hugh GOODFELLOW
Carpmaels & Ransford, London, United Kingdom
"Patents – why bother?"

08-06-11 Alexis SCHUBERT
German Patent Attorney, European Trademark & Design Attorney, Zacco GmbH, Munich, Germany
"From clones to claims - patent attorney as a career choice for scientists (The experience of a former CRG postdoc)"

Appendix 5

PRESS CLIPPING


Tendencias

Mejora el diagnóstico genético prenatal

Y, sin embargo, se mueve


Las células TF3 mejoran la reprogramación de las células precursoras neurales


ENTORNO

La lucha contra el cáncer pide ayuda a la inversión privada


Una competición busca la mejor solución para el puzle del ADN


CIENCIA

Luis Serrano


«Queremos simular un organismo vivo al completo en el ordenador»


Mito y ciencia


SOCIEDAD

Investigadores españoles obtienen el mapa genético de la leucemia


Dimisión del director del mejor centro científico de Catalunya


La pell també pateix 'jet lag'


ScienceDaily

Biological Clock Controls Activation of Skin Stem Cells


How Skin Tells Time

The behavior of skin stem cells is regulated by a 24-hour circadian clock.


+Ciència

Catalans en el rànquing dels millors


OMICS & SYSTEMS BIOLOGY

Genomics for the win against cancer?

Setting the standard


CENTENARIOS

El secreto de la longevidad


"Las células madre embrionarias y la reprogramación celular tendrán diferentes aplicaciones"


ScienceDaily

Why Does the Same Mutation Kill One Person but Not Another?


"Les hormones sexuals influeixen sobre el cablatge del cervell"


Siete científicos de Catalunya, reconocidos por Europa

El ERC amplía la financiación de los mejores proyectos


The Tradeoff of Stress

For nematode worms, a bigger stress response means a healthier, longer life, but fewer babies.


El sexe dels peixos


En busca de la MEJOR CIENCIA de España


MEDICINA

Hay demasiadas expectativas con la medicina personalizada


entrevista a mark isalan

científic del Centre de Regulació Genòmica (CRG)

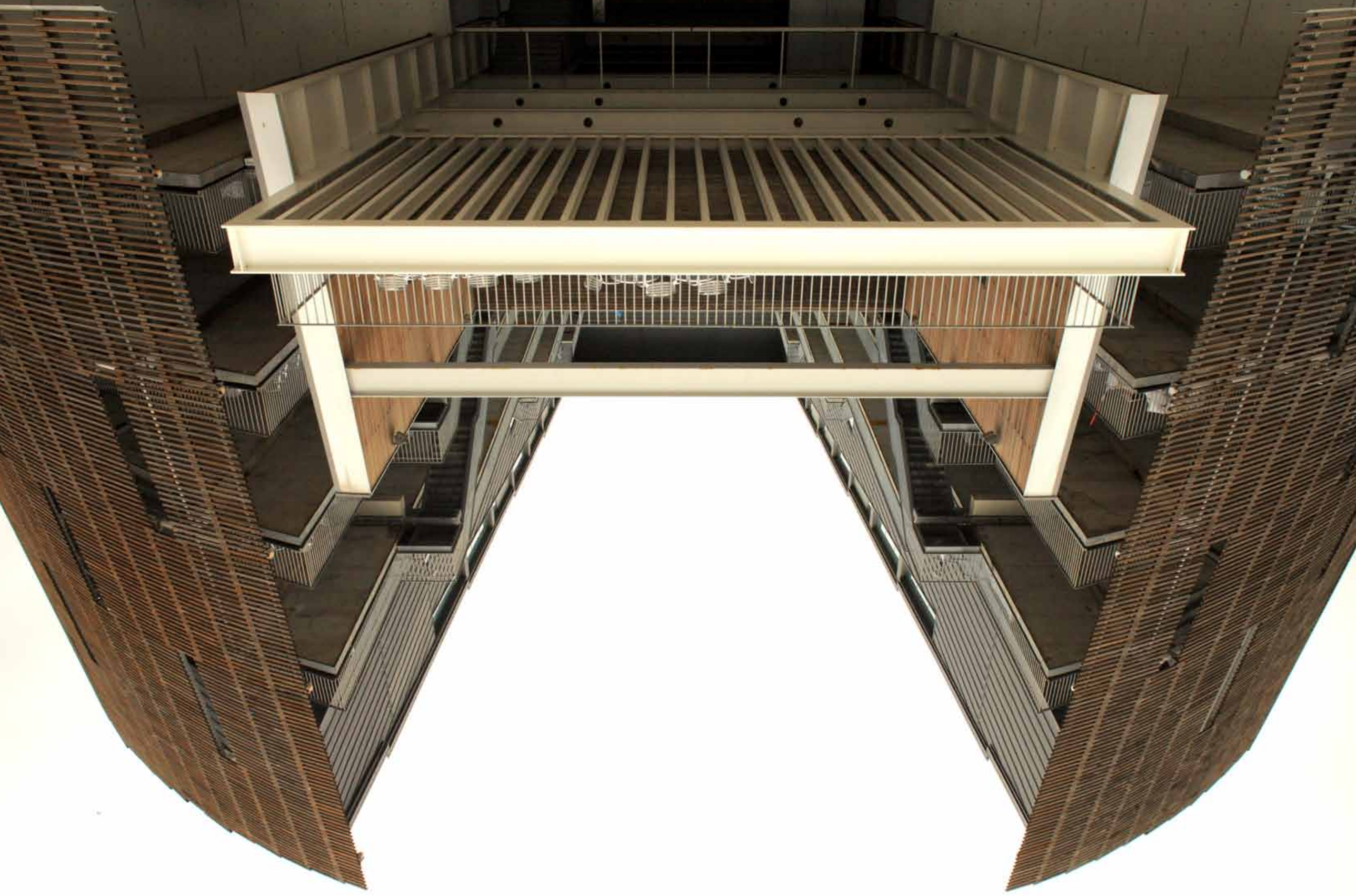

"Todos descendemos de Lucy, que tenía una sola célula"


ScienceDaily

Cell Transformation from One Type of Cell to Another


SOCIEDAD

Un gallego plantea que un virus causa la supermutación del cáncer



CRG
Centre
for Genomic
Regulation

Centre for Genomic Regulation

PRBB Building
Dr. Aiguader, 88
08003 Barcelona, Spain

Tel.: +34 93 316 01 00
Fax: +34 93 316 00 99

comunicacio@crg.eu
http://www.crg.eu