CRG
Centre for Genomic Regulation

Centre for Genomic Regulation
Annual Report 2012

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: Sergi Espada. Genoma ArtStudio SCP (www.genoma-artstudio.com)

Photography: Ivan Martí and CRG scientists

Cover photography: Esteban Rozen

Legal deposit: B-xxxxx-2013

CRG
Centre
for Genomic
Regulation

CONTENTS

FOREWORD

The year 2012 has been a difficult one for Spanish science. Due to the world economic crisis, the central and autonomic governments have been forced to cut their expenses and, in general, one of the big losers has been science. Despite this, and thanks to the strong support of the Catalan autonomic government, the CRG has managed not only to keep being an exciting and attractive place to do science, but many new endeavours are taking place.

It is important in this context to mention that the performance of the CRG was externally evaluated by the CERCA Institution (our umbrella organisation and part of our board of trustees). This evaluation praised the institute as well as recommending certain improvements that we will implement. This year the Bioinformatics & Genomics and the Genes & Disease programmes were also scientifically evaluated by both our Scientific Advisory Board (SAB) and external scientists. Based on their recommendations, the Genes & Disease programme was terminated and their scientists redistributed among other CRG programmes consistent with their scientific interests. As a result, from 2013 on the CRG will have four scientific programmes with significant critical mass. These scientific adjustments will be accompanied by a fundamental administrative reorganisation to improve efficiency.

Aside from these changes, in 2012 we decided to strengthen and develop international links with other institutes, centres and universities worldwide, as well as our collaborations with pharmaceutical and biotechnology companies. The CRG wants to play a major role in world science, not only in scientific achievement but also in guiding scientific policies and strategies. As part of this goal, we have established new collaboration agreements with local hospitals and institutes, as well as with foreign institutions.

I hope that in the years to come, the CRG together with other European institutes will help foster European competitiveness on a global scale.

Luis Serrano
Director

A LOOK BACK AT THE YEAR

A LOOK BACK AT THE YEAR

Throughout 2012, the year of CRG's 10th anniversary, the institute has been immersed in a flow of vibrant and frenetic activity that has enabled us to achieve some remarkable milestones. In a mere ten years, the CRG has become one of the leading research centres in Europe, contributing to the international recognition of Catalan and Spanish science.

In autumn, we celebrated our 10th Anniversary with a great institutional symposium that brought together prominent figures from the world of research. Nobel prize winner in chemistry, Ada Yonath, and Albert Lasker, award winners, Randy Sheckman and Tom Maniatis joined the event in recognition of the trajectory of the organisation. National and regional figures, such as Andreu Mas-Colell, Catalan Minister of Economy & Knowledge, and Carmen Vela, Spanish Science Secretary, as well as members of industry, also participated in the celebration.

From left to right:
Andreu Mas Colell, Minister of Economy and Knowledge of the Generalitat de Catalunya; Carmen Vela, Spanish Science Secretary; Luis Serrano, Director of the CRG; Jose García Montalvo, Rector of Pompeu Fabra University; Antoni Castella, Secretary for University and Research under the Ministry of Economy and Knowledge of the Generalitat de Catalunya; Jordi Camí, Director of the Barcelona Biomedical Research Park PRBB

In order to consolidate its position as a leader in biomedicine, the CRG has been strengthening its strategic alliances with institutions and companies and 2012 has been very fruitful in this regard.

> In January, the CRG and the Vall d'Hebron Research Institute (VHIR) signed an agreement for the development of collaborative projects in basic and clinical research. One of the main focuses was research on rare genetic diseases. The agreement aims to encourage an alliance between clinical and basic researchers with the objective of bringing the discoveries of basic research more in line with clinical needs and, in the opposite sense, allow clinical researchers to experience firsthand the challenges, advances and possibilities of fundamental research. The final goal is the translation of all of this knowledge into practice. Identifying the current clinical necessities of medical centres is vitally important for the partnership between the basic and clinical areas.

> In March, the CRG and Sanofi entered into a master collaborative research agreement. The main goal is to share expertise and know-how to explore new ways for evaluating translational medicine from basic research to patient benefits. Within the frame of the agreement, the CRG and Sanofi have already initiated a first set of projects to discover innovative therapeutic approaches for infectious diseases, develop novel delivery systems using synthetic biology, decipher disease-relevant cell trans-differentiation pathways, and identify original targets from unexploited genomic transcription mechanisms.

> In April, the CRG and the National Centre for Genomic Analysis (CNAG) signed a collaboration agreement with the aim of boosting research in the field of genomic analysis. The first fruit of this alliance was the Structural Genomics research group of CNAG, led by researcher Marc A. Marti-Renom, joining the Gene Regulation, Stem Cells and Cancer programme at the CRG. With this dual affiliation, the Structural Genomics group will have better resources to develop their science both experimentally and computationally.

> In May, the CRG and the ACE Foundation signed a new collaboration agreement and united in research into Alzheimer's and neurodegenerative diseases. The CRG will study genetic basis as a possible cause of the development of neurodegenerative diseases and the ACE Foundation will contribute its expertise in the diagnosis and treatment of these diseases. Specifically, scientific collaboration is expected in the area of genomics as well as functional studies of cases of neurodegenerative diseases.

> In August, the CRG, after signing a collaboration agreement with the National Centre for Biological Sciences (NCBS) in Bangalore, India, announced joint postdoctoral fellowships through the NCBS-inStem-CRG Postdoctoral Fellows Programme. This programme aims to attract postdoctoral fellows of any nationality to link ongoing collaborations between researchers at the NCBS-inStem and the CRG to develop new research directions. The primary goal of this programme is to provide a unique training experience for postdoctoral scientists who want to work at the interface between fundamental, interdisciplinary areas of biology with the possibility of using pioneering technology to create applications and new research tools.

A very significant and remarkable milestone for the CRG in 2012 was its accreditation as a "Severo Ochoa Centre of Excellence". This award from the Spanish Ministry of Economy and Knowledge aims to "identify and promote the Spanish centres and research units which are among the best in the world in their field." The distinction is valid for the next four years and involves 1 million euros of funding per year for the centre. It will reinforce the research capacities of the institute, providing access to other competitive grant programmes and large research facilities. Moreover, this recognition will also improve the CRG's visibility for obtaining new sponsorships. The choice was made by an international scientific committee, with more than 70 scientists, world-recognised in their respective areas, from 12 different countries.

In addition, in 2012 the CRG, together with the VIB (Belgium), launched the Core Facilities Excellence Alliance "Core For Life" (www.core4life.eu), which also includes the EMBL, MPI-CBG (Dresden, Germany), IMP and CSF (Vienna, Austria), as well as the Functional Genomics Centre Zurich.

The SCImago Institution Rankings (SIR) World Report 2012, covering the years 2006-2010, classifies the CRG in 13th position (according to the Q1 indicator, health sector) out of over 3,290 research institutions around the world. In Europe only three other research centres in the health sector 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 plateaued since 2009 reflecting the growth freezing experienced that year. In 2012, 188 papers were published in international journals with an average impact factor of 10.81, and 88 seminars were held by top-level invited speakers. Many of these activities were picked up by the media, and the CRG hit the news (newspapers, radio, TV) on 760 occasions.

Finally some of our young and senior scientists have obtained awards and honours in recognition of the excellence of their science. Ben Lehner obtained the Banc Sabadell Prize for Biomedical Research and the 2012 Catalan National Research Award for Young Talent; Pedro Carvalho and Fyodor Kondrashov received the International Early Career Award from the Howard Hughes Medical Institute; Gian Gaetano Tartaglia, Pedro Carvalho, Salvador Aznar-Benitah and Toni Gabaldón were awarded Starting Grants by the European Research Council; Vivek Malhotra was awarded the MERCK Award from the American Society of Biochemistry and Molecular Biology; and Isabelle Vernos was elected member of the Advisory Council for Science, Technology and Innovation, recently created by the Spanish Secretary of State for R&D.

Ben Lehner receiving the Banc Sabadell Prize for Biomedical Research, in June 2012.

RESEARCH

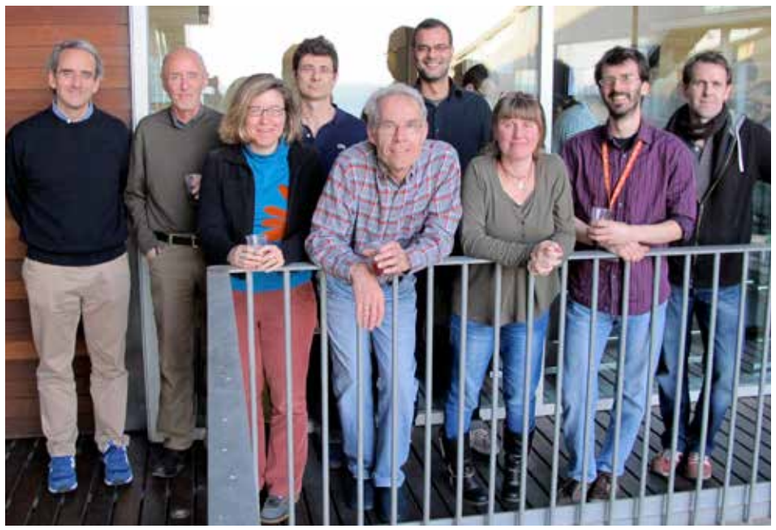
Systems Biology

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 2012, the programme was awarded many new grants, including ISBE-Infrastructure for Systems Biology in Europe (awarded by the European Commission), SwarmOrgan (from the Future and Emerging Technology call of the EC), Sinergia (from the Swiss National Science Foundation), and a variety of national projects and individual grants. Given the wide range of biological questions addressed in the programme, a highly interactive 2-day retreat was organised this year, in which postdocs and students had to propose ways to tackle 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 2012 included papers on signal transduction, gene regulatory networks, reverse engineering, multicellular patterning and fitness tradeoffs.

Gene Regulation, Stem Cells and Cancer

At the beginning of 2012, after very positive evaluations of the Gene Regulation Programme and the Differentiation and Cancer Programme, the two merged to become the Gene Regulation, Stem Cells and Cancer Programme. The main scientific goal of the merger is to optimise scientific synergies by promoting collaboration between the groups involved in mechanistic studies of gene regulation through chromatin and RNA (led by Miguel Beato, Fátima Gebauer and Juan Valcárcel) and those involved in understanding stem cell homeostasis, ageing and the onset of cancer (led by Salvador Aznar-Benitah, Maria Pia Cosma, Luciano Di Croce, Thomas Graf and Bill Keyes). This is fertile ground for obtaining very novel insights into the genetic circuits controlling key fate decisions involved in the maintenance of cell pluripotency, cell differentiation and oncogenic transformation, as well as for revealing new possibilities for inducing changes in cell identity that could have significant implications in research and medicine. The productive collaboration between the groups of Salvador Aznar-Benitah and Luciano Di Croce on the epigenetic control of stem cell differentiation is an example of this potential.



Gene Regulation, Stem Cells and Cancer group leaders and coordinators. From left to right: Juan Valcárcel, Miguel Beato, Pia Cosma, Luciano Di Croce, Thomas Graf, Salvador Aznar-Benitah, Fátima Gebauer, Guillaume Filion and Bill Keyes

Two additional groups joined the programme in 2012, providing complementary expertise to study the organisation of the genome inside the cell nucleus. Guillaume Filion joined the CRG after a very successful postdoctoral period at the Netherlands Cancer Institute, where he used novel genome-wide mapping technologies to identify five distinct types of chromatin in *Drosophila* cells. In addition, Marc Martí-Renom became the first dual appointment between the CRG and a faculty from another institute, in his case the National Centre for Genomic Analysis (CNAG, Barcelona), to strengthen the links between research institutes with complementary expertise and resources. Marc is an expert in the computational integration of large-scale datasets to predict the three dimensional structure of molecules and the organisation of genomes. Guillaume and Marc have joined forces with Miguel Beato and Thomas Graf to develop an ambitious project for the analysis of genome architecture dynamics upon hormone stimulation and during the trans-differentiation of B cells into macrophages.

A dynamic series of data seminars by PhD students and postdocs, the first joint retreat of the programme and weekly meetings of the group leaders have helped to nurture an atmosphere of scientific discussion and interdisciplinary collaboration.

Genes and Disease

The Genes and Disease Programme was created in 2002 with the aim of developing research in the field of medical genetics and genomics, from the discovery of genes involved in human disorders to the development of preventive and therapeutic strategies for disease. Work of the researchers in the Genes and Disease Programme includes analysing sequence, structural and epigenetic variants of the human genome that could participate in the predisposition of disease, and in evaluating the role of genetic and environmental modifications in disease progression. Specific collective work within the programme was focused on understanding the function of genes with potential implications for mental retardation, psychiatric disorders and neurodegeneration, by using bioinformatics, as well as cellular and animal model approaches. Below are some of the achievements of the investigators during 2012.

The group of Susana de la Luna showed that deregulation of the apoptotic response in differentiating neurones participates in the neuropathology of diseases that display DYRK1A gene-dosage imbalance effects, such as Down Syndrome. Mara Dierssen's group revealed that the CHRNA5/A3/B4 genomic cluster produces nicotine addiction by modifying the activity of brain regions responsible for the balance between reward and aversion and that this disrupts working memory and impulsivity. The group of Stephan Ossowski has developed pipelines for the analysis of exome sequencing data which have led to the identification of genes responsible for several human disorders, such as KLHL3 mutations in familial hyperkalemic hypertension. Xavier Estivill's group identified alterations in several miRNAs in early stages of Parkinson's disease, and a new pathogenic mechanism for neurotoxic activity in Huntington's disease, which involves small CAG-repeated RNAs.

As a result of the CRG reorganisation, the Genes and Disease Programme was terminated at the end of 2012. Mara Dierssen was affiliated to the Systems Biology programme; Susana de la Luna to the Gene Regulation, Stem Cells and Cancer programme; and Stephan Ossowski and Xavier Estivill to the Bioinformatics and Genomics programme.

Bioinformatics and Genomics

In 2012, collaborations within the programme have grown substantially, as a result of a confluence of interests: Cedric Notredame and Roderic Guigo have both participated in the ENCODE project; Roderic Guigo specifically coordinated the RNA analysis working group within the project; Fyodor Kondrashov and Cedric Notredame collaborated in a landmark study that attempted to quantify the impact of epistasis in constraining evolution; Toni Gabaldon and Roderic Guigo participated in a number of genome projects, notably that of the melon. There has been a convergence in the interest of the researchers in the programme in long non-coding RNAs --the biological role, and the evolutionary behaviour of which are being investigated using complementary approaches. Gian Gaetano Tartaglia, in particular, is interested in the role of lncRNAs in disease.

In addition to carrying out their own research agenda, the groups have been busy in collaborations with other CRG research groups, and they have played a very active role in the design and implementation of the CRG's scientific computing network.

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, mammalian cells, and *Xenopus* egg-based assays and genetics are used to address the specific processes of membrane trafficking, organelle dynamics, protein translocation, cell shape, spindle dynamics and cytokinesis.

Yuichi Wakana from the Malhotra lab was recruited by the Tokyo college of Pharmacy, Japan.

CORE FACILITIES

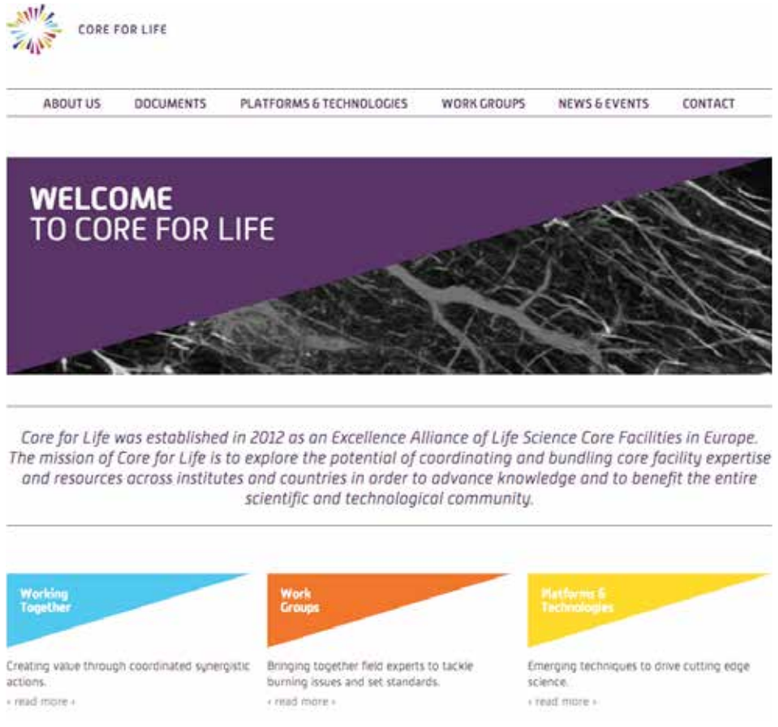
The core facilities and the technologies they offer continue to be a valued cornerstone of the research performed in the CRG. The programme currently comprises six Core Facility Units, Genomics, Proteomics, Advanced Light Microscopy, Biomolecular Screening & Protein Technologies, FACS, and Bioinformatics, as well as the Histology Service, which is only available to internal users. As in previous years, the overall activity in core facilities continues to increase.

In 2012 the Core Facilities experienced some remodelling, changes and improvements. After the closure of the CeGen Genotyping Unit at the end of 2011, the Microarrays and Ultrasequencing Units were merged into a single, strong Genomics Unit with Heinz Himmelbauer as the unit head and Anna Ferrer as the lab manager. Taking advantage of the close collaboration between the two units and the shared space, this merger is a true integration of all genomics activities at the CRG and opens new possibilities for the flexible use of resources and implementation of new technologies.

In February Eduard Sabidó joined the CRG as Head of the Joint CRG/UPF Proteomics Unit. Eduard came from ETH Zurich and has long-standing experience in mass spectrometry, biochemistry, biology and computer science. Also in February Ernesto Lowy officially became the manager of the CRG Bioinformatics Core. The major equipment upgrades in 2012 have been: 1) the installation of a new high-end cell sorter (BD Influx) and a new high-end analyser (BD LSR Fortessa) in the FACS Unit, and 2) the installation of a second super-resolution system (Leica GSD) in the Advanced Light Microscopy Unit.

In order to improve the services provided to the users, each Core Facility unit is working on individual technology development and implementation projects as time and capacity allow. In addition to these individual efforts, a cross-facility technology development project was launched to generate visibility for the CRG Core Facilities Programme while at the same time new technology was set up to provide our users with a competitive advantage over their colleagues at other institutes. The project selected is “Towards single cell (type) –omics”, a topic that fits the current developments at CRG very well, as more and more Group Leaders are asking for single cell applications, and this is conducted in collaboration with Bill Keyes on prostate cancer stem cells.

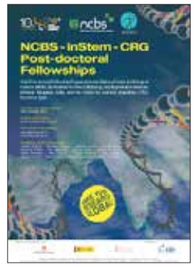
The CRG core facilities are not only well established locally, with users coming from different institutions in Barcelona and Spain (as well as from abroad), but they are also recognised partners in European initiatives. The Advanced Light Microscopy Unit is a partner in the ESFRI EuroBioimaging initiative, the Genomics Unit is a transnational access site in the ESGI European infrastructure network, and the Proteomics Unit is a transnational access and research site within the PRIME-XS European infrastructure network and the only Spanish Proteomics Facility listed on the European MERIL platform. The Head of Core Facilities is the founding member and coordinator of the Core Facilities Excellence Alliance “Core For Life” (www.core4life.eu), which also includes EMBL, VIB (Belgium), MPI-CBG (Dresden, Germany), IMP and CSF (Vienna, Austria), and the Functional Genomics Centre Zurich. Core For Life aims to share and consolidate procedures, uniting efforts in personnel training and technology validation, sharing access to facilities across institutes, and lobbying at EU level for infrastructure funding.



In December 2012 the Core Facilities Programme passed its first external review. A panel of seven core facility, research, and institute management experts from renowned international institutes spent two days at the CRG and evaluated the different units as well as the Core Facilities Programme as a whole. The result of the evaluation was very positive and the facilities were congratulated on their performance and their essential contribution to the CRG.

ADVANCED TRAINING

In 2012, the International PhD Programme continued to attract a lot of young talent from all over the world, supported both by internal and external competitive funding and the “la Caixa” International PhD Fellowship Programme. This year, the 10 candidates selected in the “la Caixa” call started their PhD thesis in September, after a highly competitive selection process involving significantly more candidates than the previous year (533 candidates from over 70 different countries). This year, the specific training offered to PhD students included 8 Advanced Seminars in Biomedical Research, in partnership with Pompeu Fabra University, as well as 20 practical scientific and technical courses organised by the CRG faculty and core facilities. The CRG PhD community has actively promoted a number of initiatives, such as the 6th annual PhD symposium (19-20 Nov) and PhD retreat (20 Nov/1 Dec), and the second joint retreat, this year with the Università San Raffaele in Milan (7-9 June) as well as the first Women in Science Symposium with leading experts on gender issues in science (9 March).



The International Postdoctoral Programme at the CRG currently hosts around 85 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 2012, 6 external fellowships and postdoctoral contracts were awarded to CRG postdocs. In the third and last INTERPOD call 2 candidates were selected. In 2012, the postdocs at the CRG also organised their own symposium (10 June). An agreement for the establishment of a joint NCBS-CRG Postdoctoral Fellowship Programme was signed with the National Centre for Biological Sciences (NCBS), the leading biomedical research institute in India and the first call was launched in September.

A highlight of the training programme is the new series Courses@CRG initiated in 2012, covering a wide range of topics, from practical scientific to technological courses, and which is also open to the scientific community. In 2012 several highly successful events were held, involving both internal and external participants:

> Summer Course Modelling for Systems Biology (1-6/07/2012) organised by J. Jaeger (CRG), H. Janssens (CRG), M. Louis (CRG), J. Sharpe (CRG)

> Foerster Resonance Energy Transfer (FRET) microscopy as a tool for functional imaging and quantitative analysis (26-28/09/2012) organised by T. Zimmermann (CRG)

> Second Generation Sequencing: Lab Methods and Data Analysis (22-26/10/2012) organised by H. Himmelbauer (CRG)

The recently established fully-equipped Teaching and Training Labs, supported by private sponsors, will be key in hosting the growing training activities of the CRG for scientists at all stages of their careers, ranging from courses and workshops to activities for school children.

Advanced Training figures

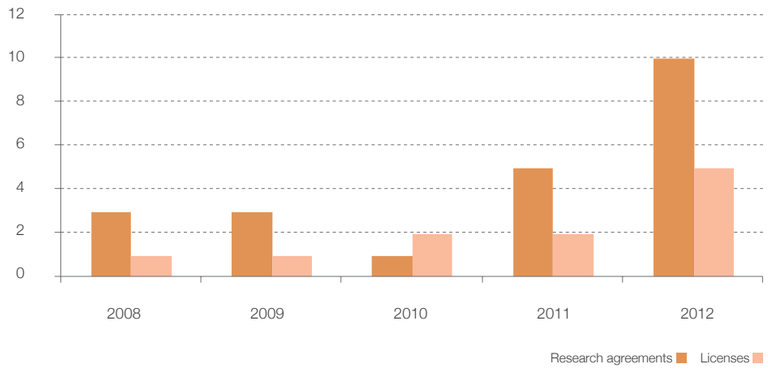
PhD students: 104 (78% foreign)
External fellowships obtained by PhD students: 9+9 “la Caixa” fellowships
Advanced Seminars offered to Masters and PhD students: 8
Practical courses offered to PhD students: 20
PhD theses defended: 21
Postdoc researchers: 85 (68% foreign)
External fellowships obtained by postdoctoral researchers: 6+2 Interpod fellowships
Courses@CRG: 3

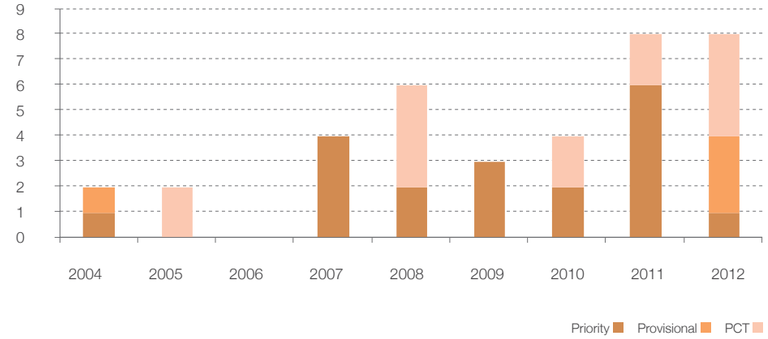
PhD Theses Defended in 2012

Name	Programme	Date	University	Thesis Project
Paola Pisano	Luciano di Croce	26/01/2012	Università San Raffaele	*Epigenetic post-translational modification of historine H3 modulate binding of adaptor proteins during embryonic stem cell differentiation*
Peggy Janich	Salvador Aznar-Benitah	16/02/2012	UPF	*The role of circadian genes in epidermal homeostasis*
Filipe Pinto	Hernán López-Schier	17/02/2012	UPF	*Development and functional regeneration of the zebrafish lateral line system*
Anne Campagna	Luis Serrano	17/02/2012	UPF	*Structural Analysis of Protein Interaction Networks*
Elisa Balducci	Mariona Arbonès	23/02/2012	UPF	*Effect of DYRK1a dose reduction on the transcriptome of the developing mouse cerebral cortex: implications on liogenesis*
Jesús Pujol	Hernán López-Schier	02/03/2012	UPF	*Neural map organisation and development in the lateral-line system*
Anna Kedzierska	Roderic Guigó	16/03/2012	UPF/UPC	*Statistical models of genomic sequences*
Joana Ribeiro	Luciano di Croce	19/03/2012	Universidade do Porto	*Transcriptional Activation of olycomb-Repressed Genes by ZRF1: Implication in ifferentiation and Cellular Senescence*
Gabriel Neurohr	Manuel Mendoza	13/04/2012	UPF	*A Midzone-Based Ruler Adjusts Chromosome Compaction to Anaphase Spindle Length*
Susanna Molas	Mara Dierssen	22/06/2012	UPF	*Nicotine addiction phenotypes in a BAC transgenic mouse model overexpressing the CHRNA5/A3/B4 genomic cluster*
Alessandro Di Tullio	Thomas Graf	13/07/2012	UPF	*Reprogramming of B cells into macrophages: mechanistic insights*
Alejandro Burga	Ben Lehner	20/07/2012	UPF	*The consequences of stochastic gene expression in the nematode Caenorhabditis elegans*
Judith Wodke	Luis Serrano	29/10/2012	UPF	*Organisation and Integration of Large-scale Datasets for Designing a Metabolic Model and Re-annotating the Genome of Mycoplasma pneumoniae*
Ana Villalba	Fátima Gebauer	09/11/2012	Universidad de Valencia	*Molecular mechanisms of cytoplasmic polyadenylation in Drosophila*
Antonious Lioutas	Isabelle Vernos	09/11/2012	UPF	*Aurora A Kinase function during anaphase*
Salvador Capella	Toni Gabaldón	16/11/2012	UPF	*Analysis of multiple protein sequence alignments and phylogenetic trees in the context of phylogenenomics studies*
Felice Alessio Bava	Juan Valcárcel/ Raúl Méndez	16/11/2012	UPF	*CPEB1 coordinates alternative 3’UTR formation with translational regulation*
Susana Iraola	Xavier Estivill	03/12/2012	UB	*Analysis of epigenetic inheritance in neurological disorders*
Carsten Kemena	Cédric Notredame	12/12/2012	UPF	*Improving the accuracy and the efficiency of multiple sequence alignment methods*
Diana Reyes	Miguel Beato	13/12/2012	UPF	*Role of MSK1 in the response of breast cancer cells to ovarian steroid hormone*
Angela Paola Krüger	Ben Lehner	14/12/2012	UPF	*Systematic and Quantitative Analysis of the Early Embryonic Development of Caenorhabditis elegans*

TECHNOLOGY TRANSFER

The most remarkable fact for 2012 is the number of new licensees and research agreements signed. Over the past years there has been a clear tendency for these numbers to increase, but in 2012 we signed more than twice as many licenses and research agreements than in the previous year. Regarding the total number of patents and PCT extensions, the number was the same as the previous year. Another remarkable fact of 2012 is that the TTO won an Interregional EU grant (ETTBio, www.ettbio.eu), the objective of which is to explore and exchange experiences to improve the efficiency of technology transfer in biomedicine. Some of our partners in this project are the Imperial College Business School (UK), the Free University of Brussels (Belgium) and the Technical University of Dresden (Germany).

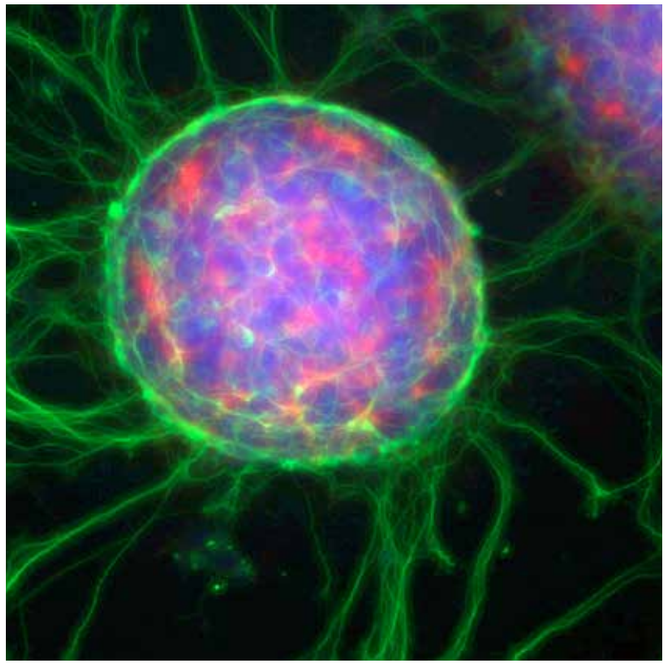




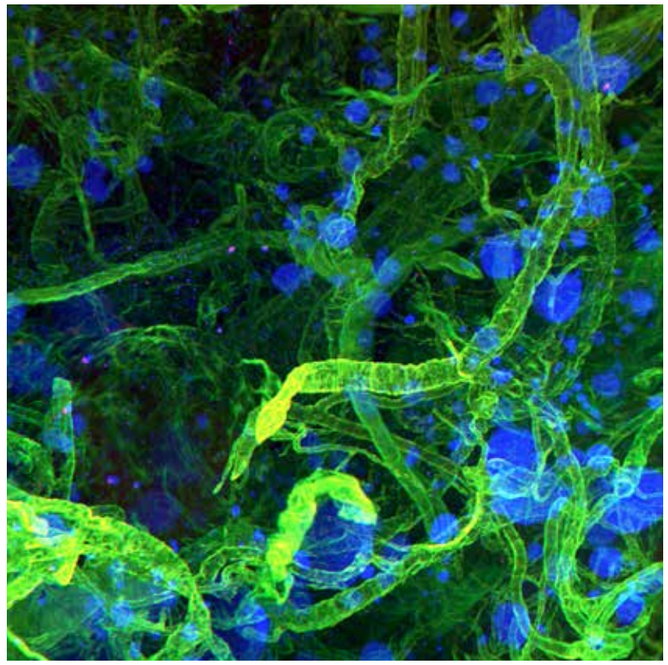
COMMUNICATION & OUTREACH

One of the strategic objectives of the CRG activity of is “to communicate and establish a dialogue with society, educating the public and taking into account their demands and needs.” To this end, during 2012 the number of activities organised continued to increase. 113 activities were held, including workshops for primary and secondary schools, summer internships, the open day, cell model contest (sculptures), science cafés, “Easy Science” conferences and participation in fairs and events for the dissemination of science.

Also in the autumn of 2012, in order to continue expanding the scientific photographic resources, we organised the second edition of the internal CRG Scientific Photography Competition. Of note was the level of participation and quality of the photographs submitted. The winning entries were announced at the Christmas party, where the authors received their prizes.



Awarded and short-listed pictures of the 2nd CRG Scientific Photography Competition:
Awarded picture (above): “Emerging stars: Small galaxies in a culture dish” (Immunostaining of clustered Retinal Ganlion cell primary cultures), by Esteban Rozen.
Short-listed picture (below): “Deep inside the pancreas” (Part of a murine pancreas imaged with an experimental light sheet microscope setup. Blood vessels in green, labeled with Asma-Cy3. Insulin producing beta-cells in blue, labeled with IR647 anti-GP. Beta-cells contribute up to 80% of the cells in Islets of Langerhans, the larger blue structures in the image. The original dimensions of the pancreatic tissue are 1.3 x 1.7 x 7.2 mm. The image is a maximum value projection of 1441 slices, so the spacing is 5 µm), by Jürgen Mayer and Jim Swoger.



The different findings published in top scientific journals, the activities organised and the recognition of the CRG as a research institute of reference in the biomedical arena led, throughout the year, to articles in the press and online media, as well as participation by scientists in radio and television programmes. In April we launched the internal news bulletin Life@CRG and in the summer we inaugurated a brand new Science & Society section on our corporate website, which includes detailed information about all our outreach activities and educational resources.

Moreover, during 2012 the presence of the CRG in social networks, which are proving to be a very effective additional dissemination tool for all activities, increased remarkably. 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. 2012 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 Masters and doctoral classes, and training courses, among others, have also been organised.

As for scientific meetings, it is worth highlighting the "QuanTissue: Quantitative Models in Cell and Developmental Biology", RECOMB 2012, Annual meeting "Chromosomes, Stem Cells and Diseases" & Satellite Workshop "Modelling 3D-Structure of Chromosomes", "Anniversary Symposium: 10 Years of Research, Education & Training at the CRG", and two editions of the Core Facilities Technology Symposia series.

CRG Communication & Outreach figures

Outreach activities: 113
Audience reached: nearly 6,000 people
Press releases & short pieces of news: 48
Written/online media appearances: 659
Radio/TV appearances: 43
Blogs: 58
International meetings: 18
High profile seminars: 88

GENERAL ADMINISTRATION

Management & Research Support

In 2012 we have implemented changes arising from the reorganisation plan of the Management & Research Support Team, begun in 2011, in order to adapt the team to the growing institute, new environmental circumstances, and also to increase efficiency.

For this reason, during the year we have been working on the automation of certain data analysis and exploitation processes. We have also consolidated the quarterly follow-up of the indicators together with their analysis by the heads of the management area departments.

The creation of the Administration Department, including Finances, Purchasing, Secretarial & Reception areas (HHRR will be added in 2013), together with the creation of the Grants & Academic Management Department that has integrated the Grants Post-award team (from the previous Finances Area) have been the highest impact changes. Work has been done, which will continue during 2013, on improving transversal processes.

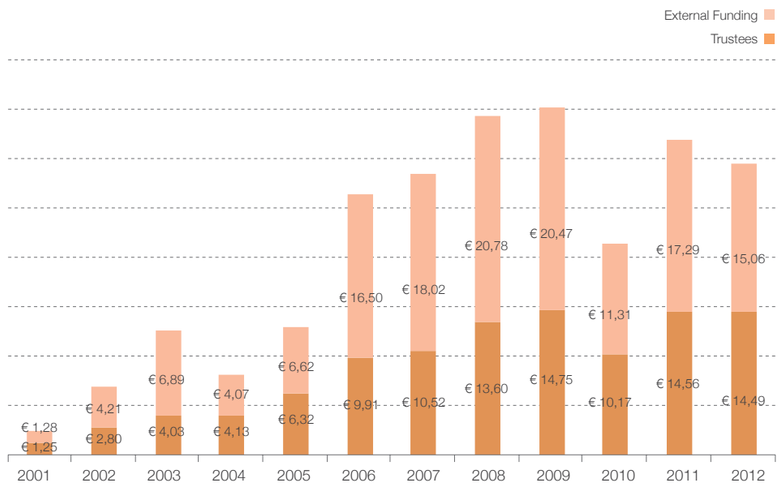
Additionally, since 2012 the CRG has been a member of the HR Strategy Group, promoted by the European Commission to support the implementation of the code of conduct that will lead to the award of excellence in HR research. To this end, the following action has been undertaken:

> Implementation of improvements within the CRG Equal Opportunity Plan.
> Action to integrate disabled people at the institute.
> Creation of the committee for conflict resolution, both personal and professional, where all categories from the centre are represented.
> A survey of values and action principles was conducted among all the CRG staff.
> Training for the management staff to acquire new skills and competences was organised focused on teamwork, interdepartmental relationships and communication.

Also, due to the economic environment and public funding cuts we have decided to set up an agreement with a Public Affairs company in order to increase the Public Affairs activity, to step up funding opportunities, strengthen relationships and enhance the CRG presence among the centres of power.

Finally, and to conclude, professionally it has been a very intense year for the CRG team, but at the same time it has also been very gratifying as we have fulfilled our goals.

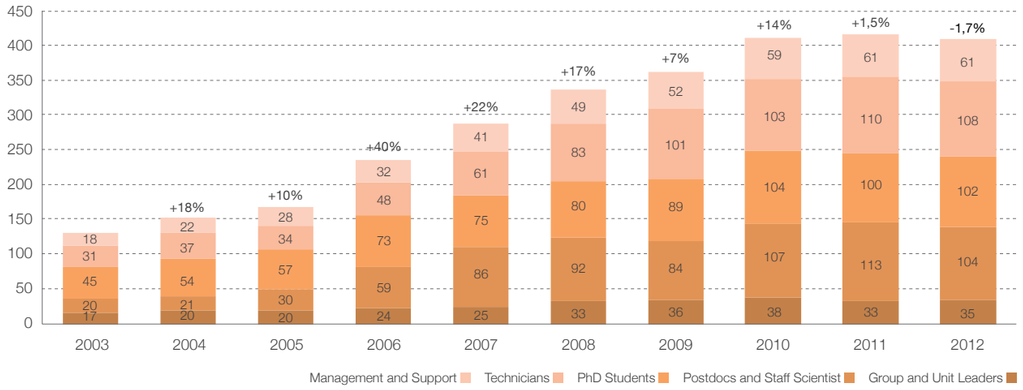
Funding evolution



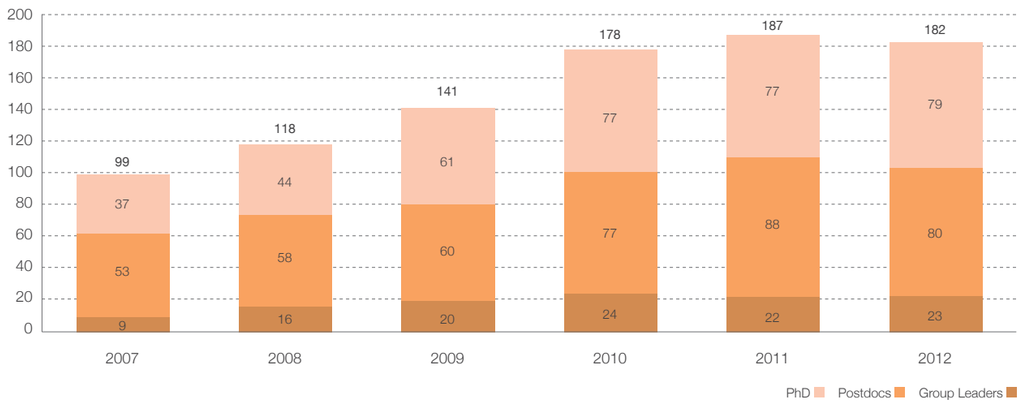
Personnel evolution

On 31st December 2012, 410 people from 39 countries were employed by the CRG.

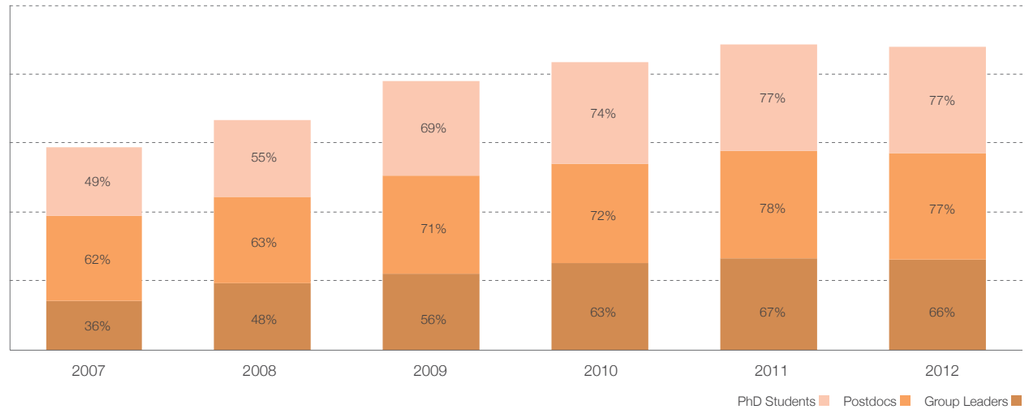
Personnel on 31st December 2012



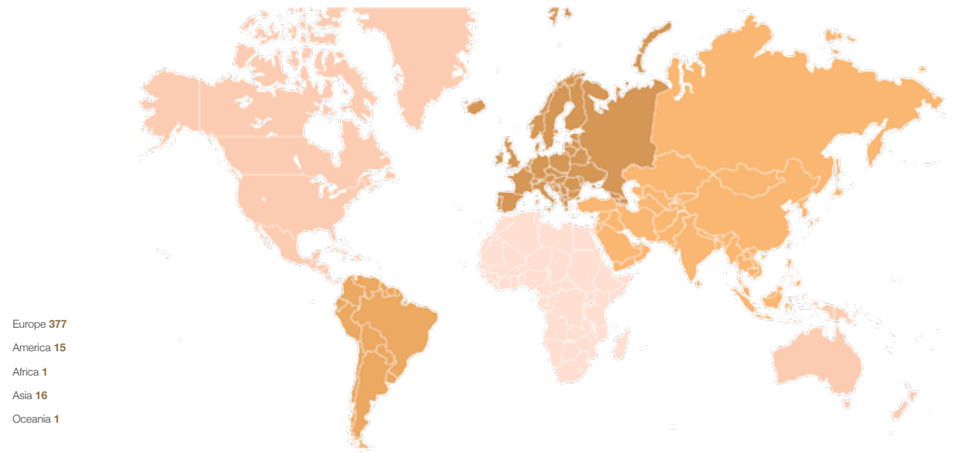
Foreign Researchers



Foreign Scientists Vs. Total



Nationalities



GRANTS AND INTERNATIONAL & SCIENTIFIC AFFAIRS

Grants & External Funding

In 2012, the CRG maintained a very successful track record in attracting competitive funding from highly prestigious funding agencies (€13.8 million), in spite of the significant decrease in national funds. The awards obtained from the European Commission (€9.4 million/10 projects, not including projects under negotiation) represent the largest share, accounting for over 68 % of the sum of external grants. With a total over €35.7 M and 55 projects (including projects under negotiation and ERA-net projects) for the period 2007-2012, the CRG is one of the leading Spanish institutes when it comes to attracting EU funds, currently occupying 12th position, only behind 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 11 grants (8 Starting and 3 Advanced Grants), according to recent statistics provided by the ERC.

Table 1. External Funding: Competitive grants

	2012		2011	
Funding Agency	**€000**	**%**	**€000**	**%**
EUROPEAN COMMISSION	9.436,12	68,4%	9.855,06	58,1%
MINISTERIO DE ECONOMIA Y COMPETITIVIDAD	2.979,81	21,6%	4.161,52	24,5%
AXA RESEARCH FUND	293,50	2,1%	0,00	0,0%
NATIONAL INSTITUTE OF HEALTH NIH	245,49	1,8%	0,00	0,0%
SWISS NATIONAL SCIENCE FOUNDATION	192,73	1,4%	153,13	0,9%
DECO	164,80	1,2%	7,00	0,0%
AGAUR - AGENCIA GESTIO D'AJUTS UNIVERSITARIS	135,64	1,0%	137,33	0,8%
HOWARD HUGHES MEDICAL INSTITUTE	0,00	0,0%	1.088,66	6,4%
EUROPEAN SCIENCE FOUNDATION	0,00	0,0%	609,25	3,6%
Other	356,27	2,6%	960,04	5,7%
TOTAL FUNDING	**13.804,35**	**100.0%**	**16.971,98**	**100,0%**
EUROPEAN COMMISSION	1.103,05		2.401,59	
MINISTERIO DE ECONOMIA Y COMPETITIVIDAD	4.005,00		0,00	
FUNDACIÓ MARATÓ DE TV3	491,88	n/a	0,00	n/a
OTHER	58,62		516,16	
TOTAL FUNDING (IN NEGOTIATIONS)	**5.658,54**		**2.917,75**	

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. Di Croce), 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 (€)
AWARDED IN 2012				
Uncovering and understanding RNA through Massively Parallel Sequencing (RNA-MAPS)	ERC Advanced Grants (AdG)	Guigó, Roderic	Individual	2.056,64
Spaciotemporal regulation of epidermal stem cells by circardian rhythms: impact on homeostasis and aging (STEMCLOCK)	ERC Starting Grants (StG)	Aznar-Benitah, Salvador	Individual	1.495,48
Biogenesis of lipid droplets and lipid homeostasis (DROPFAT)	ERC Starting Grants (StG)	Chaves Simoes de Carvalho, Pedro Nuno	Individual	1.475,28
The Role of Non-coding RNA in Protein Networks and Neurodegenerative Diseases (RIBOMYLOME)	ERC Starting Grants (StG)	Tartaglia, Gian Gaetano	Individual	1.465,35
Evolutionary genomics of long, non-coding RNAs (NONCODEVOL)	ERC Starting Grants (StG)	Gabaldón, Toni	Individual	1.302,11
The Human Early-Life Exposome – novel tools and methods for integrating early-life environmental exposures and impact on child health across Europe (HELIX)	Collaborative Project	Estivill, Xavier	Partner	898,65
Characterization of the role of DNA methylation, hydroxymethylation and TET proteins in progesterone-mediated signaling in breast cancer cells (Hydroxy-Methylation)	Marie Curie Intra-European Fellowships for Career Development (IEF)	Verde, Gaetano	Individual	176,05
Identification and manipulation of molecular pathways relevant for age-dependent tissue regeneration	AXA Research Projects	Cosma, Maria Pia	Partner	173,50
Synthetic gene regulatory networks for single-stripe gene expression (SynthStripe)	Marie Curie Intra-European Fellowships for Career Development (IEF)	Schaerli, Yolanda	Individual	168,90
Novel TNF inhibitors as therapeutic agents in autoimmune diseases, osteoporosis and cancer	Valorisation Grant (Prova't)	Serrano, Luis	Individual	157,80
Epigenetics towards systems biology (EpiGeneSys)	Network of Excellence	Filion, Guillaume	Partner	150,00
Infrastructure for Systems Biology - Europe (ISBE)	Collaborative Projects (Large-scale integrating project)	Sharpe, James	Individual	147,66
Imaging-based Systems Biology Analysis of Lymph Node Structure and Function Viral Infection	Collaborative project (Sinergia grant)	Sharpe, James	Partner	139,54
Effective Technology Transfer in Biotechnology (ETTBio)	INTERREG IVC	Rúbies, Xavier	Partner	138,65
Lanscape of Transcription in human and mouse	Collaborative Project (NIH)	Guigó, Roderic	Partner	115,74
IN NEGOTIATIONS/PENDING FINAL AWARD NOTIFICATION DURING 2012				
Excellence Centers Severo Ochoa	Institutional Project	Serrano, Luis	Individual	4.000,00
A theoretical framework for swarms of GRN-controlled agents which display adaptive tissue-like organisation (SWARM-ORGAN)	Collaborative Project	Sharpe, James	Coordinator	723,01
3D Gene Shape- Hearts and lims: Linking morphometrics with 3D analysis of gene expression patterns (3D Gene Shape)	Marie Curie International Incoming Fellowships (IIF)	Martínez, Neus	Individual	230,04
"MycoBiotics" as a new technology platform for the treatment of human respiratory andgenital tract diseases (Mico pLung)	ERC Proof of Concept	Serrano, Luis	Individual	150,00

In addition to attracting competitive grants for excellent basic research, the CRG has also succeeded in obtaining additional private funds for strategic CRG activities amounting to €1.2 million, which includes funding by "la Caixa" Foundation for the PhD Programme and valorization funds from Fundación Marcelino Botín, as detailed below.

Table 3. Other External Funding

Total external funding	2012		2011	
Type of external funding	**€000**	**%**	**€000**	**%**
Competitive Grants	**33,14**	**10,3%**	**107,81**	**9,0%**
European Commission (FP7)	9.436,12	62,7%	9.855,06	57,0%
National	3.393,10	22,5%	4.506,51	26,1%
International	975,13	6,5%	2.610,41	15,1%
Other external funding (private funding)	**1.219,19**	**8,1%**	**273,62**	**1,6%**
National	1.143,13	7,6%	70,51	0,4%
International	76,06	0,5%	203,11	1,2%
Other	**33,59**	**0,2%**	**48,10**	**0,3%**
TOTAL FUNDING	**321,72**	**100,0%**	**1.194,01**	**100,0%**

International & Scientific Affairs

During 2012 the CRG consolidated and expanded its network of scientific collaboration and partnerships, in Spain, Europe and the world.

Strategic institutional collaborations

Firstly, to promote translational research the institute has strengthened its partnerships with research intensive hospitals, through the organisation of a series of joint workshops and scientific talks; and the signing of collaboration agreements.

Notably, scientific collaborations with the Vall d´Hebron Research Institute (VHIR) are successful and diverse, and include research on diseases such as familial Chiari, congenital myasthenia and congenital ataxia; Erwin sarcoma and human squamous cell carcinomas (SCC); as well as Parkinson's Disease.

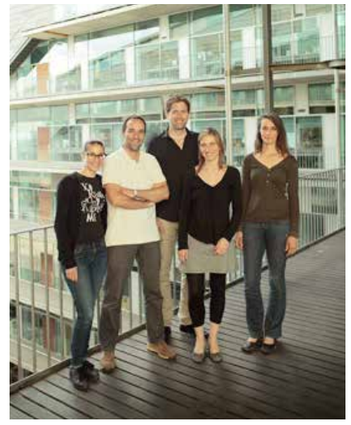
Secondly, the CRG has developed strong interdisciplinary collaborations with other institutes in Barcelona and its surroundings; such as the Institute of Chemical Research of Catalonia (ICIQ), through the organisation of a joint meeting to explore synergies between biology and chemistry; and the Barcelona Supercomputing Centre (BSC), through three encounters to exchange good research/project management practices as well as explore scientific collaboration to link biology with computational science.

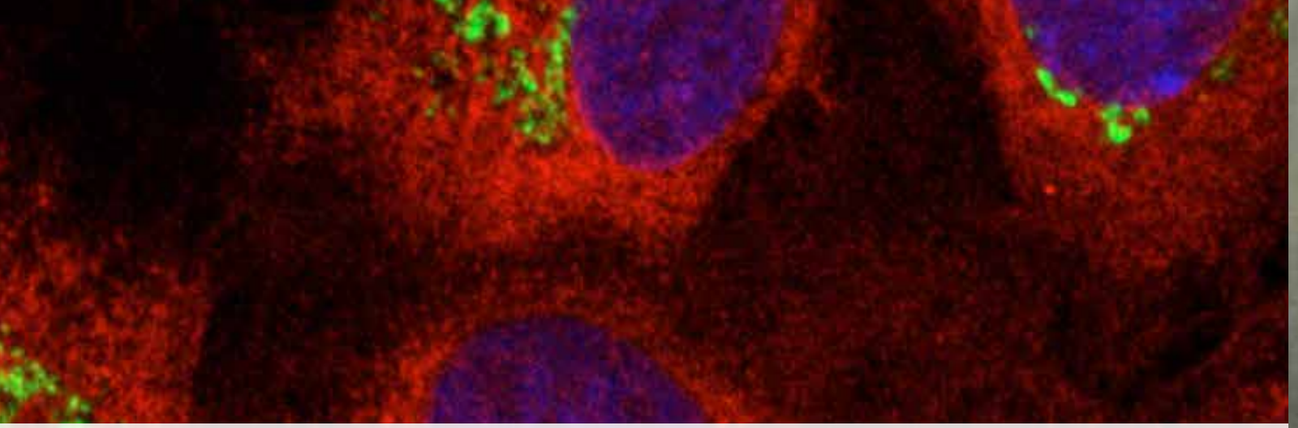
Thirdly, strategic scientific collaborations with other institutes both in Europe and further afield have been consolidated and explored; including the organisation of a joint workskhop with scientists from the Institute of Human Genetics (IGH) in Montpellier, France; the visit of a CRG delegation to Academia Sinica in Taiwan; the signing of a collaboration agreement with the Riken Centre for Developmental Biology (CDB) in Japan; and the planning of a joint exchange programme with Wits University/Sydney Brenner Institute for Molecular Biosciences in South Africa, sponsored by Novartis.

Finally, the CRG has led the establishment of **EU-LIFE**, a European Strategic Alliance of Life Sciences Institutes for Excellence. The objectives of this alliance, coordinated by the CRG, are to promote excellence in research, to promote a better integration among European research life sciences institutes, and to develop and share best practices in research, research management and training. The founding partners of this alliance include the VIB (Belgium), Institut Curie (France), the Gulbenkian Foundation (Portugal), the European Institute of Oncology (Italy), the Max Delbrück centre for Molecular Medicine (Germany), and the CeMM (Austria).

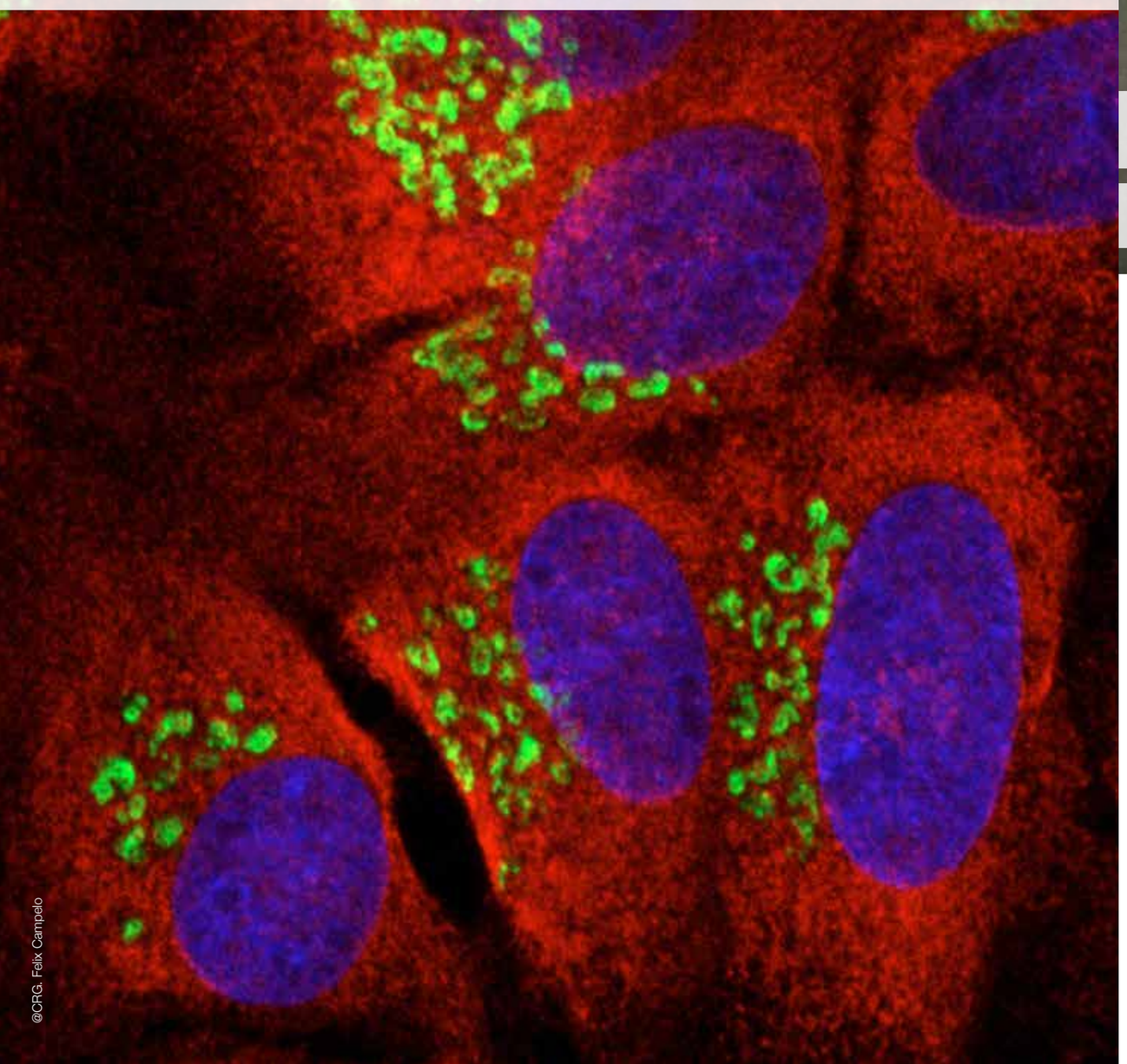
Coordination of collaborative scientific projects

The CRG is leading several Spanish and large European collaborative projects; and is therefore contributing to advancing knowledge in diverse fields of biology and biomedical research, ranging from systems biology, cancer, epigenetics, cell trafficking, and rare diseases to medical genomics. The current portfolio includes four Spanish Consolider projects and seven European projects. Three projects started in 2012, and one was awarded in 2012 (see section on "External funding").





SCIENTIFIC HIGHLIGHTS



@CRG, Felix Campelo

UNMASKING CANCER

The Chromatin and Gene Expression lab at the CRG reveals that the PARP-1 enzyme activated by kinase CDK2 is necessary for the hormonal induction of the genes responsible for the proliferation of abnormal cells in breast cancer

It is one of the most commonly heard words and engenders much respect and fear when it comes to health. Cancer. And the fact is, at present, it is one of the main causes of death around the world, according to the World Health Organisation (WHO). The term “cancer” involves a group of diseases that can affect any part of the body, and which are characterised by a rapid proliferation of abnormal cells that extend far beyond their usual limits and which can even invade adjacent areas (metastasis).

Fortunately, cancer is also one of the most studied diseases. Advances in genomic techniques have allowed researchers begin to understand the cellular and molecular mechanisms that are disrupted in cancer cells. “It is like a constantly moving jigsaw puzzle”, says Miguel Beato, director of the Chromatin and Gene Expression lab at the CRG.

This group of scientists is responsible for studying how steroid hormones activate the proliferation of cancer cells, particularly in breast cancer. Breast cancer cells are sensitive to oestrogen and progesterone -hormones involved in the menstrual cycle, pregnancy and embryo development. If the relevant receivers are blocked, which is what happens in current treatments, many tumours stop growing, but over time they develop resistance.

Therefore, more knowledge about the mechanisms that activate the proliferation of tumour cells is needed in order to inhibit them. In this regard, it is known that steroid hormones are able to activate or repress genes in cells. To do this, they pass through the cell membrane and bind to the receptor. Hormone and receptor reach the cell nucleus and attach to DNA sequences; this interaction dictates which genes

to activate or inhibit. But it is not so simple, because the DNA is packaged into chromatin (a combination of DNA, histones, and other proteins located in the nucleus of eukaryotic cells), which acts as a barrier. The steroid hormone, via its receptor, activates various enzymes that initiate the opening of chromatin.

In September 2012, the scientists of the Chromatin and Gene Expression lab at the CRG published a study in the journal Genes & Development, which spoke of the important role of a little-studied enzyme in this context, poly-ADP-ribose polymerase-1. PARP-1 (its abbreviation) belongs to a family of enzymes that are involved in the regulation of many cellular processes, but which are best known for participating in the repair of cuts in DNA. They synthesise a third nucleic acid, PAR, which has a similar shape to that of a fir tree with many branches.

“Until now, it wasn’t known that steroid hormones activated PARP-1”, explains Beato. The researchers have discovered that it is the hormonal activation of the CDK2 kinase, which phosphorylates and activates PARP-1, that modifies the histone H1 (histones are proteins on which the DNA is wound to stay compacted) and the displacement of the chromatin. This is how it gains access to the information contained in the gene.

According to the first author of the study, Roni Wright, a postdoctoral researcher in the CRG laboratory, “If PARP-1 did not act, many of the progesterone target genes would not be activated”. “What Roni saw”, states Beato, “is that 5 minutes after adding hormone to the cells, the nucleus is filled with this PAR nucleic acid. And that grabbed our attention, because we had never seen such a dramatic consequence of any other hormonal effect.”

Thanks to the results of this work, which has been the subject of reviews and editorials in various publications, we are a little closer to understanding the mechanism by which hormones stimulate the growth of certain tumours. What this really means is that this knowledge can be used to design pharmacological strategies to control the development of hormone-dependent cancers, such as breast or prostate cancer.

Reference work:

Wright RH, Castellano G, Bonet J, Le Dily F, Font-Mateu J, Ballaré C, Nacht AS, Soronellas D, Oliva B, Beato M.
“CDK2-dependent activation of PARP-1 is required for hormonal gene regulation in breast cancer cells.”
Genes Dev, 26(17):1972-83 (2012)

SCIENTIFIC HIGHLIGHTS

Gene Regulation, Stem Cells and Cancer

A new study by CRG Research Group helps to explain how aging occurs and how this may lead to the development of cancer

WHEN AGING IS A PROTECTING STRATEGY

“Forever young, I want to be forever young […] It’s so hard to get old without a cause/ I don’t want to perish like a fading horse /”, used to sing the German band Alphaville in the 80s, when that song became almost a hymn. And they were not the only ones to dream about eternal youth.

Since Ancient times, humans have tried to find magical potions to stay young. It is said Cleopatra used to bath in female donkey’s milk to preserve her skin from aging. At 19th century, European aristocrats, kings and queens used to eat powder obtained from battering Egyptian mummies believing it would heal them from any malady and prevent them from getting old. And Oscar Wilde well depicted the infatuation of society of youth and beauty in *The Picture of Dorian Gray*.

Nevertheless, despite the efforts of art and also science, the formula for eternal youth is still unknown. In fact, we do not even fully understand the biological reasons behind the very complex process of getting old, a period most of us fear, because it normally involves the loss of functional capacity over time and also the appearance of many diseases.

Why and how we age is still, even today, a scientific mystery. But a new study from researchers at the CRG, published in *Genes & Development*, may have gained some insight in this question and found an important clue to understand how the process occurs and how it might facilitate diseases such as cancer to develop. They have based their research on skin, as it is one of the most obvious tissues to undergo aging.

More but less efficient

As it happens in the rest of our body, skin is constantly renovating. Dead or damaged cells are replaced by healthy new ones. This lifelong process is possible due to stem cells, which have a unique ability to grow, divide and differentiate into fully functioning specialized cells, giving rise to the whole tissue; they also withstand stress and damage better than other types of cells.

However, as we grow older, this procedure seems to start failing: wrinkles appear, we lose hair growth, stains appear, and our skin starts getting thinner and is less capable to heal wounds as fast as before. But, why? Do skin stem cells stop doing their job? Do they begin making mistakes?

These are the questions researchers at CRG, led by Bill Keyes, Head of Mechanisms of Cancer and Aging Group, posed themselves and served as a starting point for their investigation. Using a particular mouse model, they found for the first time that major changes occur in these stem cells during aging, causing them to actually lose their ability to function properly. "It was known to happen in other systems like in blood or brain, but not in the skin", Keyes points out.

The research group looked at the stem cells and saw their number increased during aging. "We tested their capacity and found that it decreased, which suggested that something was wrong" Keyes explains. "If there were more stem cells, it might be because they were not functioning properly, and were produced in higher numbers to try to compensate".

These findings strongly support the idea that changes in stem cell performance may be actually driving the aging process. Also, they open the door for future studies that may help to alleviate aspects of the growing older procedure.

A balance matter

CRG Researchers went further and managed to link the aging process with diseases such as cancer. They demonstrated that the entire skin changes and produces many different proteins that mediate inflammation, which is a cellular response to stress or damage; when inflammation happens in low levels, it is involved in repairing tissue damage, whereas when it is produced both in chronic long term or in high levels, it becomes a serious problem that can lead to the proliferation of a tumour.

Keyes' group found out that when proteins dealing with inflammation are abnormally produced, that contributes to the decline of stem cell function. "What we found was that the signature within the stem cell suggested it is surrounded by inflammation, and harbors pro-tumorigenic alterations. So it is possible that the aged tissue may be already on its way to becoming a tumour", Keyes declares.

If the aged tissue is almost becoming a cancer, then it is possibly protecting itself by inhibiting from growing and dividing, which, in turn, can lead to a tumour. But when stem cells stop growing, they cannot do their normal function of repairing the damaged tissue; so, in a way, protection from cancer could be causing aging.

In an effort to make it clear, Bill Keyes compares this vicious cycle with a typical market scales: "Imagine the following analogy: you have a weighing scales, on one side you have pieces representing cancer, and on the other side, other ones representing aging. When you have a lot of tumour on one plate, then there is less cellular aging on the other plate. Whereas if you have too much aging, then this might be as a result of protection from cancer. Then anything that causes a deregulation of normal processes such as stem cell behaviour may tip the scales in one direction or the other".

Altogether, these findings help to explain what is likely a major cause of the aging process and how this develops. Moreover, having identified inflammation as a cause of stem cell dysfunction uncovers likely causes in the development of general cancer, not just skin cancer. CRG researchers suspect that if they can figure out what's working in skin, it may be applicable to other tissues such as intestine, muscle or heart. "If we can make each tissue function better, there would be less cancer and people could age healthier. The goal is not to increase lifespan, but healthspan".

Reference work:
Doles J, Storer M, Cozzuto L, Roma G, Keyes WM.
"Age-associated inflammation inhibits epidermal stem cell function."
Genes Dev, 26(19):2144-53 (2012).

IN THE BOWELS OF HUNTINGTON'S DISEASE

The Genetic Causes and Diseases research group at the CRG has discovered the key importance of RNA in the pathogenesis of this neurodegenerative disease

Spasmodic and uncontrollable body movements, difficulties in swallowing, loss of reasoning capacity and memory, language changes, even personality disorders: these are the most common symptoms Huntington's disease (also called Huntington's chorea and previously known as St. Vitus dance), a neurodegenerative disease that affects the basal ganglia of the brain and that appears, on average, at around 40 years of age.

When someone is diagnosed with Huntington's disease, they are being handed a death sentence. And, despite the fact that it was first described in 1872 by the North American doctor George Huntington and that its genetic base has been known for 20 years, there is still no cure. The patient passes away between 10 and 15 years after the first symptoms are triggered. Medicines are not efficient in stopping the progression of the disorder, they only serve to reduce or diminish the symptoms. Now, scientists at the CRG have made a discovery that will enable an advance in the development of treatments for this rare disease.

The gene implicated in Huntington's chorea is HTT (Huntingtin), located on the short arm of chromosome 4. This gene has three repeated nucleotides (cytosine, adenine and guanine: CAG), that code for the synthesis of the amino acid glutamine in the composition of the Huntingtin protein.

The genetic code of gene HTT contains many repetitions of this triplet, generally between 10 and 28 in healthy people. But a person with Huntington's has between 36 and 120 CAG repeats. This means it has more glutamines than it should. There is a very clear correlation between the number of CAG repetitions, the age of onset of the disease and the seriousness of the symptoms. This means the more

CAG repeats, the greater the probability of developing this neurodegenerative disorder at a younger age. Over last the 20 years, the scientists who have studied the mechanism by which the disease is produced have seen that “glutamine, in excess, has a toxic effect on the cell that causes it to generate apoptosis (programmed cell death) in some nuclei of the brain, producing neurodegeneration in these regions”, explains Xavier Estivill, head of the Genetic Causes and Disease research group at the CRG.

Although up to now efforts had been centred on blocking this anomalous protein, the CRG researchers began, some years ago, to look into whether the ribonucleic acid (RNA) might play an important role in the pathogenesis of Huntington’s disease. And it does. A paper by Eulàlia Martí, Mónica Báñez-Coronel and Estivill himself, in the journal *PLoS Genetics*, provides evidence of the toxicity of the RNA for the cell, confirming that it was the cause of the neurodegeneration.

“We have found that the abnormal RNA, by itself, without manufacturing protein, gives toxicity to the cells. And in addition, when we eliminate this abnormal RNA, we take away the toxicity as well”. The scientists have demonstrated this in both samples from the brains of patients affected with Huntington’s and mouse models of the disease.

Estivill likens the phenomenon to what in everyday life would be the expansion and contraction of an accordion, or the typical slinky that kids play with. “We have seen that in all cases the machinery of micro RNA, which is RNA that has a regulating function in the expression of the genes of our genome, is altered. Toxic micro RNA is produced that contains CAG repetitions. However, when we block this expansion, we eliminate the toxicity. This is a revolution, a huge change, with respect to what was known up to now”.

Since, in 1993, the mutation causing Huntington’s chorea was identified by several research groups united in an international consortium, it has been accepted and demonstrated that the abnormal protein is toxic. The CRG group is the first to be able to demonstrate that the RNA, with the expansion, is sufficient to trigger neurodegenerative changes.

The discovery will allow, from now on, concrete therapeutic solutions to be found for this hereditary disease (the descendants of a patient have a 50% probability of inheriting the mutation and developing the disorder) which, very slowly, over a 15 to 20 year period, produces cognitive, psychiatric and motor alterations. The discovery is also useful for the study other neurodegenerative diseases that have the same sequence of CAG expansion, but which affect different genes, resulting in lethal consequences in the neuronal physiology of the person, such as, for example, inherited dominant autosomal ataxia. The researchers are also exploring other nucleotide expansions associated to other diseases.

Reference work:

Bañez-Coronel M, Porta S, Kagerbauer B, Mateu-Huertas E, Pantano L, Ferrer I, Guzmán M, Estivill X, Martí E.
“A pathogenic mechanism in Huntington’s disease involves small CAG-repeated RNAs with neurotoxic activity.”
PLoS Genet, 8(2):e1002481 (2012).

SCIENTIFIC HIGHLIGHTS

Bioinformatics and Genomics

THE IMPORTANCE OF JUNK DNA

Project ENCODE, propelled by more than 400 researchers worldwide, including a score from the CRG, reveals the usefulness of so-called “junk DNA” in the human genome

In April 2000, the announcement of a first draft of the human genome sequence was to mark the beginning of a revolution in the field of biomedicine. Scientists worldwide, united in the Human Genome Project, financed by the Department of Energy and the US National Institutes of Health, and equipped with a budget of 280 million dollars annually, had taken only 10 years to map that which makes us unique. Our knowledge of diseases has become much more exhaustive since then.

Currently, one of the most important international research projects on genomics is ENCODE (Encyclopaedia of DNA Elements). Led by the National Genome Research Institute (NHGRI) in the United States and the EMBL-European Bioinformatics Institute (EMBL-EBI) in the UK, the aim is to decipher the instructions coded into the human genome and understand how they are interpreted in different ways in different cell types.

After more than 1,600 experiments carried out since the pilot phase began in 2003, the researchers have discovered that the so-called “junk DNA” (the DNA that does not code for proteins) has an important role. Until now, it had been thought that a large part of this DNA had no biological function, that the instructions for us were written in a very restricted part of the genome, somewhere between 10 and 15%, and that the rest was useless. But the ENCODE project has revealed that some of this remaining part also encodes these instructions.

Roderic Guigó, coordinator of the Bioinformatics and Genomics at the CRG and lecturer at Pompeu Fabra University (UPF), explains that “junk DNA” is like a huge electrical control panel with millions of switches that regulate the activity of our genes. The genes are regions of the genome that determine our

biological characteristics: the colour of our eyes, our height, muscle mass, *etc*. The genome is identical in every cell of our body; nevertheless, the cells specialise, adopt different morphologies and functions depending on where they are: in the brain, the skin, the lungs, for example. And it is precisely because the genes, within these regions, work or do not work.

Returning to the analogy of the electrical control panel, it is as if the lights this panel works could be switched on or off. If we go to the lights for the bones, there will be thousands lit; the same if we look at the cells of the muscles. It is the combination of these lights that causes the genome to work in a different way for each cell type. Controlling whether the lights are on or off are "switches", without which the genes would not work and mutations in these regions would cause disease.

"The expression of a gene, a light, does not depend on any single switch, but rather a combination of switches that, in addition, cause a light to be more or less intense", states Guigó. We could draw a parallel with the situation that takes place when we have two switches in the living room to turn the same light on or off, for example; sometimes, when you turn off one side, if the other is not on, it does not stay lit. "There must be a combination of several switches so that, depending on what position they are in, they produce a more or less intense light or are even turned off".

There are more than 400 scientists worldwide that work jointly on the project, and who have produced 30 interconnected open-access articles published in the journals *Nature, Genome Biology* and *Genome Research*.

Around twenty of these researchers are from the CRG, and they coordinate their work with scientists from California, the UK, Switzerland, Singapore and Japan. They focus their efforts on the analysis of the transcriptional activity of the genome. The extension of the instructions coded for in the genome is triggered by the transcription of DNA into messenger RNA, which is later translated into protein, and proteins are what perform the main biological functions. "In particular we have studied where in the genome the RNA is produced and what type it is. RNA biology will be increasingly crucial for basic research and for technical applications in biology and medicine in particular", concludes the coordinator of the Bioinformatics and Genomics programme at the CRG.

The ENCODE project, with an investment of over 185 million dollars, entered its third phase in October 2012, which is due to last a further four years. Up to that point, four million gene "switches" had been identified, which will help to locate very specific areas of human diseases. The objective, in the long term, is to explore the entire cell diversity.

Reference work:

The ENCODE Project Consortium, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M.
"An integrated encyclopedia of DNA elements in the human genome."
Nature, 489(7414):57-74 (2012).

SCIENTIFIC HIGHLIGHTS

Cell and Developmental Biology

UNDERSTANDING THE NORMAL FUNCTION OF OUR INTERNAL POST OFFICE

Scientists at the CRG discover the mechanism by which the 'post office' (the Golgi apparatus) in our cells sends mail (lipids and proteins) by couriers (transport carriers) to their respective destinations.

Cells are factories that build basic ingredients and release them into the extracellular space so our bodies can work. The list of important ingredients, such as proteins and lipids, includes collagen for cell-cell attachment and the mineralisation of bones, insulin for regulating carbohydrate and fat metabolism, neurotransmitters, hormones, growth factors, and many others. How is this sorting and delivery performed by a cell? Vivek Malhotra, a leader in the study of cell compartmentation and protein secretion, has been trying to understand this question for over two decades.

Each cell contains an organelle -the Golgi apparatus- that is responsible for sorting and transporting proteins to various other organelles within the cell, and releasing (or secreting) them in order to control key functions outside the cell. In a way, the function of the Golgi apparatus is similar to that of a post office, but instead of sorting mail and packages, the Golgi apparatus sorts proteins, and instead of using couriers and vehicles, it employs membrane-bounded transport carriers, or vesicles, to deliver the proteins to their respective destinations.

Previous work carried out by the group revealed the importance of a kinase called PKD, which has a number of Golgi membrane specific targets that control the levels of a number of different lipids. In that paper they explained how they directly tested the requirement of a lipid called Sphingomyelin (SM) in the transport of cargo from the Golgi apparatus. SM is produced from ceramide and binds cholesterol.

In a new paper published in the EMBO journal, Malhotra's group report that the concentration of SM in the Golgi apparatus determines the ability of this compartment to generate transport carriers. Malhotra

and his colleagues devised a method to control the lipid levels of the Golgi membranes and found that the ability to generate transport carriers was directly dependent on the concentration of SM. Reducing SM levels inhibited the biogenesis of transport carriers and blocked protein secretion. These new findings will also help in resolving a major controversy in the field of membrane biology. SM assembles with cholesterol into specific domains called rafts that are proposed to regulate key signalling events. However, it has not yet been possible to see rafts in the membranes and their function is unclear. The findings of Malhotra's group suggest the existence of SM-enriched domains in the Golgi membranes and more importantly reveal their function in vesicle biogenesis.

Above all, these new findings will help in the understanding of the mechanism by which the post office of the cell (the Golgi apparatus) generates specific couriers (transport vesicles) to transport mail (lipids and proteins) to their respective destinations both internally and to the exterior of the cell.

Reference work:

Duran JM, Campelo F, van Galen J, Sachsenheimer T, Sot J, Egorov MV, Rentero C, Enrich C, Polishchuk RS, Goñi FM, Brügger B, Wieland F, Malhotra V.
"Sphingomyelin organization is required for vesicle biogenesis at the Golgi complex."
EMBO J, 31(24):4535-46 (2012).

SCIENTIFIC HIGHLIGHTS

Systems Biology

A new study by a CRG Group sheds light on how fingers are formed and demonstrates the process may be explained by a Turing-type patterning mechanism

WHY IT'S EASIER TO HAVE AN EXTRA FINGER THAN AN EXTRA ARM

Sometimes, having an extra pair of arms would be really useful. Imagine how many things you could do at the same time! Think about dinner, when you are feeding the baby while trying to prepare a Spanish Omelette and the phone rings –it's your mum, you have to pick it up - and accidentally you spill a glass of milk. What a mess! You get the point - four arms would surely be more practical than just two. Like those Buddhist gods and goddess, who have two, three or even four pair of hands!

But, although you might imagine many advantages, unfortunately Nature has only equipped us with two upper limbs. Probably because, apart from that comical kitchen scenario, having more than a pair of arms could pose an evolutionary disadvantage (imagine trying to balance or to walk straight?); and it's the same for the head or legs.

However, people with a pair of extra fingers in their hands or feet are quite common. The actress Halle Berry has one extra toe on each foot, and it is said the English Queen Anne Boleyn had 11 fingers. Not so rare: it affects one in 500 live births. But, why? How is it we can have multiple fingers while it is quite rare to get extra arms, or more than one head ?

A new study by the CRG Multicellular Systems Biology Research Group in collaboration with CSIC and published in *Science* magazine sheds light on this question and has gained some insight in how fingers are formed. According to this investigation, a British Mathematician, Alan Turing, may have the answer.

Like the sand in the desert

Most of our body is set up in a very hierarchical way based on *positional information*; when the embryo is being formed in the womb, cells know what they have to do due to chemicals, like gene products, which are active in different areas. Each cell detects different concentrations of those chemicals and knows whether it must start making the head, or an arm.

For a long time, this principle based on positional information was believed to be involved everywhere in the body. But recently it started to be suspected there were some places where it does not work like that. And the fingers was one of those exceptions. The paper published by CRG research group contributes to the belief that control of how many fingers you have, is not controlled by positional information, but by a local spontaneous self-organising cellular system.

Have you ever looked at flocks of starlings flying in the sky? They form incredible patterns, move all at once, and each individual knows what to do just looking at the other birds flying around. There's no one bird in control – there's no boss. The self-organising cellular systems in your embryonic hand resemble those flocks of birds in the sense that there is no one in command - they form their pattern spontaneously. But there is also a difference from the birds, as your finger cells always form an alternating pattern: digit, non digit, digit, non digit, etc. The very same thing happens with the desert's dunes; due to wind interaction with sand particles, ridges are formed; they are perfectly periodic patterns, like, also, stripes on zebras or tigers. Finger formation seems to be a similar interaction between the genes and the cells; they start with a uniform arrangement and then spontaneously start forming a pattern. And the first person who discovered how these patterns form in biological systems was the British visionary Alan Turing; in 1952 he formulated a mathematical model for the formation of repeated structures in biological systems.

Why you cannot have two heads

In collaboration with a CSIC group based in Santader using mouse mutants, the CRG researchers worked on a computer model in order to gather stronger evidence relating the hand question to the Turing system. "The thing about the Turing system is that the wavelength of the pattern should be fixed all the way across, meaning that your fingers would all be roughly the same width. This does not apply later when you are an adult, but when the pattern is first formed, as an embryo. And at that point, the width of each finger is exactly the same", James Sharpe, Multicellular Systems Biology Group Leader and corresponding author of the paper, points out.

The research demonstrated for the first time that mutations exist which change the wavelength of the fingers across the whole hand. They found there is a particular group of genes, called Hox genes, which among their many roles appear to control the wavelength – how thick the fingers and spaces between them are. "The contribution from our lab was quantifying these data, measuring how the wavelengths change in all these mutants, and then exploring whether these mutant digital arrangements can be explained by a computer simulation implementing this Turing type of system", Sharpe explains.

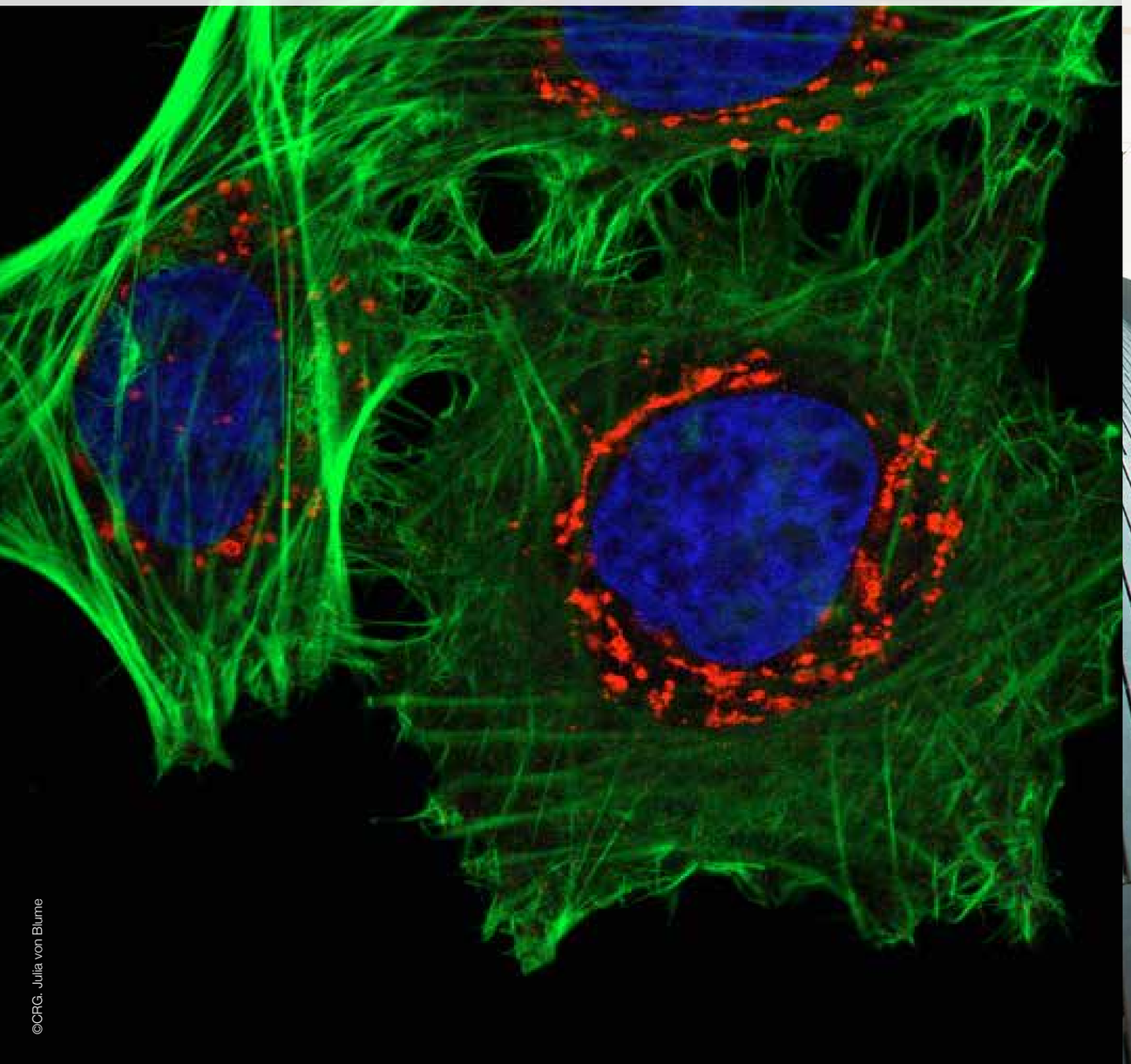
The good news is the computer model developed by the CRG group can indeed explain these results and therefore strengthens the idea that this pattern forming process is a Turing type of mechanism. "This is why you can have 6 or 7 fingers by accident. If each finger had its own position, as it is the case of head or arms, there would not exist a 'position six' for example. So you could not have a sixth finger".

The ultimate aim of CRG group is to get a better understanding of how any muticellular system works. In Jame Sharpe's words, "One of the long term reasons for trying to comprehend all this is organ regeneration. Our research is related to basic principles about how cells know where they are and organise themselves and build something together. We choose the limbs because they are easy to study, but the knowledge gained should be relevant to the whole body".

Reference work:

Sheth R, Marcon L, Bastida MF, Junco M, Quintana L, Dahn R, Kmita M, Sharpe J, Ros MA.
"Hox genes regulate digit patterning by controlling the wavelength of a Turing-type mechanism"
Science, 338(6113):1476-80 (2012).

SCIENTIFIC REPORT



©CRG. Julia von Blume

GENE REGULATION, STEM CELLS AND CANCER

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 (until November 2012), Elias Bechara, Panagiotis Papasaikas Lorena Zubovic (from November 2012)

Students: Felice Alessio Bava (joint student with the group of Raúl Méndez, IRB, Barcelona, until November 2012), Camilla Ianonne, Elena Martín (since October 2012), Juan Ramón Tejedor, Luisa Vigevani

Technicians: Belén Miñana, Vasiliki Michaki, Anna Ribó, Diana Stork

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. 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? Using *in vitro* and *in vivo* depletion, as well as reconstitution assays using purified components, we identified 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. 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. 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.

2. CPEB1 coordinates alternative 3'UTR formation with translational regulation (collaboration with the groups of Raúl Méndez, IRB, Barcelona and of Roderic Guigó, CRG)

Cytoplasmic Polyadenylation Element Binding protein 1 (CPEB1) is an RNA-binding protein that regulates mRNA polyadenylation in the cell cytoplasm and consequently mRNA translation. We have found that CPEB1 shuttles to the nucleus, where it co-localizes with splicing factors and regyulates alternative splicing and alternative polyadenylation. CPEB1 binding to pre-mRNAs modulates alternative splicing by preventing U2AF65 recruitment. Regulation of polyadenylation allows CPEB1 to mediate shortening of the 3' Untranslated Region (UTR) of hundreds of mRNAs, thereby modulating their translation efficiency in the cytoplasm. CPEB1-mediated 3'UTR shortening correlates with cell proliferation and tumorigenesis.

3. Co-option of the piRNA pathway for germline-specific regulation of *C. elegans* TOR expression and alternative splicing (in collaboration with the groups of Ben Lehner, CRG, and of Julian Ceron, IDIBELL, Barcelona)

Many eukaryotic genes contain embedded antisense transcripts and repetitive sequences that are widely assumed to represent accumulated 'genetic junk' without functional consequences. Using the nematode *C. elegans* as a model system, we have found that such elements can be regulatory precisely because they are recognized as 'non-self' RNA by piRNAs and Argonaute proteins. In particular, activation of a 3' splice site of Target of Rapamycin pre-mRNA (CeTOR, let-363) in the male germline of *C. elegans* coincides with male germline-specific expression of an upstream antisense transcript and with accumulation of endo-siRNAs against a Helitron transposon located within the first intron of the antisense gene. Mutation of the germline-specific piwi homolog prg-1 prevents the generation of endo-siRNAs, affects CeTOR mRNA levels and alternative 3' splice site selection and compromises fertility. Conversely, mutation of another Argonaute protein, csr-1, involved in licensing of endogenous transcripts, has opposite effects on the accumulation of endo-siRNAs and splice site selection (Figure 1). Consistent with a physiological function for endo-siRNAs in splice site selection, alternative splicing regulation and fertility are partially restored in prg-1 mutants by exogenous supply of double stranded RNA corresponding to the CeTOR intron that overlaps with the antisense transcript. We propose that epigenetic recognition of a 'non-self' intronic transposon by the piRNA system provides important physiological regulation of a host gene and may offer a general mechanism by which internal transcripts influence the evolution of gene expression.

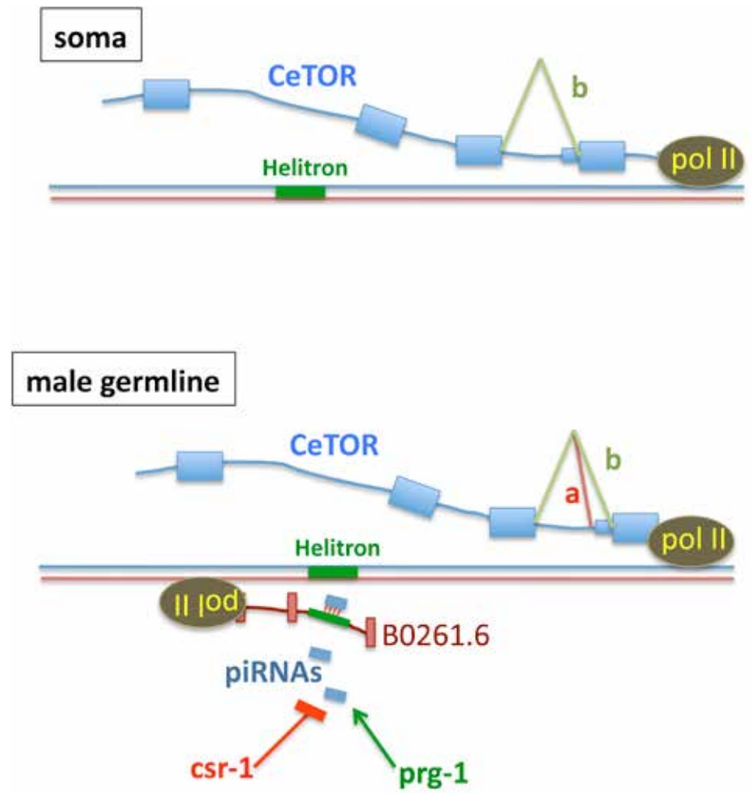


Figure 1.
Model for male germline-specific regulation of CeTOR alternative splicing by endo-siRNAs. *In somatic tissues, CeTOR pre-mRNAs are spliced to 3' splice site b. Expression of the B026.1 antisense transcript during spermatogenesis offers sequences complementary to piRNAs directed against a Helitron transposon located within the first intron of the antisense transcript. The antagonistic effects of Argonaute proteins Prg-1 and Csr-1 influence piRNA synthesis and/or function, which in turn modulate CeTOR expression and activation of 3' splice site a.*

PUBLICATIONS

Tavanez JP, Madl T, Kooshapur H, Sattler M and Valcárcel J
"hnRNP A1 proofreads 3' splice site recognition by U2AF."
Molecular Cell, 45:314-329 (2012).

Bonnal S, Vigevani L and Valcárce, J
"The splicesosome as a target of antitumour drugs."
Nature Reviews in Drug Discovery, 11:847-859 (2012).

Papasaikas P and Valcárcel J.
"Evolution. Splicing in 4D."
Science, 338:1547-1548 (2012).

Ramsay AJ, Rodríguez D, Villamor N, Kwarciak A, Tejedor JR, Valcárcel J, López-Guillermo A, Martínez-Trillos A, Puente XS, Campo E, López-Otín C and Quesada V.
"Frequent somatic mutations in components of the RNA processing machinery in chronic lymphocytic leukemia."
Leukemia. 2012 Nov 28. doi: 10.1038/leu.2012.344. [Epub ahead of print]

Auboeuf D, Carmo-Fonseca M, Valcarcel J, Biamonti G.
"Alternative splicing and cancer."
J Nucleic Acids, 2012:363809. Epub 2012 May 20.

Kulis M, Heath S, Bibikova M, Queirós AC, Navarro A, Clot G, Martínez-Trillos A, Castellano G, Brun-Heath I, Pinyol M, Barberán-Soler S, Papasaikas P, Jares P, Beà S, Rico D, Ecker S, Rubio M, Royo R, Ho V, Klotzle B, Hernández L, Conde L, López-Guerra M, Colomer D, Villamor N, Aymerich M, Rozman M, Bayes M, Gut M, Gelpí JL, Orozco M, Fan JB, Quesada V, Puente XS, Pisano DG, Valencia A, López-Guillermo A, Gut I, López-Otín C, Campo E, Martín-Subero JI.
"Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia."
Nat Genet, 44(11):1236-42 (2012).

Bava F-A, Eliscovich C, Ferreira PG, Miñana B, Ben-Dov C, Guigó R, Valcárcel J and Méndez R.
"CPEB1 coordinates alternative 3' UTR formation with translational regulation."
Nature, in press.

Reviews

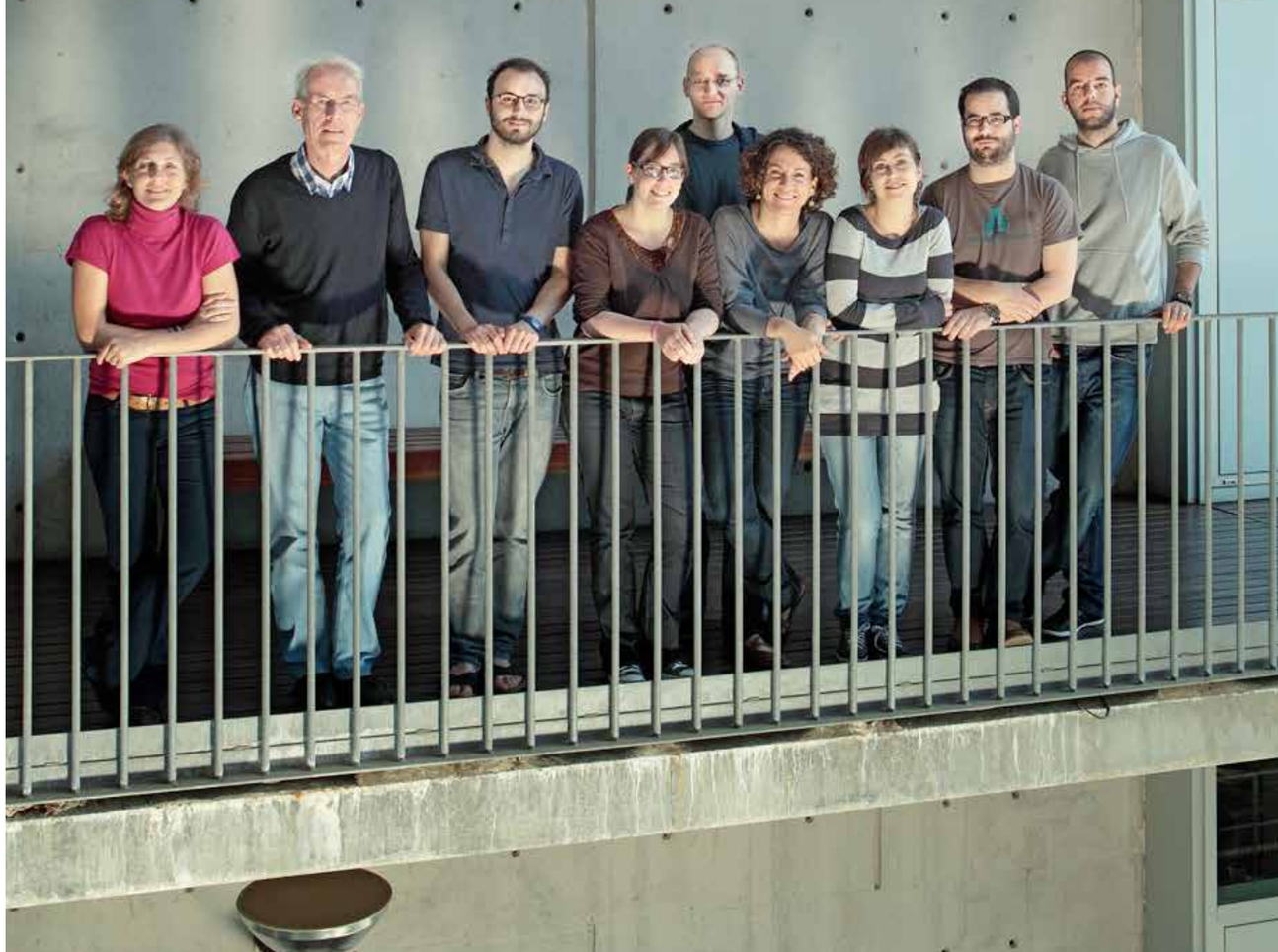
Weill L, Belloc E, Bava FA, Méndez R.
"Translational control by changes in poly(A) tail length: recycling mRNAs."
Nat Struct Mol Biol, 19(6):577-85 (2012).

GENE REGULATION, STEM CELLS 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), Jasna Medvedovic (BLUEPRINT, as of September 2012), Jose Luis Sardina (CRG, as of October 2012)
PhD Students:	Luisa de Andres (CRG), Alessandro DiTullio (CRG), Francesca Rapino ("la Caixa") Bruno Di Stefano (« la Caixa »)
Technicians:	Clara Berenguer (Plan Nacional, until September 2012), Jose Francisco Infante (CRG, until September 2012), Alai Urrutikoetchea (Plan E, until September 2012)



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 the activation 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 knockdown 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. For this purpose we collected samples of uninduced (0h) as well as cells induced for 3h, 12h, 24h and 48h and performed ChIPseq experiments. These included the binding of the myeloid regulators C/EBPa, C/EBPb and PU.1 as well as PolII to DNA and also the chromatin marks H3K4me1, HeK4me3, H3K27me3, H3K27Ac and H3K9me3. Our data show that establishment of the macrophage program involves the activation of two enhancer pathways, depending on whether C/EBPα or its partner PU.1 binds to DNA first. These two enhancer-binding modes dictate distinct kinetics of histone modification and nucleosome positioning changes. Importantly, genes associated with the two enhancer classes differ in their upregulation kinetics, both during induced lineage reprogramming and normal differentiation, and may exert distinct biological functions.

We have also studied the silencing of the B cell program, and did ChIPseq analyses of DNA binding of the B cekll regulators E2A and Ebf1 during transdifferentiation, as well as Pax5 and Foxo1 in preB cells. We found that silencing of the B cell program correlates with the transient binding of C/EBPa to enhancers of key B cell regulators, manifesting itself with a rapid decrease in PolII binding. We are now investigating how this works at the molecular level.

2. Transdifferentiation of human lymphoma cells and a potential therapeutic application

In another line of research we asked whether human lymphoma B cells could 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 the majority of 23 lymphoma and leukemia cell lines tested responded with a complete or partial change in the expression of the B cell marker CD19 and the macrophage marker Mac-1. Two lines in whch high levels of C/EBPa expression could be maintained, the Burkit lymphoma line Seraphina and the B cell precursor acute leukemia RCV-ACH, could be stably transdifferentiated into functional, quiescent macrophages. Remarkably, the conversion 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 tumors after transplantation into immunodeficient mice. We found (in collaboration with J. A. Martinez Climent, U. de Navarra) using both the Burkitt's lymphoma cells line and the immature B cell leukemia line stably expressing C/EBPaER that tumor formation could be inhibited by inducing their transdifferentiation into macrophages. Moreover, tumor formation could be delayed even when administering the inducer *in vivo*, 10 days after inoculation of untreated, inducible tumor cells.

3. Screening for compounds that induce transdifferentiation in a Burkitt lymphoma cell line

The results described above suggest that it might be possible to find drugs that can induce transdifferentiation as potentially new therapeutic drugs for leukemias and lymphomas. We therefore first developed a reporter system consisting of our Seraphina derived beta estradiol/tamoxifen inducible C/EBPaER containing cell line BLaER1 line containing the lysozyme promoter driving the expression of tdTomato. These cells turn red when they are induced to transdifferentiate into macrophages. Using this line together with the screening facility at the CRG we screened a total of 1,693 compounds, consisting of pharmacologically active compounds, a phytochemical libray as well as drugs targetingknown signaling pathways as well as cell reprpogramming. Of 30 primary hits 21 were discarde because they turned out to be estradiol analogs. Of the remaining 9 3 looked promising and validation experiments are ongoing. A similar screen is ongoing in collaboration with Sanofi Aventis in Toulouse, and one compound (an Aurora kinase inhibitor) is being tested for its effect on mouse and human B cell lines. If an effective compound could be found it might eventually be possible to develop it into a drug that could be tested for its inhibitory potential for patient derived cells, perhaps in combination with drugs currently used in the clinic that work based on their ability to induce apoptosis.

PUBLICATIONS

Kallin EM, Rodrıguez-Ubreva J, Christensen J, Cimmino L, Aifantis I, Helin K, Ballestar E and Graf T.
"Tet2 Facilitates the Derepression of MyeloidTarget Genes during CEBPa-Induced Transdifferentiation of Pre-B Cells."
Molecular Cell, 48:266–276 (2012).

Di Tullio A and Graf T.
"C/EBPα bypasses cell cycle-dependency during immune cell transdifferentiation"
Cell Cycle 11(14):2739-2746 (2012).
See also Commentary by M.Stadtfeld, Cell Cycle 11(17):3155 (2012).

Zhang L, Orban M, Lorenz M, Barocke V, Braun D, Urtz N, Schulz C, von Brühl ML, Tirniceriu A, Gaertner F, Proia RL, Graf T, Bolz SS, Montanez E, Prinz M, Müller A, von Baumgarten L, Billich A, Sixt M, Fässler R, von Andrian UH, Junt T, Massberg S.
"A novel role of sphingosine 1-phosphate receptor S1pr1 in mouse thrombopoiesis."
J Exp Med, 209(12):2165-81 (2012).

Adams D, Altucci L, Antonarakis SE, Ballesteros J, Beck S, Bird A, Bock C, Boehm B, Campo E, Caricasole A, Dahl F, Dermitzakis ET, Enver T, Esteller M, Estivill X, Ferguson-Smith A, Fitzgibbon J, Flicek P, Giehl C, Graf T, Grosveld F, Guigo R, Gut I, Helin K, Jarvius J, Küppers R, Lehrach H, Lengauer T, Lernmark Å, Leslie D, Loeffler M, Macintyre E, Mai A, Martens JH, Minucci S, Ouwehand WH, Pelicci PG, Pendeville H, Porse B, Rakyan V, Reik W, Schrappe M, Schübeler D, Seifert M, Siebert R, Simmons D, Soranzo N, Spicuglia S, Stratton M, Stunnenberg HG, Tanay A, Torrents D, Valencia A, Vellenga E, Vingron M, Walter J, Willcocks S.
"BLUEPRINT to decode the epigenetic signature written in blood."
Nat Biotechnol, 30(3):224-6 (2012).

Rodríguez-Ubreva J, Ciudad L, Gómez-Cabrero D, Parra M, Bussmann LH, di Tullio A, Kallin EM, Tegnér J, Graf T, Ballestar E.
"Pre-B cell to macrophage transdifferentiation without significant promoter DNA methylation changes."
Nucleic Acids Res, 40(5):1954-68 (2012).

Reviews

de Andrés-Aguayo L, Varas F, Graf T.
"Musashi 2 in hematopoiesis."
Curr Opin Hematol, 19(4):268-72 (2012).

GENE REGULATION, STEM CELLS AND CANCER

Group: Chromatin and Gene Expression

Group Structure:

Group Leader: Miguel Beato

Postdoctoral Fellows: Francois Le Dily, Gaetano Verde, Guillermo Vicent (Staff Scientist), Roni Wright, Roser Zaurin

PhD Students: Andy Pohl, Diana Reyes, Michael Wierer, Gireesh Bogu (since August, co-mentored with M. Marti-Renom)

Technician/s: Jofre Font, A. Silvina Nacht, Daniel Soronellas (bioinformatics)

Visitors: Alejandro La Greca, Buenos Aires (until September)

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 chromatin structure and gene expression. Steroid hormones signal to chromatin via binding of their intracellular receptors, a fraction of which is attached to the cell membrane and cross-talks with kinase signalling pathways, which ultimately facilitate their interaction of nuclear hormone receptors with the target sites in nucleosomes and the subsequent chromatin remodelling. In genome wide studies we found that two consecutive cycles of remodelling involving kinases, PARP1, histone modifying enzymes and ATP-dependent complexes lead to consecutive displacement of histones H1 and H2A/H2B dimers. Inhibiting some of theses kinases selectively blocks proliferation of breast cancer cells. We are studying how the topological information that organizes chromatin in the cell nucleus modulates the hormonal response and want to use the hormonal signalling network as target for combinatorial control of cancer cell proliferation.

RESEARCH PROJECTS

1. Global analysis of chromatin structure and dynamics during hormonal gene regulation

F. Le Dily, S. Nacht, A. Pohl, D. Soronellas, G. Verde, G. Vicent, R. Wright, R.Zaurin

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. We know that before hormone addition the potential Progesterone Receptor (PR) binding sites are partly occupied by PR and an associated repressive complex containing HP1γ and LSD1 (Vicent *et al.*, submitted). The nucleosome occupancy at PR binding sites changes significantly after hormone treatment, which also generates a large number of new DNase I hypersensitive sites. In ChIP-seq experiments with an antibody to PR we discovered a preferential location of PR binding sites over nucleosomes, which are remodeled after hormone treatment lading to displacement of histone H1 and H2A/H2B (Ballaré *et al.*, *Mol. Cell*, published on line Nov 2012) (Fig. 1). The role of histone arginine methylation and citrullination in this process is another question we are exploring (Zaurin, unpublished).

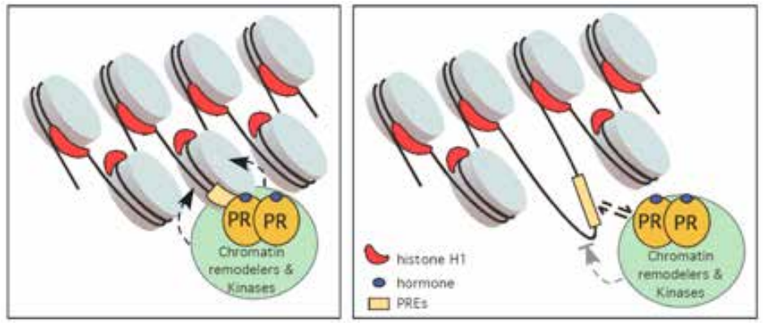


Figure 1. **Model for functional PR binding in response to progestin stimulation.**
(Left) Functional PREs are organized in nucleosomes. Upon hormone binding, the activated PR reaches the target PREs in association with histone tail modifiers and chromatin remodeling complexes that contact the histones stabilizing PR binding.
(Right) In PREs that are not organized in nucleosomes, histone tail modifiers and nucleosome remodelers cannot interact with histones; therefore, binding of PR is weaker and not functional.

Hormonal gene regulation and histone H1 displacement also require activation of the Poly(ADP-ribose) polymerases PARP1 via phosphorylation by CDK2 (Wright *et al.*, *Genes Dev* 2012). PARP1 activation leads to transient accumulation of PAR in the cell nucleus accompanied by a 50% decrease in cellular NAD+ content within 15-30 min of hormone addition. The role of PAR and its degradation in chromatin remodelling and gene regulation will be one of the focuses of the lab during the next year. We are also studying the changes in DNA methylation and hydroxymethylation in response to hormone and want

to explore whether the persistence of these changes influences DNaseI sensitivity as part of the cells epigenetic memory (Verde *et al.*, unpublished).

We have started 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. Initial results suggest that the consecutive Topological Associating Domains (TADs) that organize the whole human genome may represents units of hormonal regulation (Le Dily *et al.*, unpublished). In this context we are also analyzing he role of long non-coding RNAs (Gireesh, unpublished).

2. Signalling by progesterone to chromatin via kinase cascades

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. Recently we discovered a non-mitotic role of PLK1 in response to estrogens and progesterone, as co-activator of their respective receptors (Wierer *et al.*, submitted). 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 (*Wright*, unpublished). 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 to regulation of different gene cohorts as studied by oligonucleotide arrays and RNA-seq.

3. Role of steroid hormones in breast and endometrial cancer

D. Reyes, A. La Greca

We have investigated the function of MSK1 in the proliferative response induced by estrogens and progesterone in breast cancer cells T47D and MCF7. We found that both in cell culture and in mouse xenografts inhibiting MSK1 chemically or depleting cells of MSK1 using shRNA selectively inhibits proliferation of breast cancer cells in response to estrogens or progestins but not in response to other growth stimuli (Reyes *et al.*, *Oncogene*, in press). Similar experiments are being performed with endometrial cancer cells, in which progestins regulate a different cohort of genes (La Greca, unpublished).

PhD Thesis: *Diana Reyes* "Searching for a functional relationship between the breast cancer susceptibility gene BRCA1 and the progesterone receptor in breast cancer cells" *Doctor in Biology*, Universitat Pompeu Fabra (2012).

PUBLICATIONS

Wright RHG, Castellano G, Bonet J, Le Dily F, Font-Mateu J, Ballaré C, Nacht SA, Soronellas D, Oliva B, Beato M.
"CDK2 dependent activation of PARP-1 is required for hormonal gene regulation in breast cancer cells"
Genes & Dev, 26:1972-1983 (2012).
Included in *"Cancer Discovery"* News Archive Sept 13, 2012.
Selected by Faculty1000

Ballaré C, Castellano G, Gaveglia L, Althammer S, González-Vallinas J, Eyras E, Le Dily, Zaurin R, Soronellas D, Vicent GP, Beato M.
"Nucleosome driven transcription factor binding and gene regulation"
Mol Cell, doi: 10.1016/j.molcel.2012.10.019. Epub 2012 Nov 21.
Selected by Faculty1000

Dreszer TR, Karolchik D, Zweig AS, Hinrichs AS, Raney BJ, Kuhn RM, Meyer LR, Wong M, Sloan CA, Rosenbloom KR, Roe G, Rhead B, Pohl A, Malladi VS, Li CH, Learned K, Kirkup V, Hsu F, Harte RA, Guruvadoo L, Goldman M, Giardine BM, Fujita PA, Diekhans M, Cline MS, Clawson H, Barber GP, Haussler D, James Kent W.
"The UCSC Genome Browser database: extensions and updates 2011."
Nucleic Acids Res, 40(Database issue):D918-23 (2012).

Rosenbloom KR, Dreszer TR, Long JC, Malladi VS, Sloan CA, Raney BJ, Cline MS, Karolchik D, Barber GP, Clawson H, Diekhans M, Fujita PA, Goldman M, Gravell RC, Harte RA, Hinrichs AS, Kirkup VM, Kuhn RM, Learned K, Maddren M, Meyer LR, Pohl A, Rhead B, Wong MC, Zweig AS, Haussler D, Kent WJ.
"ENCODE whole-genome data in the UCSC Genome Browser: update 2012."
Nucleic Acids Res, 40(Database issue):D912-7 (2012).

Reyes D, Ballare C, Castellano G, Soronellas D, Bago JR, Blanco J, Beato M.
"Activation of Mitogen- and Stress-activated Kinase 1 is required for proliferation of breast cancer cells in response to estrogens and progesterone"
Oncogene, in press

Reviews

Beato M, Vicent GP.
"Impact of chromatin structure and dynamics on PR signaling. The initial steps in hormonal gene regulation."
Mol Cell Endocrinol, 357:37-42 (2012).

Wright RHG, Beato M.
"PARty promoters: Hormone dependent gene regulation requires CDK2 activation of PARP1"
Cell Cycle, 11:4291-3 (2012). Feature.

Vicent GP, Zaurin R, Ballaré C, Nacht AS, Wright RHG, Le Dily F, Castellano C, Font-Mateu J, Beato M.
"Progesterone signaling to chromatin in breast cancer cells. Two initial cycles of remodelling."
In: *Advances in rapid sex-steroid action: new challenges and new chances in breast and prostate cancers,* Ed. G. Castoria & A. Migliacio, Springer, Part 1, 19-29 (2012).

GENE REGULATION, STEM CELLS 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: Giacoma Simonte, Joao Frade, Elisa Pedone, Maria Aurelia Ricci, Francesco Aulicino, Francesco Sottile, Ilda Theka

Master Students: 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 reprogram 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 reprogram differentiated cells. We are using reverse engineering and forward algorithms to infer the natural network of interactions surrounding the genes involved in the reprogramming of somatic cells. Furthermore, we are developing 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.

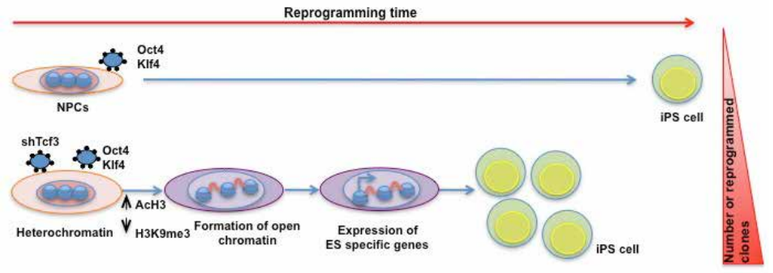


Figure Legend:
Silencing of Tcf3 induces massive epigenome modifications and increases reprogramming efficiency. From Luis, Ombrato et al. PNAS 2011

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. We are carrying out transplantation of perturbed (Wnt-activated or repressed) adult stem cells into a variety of drug-induced or genetically modified damaged organs. Short-term and long-term regeneration is currently studied. Genetic approaches are used to evaluate cell fusion, reprogramming and regeneration in the tissues analyzed.

PUBLICATIONS

Ombrato L, Lluis F, Cosma MP.
"Regulation of self-renewal and reprogramming by TCF factors."
Cell Cycle, 11(1):39-47 (2012).

Reviews

Lluis F, Cosma MP.
"Resetting Epigenetic Signatures to induce Somatic Cell Reprogramming."
Cell Mol Life Sci, doi: 10.1007/s00018-012-1137-8. Epub 2012 Aug 30.

GENE REGULATION, STEM CELLS 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, Elisabeth Simboeck, Pedro Vizan (since September 2012), Milos Pjanic (since September 2012)
PhD Students:	Santiago Demajo, Sophia Teichmann (until October 2012), Joana Ribeiro (until March 2012), Paola Pisano (until October 2012), Jain Payal, Alexandra Santanach
Technician:	Arantxa Gutierrez, Cecilia Ballaré
Visiting Scientist:	Livia Caizzi, Chiara Laezza

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. Variation in Polycomb complexes assambly

We performed thorough ChIP-seq analyses of several components of the PRC 1 and PRC2. In particular, we monitored the occupancy of the Cbx family of proteins (Figure 1) in proliferating mES cells and differentiating mES cells.

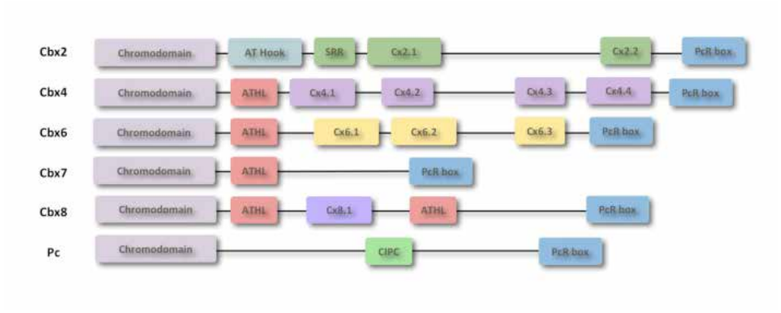


Figure 1.
The Cbx protein family.

Our biochemical and proteomic data indicate that, in proliferating mES cells, Cbx7 is the main chromodomain-containing protein of PRC1. In agreement with this, we identified 2,349 Cbx7 target genes, of which 96.7% were co-occupied by Ring1B (Figure 2). A large number (96.7%) were co-occupied by Suz12 (PRC2), thus suggesting that the majority of genomic loci occupied by PRC1 are also occupied by PRC2.

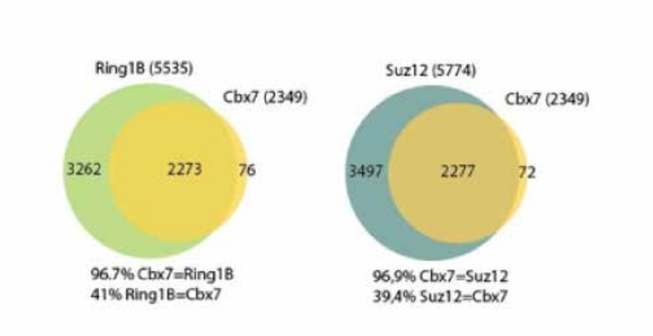


Figure 2.
ChIP-seq analysis of Cbx7 target genes and their overlap with genes containing Ring 1B and Suz12.

Gene ontology (GO) analyses of the list of genes co-occupied by Cbx7 indicated that they are involved in developmental processes; this includes factors involved in ectoderm and mesoderm specification. Interestingly, pathways essential for early developmental processes that are commonly deregulated in cancer, such as Hedgehog, Wnt, TGFb, MAPK, and those involved in calcium signalling, were highly represented within the cohort of Cbx7 targets.

On the other hand, biochemical purification of Cbx6 indicates that it did not interact with the PRC1 complex. ChIP-seq analysis of Cbx6 in mES cells indicated that it localized at 819 target genes. Only 5% of genes occupied by Cbx7, Ring1B, Suz12, H3K27me3, or H3K9me3 were also co-occupied by Cbx6 (Figure 3). Despite the lack of PRC1/2 binding to Cbx6 target genes, the majority of these were not transcribed in mES cells. GO analysis indicated that Cbx6 target genes are involved mainly in cellular, metabolic, and protein modification processes. Intriguingly, Cbx6 binding to chromatin occurs predominantly at TSSs, but spreads up to ±50 kb from the TSS.

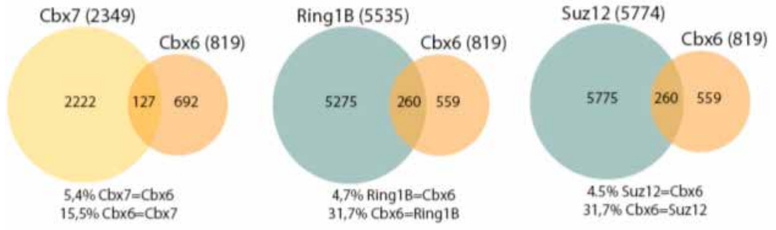


Figure 3.
ChIP-seq analysis of Cbx7 target genes and their overlap with genes with Ring 1B, Suz12, and/or H3K27me3.

We also analyzed whether Cbx7 plays a role during the differentiation of **mES cells** into **embryoid bodies (EBs)**. Intriguingly, Cbx7 expression was downregulated during EB formation, while Cbx2 and Cbx4 were upregulated. We biochemically identified two distinct PRC1 complexes in EBs: one with Cbx2, and the other with Cbx4. This prompted us to speculate that (1) PRC1 complexes containing either Cbx2 or Cbx4 might co-regulate the same set of genes; and (2) two different PRC1 complexes, containing either Cbx2 or Cbx4, might target a specific cohort of genes during EB differentiation. ChIP-seq analyses for both Cbx2 and Cbx4. 96.4% of Cbx4 (7,160) and 99% of Cbx2 (1,400) target genes showed enrichment for other PRC1 components (Figure 4).

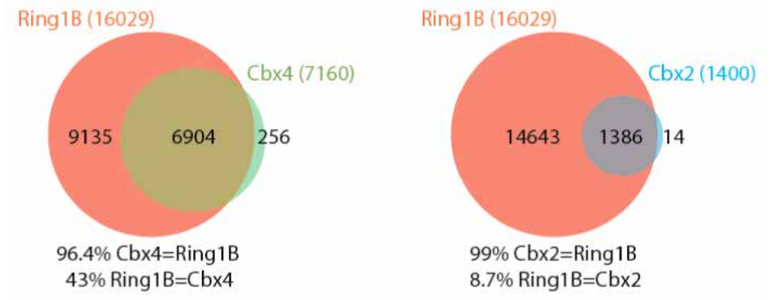


Figure 4.
Overlap between target genes of Ring1B, Cbx2, and/or Cbx4.

Finally, GO analysis of each set of ChIP-seq data revealed that Ring1B, Cbx2, and Cbx4 target genes relate to **early development and specification** to the three germ layers, albeit with an overrepresentation of the mesoderm lineage for Cbx4 and Cbx2.

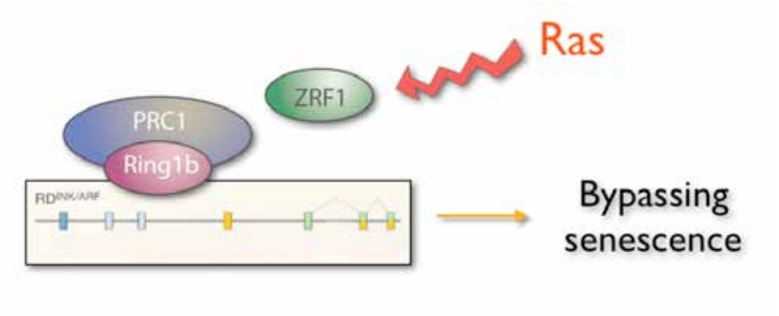
2. ZRF1 and histone H2A ubiquitination

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.



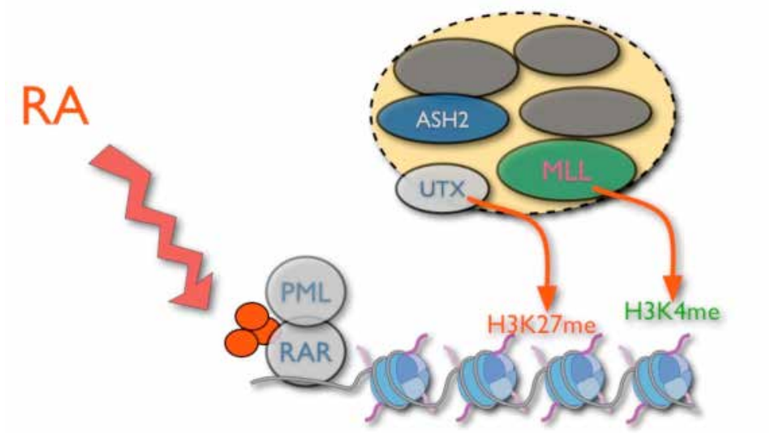
3. Role of histone demethylases in leukemia

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-RARa 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.



PUBLICATIONS

Morey L, Aloia L, Cozzuto L, Benitah SA, and Di Croce L.
"RYBP and Cbx7 define specific biological functions of PRC1 complexes in mouse embryonic stem cells."
Cell Rep, doi: 10.1016/j.celrep.2012.11.026. Epub 2012 Dec 27.

Ribeiro J, Morey M, Gutierrez A, Mejetta S, Mas A, Richly H, Benitah SA, Keyes WM, and Di Croce L.
"ZRF1 controls oncogene-induced senescence through the INK4-ARF locus."
Oncogene, doi: 10.1038/onc.2012.241. Epub 2012 Jun 25.

Soncini M, Santoro F, Gutierrez A, Frigè G, Romanenghi M, Botrugno OA, Pallavicini I, Pelicci P, Di Croce L, Minucci S.
"The DNA demethylating agent decitabine activates the TRAIL pathway and induces apoptosis in acute myeloid leukemia."
Biochim Biophys Acta, doi: 10.1016/j.bbadis.2012.10.001. Epub 2012 Oct 6.

Ballare C, Lange M, Lapinaite A, Martin GM, Morey L, Pascual G, Liefke R, Simon B, Shi Y, Gozani O, Carlomagno T, Benitah SA, and Di Croce L.
"Phf19 links methylated lysine 36 of of histone H3 to regulation of Polycomb activity."
Nat. Struct. Mol. Biol, 19:1257-65 (2012).

Herranz N, Dave N, Lorenz-Fonfría V, Millanes-Romero A, Morey L, Jeronimo C, Di Croce L, García de Herreros A, and Peiró S.
"Lysil Oxidase Like 2 LOXL2 is a Histone H3 lysine 4 deaminase."
Mol Cell, 46:369-376 (2012).

Creppe C, Janich P, Cantariño N, Noguera M, Valero V, Musulén E, Posavec M, Douet J, Caballero J-M, Sumoy L, Di Croce L, Benitah SA, and Buschbeck M.
"MacroH2A1 regulates the balance between self-renewal and differentiation commitment in embryonic and adult stem cells."
Mol Cell Biol, 32:1442-52 (2012).

Morey M, Pascual G, Cozzuto L, Roma G, Wutz A, Benitah SA, and Di Croce L.
"Non-Overlapping Functions of Cbx Proteins in Embryonic Stem Cells."
Cell Stem Cell, 10:47-62 (2012).

Proto MC, Gazzerro P, Di Croce L, Santoro A, Malfitano AM, Pisanti S, Laezza C, Bifulco M.
"Interaction of endocannabinoid system and steroid hormones in the control of colon cancer cell growth."
J Cell Physiol, 227:250-58 (2012).

Creppe C, Cantarino N, Janich P, Valero V, Di Croce L, Aznar-Benitah S, Buschbeck M.
"The histone variant macroH2A regulates the commitment of embryonic and adult stem cells."
FEBS Journal, 279(1):189-189 (2012).

Reviews

Luis NM, Morey L, Di Croce L, Benitah SA.
"Polycomb in Stem Cells: PRC1 Branches Out."
Cell Stem Cell, 11:16-21 (2012).

Uribesalgo I, Benitah SA, and Di Croce L.
"From oncogene to tumor suppressor: Myc's dual role in leukemia."
Cell Cycle, 11:1757-64 (2012).

Uribesalgo I, Ballaré C, and Di Croce L.
"Polycomb regulates NF-kB signalling through miRNA in cancer."
Cancer Cell, 21:5-7 (2012).

GENE REGULATION, STEM CELLS AND CANCER

Group: Regulation of Protein Synthesis in Eurkaryotes

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. We have exploited translation-competent extracts from *Drosophila* embryos to study mechanistically translational control events important for early development and viability. We have identified RBP complexes involved in regulation, have uncovered an unexpected layer of sex-specificity in the binding of some RBPs, and have taken steps towards the understanding of the role of conserved RBPs in cancer progression.

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). 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 genome-wide analysis. The results have revealed a large degree of sex-specific target binding by UNR, which is mainly explained by alternative processing of UTRs in target transcripts (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.

UNR is highly conserved in mammals, and the nature of its target transcripts suggests an involvement of UNR in cancer progression. We are defining the targets of mammalian UNR in specific tumor settings, and the effects of its depletion or over-expression on tumor growth.

We have also identified additional components of the *msl2* silenced complex using a novel technology called GRAB (GST-pulldown and RNA Affinity Binding) (Figure 1), and are currently characterizing their molecular roles. We are specially interested in factors that –as UNR- are highly conserved in mammals.

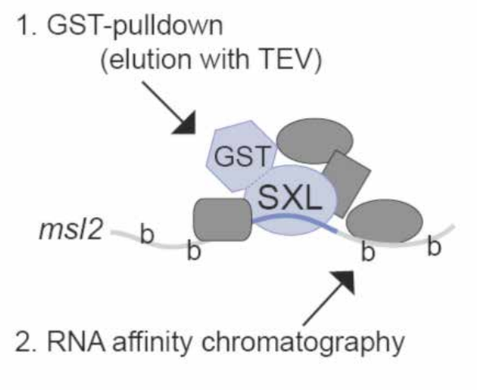


Figure 1. **Using GRAB to identify components of the silenced SXL:*msl2* mRNP**
Recombinant GST-SXL and biotinylated RNA are incubated with Drosophila embryo extracts. RBP complexes are allowed to form, and are first selected by GST-pulldown followed by RNA affinity chromatography using streptavidine beads.

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 poly(A) tail elongation. 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 several co-existing machineries for cytoplasmic polyadenylation, notably distinct from that known in vertebrates. Our current research focuses on analyzing the specific roles of the components of these machineries on 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, 18:53-64 (2012).

Gebauer F, Preiss T and Hentze MW.
"Translational control: from cis-regulatory elements to complex RNPs and back."
Cold Spring Harb. Perspect. Biol., 4:a012245 (2012).

Gebauer F.
"Versatility of the translational machinery during stress: changing partners to keep dancing."
Cell Research, 22:1634-6 (2012).

Szostak E and Gebauer F.
"Translational control by 3' UTR binding proteins."
Brief Funct Genomics, doi: 10.1093/bfgp/els056. Epub 2012 Nov 28.

Reviews

Wurth L.
"Versatility of RNA-binding proteins in cancer"
Comparative and Functional Genomics, 2012:178525.

Book chapters

Coll O, Villalba A and Gebauer F.
"Cytoplasmic polyadenylation assays. Methods in Molecular Biology"
In: *Polyadenylation: Methods and Protocols*, Humana Press, in press.

GENE REGULATION, STEM CELLS 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, Lorenzo Rinaldi

Technicians: Gloria Pascual

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.**

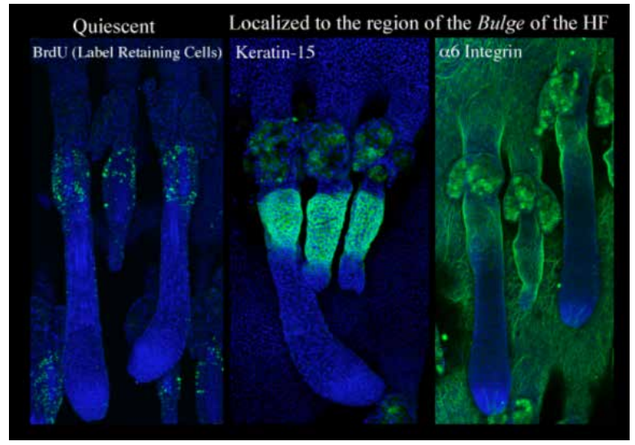


Figure 1:
Epidermal stem cells are located at a region of the hair follicle known as the bulge. They are relatively quiescent and strictly positioned at their niche (Immunostaining and pictures by Peggy Janich).

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?

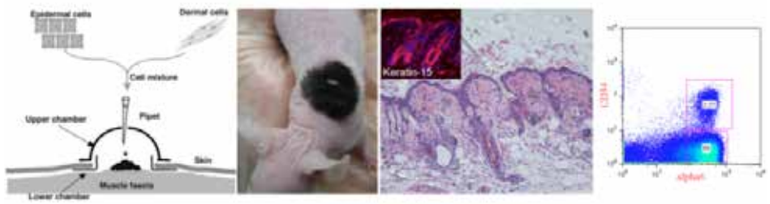


Figure 2:
Epidermal stem cells are the only population that can regenerate the entire epidermal compartment. Epidermal stem cells transplanted onto nude mice can regenerate a homeostatic functional epidermis, hair follicles and sebaceous glands. A niche of epidermal stem cells is re-established once 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 and Hospital Vall D´Hebron 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

Lapinaite A, Lange M, Marti GM, Morey L, Ballare C, Simon B, Gozani O, Carlomagno T, Benitah SA, Di Croce L.
"Phf19 links methylated lysine 36 of histone H3 to regulation of Polycomb activity."
Nat Struct Mol Biol, 19(12):1257-65 (2012).

Frye M, Benitah SA.
"Chromatin regulators in mammalian epidermis."
Semin Cell Dev Biol, 23(8):897-905 (2012).

Benitah SA.
"Defining an epidermal stem cell epigenetic network."
Nat Cell Biol, 14(7):652-3 (2012).

Morey L, Pascual G, Cozzuto L, Roma G, Wutz A, Benitah SA, Di Croce L.
"Non overlapping functions of Cbx proteins in embryonic stem cells."
Cell Stem Cell, 10(1):47-62 (2012).

Uribesalgo I, Benitah SA*, Di Croce L*.
"From oncogene to tumor suppressor: Myc's dual role in leukemia."
Cell Cycle, 11(9):1757-64 (2012).
(*co-corresponding authors)

Creppe C, Janich P, Cantariño N, Valero V, Musulén E, Posavec M, Douet J, Di Croce L, Benitah SA, Buschbeck M.
"MacroH2A1 regulates the balance between self-renewal and differentiation commitment in stem cells."
Mol Cell Biol, 32(8):1442-52 (2012).

Creppe C, Cantarino N, Janich P, Valero V, Di Croce L, Aznar-Benitah S, Buschbeck M.
"The histone variant macroH2A regulates the commitment of embryonic and adult stem cells."
FEBS Journal, 279(1):189-189 (2012).

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 (2012).

Ribeiro JD, Morey L, Gutierrez, Mejetta S, Richly H, Benitah SA, Keyes WM, Di Croce L.
"ZRF1 controls oncogene-induced senescence through the INK4-ARF locus."
Oncogene, doi: 10.1038/onc.2012.241. Epub 2012 Jun 25.

Morey L, Aloia L, Cozzuto L, Benitah SA, and Di Croce L.
"RYBP and Cbx7 define specific biological functions of PRC1 complexes in mouse embryonic stem cells."
Cell Rep, doi: 10.1016/j.celrep.2012.11.026. Epub 2012 Dec 27.

Reviews

Luis N, Morey L, Di Croce L, Benitah SA.
"Polycomb in stem cells: PRC1 branches out."
Cell Stem Cell, 11(1):16-21 (2012).

Benitah SA, Frye M.
"Stem cells in ectodermal development."
J Mol Med, 90(7):783-90 (2012).

GENE REGULATION, STEM CELLS 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 posess 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 developmet, focusing on identifying the mechanisms involved and the processes by which aberrant stem cell proliferation may favour tumor development.

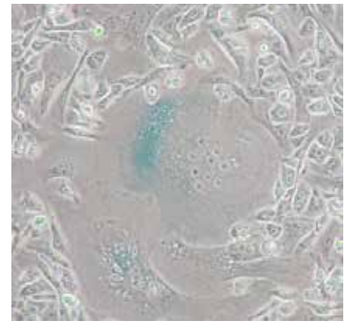


Figure 1.
Senescence-associated β-galactosidase staining identifies senescent cells (blue stain). Note the characteristic morphology, including enlarged flat shape and multiple nuclei. Multiple smaller transformed cells surround.

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.

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 one study to address this challenge, we took the approach to study purified skin stem cells isolated directly from young and old animals, and performed high-throughput RNA-sequencing on these cells to identify the potential molecular changes associated with aging at the cellular level. Surprisingly, we discovered that during normal aging, the number of stem cells in the skin significantly increases in number, but that these cells decrease in their functional capacity both *in vitro* and *in vivo*. By analysing our profiling data, we uncovered that these cells reside in an increasingly pro-inflammatory environment, which we demonstrated could functionally impair stem cell function with age. Interestingly, we were further able to demonstrate that this inflammatory signal was coming from the aging epithelial cells of the basal layer of the skin. Together, this study identified for the first time that stem cells in the skin are subject to age-associated alterations and implicates global alterations in the tissue microenvironment as a critical mediator.

Additionally, 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

Doles J, Storer MA, Cozzuto L, Roma G and Keyes WM.
"Age-associated inflammation inhibits epidermal stem cell function."
Genes & Development, 26(19):2144-53 (2012).

Ribero JD, Morey L, Gutierrez A, Mas A, Mejetta S, Richly H, Aznar Benitah S, Keyes WM and Di Croce L.
"ZRF1 controls oncogene-induced senescence through the INK4-ARF locus."
Oncogene, doi: 10.1038/onc.2012.241. Epub 2012 Jun 25.

Doles J and Keyes WM.
"Epidermal stem cells undergo age-associated changes."
Aging Journal, in press.

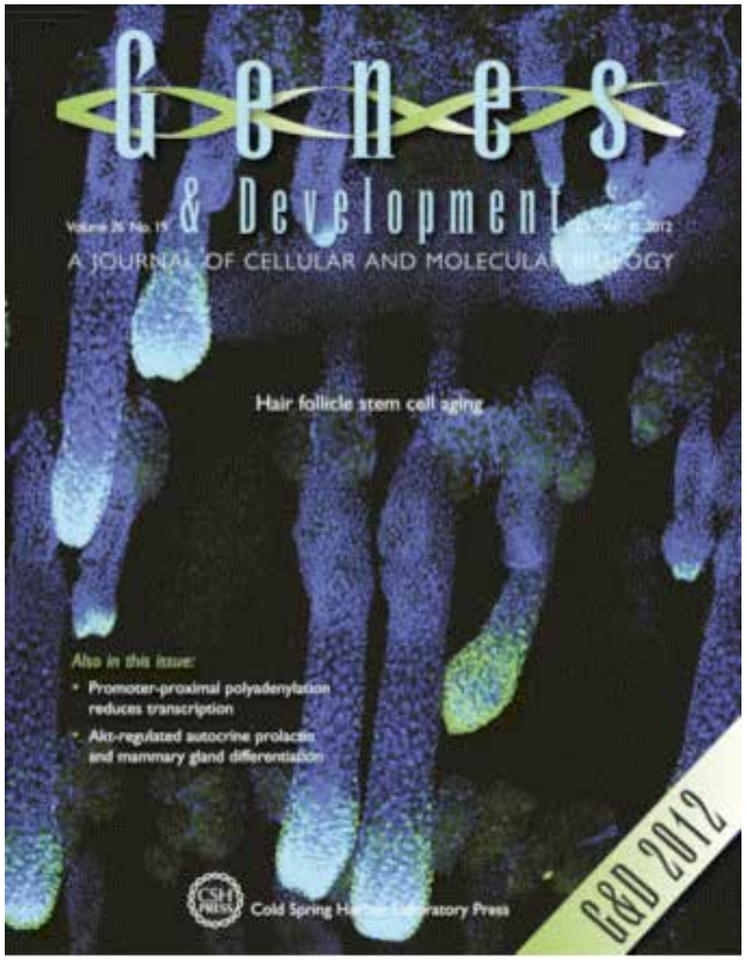


Figure 2.
Our study identifying age-associated changes in hair-follicle stem cells was selected as the cover picture in the journal Genes and Development. Shown is a wholemount immunostaining of the hair follicle labelled with CDP (green), a marker for actively growing hair follicles, as well as DAPI (blue) to visualize the nuclei and highlight the hair shaft.

GENE REGULATION, STEM CELLS AND CANCER

Group: Genome Architecture

Group Structure:

Group Leader: Guillaume Filion

Lab Manager: Heng-Chang Chen

Postdoctoral Researchers: Olivera Vujatovic

PhD Students: Marc Corrales, Ana Catarina Vidinhas De Oliveira

Students: Ricky Lim

SUMMARY

The group is interested in the regulatory genome. Contrary to the coding genome, the information of the regulatory genome is context-dependent. For example, the same promoter can have different levels of activity, the same enhancers can activate one gene or another, depending on available transcription factors and on the local chromatin marks. So the DNA sequence is not enough to understand the function of regulatory sequences. Our research lines focus on the influence of the chromatin context on transcription; and on the co-evolution between the genome and its chromatin context.

RESEARCH PROJECTS

1. Invasion landscapes

Distinct chromatin types can “fight” for a genome territory. This was suggested by an early observation of *Drosophila* genetics called Position Effect Variegation (PEV). In mutant PEV strains, the white gene is translocated near a centromere, which is coated in heterochromatin, rich in repeated sequences and poor in genes. Even though the sequence of the gene is not mutated, these flies have a characteristic phenotype with mottled eye color. This reflects random expression levels, mirroring the invasion of heterochromatin into the territory of the gene.

Intriguingly, some genes seem to be immune to PEV (in particular the ones that naturally map to the centromeres). So the question remains if and how the phenomenon happens in physiological conditions. How can a chromatin type invade a territory and replace the local chromatin? And what is the role of the DNA sequence in allowing this transition to take place?

To better understand these phenomena, we develop a technology that we call TRiP (see below), for Thousand Reporters in Parallel. Inspired by PEV of the white gene, the aim of TRiP is to place a gene in a new chromatin context at random and measure the impact on its transcription. The novelty of TRiP is that it allows to do this in a genome-wide and high-throughput manner. If the sequence of the gene fully determines it activity, the transcription will be even across the genome. Otherwise, if the chromatin context has a say, the transcription will vary. In short, TRiP allows us to map the regions where a gene can invade the local chromatin and maintain its expression level, and regions where the local chromatin shuts down its expression.

The TRiP technology

The TRiP technology consists of three steps: barcoding PCR, Sleeping Beauty transposition, and high throughput sequencing. Barcoding PCR is a technique developed by our team to generate libraries of tagged plasmids making every molecule unique. The principle of barcoding PCR is to insert 20 random nucleotides at a chosen location of a plasmid. This generates a library with a sufficient complexity to ensure that we can identify every molecule uniquely.

This technique allows us to construct a library of barcoded Sleeping Beauty transposons that are integrated in a cell population. The chances of two integrations having the same barcode are negligible. The constant part of the transposon consists of the promoter of an endogenous gene driving the expression of GFP and Sleeping Beauty transposase binding sites necessary for transposition. Such a library and a Sleeping Beauty transposase expression plasmid a co-transfected in a population of cells, which results in a heterogeneous population expressing the transgene at different levels, depending on the integration site.

The barcode is used in two ways. First, barcodes are mapped by inverse PCR followed by high throughput sequencing, which allows us to associate a barcode with a unique integration site. In addition, the barcode is actually placed in the mRNA of the GFP gene, which allows us to distinguish transcripts produced by different transposons and measure their individual expression by quantitative high throughput sequencing. By combining both informations we know whether a promoter can drive the expression of GFP at a particular site and in a particular chromatin context.

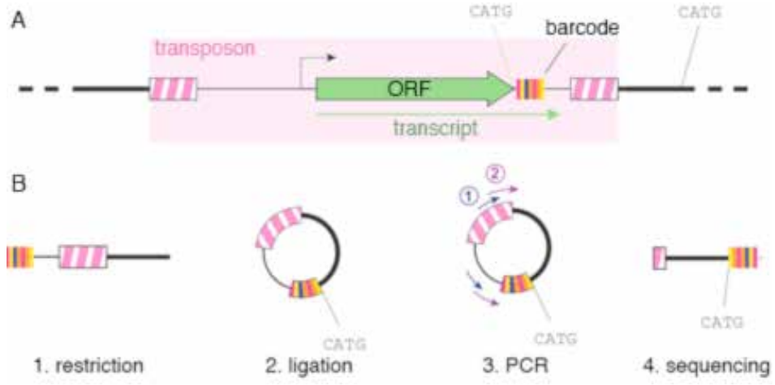


Fig .1
Principle of the TRiP technology. *A. Reporters are inserted at thousands of locations in the cell population. The transposons harbour a barcode inserted in the 3'UTR of the reporter gene, unique for every insertion. B. The integrations are recovered by inverse PCR. The genome is digested by the 4-cutter NlaIII and ligated to favor intra-molecular ligations. PCR is carried out with primers on transposon parts to amplify the insertion points, which are sequenced by high throughput methods.*

PUBLICATIONS

Steglich B, Filion GJ, van Steensel B, Ekwall K.
"The inner nuclear membrane proteins Man1 and Ima1 link to two different types of chromatin at the nuclear periphery in S. pombe."
Nucleus. 2012 Jan-Feb;3(1):77-87. (*)

(*) This publication results from the work of Dr. Guillaume J. Filion at the Division of Gene Regulation, Netherlands Cancer Institute, Amsterdam, The Netherlands.

GENE REGULATION, STEM CELLS AND CANCER

Group: Structural Genomics
Group with dual affiliation CRG-CNAG

Group Structure:

Group Leader:	Marc A. Marti-Renom
Staff Scientist:	Davide Baù
Postdoctoral Fellows:	François Serra
PhD Students:	David Dufour, Gireesh K. Bogu, Francisco Martínez

SUMMARY

We are interested in the molecular mechanisms that regulate cell fate. To study such mechanisms, we employ the laws of physics and the rules of evolution to develop and apply computational methods for predicting the 3D structures of macromolecules and their complexes. Our immediate research aims are:

> Determine the three-dimensional structure of the human genome at high-resolution in the interphase nucleus.
> Predict the structure of as many non-coding RNA molecules as possible by computational means.
> Apply our computational methods to the discovery of new drugs against neglected diseases.

The immediate aims outlined above have a very strong development component. New computational methods for the analysis and high-resolution prediction of molecular structure, interaction and function will be needed to achieve them. Additionally, a collaborative environment, which allows a rapid conversion of the newly developed methods to interesting biological applications, is essential for the success of such aims.

RESEARCH PROJECTS

1. Structure determination of genomes

D. Baù, F. Serra

The three-dimensional organization of the genome plays important, yet poorly understood roles in gene regulation. Detailed insights into chromosome conformation will greatly contribute to a more complete characterization of genome regulation. The spatial organization of chromosomes is reflected in, and driven by, cis- and trans interactions between genomic elements. We have recently developed a hybrid method (computational and experimental) based on the hypothesis that the spatial conformation of chromosomes can be determined by using comprehensive *in vivo* chromatin interaction data sets. Our collaborator, Dr. Job Dekker from the University of Massachusetts Medical School (US), have developed the 5C/HiC technology, which we are combining with the Integrative Modeling Platform (IMP) to determine the higher-order chromatin folding of genomic domains (Figure 1) and whole genomes.

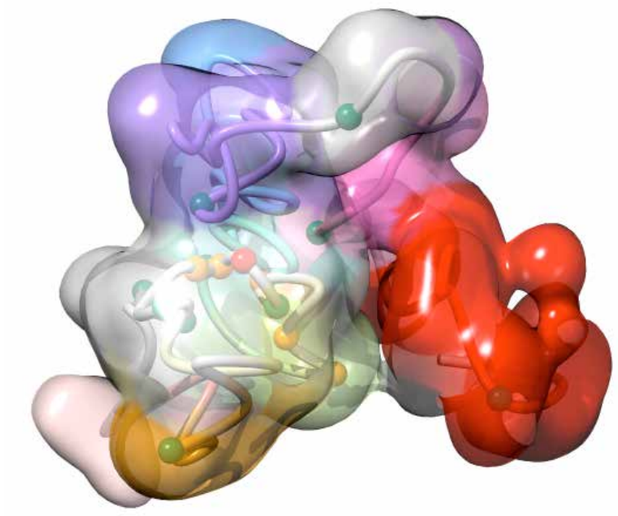


Figure 1.
Structure of the α-globin domain in the chromosome 16 in human fibroblast cells. (Baù, D. et al. The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules. 2011. Nat Struct Mol Biol 18, 107–114.)

2. RNA structure prediction

D. Dufour

The view of RNA as simple information transfer molecule has been continuously challenged since the discovery of ribozymes. Functional RNA molecules fold into their native three-dimensional conformations essential for performing their biological activity. Despite advances in understanding the folding and unfolding of RNA molecules, our knowledge of the atomic mechanism by which RNA molecules adopt their biological active structures is still limited. Moreover, we have limited knowledge on how non-coding RNA molecules regulate gene expression in health and disease. Over the past years, we have established a line of research dedicated to RNA. We have been making use of structural data in the Protein Data Bank to develop new methods for aligning RNA structures, assigning function to new RNA structures, and determining the limits of automatic RNA comparative structure prediction.

3. Open Source drug discovery projects for tropical diseases

F. Martínez

We have been pioneers in applying computational methods within an open source framework. In 2004, we co-initiated the Tropical Disease Initiative (TDI) with the aim of finding cures for tropical diseases such as malaria, leishmaniasis, denge, or tuberculosis. The main idea behind TDI is to promote research in a decentralized, Web-based, community-wide effort where scientists from laboratories, universities, institutes, and corporations can work together for a common cause. TDI provides, in our opinion, a perfect venue where our research can be tested and applied with immediate repercussion and publicity.

PUBLICATIONS

Dufour D and Marti-Renom MA.
"Modeling of RNA sequences"
Encyclopedia of Biophysics, in press.

Baeza C, Marti-Renom MA and Mingarro I.
"Structure-based statistical analysis of transmembrane segments"
European Biophysics Journal, in press.

Dufour D, Capriotti E and Marti-Renom MA.
"Computational methods for RNA structure prediction and analysis."
PanStanford Publishing, in press.

Bañó-Polo M, Baeza-Delgado C, Orzáez M, Marti-Renom MA, Abad C and Mingarro I.
"Polar/Ionizable residues in transmembrane segments: effects on helix-helix packing."
PLoS ONE, 7:e44263 (2012).

Baù D and Marti-Renom MA.
"Genome structure determination via 3C-based data integration by the Integrative Modeling Platform"
Methods, 58:300-306 (2012). (*)

(*) This publication results from the work of Dr. Marc A. Marti-Renom at Structural Genomics Team, Genome Biology Group, National Center for Genomic Analysis (CNAG), Barcelona, Spain.

GENES AND DISEASE

Group: Genetic Causes of Disease

Group Structure:

Group Leaders: Xavier Estivill

Staff Scientist: Eulàlia Martí

Scientific Officer: Mònica Gratacòs (until February 2012)

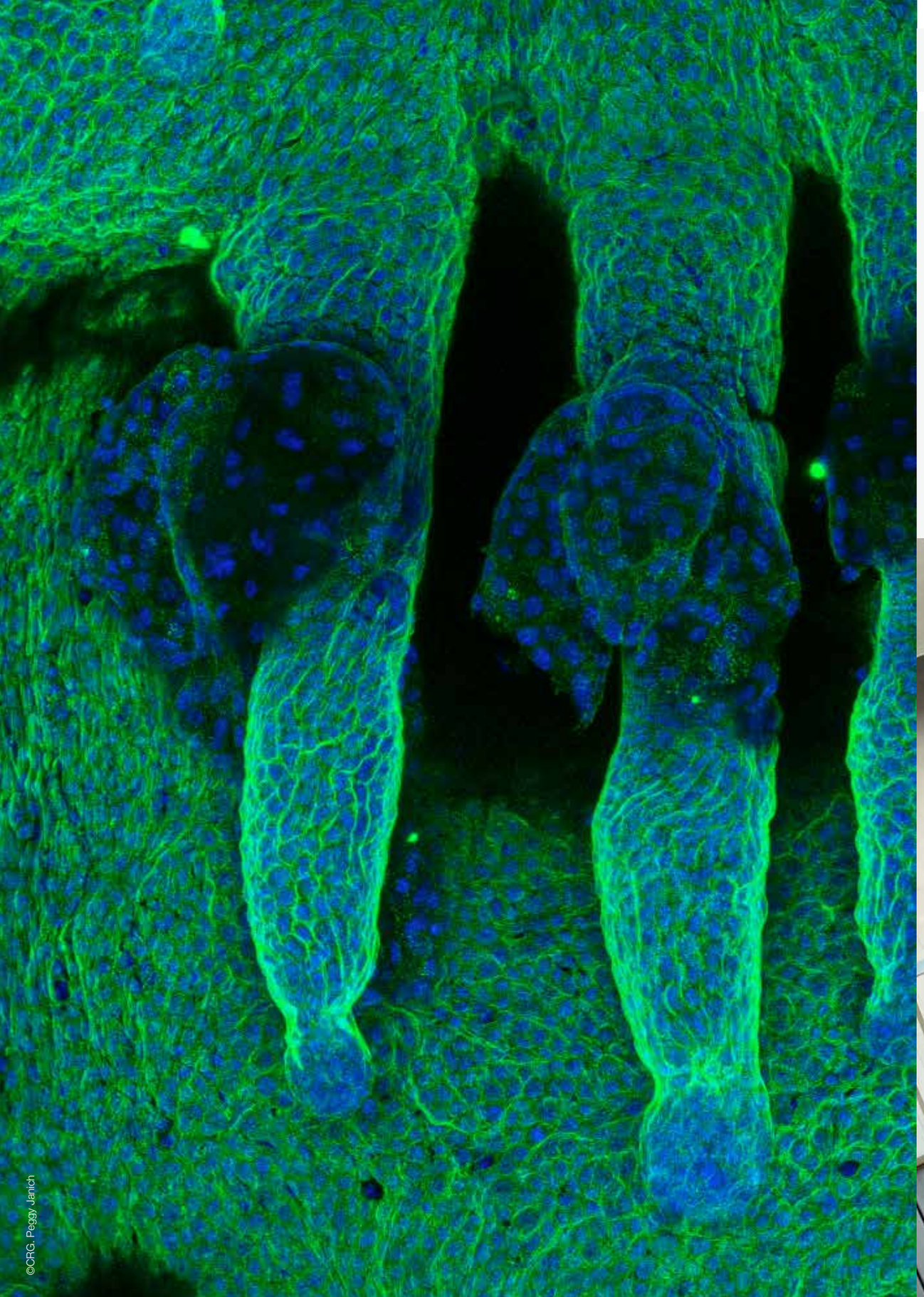
Postdoctoral Fellows: Mónica Bañez-Coronel, Mariona Bustamante (CREAL), Georgia Escaramís (CIBERESP) Marc Friedlander, Hyun Hor, Charlotte Hor-Henrichsen, Esther Lizano, Raquel Rabionet, Anna Houben (since January 2012), Aparna Prasad (since November 2012)

PhD Students: Johanna Aigner, Laia Bassaganyas, Elisa Docampo, Elisabet Mateu, Daniel Trujillano, Nàdia Vilahur (CREAL)

Graduate Students: Laura Domènech, Joan Pallarès

Technicians: Justo González (until December 2012), Birgit Kagerbauer, Marta Morell, Anna Puig, Cristian Tornador,

Visiting Scientists: Tim Spector (King's College London, UK)



SUMMARY

We are interested in the evaluation of how different types of genetic variants (single nucleotide variants, SNVs; structural variations (SVs), mainly copy number variants or CNVs; and insertion/deletion variants, including transposable elements) contribute to human disorders. We are also exploring the contribution of epigenetic modifications and non-coding RNA pathways in complex diseases, mainly neuropsychiatric and neurodegenerative disorders. We approach these questions through very well characterized cohorts of patients and the use of high-throughput genomic platforms and functional studies including longitudinal studies of patients at different time-points and cohorts in which exposure to environment has been monitored.

RESEARCH PROJECTS

1. Structural variations, environment and disease

J. Aigner, L. Bassaganyas, M. Bustamante, G. Escaramís, J. González, C. Hor-Henrichsen, M. Morell, R. Rabionet, C. Tornador, N. Vilahur

We characterize SVs as genomic regions potentially involved in phenotype variation, including disease. Some of the targeted regions contain CNVs, inversions, segmental duplications or transposable sequences. The 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 already characterized the genome and transcriptome of the first ten cases of CLL and have developed new tools for the bioinformatic characterization of SV in this common type of leukaemia. We have performed a deep characterization of one case of chromotripsis by carrying out a longitudinal analysis of SVs and point mutations over a period of eleven years of disease evolution, including the analysis of pre-treatment samples, post-treatment and relapsed samples. To identify SVs from whole genome sequencing data, we have developed the PeSV-Fisher pipeline, which includes paired-end mapping (PEM) and depth of coverage (DOC) strategies to identify SVs. In collaboration with the group of Stephan Ossowski, we are also developing a method for CNVs detection from multi-sample exome sequencing data to detect heterozygous deletions and duplications as well as higher order genome amplifications, even if more than two alleles exist in the population. We are also working in the identification of germ line mutations (SNVs and SVs) in the CLL project with the aim identify genetic factors that drive the development of several cancer types at their early stages of development. The project will provide gold standard information on the spectrum of SV of the human genome of importance for many human diseases and phenotypic traits.

The group had made specific progress in the dissection of genetics of psoriasis-related phenotypes. We have identified a common CNV that involves the deletion of two genes (LCE3B and LCE3C) in a significant fraction of patients with psoriasis. We have found that patients with rheumatoid arthritis and psoriatic arthritis also have a higher frequency of deletions of LCE3B and LCE3C. Our group has evaluated the effect of CNVs in two broader phenotypes, fibromyalgia and reproductive outcomes, identifying the association of a specific CNV with adverse reproductive outcomes and with fibromyalgia, in which we have identified an association with a CNV in a subset of samples showing highest levels of pain and lower levels of comorbidities.

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. Several disorders (cardiac electric conduction, idiopathic juvenile arthritis, cystic fibrosis, mitral valve prolapse, Parkinson diasese, dominant ataxia and essential tremor) are currently under investigation. The group is also implementing whole-exome sequencing as a diagnostic tool in the clinical setting.

2. Genetic variants responsible for psychiatric disorders susceptibility

E. Docampo, M. Gratacòs, G. Escaramís, 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 in genes known to be involved in the pathophysiology of these diseases. In particular, we have characterized the role of the BDNF Val66Met variant in melancholic depression and obsessive-compulsive disorder, and we have performed further studies on the FTO gene in bulimia and anorexia nervosa. Furthermore, we are participating in a large GWAS for anorexia that is currently in the last phases of the analysis. We have studied 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, where we have detected an association of NRXN3 variants with smoking behavior. 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.

In an effort to find both rare and common variants involved in OCD, we are sequencing the exome of patients with severe and early onset of the disorder, including patients with family history of OCD, early age at onset, severe cases, and patients with good prognosis. The high depth of sequencing achieved allowed us to confidently mine the data for novel SNVs and short InDels discovery. After various filtering steps, we have selected a panel of genes that we have resequenced in a large cohort of cases and controls. All this resulted in the identification various genes and pathways that have a higher number of damaging SNVs in OCD patients than in on-OCD subjects. The approach used in OCD is being applied also to fibromyalgia cases that are very well characterized at the clinical level, with a very close followup of disease evolution.

3. Small non-coding RNAs in neurological-neurodegenerative disorders

M. Bañez-Coronel, M. Friedländer, B. Kagerbauer, E. Lizano, E. Mateu, E. Martí

We have characterized small non-coding RNAs (sncRNAs), in brain samples from individuals at different evolutionary stages of Parkinson's disease (PD) and control age-matched individuals, using illumina-based high-throughput sequencing. We have sequenced small-RNAs of the amygdala of individuals with diagnosed PD (Braak stages 4-5), Individuals at Braak stages 2-3 and control individuals without major histopathological lesionas and no neurological signs. The individuals at Braak stages 2-3 are considered pre-motor or pre-clinical cases of PD, showing incipient Lewy body (LB) lesions at certain brain areas, although they did not report neurological abnormalities. Our analyses suggest a strong sncRNA expression deregulation at pre-motor cases, including miRNAs and sncRNAs derived from tRNAs, snoRNAs and lncRNAs. Importantly, around 50 % of the differentially-expressed sncRNAs (miRNAs and others) were also deregulated in the motor cases, suggesting that the expression of sncRNAs is an early phenomena that may underlie epigenetic gene expression perturbations. Importantly, the majority of sncRNAs, specially miRNAs, were downregulated in both in pre-motor and motor cases. Preliminary analyses sncRNA sequencing data of the frontal cortex of PD cases (Brak stages 4-5) that show few LB pathology revealed a considerable number of overlapping deregulated miRNAs, further indicating that deregulation is not strictly linked to the degree of histopathological lesions.

The group is also studying the role of small non-coding RNAs in neurodegeneration associated to triplet repeat expansion diseases such as Huntington's disease (HD) and fragile X-associated tremor ataxia syndrome (FXTAS). The group is evaluating the role of small CAG and CGG repeated RNA (sCAG and sCGG) in generating toxic effects in neurons. During this year we have been evaluating the mechanisms associated to the protective effects of anti-sCAG molecules in HD. Intra-striatal injection of anti-sCAG into R6/2 HD mouse model resulted in an improvement of the motor deficits. We are currently investigating the mechanisms associated to the improvement of the motor deficits, including evaluation of early transcriptional changes. Transcriptomic modulation by sCAG is also being evaluated in a neuronal cell model. Finally, we have been involved in the evaluation of the transcriptional perturbations associated to CGG expansion in FMR1 5'-UTR in FXTAS patients.

PUBLICATIONS

Articles

Alonso P, Gratacós M, Segalàs C, Escaramís G, Real E, Bayés M, Labad J, López-Solà C, Estivill X, Menchón JM.
"Association between the NMDA glutamate receptor GRIN2B gene and obsessive-compulsive disorder."
Psychiatry Neurosci, 37(4):273-81 (2012).

Bañez-Coronel M, Porta S, Kagerbauer B, Mateu-Huertas E, Pantano L, Ferrer I, Guzmán M, Estivill X, Martí E.
"A pathogenic mechanism in Huntington's disease involves small CAG-repeated RNAs with neurotoxic activity."
PLoS Genet, 8(2):e1002481 (2012).

Bradfield JP, Taal HR, Timpson NJ, Scherag A, Lecoeur C, Warrington NM, Hypponen E, Holst C, Valcarcel B, Thiering E, Salem RM, Schumacher FR, Cousminer DL, Sleiman PM, Zhao J, Berkowitz RI, Vimaleswaran KS, Jarick I, Pennell CE, Evans DM, St Pourcain B, Berry DJ, Mook-Kanamori DO, Hofman A, Rivadeneira F, Uitterlinden AG, van Duijn CM, van der Valk RJ, de Jongste JC, Postma DS, Boomsma DI, Gauderman WJ, Hassanein MT, Lindgren CM, Mägi R, Boreham CA, Neville CE, Moreno LA, Elliott P, Pouta A, Hartikainen AL, Li M, Raitakari O, Lehtimäki T, Eriksson JG, Palotie A, Dallongeville J, Das S, Deloukas P, McMahon G, Ring SM, Kemp JP, Buxton JL, Blakemore AI, Bustamante M, Guxens M, Hirschhorn JN, Gillman MW, Kreiner-Møller E, Bisgaard H, Gilliland FD, Heinrich J, Wheeler E, Barroso I, O'Rahilly S, Meirhaeghe A, Sørensen TI, Power C, Palmer LJ, Hinney A, Widen E, Farooqi IS, McCarthy MI, Froguel P, Meyre D, Hebebrand J, Jarvelin MR, Jaddoe VW, Smith GD, Hakonarson H, Grant SF; Early Growth Genetics Consortium.
"A genome-wide association meta-analysis identifies new childhood obesity loci."
Nat Genet, 44(5):526-31 (2012).

Bustamante M, Danileviciute A, Espinosa A, Gonzalez JR, Subirana I, Cordier S, Chevrier C, Chatzi L, Grazuleviciene R, Sunyer J, Ibarluzea J, Ballester F, Villanueva CM, Nieuwenhuijsen M, Estivill X, Kogevinas M.
"Influence of fetal glutathione S-transferase copy number variants on adverse reproductive outcomes."
BJOG, 119(9):1141-6 (2012).

Cardoner N, Soria V, Gratacòs M, Hernández-Ribas R, Pujol J, López-Solà M, Deus J, Urretavizcaya M, Estivill X, Menchón JM, Soriano-Mas C.
"Val66met BDNF genotypes in melancholic depression: effects on brain structure and treatment outcome."
Depress Anxiety, doi: 10.1002/da.22025. Epub 2012 Nov 16.

Dannemann M, Nickel B, Lizano E, Burbano HA, Kelso J.
"Annotation of primate miRNAs by high throughput sequencing of small RNA libraries."
BMC Genomics, 13:116 (2012).

Dannemann M, Prüfer K, Lizano E, Nickel B, Burbano HA, Kelso J.
"Transcription factors are targeted by differentially expressed miRNAs in primates."
Genome Biol Evol, 4(4):552-64 (2012). (*)

Docampo E, Ribasés M, Gratacòs M, Bruguera E, Cabezas C, Sánchez-Mora C, Nieva G, Puente D, Argimon-Pallàs JM, Casas M, Rabionet R, Estivill X.
"Association of Neurexin 3 polymorphisms with smoking behavior."
Genes Brain Behav, 11(6):704-11 (2012).

Duarte-Salles T, Mendez MA, Morales E, Bustamante M, Rodríguez-Vicente A, Kogevinas M, Sunyer J.
"Dietary benzo(a)pyrene and fetal growth: effect modification by vitamin C intake and glutathione S-transferase P1 polymorphism."
Environ Int, 45:1-8 (2012).

Esko T, Mezzavilla M, Nelis M, Borel C, Debniak T, Jakkula E, Julia A, Karachanak S, Khrunin A, Kisfali P, Krulisova V, Aušrelé Kučinskienė Z, Rehnström K, Traglia M, Nikitina-Zake L, Zimprich F, Antonarakis SE, Estivill X, Glavač D, Gut I, Klovins J, Krawczak M, Kučinskas V, Lathrop M, Macek M, Marsal S, Meitinger T, Melegh B, Limborska S, Lubinski J, Paolotie A, Schreiber S, Toncheva D, Toniolo D, Wichmann HE, Zimprich A, Metspalu M, Gasparini P, Metspalu A, D'Adamo P.
"Genetic characterization of northeastern Italian population isolates in the context of broader European genetic diversity."
Eur J Hum Genet, 2012 Dec 19. doi: 10.1038/ejhg.2012.229. [Epub ahead of print]

Fernàndez-Castillo N, Roncero C, Grau-Lopez L, Barral C, Prat G, Rodriguez-Cintas L, Sánchez-Mora C, Gratacòs M, Ramos-Quiroga JA, Casas M, Ribasés M, Cormand B.
"Association study of 37 genes related to serotonin and dopamine neurotransmission and neurotrophic factors in cocaine dependence."
Genes Brain Behav, in press.

Fernández-Navarro P, Vaquero-Lorenzo C, Blasco-Fontecilla H, Díaz-Hernández M, Gratacòs M, Estivill X, Costas J, Carracedo Á, Fernández-Piqueras J, Saiz-Ruiz J, Baca-Garcia E.
"Genetic epistasis in female suicide attempters."
Prog Neuropsychopharmacol Biol Psychiatry, 38(2):294-301 (2012).

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, 27(5):386-90 (2012).

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, 43(2):897-909 (2012).

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 (2012).

Horikoshi M, Yaghootkar H, Mook-Kanamori DO, Sovio U, Taal HR, Hennig BJ, Bradfield JP, St Pourcain B, Evans DM, Charoen P, Kaakinen M, Cousminer DL, Lehtimäki T, Kreiner-Møller E, Warrington NM, Bustamante M, Feenstra B, Berry DJ, Thiering E, Pfab T, Barton SJ, Shields BM, Kerkhof M, van Leeuwen EM, Fulford AJ, Kutalik Z, Zhao JH, den Hoed M, Mahajan A, Lindi V, Goh LK, Hottenga JJ, Wu Y, Raitakari OT, Harder MN, Meirhaeghe A, Ntalla I, Salem RM, Jameson KA, Zhou K, Monies DM, Lagou V, Kirin M, Heikkinen J, Adair LS, Alkuraya FS, Al-Odaib A, Amouyel P, Andersson EA, Bennett AJ, Blakemore AI, Buxton JL, Dallongeville J, Das S, de Geus EJ, Estivill X, Flexeder C, Froguel P, Geller F, Godfrey KM, Gottrand F, Groves CJ, Hansen T, Hirschhorn JN, Hofman A, Hollegaard MV, Hougaard DM, Hyppönen E, Inskip HM, Isaacs A, Jørgensen T, Kanaka-Gantenbein C, Kemp JP, Kiess W, Kilpeläinen TO, Klopp N, Knight BA, Kuzawa CW, McMahon G, Newnham JP, Niinikoski H, Oostra BA, Pedersen L, Postma DS, Ring SM, Rivadeneira F, Robertson NR, Sebert S, Simell O, Slowinski T, Tiesler CM, Tönjes A, Vaag A, Viikari JS, Vink JM, Vissing NH, Wareham NJ, Willemsen G, Witte DR, Zhang H, Zhao J; Meta-Analyses of Glucose- and Insulin-related traits Consortium (MAGIC), Wilson JF, Stumvoll M, Prentice AM, Meyer BF, Pearson ER, Boreham CA, Cooper C, Gillman MW, Dedoussis GV, Moreno LA, Pedersen O, Saarinen M, Mohlke KL, Boomsma DI, Saw SM, Lakka TA, Körner A, Loos RJ, Ong KK, Vollenweider P, van Duijn CM, Koppelman GH, Hattersley AT, Holloway JW, Hocher B, Heinrich J, Power C, Melbye M, Guxens M, Pennell CE, Bønnelykke K, Bisgaard H, Eriksson JG, Widén E, Hakonarson H, Uitterlinden AG, Pouta A, Lawlor DA, Smith GD, Frayling TM, McCarthy MI, Grant SF, Jaddoe VW, Jarvelin MR, Timpson NJ, Prokopenko I, Freathy RM; Early Growth Genetics (EGG) Consortium.
"New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism."
Nat Genet, 45(1):76-82 (2012).

Krug T, Gabriel JP, Taipa R, Fonseca BV, Domingues-Montanari S, Fernandez-Cadenas I, Manso H, Gouveia LO, Sobral J, Albergaria I, Gaspar G, Jiménez-Conde J, Rabionet R, Ferro JM, Montaner J, Vicente AM, Silva MR, Matos I, Lopes G, Oliveira SA.
"TTC7B emerges as a novel risk factor for ischemic stroke through the convergence of several genome-wide approaches."
J Cereb Blood Flow Metab, 32(6):1061-72 (2012).

Louis-Dit-Picard H, Barc J, Trujillano D, Miserey-Lenkei S, Bouatia-Naji N, Pylypenko O, Beaurain G, Bonnefond A, Sand O, Simian C, Vidal-Petiot E, Soukaseum C, Mandet C, Broux F, Chabre O, Delahousse M, Esnault V, Fiquet B, Houillier P, Bagnis CI, Koenig J, Konrad M, Landais P, Mourani C, Niaudet P, Probst V, Thauvin C, Unwin RJ, Soroka SD, Ehret G, Ossowski S, Caulfield M; International Consortium for Blood Pressure (ICBP), Bruneval P, Estivill X, Froguel P, Hadchouel J, Schott JJ, Jeunemaitre X.
"KLHL3 mutations cause familial hyperkalemic hypertension by impairing ion transport in the distal nephron."
Nat Genet, 44(4):456-60, S1-3 (2012).
Erratum in: Nat Genet, 44(5):609 (2012).

Martín-Santos R, Gelabert E, Subirà S, Gutierrez-Zotes A, Langorh K, Jover M, Torrens M, Guillamat R, Mayoral F, Canellas F, Iborra JL, Gratacos M, Costas J, Gornemann I, Navinés R, Guitart M, Roca M, DE Frutos R, Vilella E, Valdés M, Esteve LG, Sanjuan J.
"Research letter: is neuroticism a risk factor for postpartum depression?"
Psychol Med, 42(7):1559-65 (2012).

Mercader JM, González JR, Lozano JJ, Bak M, Kauppinen S, Sumoy L, Dierssen M, Fernández-Aranda F, Visa J, Gratacòs M, Estivill X.
"Aberrant brain microRNA target and miRISC gene expression in the anx/anx anorexia mouse model."
Gene, 497(2):181-90 (2012).

Morales E, Bustamante M, Vilahur N, Escaramis G, Montfort M, de Cid R, Garcia-Esteban R, Torrent M, Estivill X, Grimalt JO, Sunyer J.
"DNA hypomethylation at ALOX12 is associated with persistent wheezing in childhood."
Am J Respir Crit Care Med, 185(9):937-43 (2012).

Müller TD, Greene BH, Bellodi L, Cavallini MC, Cellini E, Di Bella D, Ehrlich S, Erzegovesi S, Estivill X, Fernández-Aranda F, Fichter M, Fleischhaker C, Scherag S, Gratacòs M, Grallert H, Herpertz-Dahlmann B, Herzog W, Illig T, Lehmkuhl U, Nacmias B, Ribasés M, Ricca V, Schäfer H, Scherag A, Sorbi S, Wichmann HE, Hebebrand J, Hinney A.
"Fat mass and obesity-associated gene (FTO) in eating disorders: evidence for association of the rs9939609 obesity risk allele with bulimia nervosa and anorexia nervosa."
Obes Facts, 5(3):408-19 (2012).

Peñas-Lledó EM, Dorado P, Agüera Z, Gratacós M, Estivill X, Fernández-Aranda F, Llerena A.
"CYP2D6 polymorphism in patients with eating disorders."
Pharmacogenomics J, 12(2):173-5 (2012).

Real E, Gratacòs M, Labad J, Alonso P, Escaramís G, Segalàs C, Subirà M, López-Solà C, Estivill X, Menchón JM.
"Interaction of SLC1A1 gene variants and life stress on pharmacological resistance in obsessive-compulsive disorder."
Pharmacogenomics J, 2012 Jul 10. doi: 10.1038/tpj.2012.30. [Epub ahead of print]

Santos JL, Saus E, Smalley SV, Cataldo LR, Alberti G, Parada J, Gratacòs M, Estivill X.
"Copy number polymorphism of the salivary amylase gene: implications in human nutrition research."
J Nutrigenet Nutrigenomics, 5(3):117-31 (2012).

Taal HR, St Pourcain B, Thiering E, Das S, Mook-Kanamori DO, Warrington NM, Kaakinen M, Kreiner-Møller E, Bradfield JP, Freathy RM, Geller F, Guxens M, Cousminer DL, Kerkhof M, Timpson NJ, Ikram MA, Beilin LJ, Bønnelykke K, Buxton JL, Charoen P, Chawes BL, Eriksson J, Evans DM, Hofman A, Kemp JP, Kim CE, Klopp N, Lahti J, Lye SJ, McMahon G, Mentch FD, Müller-Nurasyid M, O'Reilly PF, Prokopenko I, Rivadeneira F, Steegers EA, Sunyer J, Tiesler C, Yaghootkar H; Cohorts for Heart and Aging Research in Genetic Epidemiology Consortium, Breteler MM, Decarli C, Breteler MM, Debette S, Fornage M, Gudnason V, Launer LJ, van der Lugt A, Mosley TH Jr, Seshadri S, Smith AV, Vernooij MW; Early Genetics & Lifecourse Epidemiology Consortium, Blakemore AI, Chiavacci RM, Feenstra B, Fernandez-Banet J, Grant SF, Hartikainen AL, van der Heijden AJ, Iñiguez C, Lathrop M, McArdle WL, Mølgaard A, Newnham JP, Palmer LJ, Palotie A, Pouta A, Ring SM, Sovio U, Standl M, Uitterlinden AG, Wichmann HE, Vissing NH, DeCarli C, van Duijn CM, McCarthy MI, Koppelman GH, Estivill X, Hattersley AT, Melbye M, Bisgaard H, Pennell CE, Widen E, Hakonarson H, Smith GD, Heinrich J, Jarvelin MR, Jaddoe VW; Early Growth Genetics Consortium.
"Common variants at 12q15 and 12q24 are associated with infant head circumference."
Nat Genet, 44(5):532-8 (2012).

Tsoi LC, Spain SL, Knight J, Ellinghaus E, Stuart PE, Capon F, Ding J, Li Y, Tejasvi T, Gudjonsson JE, Kang HM, Allen MH, McManus R, Novelli G, Samuelsson L, Schalkwijk J, Ståhle M, Burden AD, Smith CH, Cork MJ, Estivill X, Bowcock AM, Krueger GG, Weger W, Worthington J, Tazi-Ahnini R, Nestle FO, Hayday A, Hoffmann P, Winkelmann J, Wijmenga C, Langford C, Edkins S, Andrews R, Blackburn H, Strange A, Band G, Pearson RD, Vukcevic D, Spencer CC, Deloukas P, Mrowietz U, Schreiber S, Weidinger S, Koks S, Kingo K, Esko T, Metspalu A, Lim HW, Voorhees JJ, Weichenthal M, Wichmann HE, Chandran V, Rosen CF, Rahman P, Gladman DD, Griffiths CE, Reis A, Kere J; Collaborative Association Study of Psoriasis (CASP); Genetic Analysis of Psoriasis Consortium; Psoriasis Association Genetics Extension; Wellcome Trust Case Control Consortium 2, Nair RP, Franke A, Barker JN, Abecasis GR, Elder JT, Trembath RC.
"Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity."
Nat Genet, 44(12):1341-8 (2012).

Vilahur N, Molina-Molina JM, Bustamante M, Murcia M, Arrebola JP, Ballester F, Mendez MA, Garcia-Esteban R, Guxens M, Santa Marina L, Tardón A, Sunyer J, Olea N, Fernandez MF.
"Male specific association between xenoestrogen levels in placenta and birthweight."
Environ Int, doi: 10.1016/j.envint.2012.10.004. Epub 2012 Dec 21.

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, 23(1):23-32 (2012).

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 (2012).

Letters

Sullivan P; 96 Psychiatric Genetics Investigators.
"Don't give up on GWAS."
Mol Psychiatry, 17(1):2-3 (2012).

Adams D, Altucci L, Antonarakis SE, Ballesteros J, Beck S, Bird A, Bock C, Boehm B, Campo E, Caricasole A, Dahl F, Dermitzakis ET, Enver T, Esteller M, Estivill X, Ferguson-Smith A, Fitzgibbon J, Flicek P, Giehl C, Graf T, Grosveld F, Guigo R, Gut I, Helin K, Jarvius J, Küppers R, Lehrach H, Lengauer T, Lernmark Å, Leslie D, Loeffler M, Macintyre E, Mai A, Martens JH, Minucci S, Ouwehand WH, Pelicci PG, Pendeville H, Porse B, Rakyan V, Reik W, Schrappe M, Schübeler D, Seifert M, Siebert R, Simmons D, Soranzo N, Spicuglia S, Stratton M, Stunnenberg HG, Tanay A, Torrents D, Valencia A, Vellenga E, Vingron M, Walter J, Willcocks S.
"BLUEPRINT to decode the epigenetic signature written in blood."
Nat Biotechnol, 30(3):224-6 (2012).

(*) This publication results from the work of Dr. Esther Lizano at the Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

GENES AND DISEASE

Group: Gene Function

Susana de la Luna is an ICREA Research Professor.

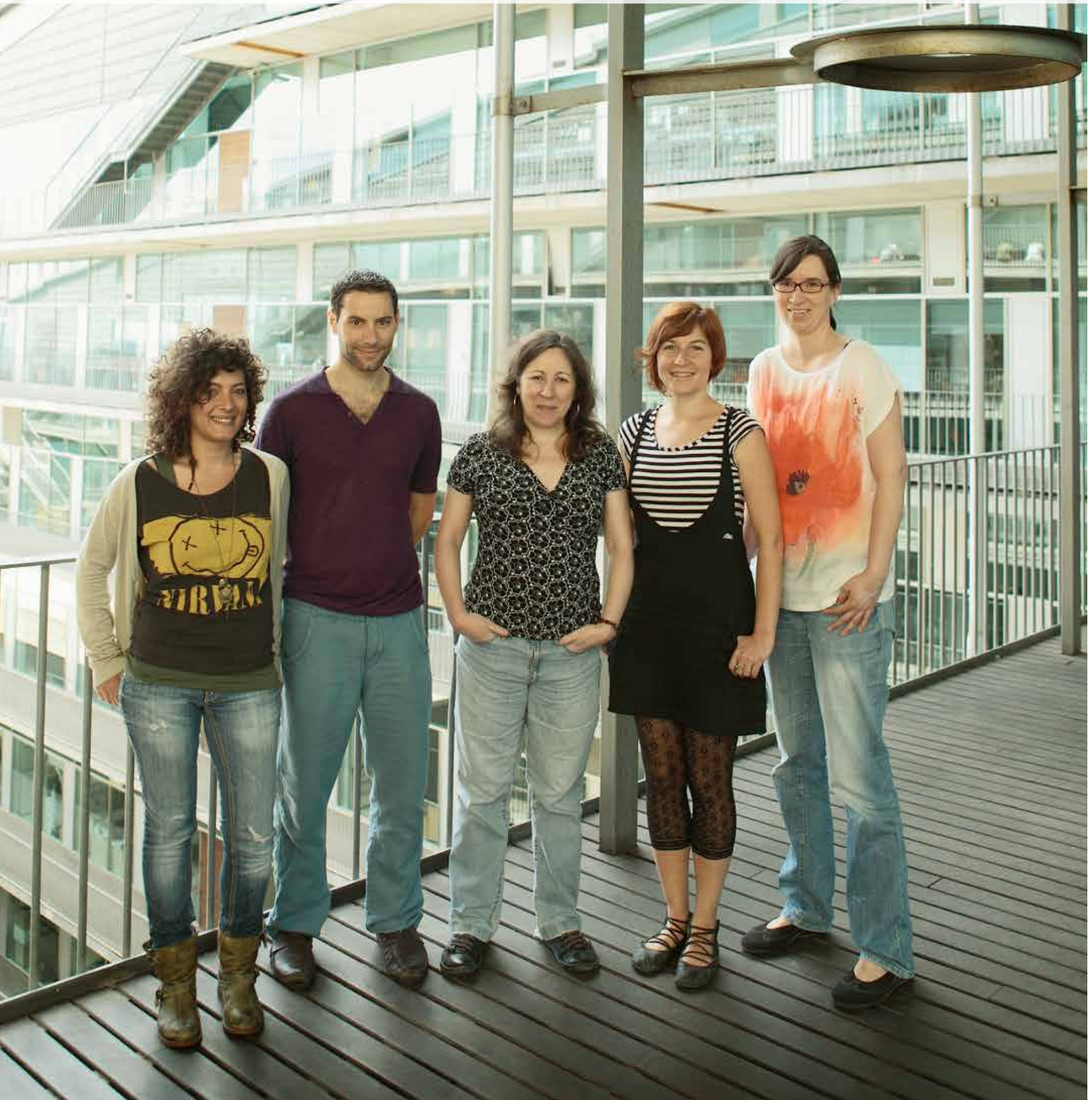
Group Structure:

Group Leader: Susana de la Luna

Postdoctoral: Esteban Rozen (CIBERER researcher)

PhD Students: Chiara Di Vona, Julia Rowenstrunk

Technicians: Alicia Raya, Krisztina Arató



SUMMARY

Dual-specificity tyrosine-regulated kinases or dual-specificity yak-related kinases (DYRKs), belong to the CMGC group, which also includes cyclin-dependent kinases (CDKs), mitogen-activated protein kinases (MAPKs), glycogen synthase kinase-3 (GSK3), CDK-like kinases, serine/arginine-rich protein kinases, cdc2-like kinases and RCK kinases. Members of the DYRK family are found in four of the five main taxa (animalia, plantae, fungi and protista), and all DYRK proteins studied to date share common structural, biochemical and functional properties with their ancestors in yeast. The mammalian DYRK subfamily consists of five members: DYRK1A, DYRK1B (also named Mirk), DYRK2, DYRK3 (also named REDK) and DYRK4, and studies of their function indicate that they participate in several signaling pathways critical for developmental processes and cell homeostasis. In particular, *DYRK1A* plays an important role in mammalian brain development. Both its overexpression, as part of the Down syndrome critical region, as well as its haploinsufficiency have been linked to neurodevelopmental alterations in humans. In fact, haploinsufficiency of DYRK1A is proposed to be a distinctive clinical syndrome including, primary microcephaly, intra-uterine growth retardation and behavioural problems.

Our group is interested in understanding the biological roles of DYRK1A and of the mechanisms of its regulation to have an insight on how alterations in DYRK1A gene dosage result in pathological phentotypes. Given that the commonalities and differences between DYRK family members are not well understood, and that it is still an open question whether distinct DYRKs can functionally compensate DYRK1A, we also include differnt DYRK family members in our studies.

RESEARCH PROJECTS

1. Regulation of DYRK1A

As with all DYRKs, activation of DYRK1A is the result of an autophosphorylation event on an essential tyrosine in the activation loop, which produces a constitutively active kinase. Genetic evidence indicates that small changes in DYRK1A expression levels have severe phenotypic consequences, pointing to a tight control of DYRK1A levels and further suggesting that the function of DYRK1A may be controlled by subtle changes. We have continued our studies on the regulation of the subcellular localization of DYRK proteins to identify potential regulatory mechanisms for DYRK1A, and eventually for other DYRKs, as a way of controlling substrate accessibility. Previous work by our group identified the histidine-rich segment within the DYRK1A carboxy-end as a signal for targeting to the subnuclear compartment known as nuclear speckles or the splicing factor compartment (Figure 1), and we have initiated a search for proteins responsible for this recruitment. Moreover, we have extended our findings in DYRK1A to its paralogous DYRK1B, and found a striking conservation of the protein sequences involved in nuclear import for both the bipartite nuclear localization signal (NLK) within the non-catalytic N-terminus and the complex NLK within the catalytic domain. The sequence conservation results in a similar ability for both DYRK1A and DYRK1B to interact with distinct members of the importin alpha family.

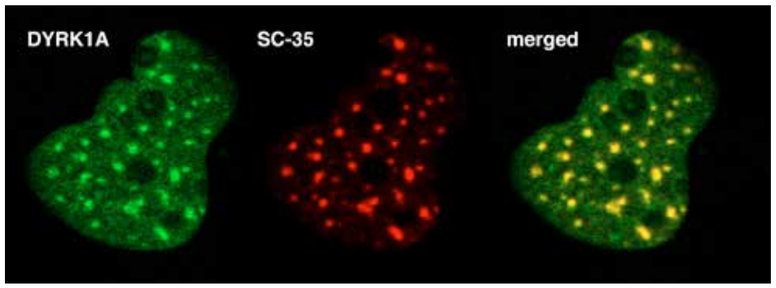


Figure 1.
DYRK1A accumulates in the splicing factor compartment (also known as nuclear speckles).

2. Towards a comprehensive catalog of DYRK1A substrates

DYRK1A gene dosage sensitivity might be linked to its ability 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. A well-defined interactoma will also contribute to set up a comprehensive catalog for DYRK1A substrates. To this aim, we are including not only binary, direct protein-protein interaction approaches as yeast-two hybrid screens, but also large-scale identification of DYRK1A protein interaction complexes based on affinity purification of endogenous DYRK1A in combination with mass spectrometric detection of bound proteins. Our goal is to systematically describe all physical interactions of DYRK1A. This should lead us to define DYRK1A interaction networks and to study how they are organized depending on the cell type studied or the cellular status (for instance, proliferation vs differentiation). The combination of the interaction data with data coming from quantitative phosphoproteomics approaches will provide us with tools to predict DYRK1A specific signaling networks and therefore to understand DYRK1A involvement in cellular processes.

GENES AND DISEASE

Group: Neurobehavioral Phenotyping of Mouse Models of Disease

Group Structure:

Group Leader:	Mara Dierssen
PhD Students:	Susanna Molas, Davide D'Amico, Meritxell Pons, Debora Masini, Silvina Catuara, Jose Antonio Espinosa, Laura Xicota (collaborator)
Postdoctoral fellows:	Monica Joana Pinto Dos Santos, Maria del Carmen de Agustín Pavón (until January 2012), Thomas Gener
Technicians:	María Martínez de Lagrán Cabredo, Tiziana Pederzani, Susanna Molas (since December 2012)
Mouse phenotyping:	Ignasi Sahún Abizanda (until June 2012)
Master/Graduate Students:	C. Faviez (2012), S. Padrell (2012), E. Passalacqua (2012), V. Brugada (2012), M. Fructuoso (2012), C. Duque (2012), O. Roige (2012)
Visiting scientists:	Jorge Busciglio, University of Irvine, California, USA (July-October 2012), Oliver Stork, University of Magdeburg, Germany (January-July, 2012)
Visiting fellows:	Jacobo Chamorro (UAB), Dor Konforti (University of Jerusalem)

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

1. Intellectual disability

Down syndrome (DS) is characterized by intellectual disability and results from one of the most complex genetic perturbations compatible with survival, trisomy of human chromosome 21 (HSA21). The main questions in the field are in what way deregulation of a myriad of different genes in trisomy 21 perturbs the structure and function of the brain and how environmental factors (epigenetic modifications, effects of environmental interventions) can modify the expression of the phenotype. My lab uses a number of genetically modified mouse models target these questions focusing on the role of selected candidate genes. We also use the Fmr1 knockout mouse, a model of Fragile X syndrome, to perform preclinical studies using drugs targeting small RhoGTPases family members. These proteins are involved in dendritic spine remodelling in synaptic plasticity, and may thus be useful therapeutic agents in FXS. Brain and behavior phenotypic analyses in these mouse models help us to identify the timing and location of divergence between trisomic and euploid individuals and pinpoint important stages in the disorder's natural history.

2. Neuropsychiatric disorders

A. Panic disorder: why the brain works in "anxious mode"?

Panic disorder is a major cause of medical attention with substantial social and health service cost. Based on pharmacological studies, research on its etiopathogenesis has focused on the dysfunction of specific neurotransmitter systems. We proposed that cognitive processes are essential for development of complex anxiety responses and thus, genes involved in synaptic plasticity and remodelling could be good candidates. Using a genetically modified mouse model (TgNTRK3) we demonstrated that NTRK3 gene, which encodes the high affinity receptor of neurotrophin 3, TrkC, confers increased risk for panic disorder (Dierssen *et al.*, 2006). We have addressed the mechanisms that underlie the pathogenetic contribution of TrkC to panic and we discovered that TgNTRK3 mice present an inability to correctly identify the fear-related information (Sahún *et al.*, 2007; Amador *et al.*, 2009; Gallego *et al.*, 2010) and that the three main brain structures that regulate fear (amygdala, hippocampus and prefrontal cortex) present an altered activation pattern (D'Amico, PhD thesis; D'Amico D., *et al.*, in preparation). We are addressing putative molecular players, such as MAPK signalling cascade, that may also be good candidates for therapy (Santos M., *et al.*, in preparation).

B. Nicotine dependence: why expression changes of ion channel genes increase tobacco consumption?

The search for determinants of tobacco use initiation and maintenance, the development of nicotine dependence, and levels of cigarette consumption has broadened over the last five years to include a variety of candidate genes, in addition to sociodemographic, psychosocial, and environmental risk factors. Recent studies have revealed that sequence variants in genes encoding the α3/α5/β4 nicotinic acetylcholine receptor subunits are associated with nicotine dependence. Nicotinic acetylcholine receptors are ligand-gated pentameric ion channels that account for the effects of nicotine. We created transgenic mice overexpressing the three subunits and could demonstrate that the CHRNA5/A3/B4 genomic cluster is responsible for nicotine addiction through modifying the activity of brain regions responsible for the balance between the rewarding and the aversive properties of this drug (Gallego *et al.*, 2012, a, b). Our results also revealed results reveal the involvement of α3/α5/β4 nicotinic receptor subunits in working memory and impulsivity, two behavioural traits directly related to the vulnerability to develop nicotine dependence (Viñals *et al.*, 2012).

C. Obsessive-compulsive behavioural components of obesity: phenotypic roadmap of obesity pandemics

The increase in the incidence of obesity and eating disorders has promoted research aimed at understanding the aetiology of abnormal eating behaviours. Apart from metabolic factors, obesity is caused by overeating. Clinical reports have led to the suggestion that some individuals may develop addictive-like behaviours when consuming palatable foods, and thus, compulsive eating plays a similar dominant role in obesity as compulsive drug taking does in drug addiction. We have developed a behavioural model of the obsessive-compulsive elements involved in eating disorders (J. McDonald *et al.*, in preparation; Heyne *et al.*, 2009; Mercader *et al.*, 2012) and in the bioinformatic analysis of the behavioural structure. Mice that are given free choice between standard chow and a palatable, chocolate-containing diet develop distinct signs of compulsive food taking that appear at an early stage. These include the inability to adapt intake behaviour in periods of limited or bitter-tasting chocolate access, and changes in fine structure of feeding (duration, distribution and recurrence of feeding bouts).

PUBLICATIONS

Articles

Garriga-Canut M, Agustín-Pavón C, Herrmann F, Sánchez A, Dierssen M, Fillat C, Isalan M.
"Synthetic zinc finger repressors reduce mutant huntingtin expression in the brain of R6/2 mice."
Proc Natl Acad Sci U S A, 109(45):E3136-45 (2012).

Altafaj X, Martín E, Ortiz-Abalia J, Valderrama A, Lao-Peregrin C, Dierssen M, Fillat C.
"Normalization of Dyrk1A expression by AAV2/1-shDyrk1A attenuates hippocampal-dependent defects in the Ts65Dn mouse model of Down syndrome."
Neurobiology of Disease, doi: 10.1016/j.nbd.2012.11.017. Epub 2012 Dec 5.

Carrasco P, Sahun I, McDonald J, Ramirez S, Jacas J, Gratacós E, Sierra AY, Serra D, Herrero L, Acker-Palmer A, Hegardt FG, Dierssen M, Casals N.
"Ceramide levels regulated by carnitine palmitoyl transferase 1C control dendritic spine maturation and cognition."
Journal of Biological Chemistry, 287(25):21224-32 (2012).

*Dierssen M, Fedrizzi L, Gomez-Villafuertes R, Martinez de Lagran M, Gutierrez-Adan A, Sahun I, Pintado B, Oliveros JC, Dopazo XM, Gonzalez P, Brini M, Mellstrom B, Carafoli E, Naranjo JR.
"Reduced Mid1 expression and delayed neuromotor development in daDREAM transgenic mice."
Frontiers in Molecular Neuroscience, 5:58 (2012)
(*corresponding author)

Mercader JM, González JR, Lozano JJ, Bak M, Kauppinen S, Sumoy L, Dierssen M, Fernández-Aranda F, Visa J, Gratacòs M, Estivill X.
"Aberrant brain microRNA target and miRISC gene expression in the anx/anx anorexia mouse model."
Gene, 497:181-90 (2012).

Gallego X, Ruiz J, Valverde O, Molas S, Robles N, Sabriá J, Crabbe JC, Dierssen M.
"Transgenic Over Expression of Nicotinic Receptor Alpha 5, Alpha 3, and Beta 4 Subunit Genes Reduces Ethanol Intake in Mice."
Alcohol, 46(3): 205-15 (2012).

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, 22(12):2867-77 (2012).

Gallego X, Molas S, Amador-Arjona A, Marks MJ, Robles N, Murtra P, Armengol LI, 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."
AminoAcids, 43: 897-909 (2012).

Viñals X, Molas S, Gallego X, Fernández-Montes RD, Robledo P, *Dierssen M, Maldonado R.
"Overexpression of a3/a5/b4 nicotinic receptor subunits modifies impulsive-like behaviour."
Drug and Alcohol Dependence, 122:247-52 (2012).
(*corresponding author)

Reviews

Dierssen M.
"Down syndrome: the brain in trisomic mode."
Nature Reviews Neuroscience, 13(12):844-58 (2012).

de la Torre R, Dierssen M.
"Therapeutic approaches in the improvement of cognitive performance in Down syndrome: past, present, and future."
Prog Brain Res, 197:1-14 (2012).

Books

Dierssen M, de la Torre R.
"Down Syndrome: From Understanding the Neurobiology to Therapy."
In: Progress in Brain Research, 197:2-262 (2012).

PATENT

Patent application: Compounds for the treatment of neurodegenerative diseases.
PCT/ES2012/070020

GENES AND DISEASE

Group: Genomic and Epigenomic Variation in Disease

Group Structure:

Group Leader: Stephan Ossowski

Postdoctoral Fellows: Oliver Drechsel, Charlotte Hor

PhD Students: Luis Zapata, Rubayte Rahman (since February 2012), Sergio Espeso (since June 2012), Shalu Jhanwar (since September 2012)

Technicians: Daniela Bezdan

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 ExomeCRG. 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 centres 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

Articles

Paschold A, Jia Y, Marcon C, Lund S, Larson NB, Yeh CT, Ossowski S, Lanz C, Nettleton D, Schnable PS, Hochholdinger F.
"Complementation contributes to transcriptome complexity in maize (Zea mays L.) hybrids relative to their inbred parents."
Genome Res, doi: 10.1101/gr.138461.112. Epub 2012 Oct 19. (*)

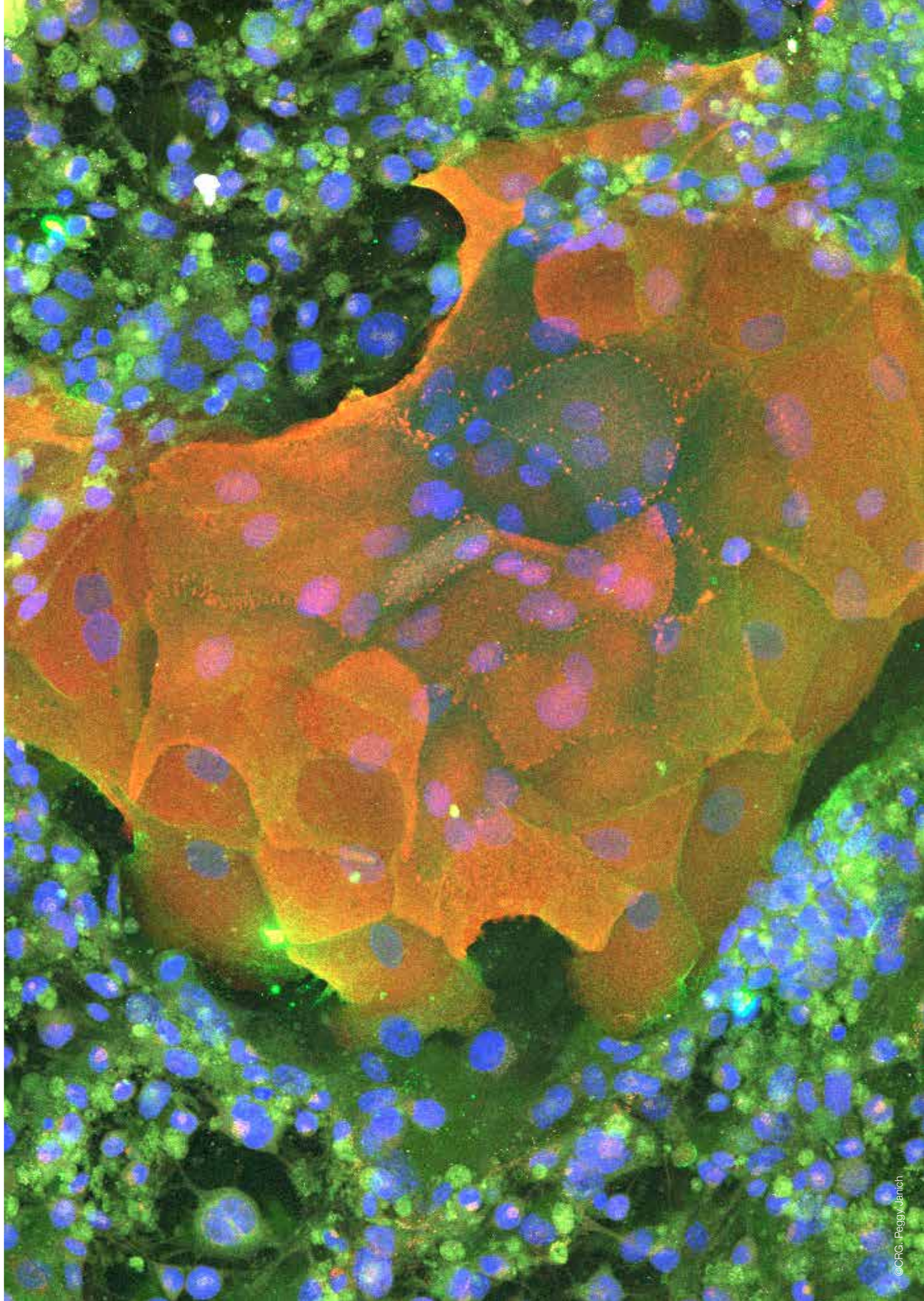
Kopischke M, Westphal L, Schneeberger K, Clark R, Ossowski S, Wewer V, Fuchs R, Landtag J, Hause G, Dörmann P, Lipka V, Weigel D, Schulze-Lefert P,Scheel D, Rosahl S.
"Impaired sterol ester synthesis alters the response of Arabidopsis thaliana to Phytophthora infestans."
Plant J, doi: 10.1111/tpj.12046. Epub 2012 Dec 12.(*)

Sorge S, Ha N, Polychronidou M, Friedrich J, Bezdan D, Kaspar P, Schaefer MH, Ossowski S, Henz SR, Mundorf J, Rätzer J, Papagiannouli F, Lohmann I.
"The cis-regulatory code of Hox function in Drosophila."
EMBO J, 31(15):3323-33 (2012). (*)

Rutter MT, Roles A, Conner JK, Shaw RG, Shaw FH, Schneeberger K, Ossowski S, Weigel D, Fenster CB.
"Fitness of arabidopsis thaliana mutation accumulation lines whose spontaneous mutations are known."
Evolution, 66(7):2335-9 (2012). (*)

Louis-Dit-Picard H, Barc J, Trujillano D, Miserey-Lenkei S, Bouatia-Naji N, Pylypenko O, Beaurain G, Bonnefond A, Sand O, Simian C, Vidal-Petiot E (...) Estivill X, Froguel P, Hadchouel J, Schott JJ, Jeunemaitre X.
"KLHL3 mutations cause familial hyperkalemic hypertension by impairing ion transport in the distal nephron."
Nature Genetics, 44(4):456-60 (2012).
Erratum in: Nat Genet, 44(5):609 (2012).

(*) 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, Pedro Gabriel Dias Ferreira, David González, Angelika Merkel Andrea Tanzer, Dmitry Pervouchine, Marta Melé, Michael Sammeth

PhD Students: Marco Mariotti, Joao Curado, Didac Santesmases, Alessandra Breschi

Technicians: Julien Lagarde, Francisco Câmara, Maik Röder, Carmen Arnán Emilio Palumbo, Maria Sanz

Students: Jean Monlong, Chiara Medoro, Dario Cecchi, Anaïs Gouin, Carles Boix Maitena Tellaeche

Visiting scientists: Timothy Mercer

SUMMARY

Research in our group (http://big.crg.cat/bioinformatics_and_genomics) 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 are continually working on the development of geneid, an *ab initio* gene prediction program, and SGP2 a comparative gene finder. We currently have more than fifty parameter files for gene prediction in a wide range of species spanning all Kingdoms of life. These parameter files are available here: http://genome.crg.es/software/geneid/index.html#parameters

In this past year we started collaboration with the University of Bristol (UK) on the genome annotation of two Hymenoptera species, *D.quadriceps* (Dino ant) and *P.canadensis* (a wasp species). In these two projects in addition to obtaining geneid and sgp2 parameter files we are also using a well-established gene annotation combiner pipeline (EVM –evidence modeler-) in combination with PASA (a program to assemble spliced alignments) as tools to annotate these organisms. The EVM/PASA pipelines combine *ab initio* or evidence-based gene predictions (such as geneid and sgp2) with protein homology and transcript evidence to potentially improve protein-coding genome annotation significantly.

We also continued our collaboration with the Baylor College of Medicine (Houston, USA) in the annotation of several hymenoptera species (*B.terrestris/B.impatiens/A.mellifera*). The annotation of these genomes has recently been completed and will incorporate both geneid and sgp2 gene predictions for all three species. We are currently assisting in the preparation of a draft of a manuscript describing the latest annotation of the *A.mellifera* with a planned submission date in the first quarter of 2013.

In the course of 2012, and as members of the INB ("Instituto Nacional de Bioinformatica") we also continued/concluded our participation in three large-scale genome annotation projects:

1- Annotation of the *Cucumis metuliferus* (melon) genome. As members of the melonomics consortium we have developed melon-specific geneid and SGP2 parameter files which were used on the final pipeline to annotate the *Cucumis metuliferus* (melon) genome. The work on this project was completed in 2012 and resulted in a PNAS publication (Garcia-Mas J. *et al.*, 2012) It is important to note that this was the largest Spanish-funded genome sequencing/annotation project to date.

2- Sequencing and Annotation of the Common Bean Genome: Maximization of the Latin American/Iberian natural resources (PhasIbeam consortium). On this ongoing project we have so far developed a preliminary set of geneid and SGP2 parameter files that seem to predict genes in this species with high accuracy. We are also involved in the coordination of this project and have collaborated in the creation of both a google group (http://groups.google.com/group/phasibeam?hl=en and a wiki (http://phasibeam.crg.eu/wiki/Main_Page) in order to help organize and centralize this international project. We are also assisting our collaborators at the CNAG in obtaining a final annotation. Furthermore we are going use the *in house* tool GRAPE (General RNASeq Analysis Pipeline Environment) to also study the differential expression of genes in different conditions. We are currently in the final stages of this project and expect to have a draft manuscript for publication completed by early March 2013.

3- Assembly, annotation and comparative analysis of Iberian and European lynx genomes. We are also continuing with our participation in the Iberian lynx (a threatened species) sequencing/annotation effort. A final assembly of this genome should be made available by our CNAG collaborators by the middle of February 2013. Our task will be to assist in the annotation of the genome by generating both geneid and sgp2 predictions for this species. We will also study the differential expression of genes in different tissues by using the *in house* pipeline GRAPE.

Finally, the work generated from another genome sequencing/annotation project we have participated in ("Annotation of the *Solanum lycopersicum* (tomato) genome") was also published in the journal *Nature* in 2012 (Tomato Genome Consortium, 2012)

2. Prediction of selenoproteins

Particularly difficult in eukaryotic genomes is the prediction of selenoprotein genes, because selenocysteine is specified by the UGA codon, normally a stop codon (Mariotti *et al*., 2012). We have been developing computational methods for selenoprotein prediction since early 2000. In the last year we have continued our work on selenoproteins that have been lost independently in a number of different insect lineages to pinpoint the mechanisms of selenoprotein extinction in insects. We have now genome assemblies for 8 *Drosophila* species from the Saltans group, which we identified as interesting for selenoproteins from preliminary PCR experiments. During 2012, we worked to obtain a robust phylogenetic tree, defining the topology of the Saltans group and also of all other *Drosophilas*, allowing us to draw the full history of events leading to Sec extinction. Searching selenoproteins and selenocysteine machinery genes, we found an interesting panorama of selenoprotein presence: although most of the saltans species possess the full selenoprotein repertoire, 3 species were found to have lost selenocysteine: *D.sturtevanti* and *D.milleri* (clustering together phylogenetically, suggesting a common selenoprotein extinction before their split), and *D.neocordata*. This is the most interesting species for selenoproteins: it appears to have lost or converted to cysteine 2 selenoproteins, while a third one remains apparently intact. This last selenoprotein is SPS2, whose function is the production of selenocysteine itself. The rest of Sec machinery is almost entirely intact, and yet some features of these genes strongly suggest that selenocysteine cannot be coded anymore in this species, and thus even the SPS2 gene cannot functional. We believe that a very recent extinction took place in this species. We may predict that with enough time, Sec machinery will further degenerate here too. This may provide a powerful tool to test neutral models of evolutionary theory. Observing the selenoprotein presence in all *Drosophilas* and in particular in *D.neocordata*, we formulated a model of selenoprotein extinction, in which the selenoprotein machinery can degenerate only after the selenoproteins have been already lost or converted to cysteine homologues. SPS2, playing both as selenoprotein and as selenoprotein machinery, must thus be always the last selenoprotein to be lost, which is what we observe in *D.neocordata*. We then focused on finding the causes of selenoprotein extinctions in the Saltans group, and we noticed that this group is peculiar among *Drosophilas* for several aspects, particularly for GC content and codon bias. We believe that here the selenoprotein extinctions are a downstream effect of a more general process acting in this lineage, involving probably a generalized accelerated evolution of protein coding genes.

To further investigate this, and hoping to gain insights on this general process, we proceeded to annotate all protein coding genes in our genomes. We used a profile-based approach, adapting the program selenoprofiles for this purpose, and we complemented it with *ab initio* predictions by geneid. We are now using these annotations for a few functional analysis, whose aim is to find functional classes that behave in a peculiar way in this specific lineage, either with an overrepresentation in duplicated or lost genes, or a striking pattern of differential expression when compared to non-saltans species, or a particularly high or low evolution rate.

3. Methods for transcriptome analysis and the ENCODE project

The field of transcriptomics has recently been given a huge boost from the use of "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. Therefore 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 (P. Ribeca, now at the CNAG; http://big.crg.cat/services/gem_genome_multi_tool_library), the Flux Capacitor (M. Sammeth; http://big.crg.cat/services/flux_capacitor) for transcript quantification, and NextGeneid for de novo transcript modeling and discovery (T. Alioto, now at the CNAG). GEM has very powerful split mapping module, making it particularly appropriate for RNASeq mapping (Marco-Sola *et al*., 2012.) The Flux Capacitor is a program to produce transcript quantifications from RNASeq data. It takes as input an annotation on a reference genome and a set of RNASeq reads mapped into this genome (Griebel *et al*., 2012)

In the past year we have continued working towards the incorporation of 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 (D. Gonzalez) to transparently produce several useful types of information from the RNASeq data. This pipeline includes quality control of the reads, read alignment using GEM and expression analysis information. It generates transcript models by running cufflinks, transcript quantifications in the form of RPKM (flux capacitor) or read counts, splicing information (novel junctions and exon inclusion index, fusion information (from split mapping and paired end mapping). The pipeline stores the information produced in a database that also allows us to track the pipeline parts that have been completed or those that are out of date. Furthermore, a common interface to both, the data and the analysis results has been developed (M. Röder & J. Lagarde; http://rnaseq.crg.cat/). GRAPE has been used to process the large amount of RNASeq data produced by the ENCODE project as well as in various other projects involving the analysis of large amount of RNASeq data it. GRAPE was released as open-source software in early 2012 (http://big.crg.cat/services/grape), and is being used in production at a number of sites. The development of GRAPE has also been described in a recent publication (Knowles *et al*., 2013). The current 1.9.x releases are in a stable state and will remain in maintenance mode from now on. Work on version 2.0 of Grape has already started in collaboration with the CNAG and commercial third-party developers. An open-source development community has been formed, and help and support options are available through a mailing list.

The GRAPE dashboard, under development, offers a point of entry to the data and results through a schematic representation of the project's experimental design (Figure 1). The dashboard is automatically kept synchronized with data released by the ENCODE Data Coordination Center, but can also list extra file entries absent from that source. It is a widely used resource in the community (5,000 server hits per month on average).

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). Moreover, by analyzing the ENCODE data using these methods, we have found that a large fraction of the variability in transcript abundance can be simply explained by the variability in global gene expression (The ENCODE RNA Group, 2012).

In addition to this, we developed a methodology to quantify complex alternative splicing events and applied it to the analysis of the ENCODE data (Pervouchine *et al*., 2012). 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 (Howald *et al*., 2012).

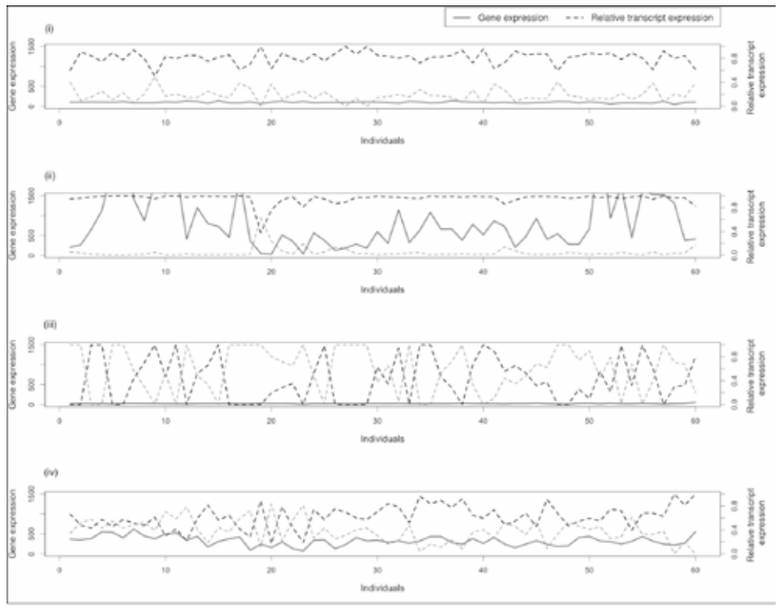


Figure 2
Variability in gene expression versus variability in splicing ratios. Behavior of four different genes regarding expression and splicing variability in human populations. Four genes with two splice forms each have been selected to illustrate possible extreme cases: (i) Low variability in both 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)

4. Modeling splicing from chromatin

In the past year we have used a statistical framework similar to that used to model transcription activity to help us develop predictive models to explore the relationship between levels of histone modifications in the vicinity of exons and the corresponding exon inclusion levels. The contribution of chromatin modifications to splicing is not expected to be dominant, and as a consequence, the predictive power of the exon inclusion models is smaller than that of the models of transcriptional activity. Nevertheless, a number of chromatin marks are identified consistently across cell lines as having a significant association with exon inclusion levels. Some of these markers seem to have a strong positive association with exon inclusion, others while others have a clearly negative association. Importantly, the models inferred in a particular cell type are, in general, also quite accurate in the other cell lines, suggesting that the relationships between histone modifications and alternative splicing being uncovered may be general in scope. Results derived from this work are being compiled to be submitted to a specialized journal in early 2013.

Moreover, in order to validate our results we are also performing wet lab experiments on a subset of our predictions that could confirm our genome-wide results. In order to run these experimental analyses we are continuing our close collaboration with the group of Juan Valcárcel from the CRG's Gene Regulation, Stem Cells and Cancer program, whose work focuses on the investigation of the mechanisms through which splice signals are recognized and processed. Concurrently, we also started to set up in our own wet lab a new *in vivo* system that will allow us to test the directionality of such a relationship between chromatin and splicing.

5. Long Noncoding RNAs with enhancer-like function in human cells

Long Noncoding RNAs (lncRNA) constitute a large portion of the mammalian transcriptome even though their biological functions remained unclear. In this past year we have developed a new custom microarray platform to measure lncRNA expression. We are using these arrays with various collaborators to study lncRNA involvement in (1) haematopoeitic cell differentiation (with Thomas Graf of the CRG), (2) cardiac progenitor cell differentiation (with Thierry Pedrazzinni at the U. of Lausanne), (3) Mathijs Voorhoeve (Duke NUS, Singapore), (4) Leah Vardy (IMB, Singapore). In parallel, we are developing tools to study candidate lncRNAs in loss-of-function assays. We are presently writing a manuscript describing the findings from the collaboration with Pedrazzini. 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. We have been developing a novel reporter plasmid strategy to systematically test the enhancer function of lncRNAs (with Leonardo da Vinci student, Dario Cecchi). In 2012 we published a manuscript in Genome Research, within the ENCODE project consortium, describing the gene structure and expression of the largest collection of human lncRNAs to date: "The GENCODE v7 catalogue of human long non-coding RNAs: Analysis of their gene structure, evolution and expression" (Derrien *et al*., 2012). 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.

Furthermore, we have been performing studies based on the increasing evidence that long non-coding RNA (lncRNA) may play a relevant role in the regulation of several cellular processes. 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. To date, there has not yet been any comprehensive attempt to assess events of positive selection in the recent evolution of lnRNAs. For this purpose Marta Mele *et al.* have compared levels of intraspecific variation of lncRNAS within the human lineage and compared similar intraspecific patterns to their closest living relatives: the non-human great Apes. We are currently studying the patterns of genetic variation of around 10,000 annotated lncRNAs in the human genome across all the major human and Great Ape populations (including full genome data for ~100 great ape genomes--great Ape diversity Consortium). 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.

PUBLICATIONS

Adams D, Altucci L, Antonarakis SE, Ballesteros J, Beck S, Bird A, Bock C, Boehm B, Campo E, Caricasole A, Dahl F, Dermitzakis ET, Enver T, Esteller M, Estivill X, Ferguson-Smith A, Fitzgibbon J, Flicek P, Giehl C, Graf T, Grosveld F, Guigo R, Gut I, Helin K, Jarvius J, Küppers R, Lehrach H *et al*.
"BLUEPRINT to decode the epigenetic signature written in blood."
Nature Biotechnology, 30(3):224-6 (2012).

Derrien T, Guigo R, Johnson R.
"The Long Non-Coding RNAs: A New (P)layer in the "Dark Matter"."
Frontiers in Genetics, 2:107 (2012).

Derrien T, Estelle J, Marco Sola S, Knowles DG, Raineri E, Guigo R, Ribeca P.
"Fast computation and applications of genome mappability."
PLoS One, 7(1):e30377 (2012).

Derrien T. *et al*. (incl. R. Guigo, A. Tanzer, S. Djebali, H.Tilgner, D. Martin, D. Gonzalez, J. Lagarde, R. Johnson and A. Merkel)
"The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression."
Genome Research, 22(9):1775-89 (2012).

Djebali S, Lagarde J, *et al.* (including Guigo, R and Martin D).
"Evidence for transcript networks composed of chimeric RNAs in human cells."
PLoS One, 7(1):e28213 (2012).

Dong X, Greven MC, Kundaje A, Djebali S, Brown JB, Cheng C, Gingeras TR, Gerstein M, Guigo R, Birney E, Weng Z.
"Modeling gene expression using chromatin features in various cellular contexts."
Genome Biology, 13(9):R53 (2012).

Frenkel-Morgenstern M, Lacroix V, Ezkurdia I, Levin Y, Gabashvili A, Prilusky J, Del Pozo A, Tress ML, Johnson R, Guigo R, Valencia A.
"Chimeras taking shape: Potential functions of proteins encoded by chimeric RNA transcripts."
Genome Research, 22(7):1231-42 (2012).

Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, Gonzalez VM, Henaff E, Camara F, Cozzuto L, Lowy E, Alioto T, Gabaldon T, Roma G, Guigo R, Casacuberta JM, Arus P, Puigdomenech P.
"The genome of melon (Cucumis melo L.)."
Proceedings of the National Academy of Sciences of the United States of America, 109(29):11872-7 (2012).

Gonzalez-Porta M, Calvo M, Sammeth M, Guigó R.
"Estimation of alternative splicing variability in human populations."
Genome Research, 22(3):528-38 (2012).

Griebel T, Zacher B, Ribeca P, Raineri E, Lacroix V, Guigo R, Sammeth M.
"Modelling and simulating generic RNA-Seq experiments with the flux simulator."
Nucleic Acids Research, 40(20):10073-83 (2012).

Howald C, Tanzer A, Chrast J, Kokocinski F, Derrien T, Walters N, Gonzalez JM, Frankish A, Aken BL, Hourlier T *et al*.
"Combining RT-PCR-seq and RNA-seq to catalog all genic elements encoded in the human genome."
Genome Res., 22:1698–1710 (2012).

Johnson R, Richter N, Bogu GK, Bhinge A, Teng SW, Choo SH, Andrieux LO, de Benedictis C, Jauch R, Stanton LW.
"A genome-wide screen for genetic variants that modify the recruitment of REST to its target genes."
PLoS Genetics, 8(4):e1002624 (2012).

Johnson R.
*"Long non-coding RNAs in Huntington"*s disease neurodegeneration."
Neurobiology of Disease, 46(2):245-54 (2012).

Johnson R, Noble W, Tartaglia GG, Buckley NJ.
"Neurodegeneration as an RNA Disorder."
Prog Neurobiol, 99(3):293-315 (2012).

Kedzierska AM, Drton M, Guigó R, Casanellas M.
"SPIn: model selection for phylogenetic mixtures via linear invariants."
Molecular Biology and Evolution, 29(3):929-37 (2012).

Marco-Sola S, Sammeth M, Guigó R, Ribeca P.
"The GEM mapper: fast, accurate and versatile alignment by filtration."
Nature Methods, 9(12):1185-8 (2012).

Mariotti M, Ridge PG, Zhang Y, Lobanov AV, Pringle TH, Guigo R, Hatfield DL, Gladyshev VN.
"Composition and evolution of the vertebrate and Mammalian selenoproteomes."
PLoS ONE, 7(3):e33066 (2012).

Mouse ENCODE Consortium (incl. R. Guigo, S. Djebali, J.Lagarde).
"An encyclopedia of mouse DNA elements (Mouse ENCODE)."
Genome Biology, 13(8):418 (2012).

Ng SY, Johnson R, Stanton LW.
"Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors."
EMBO Journal, 31(3):522-33 (2012).

Pervouchine DD, Knowles DG, Guigó R.
"Intron-Centric Estimation of Alternative Splicing from RNA-seq data."
Bioinformatics, doi: 10.1093/bioinformatics/bts678. Epub 2012 Nov 21.

Plass M, Codony-Servat C, Ferreira PG, Vilardell J, Eyras E.
*"RNA secondary structure mediates alternative 3"*ss selection in Saccharomyces cerevisiae."
RNA, 18(6):1103-15 (2012).

The ENCODE Project Consortium, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M.
"An integrated encyclopedia of DNA elements in the human genome."
Nature, 489(7414):57-74 (2012).

The ENCODE RNA Group, including Djebali S, Lagarde J, Merkel A, Tanzer A, Röder M, Ferreira P, Tilgner H, Gonzalez D, Skancke J, Curado J, Derrien T, Ribeca P, Alioto T, Sammeth M, Kingswood C, Johnson R, and R. Guigó.
"Landscape of transcription in human cells."
Nature, 489(7414):101-8 (2012).

Tilgner H, Knowles DG, Johnson R, Davis CA, Chakrabortty S, Djebali S, Curado J, Snyder M, Gingeras TR, Guigó R.
"Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs."
Genome Res, 22(9):1616-25 (2012).

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N, Balasubramanian S, Pei B, Tress M, Rodriguez JM, Ezkurdia I, van Baren J, Brent M, Haussler D, Kellis M, Valencia A, Reymond A, Gerstein M, Guigó R, Hubbard TJ.
"GENCODE: The reference human genome annotation for The ENCODE Project."
Genome Research, 22(9):1760-1774 (2012).

Cheng C, Alexander R, Min R, Leng J, Yip KY, Rozowsky J, Yan KK, Dong X, Djebali S, Ruan Y, Davis CA, Carninci P, Lassman T, Gingeras TR, Guigó R, Birney E, Weng Z, Snyder M, Gerstein M.
"Understanding transcriptional regulation by integrative analysis of transcription factor binding data."
Genome Research, 22(9):1658-67 (2012).

Alemany A, Mossa A, Junier I, Ritort F.
"Experimental free-energy measurements of kinetic molecular states using fluctuation theorems."
Nature Physics, 8(9):688-694 (2012).

The Tomato Genome Consortium. (incl. F. Camara and R. Guigo)
"The tomato genome sequence provides insights into fleshy fruit evolution."
Nature, 485(7400):635-41 (2012).

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, Maria Hatzou

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” 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. [1-7]

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. This presents us with a major challenge, at a time where accurate non-coding alignments have never been in such dire need. This is well illustrated by the rapid growth of the non-coding component of Gencode [4], in which we contributed (in collaboration with Roderic Guigo) to the report of over 7.000 new non-coding transcripts. We are now actively investigating how such experimental data can be used to transfer information onto un-annotated genomes, using a combination of homology based and experimental RNA-Seq data. This new framework, named PipeR was used to do the evolutionary profiling of the Genecode 7 human lncRNA complement (Figure 1) and is now further developed as an annotation platform.

PipeR depends heavily on 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.

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. 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 through the development of very large scale multiple sequence alignments. We recently reported a new flavour of the T-Coffee package able to align up to a 100.000 protein sequences [7]. These models were shown by the group of Fyodor Kondrashov to support the elaboration of sophisticated models establishing the role of epistasis in the process of molecular evolution. We are also addressing the issue of generating very accurate alignments using either evolutionary information [6] or structural information [2]. Such alignments are also meant to support detailed structural analysis and one of our mid term objectives is to show how structure based analysis can be used to determine complex evolutionary histories poorly supported by the comparison of distantly related sequences.

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 characterization 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 [3] 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 ChIp-Seq datasets (done on transcription factors) and more traditional comparative genomics processes.

PUBLICATIONS

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 (2012).

Magis C, van der Sloot AM, Serrano L and Notredame C.
"An improved understanding of TNFL/TNFR interactions using structure-based classifications."
Trends Biochem Sci, 37(9):353-363 (2012).

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, 40(7):e52 (2012).

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG *et al*.
"The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression."
Genome Res, 22(9):1775-89 (2012).

ENCODE consortium.
"An integrated encyclopedia of DNA elements in the human genome."
Nature, 7414(489):57-74 (2012).

Chang JM, Di Tommaso P, Taly JF and Notredame C.
"Accurate multiple sequence alignment of transmembrane proteins with PSI-Coffee."
BMC Bioinformatics, 13(Suppl 4):S1 (2012).

Breen MS, Kemena C, Vlasov PK, Notredame C and Kondrashov FA.
"Epistasis as the primary factor in molecular evolution."
Nature, 490(7421):535-8 (2012).

Orobitg M, Cores F, Guirado F, Kemena C, Notredame C, Ripoll, A.
"Enhancing the scalability of consistency-based progressive multiple sequences alignment applications."
Proceedings of the 2012 IEEE 26th International Parallel and Distributed Processing Symposium (IPDPS 2012), art. no. 6267825, pp. 71-82 (2012).

Arnold P, Erb I, Pachkov M, Molina N, van Nimwegen E.
"MotEvo: Integrated bayesian probabilistic methods for inferring regulatory sites and motifs on multiple alignments of DNA sequences."
Bioinformatics, 28(4):487-94 (2012).

Su EC, Chang JM, Cheng CW, Sung TY, Hsu WL.
"Prediction of nuclear proteins using nuclear translocation signals proposed by probabilistic latent semantic indexing."
BMC Bioinformatics, 13 Suppl 17:S13 (2012).

BIOINFORMATICS AND GENOMICS

Group: Comparative Genomics

Group Structure:

Group Leader: Toni Gabaldón

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

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

Technicians: Marina Marcet-Houben, Ester Saus

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 have 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. Evolutionary genomics of long, non-coding RNAs

Recent genomics analyses have facilitated the discovery of a novel major class of stable transcripts, now called long non-coding RNAs (lncRNAs). A growing number of analyses have implicated lncRNAs in the regulation of gene expression, dosage compensation and imprinting, and there is increasing evidence suggesting the involvement of lncRNAs in various diseases such as cancer. Despite recent advances, however, the role of the large majority of lncRNAs remains unknown and there is current debate on what fraction of lncRNAs may just represent transcriptional noise. Moreover, despite a growing number of lncRNAs catalogues for diverse model species, we lack a proper understanding of how these molecules evolve across genomes. Evolutionary analyses of protein-coding genes have proved tremendously useful in elucidating functional relationships and in understanding how the processes in which they are involved are shaped during evolution. Similar insights may be expected from a proper evolutionary characterization of lncRNAs, although the lack of proper tools and basic knowledge of underlying evolutionary mechanisms are a sizable challenge. Our group has embarked into a project, funded by the European Research Council that aims to combine state-of-the-art computational and sequencing techniques in order to elucidate what evolutionary mechanisms are shaping this enigmatic component of eukaryotic genomes. The first goal is to enable large-scale phylogenomic analyses of lncRNAs by developing, for these molecules, methodologies that are now standard in the evolutionary analysis of protein-coding genes. The second goal is to explore, at high levels of resolution, the evolutionary dynamics of lncRNAs across selected eukaryotic groups for which novel genome-wide data will be produced experimentally. In particular, we plan to exploit RNAseq experiments and, most importantly, recently developed sequencing techniques that enable obtaining genome-wide footprints of RNA secondary structure. This will allow us to trace the conservation of lncRNA structures across species, information that, in turn, will be instrumental for the informed development of phylogenetic algorithms. Finally, this dataset will be used to test the impact on lncRNAs evolution of processes such as, among others, purifying and positive selection, gene duplication, co-evolution and gene conversion. The global objective is to understand how lncRNAs evolve across species, and how this relates to the evolution of protein-coding genes and the biological processes to which they may be related. Potential outcomes of this research are the discovery of novel evolutionary mechanisms, the development of lncRNA-tailored phylogenetic algorithms, and the prediction of potential functional associations between lncRNAs, proteins, and biological processes.

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 and 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.

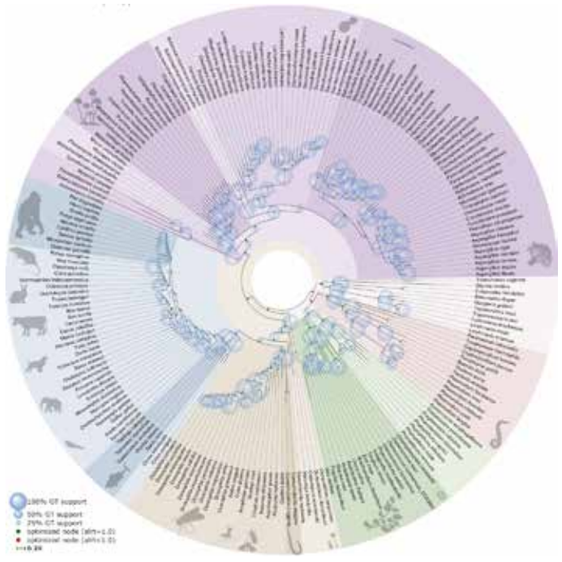


Figure 1.
Eukaryotic tree of life comprising 242 species with fully-sequenced genomes. We developed this using a new technique developed in the group called Nested Phylogenetic Reconstruction. Figure is represented using our tool ETE (http://ete.cgenomics.org). Spheres in nodes represent congruency with the reconstructed node with collections of individual gene trees (GT: Gene Tree support).

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

Ollivier M, Gabaldón T, Poulain J, Gavory F, Leterme N, Gauthier JP, Legeai F, Tagu D, Simon JC, Rispe C.
"Comparison of gene repertoires and patterns of evolutionary rates in eight aphid species that differ by reproductive mode."
Genome Biol Evo, 4(2):155-67 (2012).

Bohle MH and Gabaldón T.
"Selection of marker genes using whole-genome DNA polymorphism analysis."
Evol. Bioinf, 8:161-169 (2012).

Dessimoz C, Gabaldón T, Roos DS, Sonnhammer EL, Herrero J and the Quest for Orthologs Consortium.
"Toward Community Standards in the Quest for Orthologs."
Bioinformatics, 28(6):900-904 (2012).

Lee JH, Wan KL, Mohd-Adnan A, Gabaldón T.
"Evolution of the ferritin family in vertebrales."
Trends in Evol Biol, 4:e3 (2012).

Capella-Gutíerrez S, Marcet-Houben M, Gabaldón T
"Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi."
BMC Biology, 10(1):47 (2012).

Piškur J, Ling Z, Marcet-Houben M, Ishchuk OP, Aerts A, Labutti K, Copeland A, Lindquist E, Barry K, Compagno C, Bisson L, Grigoriev IV, Gabaldón T, Phister T.
"The genome of wine yeast Dekkera bruxellensis provides a tool to explore its food-related properties."
Int J Food Microbiol, 157(2):202-9 (2012).

Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, González VM, Hénaff E, Câmara F, Cozzuto L, Lowy E, Alioto T, Capella-Gutiérrez S, Blanca J, Cañizares J, Ziarsolo P, Gonzalez-Ibeas D, Rodríguez-Moreno L, Droege M, Du L, Alvarez-Tejado M, Lorente-Galdos B, Melé M, Yang L, Weng Y, Navarro A, Marques-Bonet T, Aranda MA, Nuez F, Picó B, Gabaldón T, Roma G, Guigó R, Casacuberta JM, Arús P, Puigdomènech P.
"The genome of melon (Cucumis melo L.)."
PNAS, 109(29):11872-11877 (2012).

Valach M, Pryszcz LP, Tomaska L, Gacser A, Gabaldón T, Nosek J.
"Mitochondrial genome variability within the Candida parapsilosis species complex."
Mitochondrion, 12(5):514-519 (2012).

Silva LL, Marcet-Houben M, Nahum LA, Zerlotini A, Gabaldón T, Oliveira G.
"The Schistosoma mansoni phylome: using evolutionary genomics to gain insight into a parasite's biology."
BMC Genomics, 13:617 (2012).

Marcet-Houben M, Ballester AR, de la Fuente B, Harries E, Marcos JF, González-Candelas L, Gabaldón T.
"Genome sequence of the necrotrophic fungus Penicillium digitatum, the main postharvest pathogen of citrus."
BMC Genomics, 13:646 (2012).

Gabaldón T.
"A genomics perspective on the evolutionary emergence of pathogenesis."
Mycoses, 55(4):10-10 (2012).

De Vienne DM, Ollier S, Aguileta G.
"Phylo-MCOA: A fast and efficient method to detect outlier genes and species in phylogenomics using multiple co-inertia analysis."
Molecular Biology and Evolution, 29(6):1587-1598 (2012).

Petit E, Giraud T, de Vienne DM, Coelho MA, Aguileta G, Amselem J, Kreplak J, Poulain J, Gavory F, Wincker P, Young SK, Cuomo C, Perlin MH, Hood ME.
"Linkage To The Mating-Type Locus Across The Genus Microbotryum: Insights Into Nonrecombining Chromosomes."
Evolution, 66(11):3519-33 (2012).

Mohd-Padil H, Mohd-Adnan A, Gabaldón T.
"Phylogenetic Analyses Uncover a Novel Clade of Transferrin in Nonmammalian Vertebrates."
Mol Biol Evol, in press

Reviews

Miranda-Saavedra D, Gabaldón T, Barton GJ, Langsdey G, Doering C.
"The kinomes of apicomplexan parasites."
Microbes and Infection, 14(10):796-810 (2012).

BIOINFORMATICS AND GENOMICS

Group: Evolutionary Genomics
Fyodor Kondrashov is an ICREA Research Professor.

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, Carla Bello (after a Master Degree done in part in our lab)
Undergraduate and high-school volunteers:	Oriol Pich i Roselló, Aina Solà i Boloix

SUMMARY

In 2012 our laboratory published in several different themes of evolutionary genomics. Overall, our laboratory is interested in the study of evolutionary biology in the broader sense. We study different organisms and mechanisms, on a fundamental level as well as focusing on observation-based issues of genomes. Last year our work focused on issues of phylogeny, impact of error rates on evolution and studying the impact of epistasis in protein evolution.

RESEARCH PROJECTS

1. Epistasis in evolution

One of the central ideas in evolutionary biology and genetics is that of an abstract representation of the entire collection of genetic material as a complex function of biological fitness. Such a genotype to phenotype representation is called a fitness landscape because its caricature representation in a two-dimensional space resembles a topographical map, or a landscape. In theory, the fitness landscape is an N-dimensional function where N stands for the number of independently-acting genetic elements in the genome, such as genes or regulatory elements. In practice, we still have a very poor idea about many different aspects of fitness landscapes representing genomes in nature.

Historically, fitness landscapes were modelled as a unidimensional function of the number of alleles (or mutations) that is carried by the genome. Under unidimensional fitness function the only parameter that is important to determine the state of the organism is the number of bad (or good, or a difference of the two) mutations. However, this implies that the effect of a single mutation is self-determined, meaning that other genetic factors do not affect how good or bad a specific mutation is. The contribution of mutations to fitness, however, many not be independent in nature as the effect of a mutation is often observed to be very different in a different genetic "background". This means that a mutation has a vastly different impact on fitness depending on what other mutations are present in the genome at the same time.

Representing interactions between mutations requires a multidimensional fitness function. However, it is impractical and may be impossible to picture and study N-dimensional representations of fitness landscapes. For that reason, researchers studying fitness landscapes often formulate questions regarding general aspects of its shape. The past year we published two such studies of the general aspects of fitness landscapes. One of them was focused on a very broad, large scale depiction of fitness landscape and the other on a fine scale.

2. Evolutionary irreversibility

One of the ways to study fitness landscapes is to focus on instances when the fitness impact of one substitution is different in different species. In this study we looked at instances when an ancestral amino acid state, which must have corresponded to a state of high fitness, is identical to a disease mutation in the modern day humans. Such instances represent cases of inherent evolutionary irreversibility because, at the present moment, reversal to a genetic state realized in our ancestors is impossible due to the drastic nature of the phenotypes caused by such mutations (Figure 1). We find that such instances are relatively common, with an estimated of 10-40% of all amino acid substitutions corresponding to genetically irreversible states. Thus, the fitness landscape of proteins involved in genetic pathologies is rugged on a local scale, with the fitness impact of individual substitutions changing across observable evolutionary time.

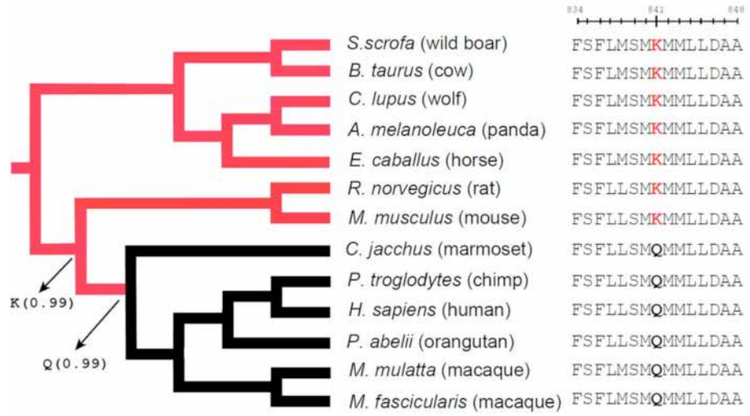


Figure 1.
An example of an irreversible state in the ABAC4 gene. The Q841K mutation causes Stargardt disease, an early onset muscle degenerative disorder leading to blindness, with the pathogenic K amino acid state corresponding to the ancestral state of primate species.

3. Epistasis in molecular evolution

In this study we looked at the factor of genetic non-independence, epistasis, in the course of protein evolution. We constructed large multiple alignments from Metazoa for 13 mitochondrially-encoded and 2 nuclear-encoded proteins and from Viriplantae for Rubisco, an enzyme encoded in the chloroplast. We then analyzed the average number of amino acids per site found in each alignment. After taking into account multiple confounding factors, such as quality of alignment and sequence polymorphism, we obtained that on average ~8 different amino acids are found per site. In a world were the impact of an amino acid state is independent from all other factors this implies if an amino acid state is observed in one species it should also be benign in another. Thus, if a horse has R in a site and an elephant a G in the orthologous site then the R <-> G substitutions should be uninhibited in the short term evolution. Thus, the non-epistatic evolutionary rate should be roughly 7/19, as on average one amino acid should be free to evolve into the 7 other amino acids present in the genome out of the 19 total possible substitutions. In contrast to this prediction, we find that the observed rate of short-term evolution is almost one order of magnitude slower than predicted based on average amino acid usage across sites. Excluding environmental adaptation as a possible explanation we conclude that epistasis, or genetic interactions, are a factor in evolution of a majority of amino acid substitutions in molecular evolution.

PUBLICATIONS

Soylemez O, Kondrashov FA.
"Estimating the rate of irreversibility in protein evolution."
Genome Biol Evol, 4(12):1213-22 (2012).

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, 29(2):531-44 (2012). (*)

Derelle R, Lang BF.
"Rooting the eukaryotic tree with mitochondrial and bacterial proteins."
Mol Biol Evol, 29(4):1277-89 (2012).

Povolotskaya IS, Kondrashov FA, Ledda A, Vlasov PK.
"Stop codons in bacteria are not selectively equivalent."
Biol Direct. 2012 Sep 13;7:30. doi: 10.1186/1745-6150-7-30.

Breen MS, Kemena C, Vlasov PK, Notredame C, Kondrashov FA.
"Epistasis as the primary factor in molecular evolution."
Nature, 490(7421):535-8 (2012).

Warnecke T.
"Loss of the DnaK-DnaJ-GrpE chaperone system among the Aquificales."
Mol Biol Evol, 29(11):3485-95 (2012).

Kondrashov FA.
"Gene duplication as a mechanism of genomic adaptation to a changing environment."
Proc Biol Sci, 279(1749):5048-57 (2012).

Warnecke T, Supek F, Lehner B.
"Nucleoid-associated proteins affect mutation dynamics in E. coli in a growth phase-specific manner."
PLoS Comput Biol, 8(12):e1002846 (2012).

(*) 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

BIOINFORMATICS AND GENOMICS

Group: Gene Function and Evolution

Group Structure:

Group Leader: Gian Gaetano Tartaglia

Postdoctoral Fellows: Benedetta Bolognesi, Andreas Zanzoni (Sept-Dec 2012)

Technicians: Silvia Rodriguez, Maria de las Nieves Lorenzo

PhD Students: Federico Agostini, Davide Cirillo, Petr Klus and Domenica Marchese

SUMMARY

We are interested in the regulation of proteins that have high propensity to misfold. Indeed, the presence of misfolded or incompletely folded proteins, which almost lack functional activity, not only represents an energetic drain on the cell but result in the accumulation of aggregates that range from inclusion bodies in bacteria to amyloid fibrils in mammals[1]. We study protein chaperones[2], which protect against formation of toxic aggregates, and identify novel protein-RNA interactions that could be involved in regulating protein expression[3]. Indeed, long non-coding RNA have emerged as key players involved in the control of transcriptional and post-transcriptional gene regulatory pathway[4]. Only a limited number of functional lncRNAs has been identified so far, but great regulatory potential is envisaged[5].

RESEARCH PROJECTS

1. New developments of the catRAPID method

catRAPID exploits physicochemical properties of nucleotide and amino acid chains such as secondary structure, hydrogen bonding and van der Waals' propensities to predict protein–RNA associations with a confidence of 78% or higher[6]. In the original implementation of the method, we could calculate protein interactions with transcripts < 3 kb. In order to investigate Xist, which is 16–19 kb long and represents the largest non-coding transcript with known function, we developed an extension of the algorithm[4]. In addition to the fine calculation of protein–RNA interactions (interaction propensity), we can estimate the specificity of associations (interaction strength) and identify binding regions in transcripts (interaction fragments)[3]. These new developments are introduced to facilitate the characterization of protein interactions with long non-coding RNA and guide future experimental design. A new version of our web servers was released at http://tartaglialab.crg.cat/.

2. X-Inactivation

To date, the precise mechanisms underlying localization of the long noncoding RNA Xist on the X-chromosome remain poorly understood[4].

We used our theoretical framework, catRAPID, to investigate Xist interactions with a number of epigenetic modifiers as well as transcription and splicing factors including SUZ12, EZH2, YY1, SAF-A, SFRS1 and SATB Our calculations suggest that localization and confinement of Xist are finely regulated by multiple factors acting at the interface between chromosome X and the nuclear matrix. Our results are compatible with a model in which following X-chromosome docking mediated by YY1, matrix-associated proteins SAF-A and SATB1 recruit the 5′-half of Xist and drive the translocation in cis of the Xist–PRC2 complex. We also applied our method to SFRS1's interactome, showing that catRAPID predicts CLIP-seq-binding sites with great accuracy[4].

3. Neurodegenerative diseases

Although neurodegenerative diseases are traditionally described as protein disorders leading to amyloidosis, very recent evidence indicates that protein–RNA associations are involved in a number of neuropathies[7]. In Huntington's disease, ataxias, and myotonic dystrophy, primary transcripts containing expanded trinucleotide regions form intranuclear *foci* where proteins are sequestered and inactivated. In Creutzfeldt-Jakob disease, the RNA ability to interact with proteins facilitates conversion of the α-helix-rich prion protein (PrP^{C}) into its infectious β-structure-rich insoluble conformer (PrP^{Sc}). In Alzheimer's disease (AD), translation of aggregation-prone proteins is regulated by iron-dependent ribonucleoprotein interactions. In many cases, aggregation of RNA-binding proteins is triggered by defective transcription and affects processing of RNA molecules, which leads to progressive cell death.

We studied the metabolic signature associated with Fragile X Syndrome and analyzed FMRP associations with *SOD1* and *APP*, highlighting possible links with Amyotrophic Lateral Sclerosis and AD (Figure 1)[3]. We investigated the X-chromosome disorder FXTAS, which is caused by CGG expansions in the *FMR-1* untranslated region and characterized key players involved in protein sequestration. We predicted TDP-43 interactions with several ncRNAs displaying changes of expression levels upon TDP-43 and identified a set of vault-associated and natural antisense transcripts that could be linked to clinical manifestations of TDP-43 proteinopathy. We also studied the ability of FMRP and TDP-43 to regulate their own expression levels through autogenous interactions, characterizing their binding sites in great detail. We analyzed the interaction between IRP-1 and Alzheimer's transcript *APP* and investigated the interaction between IRP-1 and an IRE-like region of α-synuclein mRNA, which represents a link to the iron-pathway deregulation associated with Parkinson's disease. Finally, we investigated the ability of RNA aptamers to bind to aggregation-prone regions of prions, which shows that approach could be useful for screening of RNA-based therapeuticals[3].

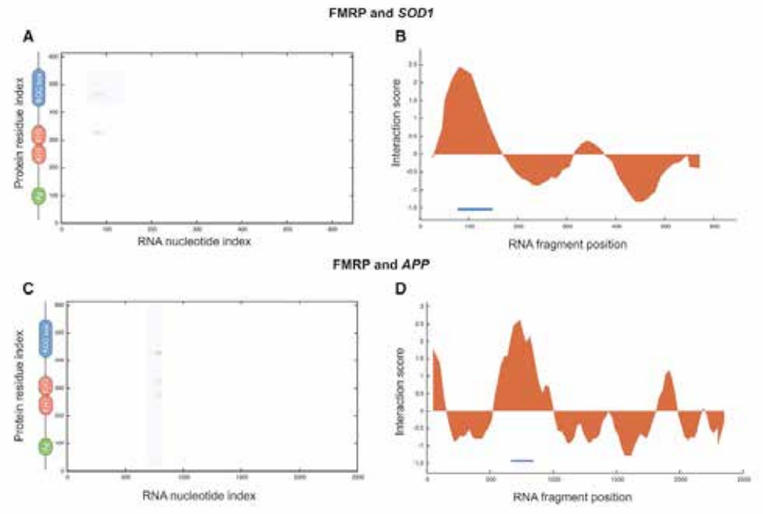


Figure 1.
Fragile Mental Retardation Protein FMRP interacts with Superoxide Dismutase SOD1 and Amyloid Precursor Protein APP transcripts. *Interaction maps of FMRP binding to: a) SOD1 and c) APP (FMRP secondary structure elements are displayed next to the "protein residue index" axis; blue areas indicate experimentally validated interactions). RNA interaction profiles for FMRP interactions with: b) SOD1 and d) APP (blue lines indicate experimentally identified binding regions). Our findings suggest that Fragile Mental Retardation is linked with Alzheimer's disease and Amyotrophic Lateral Sclerosis[3].*

4. Chaperone networks

Cellular chaperone networks prevent potentially toxic protein aggregation and ensure proteome integrity. We used *Escherichia coli* as a model to understand the organization of these networks, focusing on the cooperation of the DnaK system with the upstream chaperone Trigger Factor and the downstream GroEL[2].

DnaK interacting proteins cover a wide range of cellular functions, prominently including DNA replication, recombination and repair and cell division and chromosome partitioning DnaK interactors have a also significant preference for proteins involved in translation, ribosomal structure, and biogenesis and include ribosomal proteins. To determine whether the proteins enriched on DnaK have a high propensity to aggregate, we took advantage of a recent study on solubility of *E. coli* proteins upon *in vitro* translation in the absence of chaperones[1]. Indeed, DnaK biding proteins are more aggregation prone upon translation than average proteins of soluble cell lysate. This is consistent with the finding that the enriched proteins frequently display pI values close to neutral pH Moreover, these proteins are predicted to bury amino acid residues less effectively from solvent than the less enriched DnaK interactors and average soluble proteins[2] (Figure 2).

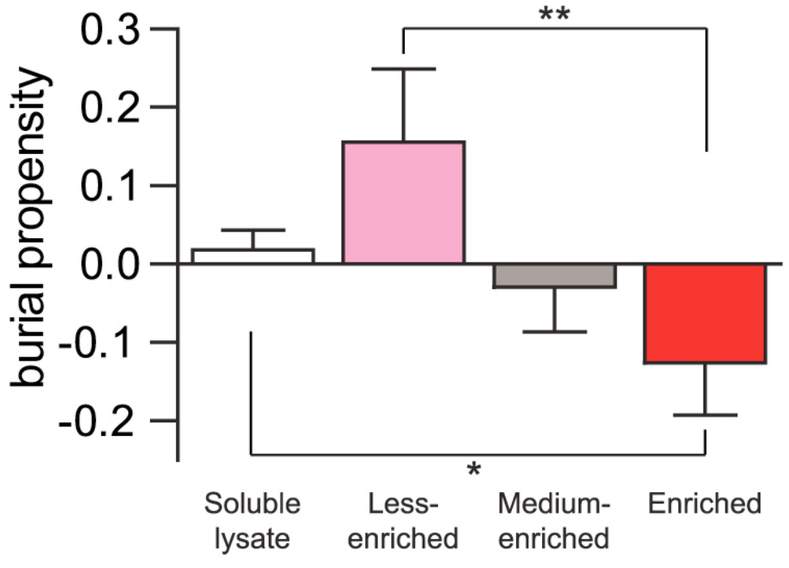


Figure 2.

DnaK substrates. *DnaK interactors (i.e., "medium" and "less" enriched substrates) do no bury hydrophobic regions. The lack of burial correlates with a high propensity to populate dynamic intermediates exposing hydrophobic residues. DnaK help these proteins to fold into their native states[2].*

1. Agostini F, Vendruscolo M, Tartaglia GG. *"Sequence-based prediction of protein solubility."* J Mol Biol, 421:237–241 (2012).
2. Calloni G *et al*. *"DnaK functions as a central hub in the E. coli chaperone network."* Cell Rep, 1:251–264 (2012).
3. Cirillo D *et al*. *"Neurodegenerative diseases: Quantitative predictions of protein-RNA interactions."* RNA, 19:129–140 (2013).
4. Agostini F, Cirillo D, Bolognesi B, Tartaglia GG. *"X-inactivation: quantitative predictions of protein interactions in the Xist network."* Nucleic Acids Res, doi:10.1093/nar/gks968 (2012).
5. Cirillo D, Agostini F, Tartaglia GG. *"Predictions of protein–RNA interactions."* Wiley Interdisciplinary Reviews: Computational Molecular Science n/a–n/a (2012). doi:10.1002/wcms.1119
6. Bellucci M, Agostini F, Masin M, Tartaglia GG. *"Predicting protein associations with long noncoding RNAs."* Nat. Methods, 8:444–445 (2011).
7. Johnson R, Noble W, Tartaglia GG and Buckley NJ. *"Neurodegeneration as an RNA disorder."* Prog. Neurobiol, 99:293–315 (2012).

PUBLICATIONS

Cirillo D, Agostini F, Klus P, Marchese D, Rodriguez S, Bolognesi B, Tartaglia GG.
"Neurodegenerative diseases: Quantitative predictions of protein-RNA interactions." RNA, doi: 10.1261/rna.034777.112. Epub 2012 Dec 21.

Agostini F, Cirillo D, Bolognesi B, Tartaglia GG.
"X-inactivation: quantitative predictions of protein interactions in the Xist network."
Nucleic Acids Res, doi: 10.1093/nar/gks968. Epub 2012 Oct 22.

Johnson R, Noble W, Tartaglia GG, Buckley NJ.
"Neurodegeneration as an RNA disorder."
Prog Neurobiol, 99(3):293-315 (2012).

Roodveldt C, Andersson A, De Genst EJ, Labrador-Garrido A, Buell AK, Dobson CM, Tartaglia GG, Vendruscolo M.
"A rationally designed six-residue swap generates comparability in the aggregation behavior of α-synuclein and β-synuclein."
Biochemistry, 51(44):8771-8 (2012).

Calloni G, Chen T, Schermann SM, Chang HC, Genevaux P, Agostini F, Tartaglia GG, Hayer-Hartl M, Hartl FU.
"DnaK functions as a central hub in the E. coli chaperone network."
Cell Rep., 1(3):251-64 (2012).

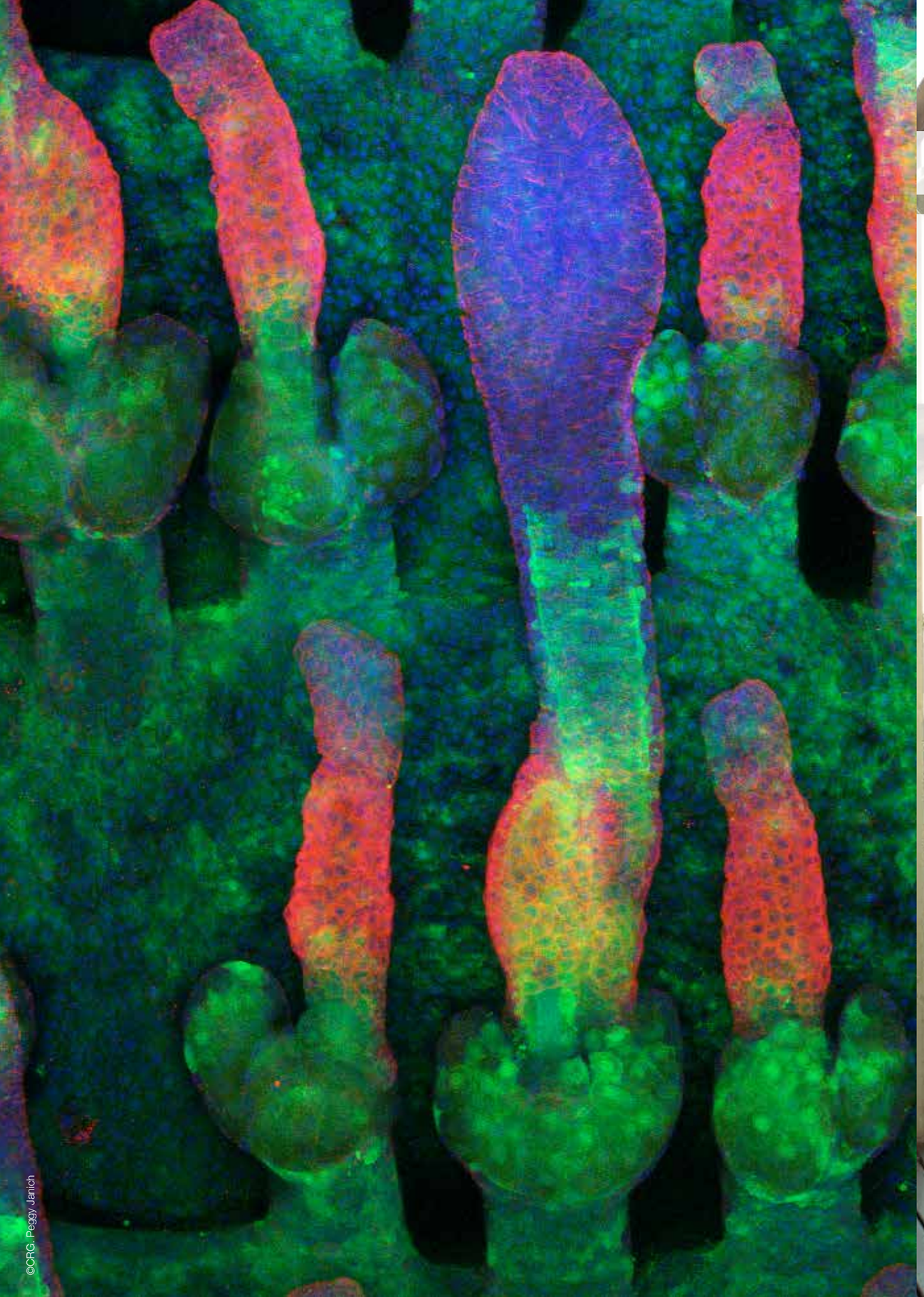
Speretta E, Jahn TR, Tartaglia GG, Favrin G, Barros TP, Imarisio S, Lomas DA, Luheshi LM, Crowther DC, Dobson CM.
"Expression in drosophila of tandem amyloid β peptides provides insights into links between aggregation and neurotoxicity."
J Biol Chem, 287(24):20748-54 (2012). (*)

Agostini F, Vendruscolo M, Tartaglia GG.
"Sequence-based prediction of protein solubility."
J Mol Biol, 421(2-3):237-41 (2012).

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 (2012). (*)

Iglesias-Platas I, Martin-Trujillo A, Cirillo D, Court F, Guillaumet-Adkins A, Camprubi C, Bourc'his D, Hata K, Feil R, Tartaglia G, Arnaud P, Monk D.
"Characterization of novel paternal ncRNAs at the Plagl1 locus, including Hymai, predicted to interact with regulators of active chromatin."
PLoS One, 7(6):e38907 (2012).

(*) 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

Group Structure:

Group Leader: Vivek Malhotra

Scientific Staff: Sandra Mitrovic

Postdoctoral Fellows: Felix Campelo, David Cruz, Amy Curwin, Juan Duran, Patrik Erlmann, Cristina Nogueira, Josse van Galen, Julien Villeneuve, Yuichi Wakana (until March 2012), Thomas Bohnacker (until February 2012)

PhD Students: Caroline Bruns, Ana Catarina Vidinhas De Olivera (until August 2012)

Technicians: Anne-Marie Alleaume (until May 2012), 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 regulates the association of CERT and OSBP by phosphorylation. Altogether this suggests that PKD regulates the levels of PI4P, ceramide and sterols at the TGN, and important for transport carrier biogenesis. We have tested this hypothesis directly by affecting the levels of the ceramide product Sphingomyelin at the TGN. Perturbing the balance of Sphingomyelin inhibited transport carrier formation at the Golgi membranes without affecting the fusion of incoming carriers (Duran *et al.*, 2012). These findings highlight the role of lipid homeostasis in transport carrier formation. 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 (CARTS) that were analyzed by mass spectrometry, which led to the characterization of a number of new proteins (Wakana *et al., JCB* 2012). 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.

2. New transport components

Amy Curwin, Julien Villeneuve, Patrik Erlmann, Maria Ortega, David Cruz, Felix Campelo

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; Curwin *et al.*, 2012). The cofilin mediated cargo sorting involves a transmembrane Ca^{2+} ATPase (von Blume *et al.*, 2011) and a soluble-lumenal protein of the Golgi membranes called Cab45 (von Blume *et al.*, 2012).

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.

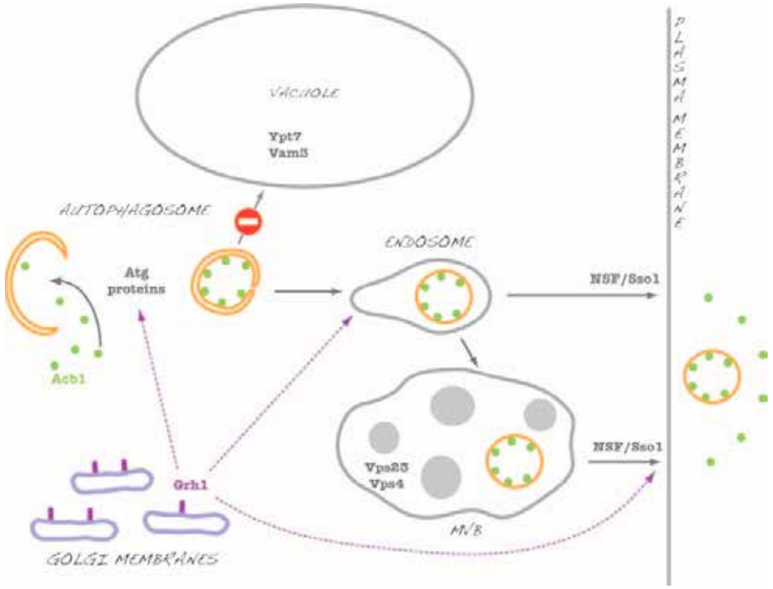


Figure 1.
Cofilin dependent actin depolymerization at the TGN is required for SPCA1 mediated Ca^{2+} pumping into the lumen TGN, regulating the sorting of a subset of secretory cargo.

3. Unconventional protein 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). Our new findings have revealed of a new compartment called CUPS that assembles near the ER and required for the unconventional secretion of soluble cytoplasmic proteins (Bruns *et al., JCB* 2011). We are using yeast and mammalian cells to reveal the mechanism of unconventional secretion.

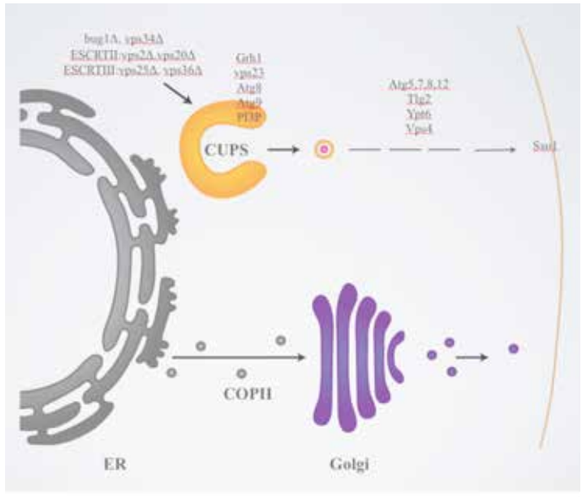


Figure 2.
The unconventional secretion pathway. The COPII mediates export of conventional secretory cargo from the ER to the Golgi membranes is shown in the bottom half of the diagram. The export of cargo from the Golgi apparatus is mediated by a number of different classes of vesicles including CARTS. The unconventional protein secretion is independent of ER and the Golgi apparatus. A compartment called CUPS containing Grh1, Atg8, Atg9 and PI3P assembles near the ER-exit site upon Glucose starvation in yeast. We suggest that this compartment collects secretory cargo such as Acb1, which is then packed into a double membrane composed autophagosome like vesicle (secretory autophagosome). The secretory autophagosomes then deliver the cargo by a process that requires a number of gene products including the cell surface t-SNARE Sso1 to the extracellular space. The biogenesis of CUPS requires a number of components including bug1, Vps34, Vps2, Vps20, Vps25 and Vps36. A genome wide screen is in progress is in progress to identify all the components required for the biogenesis of CUPS.

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. We have found that a kinase called Myt1 is the substrate of MEK1 dependent Golgi apparatus fragmentation (Villenueve *et al., EMBO J.* 2012) and required for controlling the timing of mitotic entry. A combination of *in vitro* approaches and system wide siRNA is being used to further dissect the mechanism of Golgi complex specific cell cycle checkpoint.

PUBLICATIONS

Articles

von Blume J, Alleaume AM, Kienzle C, Carreras-Sureda A, Valverde M, Malhotra V.
"Cab45 is required for Ca2+-dependent secretory cargo sorting at the trans-Golgi network."
Journal of Cell Biology,199(7):1057-66 (2012).

Villeneuve J, Scarpa M, Ortega-Bellido M, Malhotra V.
"MEK1 inactivates Myt1 to regulate Golgi membrane fragmentation and mitotic entry in mammalian cells."
EMBO Journal, doi: 10.1038/emboj.2012.329. Epub 2012 Dec 14.

Duran JM, Campelo F, van Galen J, Sachsenheimer T, Sot J, Egorov MV, Rentero C, Enrich C, Polishchuk RS, Goñi FM, Brügger B, Wieland F, Malhotra V.
"Sphingomyelin organization is required for vesicle biogenesis at the Golgi complex."
EMBO Journal, 31(24):4535-46 (2012).

Wakana Y, van Galen J, Meissner F, Scarpa M, Polishchuk RS, Mann M, Malhotra V.
"A new class of carriers that transport selective cargo from the trans Golgi network to the cell surface."
EMBO Journal, 31(20):3976-90 (2012).

Venditti R, Scanu T, Santoro M, Di Tullio G, Spaar A, Gaibisso R, Beznoussenko GV, Mironov AA, Mironov A Jr, Zelante L, Piemontese MRNotarangelo A, Malhotra V, Vertel BM, Wilson C, De Matteis MA.
"Sedlin controls the ER export of procollagen by regulating the Sar1 cycle."
Science, 337(6102):1668-72 (2012).

Curwin AJ, von Blume J, Malhotra V.
"Cofilin mediated sorting and export of specific cargo from the Golgi apparatus in yeast."
Molecular Biology of the Cell, 23(12):2327-38 (2012).

Malhotra V.
"COPII Vesicles Get Supersized by Ubiquitin."
Cell, 149(1):20-1 (2012).

Rabouille C, Malhotra V, Nickel W.
"Diversity in unconventional protein secretion."
J Cell Sci, 125(Pt 22):5251-5 (2012).

Campelo F, Cruz A, Pérez-Gil J, Vázquez L, Hernández-Machado A.
"Phase-field model for the morphology of monolayer lipid domains."
Eur Phys J E Soft Matter, 35(6):49 (2012).

Reviews

Campelo F, Malhotra V.
"Membrane fission: the biogenesis of transport carriers."
Annual Review of Biochemistry, 81:407-27 (2012).

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

Staff Scientist: Sylvain Meunier

Postdoctoral Fellows: Roser Pinyol (until September 2012), Marti Badal (until May 2012), Nathalie Brouwers, Mónica Rodríguez (since February 2012), Maria Ana Gomez-Ferreria (since October 2012)

Students: Antonios Lioutas, Jacopo Scrofani, Tommaso Cavazza

Trainees: Rebecca Adikes (until June 2012), Eduardo Goicoechea Serrano (July-August 2012) Artur Ezquerra (since November 2012)

Technicians: Nuria Mallol

SUMMARY

Research in my lab is directed at understanding the role microtubules in cell organization and function. Our current focus is cell division. To get more insights into this process we aim at identifying novel players in spindle assembly and studying their function and regulation during mitosis. 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. The combination of experiments in this unique experimental system (Karsenti and Vernos, 2001), in tissue culture cells and *in vitro* with purified components offers a powerful approach to get at the mechanism driving spindle assembly and function.

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, the small GTPase Ran in its GTP bound form plays an important role in the spatial regulation of spindle assembly (Gruss and Vernos, 2004).

This year we have obtained novel insights on the mechanism controlling microtubule nucleation during M-phase through the phosphorylation the γ-tubulin-ring complex (γ-TURC) adaptor protein NEDD1 by two mitotic kinases, the NIMA family Nek9 and Aurora A.

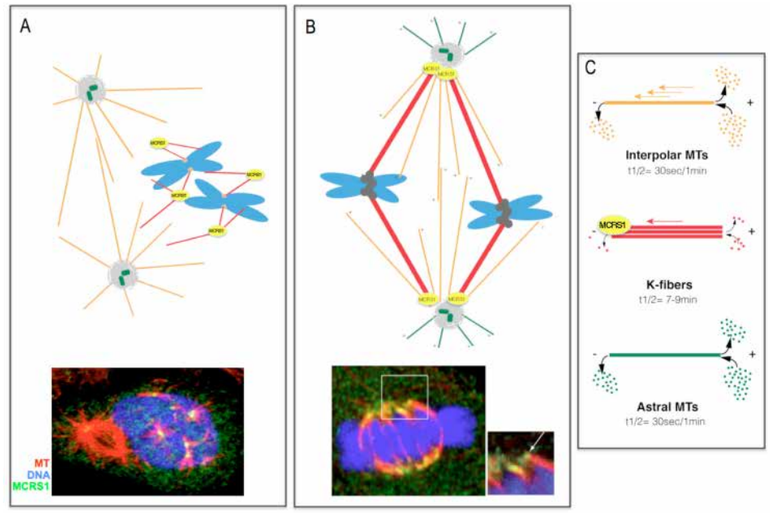


Figure 1

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. The γ-TURC is essential for driving microtubule nucleation. This year we found that NEDD1, a γ-TURC associated protein, is a phosphorylation target for two mitotic kinases: the NIMA family kinase Nek9 and Aurora A. Phosphorylation of NEDD1 by Nek9 at a single site drives the recruitment of the γ-TURC to the centrosome whereas its phosphorylation by Aurora A on a different site is essential for RanGTP acentrosomal microtubule nucleation (Sdelci and Schutz et al. Current biology 2012; Pinyol et al, Current Biology 2013)
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-fibers (B) (MTs are shown in red, DNA in blue). For more details see Meunier and Vernos (2012).

1. The role of the NIMA family kinase Nek9 during mitosis

Martin Schutz, Roser Pinyol

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-TURC and as a result the MT nucleation activity of the centrosome increases. In collaboration with the group of Dr Joan Roig (IRB, Barcelona) we have shown that Nek9 phosphorylates NEDD1 at a single residue and thereby controls the targeting of the microtubule nucleation machinery to the centrosome in mitotic cells.
(Sdelci and Schutz *et al*., 2012)

2. The RanGTP pathway in MT nucleation and stabilization in M-phase

Roser Pinyol, Sylvain Meunier, Jacopo Scroffani, Tommaso Cavazza

After nuclear envelope breakdown, a centrosome independent pathway relying on a RanGTP 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 want to understand the mechanism driving RanGTP dependent microtubule nucleation. We found that the γ-TURC adaptor protein NEDD1 interacts with Aurora A and becomes phosphorylated at a single residue. Through a combination of approaches in Hela cells and *Xenopus* egg extract we showed that phosphorylation at this site is essential and specific for microtubule assembly induced by RanGTP and the mitotic chromosomes.
(Pinyol *et al*., 2012)

3. The function of the mitotic kinase Aurora A during anaphase and telophase

Antonios Lioutas

The mitotic Aurora A kinase has been extensively characterized by different groups including ours (Peset I. *et al.*, 2005; Sardon *et al.*, 2008; 2009; 2010) more specifically during mitotic entry, centrosome maturation and separation and bipolar spindle assembly. Although Aurora A is degraded after mitosis has been completed, very little is known about its possible role(s) after metaphase. We are examining this question in human culture cells using inhibitors for the Aurora kinases (manuscript in preparation).

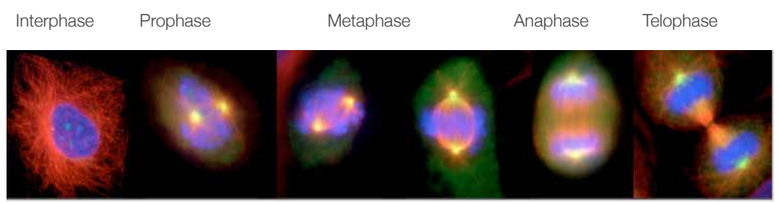


Figure 2
Immunufluorescence analysis of HeLa cells shows that the kinase Aurora A is associated to the centrosome and the spindle throughout mitosis.
MTs are labelled in red, DNA in blue and Aurora A in green.

4. The role of molecular motors in spindle assembly and chromosome movements

Nathalie Brouwers, Rebekka Adikes

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

Sdelci S, Schütz M, Pinyol R, Bertran MT, Regué L, Caelles C, Vernos I and Roig J.
"Nek9 Phosphorylation of NEDD1/GCP-WD Contributes to Plk1 Control of γ-Tubulin Recruitment to the Mitotic Centrosome"
Current Biology, 22(16):1516-23 (2012).

Meunier S and Vernos I.
"Microtubule assembly during mitosis - from distinct origins to distinct functions?"
Journal of Cell Science, 125(Pt 12):2805-14 (2012).

Pinyol R, Scrofani J and Vernos I.
"The Role of NEDD1 Phosphorylation by Aurora A in Chromosomal Microtubule Nucleation and Spindle Function."
Curr Biol, doi: 10.1016/j.cub.2012.11.046. Epub 2012 Dec 27.

CELL AND DEVELOPMENTAL BIOLOGY

Group: Coordination of Cytokinesis with Chromosome Segregation

Group Structure:

Group Leader: Manuel Mendoza

Postdoctoral Fellows: Alexandre Vendrell (until August 2012), Petra Stockinger (shared with J. Solon lab)
Arun Kumar (since August 2012)

PhD Students: Aina Masgrau, Iris Titos, Nuno Amaral, Francesca DiGiovanni (since May 2012)
Michael Maier (since September 2012), Gabriel Neurohr (until April 2012)

Technician: Trinidad Sanmartin

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. How cells respond to errors in chromosome segregation: the NoCut checkpoint

Nuno Amaral and Arun Kumar

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 is proving instrumental in dissecting the molecular pathways coordinating chromosome segregation and cell division.

2. The role of Boi1 and Boi2 during abscission and polarized growth

Aina Masgrau and Michael Maier

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.

3. Regulation of chromosome condensation

Iris Titos and Francesca Di Giovanni

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 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 adjusting their level of condensation during anaphase. 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). We are currently addressing the mechanism of chromosome hyper-condensation with 3C-based approaches.

4. 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.

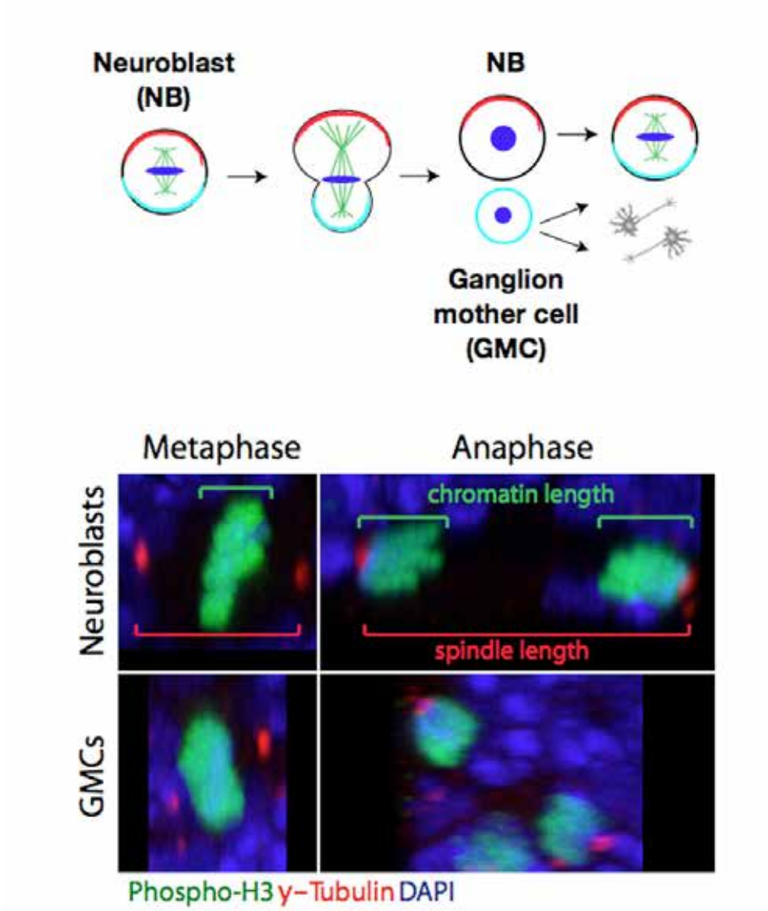


Fig. 1.
Characterization of chromosome compaction and spindle length, in neuroblasts and GMCs in Drosophila larvae. Brains were stained with antibodies to visualize mitotic chromosomes (phospho-histone H3), spindle poles (gamma tubulin) and DNA (DAPI). Image by Petra Stockinger.

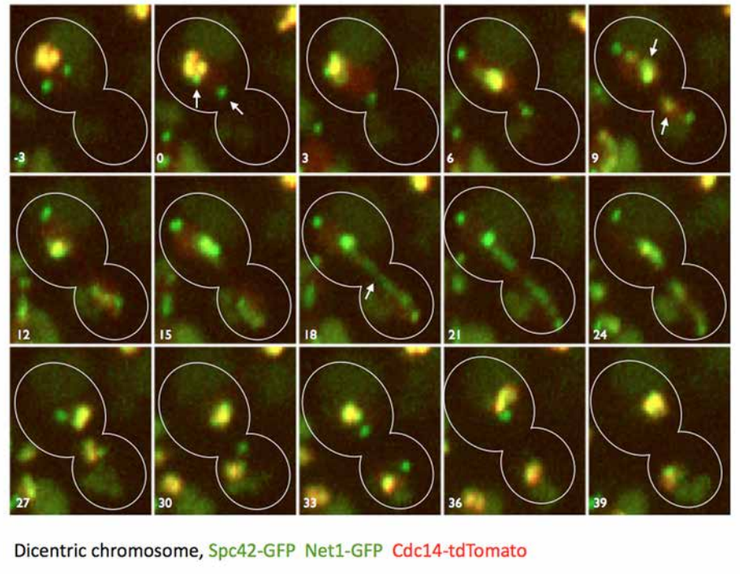


Fig. 2.
Segregation of a dicentric chromosome in a budding yeast cell, imaged thanks to Net1-GFP labeling of the rDNA locus and Spc42-GFP labeling of spindle poles. Cell cycle progression is visualized thanks to Cdc14-tomato, a protein that is released from the nucleolus during anaphase. Image by Nuno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

CELL AND DEVELOPMENTAL BIOLOGY

Group: Biomechanics of Morphogenesis

Group Structure:

Group Leader: Jérôme Solon

Postdoctoral Fellows: Laure Saias (until December 2012), Kai Dierkes, Petra Stockinger (shared with M. Mendoza lab)
Philippe-Alexandre Pouille (until January 31/01/2012)

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.

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). Interestingly, 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.

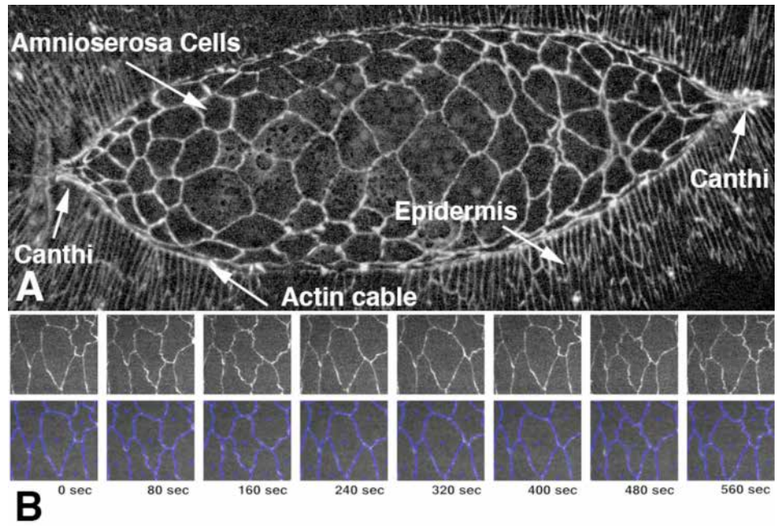


Fig 1
***(A) Armadillo-GFP highlighting cell membranes during Drosophila dorsal closure**. The epidermal tissues converge from the lateral part (top and bottom of the picture) to the dorsal part (in the center of the picture) of the embryo. The dorsal part is covered with amnioserosa cells, which provide 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).*

Understanding the mechanism regulating cellular contraction
We are interested in investigating the molecular regulation of the amnioserosa contractions, and what molecular mechanisms translate these contractions into tissue movement.

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 (Fig 1 B). Recent experimental observations performed in the lab indicate that the dynamical properties of the acto-myosin cortex per se are at the origin of the emergence of the cellular pulses. In collaboration with Guillaume Salbreux (Max Planck Institute Physik Komplex System, Dresden), We are developing a biophysical model of AS tissue contraction that we are challenging experimentally.

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 (Fig 2-A). We are combining this technique with regular confocal imaging to extract *in vivo* the 3D dynamics of the amnioserosa contractions (Fig 2-B).

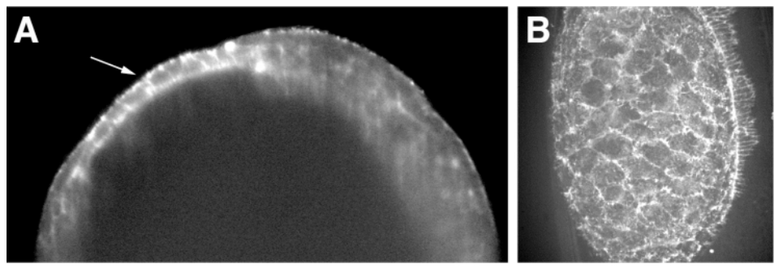


Fig 2.
Three-dimensional imaging of DC. (A) *Live SPIM imaging of a Drosophila embryo in the dorsal ventral axis. The arrow indicates the apical surface the amnioserosa tissue covering the dorsal part of the embryo.* ***(B)*** *Live imaging of the apical surface of the amnioserosa tissue covering the 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. The observed decrease in volume regulates the process and is the consequence of apoptotic activation. This unusual mechanism allows the progression of closure without remodelling of the tissue covering the gap and of underlying biological structures. We want now to understand the mechanisms at the origin of apoptotis activation and epidermis progression.

Tissue fusion by zipping and actin cable formation
Completion of closure requires the additional formation of an actin cable surrounding the dorsal gap and the zipping of the two epithelial layers at the canthi of the opening. We are interested in understanding the regulation of both actin cable formation and zipping progression and also the contribution of these two mechanisms to the process.

2. 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.

3. Methods development: *In vivo* force induction and cellular mechanical properties measurement

Additionally, we are developing experimental methods to extract the biomechanical properties of the cell and tissues in living organisms. Applying mechanical forces on tissues and measuring the responding forces exerted by individual cells remains challenging.

We have developed an *in vivo* force induction system based on the induction of a magnetic force on a magnetic particle injected in an individual cell of the *Drosophila* embryo. We are able to exert sufficient forces to deform cells and tissues. Therefore, we can directly measure mechanical properties of cells and their variation over time in a developing organism.

4. Mechanics of chromosome condensation during *Drosophila* development

We are developing a collaborative project between the lab of Manuel Mendoza and our lab. Following Manuel's work in yeast, we have found that the amplitude of chromosome condensation correlates with cell and spindle size in different *Drosophila* cell types. This scaling suggests interesting feedback mechanisms between chromosome length and spindle/cell size to insure proper cell division. We are interested in revealing these mechanisms by using an interdisciplinary approach.

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

Colombelli J, Solon J.
"Force communication in multicellular tissues addressed by laser nanosurgery."
Cell Tissue Res, doi: 10.1007/s00441-012-1445-1. Epub 2012 May 24.

CELL AND DEVELOPMENTAL BIOLOGY

Group: Organelle Biogenesis and Homeostasis

Group Structure:

Group Leader: Pedro Carvalho

Postdoctoral Fellows: Ombretta Foresti, Cristina Miró (since September 2012)

PhD Students: Alexandra Grippa, Annamaria Ruggiano, Lisa Johnsen (since September 2012)
Victoria Rodríguez (since December 2012)

Technicians: Josep Pareja (until September 2012), Laura Buxó (since May 2012)
Robert Oliete (since October 2012)

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.

Although misfolded proteins are the primary substrates of ERAD, some folded, active proteins can also be targeted by ERAD-dependent degradation upon a certain stimulus. Therefore in addition to quality control, the ERAD pathway is also well positioned to modulate other cellular processes. The best known example of regulated ERAD involves the feedback degradation of HMG-CoA reductase, a key enzyme in the mevalonate pathway required for sterol biosynthesis, by the ER-associated degradation (or ERAD) ubiquitin ligase Hrd1/Gp78.

We have recently developed a quantitative proteomics approach to identify ERAD substrates. As part of this effort we have uncovered a novel posttranslational feedback system that involves the degradation of the squalene monooxygenase Erg1 by Doa10, another ERAD ubiquitin ligase, and that is essential for sterol homeostasis. This finding reveals a surprisingly intimate connection between sterol regulation and the ERAD pathway that can potentially be exploited for the therapy of hypercholesterolemia.

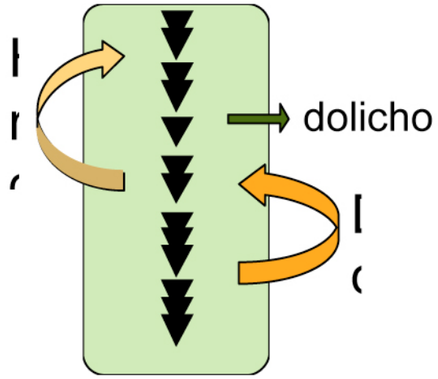


Fig1
ERAD-dependent negative feedback loops controlling sterol biosynthesis.

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.

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.

a- **Characterization of ER protein complexes required for LD formation**

Our attention has been focused on the role of a highly conserved protein called Seipin. Mutations in the yeast Seipin Fld1 lead to a very strong lipid droplet phenotype while germline mutations of Seipin in humans are the leading cause of congenital generalized lipodystrophy. We have identified novel binding partners of Fld1 and demonstrated that these act together in a complex during lipid droplet formation. More importantly, we found that deletion of Fld1 or its binding partners lead to defects of nuclear envelope organization primarily due to aberrant accumulation of phasphatidic acid (PA). PA is the main precursor for the synthesis of all major cellular phospholipids

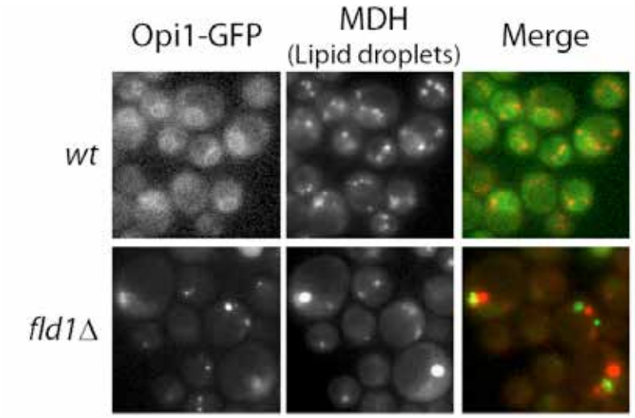
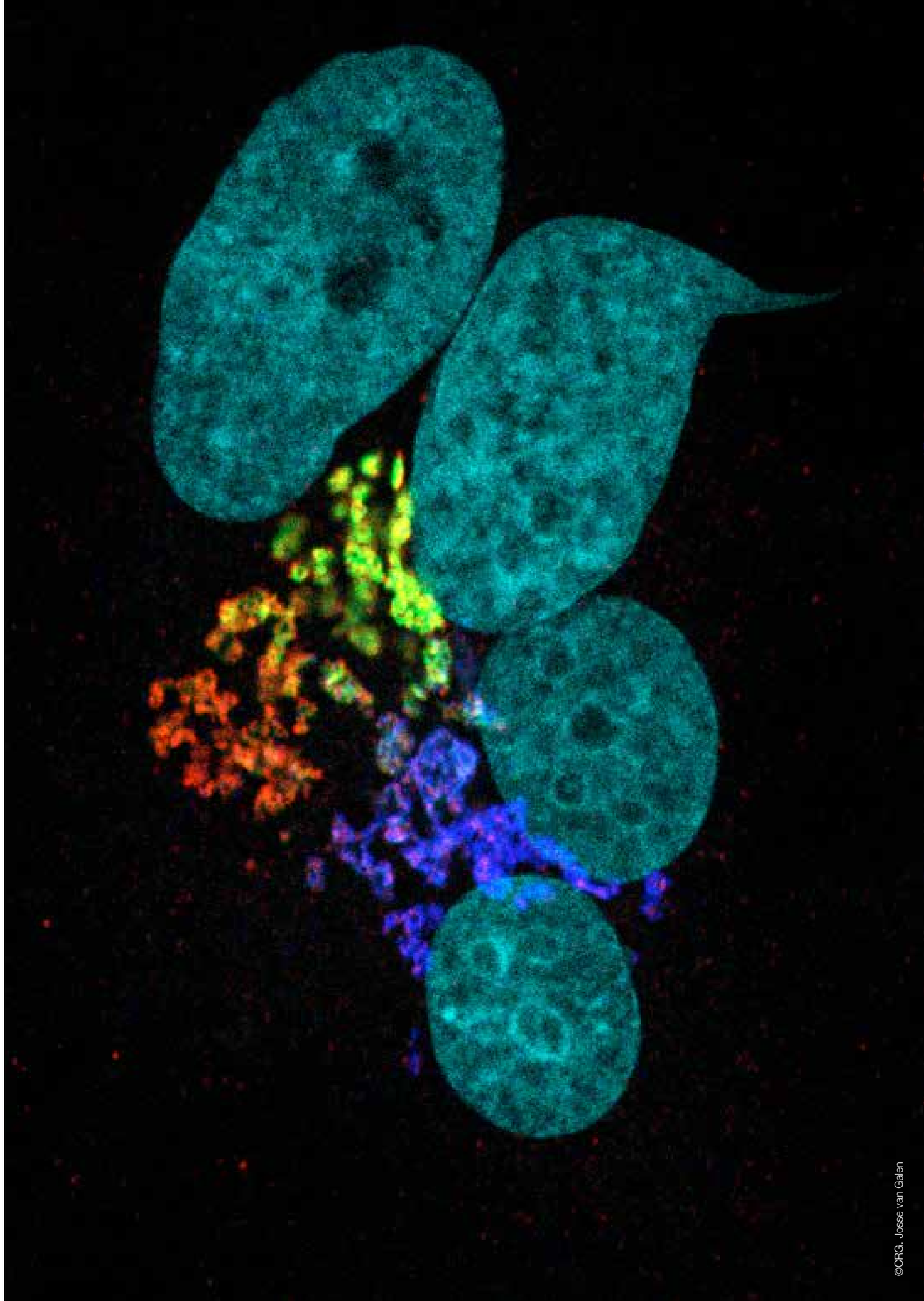


Figure 2

Localization of the transcriptional repressor Opi1-GFP. In wt cells, Opi1-GFP expressed from endogenous Opi1 localizes to the nuclear envelope. In fld1Δ mutant, Opi1-GFP concentrates as foci in the nuclear envelope in the proximity of LDs (stained by MDH).

as well as triglycerides, a major neutral lipid in LDs. In addition, in yeast cells, the levels of PA at the nuclear envelope are critical in controlling lipid biosynthesis: high PA levels recruit the transcriptional repressor Opi1 and lipid biosynthesis is turned on; under conditions of low PA, Opi1 is released and travels to the nucleus where it shuts down phospholipid synthesis. Whether an analogous feedback system involving PA is also present in mammals is unclear. We are currently investigating the mechanisms by which Fld1 regulates the levels of PA at the nuclear envelope. In parallel we are analysing the distribution of PA and certain PA-binding proteins in fibroblasts derived from lipodystrophic patients with Seipin mutations. We hope these studies to unravel important insight into the mechanisms of LD formation and lipid homeostasis under normal conditions and also in disease situations.



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: Neus Martínez, James Cotterell, Andreea Munteanu, Marco Musy

Technicians: Laura Quintana, Martina Niksic, Alexandre Robert, Lucia Russo

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

SUMMARY

The Sharpe lab has 2 primary goals:

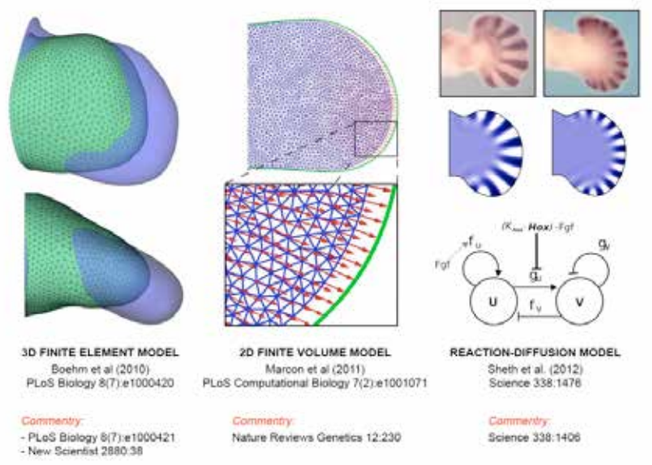
(1) To further our understanding of *organogenesis as a complex system*, by bringing together a diverse range of techniques from biology, physics, imaging and computer science. Within this general theme, we focus on two aspects: (a) We are studying a well-characterised standard model of development – the vertebrate limb (using both mouse and chick). The goal is to combine experimental data (especially 3D data sets using optical projection tomography) into a computational framework, so that we can explore and test mechanistic hypotheses about how this example of organogenesis works. Using this approach we are studying both the physical morphogenesis (eg. Boehm *et al.,* 2010, Marcon *et al.*, 2011) and also the genetic patterning mechanisms (eg. Sheth *et al*., 2012). (b) In addition to this specific model system, we are also interested in the theoretical principles by which gene regulatory networks can create controlled spatial patterns in multicellular contexts (eg. Cotterell *et al*., 2010)..

(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. Computational modeling of limb bud development

We have created various computer simulations, which help to explore different aspects of limb bud development. We particularly focus on finite-element modeling, which allows us to represent the developing organ at the tissue level. Tissue movement maps (both in 2D and 3D) are now serving as the framework within which to simulate and explore the gene network designs responsible for controlling limb development. Over the last year we have modelled how self-organising reaction-diffusion networks can control formation of the skeletal pattern during limb bud development. In particular, we have explored how an accurate pattern may depend on a close integration of a Turing-type patterning mechanism with positional information provided by genes like Hox genes and Fgf signaling (Sheth *et al*., 2012, *Science* 338:1476).



The Sharpe lab is constructing various 2D and 3D computer models of limb development. These have allowed us to explore and dispute the proliferation gradient hypothesis (left), create a realistic map of tissue movements in 2D (centre), and explain how Hox genes affect a Turing-type mechanism to explain the basic patterning of digits (right).

2. 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 that 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.

3. 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 2012 we started our second Sinergia grant from the Swiss National Science Foundation). During the last year we have continued to make improvements to the technology, including the ability to perform high-resolution images of larger regions of tissue. 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.

PUBLICATIONS

Marcon L, Sharpe J.
"Turing patterns in development: what about the horse part?"
Curr Opin Genet Dev, 22(6):578-84 (2012).

Sheth R, Marcon L, Bastida MF, Junco M, Quintana L, Dahn R, Kmita M, Sharpe J, Ros MA.
"Hox genes regulate digit patterning by controlling the wavelength of a Turing-type mechanism."
Science, 338(6113):1476-80 (2012).

Sanders PG, Cotterell J, Sharpe J, Isalan M.
"Transfecting RNA quadruplexes results in few transcriptome perturbations."
RNA Biol, doi: 10.4161/rna.22781. Epub 2012 Dec 12.

Mayer J, Swoger J, Ozga AJ, Stein JV, Sharpe J.
"Quantitative Measurements in 3-Dimensional Datasets of Mouse Lymph Nodes Resolve Organ-Wide Functional Dependencies."
Computational and Mathematical Methods in Medicine, 2012:128431 (2012).

Olarte OE, Licea-Rodriguez J, Palero JA, Gualda EJ, Artigas D, Mayer J, Swoger J, Sharpe J, Rocha-Mendoza I, Rangel-Rojo R, Loza-Alvarez P.
"Image formation by linear and nonlinear digital scanned light-sheet fluorescence microscopy with Gaussian and Bessel beam profiles."
Biomedical Optics Express, 3(7):1492-505 (2012).

Cheddad A, Svensson C, Sharpe J, Georgsson F, Ahlgren U.
"Image Processing Assisted Algorithms for Optical Projection Tomography."
IEEE Transactions on Medical Imaging, 31(1):1-15 (2012).

Nowlan N, Dumas G, Tajbakhsh S, Prendergast P, Murphy P.
"Biophysical stimuli induced by passive movements compensatefor lack of skeletal muscle during embryonic skeletogenesis."
Biomechanics and Modeling in Mechanobiology, 11(1-2):207-19 (2012).

Briscoe J, Sharpe J.
"Genetics of system biology."
Curr Opin Genet Dev, 22(6):523-6 (2012).

Reviews

Ludewig B, Stein JV, Sharpe J, Cervantes-Barragan L, Thiel V, Bocharov G.
"A global "imaging" view on systems approaches in immunology."
Eur J Immunol, 42(12):3116-25 (2012).

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 with Yogi Jaeger), Jae-Seong Yang

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

Technicians: Sira Martinez, Tony Ferrar, Violeta Beltran, Hannah Benisty

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, etc.) 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 2012 we have in collaboration with Anne Claude Gavin and Peer Bork's group determined the complete phosphrylome and acetylome of *M. pneumoniae* (Van Noort *et al.*, 2012). Also we have identified a new group of small RNAs in bacteria that we denominate them tssRNA. These RNAs are different from abortive ones and they mapped precisely the beginning of transcription (Yus *et al.*,2012) (Figure 1). This year in collaboration with Sanofi we have started the engineering of *M. pneumonia* to treat human diseases.

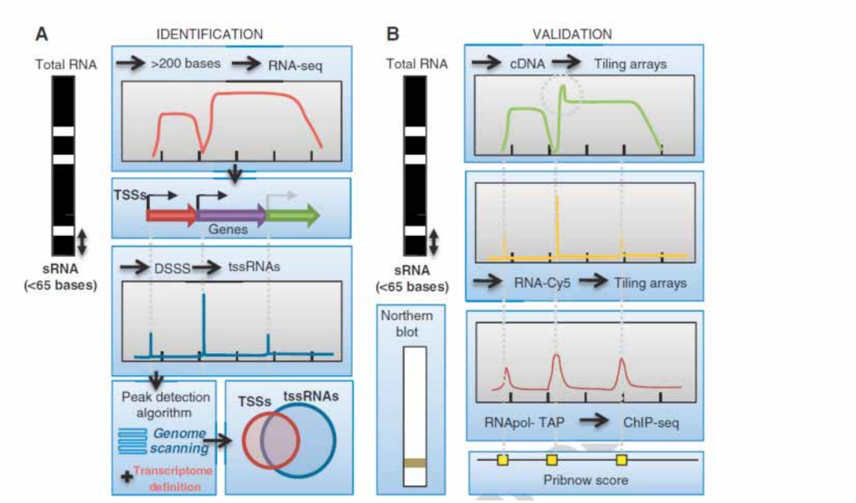


Figure 1.
Bacterial tssRNA identification. *A. Methodology to detect tssRNAs. B. tssRNAs map preferentially to the promoter regions. Upper panel: the mapping locations of various tssRNAs; middle panel: mRNA levels of the genomic region measured by DSSS (Vivancos et al.); lower panel: the coexpression matrix of the nine genes composing an operon located on the shown genome area. tssRNAs are located at the TSS of the different alternative transcripts of this operon.*

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.

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 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 2012 we have developed a new webserver tool (SAPIN) that can be used to determine competing interactions in a protein network using structural reconstruction (Yang *et al.*, 2012) (Figure 2). We have also shown how changing protein specificity in promiscuous protein interactions can increase signal output without affecting affinity (Szegezdi *et al.*, 2012), as well as design mutations that decrease protein-receptor promiscuity (Kimberley *et al.*,2012).

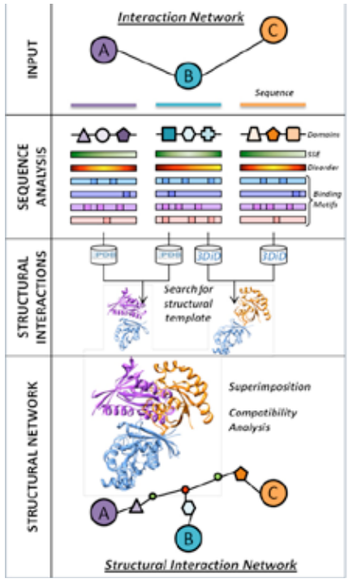


Figure 2
Overview of the SAPIN pipeline. *A) As an input, the pipeline takes a network in this case with three proteins A, B, and C and two interactions A-B and A-C, and the related protein sequences. B) First, the sequences are analyzed to determine the domain composition, Secondary Structure Elements (SSEs), disordered regions, binding motifs and phosphorylation sites. C) Second, the available structural data is mapped to identify a potential structural template for each interaction. D) Finally, if a protein has at least two structural interacting partners, the interactions are superimposed on the reference domain (in this case the domain from protein A, in blue) and the interacting domains are analyzed for clashing. This is showed in a final structural interaction network by adding nodes for the domains involved in the interactions, and for the interfaces through which the binding takes place.*

PUBLICATIONS

Yang JS, Campagna A, Delgado J, Vanhee P, Serrano L, Kiel C.
"SAPIN: A framework for the structural analysis of protein interaction networks"
Bioinformatics, 28(22):2998-2999 (2012).

Simões-Correia J, Figueiredo J, Lopes R, Stricher F, Oliveira C, Serrano L, Seruca R.
"E-cadherin destabilization accounts for the pathogenicity of missense mutations in hereditary diffuse gastric cancer."
PLoS ONE, 7(3):e33783 (2012).

Van Noort V, Seebacher J, Bader S, Mohammed S, Vonkova I, Betts MJ, Kühner S, Kumar R, Maier T, O'Flaherty M, Rybin V, Schmeisky A, Yus E, Stülke J, Serrano L, Russell RB, Heck AJR, Bork P, Gavin AC.
"Cross-talk between phosphorylation and lysine acetylation in a genome-reduced bacterium."
Molecular Systems Biology, 8:571 (2012).

Kimberley FC, Van Der Sloot AM, Guadagnoli M, Cameron K, Schneider P, Marquart JA, Versloot M, Serrano L, Medema JP.
"The design and characterization of receptor-selective APRIL variants."
Journal of Biological Chemistry, 287(44):37434-37446 (2012).

Yus E, Güell M, Vivancos AP, Chen WH, Lluch-Senar M, Delgado J, Gavin AC, Bork P, Serrano, L.
"Transcription start site associated RNAs in bacteria."
Molecular Systems Biology, 8:585 (2012).

Szegezdi E, Van Der Sloot AM, Mahalingam D, O'Leary L, Cool RH, Muñoz IG, Montoya G, Quax WJ, De Jong S, Samali A, Serrano L.
"Kinetics in signal transduction pathways involving promiscuous oligomerizing receptors can be determined by receptor specificity: Apoptosis induction by TRAIL."
Molecular and Cellular Proteomics, 1(3):M111.013730 (2012).

Serrano L.
"A quantitative systems biology study on a model bacterium."
FEBS Journal, 279(1):41-41 (2012).

Bergman JE, Janssen N, van der Sloot AM, de Walle HE, Schoots J, Rendtorff ND, Tranebjaerg L, Hoefsloot LH, van Ravenswaaij-Arts CM, Hofstra RM.
"A novel classification system to predict the pathogenic effects of CHD7 missense variants in CHARGE syndrome."
Hum Mutat, 33(8):1251-60 (2012).

Reviews

Kiel C, Serrano L.
"Structural data in synthetic biology approaches for studying general design principles of cellular signaling networks."
Structure, 20(11):1806-1813 (2012).

Magis C, van der Sloot AM, Serrano L, Notredame C.
"An improved understanding of TNFL/TNFR interactions using structure-based classifications."
Trends in Biochemical Sciences, 37(9):353-363 (2012).

Kiel C, Serrano L.
"Challenges ahead in signal transduction: MAPK as an example."
Current Opinion in Biotechnology, 23(3):305-314 (2012).

Lamond AI, Uhlen M, Horning S, Makarov A, Robinson CV, Serrano L, Hartl FU, Baumeister W, Werenskiold AK, Andersen JS, Vorm O, Linial M, Aebersold R, Mann M.
"Advancing cell biology through proteomics in space and time (PROSPECTS)."
Molecular and Cellular Proteomics, 11(3):O112.017731 (2012).

Képès F, Jester BC, Lepage T, Rafiei N, Rosu B, Junier I.
"The layout of a bacterial genome."
FEBS Lett, 586(15):2043-8 (2012).

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, Yolanda Schaerli, Manjunatha Kogenaru, Luis Diambra
Students:	Diego Barcena, Vivek Senthivel
Technicians:	Rebecca Baumstark, Carmen Agustin Pavon, Maria Lluisa Gili

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 behave in the ways 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. 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. & Choo, Y. *Nature Biotechnology*, 19:656-60; 2001). Thus, we aim to contribute to the field of endogenous gene repair using zinc finger nucleases.

The p53 pathway is the most common mutation in human cancer and, following our recent study on p53 gene repair (*PLoS One*, 6(6):e20913; 2011) we followed-up with a review explaining the engineering methods 'Zinc finger nucleases: how to play two good hands' (*Nature Methods*, 9:32-35; 2012).

Our work on Huntington's disease reached a point where we successfully observed disease symptom improvements in mice, after treatment with zinc finger proteins (ZFPs) targeted to repress the mutant gene (PNAS, 109(45):E3136-45; 2012). We used a gene therapy approach to deliver therapeutic viruses directly to mouse brains.

2. The seven ways to make a stripe (Y. Schaerli & Sharpe Group)

As well as our ongoing work on top-down large scale gene network rewiring in *E. coli* ("*Evolvability and hierarchy in rewired bacterial gene networks*". *Nature*, 452(7189):840-5; 2008) new manuscript *in prep.*: "*The propagation of perturbations in shuffled gene networks*"), we have developed a bottom-up project in collaboration with the group of James Sharpe:

We have successfully engineered stripe patterns in bacteria in response to a morphogen gradient, as in the classical "French Flag" paradigm of developmental pattern formation. We have developed a platform for engineering many kinds of 3-node network topologies. These have been used to build stripe forming networks that have distinct mechanisms. The computational study (Cotterrell and Sharpe, "*An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients*". *Molecular Systems Biology,* 6:425; 2010) showed that there are only 6 or 7 distinct mechanisms for stripe formation, and some of them have not yet been discovered in nature. We have built some of these networks and have shown that computationally-predicted networks can indeed function in synthetic biology.

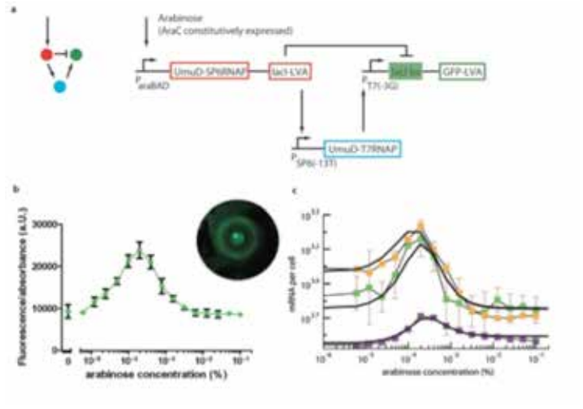


Figure 1:
Stripe formation from an incoherent feed-forward type 3 mechanism a) General topology of the network and its implementation in the network scaffold. b) Cells carrying the network display a stripe of fluorescence in an arabinose gradient. Inset: the cells were spread out on an agar plate containing a central paper disk, from which an arabinose gradient was established. After growth, a ring of fluorescent cells was detected: a stripe. c) Related networks were built by changing promoter strengths. They behaved as predicted by the in silico model (experimental data are plotted in colours, with error bars; computer model predictions are shown by thick black lines).

PUBLICATIONS

Isalan M.
"Zinc finger nucleases: how to play two good hands."
Nature Methods, 9:32–34 (2012).

Garriga-Canut M, Agustín-Pavón C, Herrmann F, Sánchez A, Dierssen M, Fillat C, Isalan M.
"Synthetic zinc finger repressors reduce mutant huntingtin expression in the brain of R6/2 mice."
PNAS, 109(45):E3136-45 (2012).

Mark Isalan.
"Systems biology: A cell in a computer."
Nature, 488(7409):40-1 (2012).

Peccoud J, Isalan M.
"The PLOS ONE Synthetic Biology Collection: Six Years and Counting."
PLoS ONE, 7(8):e43231 (2012).

Isalan M, Schaerli Y.
"Biología sintética en busca de andamios / Synthetic biology in search of a scaffold. "
Dossier científico SEBBM (Sociedad Española de Bioquímica y Biología Molecular) (2012).

Sanders PG, Cotterell J, Sharpe J, Isalan M.
"Transfecting RNA quadruplexes results in few transcriptome perturbations."
RNA Biol, doi: 10.4161/rna.22781. Epub 2012 Dec 12.

Isalan M, Schaerli Y.
"Building synthetic gene circuits from combinatorial libraries: screening and selection strategies."
Molecular BioSystems, in press.

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.

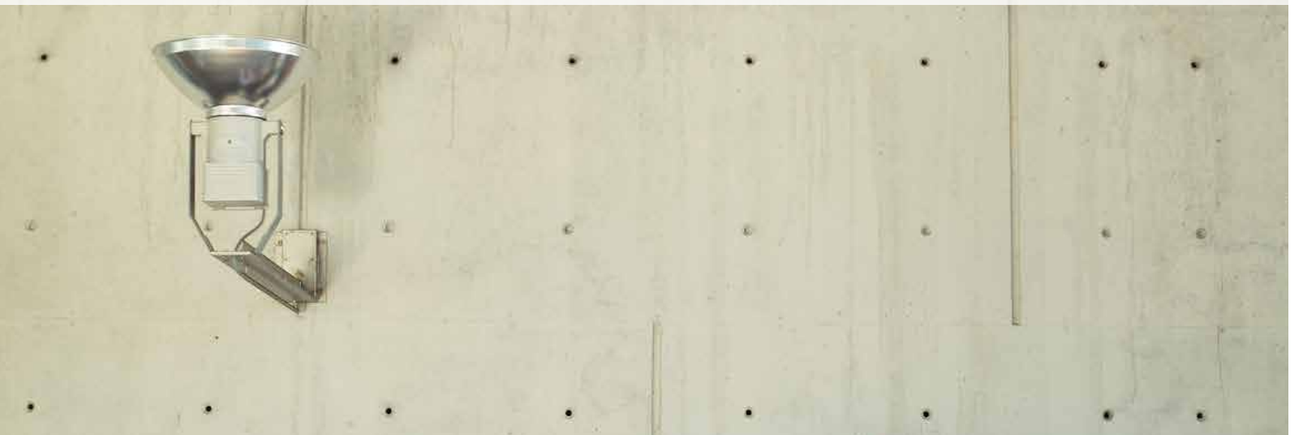
Group Structure:

Group Leader: Ben Lehner

Postdoctoral Researchers: Jennifer Semple, Rob Jelier, Mirko Francesconi, Solip Park, Fran Supek

PhD Students: Adam Klosin, Janet Melling, Kadri Reis

Technician: Laura Biondini



SUMMARY

Our main interest is in the biology of individuals: what are the causes of phenotypic variation amongst individuals and how can we predict this variation? Conversely, we are also interested in how phenotypic variation is constrained, both during development and during evolution. Being able to predict how individuals vary is both a fundamental challenge for biology, and one that is central for the development of personalised medicine: a patient does not want to know the typical outcome of a mutation that they carry, they want to know what will actually happen to them. We primarily use *C. elegans* as a model system but also increasingly are testing some of our ideas about genetics and phenotypic robustness using data from cancer genome projects.

RESEARCH PROJECTS

1. Inter-individual variation in the outcome of mutations and epigenetics

To what extent can variation in phenotypic traits such as disease risk be accurately predicted in individuals? Our recent studies using model organisms have both directly tackled the challenge of accurately predicting phenotypic variation from individual genome sequences ('whole-genome reverse genetics') and the reasons why, in many cases, this is impossible. These studies argue that only by combining genetic knowledge with *in vivo* measurements of biological states will it be possible to make accurate genetic predictions for individual humans.

Continuing this work, we are further dissecting the causes of inter-individual variation in the outcome of mutations and phenotypic variation more generally, and are also tackling the epigenetic mechanisms by which inter-individual variation is transmitted (Figure 1).

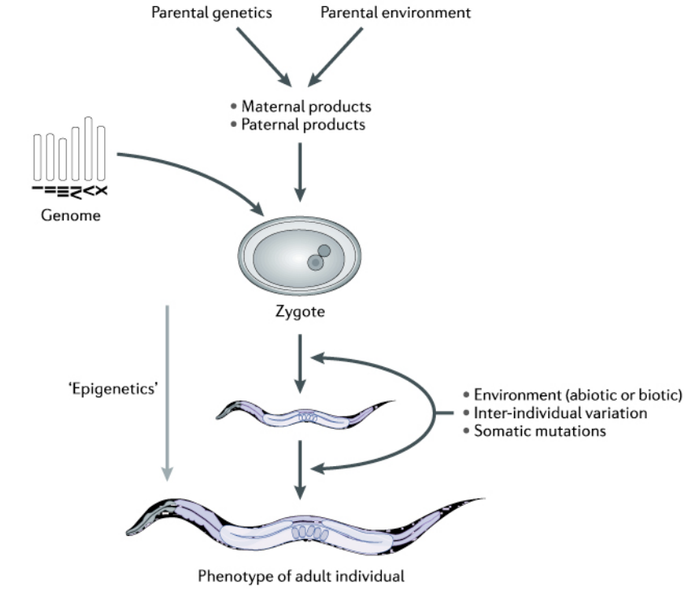


Figure 1.
Influences on inter-individual phenotypic variation in animals. We are particularly interested in understanding the reasons why the inherited genotype of an individual is insufficient to accurately predicts its phenotype.

2. Cancer genomes

A second focus of our current research is to use cancer genome sequencing data (which is by far the largest genotype-to-phenotype sequence resource in any species) to address some basic questions in genetics. Moreover, we are also applying some of the insights derived from model organisms to understanding cancer biology and how to specifically target cancer cells.

For example, in a recent publication we used cancer genome data to address the question of how mutation rates vary across the human genome. Cancer genomes provide the first genome-scale information on somatic mutation processes. We found that regional mutation rates vary considerably across the human genome, and that this regional variation is closely related to how DNA is packaged into chromatin. This strong association between mutation rates and chromatin organization is upheld in samples from different tissues and for different mutation types. This suggests that the arrangement of the genome into heterochromatin- and euchromatin-like domains is a dominant influence on regional mutation-rate variation in human somatic cells.

PUBLICATIONS

Schuster-Böckler, Lehner B.
"Chromatin organisation is a major influence on regional mutation rates in human cancer cells."
Nature, 488:504-507 (2012).

Casanueva MO, Burga A, Lehner B.
"Fitness trade-offs and environmentally-induced mutation buffering in isogenic C. elegans."
Science, 335:82-85 (2012).

Semple JI, Biondini L, Lehner B.
"Generating transgenic nematodes by bombardment and antibiotic selection."
Nature Methods, 9:118-9 (2012).

Warnecke T, Supek F, Lehner B.
"Nucleoid-Associated Proteins Affect Mutation Dynamics in E. coli in a Growth Phase-Specific Manner."
PLoS comput Biol, 8(12):e1002846 (2012).

Vavouri T, Lehner B.
"Human genes with CpG island promoters have a distinct transcription-associated chromatin organization."
Genome Biol, 13(11):R110 (2012).

Lehner B, Park S.
"Cancer: exploiting collateral damage."
Nature, 488(7411):284-5 (2012).

Burga A, Lehner B.
"Beyond genotype to phenotype: why the phenotype of an individual cannot always be predicted from their genome sequence and the environment that they experience."
FEBS J, 279(20):3765-75 (2012).

Book chapters

Fraser AG, Lehner B.
"Systems biology of Caenorhabditis elegans."
In: The Handbook of Systems Biology (eds. Walhout, Dekker, Vidal, 2012).

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
Technicians:	Octavi Domingo, Mariana Lopez-Matas, Moraea Philipps, Vani Rajendran
Postdoctoral Fellows:	Alex Gomez-Marin, Daeyeon Kim, Sam Reid
PhD Students:	Andreas Braun, Julia Riedl, Aljoscha Schulze
Undergraduate Students:	Sinja Kraus, Ana Jimenez Pascual

SUMMARY

To survive in an ever-changing environment, animals are confronted with the need to continuously process vast amounts of sensory information. Which features within this ocean of signals are actually captured by the sensory system? We address this question by studying the olfactory system of the fruit fly *Drosophila melanogaster* larva. We aim at understanding how orientation decisions come about in terms of neural circuit computation. Our research combines experimental and computational approaches to define the sensorimotor integration that converts sensory inputs into motor outputs. To this end, we have developed tools to track behaviour in real time and to monitor the activity of targeted neurons. Using optogenetics, we create virtual sensory realities to test mechanistic hypotheses about the spatio-temporal integration of olfactory cues. We seek to identify the neural circuits involved in the processing of olfactory information and in its conversion into orientation decisions. Finally, we examine the logic underlying the integration of sensory inputs arising from different modalities. 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. Individual odorant receptors have overlapping, yet distinct, ligand tuning properties. Each OSN can be viewed as a distinct information channel to the olfactory system. To disentangle the contribution of single OSNs to the representation of odour stimuli, we have invented a novel method to perform *in vivo* extracellular recordings from identified OSNs. To separate odour-driven signals from irrelevant background, we exploit optogenetics to identify spikes of interest throughout the recording.

We started off by comparing the activity elicited in identified OSNs in response to different odours, and subsequently extended this approach by using a microfluidics control system that allows us to accurately control the odour delivery in time and space. We are investigating the computational principles underpinning the encoding of dynamical odour stimuli comparable to those experienced by a larva during free behaviour in odour gradients. We find that the activity pattern of a single OSN does not simply respond to the absolute value of the stimulus intensity, but that it also computes higher-order features related to the rate of change of the stimulus. We have devised a computational model that accurately recapitulates the activity patterns of a single OSN (Fig. 1). We found that the patterns of OSN activity capture relative changes in stimulus intensity and their temporal integration. In addition, we are aiming to clarify whether and how features related to the quality of an odour are represented by single functional OSNs.

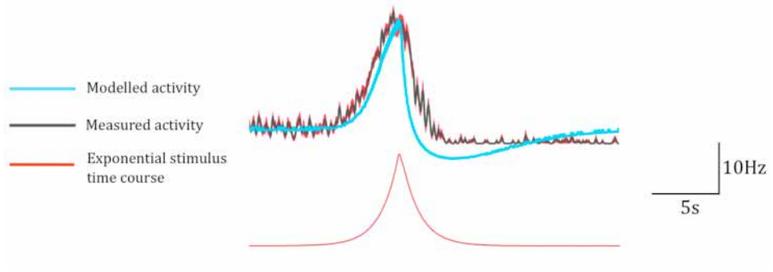


Figure 1:
Quantitative model predicting the firing rate of a single OSN in response to a dynamic odour stimulation (red trace at the bottom). The firing rate elicited by an exponential odour ramp is predicted by the model (blue trace) in excellent agreement with the experimental data (black trace).

2. Remote control of orientation behaviour in virtual reality experiments

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

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. 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. In more than 75% of the cases, a turn is implemented toward the side of higher concentrations. Larvae appear to perform a spatio-temporal comparison of differences in concentration measured during lateral head movements (head casting), which suggests the participation of a spatial short-term memory to the decision-making process guiding larval chemotaxis.

To test this model, we combine real-time tracking, electrophysiology, optogenetics and quantitative analysis of behaviour. 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 are activated according to the behavioural history of the larva (head casting, running or turning). Using optogenetics in conjunction with electrophysiology, we systematically explore the coding space of a single OSN and subsequently study the effect of synthetic patterns of OSN activity (Fig. 1) on the olfactory behaviour of freely moving larvae. We are now in position to investigate the computational bases of spatiotemporal integration during orientation behaviour by reverse engineering odour evoked activity patterns with light. To date, we have managed to facilitate and suppress the onset of turns by associating runs with either decreases or increases in light intensity. Our goal is to develop a mathematical model that accurately predicts behavioural decisions from olfactory signals.

3. Mapping new neuronal centres participating in orientation behaviour

Sam Reid & Julia Riedl

Very little is known about the circuit computations that transform sensory information into behavioural output. Whereas the first- and second-order neurons of the fly olfactory circuit are well characterized, our knowledge about the downstream neurons integrating dynamic olfactory information remains elusive. To uncover the neural circuits involved in the sensorimotor transformations directing chemotaxis, we have conducted a large-scale behavioural screen. 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 from the Drosophila Genetic Resource Center (so-called NP collection) were crossed with a UAS reporter expressing tetanus toxin (TNT), a protein that efficiently silences neuronal activity by impairing synaptic transmission.

Based on a behavioural quantification and a detailed assessment of expression patterns, we focused on the most promising NP[Gal4] driver line covering a distinct population of neurons with a deficit specific to chemotaxis. This line shows altered integration of odour information during forward locomotion, failing to initiate reorientation at the correct time point (Fig. 2). It is impaired in the transition from runs to turns, 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.

In parallel, we are taking part in the collaborative Larval Olympiad Project at Janelia Farm (HHMI). We have developed a behavioural assay that is both high-throughput and high-resolution. This system has been used to screen another collection of 800 Gal4 lines that label small sub-regions of the larval brain. The results of this screen should reveal additional neurons contributing to the control of chemotaxis. Hypotheses about circuit-function relationships are being tested in an on-going secondary screen. Our goal is to identify the interneurons involved in the sensorimotor processes that direct the production of distinctive and predictable motor responses.

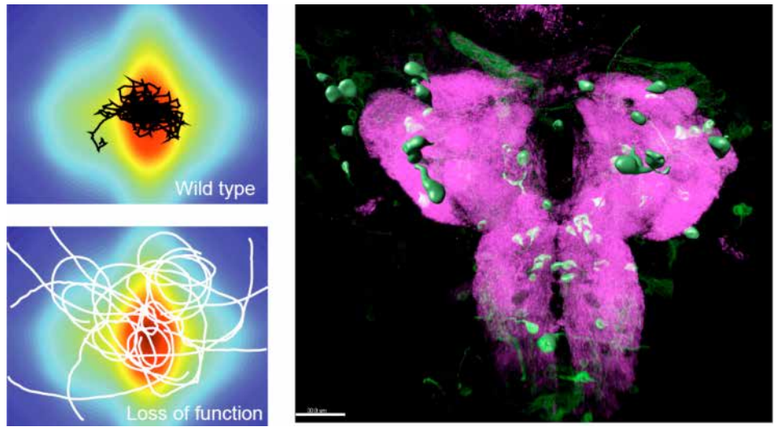


Figure 2:
Identification of new neurons participating in the processing of olfaction information. ***(Left)*** *Chemotaxis assay featuring a radially symmetrical odour gradient. Concentration reported as a false colour code where blue corresponds to the minimum intensity and red to the maximum intensity under the single odour source (centre of plate). Illustration of the chemotactic behaviour of wild type larvae (black traces, top panel) and the NP[Gal4] line with the loss-of-function phenotype line isolated in our screen (white traces, bottom panel). Each plot displays 12 superimposed trajectories corresponding to a different animal. (Right) Larval brain showing the expression pattern of the NP[Gal4] line with the loss-of-function phenotype illustrated in the left panel (green channel: GFP driven by the NP[Gal4] line; magenta channel: neuropile staining). Scale bar corresponds to 30 µm.*

4. Neural computation underpinning multisensory integration in the larva

Andreas Braun, Octavi Domingo & Daeyeon Kim

To survive in a complex and changing environment, our brain has evolved to construct a coherent representation of the external world. This requires combining information conveyed by noisy sensors reporting different features of the world — a process called multisensory integration. This integration mechanism should be flexible because sensory information is context-dependent and often corrupted by noise. Neural theories have shown that probabilistic inference offers a way to satisfy this necessity. An optimal solution is reached when each cue contributes to the final estimate in proportion to its reliability. Until now, this behaviour has been demonstrated in humans and primates where the underlying neural principles remain a mystery. Even though multisensory integration is commonly seen as a hallmark of cortical functions, this process could be in theory implemented with more basic neural networks.

We aim to demonstrate that the *Drosophila* larva is capable of multisensory integration. We investigate the mechanisms allowing larvae to exploit different sources of sensory information to direct navigational decision-making. Using a Bayesian probabilistic framework, we make predictions about the enhancement in behaviour observed upon integration of multi-sensory inputs. Our results suggest that even simple organisms are capable of representing probability distributions over sensory variables. The rationale for conducting these experiments in the larvae is that the powerful tools available for *Drosophila* and the numerical simplicity of the larval nervous system give us the opportunity to understand the neurobiology of multisensory integration and the origin of variability in behaviour.

PUBLICATIONS

Reviews

Gomez-Marin A, Louis M.
“Active sensation during orientation behavior in the Drosophila larva: more sense than luck.”
Curr Opin Neurobiol, 22:208-215 (2012).

Book chapters

Louis M, Phillips M, Lopez-Matas M, Sprecher S.
“Behavioral Analysis of Navigation Behaviors in the Drosophila Larva.”
In: *The Making and Un-Making of Neuronal Circuits in Drosophila*, Volume 69, B.A. Hassan, ed. (Humana Press), pp. 163-19 (2012).

Riedl J, Louis M.
“Behavioral Neuroscience: Crawling is a No-brainer for Fruit Fly Larvae.”
Curr Biol, 22:R867-R869 (2012).

Gomez-Marin A, Partoune N, Stephens GJ, Louis M.
“Automated Tracking of Animal Posture and Movement during Exploration and Sensory Orientation Behaviors.”
PLoS One, 7:e41642 (2012).
Correction in: PLoS One, 7(10); Error ir figure 2

SYSTEMS BIOLOGY

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

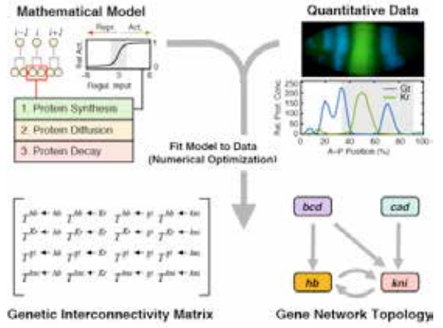
Group Structure:

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
Undergraduate Students:	Kolja Becker, Elisa Beltrán

SUMMARY

Our laboratory is interested in the origin and evolution of biological form. The diversity of body shapes in nature is overwhelming. Nevertheless, there are regularities and recurring motifs, such as segmented body plans, or striped and dotted skin patterns. These patterns seem to occur more often than others. This is because they frequently arise in populations and therefore are more likely to be selected during evolution. However, we do not understand the orgin of these beneficial phenotypes. 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 interactions. Therefore, its study requires a systems-biology approach.

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



Our work is based on a computational method (reverse-engineering), which combines quantification of gene expression with mathe-matical modelling using gene circuits (Fig. 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 two particularly well studied gene regulatory networks (Fig. 2): the gap gene system involved in segment determination, and the thoracic patterning network involved in positioning sensory bristles during early insect development. We compare gene networks across different species of dipterans (flies, midges, and mosquitoes) to learn which interactions are conserved, and which have diverged between species. We plan to test these predictions experimentally, using RNA interference (RNAi), and trace changes in regulatory interactions to molecular changes in regulatory sequences.

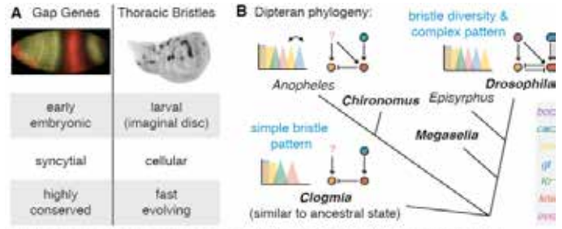


Figure 2. (A) Developmental gene regulatory networks studied in this project. (B) Comparison of networks across dipteran species: The thoracic bristle patterning network has evolved from simple to complex, diversified patterns. The gap network shows convergent evolution: gap domains are shown schematically (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. Species to be used in this project in bold.

RESEARCH PROJECTS

1. A quantitative, comparative study of gap gene regulation in dipterans

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

Aim. As the main line of investigation in our group, we aim to create gene circuit models of the gap gene network in three species of dipterans: *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. Gene circuits are then obtained by fitting models to data (Fig. 1). The resulting models predict gene network topologies that allow us to test, which aspects of the network have been conserved, and which ones diverged during evolution.

Results. We have characterized the early development of all three species. We have established RNAi protocols in *Megaselia* (Fig. 4) and are doing the same for *Clogmia*. We have created early embryonic transcriptomes (in review at BMC Genomics) 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 *Drosophila*, *Megaselia*, and *Clogmia* using colorimetric *in situ* hybridization (Fig. 3). We have developed a novel data quantification pipeline for our mRNA data (published in PLoS ONE this year), which was used to create expression databases and integrated expression patterns for all three species. These data are now being used to fit gene circuit models, and will serve as a basis for an *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 (published this year in *PLoS Computational Biology*). We are now obtaining models for the other two species.

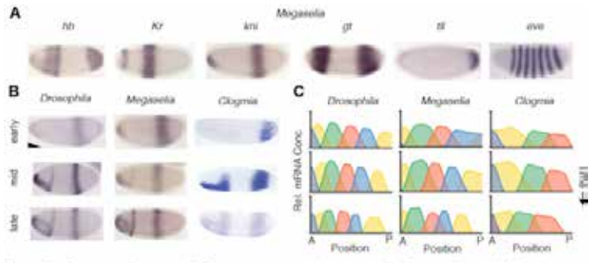


Figure 3. Gap gene datasets. (A) Gap gene expression patterns in *Megaselia* at mid blastoderm. (B) Comparison of *kni* expression over time in all three species. (C) Integrated datasets for *Drosophila*, *Megaselia* and *Clogmia*. Embryos stained by *in situ* hybridisation. A = anterior, P = posterior.

2. 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 obtaining the *Clogmia* models that will be used as the starting point for this study.

3. Modeling Gap Gene Translation in *Drosophila melanogaster*

Kolja Becker

Aim. In this project, we wanted to establish whether regulation of gap genes requires post-transcriptional regulation. Furthermore, translational models allow us to adapt, test, and improve various algorithms used for model fitting and analysis.

Results. We have created models for gap gene translation, which show that post-transcriptional regulation is not required for pattern formation, but is required for controlling the precise levels of gap gene products in the embryo. Our work shows that our model fitting procedure results in unique and robust solutions, which yields precise predictions of model parameters, such as gap protein production and diffusion rates.

4. A systems-level analysis of *giant* (*gt*) regulation in *Drosophila melanogaster*

Astrid Hörmann

Aim. Since we still lack a precise mechanistic understanding of eukaryotic gene regulation, 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 studying the regulation of the gap gene *giant (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.

5. 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 has been published in *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.

6. Inferring the thoracic bristle patterning network of *Drosophila*

Bárbara Negre, Elisa Beltrán

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

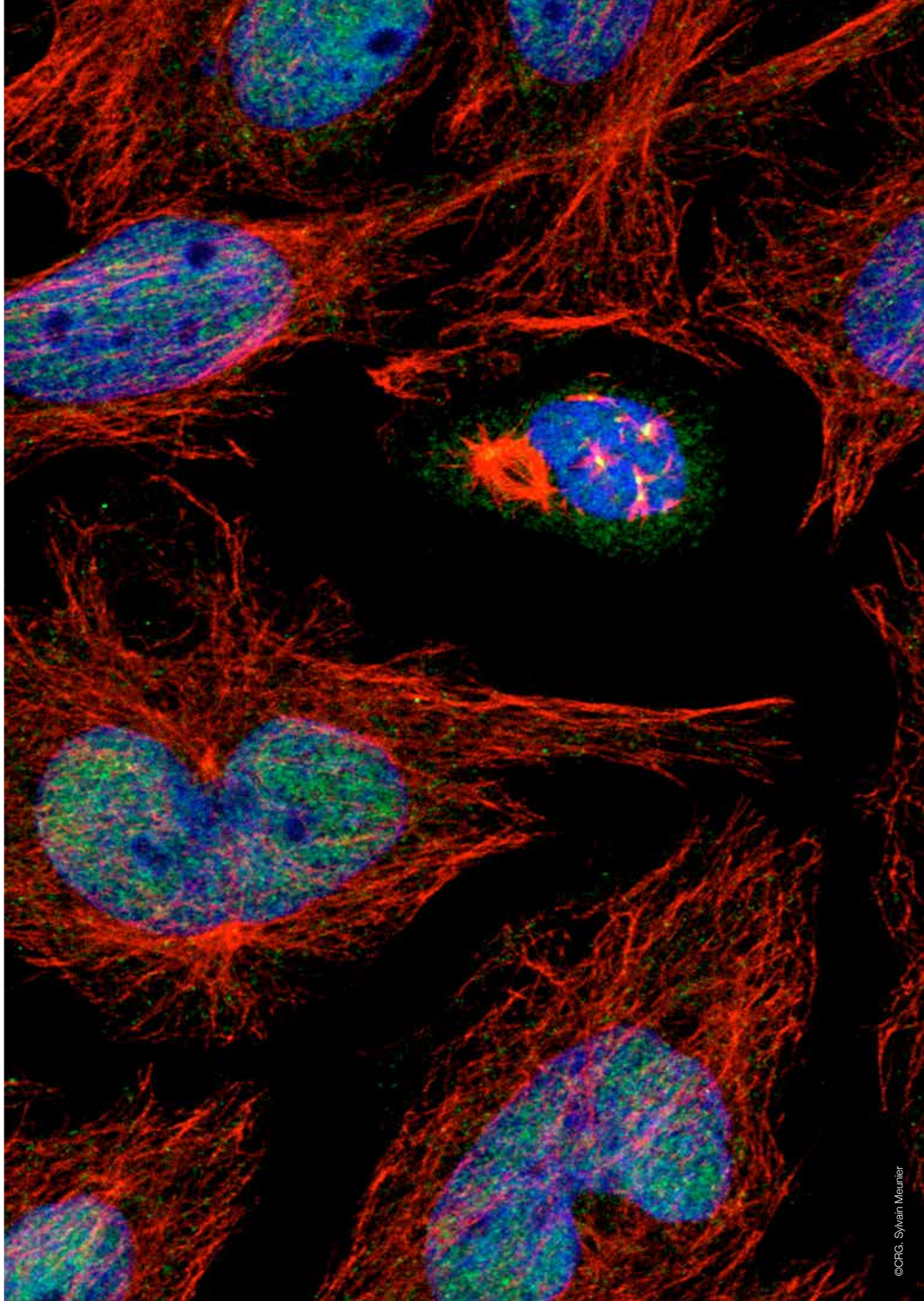
Jaeger J, Crombach A.
"Life's Attractors: Understanding Developmental Systems Through Reverse Engineering and In Silico Evolution. "
Advances in Experimental Medicine and Biology, 751:93-119 (2012).

Crombach A, Wotton KR, Cicin-Sain D, Ashyraliyev M, Jaeger J.
"Efficient reverse-engineering of a developmental gene regulatory network."
PLoS Computational Biology, 8(7):e1002589 (2012).

Crombach A, Cicin-Sain D, Wotton KR, Jaeger J.
"Medium-Throughput Processing of Whole Mount In Situ Hybridisation Experiments into Gene Expression Domains."
PLoS ONE, 7(9): e46658 (2012).

Jaeger J, Manu, Reinitz J.
"Drosophila blastoderm patterning."
Current Opinion in Genetics and Development, 22(6):533-41 (2012).

Janssens H, Crombach A, Richard Wotton K, Cicin-Sain D, Surkova S, Lu Lim C, Samsonova M, Akam M, Jaeger J.
"Lack of tailless Leads to an Increase in Expression Variability in Drosophila Embryos."
Developmental Biology, in press.



©CRG, Sylvain Meunier

CORE FACILITIES

Unit: Genomics Unit

Unit Structure:

Head of the Unit: Heinz Himmelbauer

Laboratory Manager: Anna Ferrer

Laboratory Technicians: Anna Menoyo, Maik Zehnsdorf, Núria Andreu (maternity leave until September 2012)
Irene González, Ester Cuenca (until April 2012), Maria Aguilar, Inês Guimaraes
Magda Montfort

Bioinformaticians: Debayan Datta, Manuela Hummel (maternity leave March-October 2012)
Sarah Bonnin (sabbatical from November 2012), Juliane Dohm (MPI Berlin)
Sarah Jäger (MPI Berlin, from October 2012)

PhD Student: André Minoche (MPI Berlin)

SUMMARY

The scope of the Genomics Unit is to provide excellent, state-of-the-art services in the fields of microarray technology and next-generation sequencing. The Unit utilizes different sequencing platforms (Illumina Genome Analyzer, Illumina HiSeq2000, Roche FLX). Array-based work focuses on the Agilent platform. In addition to providing service, the Unit performs its own research in the genomics field and is dedicated to test, implement, and to develop new protocols (wet lab and bioinformatics) to advance the analysis of genes and genomes.

SERVICES

A large number of different questions in biology and in biomedical sciences can be addressed using high-throughput sequencing or microarrays as tools. For performing sequencing services, the Unit is equipped with two Illumina Genome Analyzer IIx sequencers, one Illumina HiSeq2000 instrument and one FLX sequencer from 454/Roche. No new sequencing instruments were purchased during the year. In 2012, the Unit performed 14 Illumina GA sequencing runs (-67% compared to 2011). On the Hiseq2000, 64 runs were performed (+12%). 21 sequencing runs were performed on the Roche/FLX sequencer (-52%). Illumina GA sequencing was phased out at the end of 2012 since it was no longer competitive due to the 5 to 10-fold higher data output achieved by Illumina HiSeq sequencing.

The second generation sequencing platform at the CRG is certainly unique in offering a wide range of state-of-the-art protocols, encompassing 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. New protocols were set up for the preparation of amplification-free genomic libraries and robust protocols for ChIP-seq library pooling were implemented. Also, two protocols for directional sequencing of polyadenylated transcripts were tested, and the Illumina v 1.5 protocol for preparation of small RNA libaries was set up for routine service work.

In the microarray service, 141 microarray slides were hybridised and processed, most of them for gene expression profiling on the Agilent platform. A few slides were for miRNA profiling (Agilent), and for CGH on Nimblegen arrays. The production of Nimblegen arrays, however, was discontinued by the manufacturer at the end of 2012, so that CGH services will be continued with the Agilent platform. In addition, experimental protocols were set up for the Fluidigm-Biomark platform which enables highly parallelized real-time PCR, in a pilot project together with Thomas Graf's group.

For microarrays, services offered to users include the quality control of DNA and RNA samples, probe preparation and slide hybridisation, primary data quality control and, by request, advanced data analysis. For sequencing services, the Unit staff checks quality and quantity of the supplied samples, prepares sequencing libraries, and performs sequencing runs. After basecalling, the quality of the sequencing reads is confirmed using custom scripts and public tools, before being made available as fastq sequence data files.

RESEARCH PROJECTS

The research of the Genomics Unit focuses on the development of new procedures, both in the lab, and for data analysis.

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*) based on combining data from three different sequencing technologies (Illumina, 454, Sanger). While such an approach provides highly accurate, high-quality genome assemblies consisting of large scaffolds, the costs to prepare the necessary input data are substantial. We therefore tested if Illumina sequencing (combination of paired end and mate-pair sequences) on its own is sufficient to prepare high quality genome drafts. For the sugar beet reference genotype, we assembled Illumina data only (paired-end data plus mate-pair data). We obtained an assembly similar in size as the reference, but with 10-fold smaller N50 length. Thus, while overall quality of Illumina-based assemblies is good, such assemblies tend to be more fragmented. Next, we analysed if heterozygous genomes can be assembled well from Illumina data. Mixing data *in silico* from two homozygous lines clearly showed that this is not the case. In practise, we made similar ob-

servations in cooperation with Roderic Guigó. When sequencing *Drosophila* genomes, the scaffold N50 length varied between 5kb and 440kb, depending on inbreeding status of the sequenced strains. In contrast, when sequencing the genome of a *haploid* animal, we obtained a scaffold N50 length of 1300kb.

The analysis of genomic variation is generally addressed by re-sequencing, i.e. the generation of short-read data followed by mapping of reads against a reference sequence. Such an approach will fail on genomic regions too diverged from the reference. We therefore compared two different strategies for variation discovery, a) read mapping and b) scaffold mapping. While read mapping allowed interrogation of highly fragmented regions, mapping success decreased within regions that were locally rich in variants. Accordingly, we observed that read mapping covered more positions in the reference, while scaffold mapping discovered more variants. Thus, the two approaches proved to be complementary, and both should be considered in cases where regionally high variation is expected.

Gene expression measurements taken on tissues or on cell samples provide an average over many cells. In many cases, measuring the transcriptional activities of single cells is desirable. However, such analyses are technically demanding. We have therefore initiated projects aiming at the analysis of transcriptomes from small amounts of samples. So far, expression profiling has been successful from 100 pg of RNA, corresponding to the RNA content of 10 cells.

1a) 1b)

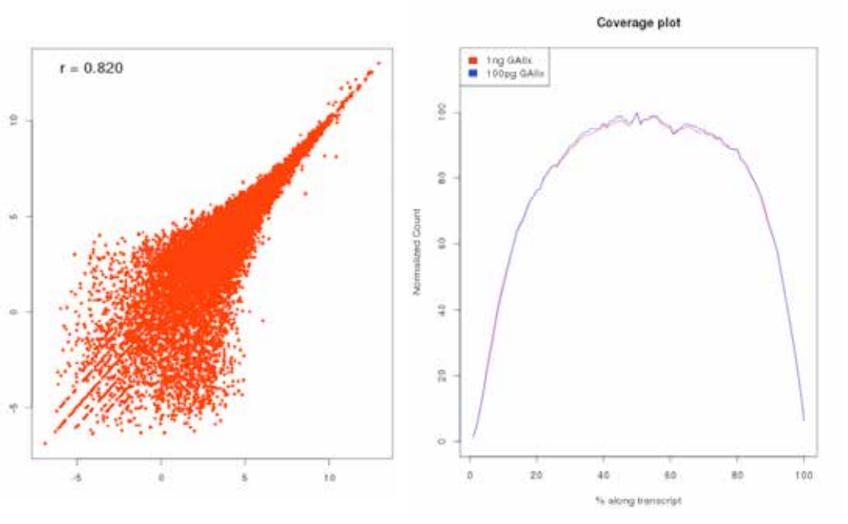


Fig.1
a) Low-RNA transcriptome sequencing experiment. Illumina libraries were prepared from 100 pg of total RNA (y-axis) or 1 ng of total RNA (x-axis), and sequencing was performed on the Illumina Genome Analyzer II x. Reads were mapped with GEM against a set of reference transcripts, uniquely matching hits (0 mismatches) were collected and rpkm values were calculated for each transcript. log2 transformed rpkm values were plotted. Good correlation between the two experiments is observed.

b) Normalised transcript coverage plot. Even coverage is observed over the entire transcript length, without 3' or 5' bias.

PUBLICATIONS

Wollrab C, Heitkam T, Holtgräwe D, Weisshaar B, Minoche AE, Dohm JC, Himmelbauer H, Schmidt T.
"Evolutionary reshuffling in the Errantivirus lineage Elbe within the Beta vulgaris genome."
Plant J, 72(4):636-51 (2012).

Pin PA, Zhang W, Vogt SH, Dally N, Büttner B, Schulze-Buxloh G, Jelly NS, Chia TYP, Mutasa-Göttgens ES, Dohm JC, Himmelbauer H, Weisshaar B, Kraus J, Gielen JJL, Lommel M, Weyens G, Wahl B, Schechert A, Nilsson O, Jung C, Kraft T, Müller AE.
"The role of a pseudo-response regulator gene controls lifecycle adaptation in beet."
Current Biology, 22:1095-2101 (2012).

Dohm JC, Lange C, Holtgräwe D, Rosleff Sörensen T, Borchardt D, Schulz B, Lehrach H, Weisshaar B, Himmelbauer H.*
"Palaeohexaploid ancestry for Caryophyllales inferred from extensive gene-based physical and genetic mapping of the sugar beet genome."
Plant J, 70:528-540 (2012).
Correction: Plant J, 71(1):182 (2012).

Tsend-Ayush E, Kortschak RD, Bernard P, Lim SL, Ryan J, Rosenkranz R, Borodina T, Dohm JC, Himmelbauer H, Harley VR, Grützner F.
"Identification of mediator complex 26 (Crsp7) on platypus X1 and Y5 sex chromosomes: a candidate testis determining genes in monotremes?"
Chromosome Res, 20:127-138 (2012).

Berndt H, Harnisch C, Rammelt C, Stöhr N, Zirkel A, Dohm JC, Himmelbauer H, Tavanez JP, Hüttelmaier S, Wahle E.
"Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming."
RNA, 18:958-972 (2012).

Montagut C, Dalmases A, Bellosillo B, Crespo M, Pairet S, Iglesias M, Salido M, Gallen M, Marsters S, Tsai SP, Minoche A, Somasekar S, Serrano S, Bellmunt J, Himmelbauer H, Rovira A, Settleman J, Bosch F, Albanell J.
"Identification of a mutation in the extracellular domain of the epidermal growth factor receptor conferring cetuximab resistance in colorectal cancer."
Nature Medicine, 18:221-223 (2012).

Menzel G, Krebs C, Diez M, Holtgräwe D, Weisshaar B, Minoche AE, Dohm JC, Himmelbauer H, Schmidt T.
"Survey of sugar beet (Beta vulgaris L.) hAT transposons and MITE-like hATpin derivatives."
Plant Mol Biol, 78:393-405 (2012).

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, 195:507-523 (2012).

Morales E, Bustamante M, Vilahur N, Escaramis G, Montfort M, de Cid R, Garcia-Esteban R, Torrent M, Estivill X, Grimalt JO, Sunyer J.
"DNA Hypomethylation at ALOX12 is Associated with Persistent Wheezing in Childhood."
American Journal of Respiratory and Critical Care Medicine, 185(9):937-43 (2012).

CORE FACILITIES

Unit: CRG/UPF Proteomics Unit

Unit Structure:
Head of the Unit: Eduard Sabidó

Scientists: Eva Borràs (UPF), Cristina Chiva (UPF), Guadalupe Espadas-Garcia (CRG)
Francesco Mancuso (CRG), Ester Dionís (UPF), Mireia Ortega (CRG)

SUMMARY

The CRG/UPF Proteomics Unit is a joint effort of Universitat Pompeu Fabra (UPF) and the Center of Genomic Regulation (CRG) to create an innovative core facility that provides high quality proteomics services to its final users, by providing proper expertise and advice, and by developing new methods and techniques.

The Proteomics Unit provides full service in a variety of proteomics applications including sample preparation, protein quantification, identification of post-translational modifications, and data analysis, among others. In addition to the services provided to the research community, the Proteomics Unit also promotes internal technology-driven research as an essential task to keep the unit at the forefront of the proteomics field.

As a reference proteomics center, the Unit has state-of-the-art equipment consisiting on high resolution liquid chromatography and electrophoresis systems, and three advanced mass spectrometers to identify and quantify the proteins of interest. Specifically, the Proteomics Unit currently has two LTQ-Orbitrap mass spectrometers, a Q-Trap mass spectrometer, three nanoLC-HPLC, and a microLC-HPLC, thus being one of the best well-equipped units in Europe. The instruments available at the Proteomics Unit cover the most advanced techniques of the proteomics field, and are accessible to national and international scientists.

The Proteomics Unit is part of the *ProteoRed* and "*Proteomics Research Infrastructure Maximizing knowledge EXchange and access (PRIME-XS)*", a prestigious international consortium of twelve partners that grants access to state-of-the-art proteomics technology to the European biological and biomedical research community. Moreover, the CRG/UPF Proteomics Unit has refecently been acknowledged as a European Reference Proteomics Unit by the EU MERIL initiative.

SERVICES

The Proteomics Unit is offering a wide range of services to its users that cover many different types of sample preparation and fractionation, post-translational modification enrichment, depletions, protein identification and quantitation, experimental design, and data analysis.

Detailed list of services currently offered by the Proteomics Unit:

Sample Preparation
> Protein Digestion (FASP, In-solution, In-gel)
> Chromatographic Protein and Peptide Separation (Reverse Phase, Strong Cation Exchange, Strong Anion Exchange, ERLIC, Size exclusion)
> Polyacrylamide Gel Protein Separation (1-Dimension and 2-Dimensions).
> Staining Polyacrylamide Gels (Coomassie, Silver)
> De-Staining Polyacrylamide Gels (Coomassie, Silver)
> Phospho-Peptide enrichment (TiO2, IMAC)
> Plasma depletion (IgG+Alb, Top20, glyco-enrichment)

Identification techniques
> Protein Identification by nLC-MS/MS
> Protein Identification by 2-Dimension Polyacrylamide Gel Protein Separation
> Identification of Protein post-translational modifications (phosphorylation, glycosilation, methylation, acetylation, ubiquitination)
> Molecular weight determination of peptides and proteins

Quantification services
> Protein quantitation by label-free strategies of complex mixtures
> Protein quantitation by non-isobaric chemical labeling of complex mixtures (SILAC, Dimethyl, AQUA Peptides)
> Protein quantitation by isobaric chemical labeling of complex mixtures (iTRAQ/TMT)

Targeted proteomics
> Selected Reaction Monitoring Assay development
> Protein quantitation by Selected Reaction Monitoring

Experimental design and Data analysis
> Experimental design for mass spectrometric studies
> Mass spectrometric data analysis and interpretation

RESEARCH PROJECTS

The Proteomics Unit promotes internal technology-driven research as an essential task to develop new applications and thus improve the services offered to the users.

In this regard, during 2012 the Proteomics Unit has been developing new-targeted proteomics methods to quantify post-translational modifications, the re-assignment of unidentified good-quality spectra, and the assessment of protein-protein interactions using cross-linking molecules. These research activities have been conducted in the framework of the multi-partner European consortium *PRIME-XS*, which promotes proteomics method development within Europe and grants access to state-of-the-art proteomics technology to the biological and biomedical research community.

Moreover, this last year the Proteomics Unit was also awarded with another multi-partner European project, the *The Human Early-Life Exposome (HELIX)*. This project, that will officially start beginning of 2013, which aims to combine remote sensing/GIS-based spatial methods with omics-based data of selected cohorts to characterize early-life exposure to a wide range of environmental hazards, by integrating the outcome results with data on major child health outcomes.

TRAINING AND DISSEMINATION ACTIVITIES

User training is an essential part of the Proteomics Unit services and it is considered a true investment that leads to mid-term benefits in terms of improved scientific results and higher research quality.

The Proteomics Unit has actively participated in several training and dissemination activities during the year for both users and members of the Unit, which includes teaching in the PhD Course *Genes and Cell Functions* of Universitat Pompeu Fabra, and the organization of the "*CRG Core Facilities Technology Symposium: "Applying proteomics to life sciences: from ions to biology*"

PUBLICATIONS

García-Santamarina S, Boronat S, Calvo IA, Rodríguez-Gabriel M, Ayté J, Molina H, Hidalgo E.
"Is oxidized thioredoxin a major trigger for cysteine oxidation? Clues from a redox proteomics approach."
Antioxid Redox Signal, doi: 10.1089/ars.2012.5037. Epub 2012 Dec 10.

Mancuso F, Bunkenborg J, Wierer M, Molina H.
"Data extraction from proteomics raw data: An evaluation of nine tandem MS tools using a large Orbitrap data set."
J Proteomics, 75(17):5293-303 (2012).

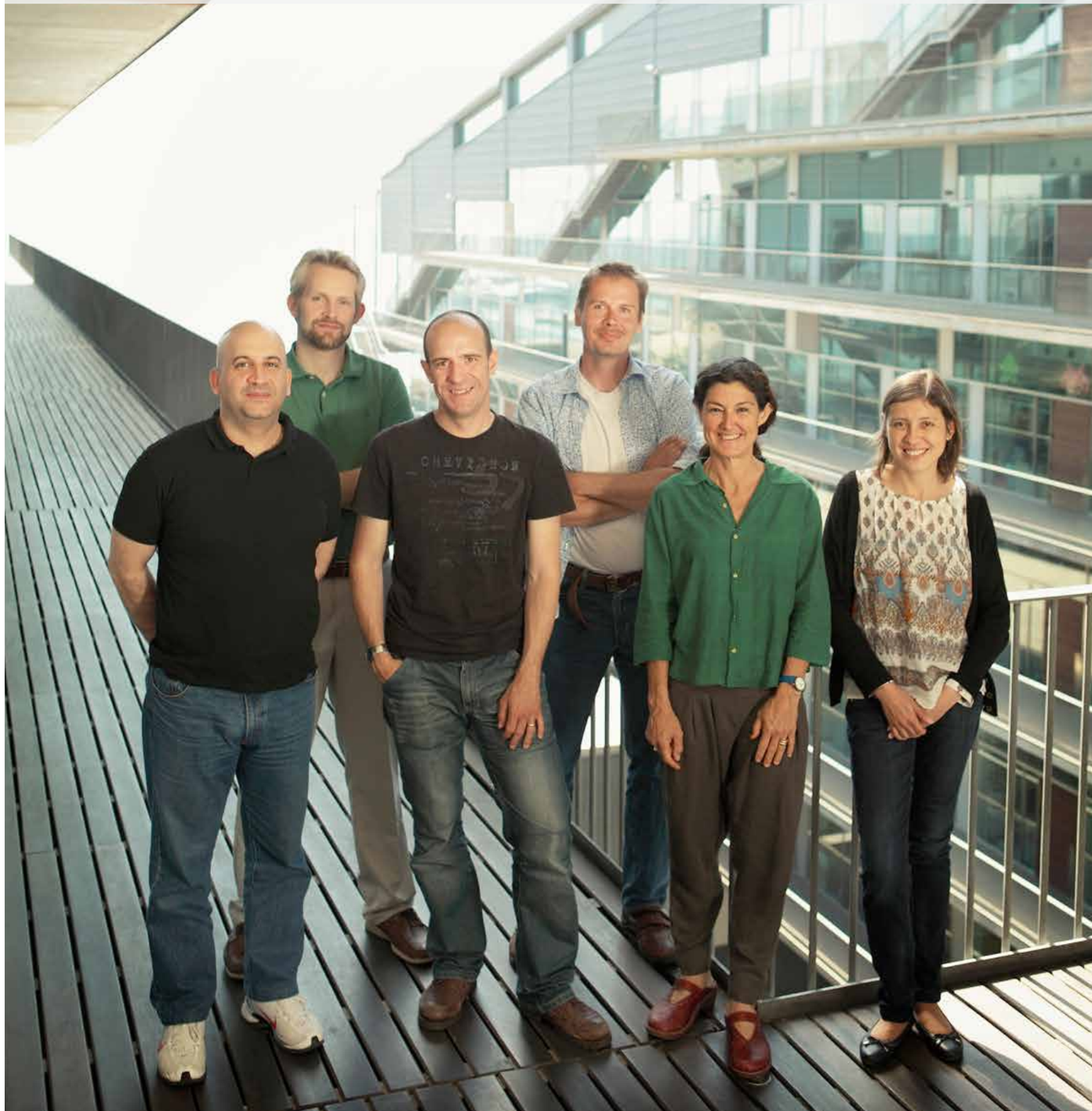
Nikolovski Z, De La Torre C, Chiva C, Borràs E, Andreu D, Ventura R, Segura J.
"Alterations of the erythrocyte membrane proteome and cytoskeleton network during storage - a possible tool to identify autologous blood transfusion."
Drug Test Anal, 4(11):882-90 (2012).

CORE FACILITIES

Unit: 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 2012 to provide instrumentation at the forefront of imaging technology. In autumn the first Leica Ground State Depletion (GSD) microscope was installed in the unit. This instrument is a laser microscope capable of multi-color Total Internal Reflection Fluorescence (TIRF) imaging and is adapted for super-resolution light microscopy approaches that are based on the accurate localization of single fluorophores. These approaches comprise GSD, Photoactivated Localization Microscopy (PALM) and Stochastic Optical Reconstruction Microscopy (STORM).

Also in autumn, the pulsed infrared laser for multi-photon imaging was upgraded with prechirp technology for the compensation of short pulse dispersion during the passage through the optical system. This allows deeper penetration into tissue samples and the collection of signals from deeper levels of thick specimens on the upright and the inverted microscopes connected to the laser.

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 2012 reached 16600 hours corresponding to approximately 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 the earlier years, reflecting the continuing high need for light microscopy by CRG, UPF and PRBB researchers. During the year, 134 users from 24 CRG research groups and 30 users from 16 UPF CEXS groups have used the unit. Additionally the unit was used by 30 users from 15 groups of other PRBB institutes. On average, 89 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 Europe as well in South America (Chile, Brazil).

In September the ALMU organized a three day international workshop on Foerster Resonance Energy Transfer for 25 participants from seven countries. The workshop featured lectures by four invited experts in the field as well as hands-on training sessions and the possibility to discuss in small groups with the speakers. a latest generation Leica SP8 microscope was for the first time ever set up in Spain by Leica for the imaging workshop.

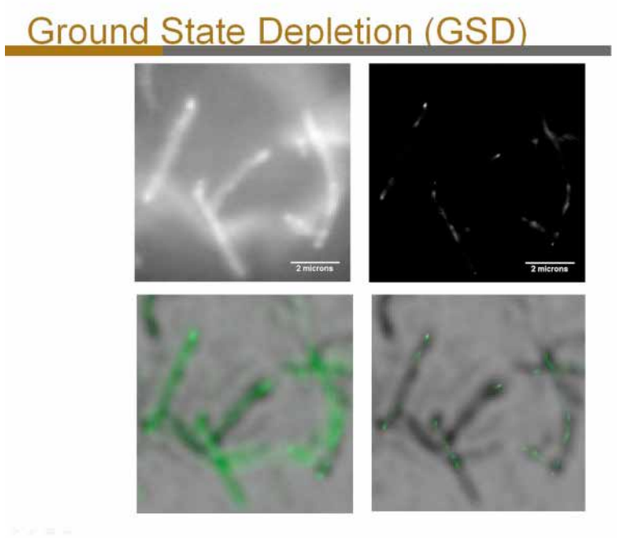
The ALMU is participating in EuroBioImaging, an initiative of the biomedical imaging field that is currently in the preparatory phase of the ESFRI roadmap. Timo Zimmermann is currently the national contact person for Spain for biological imaging in this initiative and involved in the workgroups for general and special access. He is very actively involved in the activities of the Spanish Advanced Light Microscopy Network (Red Española de Microscopía Óptica Avanzada, REMOA), which had its first scientific congress this year at the University of Barcelona.

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.

PUBLICATIONS

Kopp K, Buntru A, Pils S, Zimmermann T, Frank R, Zumbusch A, Hauck CR.
"Grb14 is a negative regulator of CEACAM3-mediated phagocytosis of pathogenic bacteria."
J Biol Chem, 287(46):39158-70 (2012).

Seitz A, Terjung S, Zimmermann T, Pepperkok R.
"Quantifying the influence of yellow fluorescent protein photoconversion on acceptor photobleaching-based fluorescence resonance energy transfer measurements."
J Biomed Opt, 17(1):011010 (2012).



Super-resolution image of a YFP fusion protein of a bacterial transcription factor in Mycobacterium bovis. The resolution is improved beyond the diffraction limit by using Ground State Depletion (GSD) microscopy, a technique in which fluorophores are enriched in transient dark states and are imaged as sparse signals upon their return to the fluorescent state. By sequentially locating the accurate positions of single fluorophores an image that is much better resolved than the diffraction limited fluorescence image can be created. Sample: Carlo Carolis, Luis Serrano Group

CORE FACILITIES

Unit: Biomolecular Screening and Protein Technologies Unit

Unit Structure:
Head of the Unit: Renza Roncarati

Technicians: Miriam Alloza, Carlo Carolis, Raul Gomez, Maria Llamas, Silvia Speroni

SUMMARY

The Biomolecular Screening & Protein Technologies Unit (BMS&PT) provides researchers at CRG/PRBB and external institutions with a high quality technology platform to perform protein production on milligram scale and automated processes in the context of biochemical, cellular and image-based high content screening assays.

Through staff-assisted screening, users will have access to facility equipment including advanced liquid handler workstations (CaliperLS SciClone ALH 3000, TECAN Freedom EVO 200, CaliperLS Zephyr), instruments for biophysical characterization (Isothermal Titration Calorimeter VP-ITC Microcal, Jasco 815 Spectrometer), protein and nucleic acids analyzer (Labchip GXII).

In addition, the Unit is equipped with a human whole genome siRNA library, as well as three chemical compound libraries of small molecules and natural products.

MAJOR PROJECTS AND ACHIEVEMENTS

In 2012 the screening activities of the BMS&PT unit were focused on a project requested by the laboratory of Thomas Graf aiming. The aim of the project is the identification of small organic molecules able to induce human leukemic B cell transdifferentiation into macrophages. Assay optimization was completed, including definition of the image acquisition protocol and analysis, testing of optimal cell density, maximal tolerated concentration of DMSO, plate uniformity, assay variability and stability. A library of approx. 2000 compounds was screened in this cell-based assay.

During the last year, the BMS&PT Unit has completed 120 protein expression assignments according to the requests of 9 CRG groups (Ossowski, Serrano, Malhotra, Carvalho, Cosma, Isalan, Gebauer, Zimmerman and Kondrashov), 1 IMIM group (Clara Montagau) and 1 UPF group (Inaki Ruiz Trillo). The experiments aimed at determining optimal expression conditions and purifying target proteins for antibody production, functional assays and/or biophysical characterization. A new pipeline for protein complex production has been implemented and used for a project of Isabelle Vernos lab.

In collaboration with the Ossowski lab the Unit also performed a screening of antibodies directed against modified histones using the DotBlot technique. The best antibodies were identified and will be employed in further experiments.

The Unit is also involved in the European FP7 project PROSPECTS (Proteomics in Time and Space), in which the Serrano lab is partner. One part of the project aims to quantitatively characterize the Wnt signaling pathway, an important pathway that regulates various developmental and cellular functions. Currently, at least 44 proteins have been shown to be involved in the Wnt signal transduction pathway. The corresponding cDNAs were cloned in suitable expression vectors, the levels of expression were tested in two different bacterial strains comparing different expression conditions and finally 15 proteins of the Wnt signaling pathway were successfully purified.

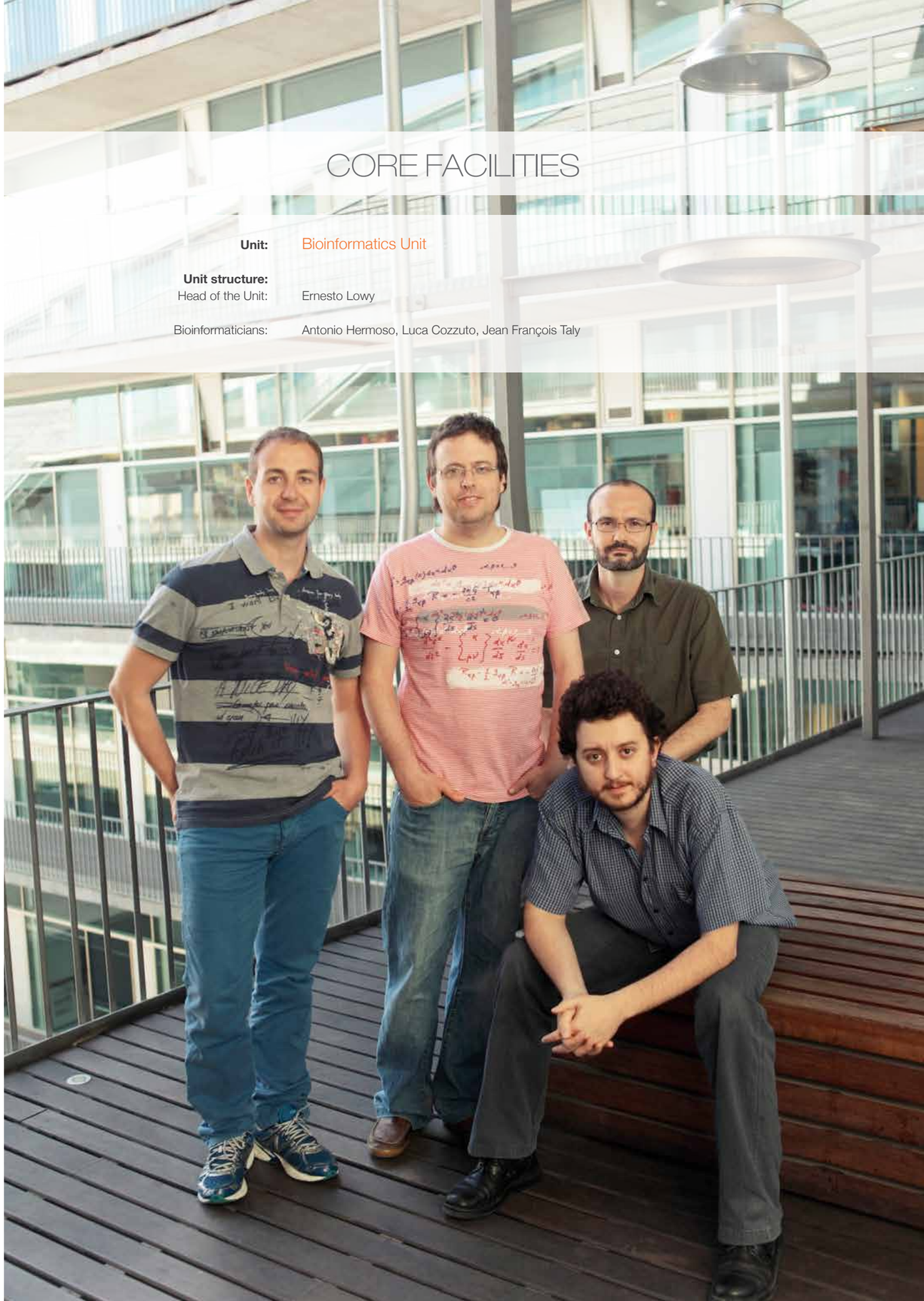
Within the reporting period, the Unit has also worked on the optimization of the pipeline for protein production in *E. coli*, increasing its efficiency and throughput, and significantly reducing the lead-time for the identification of soluble proteins. For instance, the parallel processing of multiple protein targets has been obtained by a reliable and robust 2-step automated purification protocol using the ÄKTAxpress™ liquid chromatography, allowing the production of milligram amounts of up to 8 pure His-tagged recombinant proteins in a few hours. Integration of the Gibson cloning method in our standard cloning pipeline has been concluded. Automation using our robotic platform of this strategy allows us to provide of HTP cloning to the scientific community.

Upon request the Unit has also tested the possibility to implement Synthetic Genetic Array (SGA) analysis in yeast using the available liquid handlers. SGA was initially developed a decade ago and has since been used by many groups. Due to the large number of precise replication steps in SGA analysis, robots equipped with pin-tools are widely used to perform the colony manipulations on microplates at standard densities (96-well, 384-well etc). These machines are quite expensive and there are only a few systems specifically designed for such analysis. In our facility several liquid-handling robots exist, but they are not designed for colony manipulation. Therefore, the challenge was to adapt a general-purpose liquid handler in order to perform SGA assays. Based on the results obtained so far this can be done, although

not with comparable precision as compared to the specifically designed robots. The final validation of the method will be performed in June in collaboration with the interested laboratory of Pedro Carvalho, and according to the agreed timelines.

In the first part of the year, experiments were also performed to test a new class of instruments, which combine label-free detection with high sensitivity for revealing live-cell responses. The underlying read-out technologies are based on impedance and optical-based label-free biosensors. The Roche xCELLigence apparatus and the Perkin Elmer Envision plate reader were evaluated. Both have demonstrated to be valuable instruments and techniques, with the xCELLigence offering the advantage to perform longer (days, weeks) recording under controlled environment conditions.

The Unit also continues offering support and usage of various equipment including the LabChip for nucleic acid and protein analysis, the Tecan Evo 200 and the Zephyr liquid handlers.



CORE FACILITIES

Unit: Bioinformatics Unit

Unit structure:
Head of the Unit: Ernesto Lowy

Bioinformaticians: Antonio Hermoso, Luca Cozzuto, Jean François Taly

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

During 2012 the unit has gone through several changes in its personnel composition. These can be summarized as follows:

> Dr Francesco Mancuso has left us and has joined the Proteomics core facility
> In July 2012, Dr Jean Francois Taly joined the Bioinformatics unit.

The activity of the unit during 2012 focused basically on two different types of services, services for the internal CRG users and services offered to external and internal users. Within the framework of the internal services just mention that we have invested time and effort in order to establish mirrors (local copies) of different scientific databases that are accessible to the users of the CRG Linux network file system. Currently, we are mirroring the main publicly available scientific databases such as the NCBI, Ensembl, UniProt, UCSC, etc. Furthermore, we have a public registry in the form of wiki describing the mirrors available, maintained releases and the deletion policy.

In addition to this and as a part of the services provided to the CRG and external users, we have continued receiving a lot of requests concerning the analysis of Next Generation Sequencing data from different types of experiments, such as ChIP-seq, RNA-seq, De-novo genome assembly, resequencing for variation detection, etc… In order to cope with the high-demand of this type of requests we have improved the pipelines we already had and have automatized some of the steps in these pipelines. Additionally, we have established a pipeline for the detection of SNPs and Indels from resequencing experiments, this pipeline is based in the Genome Analysis Toolkit (GATK [1]), which is a quite established framework for variation detection that is being used in a wide range of projects such as the 1000 genomes project [http://www.1000genomes.org/] and the The Cancer Genome Atlas [http://cancergenome.nih.gov/]. Another interesting pipeline we have been testing and finally implementing during 2012 is the Maker pipeline for the automatic annotation of protein coding genes [2]. We have implemented this pipeline in close collaboration with the groups of Roderic Guigo and Yogi Yaeger and it is being used for the automatic gene annotation of two species of flies (*Clogmia albipunctata and Megaselia abdita*).

In 2012, the unit has been also involved in 2 consortia for the sequencing and annotation of two genomes, the international effort for the sequencing of the *Phaseolus Vulgaris* genome (common bean; http://www.genoma-cyted.org/organizacion.html) and the Spanish collaborative effort to produce a draft assembly and annotation of the *Lynx pardinus* species (Iberian Lynx; http://www.lynxgenomics.eu/research/9/genome-lynx-the-unveiling). Within the framework of these projects, the unit is in charge of the functional characterization of the protein coding genes in both species and it is also in charge of the annotation of the small non-coding RNA genes.

Another substantial technology development effort the unit has been carrying out consisted on the local implementation of Galaxy (https://main.g2.bx.psu.edu/). This platform will enable the scientists without specific IT skills to perform advanced data analysis tasks. Additionally, this platform will, whenever possible, substitute the licensed software available at the CRG by open source alternatives that through the use of the Galaxy portal will be made available to the entire CRG community. In addition, this Galaxy implementation will contain some of the workflows and pipelines used in the facility and will incorporate the algorithms and tools most widely used by the scientific community, always prioritizing the ones developed at the PRBB.

The local instance of Galaxy will be available at the beginning of 2013 and will contain some of the basic bioinformatics tools that are included in the Galaxy tool shed.

As an additional mission of the unit, we have dedicated part of our resources to the implementation of scientific databases and web services for the community; I would like to stress the publication of an update [3] of the Plant Resistance Gene database (Prgdb, http://prgdb.crg.eu/wiki/Main_Page). This new release of the database incorporates a web interface based on Semantic MediaWiki technologies [http://en.wikipedia.org/wiki/Semantic_MediaWiki] enabling the scientific community to contribute with their expertise in the annotation of disease resistance genes in plants.

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. These LIMS have continued being used by the Biomolecular Screening & Protein Technologies and by the Bioinformatics unit. Furthermore, by the end of 2012 we launched in production another instance in the CRG/UPF Proteomics Unit, which has been actively used these last months by more than 30 users. As part of its development, we created different extensions for automatizing the creation of several components at once and for allowing the access to the filesystems where the instruments store experiment data.

Finally, it does worth to mention that during 2012 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 2012 the unit organized a multisession hands-on Perl programming course and in November we organized a technology symposium open to everybody named “Cloud computing: democratization of high-throughput data analysis” where we had the opportunity to learn about this exciting field from the invited speakers.

PUBLICATIONS

Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, González VM, Hénaff E, Câmara F, Cozzuto L, Lowy E, Alioto T, Capella-Gutiérrez S, Blanca J, Cañizares J, Ziarsolo P, Gonzalez-Ibeas D, Rodríguez-Moreno L, Droege M, Du L, Alvarez-Tejado M, Lorente-Galdos B, Melé M, Yang L, Weng Y, Navarro A, Marques-Bonet T, Aranda MA, Nuez F, Picó B, Gabaldón T, Roma G, Guigó R, Casacuberta JM, Arús P, Puigdomènech P.
“The genome of melon (Cucumis melo L.).”
Proc Natl Acad Sci U S A. 2012 Jul 17;109(29):11872-7. doi: 10.1073/pnas.1205415109. Epub 2012 Jul 2.

Morey L, Pascual G, Cozzuto L, Roma G, Wutz A, Benitah SA, Di Croce L.
“Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells.”
Cell Stem Cell. 2012 Jan 6;10(1):47-62. doi: 10.1016/j.stem.2011.12.006.

Doles J, Storer M, Cozzuto L, Roma G, Keyes WM.
“Age-associated inflammation inhibits epidermal stem cell function.”
Genes Dev. 2012 Oct 1;26(19):2144-53. doi: 10.1101/gad.192294.112. Epub 2012 Sep 12.

Morey L, Aloia L, Cozzuto L, Benitah SA, Di Croce L.
“RYBP and Cbx7 Define Specific Biological Functions of Polycomb Complexes in Mouse Embryonic Stem Cells.”
Cell Rep, doi: 10.1016/j.celrep.2012.11.026. Epub 2012 Dec 27.

Sanseverino W, Hermoso A, D’Alessandro R, Vlasova A, Andolfo G, Frusciante L, Lowy E, Roma G, Ercolano MR.
“PRGdb 2.0: towards a community-based database model for the analysis of R-genes in plants.”
Nucleic Acids Res, doi: 10.1093/nar/gks1183. Epub 2012 Nov 17.

CORE FACILITIES

Unit: CRG/UPF Flow Cytometry (FACS) Unit

Unit structure:
Head of the Unit: Òscar Fornas (UPF)

Operators: Erika Ramírez (CRG), Sabrina Bascones (IMIM) (until August 2012)
Eva Julià (IMIM) (since September 2012)

SUMMARY

The fast evolution of new applications available in Flow Cytometry (FCM) has multiplied the amount of research projects that, over the past decade, take advantage of this technology. Fifteen years ago, technical limitations or complexity precluded the use, for most investigators, of more than five colours simultaneously on flow phenotyping. However, the last flow decade has been characterized by the outburst of multicolour flow cytometry analysis. Nowadays, even though the technical complexity persists, state-of-the-art flow cytometers easily perform simultaneous cytometric analysis using ten colours.

In fact, the actual Multicolour Flow Cytometry Technology, making use of multi-laser excitation lines on state-of-the-arts instrumentation, more than 20 different colours simultaneously serve an extremely wide range of applications. That is the reason why demand of this technology is daily increasing. Most important applications of flow cytometry include, among others, the study of cell surface receptors, nuclear and cytoplasmic antigens, DNA content, enzyme activity, cell integrity and membrane permeability, and calcium flows. All of them can be performed in the facility.

The mission of the Flow Cytometry Unit is to provide PRBB researchers with technical expertise and training to access the state-of-the-art instrumentation, as well as technical and scientific advice to develop efficient and reliable flow cytometric assays with the highest quality control standards and productivity. The Unit supports the use of a wide range of flow cytometry applications and new ones are developed/implemented responding to the facility needs or upon user demand.

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 24%, from 5242 in 2011 to 6515 in 2012. In total, the facility had 165 users in 2012.

The only change during 2012 has been on the second operator, who was replaced by Eva Julià since Sabrina Bascones left from the facility at the end of August. She has moved to IMIM as a PhD student. Eva Julià started on September the 1st. She was working as a flow cytometry operator at IDIBELL for the last three years acquiring enough flow cytometry experience to start operating all our instrumentation since the first day.

EQUIPMENT

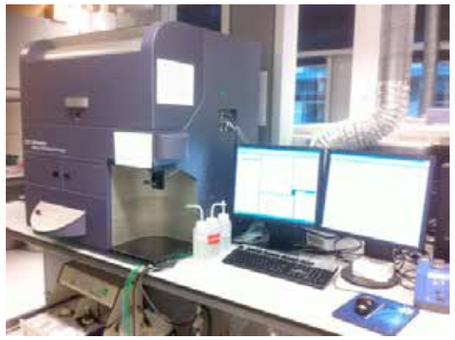
The following table shows the evolution of instrumentation during 2012. Two instruments were replaced for new ones in order to cover the demand on high-end instrumentation. Both were purchased by CRG.

The new cytometers are BD LSRFortessa analyser and BD Influx cell sorter and with these instruments the suboptimal waiting list (1-2 weeks for the analyzer and 6 weeks for the cell sorting) to use high-end instruments has gone. On the analyzers usage currently there is no waiting list, one can use required instrument when needed without planning it some days in advance. The waiting time on the analyzers usage was solved immediately after its installation. For the cell sorting, the worst situation in December was that the waiting list is not higher than one week; this is a reasonable situation.

Instrument	Laser (nm)	Detectors	Situation
BD FACScan analyser	488	**FSC/SSC + 3PMT**	Active
BD FACScalibur analyser	488 633	**FSC/SSC + 4PMT**	Active
BD FACSCanto analyser	488 633	**FSC/SSC + 6PMT**	Active
BD LSR	325 488 633	**FSC/SSC + 6PMT**	Replaced April 2012
BD LSR II analyser	325 407 488 633	**FSC/SSC + 10PMT**	Active
BD LSRFortessa analyser	407 488 561 633	**FSC/SSC + 14PMT**	Active (April 2012)
BD FACSAria II SORP cell sorter	325 407 488 561 633	**FSC/SSC + 18PMT**	Active
BD FACSDiVa cell sorter	325 488 633	**FSC/SSC + 8PMT**	Replaced April 2012
BD Influx cell sorter	325 457 488 561 633	**FSC/SSC + 14PMT**	Active (April 2012)

NEW INSTRUMENTATION

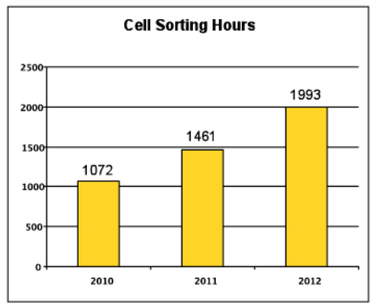
BD LSRFortessa: This cytometer is equipped with 4 laser lines and 14 fluorescence detectors and is one of the most advanced analyzer.

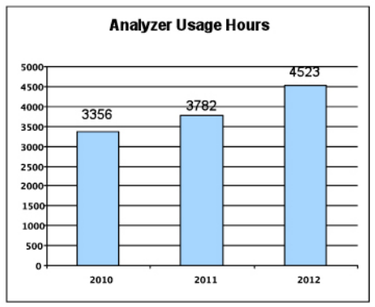


BD Influx: This cell sorter equipped with 5 laser lines and up to 14 fluorescence detectors can cover the same applications than the other one (FACSAria II) and with its special architecture improves some experiments (better cell viability and recovery) and also can perform advanced and special high resolution applications. Some users are taking advantages using this new instrument.

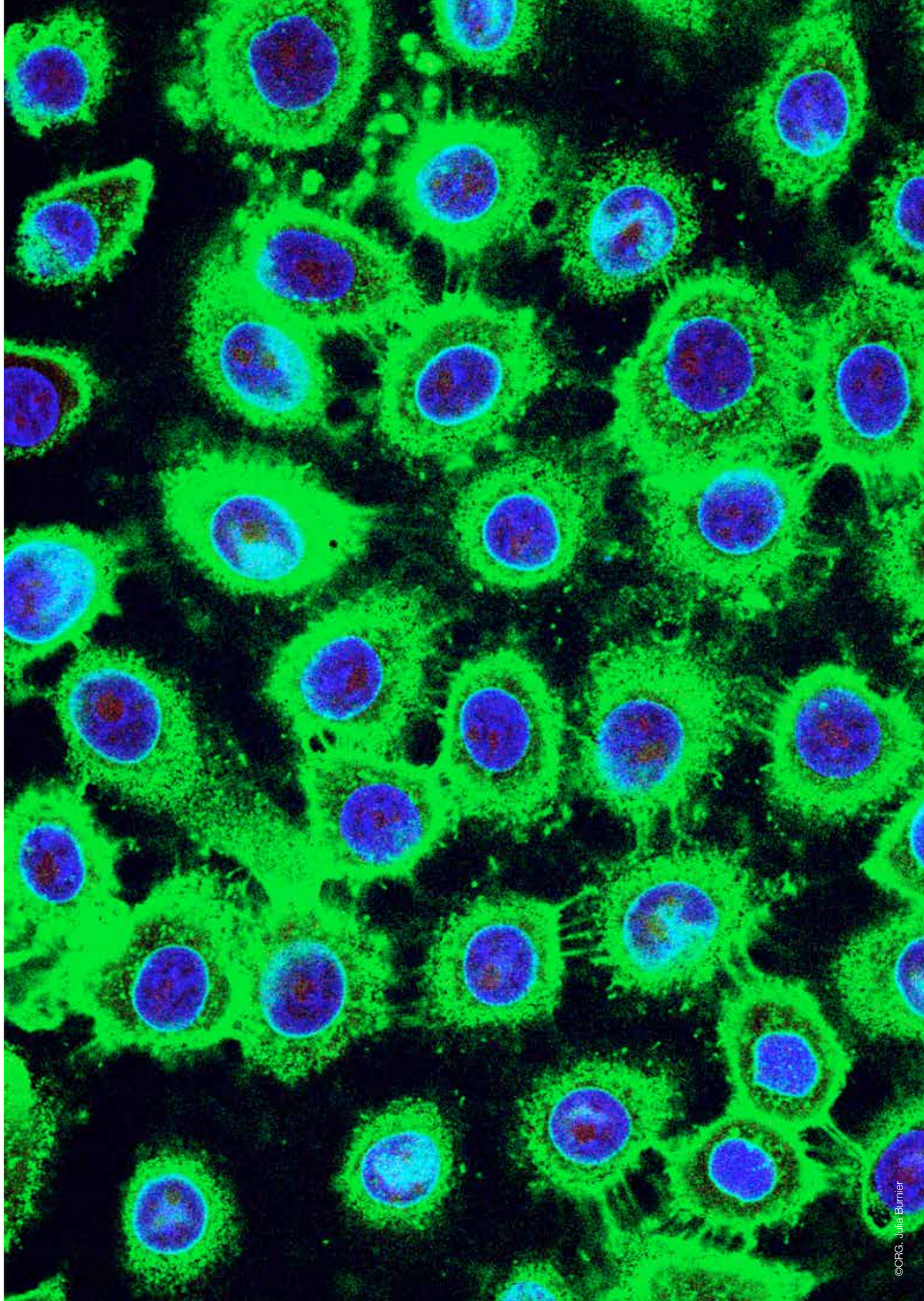
ACTIVITY

The following graphs illustrate the increasing activity for the offered services.





During 2012 the Head of the Unit has participated in master teaching courses of the UPF as well as in flow cytometry courses at other institutions in Barcelona.



CORE FACILITIES

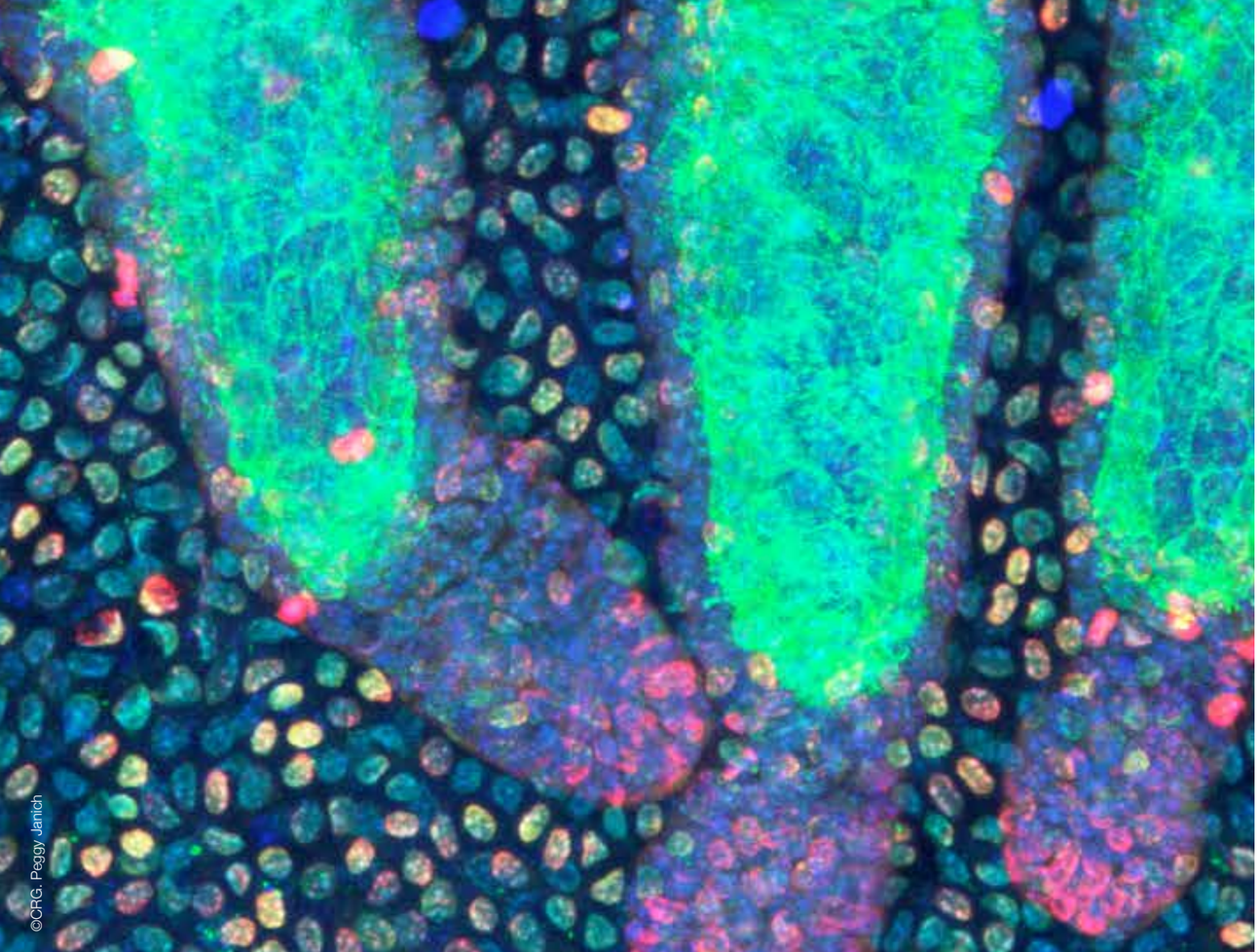
In-House Services: Histology Service

Service structure:

Service Manager: Salvador Aznar Benitah PhD; ICREA Researcher;
PI of the Epithelial and Homeostasis Laboratory;
Gene Regulation, Stem Cells & Cancer Programme (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



©CRG. Peggy Janich

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

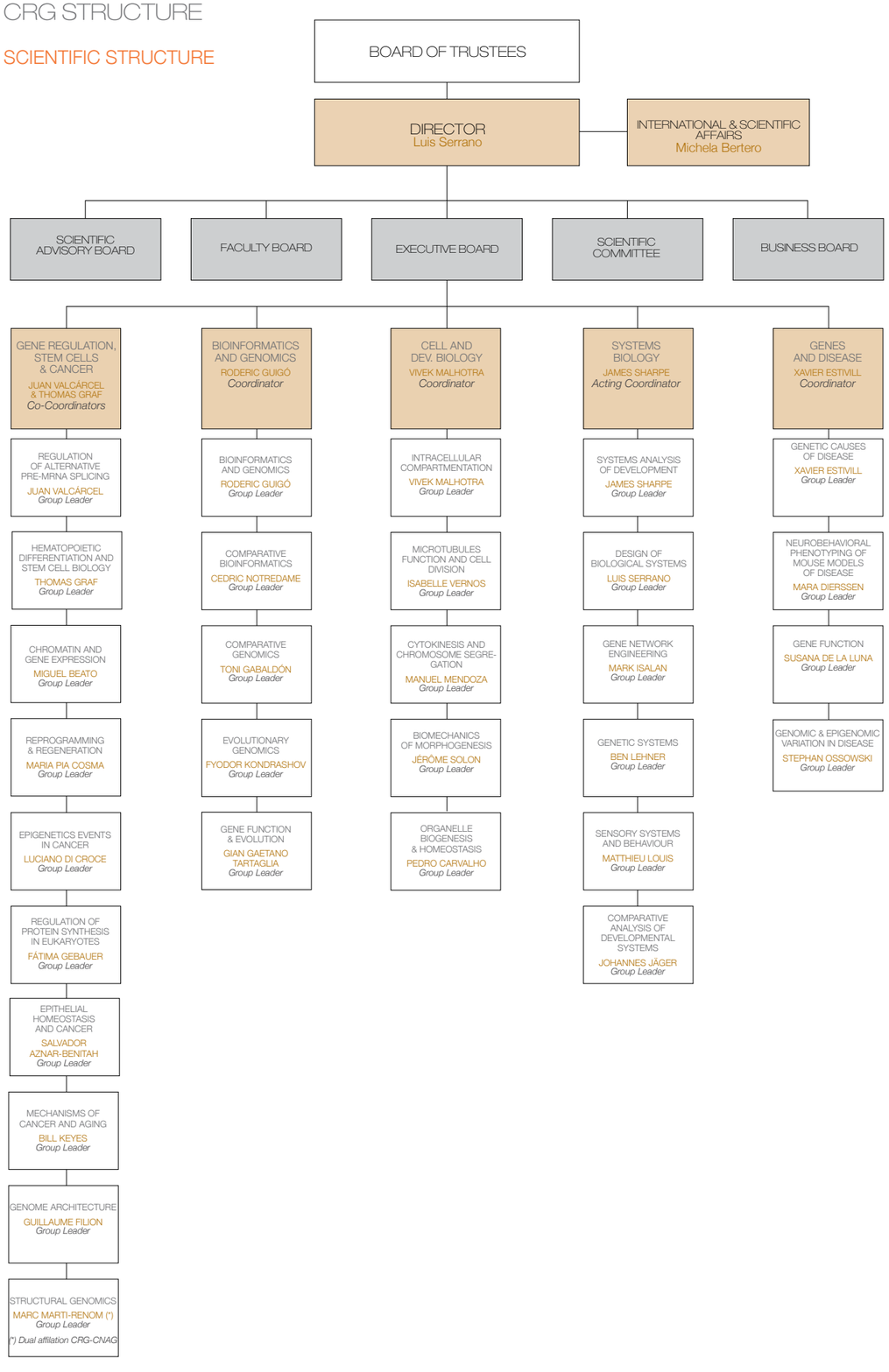
- > Two cryostats: Sectioning of frozen tissue blocks.
- > Two microtomes: Sectioning of paraffin blocks.
- > Two vibratomes: Sectioning of fresh tissue.
- > Two histological water baths: For paraffin embedding and sectioning.
- > Paraffin dispenser, hot and cold plate: Paraffin embedding.
- > Autostainer (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 Precission 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 allow 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.

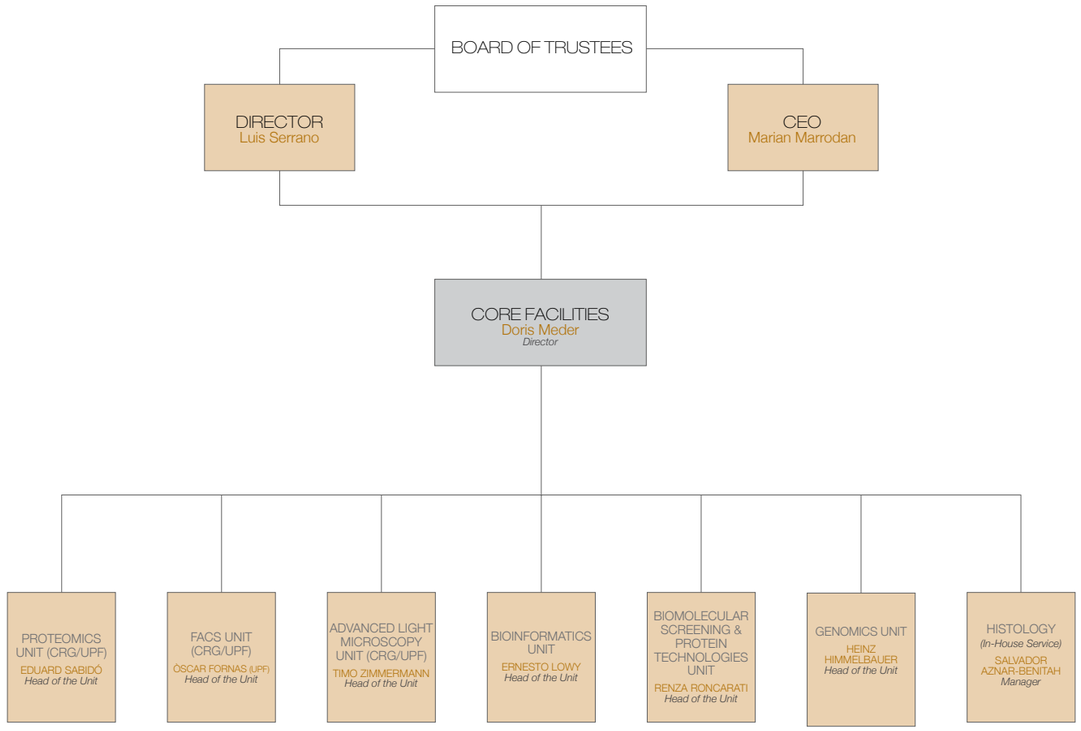
Appendix 1

CRG STRUCTURE

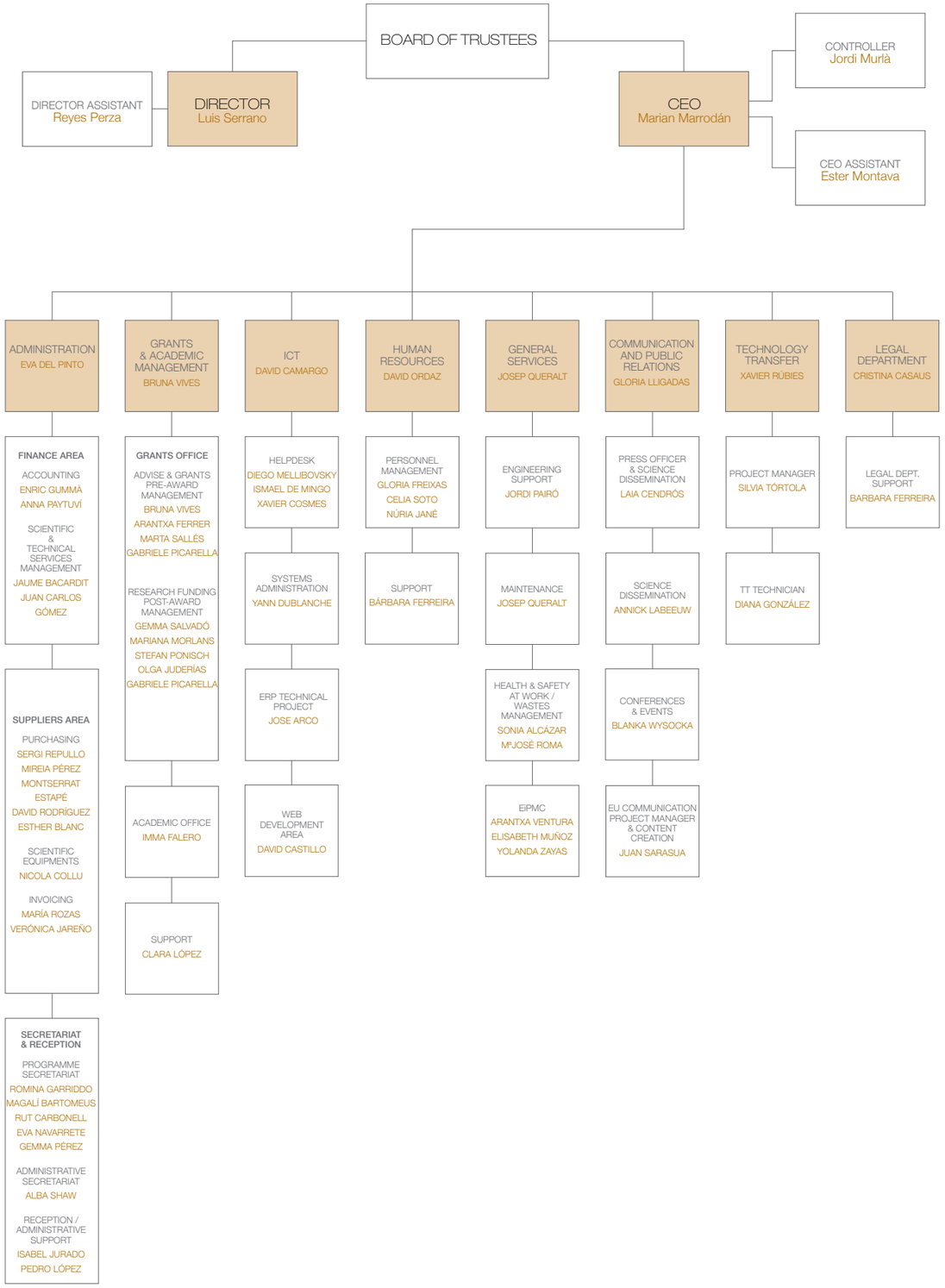
SCIENTIFIC STRUCTURE



CORE FACILITIES STRUCTURE



MANAGEMENT & RESEARCH SUPPORT STRUCTURE



SCIENTIFIC ADVISORY BOARD (SAB)

Chairman

Prof. Iain Mattaj
European Molecular Biology Laboratory (EMBL),
Heidelberg, Germany

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. Tariq Enver
UCL Cancer Institute,
London, United Kingdom

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. 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,
GP Pharm

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

PRESS CLIPPING

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

Members of the Board of Trustees:

Generalitat de Catalunya

Institut adscrit a la
UNIVERSITAT POMPEU FABRA