CRG
Centre for Genomic Regulation

Centre for Genomic Regulation
Annual Report 2010

Centres de recerca de Catalunya

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

Texts and graphics: CRG Researchers, Department of Communication and Public Relations

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

Photography: Ivan Martí

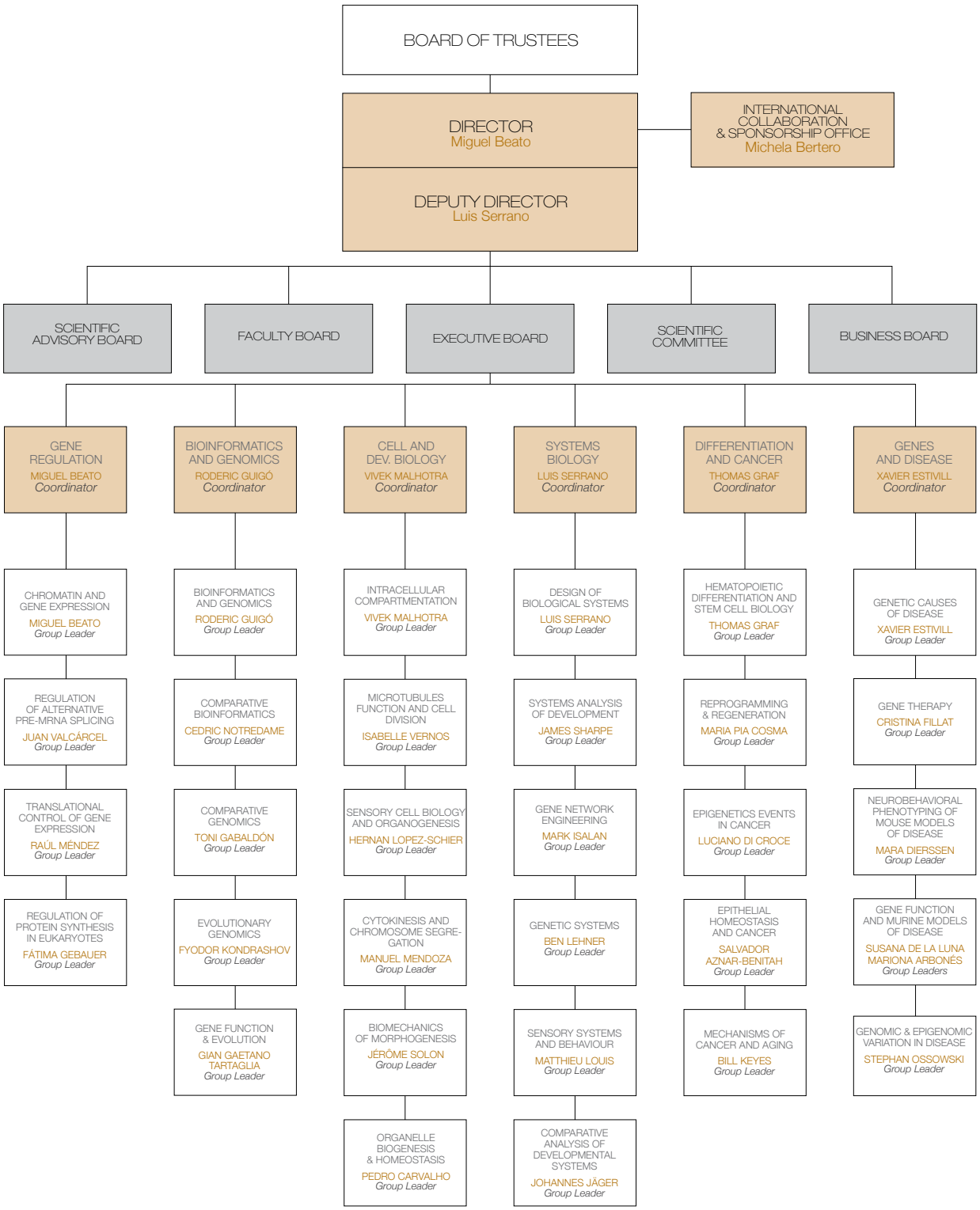
Printing: Novoprint, S.A.

Legal deposit: B-24966-2011

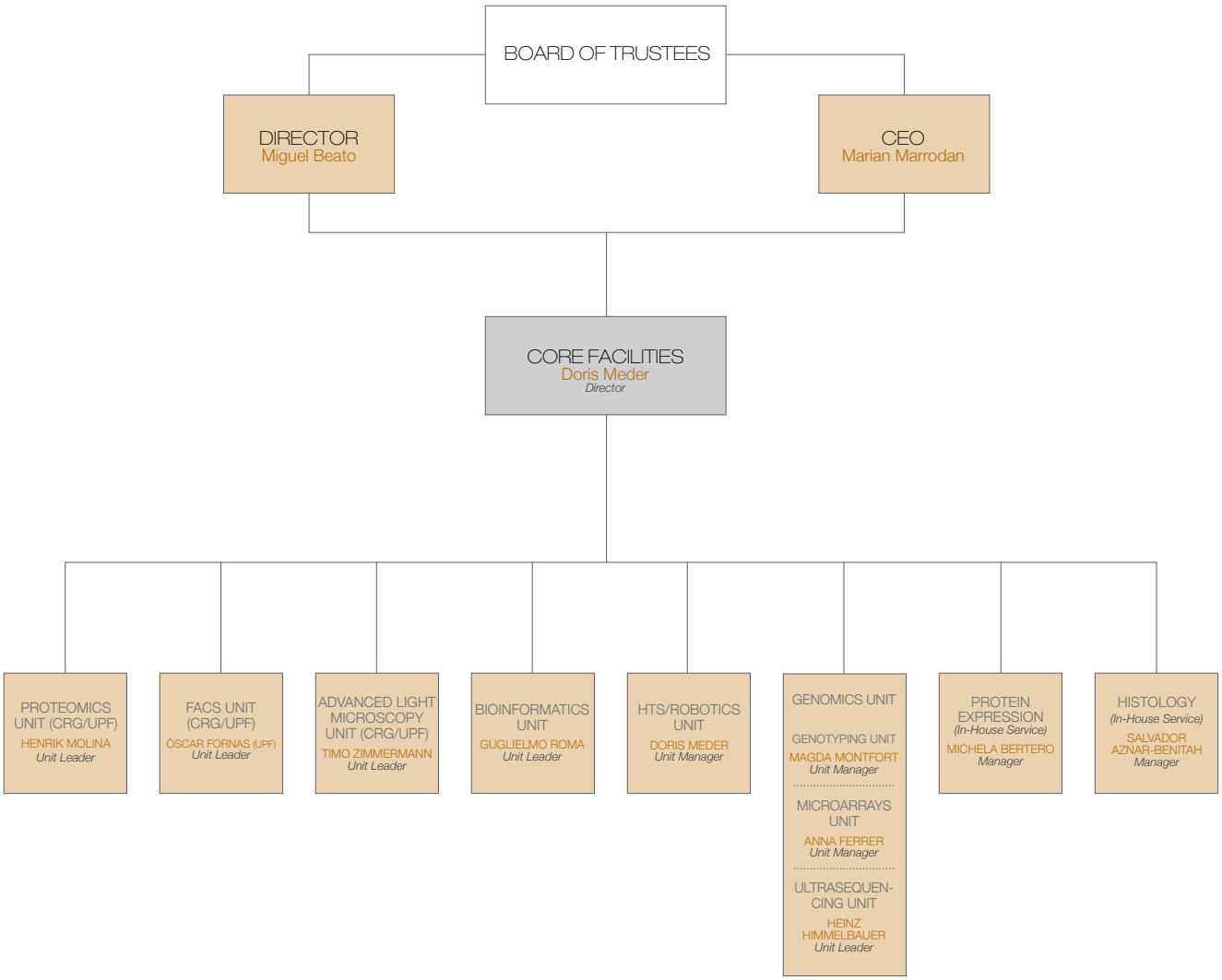

CRG
Centre
for Genomic
Regulation

CONTENTS

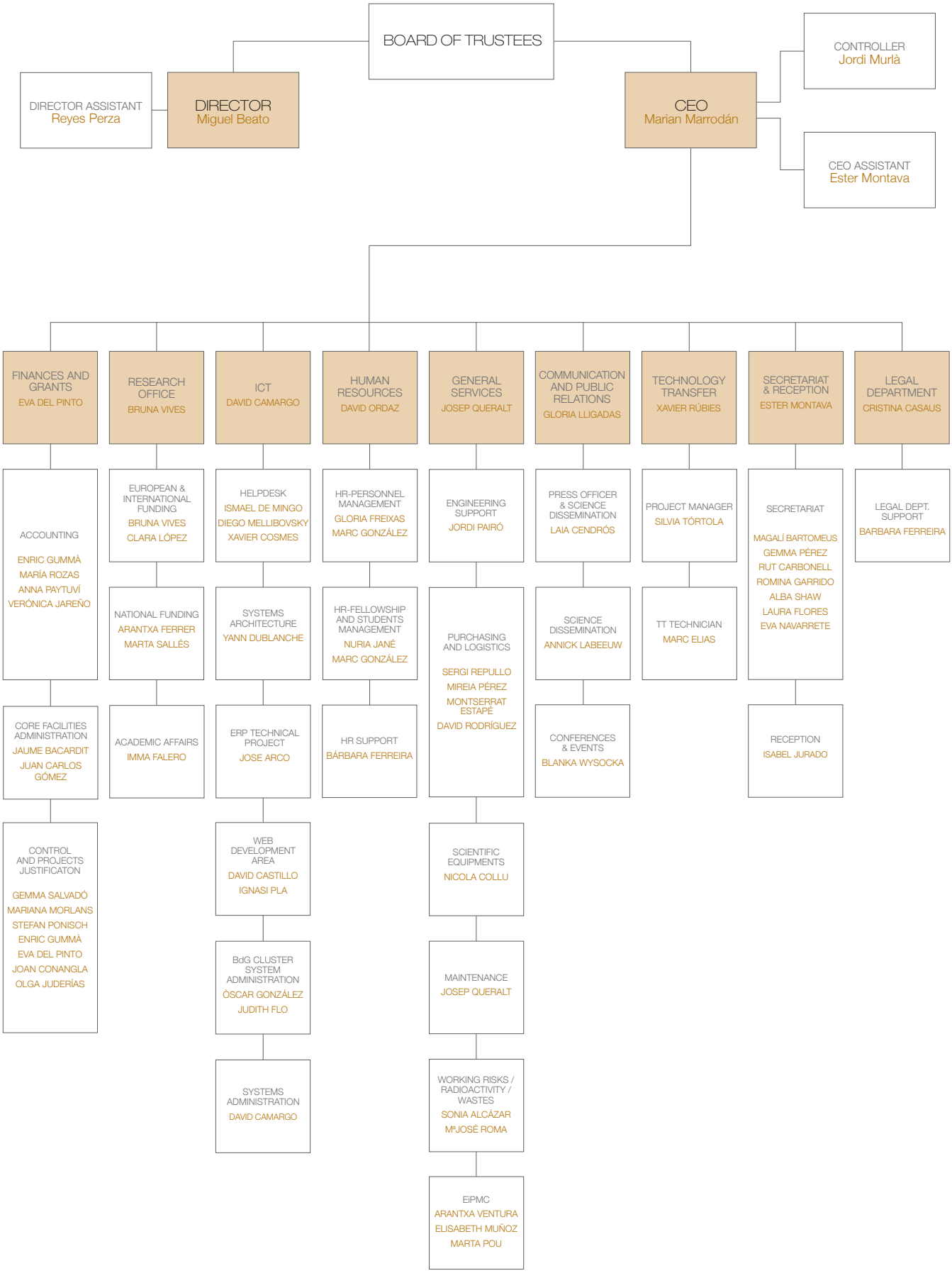
SCIENTIFIC STRUCTURE


BOARD OF TRUSTEES
DIRECTOR
Miguel Beato
INTERNATIONAL COLLABORATION & SPONSORSHIP OFFICE
Michela Bertero
DEPUTY DIRECTOR
Luis Serrano
SCIENTIFIC ADVISORY BOARD
FACULTY BOARD
EXECUTIVE BOARD
SCIENTIFIC COMMITTEE
BUSINESS BOARD
GENE REGULATION
MIGUEL BEATO
Coordinator
BIOINFORMATICS AND GENOMICS
RODERIC GUIGÓ
Coordinator
CELL AND DEV. BIOLOGY
VIVEK MALHOTRA
Coordinator
SYSTEMS BIOLOGY
LUIS SERRANO
Coordinator
DIFFERENTIATION AND CANCER
THOMAS GRAF
Coordinator
GENES AND DISEASE
XAVIER ESTIVILL
Coordinator
CHROMATIN AND GENE EXPRESSION
MIGUEL BEATO
Group Leader
REGULATION OF ALTERNATIVE PRE-MRNA SPLICING
JUAN VALCÁRCEL
Group Leader
TRANSLATIONAL CONTROL OF GENE EXPRESSION
RAÚL MÉNDEZ
Group Leader
REGULATION OF PROTEIN SYNTHESIS IN EUKARYOTES
FÁTIMA GEBAUER
Group Leader
BIOINFORMATICS AND GENOMICS
RODERIC GUIGÓ
Group Leader
COMPARATIVE BIOINFORMATICS
CEDRIC NOTREDAME
Group Leader
COMPARATIVE GENOMICS
TONI GABALDÓN
Group Leader
EVOLUTIONARY GENOMICS
FYODOR KONDRASHOV
Group Leader
GENE FUNCTION & EVOLUTION
GIAN GAETANO TARTAGLIA
Group Leader
INTRACELLULAR COMPARTMENTATION
VIVEK MALHOTRA
Group Leader
MICROTUBULES FUNCTION AND CELL DIVISION
ISABELLE VERNOS
Group Leader
SENSORY CELL BIOLOGY AND ORGANOGENESIS
HERNAN LOPEZ-SCHIER
Group Leader
CYTOKINESIS AND CHROMOSOME SEGREGATION
MANUEL MENDOZA
Group Leader
BIOMECHANICS OF MORPHOGENESIS
JÉRÔME SOLON
Group Leader
ORGANELLE BIOGENESIS & HOMEOSTASIS
PEDRO CARVALHO
Group Leader
DESIGN OF BIOLOGICAL SYSTEMS
LUIS SERRANO
Group Leader
SYSTEMS ANALYSIS OF DEVELOPMENT
JAMES SHARPE
Group Leader
GENE NETWORK ENGINEERING
MARK ISALAN
Group Leader
GENETIC SYSTEMS
BEN LEHNER
Group Leader
SENSORY SYSTEMS AND BEHAVIOUR
MATTHIEU LOUIS
Group Leader
COMPARATIVE ANALYSIS OF DEVELOPMENTAL SYSTEMS
JOHANNES JÄGER
Group Leader
HEMATOPOIETIC DIFFERENTIATION AND STEM CELL BIOLOGY
THOMAS GRAF
Group Leader
REPROGRAMMING & REGENERATION
MARIA PIA COSMA
Group Leader
EPIGENETICS EVENTS IN CANCER
LUCIANO DI CROCE
Group Leader
EPITHELIAL HOMEOSTASIS AND CANCER
SALVADOR AZNAR-BENITAH
Group Leader
MECHANISMS OF CANCER AND AGING
BILL KEYES
Group Leader
GENETIC CAUSES OF DISEASE
XAVIER ESTIVILL
Group Leader
GENE THERAPY
CRISTINA FILLAT
Group Leader
NEUROBEHAVIORAL PHENOTYPING OF MOUSE MODELS OF DISEASE
MARA DIERSSEN
Group Leader
GENE FUNCTION AND MURINE MODELS OF DISEASE
SUSANA DE LA LUNA
MARIONA ARBONÉS
Group Leaders
GENOMIC & EPIGENOMIC VARIATION IN DISEASE
STEPHAN OSSOWSKI
Group Leader

CORE FACILITIES STRUCTURE


BOARD OF TRUSTEES
DIRECTOR
Miguel Beato
CEO
Marian Marrodan
CORE FACILITIES
Doris Meder
Director
PROTEOMICS UNIT (CRG/UPF)
HENRIK MOLINA
Unit Leader
FACS UNIT (CRG/UPF)
ÒSCAR FORNAS (UPF)
Unit Leader
ADVANCED LIGHT MICROSCOPY UNIT (CRG/UPF)
TIMO ZIMMERMANN
Unit Leader
BIOINFORMATICS UNIT
GUGLIELMO ROMA
Unit Leader
HTS/ROBOTICS UNIT
DORIS MEDER
Unit Manager
GENOMICS UNIT
GENOTYPING UNIT
MAGDA MONTFORT
Unit Manager
MICROARRAYS UNIT
ANNA FERRER
Unit Manager
ULTRASEQUEN-CING UNIT
HEINZ HIMMELBAUER
Unit Leader
PROTEIN EXPRESSION
(In-House Service)
MICHELA BERTERO
Manager
HISTOLOGY
(In-House Service)
SALVADOR AZNAR-BENITAH
Manager

MANAGEMENT STRUCTURE


BOARD OF TRUSTEES
DIRECTOR ASSISTANT
Reyes Perza
DIRECTOR
Miguel Beato
CEO
Marian Marrodán
CONTROLLER
Jordi Murlà
CEO ASSISTANT
Ester Montava
FINANCES AND GRANTS
EVA DEL PINTO
RESEARCH OFFICE
BRUNA VIVES
ICT
DAVID CAMARGO
HUMAN RESOURCES
DAVID ORDAZ
GENERAL SERVICES
JOSEP QUERALT
COMMUNICATION AND PUBLIC RELATIONS
GLORIA LLIGADAS
TECHNOLOGY TRANSFER
XAVIER RÚBIES
SECRETARIAT & RECEPTION
ESTER MONTAVA
LEGAL DEPARTMENT
CRISTINA CASAUS
ACCOUNTING
ENRIC GUMMÀ
MARÍA ROZAS
ANNA PAYTUVÍ
VERÓNICA JAREÑO
EUROPEAN & INTERNATIONAL FUNDING
BRUNA VIVES
CLARA LÓPEZ
HELPDESK
ISMAEL DE MINGO
DIEGO MELLIBOVSKY
XAVIER COSMES
HR-PERSONNEL MANAGEMENT
GLORIA FREIXAS
MARC GONZÁLEZ
ENGINEERING SUPPORT
JORDI PAIRÓ
PRESS OFFICER & SCIENCE DISSEMINATION
LAIA CENDRÓS
PROJECT MANAGER
SILVIA TÓRTOLA
SECRETARIAT
MAGALÍ BARTOMEUS
GEMMA PÉREZ
RUT CARBONELL
ROMINA GARRIDO
ALBA SHAW
LAURA FLORES
EVA NAVARRETE
LEGAL DEPT. SUPPORT
BARBARA FERREIRA
NATIONAL FUNDING
ARANTXA FERRER
MARTA SALLÉS
SYSTEMS ARCHITECTURE
YANN DUBLANCHE
HR-FELLOWSHIP AND STUDENTS MANAGEMENT
NURIA JANÉ
MARC GONZÁLEZ
PURCHASING AND LOGISTICS
SERGI REPULLO
MIREIA PÉREZ
MONTSERRAT ESTAPÉ
DAVID RODRÍGUEZ
SCIENCE DISSEMINATION
ANNICK LABEEUW
TT TECHNICIAN
MARC ELIAS
CORE FACILITIES ADMINISTRATION
JAUME BACARDIT
JUAN CARLOS GÓMEZ
ACADEMIC AFFAIRS
IMMA FALERO
ERP TECHNICAL PROJECT
JOSE ARCO
HR SUPPORT
BÁRBARA FERREIRA
CONFERENCES & EVENTS
BLANKA WYSOCKA
RECEPTION
ISABEL JURADO
CONTROL AND PROJECTS JUSTIFICATON
GEMMA SALVADÓ
MARIANA MORLANS
STEFAN PONISCH
ENRIC GUMMÀ
EVA DEL PINTO
JOAN CONANGLA
OLGA JUDERÍAS
WEB DEVELOPMENT AREA
DAVID CASTILLO
IGNASI PLA
SCIENTIFIC EQUIPMENTS
NICOLA COLLU
BdG CLUSTER SYSTEM ADMINISTRATION
ÒSCAR GONZÁLEZ
JUDITH FLO
MAINTENANCE
JOSEP QUERALT
SYSTEMS ADMINISTRATION
DAVID CAMARGO
WORKING RISKS / RADIOACTIVITY / WASTES
SONIA ALCÁZAR
MªJOSÉ ROMA
EIPMC
ARANTXA VENTURA
ELISABETH MUÑOZ
MARTA POU

SCIENTIFIC ADVISORY BOARD (SAB)

Chairman

Dr. Kai Simons
Max Planck Institute of Molecular Cell Biology and Genetics,
Dresden, Germany

Members

Dr. Stylianos Emmanuel Antonarakis
Medical Genetics, University of Geneva,
Geneva, Switzerland

Dr. Michael Ashburner
European Bioinformatics Institute-EMBL Outstation, Department of Genetics,
University of Cambridge,
Cambridge, United Kingdom

Dr. Peer Bork
Structural & Computational Biology Unit, European Molecular Biology Laboratory (EMBL),
Heidelberg, Germany

Dr. Tim Hunt
Cancer Research UK,
Hearts, United Kingdom

Dr. Walter Kolch
Systems Biology Ireland & Conway Institute, University College,
Dublin, Ireland

Dr. Tony Kouzarides
The Wellcome Trust/Cancer Research UK Gurdon Institute,
Cambridge, United Kingdom

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

Dr. Joan Modolell
Center of Molecular Biology “Severo Ochoa”,
CSIC & Autonomous University of Madrid,
Madrid, Spain

Dr. Wolf Reik
Babraham Institute Cambridge, Department of Physiology, Development & Neuroscience,
University of Cambridge, United Kingdom

Dr. Veronica van Heyningen
Medical Research Council, Human Genetics Unit, Western General Hospital,
Edinburgh, United Kingdom

Dr. Marc Vidal
Dana Farber Cancer Institute,
Boston, USA

BUSINESS BOARD

President

Dr. Josep M. Taboada
Medical Director,
Sanofi-Aventis

Vice-president

Dr. Antonio Parente
President,
Grup Lipotec

Members

Dr. Pere Berga
R+D Management Director,
Almirall Prodesfarma, S.A.

Dr. Clara Campàs
Executive Vice-President,
Advancell Therapeutics

Dr. Antoni Esteve
President,
Esteve

Dr. Josep Prous, Jr.
Executive Vice President,
Prous Science

Sr. Jordi Ramentol
General Manager,
Grupo Ferrer Internacional, S.A.

Mr. Josep Lluís Sanfeliu
Co-founder partner,
Ysios Capital Partners

Sr. Fernando Turró
General Manager,
Contratas y Obras

YEAR RETROSPECT

by the Director of the CRG: Miguel Beato

During 2010 the financial crisis continued to influence the development of the CRG. Our original plan was to reach almost completion of the number of groups by the end of 2010 or beginning of 2011, but we had to postpone this plan and have limited our recruitments considerably, in order to guarantee that the activity of the CRG scientists remained relatively unaffected. In April Pia Cosma, from the TIGEM in Naples, started at the CRG in replacement for Pura Muñoz, who left the Differentiation and Cancer programme in 2009. Pia is a world expert in cell reprogramming and regeneration via stem cell fusion and had received an ERC Starting Grant and an ICREA Research Professor position. In the Bioinformatics and Genomics programme a new junior group leader, Gian Gaetano Tartaglia, from the University of Cambridge, in United Kingdom, joined the CRG in May to work in gene function and evolution.

In 2010 two other scientists of the CRG received ERC grants: Vivek Malhotra got an Advanced Grant and Manuel Mendoza got a Starting Grant. This brings the total number of ERC grants at the CRG to 7 (2 Advanced and 5 Starting), of which 6 were given to non-Spanish scientists, a sign for the capacity to compete for the best colleagues in the international arena.

In April 14, the CRG received the Narcis Monturiol Plate in recognition of its contribution to biomedical research, genomics and proteomics. The jury underlined the relevance of having established a new model of running scientific research aiming at excellence, mobility and internationality. In June 21 the neurobiologist, Mara Dierssen, was awarded the International Sisley-Jerome Lejeune Prize at the Museum for Medical History in Paris.

In July the science journalist Luis Angel Fernández finished a book entitled *The CRG: Immersion in the Century of Genomics*, which describes the life of the CRG scientists, their ways of approaching scientific questions, as well as their insertion in the local and international research and social environment. The book, published in Catalan and Spanish, was targeted to the educate citizens but had a very positive reception by the scientific community.

In February the CRG Core Facilities were presented to the scientific community and to the media in a set of talks followed by a guided visit. During the year we had other meetings focused on specific Core Facilities, including a Technology Symposium on Super-Resolution Microscopy, co-organized with ICFO. In June the CRG postdoctoral community organized its first Symposium, following the model of the PhD Students Symposium, which this year took place in November. In July the CRG, in collaboration with Biocat, organized the ENCODE Symposium, to discuss the advances and the remaining challenges of this key project for genomics. The first meeting of this community outside the US was co-organized by Roderic Guigó and was followed by a large audience. Three other meetings were organized by the CRG in October: the Leica International Course on Advanced Confocal Microscopy, the ESF-EMBO Symposium on Functional Neurobiology in Minibrains, and our IX Annual Symposium on Medical Genome Sequencing, which was a great success within the scientific community and in the media.

In the spring we interviewed 30 candidates for the 10 positions of the third call of the International PhD Programme generously financed by the “la Caixa” Foundation. The positions were filled with excellent candidates from many different nationalities, who joined the program in the fall.

A very important event this year was the evaluation of the two programmes Cell & Developmental Biology and Systems Biology, including the EMBL/CRG Systems Biology Research Unit, by a review panel composed of members of our SAB and ad hoc reviewers. The review took place in November 3 and 4 and was very successful. All group leaders and the programme coordinators received excellent reviews by the panel members, who recommended continuation and further support for the two programmes and the EMBL-CRG Partnership.

Finally another very important process was started this year, namely the search for a new CRG director. The CRG rules specify that the director cannot stay for longer than two 5-years terms. Therefore, the position was announced in *Nature* and candidates sent their applications to the chairman of our SAB, Kai Simons. The SAB evaluated the applications and proposed Luis Serrano as the most suitable candidate. The CRG Faculty accepted this proposal and Luis will take over as soon as the Board of Trustees approves the SAB choice.

At the end of 2010 there were 249 scientists (38 group leaders and unit heads, 107 staff scientists and postdoctoral fellows, 104 PhD students), 103 technicians and 59 administration personnel working at the CRG. Over 60% of the group leaders, postdocs and graduate students are foreigners.

Although the economic crisis will last for another couple of years, we are confident that the quality and strength of both our scientific community and our management team will help us to continue our work without serious damage. This optimism is supported by the return to the Catalan government of the person who created the CRG in his goal to bring international visibility to Catalan science. His nomination in December was a good Christmas present and the best way to close a complicated year.

GENE REGULATION

Coordinator: Miguel Beato

At the beginning of 2010, the programme lost Josep Vilardell, one of the founder Group Leaders, who after 9 years at the CRG earned an ICREA Research Professor position and moved to the Institute of Biomedical Research (IRB). Josep worked at defining the mechanism of splicing regulation by RPL32 in yeast and generated important results published in 2006 and 2009 in Molecular Cell. He described the characterization of the splicing-repressed complex containing L30 transcripts ("inhibited complex, IC") characterized in *in vitro* splicing and ChIP assays.

Due to cuts in our budget, we could not fill the position of senior scientist previously occupied by Ramin Shiekhattar, nor the free junior group leader position, although the positions were announced and we got excellent candidates.

During 2010 the different groups of the programme have incremented the use of genome wide studies, in particular Illumina/Solexa sequencing for studying chromatin structure, epigenetic marks, RNA transcripts and alternative splicing. A "Chromatin Group" has been created including members of three different programmes that meet once a month to report on the progress of various epigenetic projects. Moreover, groups working on splicing and chromatin in the programme have started a collaboration that includes co-mentoring of a graduate student. A sign of these multiple collaborations is the highlighted paper summarized at the end of this report in which two groups from the programme and a bioinformatics group are co-authors.

The structure of the programme at the end of 2010 was:

4 Research groups:

> Chromatin and Gene Expression (Miguel Beato, coordinator, group of the director)
> Regulation of Alternative Pre-mRNA Splicing (Juan Valcárcel)
> Translational Control of Gene Expression (Raúl Méndez)
> Regulation of Protein Synthesis in Eukaryotes (Fátima Gebauer)

Programme Secretary:

Imma Falero / Laura Flores

Numbers:

4 group leaders, 2 staff scientists, 14 postdocs, 15 students,
8 technicians and support personnel.

GENE REGULATION

Group: Chromatin and Gene Expression

Group structure:

Group Leader: Miguel Beato

Postdoctoral Fellows: Alessandra Ciociola, François Le Dily, Guillermo Vicent (Staff Scientist), Roni Wright, Marija Kundakovic (till May)

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

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

Visitors: Alejandro La Greca, Buenos Aires (November/December)

SUMMARY

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

RESEARCH PROJECTS

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

L. Gaveglia, R. Wright, F. Le Dily, A. Pohl, G. Castellano (collaboration with Christophoros Nikolau and Roderic Guigo from the Bioinformatics and Genomic programme)

We have compared the complete sets of nucleosome positions for the budding yeast (*S. cerevisiae*) in different growth conditions and found that less than 10% of the experimentally defined nucleosome positions in two published datasets were consistently positioned. This subset of well-positioned nucleosomes, when compared to the bulk, was shown to have particular properties at both sequence and structural levels. Consistently positioned nucleosomes were also shown to occur preferentially in pairs of dinucleosomes and to be surprisingly less conserved when compared to their adjacent nucleosome-free linkers (Nikolaou et al 2010).

We have performed genome wide nucleosome mapping by massive sequencing of mononucleosomal DNA (300 million unique reads per time point), from T47D-MTVL breast cancer cells treated with hormone for different time periods. The results revealed that the nucleosome occupancy near the transcription start site is significantly higher in genes that are down-regulated following hormone induction, in contrast to a lower nucleosomal occupancy in genes up-regulated by hormone (Gaveglia et al, unpublished).

2. Signalling by progesterone to chromatin via kinase cascades

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

Progesterone controls proliferation and gene expression in breast cancer cells via transient activation of the Src/Ras/Erk pathways mediated by an interaction of PR with estrogen receptor alpha, ERα. Activated Erk in the cell nucleus phosphorylates PR and Msk1, which phosphorylates histone H3 at serine 10 contributing to the activation of target genes (Vicent et al Mol Cell 2006).

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. In parallel we have used ChIP-seq to identify all genomic binding sites for PR in breast cancer cells treated with hormone for different times (Ballare et al, unpublished). Prior to hormone treatment sites where PR will bind exhibit high nucleosome occupancy and these nucleosomes become sensitive to MNase upon addition of hormone, mainly due to the hormone-dependent removal of histone H1 and dimmers of histone H2A and H2B. These results are now being integrated in a dynamic network, which should help identifying relevant nodes connecting various signalling pathways with regulation of different gene cohorts.

3. Regulation of MMTV transcription in the chromatin context

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

The group has studied the structural changes accompanying activation of MMTV promoter chromatin and how they are catalyzed. Already 1 min after progestin treatment the hNURF complex and the ASCOM complex are recruited to the MMTV promoter and required for gene activation. The MLL2/3 histone methyltransferases as part of the ASCOM complex, trimethylate histone H3 at lysine 4 anchoring the NURF complex via an interaction with the BPTF subunit (Vicent et al, in press). Simultaneously the histone demethylase KDM5B/PLU1 is displaced from the promoter. These three complexes along with the

CyclinA/CDK2 kinase complex are required for the rapid opening of chromatin by displacement of phosphorylated histone H1 (Vicent et al, in press). IP-seq and expression arrays show that H1 displacement is required for hormone induction of most hormone target genes, some of them involved in cell proliferation (Vicent et al, in press).

Within 5 minutes of progestin addition, a ternary complex of activated PR and two activated kinases, pErk1/2 and pMsk1, is recruited to the promoter and phosphorylates histone H3 at S10 (Vicent et al Mol Cell 2006). This leads to dissociation of a repressive complex containing HP1g as a prerequisite for the recruitment of BAF, an ATP-dependent chromatin-remodelling complex. BAF binding is favoured by the PCAF mediated acetylation of H3 at Lysine 14 (Vicent et al PloS Genet 2009). BAF remodelling leads to displacement of H2/H2B dimmers and facilitates NF1 access and binding of additional PR to the central HREs on the remodelled nucleosome (Vicent et al, 2010).

Thus within 5-10 min of hormone action two subsequent cycles of chromatin remodeling lead to displacement first of histone H1 and second of histones H2A/H2B dimers (Figure 1). At least three protein kinases, Erk, Msk1, CDK2, a histone methyl transferase, MLL2/3, a histone demethylase, KDM5B/PLU1, a histone acetyl transferase, PCAF, and two ATP-dependent remodelling complexes, NURF and BAF, are involved in these initial steps of hormone action. We have indications that other enzymes, including arginine methyl transferases, Poly (ADP-ribose) polymerases, histone deacetylases, histone ubiquitinating enzymes, and the proteasome also contribute to the initial remodeling of chromatin.

On the MMTV promoter progesterone induction is mediated by the reciprocal synergism between PR and the ubiquitous transcription factor NF1. The central HREs 2 and 3 are not needed for ATP-dependent H2A/H2B displacement or NF1 binding but are critical for full PR binding and MMTV transactivation (Vicent et al 2010). We found that NF1 binding to the MMTV promoter on a H3/H4 histone tetramer particle exposes the central HREs and facilitates their binding by PR, suggesting a possible mechanism for the reciprocal synergism between PR and NF1 (Vicent et al 2010).

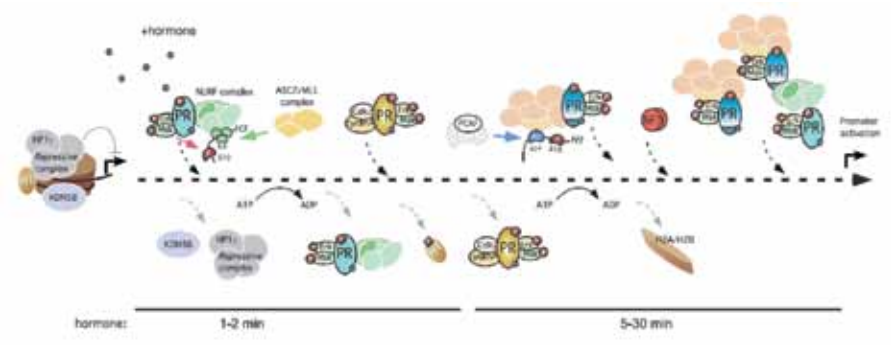


Figure 1
The MMTV promoter nucleosome B is maintained in a repressive state by binding of a repressive complex containing, among other factors HP1g and KDM5D. A complex pPR/pErk/pMsk1 including NURF and likely ASCOM is recruited to the promoter. The repressive complex containing Hp1g is displaced upon Msk1 mediated phosphorylation of serine 10 of histone H3 (Vicent et al Mol Cell 2006). The MLL2/3 HMTs methylate lysine 4 of H3 and simultaneously the HDM KDM5B/PLU1 is displaced from the promoter.
The enhanced H3K4me3 mark stabilizes NURF binding. NURF facilitates the binding of a complex of PR with CyclinA/CDK2, which phosphorylates histone H1 and promotes its displacement. Subsequently, a complex of PR with PCAF and BAF, binds to the exposed HRE1 on the MMTV nucleosome and acetylates H3 at K14, stabilizing binding of BAF. The BAF complex catalyzes the ejection of H2A/H2B dimmers, enabling NF1 binding (Vicent et al PloS Genet 2009). Binding of NF1 stabilizes the open nucleosome conformation and facilitates binding of further PR molecules and BAF remodelling complexes to the internal HREs.

4. Role of linker histone H1 subtypes in chromatin and transcription

R. Wright, F. Le Dily, A. Pohl

We found that histone H1 enhances the activation of the MMTV promoter by PR and NF1 (Koop et al EMBO J 2003) and we are now studying the role of various H1 subtypes and their phosphorylation by Cdk2 on the remodelling and transcription of MMTV chromatin. Histone H1.4 is phosphorylated by cyclin E-Cdk2, facilitating the decompaction of chromatin and transcriptional activation. A mutant of H1.4 incapable of phosphorylation by Cdk2 compacts minichromosomes similarly to wild type H1.4. Purified SWI/SNF is capable of remodeling chromatin reconstituted with wild type or mutated H1.4, but the remodeled chromatin is less efficiently transcribed in the absence of H1 phosphorylation (Clausell et al unpublished).

5. Role of steroid hormones in breast cancer and endometrial physiology

A. Ciociola, A.Mai, D. Reyes, M. Wierer (collaboration with Belen Mlñana, CRG, and the Departments of Pathology and Oncology of the Hospital del Mar).

We have studied the gene networks regulated by estrogens and progestins in primary tumour material from107 breast cancer samples from patients attending the Hospital del Mar in Barcelona, using a cDNA array with 800 selected genes. We are now using the information collected from cultured breast cancer cell lines treated with inhibitors of various kinase-signaling cascades (see point 2) to classify the tumors according to the different signaling pathways altered in each case. This set of data will be correlated with the clinical data and the evolution of the tumors in search of a gene signature distinguishing various types of tumors.

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

PUBLICATIONS

Calvo V, Beato M.
"BRCA1 counteracts progesterone action in breast cancer cells by ubiquitination leading to receptor degradation and epigenetic silencing of target genes."
Cancer Res, in press.

Nikolaou C, Althammer S, Beato M, Guigo R.
"Structural constrains revealed in consistent nucleosome positions in the genome of S. cerevisiae."
BMC Epigenet & Chromatin, 3(1):20 (2010).

Vicent GP, Nacht AS, Font-Mateu J, Castellano G, Gaveglia L, Ballare C, Beato M.
"Four enzymatic activities cooperate to displace histone H1 during the first minutes of hormonal gene activation."
Genes & Dev, in press.

Vicent GP, Zaurin R, Nacht AS, Font-Mateu J, Le Dily F, Beato M.
"NF1 synergizes with progesterone receptor on the MMTV promoter wrapped around a histone H3/H4 tetramer by facilitating access to the central hormone responsive elements."
J Biol Chem, 285:2622-2631 (2010).

Vicent GP, Nacht AS, Zaurin R, Ballare C, Clausell J, Beato M.
"Role of kinases and chromatin remodeling in progesterone signaling to chromatin".
Mol Endocrinol, 24:2088-2098 (2010).

Vallejo G, Maschi D, Mestre-Citrinovitz AC, Aiba K, Maronna R, Yohai V, Ko MS, Beato M, Saragüeta P.
"Changes in global gene expression during in vitro decidualization of rat endometrial stromal cells."
J Cell Physiol, 222(1):127-37 (2010).

GENE REGULATION

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

Juan Valcárcel is an ICREA Research Professor.

Group Structure:

Group Leader: Juan Valcárcel

Staff scientist: Sophie Bonnal

Postdoctoral Fellows: Sergio Barberán, Elias Bechara, Panagiotis Papasaikas (since November 2010), Maria Paola Paronetto, Joao Tavanez

Students: Anna Corrionero, Camilla Ianonne, Juan Ramón Tejedor

Technicians: Cecilia Albor (until September 2010), Belén Miñana, Anna Ribó

SUMMARY

The focus of our research is the process of alternative pre-mRNA splicing, which allows the production of multiple mRNAs from a single gene, greatly expanding the information content of the genomes of multicellular organisms. During 2010, we have made progress in understanding the mechanisms.

RESEARCH PROJECTS

1. Mechanism of splicing inhibition in a genetic defect leading to autoimmune disease

Pre-mRNA splice site sequences provide unequivocal information for intron excision, yet consensus splice sites are highly degenerate in higher eukaryotes, suggesting that multiple sequence arrangements can accurately signal exon/intron boundaries. We have found that the 3' splice site associated with the alternatively spliced exon 6 of the Fas receptor CD95 displays strict sequence requirements and that a mutation that disrupts this particular sequence arrangement leads to constitutive exon 6 skipping in a patient suffering from Autoimmune Lymphoproliferative Syndrome (ALPS). We find that the balance between exon inclusion and skipping is exquisitely sensitive to single nucleotide variations in the sequence RCAG/G (where R represents purine and / indicates the intron/exon boundary) or in the uridine content of the upstream polypyrimidine (Py)-tract. Biochemical experiments revealed that the ALPS patient mutation reduces U2 snRNP recruitment to the 3' splice site region and that this effect cannot be explained by decreased interaction with the U2 snRNP Auxiliary Factor U2AF, whose 65 and 35 KDa subunits recognize the Py-tract and 3' splice site AG, respectively. The results indicate that the strict architecture of Fas intron 5 3' splice site region is tuned to regulate alternative exon inclusion through modulation of U2 snRNP assembly after U2AF binding.

2. The antitumor drug spliceostain A alters the fidelity of 3' splice site recognition and induces alternative splicing changes in cell cycle control genes

Spliceostatin A (SSA) is a stabilized derivative of a Pseudomonas bacterial fermentation product that displays potent anti-proliferative and anti-tumor activities in cancer cells and animal models. The drug inhibits pre-mRNA splicing *in vitro* and *in vivo* and binds SF3b, a protein subcomplex of U2 small nuclear ribonucleoprotein (snRNP), which is essential for recognition of the pre-mRNA branch point. We have found that SSA prevents interaction of SF3b 155 kDa subunit with the pre-mRNA, concomitant with non-productive recruitment of U2 snRNP to sequences 5' of the branch point. Differences in base pairing potential with U2 snRNA in this region lead to different sensitivity of 3' splice sites to SSA and to SSA-induced changes in alternative splicing. Indeed, rather than general splicing inhibition, splicing-sensitive microarray analyses reveal specific alternative splicing changes induced by the drug, which significantly overlap with those induced by knock down of SF3b 155. These changes lead to down-regulation of genes important for cell division, including cyclin A2 and Aurora A kinase, thus providing an explanation for the anti-proliferative effects of SSA. Our results reveal a mechanism that prevents non-productive base pairing interactions in the spliceosome and highlight the regulatory and cancer therapeutic potential of perturbing the fidelity of splice site recognition (Figure 1).

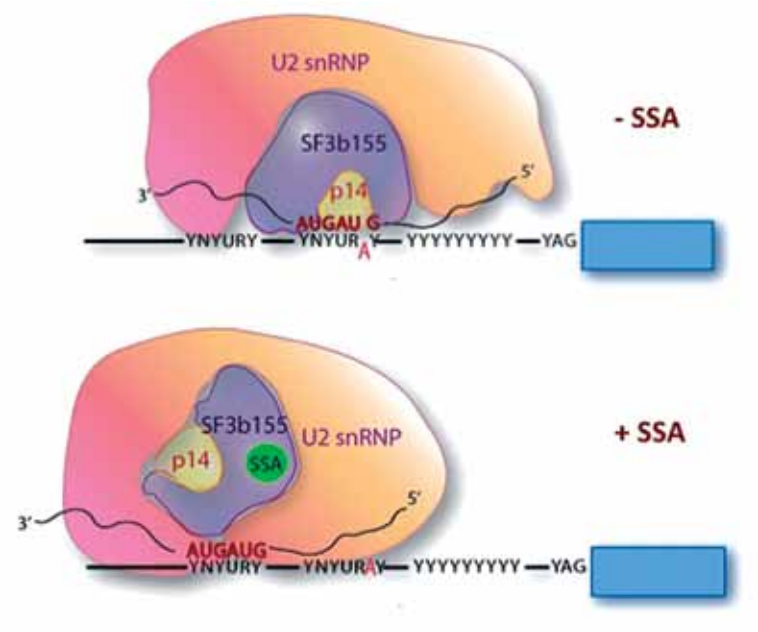


Figure 1. Model for splicing inhibition by the anti-tumor drug Spliceostatin A. *U2 snRNP assembles on the 3' splice site region of pre-mRNAs, a process that involves RNA-RNA contacts between U2 snRNA and the branch point region in the pre-mRNA, as well as RNA-protein contacts between the pre-mRNA and protein components of the snRNP, including SF3b155. In the presence of spliceostatin A (SSA) SF3b 155 fails to contact the pre-mRNA and U2 snRNA establishes base pairing interactions with pre-mRNA sequences 5' of the branch point. This leads to non-productive U2 snRNP recruitment to these decoy sites (which do not contain a bulged-out adenosine branch site and therefore cannot undergo splicing catalysis). Differential sensitivity of 3' splice sites to SSA depends on the base pairing potential with U2 snRNA of sequences 5' of the branch point. This leads to changes in alternative splicing induced by the SSA, which affect genes important for cell cycle control, thus providing an explanation for the anti-proliferative effects of the drug.*

3. Distinct regulatory programs establish widespread sex-specific alternative splicing in *Drosophila melanogaster*.

In *Drosophila melanogaster*, female-specific expression of Sex-lethal (SXL) and Transformer (TRA) proteins controls sex-specific alternative splicing and/or translation of a handful of regulatory genes responsible for sexual differentiation and behavior. We and others have detected widespread sex-biased alternative splicing in fruitflies. Bioinformatic analysis of SXL/TRA binding sites, experimental analysis of sex-specific splicing in S2 and Kc cells lines and of the effects of SXL knock down in Kc cells indicate that SXL-dependent and SXL-independent regulatory mechanisms coexist within the same cell. Additional determinants of sex-specific splicing can be provided by sex-specific differences in the expression of RNA binding proteins, including Hrp40/Squid, which we have found to be differentially expressed in male and female flies due to sex-specific alternative splicing of the gene. The significant overlap between sex-regulated alternative splicing changes and those induced by knock down of hrp40/squid and the presence of related sequence motifs enriched near subsets of Hrp40/Squid-regulated and sex-regulated splice sites indicate that this protein contributes to sex-specific splicing regulation. A significant fraction of sex-specific splicing differences are absent in germline-less *tudor* mutant flies. Intriguingly, these include alternative splicing events that are differentially spliced in tissues distant from the germline. Collectively, our results reveal that distinct genetic programs control widespread sex-specific splicing in *Drosophila melanogaster*.

PUBLICATIONS

Tejedor JR and Valcárcel J.
"Gene regulation: Breaking the second genetic code."
Nature, 465:45-46 (2010).

Tavanez J and Valcárcel J.
"A splicing mastermind for EMT."
EMBO Journal, 29:3217-3218 (2010).

Corrionero A, Raker V, Izquierdo JM and Valcárcel J.
"Strict 3' splice site sequence requirements for U2 snRNP recruitment after U2AF binding underlie a genetic defect leading to autoimmune disease."
RNA, in press.

Hartmann B, Castelo R, Miñana B, Peden E, Blanchette M, Rio DC, Singh R and Valcárcel J.
"Distinct programs establish widespread sex-specific alternative splicing in Drosophila melanogaster."
RNA, in press.

Corrionero A, Miñana B and Valcárcel J.
"Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug Spliceostatin A."
Genes & Development, in press.

Alló M, Schor IE, Muñoz MJ, de la Mata M, Agirre E, Valcárcel J, Eyras E and Kornblihtt AR.
"Chromatin and alternative splicing."
Cold Spring Harb. Symp. Quant. Biol., in press.

GENE REGULATION

Group: Regulation of Protein Synthesis in Eukaryotes

Group Structure:

Group Leader:	Fátima Gebauer
Lab manager:	Olga Coll
Postdoctoral Researchers:	Laurence Wurth, Antoine Graindorge
PhD Students:	Ana Villalba, Cristina Militti, Emilia Szostak
Master Student:	Marina García
Technician:	Anna Ribó

SUMMARY

The regulation of translation plays important roles in all aspects of cell and organismal biology, but its relevance is even higher when other mechanisms of gene regulation (e.g. transcription or splicing) are absent. This occurs during early embryonic development of many organisms. Regulation of translation is brought about by general or transcript-specific mechanisms. The latter are operated by regulators that bind to the mRNA in a sequence-specific manner, and recruit complexes that either interact directly with the translation machinery, or promote changes in the mRNA such as the elongation of the poly(A) tail. We study two examples of regulation that include these mechanisms and are essential for development of the fruitfly *Drosophila melanogaster*. We have found novel regulators, and our efforts concentrate on trying to understand how they function. We have also initiated experiments to decipher if/how one of the conserved regulators we have identified is involved in cancer progression in mammals.

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). SXL and UNR form a complex that binds to the 3' UTR of *msl2* mRNA, but they are not sufficient to repress translation (Figure 1). Using RNA-affinity chromatography, we have identified other co-factors which we are now characterizing. In addition, we have found that UNR performs opposite functions in males, as it promotes DCC binding to the X chromosome in this sex probably by binding to distinct transcripts. A genome-wide approach has revealed that indeed UNR binds to targets in a sex-specific fashion. Because some of these have been implicated in cancer progression, we are investigating whether UNR influences this process in mammals under the umbrella of the RNAReg Consolider network.

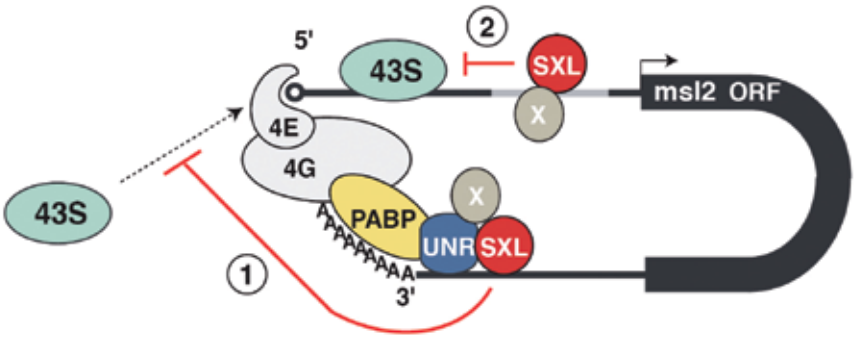


Figure 1.
Current understanding of the mechanism of translational repression of msl2 mRNA.
The translation initiation factors eIF4G, eIF4E and PABP are indicated. SXL binds to specific sites located in both the 5' and 3' UTRs of msl2 mRNA. SXL bound to the 3' UTR recruits UNR, which in turn contacts PABP and possibly other factors (X) leading to inhibition of 43S ribosomal complex recruitment (step 1). SXL bound to the 5' UTR inhibits the scanning of 43S complexes that may have escaped the 3' UTR-mediated control (step 2).

2. Translational regulation by cytoplasmic polyadenylation

Embryonic axis formation in *Drosophila* depends on the localized translation of the morphogen Bicoid at the anterior and the timely translation of the receptor Toll. Both of these events are activated by a process called cytoplasmic polyadenylation. The sequences and factors regulating cytoplasmic polyadenylation in *Drosophila* are largely unknown. Using a cell-free cytoplasmic polyadenylation system obtained from early embryos we have found that Toll mRNA is polyadenylated by a novel, non-canonical mechanism. Other transcripts follow the canonical pathway, which requires the cytoplasmic polyadenylation element (CPE) and the polyadenylation hexanucleotide. We are currently trying to isolate the machineries responsible for both canonical and non-canonical polyadenylation.

PUBLICATIONS

Coll O, Villalba A, Bussotti G, Notredame C and Gebauer F.
"A novel, non-canonical mechanism of cytoplasmic polyadenylation operates in Drosophila embryogenesis."
Genes Dev, 24:129-134 (2010).
[Faculty 1000 recommended].

Reviews

Graindorge A, Militti C and Gebauer F.
"Post-transcriptional control of X-chromosome dosage compensation."
WIREs RNA, in press.

Mihailovich M, Militti C, Gabaldón T and Gebauer F.
"Eukaryotic cold shock domain proteins: versatile regulators of gene expression."
BioEssays, 32:109-118 (2010).

Martínez-Sánchez A, Gebauer F.
"Regulation of p27(kip1) mRNA expression by microRNAs."
Prog Mol Subcell Biol, 50:59-70 (2010).

GENE REGULATION

Group: Translational Control of Gene Expression

Group structure:

Group Leader: Raul Méndez

Postdoctoral Researchers: Maria Piqué, David Pineda (Juan de la Cierva awarded), Carolina Segura, Laure Weill

Students: Alessio Bava ("la Caixa" Foundation, Graduate Student), Valeria Giangarra (Graduate Student)
Vittorio Calderone (Graduate Student), Jordina Guillen (Graduate Student)

SUMMARY

The primary interest of our group is to understand the molecular mechanisms that dictate the temporal and spatial translational control of specific mRNAs during cell cycle progression and early embryonic development. Meiotic progression and early development are programmed, at least in part, by maternally inherited mRNAs. These mRNAs are not translated en masse at any one time, or even at any one place; rather, their translation is specifically regulated by sequences located at the 3´-untranslated region (3´-UTR) of the mRNA and their binding proteins. Cytoplasmic polyadenylation is one of the most important mechanisms for regulating translation during meiotic progression and is directly controlled by the RNA-binding protein CPEB. The work of our group focuses on four questions in the area of translational regulation by cytoplasmic changes in the poly(A) tail length of mRNAs encoding for factors that drive cell cycle progression: 1) Genome-wide identification of the mRNAs that are regulated by cytoplasmic changes in their poly(A) tail length; 2) Determination of the configuration of cis-acting elements that define the temporal and spatial translational regulation by CPEB; 3) Role of the localized CPE-mediated translational regulation in meiotic progression; and 4) Identification of the cell cycle-related events regulated by the CPEB family of proteins.

RESEARCH PROJECTS

1. Cytoplasmic polyadenylation and the CPEB familly of proteins

Early animal development is directed by maternal mRNAs that are stored in the developing oocyte, until subsequent use during the late stages of meiosis or after fertilization. These mRNAs contain short poly(A) tails (~20-40 nts), and it is only when these tails are elongated in response to progesterone stimulation (to ~150 nts) that translation take place. Polyadenylation requires an element in the 3'UTR named cytoplasmic polyadenylation element (CPE). The CPE is bound by a RNA-binding protein named CPE Binding protein 1 (CPEB1). CPEB1 is the best characterized and founder member of a family of four proteins conserved in their RNA-recognition domain but distinct in their regulatory motifs (Reviewed in Mendez, 2001). In Xenopus oocytes, CPEB-1 can assemble two different functional complexes to either repress translation or to mediate cytoplasmic polyadenylation and activate translation. The switch from the «repressing-state» to the «activating-state» is driven by phosphorylation by Aurora-A and Cdc2/Plk1 kinases (Mendez, 2000a, Mendez 2000b, Mendez 2002). The unphosphorylated CPEB1 recruits a set of factors that shorten the poly(A) tail and blocks the recognition of the cap by the translational machinery. Once phosphorylated, CPEB1 changes its co-factors to recruit the cytoplasmic polyadenylation machinery and enhance the association of the translation initiation factors to the CPE-regulated mRNA. In two recent works (Pique 2008, Belloc and Mendez 2008) we have performed a systematic analysis of the combinations of cis-acting elements that define, qualitatively and quantitatively, the differential translational control of CPE-regulated mRNAs during meiosis. Overlaying this temporal control of translational activation, the same combinatorial pattern of CPEs contains the spatial information defining the oocyte subcellular localization where the mRNAs will be activated (Eliscovich 2008). All together our results indicate that the CPEBs control, in time and space, the translation of hundred of mRNAs encoding factors implicated in cell cycle, cell differentiation, angiogenesis, inflammation and other cellular functions. The CPEBs act, therefore, as hubs controlling the coordinated expression of up to 20% of the genome (Belloc et al. 2008).

Although the function of the CPEBs has been elucidated mainly in the xenopus oocytes, we have recently found that CPEB-mediated post-transcriptional regulation by phase-specific changes in poly(A) tail length is not a meiotic specific mechanism, but also required for **cell proliferation and specifically for M-phase entry in mitotically dividing cells**, where CPEB1 and CPEB4 regulate the polyadenylation of mRNAs encoding factors differentially expressed during cell cycle (Novoa et al 2010).

2. Defining the translational control circuit that regulates progression through the two meiotic divisions

Recent works from our group have contributed to define a molecular circuit to temporally regulate translational activation of mRNAs encoding the factors that mediate phase transitions during meiosis. This circuit is stabilized by a number of positive and negative feed back loops that ensure the irreversibility of the process, produce hysteric responses (switch-like transitions) and, yet, allow the key kinase activities that drive meiotic progression to oscillate between the two meiotic divisions. One essential negative

feed-back loop is mediated by the CPEB1-driven translational activation of mRNA encoding C3H-4, an ARE-binding protein that we find to accumulate in MI and whose ablation induces meiotic arrest. C3H-4, in turn, opposes CPEB activity on mRNAs containing both CPEs and AREs by recruiting the CCR4-deadenylase complex to the ARE-containing mRNAs (Belloc and Mendez 2008). However, this negative feed-back loop, required to exit the first meiotic metaphase, has to be inactivated to start the second meiotic division. We are studying how this inactivation is accomplished.

We have also defined an essential positive feed back loop where CPEB1 regulates the translation of CPEB4 mRNA. We have shown that CPEB4 can regulate the translation and mediate cytoplasmic polyadenylation of the same mRNAs that CPEB1, but at different meiotic times. Thus, while CPEB1 regulates the early maternal mRNAs during the first meiotic division CPEB4 regulates the late mRNAs during the second meiotic division. Altogether our work shows that CPEB4 expression is controlled by CPEB1 and, in turn CPEB4 replaces CPEB1 when it is degraded in anaphase I. But CPEB1 and CPEB4 are activated by different kinases and at different meiotic phases (Igea and Mendez, 2010). We are currently studying the posttranslational regulation of CPEB4, which, contrary to CPEB1, is not activated by Aurora A kinase but rather by an unknown kinase activated latter during cell cycle. Mapping the phosphorylated residues in CPEB4, characterizing how that regulates cytoplasmic polyadenylation and identifying the regulatory kinase are part of our future goals. Delineating the signal transduction pathways that regulate each CPEB will not only contribute to better define the meiotic circuit, but should be extrapolable to somatic cells and different extracellular signals.

3. Regulation of gene expression by CPEBs in somatic tissues.

a) Translational control of mitotic cell cycle

We have shown that CPEB-mediated post-transcriptional regulation by phase-specific changes in poly(A) tail length is required for M-phase entry and cell proliferation in mitotically dividing cells. This translational control is largely mediated by two members of the CPEB-family of proteins, CPEB1 and CPEB4, which regulate the polyadenylation of mRNAs encoding factors differentially expressed during cell cycle. We conclude that regulation of poly(A) tail length is not only required to compensate for the lack of transcription in specialized cell divisions, but acts as a general mechanism to control mitosis (Novoa et al., 2010). We continue to investigate the differential functions of CPEB1 and CPEB4 during mitotic divisions and their phase specific regulation by different signal transduction pathways. We are also studying in great detail their functions in chromosome segregation. In collaboration with the Isabelle Vernos group (CRG), we have shown that spindle-localized translational activation, by cytoplasmic polyadenylation, of the mRNAs encoding for TPX2 (Targeting Protein for Xenopus kinesin-like protein 2) and XKid (Xenopus Kinesin-like DNA binding protein) is essential to complete the first meiotic division and also for chromosome segregation in Xenopus oocytes (Eliscovich et al 2008). We have also found that CPEB and co-factors in the translational repression and activation complexes are asymmetrically distributed within the mitotic spindle in somatic cells. We are analyzing whether these factors generate translational gradients that define the polarity of the spindle and performing a Genome-wide identification of spindle translated mRNAs.

b) CPEB-mediated translational control in pancreatic ductal adenocarcinomas (PDAs)

In collaboration with Pilar Navarro (IMIM/PRBB) and Paco Real (CNIO), we have found that members of the CPEB family of proteins are overexpressed in tumors, resulting in abnormal regulation of mRNAs encoding pro-tumoral proteins. Moreover, nude mice injection of human pancreatic tumoral cells in which the levels of CPEB4 have been stably knocked down results in much reduced tumor growth, invasion and vascularization. Our results indicate that overexpression of CPEB1 and CPEB4 could reprogram gene expression in tumoral cells causing abnormal expression of growth, invasion and angiogenic factors, which will promote tumor growth and vascularization in PDA (Ortiz et al. Manuscript under review).

Literature cited:

Belloc and Méndez.
Nature, 452(7190):1017-1021 (2008).

Belloc E, Piqué M, Méndez R.
Biochem Soc Trans, 36(4):665-670. (2008).

Elíscovich Eliscovich C, Peset I, Vernos I, Méndez R.
Nature Cell Biology, 10(7):858-865 (2008).

Mendez and Richter.
Nat Rev Mol Cell Biol 2, 521-529. (2001).

Mendez R, Hake LE, Andresson T, Littlepage LE, Ruderman JV, Richter JD.
Nature, 404:302-307 (2000a).

Mendez R, Murthy KG, Ryan K, Manley JL, Richter JD. Mol
Cell, 6:1253-1259 (2000b).

Mendez R, Barnard D, Richter JD.
EMBO J, 21:1833-1844 (2002).

Novoa I, Gallego J, Ferreira PG, Mendez R.
Nature Cell Biology. in press.

Piqué M, López JM, Foissac S, Guigó R, Méndez R.
Cell, 132, 434-448 (2008).

PUBLICATIONS

Mejías M, García-Pras E, Gallego J, Méndez R, Bosch J and Fernández M.
"Relevance of the mTOR signaling pathway in the pathophysiology of splenomegaly in rats with chronic portal hypertension".
Journal of Hepatology, 52(4):529-539 (2010).

Novoa I, Gallego J, Ferreira PG, Méndez R.
"Mitotic cell cycle progression is regulated by CPEB-dependent translational control."
Nature Cell Biology, 12(5):447-456 (2010).

Igea A and Méndez R.
"Meiosis requires a translational positive loop where CPEB1ensues its replacement by CPEB4".
EMBO J, 29(13):2182-2193 (2010).

Book chapter

Novoa I, Eliscovich C, Belloc E and Mendez R.
"Oocyte-specific translational control mechanisms".
In: *Oogenesis: The Universal Process Marie-Helene Verlhac and Anne Villeneuve*, 2010 John Wiley & Sons, Ltd (2010).

DIFFERENTIATION AND CANCER

Coordinator: Thomas Graf

The areas of research covered within the programme are stem cell biology and regeneration as well as mechanisms of cell fate instruction, cancer and aging. The programme consists of five groups (in the order of their joining the CRG):

> Luciano Di Croce: Epigenetic mechanisms in leukemia and differentiation
> Thomas Graf: Hematopoietic stem cells, differentiation and reprogramming
> Salvador Aznar Benitah: Epithelial stem cells in the skin, cancer and circadian rhythms
> Bill Keyes: Skin stem cells, cancer and senescence
> Maria Pia Cosma: Cell fusion, reprogramming and regeneration

In the summer of 2010 Maria Pia Cosma joined the Program as a Senior Scientist, arriving from Naples, where she had worked at the Institute of Genetics and Biophysics. Pia has a broad background in transcription factor mechanisms, cell signaling and synthetic biology in yeast and mammalian cells. In the last few years she has worked on the reprogramming of somatic cells into induced pluripotent (iPS) cells, showing that activation of the wnt pathway is crucial for reprogramming. Her group also actively investigates the mechanism of cell reprogramming by cell fusion, both in cell cultures and in the eye, in a mouse model of Retinitis Pigmentosa. Pia was awarded a grant from the ERC and elected as an EMBO member in October this year.

DIFFERENTIATION AND CANCER

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

Group structure:

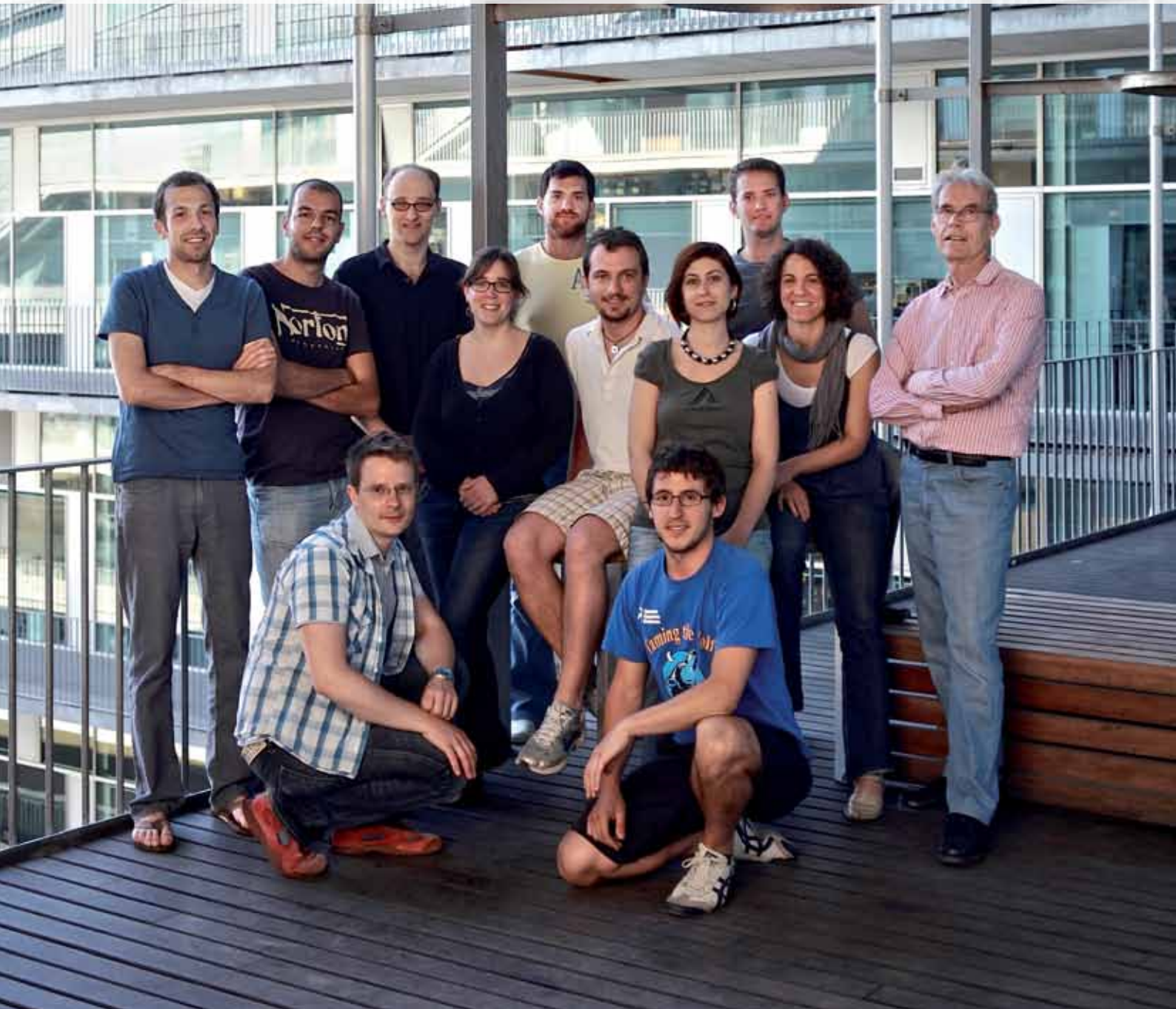
Group Leader: Thomas Graf

Staff Scientist: Chris van Oevelen

Postdoctoral Fellows: Christos Gekas (EMBO fellow), Eric Kallin (EMBO fellow)

PhD Students: Lars Bussmann (DAAD, CRG), Alessandro DiTullio (CRG), Francesca Rapino ("la Caixa" Foundation)
Luisa Irene de Andrés (CRG), Bruno Di Stefano ("la Caixa" Foundation, joined in September)

Technicians: Vanessa Chiganças (Plan Nacional, until October), Jose Francisco Infante (CRG)
Alai Urrutikoetchea (Plan E, joined in February), Clara Berenguer (Plan Nacional, joined in October)



SUMMARY

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

RESEARCH PROJECTS

1. Reprogramming of pre-B cells into macrophages reveals a novel mechanism of gene regulation

DNA methylation has critical functions in gene silencing during a multitude of biological processes including differentiation, X inactivation, transposon control, and cancer maintenance. The primary mechanism consists in the recognition of methylated CpG dinucleotides (MeC) by proteins harboring methylated DNA binding domains (MBDs), which in turn function to recruit various co-repressor complexes with chromatin-modifying potential. MeCs are established by the family of DNA methyltransferases that also maintain the marks during semi-conservative DNA replication, conferring to the cells an 'epigenetic memory'. Therefore, long-term silencing of gene expression through MeCs represents a mechanism to stably repress gene expression. Until now it has been assumed that derepression is mediated by demethylation, although evidence for the exitence of demethylases is controversial.

Recently, the Ten-eleven translocation (Tet) family of enzymes has been described that can hydroxylate the methyl group, resulting in hydroxymethylated CpGs. Of its 3 members, Tet2 is expressed at various stages of myeloid differentiation and loss-of-function mutations have been shown to be associated with several types of hematopoietic disorders, including acute myeloid leukemia (AML). However, its mechanism of action, as for Tet1 and Tet3, remains poorly understood. We have now found that Tet2 knockdowns repress the upregulation of specific myeloid genes, impairing myeloid differentiation from hematopoietic precursors and from pre-B cells induced to transdifferentiate into macrophages by C/EBPa. The expression of the human orthologues of these Tet2 target genes in acute myeloid leukemias (AML) correlated with the level of Tet2 expression, and $Tet2^{hi}$ samples were more differentiated and hypomethylated then $Tet2^{lo}$ samples. C/EBPa induced the upregulation of Tet2 during transdifferentiation and Tet2 in turn induced an increase of hydroxy-methylated cytosines at target gene promoters. Our data further show that C/EBPa and Tet2 proteins associate, suggesting that C/EBPa recruits Tet2 to its target genes, and that Tet2 function stabilizes C/EBPa binding. These results show that Tet2-type enzymes endow cells with the ability to rapidly respond to cell fate instructive transcription factors through DNA modifications and erasure of epigenetic memory.

Together, our findings have uncovered a novel mechanism of gene regulation, consisting in the activation of genes repressed by DNA methylation. The recognition of this mechanism was greatly facilitated by the use of cells that transdifferentiate, maximizing the chance of finding genes that become activated from a fully repressed state.

2. The platelet integrin alphaIIb (CD41) functions in hematopoietic stem cells

Hematopoietic stem cell (HSC) interaction with its niches is critical for self-renewal, survival and engraftment after transplantation. Whereas a lot is known about the molecular apparatus responsible for HSC self-renewal and differentiation, key components that regulate the interaction of HSCs and intermediate progenitors to the bone marrow hematopoietic niches remain poorly defined.

The classical platelet marker and fibronectin receptor integrin alphaIIb (CD41) transiently marks all emerging hematopoietic stem and progenitor cells in the embryo and remains expressed on a small subset of adult HSCs, yet its function in HSCs remains unknown. We found that, contrary to current views, CD41 expression increases drastically in aging mice to specifically mark the majority of Lin-sca1+ckit+ (LSK) flt3- and LSK CD150+ HSCs in 6 and 12 month old mice. Correspondingly, loss of CD41 in the CD41 knock-out mouse model that we described earlier (Zhang et al., Expl. Hematology, 2007) caused HSC-autonomous reductions in progenitor and mature cell numbers across all blood lineages that got exacerbated with age. CD41-deficient HSCs displayed increased apoptosis and microarray analysis revealed global reductions in their cell adhesion, motion and chemotactic repertoire. Remarkably, during aging CD41-/- HSCs rapidly and completely lost *in vivo* competitive repopulation ability when transplanted into adult irradiated hosts. Similarly, repopulation ability of wild-type HSCs was impaired when CD41 function was inhibited by blocking antibodies, suggesting a direct requirement of CD41 for HSC

homing into the bone marrow niches. Finally, we functionally assessed LSK CD34-flt3- HSCs from aged CD41YFP/+ heterozygote mice sub-fractionated on the basis of YFP expression. Surprisingly, we found that the CD41:YFP+ LT-HSC fraction contained virtually all multi-lineage repopulating activity in secondary transplants, suggesting that retention of CD41 expression by HSCs during aging is critical for the maintenance of their transplantation ability.

In summary, our studies suggest that engagement of integrin CD41 to extracellular matrix ligands in the bone marrow is required for both HSC maintenance under steady state conditions and engraftment after transplantation. I a similar situation is applicable to human HSCs, these results raise the possibility that CD41 activity can be modulated for the enhanced survival / transplantation efficiency of HSCs in a clinical setting.

3. The RNA binding protein Musashi 2 has a role in hematopoietic stem cells

Using a retroviral insertion screen for genes involved in the regulation of HSCs, we identified the Musashi 2 (msi2) gene. This gene has recently been implicated in the formation of acute leukemias and in the asymmetric division of hematopoietic progenitors. The msi2 gene encodes an RNA binding protein whose homologue in *Drosophila* shows a germ line stem cell defect when inactivated. To study its physiological function we collaborated with Thomas Floss from the GSH in Muenchen to determine the phenotype of Msi2 defective mice. The mutant mice are viable and have a functioning hematopoietic system. However, they show a decrease in the numbers of multipotent progenitors in the bone marrow, as well as of more mature cells fom the lymphoid and myeloid compartments. Interestingly, when bone marrow from these mice was transplanted competitively with bone marrow from wild type animals, Msi2 defective cells were dramatically outcompeted. This indicates that the Msi2 gene has an important function in the maintenance of functional HSCs. Current studies are directed towards understanding the mechanism of action of this gene in hematopoietic stem and progenitor cells.

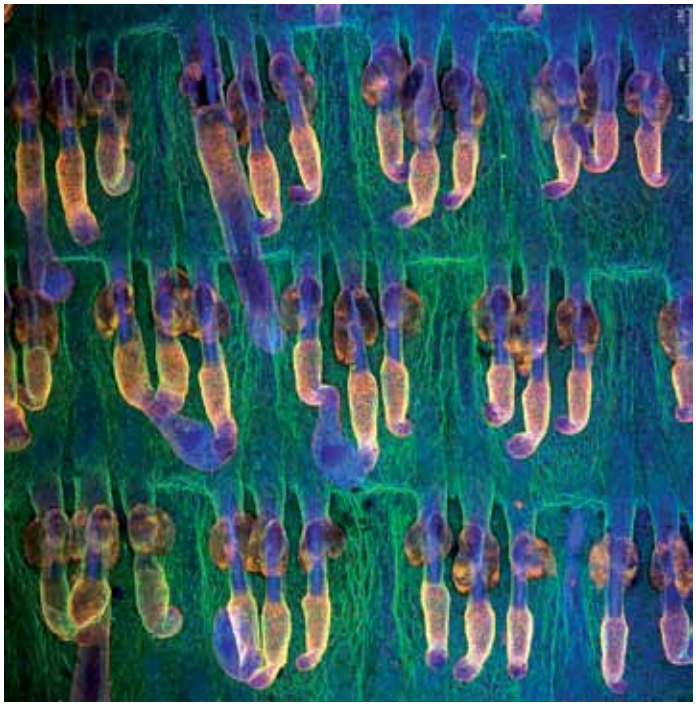


Figure 1.
This image is part of the efforts of Savador Aznar Benitah's group to characterize the molecular pathways that underlie the function of epidermal (bulge) stem cells, and how they are deregulated in cancer and aging. Shown is a section of mouse tail skin to identify the bulge stem cells, which are K15 keratin (red) and alpha6 integrin (bright green) positive. The nuclei are stained in blue.

PUBLICATIONS

Gekas C and Graf T.
"Induced pluripotent stem cell–derived human platelets: one step closer to the clinic."
J. Exp. Med., 207:2781-2784 (2010).

Singbrant S, Karlsson G, Ehingen M, Olsson K, Jaako P, Miharada KI, Stadtfeld M, Graf T, Karlsson S.
"Canonical BMP signaling is dispensable for hematopoietic stem cell function in both adult and fetal liver hematopoiesis, but essential to preserve colon architecture."
Blood, 115(23):4689-98 (2010).

Bertozzi CC, Schmaier AA, Mericko P, Hess PR, Zou Z, Chem M, Chen CY, Xu B, Lu MM, Zhou D, Sebzda E, Santore MT, Merianos DJ, Stadtfeld M, Flake AW, Graf T, Skoda R, Maltzman JS, Koretzky GA, Kahn ML.
"Platelets regulate lymphatic vascular development through CLEC-2-SLP-76 signaling."
Blood, 116(4):661-70 (2010).

Xie H, Laiosa CV, Graf T.
"Reprogramming of committed lymphoid cells by enforced transcription factor expression."
Methods Mol Biol, 636:219-32 (2010).

DIFFERENTIATION AND CANCER

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

Group structure:

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

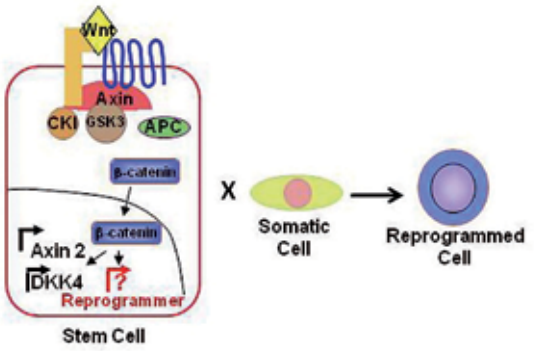
SUMMARY

Differentiation from zygotes has been considered as a unidirectional route. Recently, however, it has become clear that the reverse path is also possible: the reprogramming of somatic nuclei, i.e. the de-differentiation of somatic cells into pluripotent stem-like cells. Whether somatic cell reprogramming can occur *in vivo* in higher vertebrates and what the molecular mechanisms and genes driving reprogramming are, it remains to be defined. We have shown that activation of the Wnt/β-catenin signalling pathway enhances reprogramming of somatic cells after their fusion with stem cells. Remarkably, the activation of this signalling pathway also controls regeneration in response to damage in lower and higher vertebrates; furthermore, cell fusion is one possible mechanism of regeneration in vertebrates. Our main goals are: i.) to dissect the mechanisms of Wnt-mediated somatic cell reprogramming and ii.) to determine whether in mice activation of Wnt/β-catenin signalling controls cell-fusion mediated tissue regeneration.

RESEARCH PROJECTS

1. Identification of β-catenin targets and molecular pathways that control cell reprogramming, and analysis of their interactions via network-identification algorithms

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



Periodic accumulation of beta-catenin in ES cells enables them to reprogramme somatic cells after fusion. Taken from Lluis et al, 2008.

2. To determine whether Wnt/β-catenin-dependent reprogramming of fused cells is a mechanism of regeneration in higher vertebrates

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

PUBLICATIONS

Sanges D and Cosma MP.
"Reprogramming cell fate to pluripotency: the decision-making signalling pathways."
Int J Dev Biol, 54(11-12):1575-87 (2010).

Polito VA, Abbondante S, Polishchuk RS, Nusco E, Salvia R, Cosma MP.
"Correction of CNS defects in the MPSII mouse model via systemic enzyme replacement therapy."
Hum Mol Genet, 19(24):4871-85 (2010).

Lluis F, Pedone E, Pepe S, Cosma MP.
"The Wnt/β-Catenin Signaling Pathway Tips the Balance Apoptosis and Reprogramming of Cell-Fusion Hybrids."
Stem Cells, 28(11):1940-9 (2010).

Buono M, Visigalli I, Bergamasco R, Biffi A, and Cosma MP.
"SUMF1-mediated FGF signaling primes multilineage development Hematopoietic Stem and Progenitor Cells."
J Exp Med, 207(8):1647-60 (2010). (*)

Reviews

Buono M, and Cosma MP.
"Sulfatase activities towards the regulation of cell metabolism and signaling in mammals."
Cell Mol Life Sci, 67(5):769-80 (2010). (*)

Lluis F, Cosma MP.
"Cell-fusion-mediated somatic-cell reprogramming: a mechanism for tissue regeneration."
J Cell Physiol, 223(1):6-13 (2010). (*)

Book chapter

Cantone I, Cosma MP, Marucci L, Sanges D.
"Circuiti genici e programmabilita' della cellula".
In: *Biologia Sintetica*, Patron Editore (2010). (*)

(*) These publications result from the work of Dr. Maria Pia Cosma at Telethon Institute of Genetics and Medicine (TIGEM) and Institute of Genetics and Biophysics (IGB), CNR, Naples, Italy

DIFFERENTIATION AND CANCER

Group: Epigenetics Events in Cancer

Luciano Di Croce is an ICREA Research Professor.

Group structure:

Group Leader: Luciano Di Croce

Postdoctoral: Luigi Aloia, Lluis Morey, Holger Richly, Celia Jeronimo, Elisabeth Simboeck, Martin Lange

PhD Students: Santiago Demajo, Sophia Teichmann, Iris Uribesalgo Micás, Joana Ribeiro, Paola Pisano, Jain Payal

Technician: Arantxa Gutierrez

SUMMARY

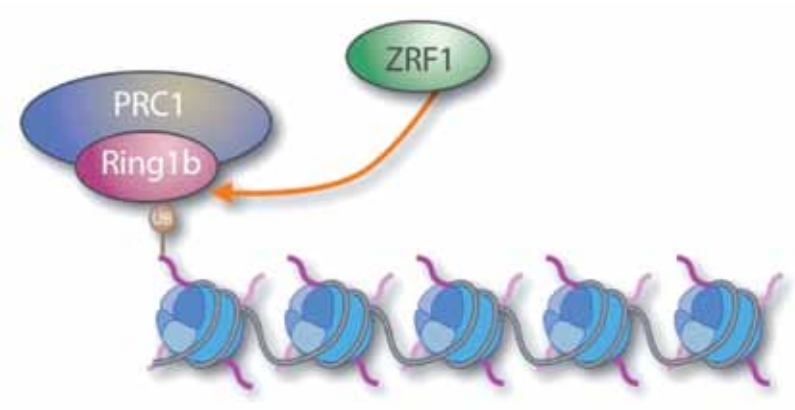
Understanding the genetic basis of cancers has been a topic of intense research, and hundreds of gene mutations have been identified that can cause carcinogenesis. However, in the past few years, increasing evidence has suggested that mutations are not the only genetic changes that lead to cancer. Indeed, 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. ZRF1

We have recently reported that the protein ZRF1 specifically binds to monoubiquitinated histone H2A and derepresses Polycomb target genes at the onset of cellular differentiation. Our results suggest that ZRF1 exerts its function in a two-step mechanism, by initially displacing the Polycomb-repressive complex 1 (PRC1) from chromatin and subsequently acting together with histone H2A-specific deubiquitinases to facilitate transcriptional activation of its target genes. These findings demonstrate an ambiguity of the epigenetic monoubiquitin mark at histone H2A. Once considered to be a hallmark of gene silencing, it is now clear that this mark can also be utilized as a recruitment platform for proteins engaged in gene activation. Genome-wide analyses demonstrate that ZRF1 is recruited to typical Polycomb target genes, thereby putting it in a position to have an impact on differentiation and animal development. This molecular mechanism for ZRF1 may represent one of the first steps in switching silenced genes to a transcriptionally active state. The strong link between the Polycomb system and differentiation suggests that ZRF1 is an important player in the occurrence of cancer. ZRF1 is often found misregulated in cancers; indeed, it ranks as one of the most upregulated genes in certain forms of leukemia.



We are now investigating the role of ZRF1 in cell cycle progression and transcription elongation, and how ZRF1 itself is regulated, and whether the re-localization of ZRF1 is regulated by specific protein interactions and/or its phosphorylation. Understanding the molecular mechanism of ZRF1 could also have the additional benefit of giving rise to developing anti-cancer drugs that could have a therapeutical use.

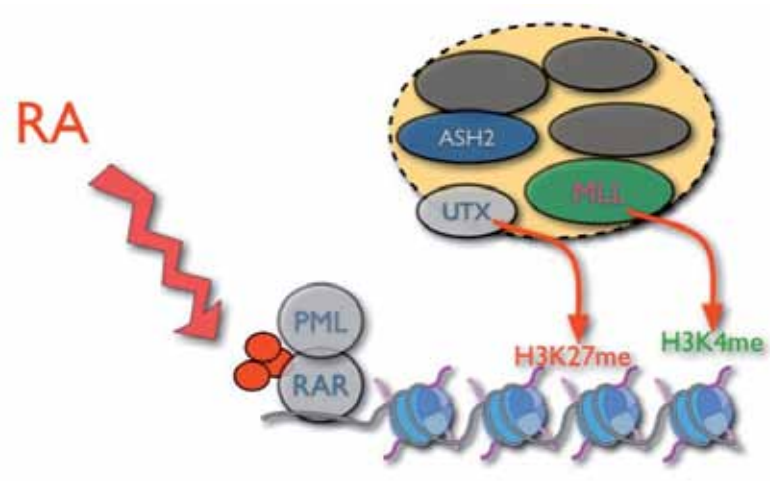
2. Role of histone demethylases in leukemia

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.

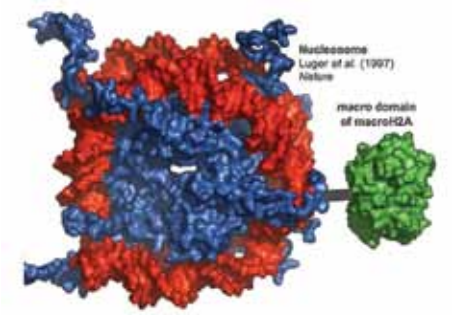
We will investigate:
(i) how Utx is recruited to promoters,
(ii) which are the target genes in the human genome,
(iii) which role has Utx in tumorigenesis.

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-RAR 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.



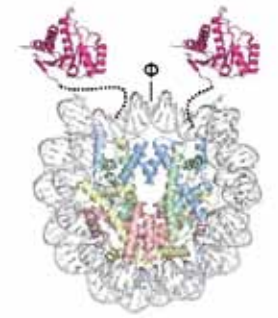
3. macroH2A and gene silencing

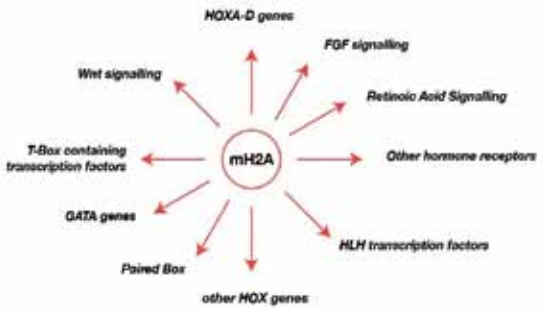
The most extensive histone modification is the complete exchange of canonical histones for variant ones. Among all known histone variants, the so-called macroH2A is the one that is the most divergent from its canonical histone and is the least understood in its function. In addition to a homologous histone domain, macroH2A possesses a large C-terminal domain of unknown function – the macro domain.
The H2A domain of macroH2A histones is ~65% identical to that of the conventional H2A.



The early observation that macroH2A is enriched on the inactive X chromosome and centrosomes suggested an involvement in gene repression and heterochromatinization. Biochemical studies using synthetic templates have indicated that macroH2A-containing nucleosomes are structurally different in the vicinity of the dyad axis, and this correlates with the inability of transcription factors to bind to DNA sites inserted nearby. These observations suggest that incorporation of macroH2A into nucleosomes could confer an epigenetic mark for gene repression. However, there are no known gene targets for macroH2A-dependent transcriptional repression, and no evidence for the mechanisms by which macroH2A could be recruited to specific genes and repress transcription *in vivo*.

Our microarray-based analysis in human male pluripotent cells uncovered occupancy of macroH2A at a large number of genes coding for key regulators of development and cell fate decisions. On these genes, presence of macroH2A is a repressive mark that locally and functionally overlaps with Polycomb repressive complex 2. We demonstrate that macroH2A contribute to the fine-tuning of temporal activation of HOXA cluster genes during neuronal differentiation. Furthermore, elimination of macroH2A function in zebrafish embryos produced severe specific phenotypes. Taken together, our data demonstrate that macroH2A constitutes an important epigenetic mark involved in the concerted regulation of gene expression programs during cellular differentiation as well as vertebrate development.



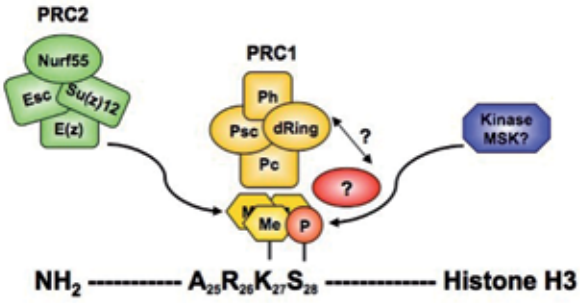


Some key questions remain still unsolved: which are the molecular mechanism involved in macroH2A deposition? Does macroH2A occupancy also control regulative regions in the genome other than gene promoters? A combination is biochemistry and genome-wide analysis are necessary to answer to those questions.

4. Cross-talk of epigenetic marks in gene regulation

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

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



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

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

PUBLICATIONS

Richly H, and Di Croce L.
"The flip side of the coin: Role of ZRF1 and histone H2A ubiquitination in transcriptional activation." Cell Cycle, in press.

Uribesalgo I, and Di Croce L.
"Dynamics of epigenetic modifications in leukemia."
Brief Funct Genomics, in press.

Richly H, Rocha-Viegas L, Domingues Ribeiro J, Demajo S, Gundem G, Nakagawa T, Rospert S, Lopez-Bigas N, Ito T, and Di Croce L.
"Transcriptional activation of Polycomb-repressed genes by the H2A-ubiquitin binding protein ZRF1."
Nature, 468:1124-28 (2010).

Simboeck E, Sawicka A, Zupkovitz G, Senese S, Winter S, Dequiedt F, Ogris E, Di Croce L, Chiocca S, Seiser C.
"A phosphorylation switch regulates the transcriptional activation of cell cycle regulator p21 by histone deacetylase inhibitors."
J Biol Chem, 285:41062-73 (2010).

Simboeck E and Di Croce L.
"HDAC1, a novel marker for benign teratomas."
EMBO J, 29:3893-95 (2010).

Buschbeck M, and Di Croce L.
"Approaching the molecular and physiological function of macroH2A variants."
Epigenetics, 5:118-23 (2010).

Review

Richly H, Lange M, Simboeck E, Di Croce L.
"Setting and resetting of epigenetic marks in malignant transformation and development."
Bioessays, 32:669-79 (2010).

DIFFERENTIATION AND CANCER

Group: Epithelial Homeostasis and Cancer

Salvador Aznar Benitah is an ICREA Researcher.

Group structure:

Group Leader:	Salvador Aznar Benitah
Postdoctoral:	Cristina Hidalgo
PhD Students:	Peggy Janich, Nuno Luis, Stefania Mejetta
Technicians:	Gloria Pascual
Masters Student:	Susann Minkwitz

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.

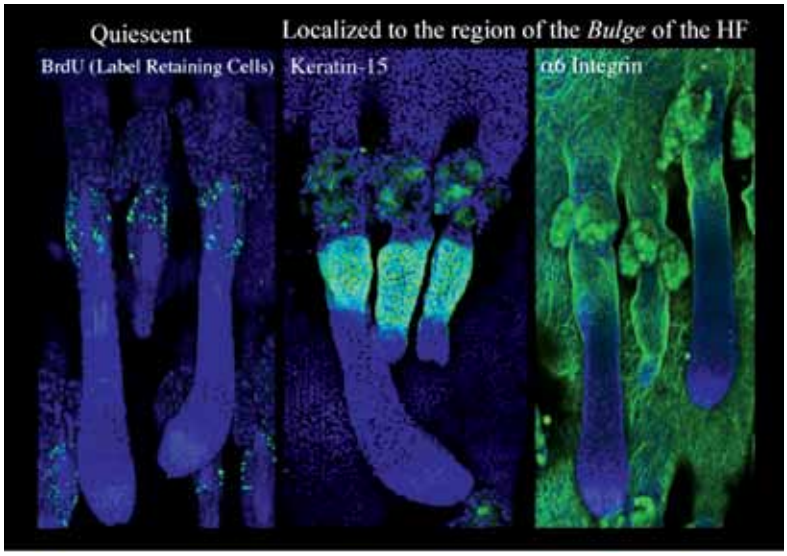


Figure 1
Epidermal stem cells are located at a region of the hair follicle known as the bulge. They are relatively quiescent and strictly positioned at their niche (Immunostaining and pictures by Peggy Janich).

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?

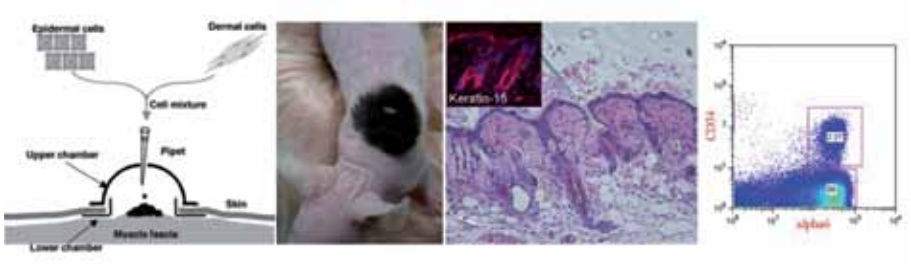


Figure 2
Epidermal stem cells are the only population that can regenerate the entire epidermal compartment. Epidermal stem cells transplanted onto nude mice can regenerate a homeostatic functional epidermis, hair follicles and sebaceous glands. A niche of epidermal stem cells is re-established once transplanted,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

Main Projects

1. Molecular and genetic mechanisms involved in epidermal self-renewal and differentiation.

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

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

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

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

2. Studying novel pathways relevant to epidermal and squamous tumour onset and progression.

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

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

DIFFERENTIATION AND CANCER

Group: Mechanisms of Cancer and Aging

Group structure:

Group Leader: Bill Keyes

Postdoctoral Researcher: Jason Doles

PhD Students: Valeria Di Giacomo ("la Caixa" Foundation fellow), Matteo Pecoraro, Mekayla Storer ("la Caixa" Foundation fellow)

Technicians: Alba Mas

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.

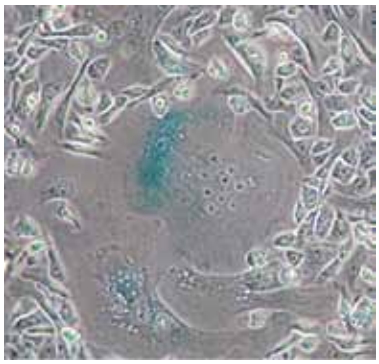


Figure 1.
Senescence-associated β-galactosidase staining identifies senescent cells (blue stain). Note the characteristic morphology, including enlarged flat shape and multiple nuclei. Multiple smaller transformed cells surround.

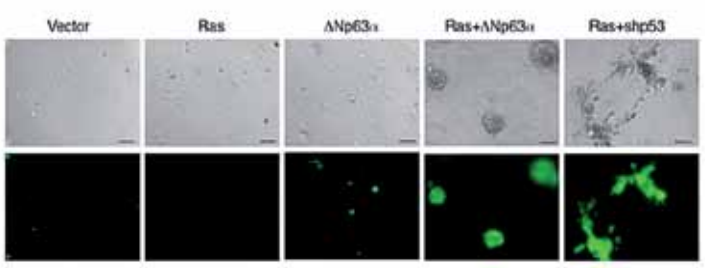
RESEARCH PROJECTS

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 2.
Cells transformed with overexpression of Ras and ΔNp63α and that induce the formation of squamous cell carcinoma in vivo, possess many stem-like properties, including an ability to form rapidly growing spheres in 3D tissue culture (above). This property is not shared by non-transformed cells, or cells transformed with Ras and a deficiency of p53.



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

PUBLICATIONS

Keyes WM*, Pecoraro M, Aranda V, Vernersson-Lindahl E, Li W,Vogel H, Garcia EL, Michurina T, Enikolopov G, Muthuswamy SK and Mills AA*
"ΔNp63α is an oncogene that targets chromatin remodeler Lsh to drive skin stem cell proliferation and tumorigenesis."
Cell Stem Cell, in press.
(corresponding authors)*

GENES AND DISEASE

Coordinator: Xavier Estivill

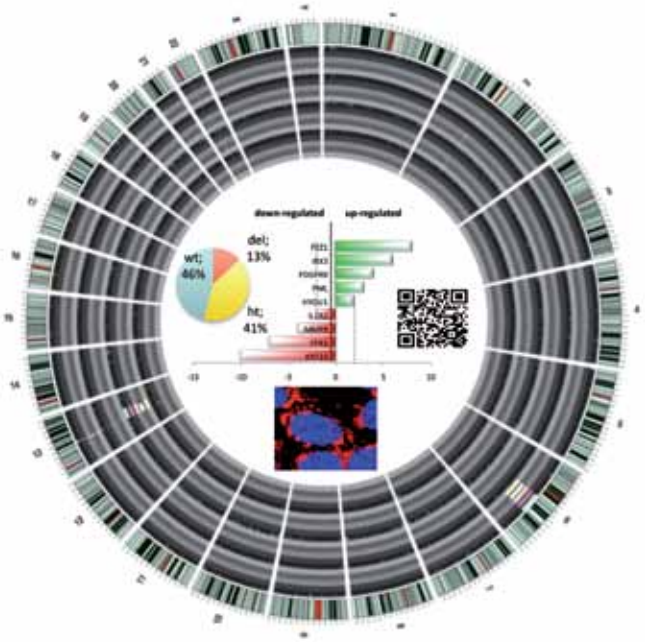
The global objective of the Genes and Disease Programme is to perform research on the molecular basis of human disease, from the discovery of genes involved in common human disorders to the development of preventive and therapeutic strategies. The Programme combines large-scale experimental approaches with advanced genetic strategies to elucidate biological determinants of common and rare human diseases.

Efforts of the Genes and Disease Programme researchers are focused on analysing sequence, structural and epigenetic variants of the human genome that could participate in the predisposition to disease. Specific collective work within the Programme is focused on understanding the function of genes with potential implications in mental retardation, psychiatric disorders and neurodegeneration, by using bioinformatics, cellular and animal model approaches. Programme researchers also work in the development of therapeutic strategies that could correct mental retardation, anxiety disorders and cancer.

Translational research is a primary concern of the Genes and Disease Programme members. To this aim, the Programme Group Leaders are part of the Rare Disorders and Epidemiology and Public Health Biomedical Research Networks, of the Spanish Ministry of Health, and other biomedical research networks at the regional, national or international levels. The genotyping core facility (CeGen-ISCIII) is associated to programme and has been supported by external funds (Instituto de Salud Carlos III, Spanish Ministry of Health). This platform provides state-of-the-art genotyping for common and rare disorders.

The programme consists of five research groups:

> Genetic Causes of Disease (Xavier Estivill)
> Gene Therapy (Cristina Fillat)
> Neurobehavioral Phenotyping of Mouse Models of Disease (Mara Dierssen)
> Gene Function and Murine Models of Disease (Susana de la Luna / Mariona Arbonés)
> Genomic and Epigenomic Variation in Disease (Stephan Ossowski)



GENES AND DISEASE

Group: Genetic Causes of Disease

Group Structure:

Group Leader: Xavier Estivill

Staff Scientist: Eulàlia Martí

Scientific Officer: Veronique Blanc (until March 2010)

Postdoctoral Fellows: Mónica Bañez-Coronel (Sara Borrell), Mariona Bustamante, Georgia Escaramís (CIBERESP), Marc Friedlander (since September 2010), Mònica Gratacòs, Esther Lizano (since April 2010) Raquel Rabionet (Ramón y Cajal), Eva Riveira (CIBERESP), José Manuel C. Tubío (since February 2010)

PhD Students: Johanna Aigner, Laia Bassaganyas, Elisa Docampo, Susana Iraola, Elisabet Mateu , Elena Miñones, Lorena Pantano, Ester Saus (until August 2010), Daniel Trujillano, Nàdia Vilahur (since September 2010)

Technicians: Anna Carreras, Birgit Kagerbauer, Marta Morell, Anna Puig (since May 2010), Cristian Tornador (since February 2010), Sergi Villatoro

SUMMARY

We are interested on how the variability of the human genome contributes to variation in phenotype. Our work aims to understand how different types of genetic variants (single nucleotide polymorphisms, SNPs; structural variations, mainly copy number variants or CNVs; and insertion/deletion variants, including transposable elements) influence common human disorders. We are also interested in studying the contribution of epigenetic modifications and non-coding RNA pathways in complex diseases, mainly neuropsychiatric and neurodegenerative disorders. We are approaching these questions through genomic platforms that interrogate SNPs, CNVs and methylation, including high-throughput sequencing for the characterisation of the exome and the transcriptome. Besides, we are developing functional studies to address the consequences of the genetic and epigenetic changes in cell physiology and pathology. The interplay between genetic changes and environmental factors is one of the major targets of our research. This is being tackled through perturbation experiments and by the study of large cohorts of patients in which epidemiological and clinical data has been assessed in highly controlled designs.

RESEARCH PROJECTS

1. Structural variations, environmental and human disease

Johanna Aigner, Laia Bassaganyas, Mariona Bustamante, Anna Carreras, Georgia Escaramís, Cristian Tornador, Raquel Rabionet, Eva Riveira, José Tubío, Nadia Vilahur, Sergi Villatoro

We characterize structural variation (SV) as genomic regions potentially involved in phenotype variation, including disease. Some of the targeted regions contain copy number variants (CNV), inversions, segmental duplications or transposable sequences. We systematically perform array comparative genomic hybridization (aCGH) experiments and use multiple technologies to further analyse CNVs at the whole-genome scale. CNVs with specific variable profiles are investigated in several common disorders. Disorders under investigation include schizophrenia, stroke, psoriasis, multiple sclerosis, asthma, Parkinson's disease and fibromyalgia.

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. The project is a collaborative effort between several groups in Spain (Clinic Hospital, University of Oviedo and CNIO, among others). The role of our group in the project is to describe the spectrum of non-coding RNAs and structural variants (SV) in normal and leukemic cells from patients with CLL (Figure 1). 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. The project will also provide gold standard information on the spectrum of SV of the human genome.

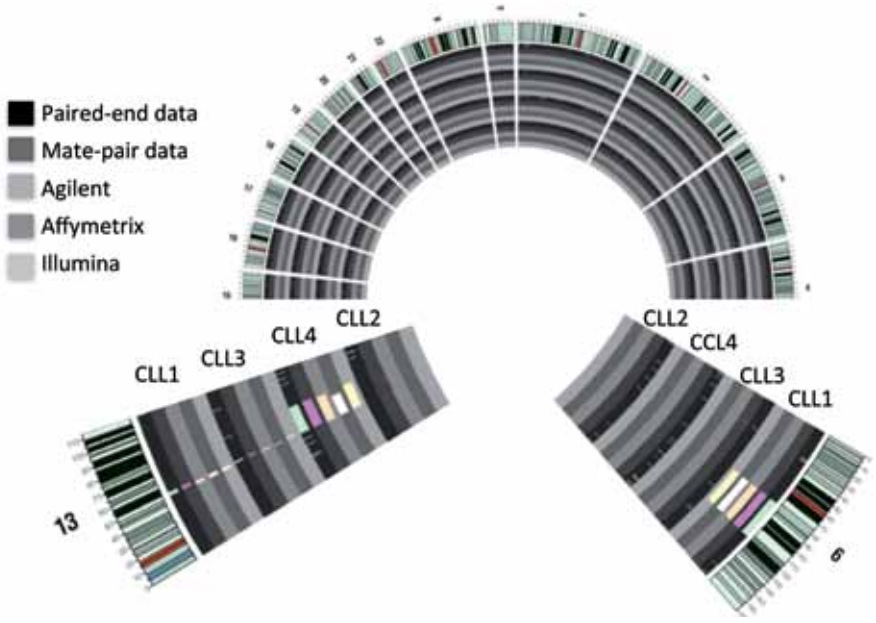


Figure 1.
Several platforms identify structural variation alterations in four cases of chronic lymphocytic leukemia.

Given the variability observed between monozygotic twins for the presence of CNVs and the presence of mosaicism in blood DNA samples from the general population, our group has been considered of interest to investigate the presence of CNVs among different areas the brain of an individual. To this end, experiments have been performed using aCGH among several brain areas (cortex, hippocampus, amygdala and hypothalamus, compared to cerebellum) of control subjects and patients affected by Alzheimer's disease.

The group has made specific progress in psoriasis, a chronic disorder of the skin affecting most ethnic groups. We have identified a common CNV that involves the deletion of two genes (LCE3B and LCE3C) in a significant fraction of patients with psoriasis. The common absence of these two genes, highly expressed in the injured epidermis, in psoriatic patients shows that skin barrier alterations play a role in the disease. This association has been replicated in several populations and constitutes one of the main genetic variants associated to psoriasis. We have also found that patients with other autoimmune disorders, such as rheumatoid arthritis and psoriatic arthritis, also have a higher frequency of deletions of LCE3B and LCE3C. We have also identified additional CNV affecting other regions of the LCE cluster (Figure 2).

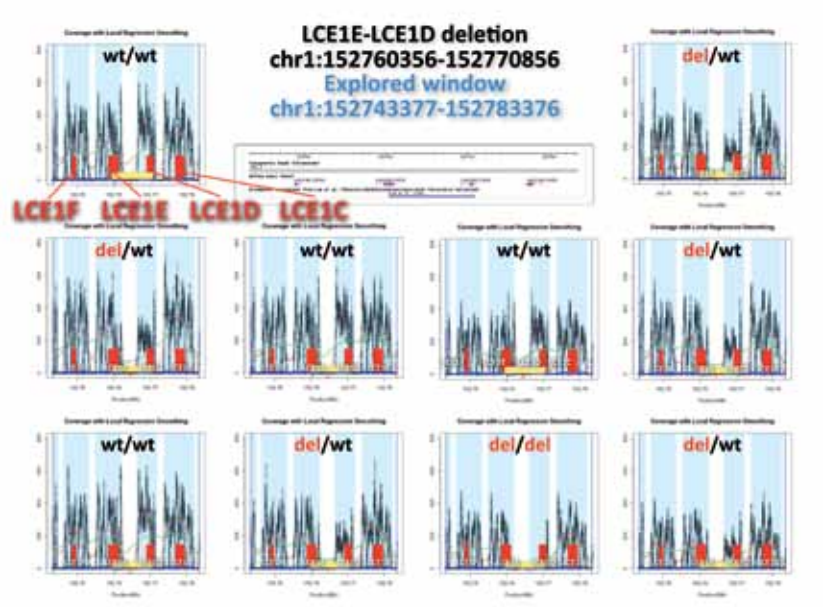


Figure 2.
Identification of a polymorphic structural variation deletion involving genes LCE1E and LCE1D by high-trhoughput sequencing after capturing genomic DNA of 10 individuals.

We have identified a CNV that consists in a 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. Moreover, we have shown a difference in the expression level of several mRNAs according to the BTNL8/BTNL3 genotype. We have found that overexpression of BTNL9 inhibits glucocorticoid-induced apoptosis, suggesting that it has a pro-survival function.

The group has a strong collaboration with investigators of the CREAL (Centre for Research in Environmental Epidemiology) to study the interaction between genetic and environment. One of the advantages of genome wide association scans (GWAS) integration is the ability to explore sources of heterogeneity such as ethnicity or environmental exposures. INMA (Infancia y Medio Ambiente) is a network of birth cohorts in Spain recruited from week twelve of pregnancy, including both small cohorts with long follow-up, as well as four new cohorts that followed a unique protocol and included 2600 children, with more than 2000 children-mother pairs having DNA available. We are currently studying samples from the INMA cohort in relation to neurodevelopment and Body Mass Index (BMI) and other phenotypes, taking into account environmental exposures.

2. Genetic variants associated to psychiatric disorders

Elisa Docampo, Mònica Gratacòs, Marta Morell, Anna Puig, Ester Saus, Daniel Trujillano

The group has developed activities in the study of substance abuse disorder, anxiety disorders (panic disorder and obsessive-compulsive disorder), eating disorders and affective disorders. In most cases, the

methodology involved the selection of variants of SNPs in candidate genes to be involved in the pathophysiology of the disease. We have studied the route of melatonin or genes that regulate circadian rhythm in patients with affective disorders, the endocannabinoid pathway in eating disorders and panic disorder, or galanin, estrogen and glutamate in obsessive compulsive disorder. In addition, we selected samples of patients with different addictions to perform an association study targeting gene family of neurexins to assess the possible involvement of variants in these genes in addictive processes. In addition to these lines of research, the group has established international partnerships to replicate the results of other investigations and contribute to studies of GWAS samples.

The group has studied the role of miRNAs and CNVs in susceptibility to psychiatric disorders. We have examined, for example, a CNV in obsessive-compulsive disorder, bipolar and anxiety disorders, and the involvement of miRNAs in anxiety disorders and insomnia associated with bipolar disorder. We have carried out functional studies for several miRNAs and have identified interesting regulation pathways for genes expressed in the central nervous system.

3. Functional genomics of neurological disorders

Mónica Bañez-Coronel, Marc Friedländer, Birgit Kagerbauer, Esther Lizano, Susana Iraola, Eulàlia Martí, Elisabet Mateu, Elena Miñones, Lorena Pantano

Small non-coding RNAs are involved in the guidance of diverse types of gene regulation, typically resulting in reduced expression of target genes. MicroRNA (miRNA) are a major class of non-coding RNAs. They are ubiquitously expressed and are believed to regulate most biological processes in a tissue- and temporal-specific manner. They are key players in the development of the central nervous system. The characterization of small RNA expression profiles in neurodegenerative disorders should provide important clues about the fine regulation of a plethora of genes that are involved in cell-specific survival of key neurones of brain regions that lead to disease.

We have characterized non-coding small RNAs, mainly miRNAs, in brain samples from individuals with Huntington disease (HD) and Parkinson disease (PD) and individuals without neuropathology. We have worked on: a/ analysis of expression profiles of miRNAs in combination with analysis of transcriptome and proteome; b/ evaluation of the functional characteristics of specific miRNAs, c/ development of bioinformatics tools for analysis of small RNAs, d/ identification small RNAs relevant to neurodegenerative disease through large scale sequencing datasets

We have identified the early downregulation of miR-34b/c in different brain areas with diverse degree of neuropathological affectation in PD. Our results further show that miR-34b/c modulate mitochondrial function and oxidative stress in neuronal cells. These results suggest that the early downregulation of miR-34b/c contribute to key neuropathological aspects in PD (Figure 3).

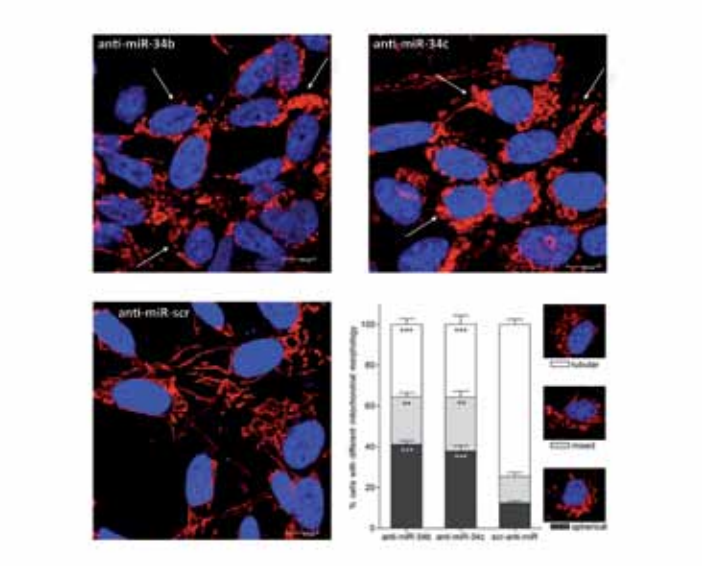


Figure 3.
Effect of depletion of miR-34b and miR-34c in mitochondrial morphology.

We have used high-throughput sequencing to characterize the small RNA transcriptome of brains of patients with Huntington's disease (HD) and individuals without neuropathological lesions. Our results show strong deregulation of miRNA and isomiR expression in HD and further suggest that altered miRNAs and isomiRs contribute to aberrant gene expression.

The group is also studying the role of small non-coding RNAs in neurodegeneration associated to triplet repeat expansion diseases such as HD and fragile X-associated tremor ataxia syndrome (FXTAS) (Figure 4). In triplet repeat expansion diseases, it has been recently shown that short CAG repeats generated from mutated long repeat hairpins, act as siRNAs and use the RNA interference machinery to trigger downstream effects. The group is evaluating the role of CAG and CGG repeated RNA in generating toxic effects in neurons.

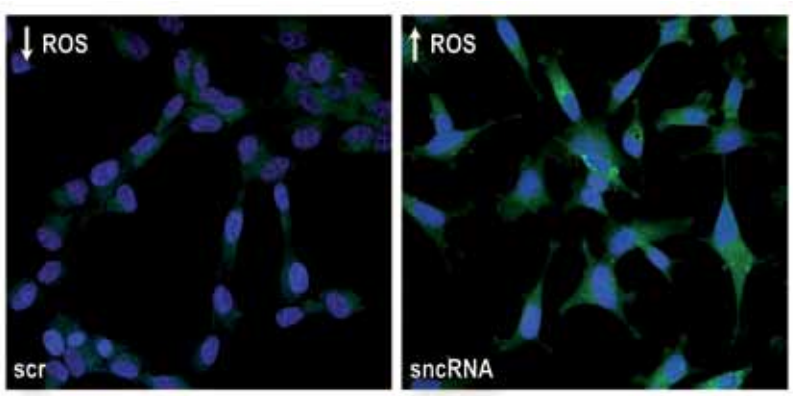


Figure 4.
Overexpression of a small non-coding RNA in oxidative stress in neuronal cell lines.

PUBLICATIONS

Articles

Guidi M, Muiños-Gimeno M, Kagerbauer B, Martí E, Estivill X, Espinosa-Parrilla Y.
"Overexpression of miR-128 specifically inhibits the truncated isoform of NTRK3 and upregulates BCL2 in SH-SY5Y neuroblastoma cells."
BMC Mol Biol, 11:95 (2010).

Biffi A, Sonni A, Anderson CD, Kissela B, Jagiella JM, Schmidt H, Jimenez-Conde J, Hansen BM, Fernandez-Cadenas I, Cortellini L, Ayres A, Schwab K, Juchniewicz K, Urbanik A, Rost NS, Viswanathan A, Seifert-Held T, Stoegerer EM, Tomás M, Rabionet R, Estivill X, Brown DL, Silliman SL, Selim M, Worrall BB, Meschia JF, Montaner J, Lindgren A, Roquer J, Schmidt R, Greenberg SM, Slowik A, Broderick JP, Woo D, Rosand J; International Stroke Genetics Consortium.
"Variants at APOE influence risk of deep and lobar intracerebral hemorrhage."
Ann Neurol, 68(6):934-43 (2010).

Genetic Analysis of Psoriasis Consortium & the Wellcome Trust Case Control Consortium 2 (including Estivill X, Riveira-Munoz E).
"A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1."
Nat Genet, 42(11):985-90 (2010).

Morales E, Guerra S, García-Esteban R, Guxens M, Alvarez-Pedrerol M, Bustamante M, Estivill X, Antó JM, Sunyer J.
"Maternal C-reactive protein levels in pregnancy are associated with wheezing and lower respiratory tract infections in the offspring."
Am J Obstet Gynecol, Epub 2010 Oct 25.

Martí E, Pantano L, Bañez-Coronel M, Llorens F, Miñones-Moyano E, Porta S, Sumoy L, Ferrer I, Estivill X.
"A myriad of miRNA variants in control and Huntington's disease brain regions detected by massively parallel sequencing."
Nucleic Acids Res, 38(20):7219-35 (2010).

Saus E, Soria V, Escaramís G, Vivarelli F, Crespo JM, Kagerbauer B, Menchón JM, Urretavizcaya M, Gratacòs M, Estivill X.
"Genetic variants and abnormal processing of pre-miR-182, a circadian clock modulator, in major depression patients with late insomnia."
Hum Mol Genet, 19(20):4017-25 (2010).

Saus E, Soria V, Escaramís G, Crespo JM, Valero J, Gutiérrez-Zotes A, Martorell L, Vilella E, Menchón JM, Estivill X, Gratacòs M, Urretavizcaya M.
"A haplotype of glycogen synthase kinase 3β is associated with early onset of unipolar major depression."
Genes Brain Behav, 9(7):799-807 (2010).

Saus E, Brunet A, Armengol L, Alonso P, Crespo JM, Fernández-Aranda F, Guitart M, Martín-Santos R, Menchón JM, Navinés R, Soria V, Torrens M, Urretavizcaya M, Vallès V, Gratacòs M, Estivill X.
"Comprehensive copy number variant (CNV) analysis of neuronal pathways genes in psychiatric disorders identifies rare variants within patients."
J Psychiatr Res, 44(14):971-8 (2010).

Gratacòs M, Escaramís G, Bustamante M, Saus E, Agüera Z, Bayés M, Cellini E, de Cid R, Fernández-Aranda F, Forcano L, González JR, Gorwood P, Hebebrand J, Hinney A, Mercader JM, Nacmias B, Ramoz N, Ribasés M, Ricca V, Romo L, Sorbi S, Versini A, Estivill X.
"Role of the neurotrophin network in eating disorders' subphenotypes: body mass index and age at onset of the disease."
J Psychiatr Res, 44(13):834-40 (2010).

Rodríguez-Santiago B, Brunet A, Sobrino B, Serra-Juhé C, Flores R, Armengol L, Vilella E, Gabau E, Guitart M, Guillamat R, Martorell L, Valero J, Gutiérrez-Zotes A, Labad A, Carracedo A, Estivill X, Pérez-Jurado LA.
"Association of common copy number variants at the glutathione S-transferase genes and rare novel genomic changes with schizophrenia."
Mol Psychiatry, 15(10):1023-33 (2010).

Donner J, Haapakoski R, Ezer S, Melén E, Pirkola S, Gratacòs M, Zucchelli M, Anedda F, Johansson LE, Söderhäll C, Orsmark-Pietras C, Suvisaari J, Martín-Santos R, Torrens M, Silander K, Terwilliger JD, Wickman M, Pershagen G, Lönnqvist J, Peltonen L, Estivill X, D'Amato M, Kere J, Alenius H, Hovatta I.
"Assessment of the neuropeptide S system in anxiety disorders."
Biol Psychiatry, 68(5):474-83 (2010).

Bergboer JG, Zeeuwen PL, Irvine AD, Weidinger S, Giardina E, Novelli G, Den Heijer M, Rodriguez E, Illig T, Riveira-Munoz E, Campbell LE, Tyson J, Dannhauser EN, O'Regan GM, Galli E, Klopp N, Koppelman GH, Novak N, Estivill X, McLean WH, Postma DS, Armour JA, Schalkwijk J.
"Deletion of Late Cornified Envelope 3B and 3C genes is not associated with atopic dermatitis."
J Invest Dermatol, 130(8):2057-61 (2010).

Soria V, Martínez-Amorós E, Escaramís G, Valero J, Crespo JM, Gutiérrez-Zotes A, Bayés M, Martorell L, Vilella E, Estivill X, Menchón JM, Gratacòs M, Urretavizcaya M.
"Resequencing and association analysis of arylalkylamine N-acetyltransferase (AANAT) gene and its contribution to major depression susceptibility."
J Pineal Res, 49(1):35-44 (2010).

Costas J, Gratacòs M, Escaramís G, Martín-Santos R, de Diego Y, Baca-García E, Canellas F, Estivill X, Guillamat R, Guitart M, Gutiérrez-Zotes A, García-Esteve L, Mayoral F, Moltó MD, Phillips C, Roca M, Carracedo A, Vilella E, Sanjuán J.
"Association study of 44 candidate genes with depressive and anxiety symptoms in post-partum women."
J Psychiatr Res, 44(11):717-24 (2010).

Naukkarinen J, Surakka I, Pietiläinen KH, Rissanen A, Salomaa V, Ripatti S, Yki-Järvinen H, van Duijn CM, Wichmann HE, Kaprio J, Taskinen MR, Peltonen L; ENGAGE Consortium.
"Use of genome-wide expression data to mine the "Gray Zone" of GWA studies leads to novel candidate obesity genes."
PLoS Genet, 6(6):e1000976 (2010).

Fonseca F, Gratacòs M, Escaramís G, De Cid R, Martín-Santos R, Fernández-Espejo E, Estivill X, Torrens M.
"Response to methadone maintenance treatment is associated with the MYOCD and GRM6 genes."
Mol Diagn Ther, 14(3):171-8 (2010).

Hüffmeier U, Estivill X, Riveira-Munoz E, Traupe H, Wendler J, Lohmann J, Böhm B, Burkhardt H, Reis A.
"Deletion of LCE3C and LCE3B genes at PSORS4 does not contribute to susceptibility to psoriatic arthritis in German patients."
Ann Rheum Dis, 69(5):876-8 (2010).

Docampo E, Rabionet R, Riveira-Muñoz E, Escaramís G, Julià A, Marsal S, Martín JE, González-Gay MA, Balsa A, Raya E, Martín J, Estivill X.
"Deletion of the late cornified envelope genes, LCE3C and LCE3B, is associated with rheumatoid arthritis."
Arthritis Rheum, 62(5):1246-51 (2010).

Soria V, Martínez-Amorós E, Escaramís G, Valero J, Pérez-Egea R, García C, Gutiérrez-Zotes A, Puigdemont D, Bayés M, Crespo JM, Martorell L, Vilella E, Labad A, Vallejo J, Pérez V, Menchón JM, Estivill X, Gratacòs M, Urretavizcaya M.
"Differential association of circadian genes with mood disorders: CRY1 and NPAS2 are associated with unipolar major depression and CLOCK and VIP with bipolar disorder."
Neuropsychopharmacology, 35(6):1279-89 (2010).

International Cancer Genome Consortium (including Estivill X).
"International network of cancer genome projects."
Nature, 464(7291):993-8 (2010).

Thorgeirsson TE et al. (including Estivill X as a collaborator)
"Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior."
Nat Genet, 42(5):448-53 (2010).

Hüffmeier U, Bergboer JG, Becker T, Armour JA, Traupe H, Estivill X, Riveira-Munoz E, Mössner R, Reich K, Kurrat W, Wienker TF, Schalkwijk J, Zeeuwen PL, Reis A.
"Replication of LCE3C-LCE3B CNV as a risk factor for psoriasis and analysis of interaction with other genetic risk factors."
J Invest Dermatol, 130(4):979-84 (2010).

Mercader JM, Fernández-Aranda F, Gratacòs M, Aguera Z, Forcano L, Ribasés M, Villarejo C, Estivill X.
"Correlation of BDNF blood levels with interoceptive awareness and maturity fears in anorexia and bulimia nervosa patients."
J Neural Transm, 117(4):505-12 (2010).

Azkona G, Amador-Arjona A, Obradors-Tarragó C, Varea E, Arqué G, Pinacho R, Fillat C, de la Luna S, Estivill X, Dierssen M.
"Characterization of a mouse model overexpressing beta-site APP-cleaving enzyme 2 reveals a new role for BACE2."
Genes Brain Behav, 9(2):160-72 (2010).

Pantano L, Estivill X, Martí E.
"SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells."
Nucleic Acids Res, 38(5):e34 (2010).

de Cid R, Ramos MD, Aparisi L, García C, Mora J, Estivill X, Farré A, Casals T.
"Independent contribution of common CFTR variants to chronic pancreatitis."
Pancreas, 39(2):209-15 (2010).

Muiños-Gimeno M, Montfort M, Bayés M, Estivill X, Espinosa-Parrilla Y.
"Design and evaluation of a panel of single-nucleotide polymorphisms in microRNA genomic regions for association studies in human disease."
Eur J Hum Genet, 18(2):218-26 (2010).

Castro-Giner F, de Cid R, Gonzalez JR, Jarvis D, Heinrich J, Janson C, Omenaas ER, Matheson MC, Pin I, Antó JM, Wjst M, Estivill X, Kogevinas M.
"Positionally cloned genes and age-specific effects in asthma and atopy: an international population-based cohort study (ECRHS)."
Thorax, 65(2):124-31 (2010).

Pedraza V, Gomez-Capilla JA, Escaramis G, Gomez C, Torné P, Rivera JM, Gil A, Araque P, Olea N, Estivill X, Fárez-Vidal ME.
"Gene expression signatures in breast cancer distinguish phenotype characteristics, histologic subtypes, and tumor invasiveness."
Cancer, 116(2):486-96 (2010).

Baca-Garcia E, Vaquero-Lorenzo C, Perez-Rodriguez MM, Gratacòs M, Bayés M, Santiago-Mozos R, Leiva-Murillo JM, de Prado-Cumplido M, Artes-Rodriguez A, Ceverino A, Diaz-Sastre C, Fernandez-Navarro P, Costas J, Fernandez-Piqueras J, Diaz-Hernandez M, de Leon J, Baca-Baldomero E, Saiz-Ruiz J, Mann JJ, Parsey RV, Carracedo A, Estivill X, Oquendo MA.
"Nucleotide variation in central nervous system genes among male suicide attempters."
Am J Med Genet B Neuropsychiatr Genet,153B(1):208-13 (2010).

Cukier HN, Rabionet R, Konidari I, Rayner-Evans MY, Baltos ML, Wright HH, Abramson RK, Martin ER, Cuccaro ML, Pericak-Vance MA, Gilbert JR.
"Novel variants identified in methyl-CpG-binding domain genes in autistic individuals."
Neurogenetics, 11(3):291-303 (2010).

Albacar G, Sans T, Martín-Santos R, García-Esteve L, Guillamat R, Sanjuan J, Cañellas F, Carot JM, Gratacòs M, Bosch J, Gaviria A, Labad A, Zotes AG, Vilella E.
"Thyroid function 48h after delivery as a marker for subsequent postpartum depression."
Psychoneuroendocrinology, 35(5):738-42 (2010).

Ma DQ, Rabionet R, Konidari I, Jaworski J, Cukier HN, Wright HH, Abramson RK, Gilbert JR, Cuccaro ML, Pericak-Vance MA, Martin ER.
"Association and gene-gene interaction of SLC6A4 and ITGB3 in autism."
Am J Med Genet B Neuropsychiatr Genet, 153B(2):477-83 (2010).

O'Roak B J, Morgan TM, Fishman D, Saus E, Alonso P, Gratacòs M, Estivill X, Teltsh O, Kohn Y, Kidd KK, Cho J, Lifton RP, State MW.
"Additional Support for the Association of SLITRK1 var321 and Tourette syndrome."
Mol Psychiatry, 15(5):447-50 (2010).

Sunyer J, Basagaña X, González JR, Júlvez J, Guerra S, Bustamante M, de Cid R, Antó JM, Torrent M.
"Early life environment, neurodevelopment and the interrelation with atopy."
Environ Res, 110(7):733-8 (2010).

Kogevinas M, Villanueva CM, Font-Ribera L, Livia D, Bustamante M, Espinoza F, Nieuwenhuijsen MJ, Espinosa A, Fernandez P, DeMarin DM, Grimalt JO, Grummt T, Marcos R.
"Genotoxic Effects in Swimmers Exposed to Disinfection By-products in Indoor Swimming Pools."
Environ Health Perspect, 118 (11):1531-1537 (2010).

Caley DP, Pink RC, Trujillano D, Carter DRF.
"Long Noncoding RNAs, Chromatin, and Development."
The Scientific World JOURNAL, 10: 90-102 (2010).

Arensburger P et al (including Tubio JM).
"Sequencing of Culex quinquefasciatus establishes a platform for mosquito comparative genomics."
Science, 330(6000):86-8 (2010).

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, 2010 Sep 28 [Epub ahead of print].

Alonso P, Gratacòs M, Segalàs C, Escaramís G, Real E, Bayés M, Labad J, Pertusa A, Vallejo J, Estivill X, Menchón JM.
"Variants in estrogen receptor alpha gene are associated with phenotypical expression of obsessive-compulsive disorder."
Psychoneuroendocrinology, 36(4):473-83 (2010).

Muiños-Gimeno M, Espinosa-Parrilla Y, Guidi M, Kagerbauer B, Sipilä T, Maron E, Pettai K, Kananen L, Navinés R, Martín-Santos R, Gratacòs M, Metspalu A, Hovatta I, Estivill X.
"Human microRNAs miR-22, miR-138-2, miR-148a, and miR-488 Are Associated with Panic Disorder and Regulate Several Anxiety Candidate Genes and Related Pathways."
Biol Psychiatry, Epub 2010 Dec 17.

Karwautz AF, Wagner G, Waldherr K, Nader IW, Fernandez-Aranda F, Estivill X, Holliday J, Collier DA, Treasure JL.
"Gene-environment interaction in anorexia nervosa: relevance of non-shared environment and the serotonin transporter gene."
Mol Psychiatry, 2010 Dec 7. [Epub ahead of print]

Riveira-Munoz E, He SM, Escaramís G, Stuart PE, Hüffmeier U, Lee C, Kirby B, Oka A, Giardina E, Liao W, Bergboer J, Kainu K, de Cid R, Munkhbat B, Zeeuwen PL, Armour JA, Poon A, Mabuchi T, Ozawa A, Zawirska A, Burden AD, Barker JN, Capon F, Traupe H, Sun LD, Cui Y, Yin XY, Chen G, Lim HW, Nair RP, Voorhees JJ, Tejasvi T, Pujol R, Munkhtuvshin N, Fischer J, Kere J, Schalkwijk J, Bowcock A, Kwok PY, Novelli G, Inoko H, Ryan AW, Trembath RC, Reis A, Zhang XJ, Elder JT, Estivill X.
"Meta-Analysis Confirms the LCE3C_LCE3B Deletion as a Risk Factor for Psoriasis in Several Ethnic Groups and Finds Interaction with HLA-Cw6."
J Invest Dermatol, Epub 2010 Nov 25.

Tubio JM, Tojo M, Bassaganyas L, Escaramis G, Sharakhov IV, Sharakhova MV, Tornador C, Unger MF, Naveira H, Costas J, Besansky NJ.
"Evolutionary dynamics of the Ty3/gypsy retrotransposons in the genome of Anopheles gambiae".
PLoS ONE, in press.

GENES AND DISEASE

Group: Gene Therapy

Group structure:

Group Leader: Cristina Fillat

Postdoctoral Fellows: Xavier Altafaj, Laura García, Maria Victoria Maliandi, Luciano Sobrevals

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

Technicians: Núria Andreu

SUMMARY

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

RESEARCH PROJECTS

1. Pancreatic cancer

Pancreatic cancer is a neoplasia with a very bad prognosis mainly due to the late diagnosis and inefficient current therapies. We take advantage of the current understanding of the mechanistic pathogenesis and of the comprehensive available data regarding inefficient therapeutic response to design alternative treatments. In the past few years we have been involved in exploring the feasibility of suicide gene-virotherapy. This system is based on the combination of two different principles, based on the attempt to achieve tumor destruction by allowing lytic adenovirus to replicate in tumor cells, multiply and disseminate, and by arming these viruses with the Herpes Simplex Thymidine Kinase gene (TK) that when combined with GCV induces apotosis mediated cell killing. This year we have gained insight in the understanding of the basic mechanisms of tumor cell killing induced by the TK/GCV system. We demonstrated the relevance of cell cycle control pathways and point out at Chk1 activation as a key factor to mediate TK/GCV cytotoxicity. Moreover we have shown the implication of E-cadherin in the modulation of the gap-junction mediated TK/GCV bystander effect. Interestingly we have also shown the key role of connexin-26 in the bystander effect of gemcitabine toxic metabolites in pancreatic adenocarcinoma. We have proved to boost the potency of adenoviruses when TK was expressed at the late phase of viral replication and we are actively working on providing tumor selectivity to those viruses. Post-transcriptional, transcriptional and viral capsid modification strategies are under development to confer selective tumor control to adenoviral replication.

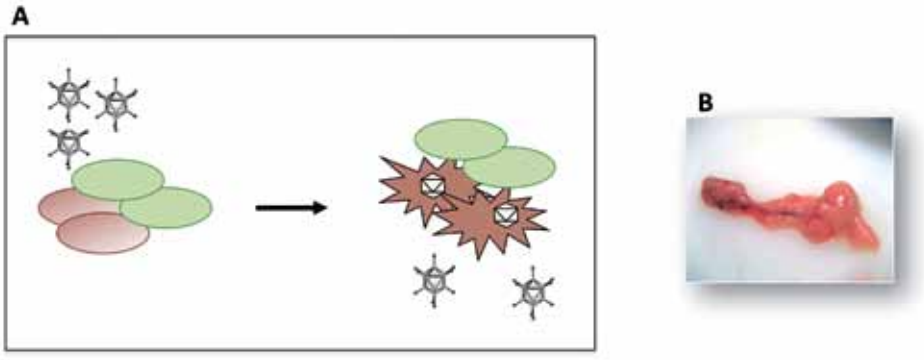


Figure 1.
Adenoviral oncoselectivity to target pancreatic tumors.
A. Scheme indicating viral lysis in tumoral cells (brown) but not in normal cells (green).
B. Orthotopic pancreatic tumor xenograft.

2. Down Syndrome

Down syndrome (DS) is a complex multi-system disorder, resulting from an extra copy of the human chromosome 21 gene HSA21. Although HSA21 contains more than three hundred genes, it has been hypothesized that the excess of specific target genes (dosage-sensitive) will have a major impact on the DS phenotype. To better understand the contribution of these specific genes, we have engineer a novel approach that takes advantage of the current knowledge from the existing mouse models. This approach is based on the rescue of a phenotype in a mouse model by the viral delivery of shRNA sequences targeting a candidate gene. A proof of principle approach was developed and shown that normalizing the expression of the Dyrk1A dosage sensitive gene in the striatum of adult TgDyrk1A by AAVshDyrk1A delivery rescued motor impairment. This approach highlighted Dyrk1A phenotypic dependence in motor alterations. Mental retardation is the more disabling trait in DS. We have preliminary evidences indicating that Dyrk1A normalization in the hippocampus attenuate specific hippocampal-dependent defects.

Studies on the molecular mechanisms by which the normalization of Dyrk1A can reverse such phenotypes are also being explored. This year we have also shown that over-expression of the HSA21 gene RCAN1 in a transgenic mouse results in marked cognitive defects. Despite the contribution triggered by the dosage-sensitive genes to the phenotype the over-representation of functional non-protein-coding elements might be involved in some of the abnormal phenotypes. On going efforts involve the development of viral gene-transfer approaches that help to evaluate the contribution of these non-coding elements to the cognitive phenotype of DS murine models.

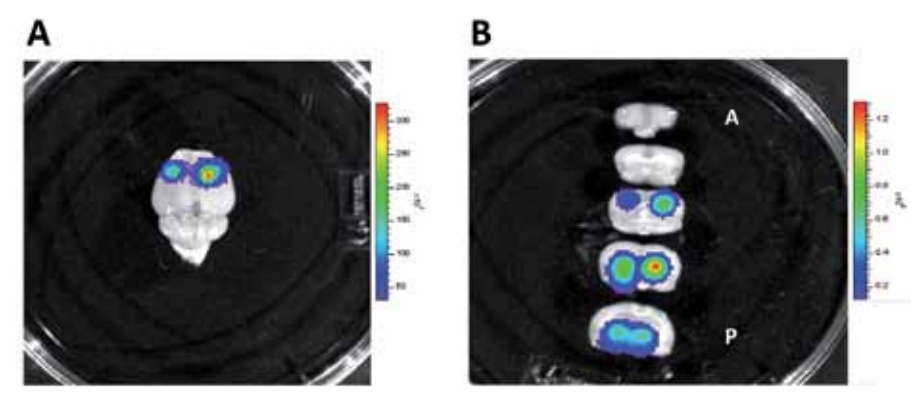


Figure 2.
Adeno-associated virus expressing Luciferase delivered to mouse striatum.
A. Whole brain.
B. Rostral to caudal coronal sections

PUBLICATIONS

Azkona G, Amador A, Obradors C, Varea E, Arqué G, Pinacho R, Fillat C, de la Luna S, Estivill S, Dierssen M.
"Characterization of a mouse model overexpressing beta-site APP-cleaving enzyme 2 reveals a new role for BACE2."
Genes Brain Behav, 9:160-72 (2010).

Albert MH, Bittner TC, Nonoyama S, Notarangelo L, Burns S, Imai K, Español T, Fasth A, Pellier I, Strauss G, Morio T, Gathman B, Noordzij JG, Fillat C, Hoenig M, Nathrath M, Meindl A, Pagel P, Wintergerst U, Fisher A, Trasher AJ, Belohradsky BH, Ochs HD.
"X-linked thrombocytopenia (XLT) due to WAS mutations: clinical characteristics, long-term outcome and treatment options."
Blood, 115: 3231-3238 (2010).

Fillat C, José A, Bofill-De Ros X, Mato-Berciano A, Maliandi MV, Abate-Daga D.
"Controlling adenoviral replication to induce oncolytic efficacy."
The Open Gene Ther J, 3:15-23 (2010).

Fillat C.
"Advances in Oncolytic Antitumour Adenoviral Therapies: Three key aspects."
The Open Gene Ther J Editorial, 3:8 (2010).

Fillat C, Dierssen M, Martínez de Lagrán M, Altafaj X.
"Insights from mouse models to understand neurodegeneration in Down Syndrome."
CNS Neurol Disor Drug Targets, 9:429-38 (2010).

Abate-Daga D, Garcia-Rodríguez L, Sumoy L, Fillat C.
"Cell cycle control pathways act as conditioning factors for TK/GCV sensitivity in pancreatic cancer cells."
BBA MolCell Res, 1803:1175-1185 (2010).

Garcia-Rodríguez L, Abate-Daga D, Rojas A, González JR, Fillat C.
"E-cadherin contributes to the bystander effect of TK/GCV suicide therapy and enhances its antitumoral activity in pancreatic cancer models."
Gene Ther, Epub 2010 Aug 19.

Ferrón SR, Pozo N, Laguna A, Aranda S, Porlan E, Moreno M, Fillat C, de la Luna S, Sánchez P, Arbonés ML, Fariñas I.
"Regulated segregation of kinase DYRK1A during asymmetric neural stem cell division is critical for EGFR-mediated biased signaling."
Cell Stem Cell, 7:367-379 (2010).

García-Rodríguez L, Pérez-Torras S, Carrió M, Cascante A, García-Ribas I, Mazo A, Fillat C.
"Connexin-26 is a key factor mediating gemcitabine bystander effect."
Mol Cancer Ther, in press.

Book chapter

Fillat C, Abate-Daga D.
"Terapia Génica. Biotecnología y Biofármacos."
Ed. Colegio Oficial de Farmacéuticos, 6:132-150 (2010).

PATENTS

Fillat C, Huch M.
"Conditionally replicating adenovirus effective in the treatment of tumors"
International Publication Date 02-09-2010, (WO2010/097419 A1).

GENES AND DISEASE

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

Group structure:

Group Leaders: Mariona Arbonés / Susana de la Luna

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

PhD Students: Krisztina Arató, Elisa Balducci, Chiara di Vona, Sonia Najas, Andrea Senna

Technicians: Erika Ramirez, Alicia Raya

SUMMARY

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

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

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

RESEARCH PROJECTS

1. DYRK1A and nervous system development

We have examined whether the anti-apoptotic effect of DYRK1A that we had previously observed in the developing retina of gain-of function *Dyrk1a* mutant mice have any impact in the establishment of retina circuitry in DS. To this aim, we have use a DS trisomic model, the Ts65Dn mouse. As in the gain of function mutants, these mice present reduced developmental cell death and, as a consequence, cellularity and electrical responses in the mature retina are significantly altered. This phenotype is due to the extra copy of *Dyrk1a* since retinas of trisomic Ts65Dn mice in which *Dyrk1a* dosage have been normalized by genetic means show normal developmental apoptosis and cellularity in the adult.

Some brain structures in *Dyrk1a*$^{+/-}$ heterozygous mutant mice present an important neuron loss. Examination of the formation of midbrain dopamine neurons, in mice with 1 or 3 functional copies of *Dyrk1a* has revealed an altered developmental apoptosis in the brain of these mutants. Similarly to our previous analysis in the retina of *Dyrk1a* transgenic models, apoptosis in the brain is increased in *Dyrk1a* loss-of function mutants and decreased in the gain-of function mutants. We have previously demonstrated that caspase-9 is phosphorylated at a major inhibitory residue, Thr125, by DYRK1A and that the levels of phosphorylated caspase-9 in the developing retina correlate with the magnitude of apoptosis. We are currently investigating whether deregulation of caspase-9 mediated apoptosis is contributing to the altered development of dopaminergic neurons in *Dyrk1a* mutant mice.

2. DYRK1A and neural stem cell regulation

The putative role of DYRK1A in controlling the homeostasis of neuronal precursor cells in the brain has been investigated. In collaboration with Isabel Fariñas (Universidad de Valencia), we have shown that DYRK1A is involved in the regulation of biased epidermal growth factor receptor (EGFR) signaling in the progeny of dividing neural stem cells of the adult subependymal zone. *Dyrk1a* heterozygous neural stem cells exhibit defects in self-renewal, EGF-dependent cell fate decisions and long-term persistence *in vivo*. At the molecular level, we found that both DYRK1A and EGFR are symmetrically or asymmetrically distributed during mitosis of neural stem cells and that EGFR asymmetry requires normal *Dyrk1a* dosage. This effect is due to DYRK1A preventing endocytosis-mediated degradation of EGFR. Our results have led to propose that symmetrical divisions play a role in the maintenance of the adult subependymal zone reservoir.

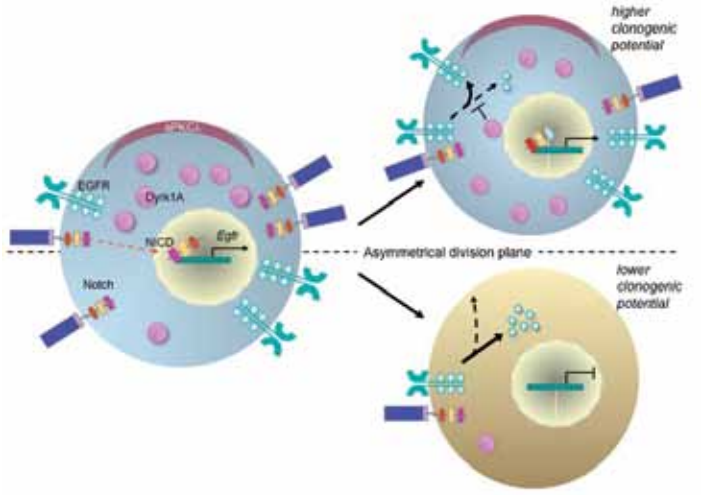


Figure 1.
DYRK1A regulation of EGFR distribution of the daughters of neural stem cells.

3. The DYRK family of protein kinases

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

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

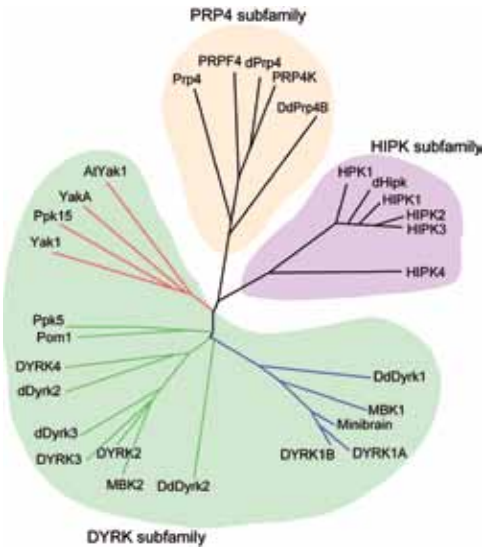


Figure 2.
DYRK family of protein kinases.

PUBLICATIONS

Arbones ML, de la Luna S.
"DYRK1A (dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1A)."
Atlas Genet Cytogenet Oncol Haematol, September 2010.
[URL: http://AtlasGeneticsOncology.org/Genes/DYRK1AID43234ch21q22.html]

Azkona G, Amador-Arjona A, Obradors-Tarragó C, Varea E, Arqué G, Pinacho R, Fillat C, de la Luna S, Estivill X, Dierssen M.
"Characterization of a mouse model overexpressing beta-site APP-cleaving enzyme 2 reveals a new role for BACE2."
Genes Brain Behav, 9:160-172 (2010).

da Costa Martins PA, Salic K, Gladka MM, Armand AS, Leptidis S, el Azzouzi H, Hansen A, Coenen-de Roo CJ, Bierhuizen MF, van der Nagel R, van Kuik J, de Weger R, de Bruin A, Condorelli G, Arbones ML, Eschenhagen T, De Windt LJ.
"MicroRNA-199b targets the nuclear kinase Dyrk1a in an auto-amplification loop promoting calcineurin/NFAT signalling."
Nat Cell Biol, 12:1220-1227 (2010).

Ferron SR, Pozo N, Laguna A, Aranda S, Porlan E, Moreno M, Fillat C, de la Luna S, Sánchez P, Arbonés ML, Fariñas I.
"Regulated segregation of kinase Dyrk1A during asymmetric neural stem cell division is critical for EGFR-mediated biased signalling."
Cell Stem Cell, 7:367-379 (2010).

Papadopoulos C, Arato K, Lilienthal E, Zerweck J, Schutkowski M, Chatain N, Müller-Newen G, Becker W, de la Luna S.
"Splice variants of the dual-specificity tyrosine phosphorylation-regulated kinase 4 (DYRK4) differ in their subcellular localization and catalytic activity."
J Biol Chem, Epub 2010 Dec 2.

Review

Aranda S, Laguna A, de la Luna S.
"DYRK family of protein kinases: evolutionary relationships, biochemical properties, and functional roles."
FASEB J, Epub 2010 Nov 3.

GENES AND DISEASE

Group: Neurobehavioral Phenotyping of Mouse Models of Disease

Group structure:

Group Leader:	Mara Dierssen
Predoctoral Students:	Carla Obradors, Susanna Molas, Davide D'Amico, Meritxell Pons
Postdoctoral Fellows:	Xavier Gallego Moreno, Monica Joana Pinto Do Santos
Technicians:	María Martínez de Lagrán Cabredo, Nuno Vasconcelos
Mouse Phenotyping:	Ignasi Sahún Abizanda, Jerome MacDonald
Master/Graduate Students:	Isabel Fernández Vargas, Ornella Spadoni, Eva Aparicio, Marc Murall, Patricia Carrasco, Mariona Font, Cristina Vilella, Aida Regi

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.

RESEARCH PROJECTS

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

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

a) We have performed a large-scale analysis of gene interactions involved in cognitive phenotype using genetically engineered aneuploid mice for overlapping mouse chromosomes homologous to HSA21 (Lopes- Pereira et al 2009, Sahún et al in preparation, coll Ionas Erb/Cedric Notredame). We have also continued the characterization of single gene transgenic models focusing on the fine dissection of different forms of learning and memory (Azkona et 2010, Dierssen et al 2011) and at the molecular levels (Azkona, PhD thesis; Toiber et al 2010; Sisley-Lejeune award 2010).

b) We are interested in dissection of the molecular mechanisms affecting neuronal network formation in primary cortical cultures (Martínez de Lagrán et al submitted) and the possible involvement of Rho signaling and actin cytoskeleton (EU project granted; M Dierssen, coordinator).

c) Epigenetic contributions to the functional and structural anomalies in DS. We have analyzed histone deacetylation patterns and DNA methylation patterns (coll. M Esteller) and the phenotypes of cortical neurons (Toiber et al 2010, G Arqué, S Martínez, M Dierssen, in preparation).

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

e) We are also collaborating with other research groups focusing on neurodegenerative disorders (Botella-López et al 2010; Fillat et al 2010; Mellström et al in preparation).

1. Neuropsychiatric disorders

a) **Panic disorder**

We address the pathogenetic mechanisms that underlie the inability of persons with anxiety disorders to correctly identify the fear-related information using genetically modified mouse models (D'Amico, PhD thesis, D'Amico et al in preparation). We also study the common neurobiological pathways responsible for co-morbid processes such as stress and drug abuse disorders (Amador-Arjona et al 2010; Gallego et al 2010), and we analyze the structure of fear memories and the contribution of different candidate genes to the sensitivity to panicogenic/panicolitic agents (M Do Santos, in preparation).

b) **Nicotine dependence**

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

c) **Obsessive-compulsive behavioral components of obesity**

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

PUBLICATIONS

Santos M, Summavielle T, Silva-Fernandes A, Teixeira-Castro A, Duarte-Silva S, Marques F, Martins L, Dierssen M, Oliveira P, Sousa N, Maciel P.
"Monoamine deficits in the brain of Mecp2-null mice suggest the involvement of the cerebral cortex in early stages of Rett syndrome."
Neuroscience, 170(2):453-67 (2010).

Azkona G, Levanon D, Groner Y, Dierssen M.
"In vivo effects of APP are not exacerbated by BACE2 co-overexpression: behavioural characterization of a double transgenic mouse model."
Amino Acids, 39(5):1571-80, (2010).

Toiber D, Azkona G, Ben-Ari S, Torán N, Soreq H, Dierssen M.
"Engineering DYRK1A over-dosage yields Down syndrome-characteristic cortical splicing aberrations."
Neurobiol Dis, 40(1):348-59 (2010).

Gallego X, Murtra P, Zamalloa T, Canals JM, Pineda J, Amador-Arjona A, Maldonado R, Dierssen M.
"Increased opioid dependence in a mouse model of panic disorder."
Front Behav Neurosci, 3:60 (2010).

Botella-López A, Cuchillo-Ibáñez I, Cotrufo T, Su San Mok, Qiao-Xin Li, Barquero M-S, Dierssen M, Soriano E Sáez-Valero J.
"Beta-amyloid controls altered Reelin expression and processing in Alzheimer's disease."
Neurobiol Dis, 37(3):682-91 (2010).

Azkona G, Amador-Arjona A, Obradors-Tarragó C, Varea E, Arqué G, Pinacho R, Fillat C, de la Luna S, Estivill X, Dierssen M.
"Characterization of a mouse model overexpressing BACE2 reveals a new role for BACE2."
Genes Brain Behav, 9(2):160-72 (2010).

Amador-Arjona A, Delgado-Morales R, Belda X, Gagliano H, Gallego X, Keck ME, Armario A, Dierssen M.
"Susceptibility to stress in transgenic mice overexpressing TrkC, a model of panic disorder."
J Psychiatr Res, 44(3):157-67 (2010).

Reviews

Gardiner K, Herault Y, Lott IT, Antonarakis S, Reeves R, Dierssen M.
"Down Syndrome: From Understanding the Neurobiology to Therapy."
J Neurosci, 30(45):14943-14945 (2010).

Dierssen M, Lott I.
"Cognitive deficits and associated neurological complications in individuals with Down's syndrome."
Lancet Neurol, 9(6):623-33 (2010).

Fillat C, Dierssen M, Martínez de Lagrán M, Altafaj X.
"Insights from Mouse Models to Understand Neurodegeneration in Down Syndrome."
CNS & Neurolog Disord Drug Targets, 9(4):429-38 (2010).

Other published work

Dierssen M.
"Una visión del arte desde la neurobiología de la discapacidad intelectual. Programas de educación especial de centros de arte contemporáneo."
Centre d'Art la Panera de Lleida, in press.

GENES AND DISEASE

Group: Genomic and Epigenomic Variation in Disease

Group structure:

Group Leader: Stephan Ossowski (since July 2010)

Technicians: Daniela Bezdan (since July 2010)

SUMMARY

Next Generation Sequencing (NGS) techniques enable us to analyze the genome, transcriptome and epigenome of an individual in a tissue or even cell specific manner at single nucleotide resolution. Different samples can be analyzed at multiple molecular or regulatory levels using the same technology platform. This allows for identification of disease specific alterations at molecular level and will likely result in optimized individual treatment of patients. Furthermore NGS techniques can be 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 the analysis of NGS data in order to detect genomic, genic and epigenomic variation related to disease or intolerance to specific treatments. We seek to develop analysis tools for related sequencing applications including genome re-sequencing and assembly, DNA and histone modification detection, transcriptome analysis and structural variant prediction. We further aim to incorporate all algorithms into a general analysis framework, allowing for quick adaptation to the challenges of coming sequencing technologies and approaches, e.g. single molecule sequencing, longer read length and further increased throughput. The direct and easy applicability of our framework to sequencing data produced at the CRG or freely available via sequencing read archives is of particular importance for us and we make use of our experience with NGS analysis in several collaborations with experimental groups working on both prokaryotes and eukaryotes.

RESEARCH PROJECTS

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

Genomic changes can lead to epigenomic variability, which can in turn give rise to gene regulatory changes, differential gene expression and susceptibility to complex diseases. In diploid genomes quantitative changes in e.g. expression or chromatin state at heterozygous loci can be observed in an allele specific manner and can be used as molecular phenotype data in expression or activity quantitative trait loci analysis (i.e. genetic variants correlated to differential expression or regulatory activity, often termed eQTL and aQTL). Thereby allele-specific gene expression, chromatin modifications or chromatin structure can reveal cis- and trans-acting regulatory factors of genes. Comparison of epigenomic profiles (chromatin states) between tissues and individuals can give insights into the cause of complex diseases and can further be used to classify disease specific genomic variation in currently non-annotated (intergenic) regions. Importantly chromatin state and expression patterns vary across tissues and can undergo drastic changes in disease cells, unveiling tissue and disease specific gene regulatory activity [2]. Recently introduced Next Generation Sequencing (NGS) technologies allow for deep sequencing of biological samples and accurate quantitative analysis of e.g. gene expression, protein binding affinity and chromatin modification at single nucleotide resolution, providing a rich resource for detection of genetic variation and correlated allele specific variability in gene regulation.

The aim of this project is to study correlations between genomic and epigenomic variability in neurological diseases using whole genome and exome sequencing, DNA-methylation and histone modification data (BS-seq, ChIP-seq), open-chromatin maps (DNAsel-seq) and chromosome conformation capture (ChIA-PET). In the first phase of the project we will produce and analyze NGS data to build an in-depth atlas of the human brain epigenome, encompassing transcriptome* and chromatin state profiles for seven tissues of healthy and diseased brains. Following a pilot study on healthy human brain samples we will expand the project to study the genome and epigenome of brains affected by Alzheimer's disease and Parkinson's disease, the two most prevalent neurodegenerative disorders.

In order to identify eQTL and aQTL we will first develop tools to detect and quantify allelic imbalances of transcriptional and epigenomic signals at heterozygous genetic variants within the same individual. Using algorithms proposed by Ernst et al. and Kharchenko et al. we would further 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. Disease specific epigenomic patterns could be interesting biomarkers for early disease diagnosis, prediction of treatment outcome, determination of causal genomic loci, and development of new treatment methods. Comparison of chromatin state and expression patterns between healthy and disease tissue might also shed light on the functional implications of intergenic variants, predicted to be

linked with complex diseases by e.g. aQTL analysis or by genome-wide association studies (GWAS). Thus, the project should provide useful information to GWAS performed on several clinical phenotypes. Because we will work on samples that have been detected at early stages of the disease through collaborative work with brain banks and experienced neuropathologists (Isidre Ferrer, CIBERNED, Research Network on Neurodegenerative Disorders), we should be able to identify early markers of disease status. This should result in the identification of biomarkers of disease development with potential application in clinical settings. Due to the universal language of epigenetic changes that will be explored in this project, we anticipate that the developed methods and analysis framework will also be applicable to the study of other complex human diseases as well as cancer. The work on the brain epigenome atlas is still in the planning phase. Sequencing of samples will start in 2011.

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

In this project we will develop novel methods for detection of structural variants (SV), copy number variants (CNV) and movement of transposable elements using Next Generation Sequencing (NGS) data from chronic lymphocytic leukemia tumors (CLL) sequenced as part of the International Cancer Genome Consortium. Further we will combine information across multiple tumor and normal samples from hundreds of CLL patients sequenced using multiple strategies (genome or exome sequencing, RNA-seq) to study driver and passenger mutations of the tumor, mutation heterogeneity as well as mutations influencing metastasis. The group is participating in the International Cancer Genomics Consortium (ICGC) with the aim to fully characterize the genome of chronic lymphocytic leukemia (CLL). The role of our group is to optimize the computational analysis of NGS data in order to obtain high quality micro-indel and structural variant predictions in normal and leukemic cells from patients with CLL. The Spanish contribution to the ICGC is a collaborative effort between several centers in Spain (Clinic Hospital, University of Oviedo and CNIO, among others). We have made substantial progress in implementing a new algorithm to detect small to medium sized indels (1 to 200bp), which are hard to detect using paired-end mapping (PEM) based strategies. We have further set up a standardized analysis pipeline including SNP, micro-indel, SV and CNV prediction, which is currently applied to data from 25 patients.

3. Transcriptome and Exome Sequencing: Key Approaches to Study Human Disease and Cancer

Human exome analysis using Next Generation Sequencing (NGS) has recently been established as a key approach to identify genetic variation causal for rare mendelian diseases. Our group in collaboration with the Estivill lab is adapting this approach to analyze hundreds of exomes from patients with complex diseases, e.g. chronic lymphocytic leukemia, psoriasis, chronic fatigue syndrome, obsessive–compulsive disorder, Alzheimer and Parkinson disease. In parallel Transcriptome sequencing (RNA-seq) will be applied to detect changes in gene regulation and expression including allele-specific regulatory changes and expression quantitative trait loci (eQTL) as well as gene fusions. In this project we will develop and optimize methods and algorithms to detect genomic variants like SNPs, indels, inversions and copy number variants from both Exome-seq and RNA-seq data and to identify eQTLs. We have made substantial progress in establishing a standardized analysis pipeline for exome sequencing, which is currently used to analyze data from the first disease (Psoriasis) and exome data from five centenarians (people above age of 100).

4. Determining Genetic Factors and Metagenomic Alterations Related to Staphylococcus aureus Infection

In this project we will study genetic factors influencing *S. aureus* infections responsible for significant morbidity and mortality in community and health care settings due to increasing frequency of antibiotic resistance. Next Generation Sequencing (NGS) technology allows for a detailed analysis of the genes and variants correlated to e.g. pathogenicity and antibiotic resistance. New algorithm will be developed to analyse genome and transcriptome NGS data from S. aureus as well as data from 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, which are currently being assembled and compared in order to detect variants causal for pathogenicity.

PUBLICATIONS

Articles

Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, et al.
"The Rate and Molecular Spectrum of Spontaneous Mutations in Arabidopsis thaliana."
Science, 327:92-94 (2010). (*)

Khraiwesh B, Arif MA, Seumel GI, Ossowski S, Weigel D, et al.
"Transcriptional control of gene expression by microRNAs."
Cell, 140:111-122 (2010). (*)

Piazza P, Bailey CD, Cartolano M, Krieger J, Cao J, et al.
"Arabidopsis thaliana leaf form evolved via loss of KNOX expression in leaves in association with a selective sweep."
Curr Biol, 20:2223-2228 (2010). (*)

Laitinen RA, Schneeberger K, Jelly NS, Ossowski S, Weigel D.
"Identification of a spontaneous frame shift mutation in a non-reference Arabidopsis thaliana accession using whole genome sequencing."
Plant Physiol, 153(2):652-4 (2010). (*)

Book chapter

Schwab R, Ossowski S, Warthmann N, Weigel D
"Directed gene silencing with artificial microRNAs."
Methods Mol Biol, 592:71-88 (2010). (*)

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

BIOINFORMATICS AND GENOMICS

Coordinator: Roderic Guigó

Researchers at the Bioinformatics and Genomics programme use computation in genome research. The programme includes five groups, those lead by Roderic Guigó, Cedric Notredame, Toni Gabaldón, Fyodor Kondrashov, and Gian Gaetano Tartaglia.

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

Current structure of the programme:

5 Research Groups:

> Genome Bioinformatics (Roderic Guigó)
> Comparative Bioinformatics (Cédric Notredame)
> Comparative Genomics (Toni Gabaldón)
> Evolutionary Genomics (Fyodor Kondrashov)
> Gene Function and Evolution (Gian Gaetano Tartaglia)

BIOINFORMATICS AND GENOMICS

Group: Bioinformatics and Genomics

Group structure:

Group Leader: Roderic Guigó

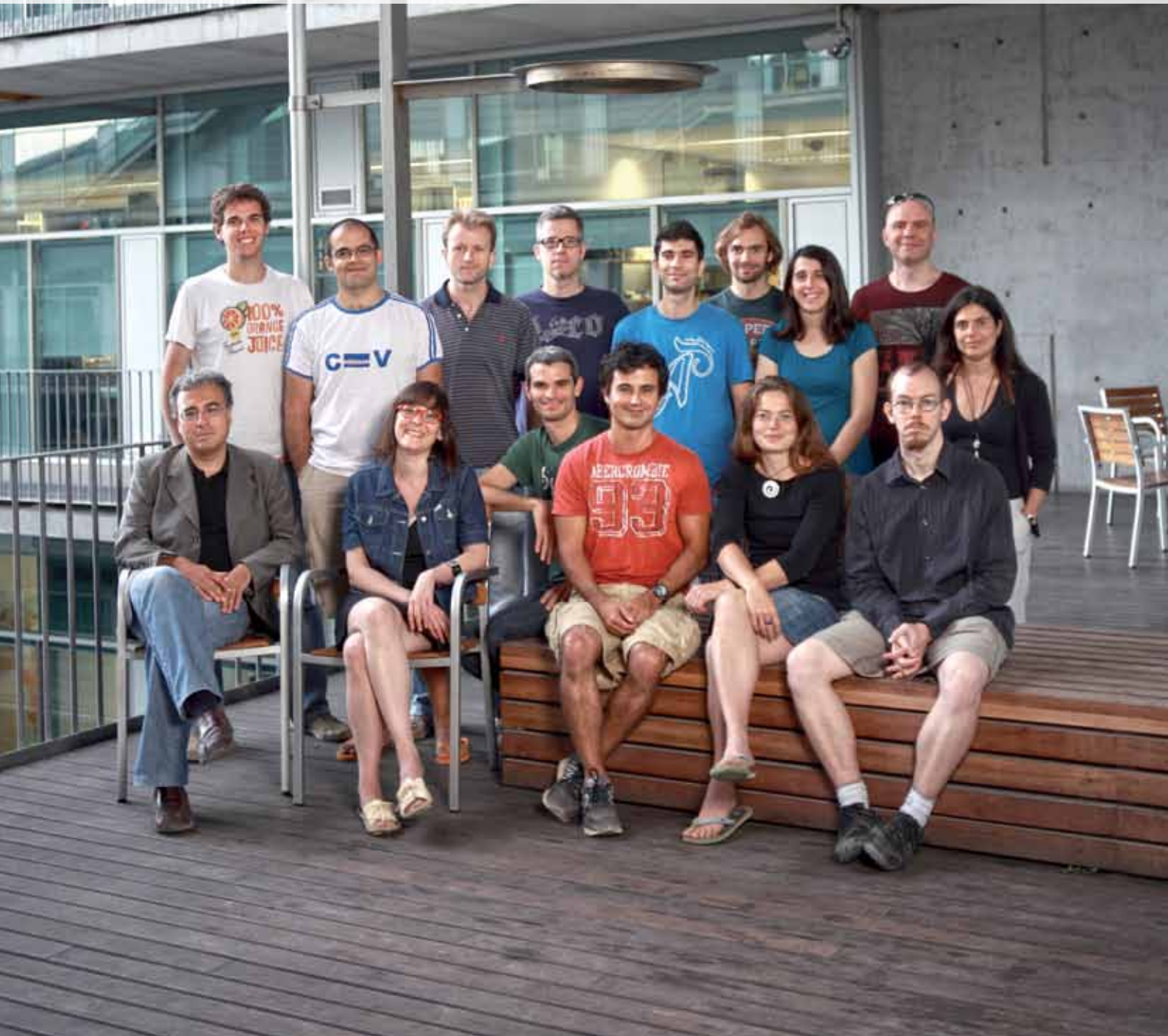
Staff scientist: Rory Johnson

Postdoctoral Fellows: Sarah Djebali, Tyler Alioto, David Martin, Michael Sammeth, Pedro Gabriel Dias Ferreira, Thomas Derrien, Paolo Ribeca, David González, Angelika Merkel, Andrea Tanzer

PhD Students: Hagen Tilgner, Anna Kedzierska, Marco Mariotti, Joao Curado, Jorgen Skancke, Maria Ortiz

Technicians: Julien Lagarde, Francisco Câmara, Colin Kingswood, Maik Röder, Carmen Arnán

Students: Alexis Grimaldi, Nadine Richter, Josue Curto, Sergi Sayols, Cinzia de Benedictis



SUMMARY

Research in the Genome Bioinformatics group focuses in the development and application of methods to identify functional domains in genomic sequences, Our group is involved in both the development of software for gene prediction in genome sequences, and in the investigation of the signals involved in gene specification. Our group has actively participated in the analysis of many eukaryotic genomes and it in involved in NIH funded ENCODE project

RESEARCH PROJECTS

1. Characterization of long non-coding RNAs

While they are poorly understood, long non-coding RNAs (lncRNAs) are emerging as central players in cell biology. lncRNAs are long, multiexonic transcripts, often polyadenylated, that exhibit the epigenetic marks typical of transcribed regions (Guttman, Amit et al. 2009). They show low level of evolutionary sequence conservation, although higher than that of the genomic background(Marques and Ponting 2009), They may be as numerous as protein coding genes (Harrow, Denoeud et al. 2006), but the precise biological role of the vast majority is unknown. We have recently shown that many lncRNAs may function by activating the expression of nearby protein-coding genes (Orom, Derrien et al. 2010). Indeed, we analyzed a set of 3,019 lnRNAs not overlapping protein-coding loci from GENCODE (REFS). Using RNAseq we have shown that lnRNAs display differential tissue expression, which is closely paralleled with their associated active or repressive chromatin signatures. Importantly, we found that the expression of LnRNAs is, in general, positively correlated with the expression of of protein coding genes in the neighborhood. Functional analysis of a number lnRNAs in multiple cell lines confirmed their role as enhancers of the expression of neighbour protein coding genes (Figure 1).

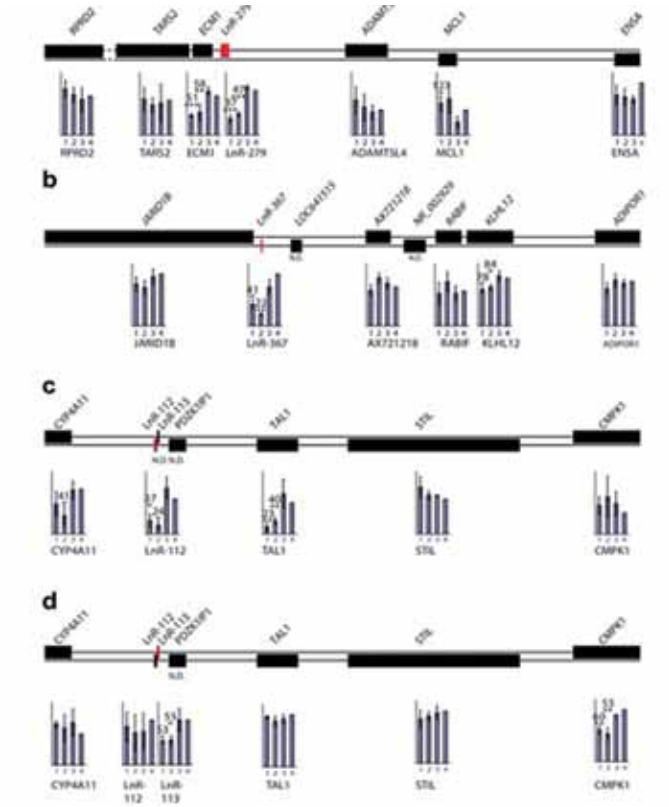


Figure 1.
*Knock-down analysis on LnRNA loci. The targeted LnRNAs are shown in red. In the bar-plots 1 and 2 represents specific siRNAs to the LnRNA in the depicted locus, 3 and 4 represents control siRNAs. a. Depletion of lnRNA LnR-279 locus in HEK293 cells results in a concomitant decrease in the expression of the neighboring EMC1 gene. Loss-of-function mutations of ECM1 results in lipoid proteinosis, a rare autosomal recessive genodermatosis24.b. Depletion of the LnR-367 locus in HeLa cells resulted in depletion of the opposite strand of the gene KLHL12 a gene known for its negative regulation of the Wnt-beta catenin pathway c. Depletion of the LnR-112 locus in MCF-7 cells resulted in a specific and potent reduction of TAL1 expression. d. Depletion of the LnR-113 locus in Jurkat cells resulted in a consistent and significant decrease in CMPK1 levels. LnR-112 and LnR-113 map within 300bp of the SCL locus, a basic helix-loop-helix protein which serves as the master regulator of hematopoiesis24. All values are relative to Gapdh expression and normalized to siRNA 4 set to 1. Error bars show standard deviation. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ by two-tailed Student's T-test.*

2. Transcriptome reconstruction from RNAseq experiments

Recent technical developments in massively parallel sequencing (also know as Next Generation Sequencing, NGS) seem to provide for the first time the throughput required for accurate transcriptome characterization. Indeed, a number of studies have demonstrated that deep RNA sequencing (RNASeq) provides access to a similar, if not greater dynamic range than microarrays, with the advantage that it allows for detection of previously unknown transcripts and the quantification of alternative transcript isoforms. RNASeq is thus quickly becoming the "de facto" standard for transcriptome characterization, spreading to very diverse application domains,. The challenges to deal with RNASeq data are, however, formidable, and they have triggered active research in Computational Biology: from short read sequence mapping (see (Li and Homer 2010) for a review) to transcript reconstruction (Guttman, Garber et al. 2010; Trapnell, Williams et al. 2010) and transcript quantification (Jiang and Wong 2009; Zheng and Chen 2009). Our group has devoted substantial effort to develop very efficient tools for RNASeq analysis. These include the GEM read aligner (*http://big.crg.cat/services/gem_genome_multi_tool_library*), the Flux Capacitor (*http://big.crg.cat/services/flux_capacitor*) for transcript quantification, and NextGeneid for "de novo" transcript modeling and discovery. The tools are still under development, but they are already part of the analysis pipeline of the ENCODE project (TheENCODEprojectConsortium 2011), and we have used them in a number of other projects. For instance, in collaboration with Manolis Dermitzakis, from the University of Geneva, we have used the Flux Capacitor to produce transcript quantifications from RNASeq data obtained in 60 individuals from the HapMap project(Montgomery, Sammeth et al. 2010). This has allowed, for the first time, the identification of eQTL responsible for the variation of abundance of individual splice forms (Figure 2).

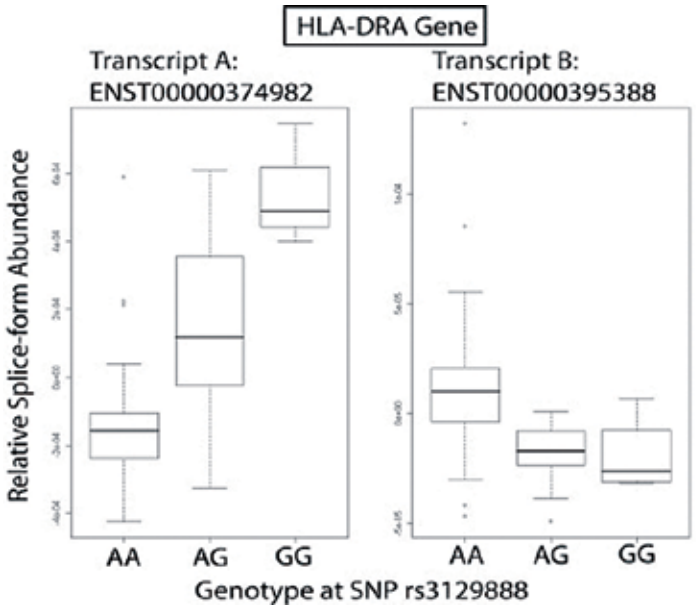


Figure 2.
eQTL with opposite effect in two alternative splice forms of the HLA-DRA gene.

3. Nucleosome positioning

The search of sequence prerequisites for nucleosome positioning has been a long-standing problem at the intersection of chromatin structure and gene regulation. With the use of structural information directly related to the primary DNA sequence we propose a new property of natural nucleosome forming sequences, which is inherently related to their intrinsic curvature and its symmetry. A measure of this property has been introduced and a corresponding method validated against novel high quality datasets of human nucleosomes, obtained through close collaboration with CRG's Chromatin and Gene Expression Group (Miguel Beato). Based on this symmetrical curvature property, we have implemented a computational ab initio method for nucleosome positioning prediction and have shown this method to be more efficient than previously published ab initio methods aiming at the same goal. The aforementioned method is already being applied in a number of analyses related to nucleosomal patterns of regions of interest in various organisms, and have lead to the discovery of structural constraints in consistent nucleosome positions. Based on these observations we have proposed a model for nucleosome positioning sequences(Nikolaou, Althammer et al. 2010) (Figure 3)

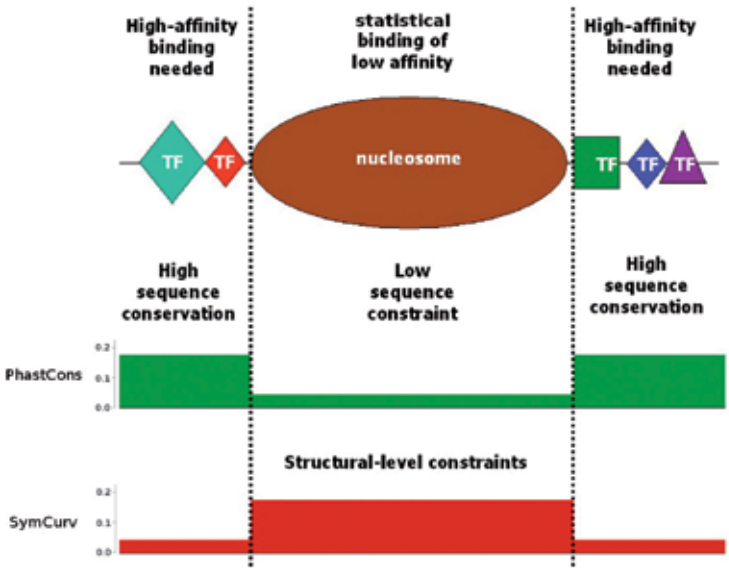


Figure 3.
***A model for nucleosome positioning sequences**. Vertical dotted lines represent consistent nucleosome boundaries. Green and red filled plots with arbitrary axes are qualitative representations of sequence conservation and SymCurv, respectively, shown here to allow better description of the model (actual data presented in Figure 4). High-affinity binding is necessary for regulatory elements occupying the surrounding space and allowing nucleosomes to occupy regions of low sequence constraint. The consistent positioning of these nucleosomes does not require high affinity but imposes a secondary level of constraints of a structural type on the underlying sequences. The symmetry of DNA curvature may be seen as one constraint of this type.*

4. Epigenomic determinants of splicing

We have found evidence that strongly suggests a direct involvement of chromatin structure in RNA splicing. We have analyzed high throughput sequencing data recently produced by Schones et al. (2008) on positioning of nucleosomes in resting and activated human CD4+ T cells. Our analyses show evidence for stable nucleosome occupancy within internal exons of human genes, a pattern absent in pseudoexons (non-repetitive intronic sequences flanked by strong splice sites), which in contrast exhibit a weak nucleosome depletion (Figure 3). Remarkably, this pattern depends on the strength of the splice sites. Indeed, the nucleosome peak is accentuated within exons with weak splice sites, while in the exons with strong splice sites, as well as in pseudoexons, a region of stable nucleosome occupancy is observed upstream of the acceptor sites. As a result, pseudoexons with strong splice sites—in which splicing does not occur despite the strength of the sites—show a pattern of nucleosome occupancy which is the mirror image of that observed on exons with weak sites—in which splicing occurs despite the weakness of the sites. These observations strongly suggest that positioning of nucleosomes influences RNA splicing(Tilgner, Nikolaou et al. 2009), and we have suggested a model in which nucleosomes would influence splicing through transcription(Tilgner and Guigo 2010) (Figure 4).

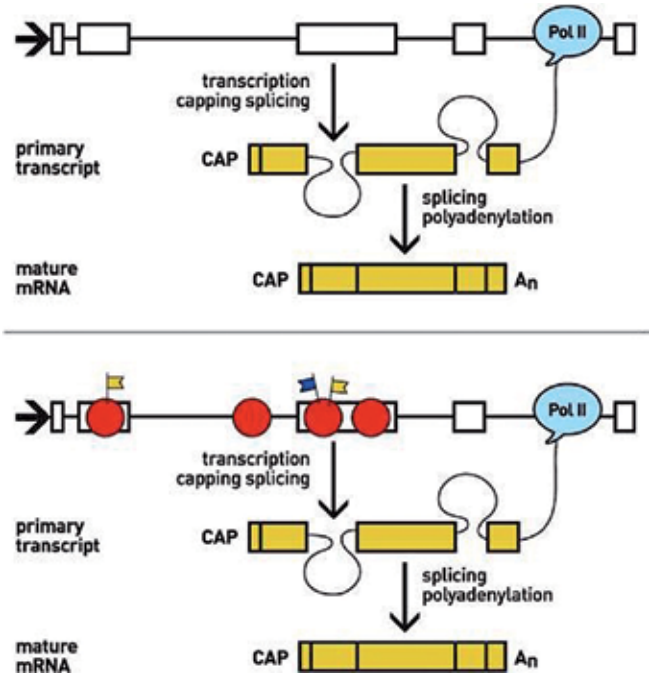


Figure 4.
Chromatin involvement in co-transcriptional splicing.
View of co-transcriptional splicing as proposed and drawn by Kornblihtt et al.7 Splicing is co-transcriptional, allowing transcription kinetics and thereby broad chromatin structure (e.g., open/closed chromatin state) to influence splicing (top). Extension of the top-panel-model, in which stable positioning of possibly single nucleosomes and histone modifications can also influence splicing. in particular nucleosome occupancy within exons will contribute to their inclusion in the mature transcript.

5. Identification of selenoproteins

Selenoproteins are a group of proteins that contain selenocysteine (Sec), a rare amino acid inserted co-translationally into the protein chain. The Sec codon is UGA, which is normally a stop codon. In selenoproteins, UGA is recoded to Sec in presence of specific features on selenoprotein gene transcripts. Due to the dual role of the UGA codon, selenoprotein prediction and annotation are difficult tasks, and even known selenoproteins are often misannotated in genome databases. Our group has had a long standing interest in selenoprotein identification and evolution. We have recently developed present an homology-based *in silico* method to scan genomes for members of the known eukaryotic selenoprotein families: selenoprofiles. The core of the method is a set of manually curated highly reliable multiple sequence alignments of selenoprotein families, which are used as queries to scan genomic sequences (Figure 5). Results of the scan are processed through a number of steps, to produce highly accurate predictions of selenoprotein genes with little or no human intervention. Selenoprofiles is a valuable tool for bioinformatic characterization of eukaryotic selenoproteomes, and can complement genome annotation pipelines (Mariotti and Guigo 2010)

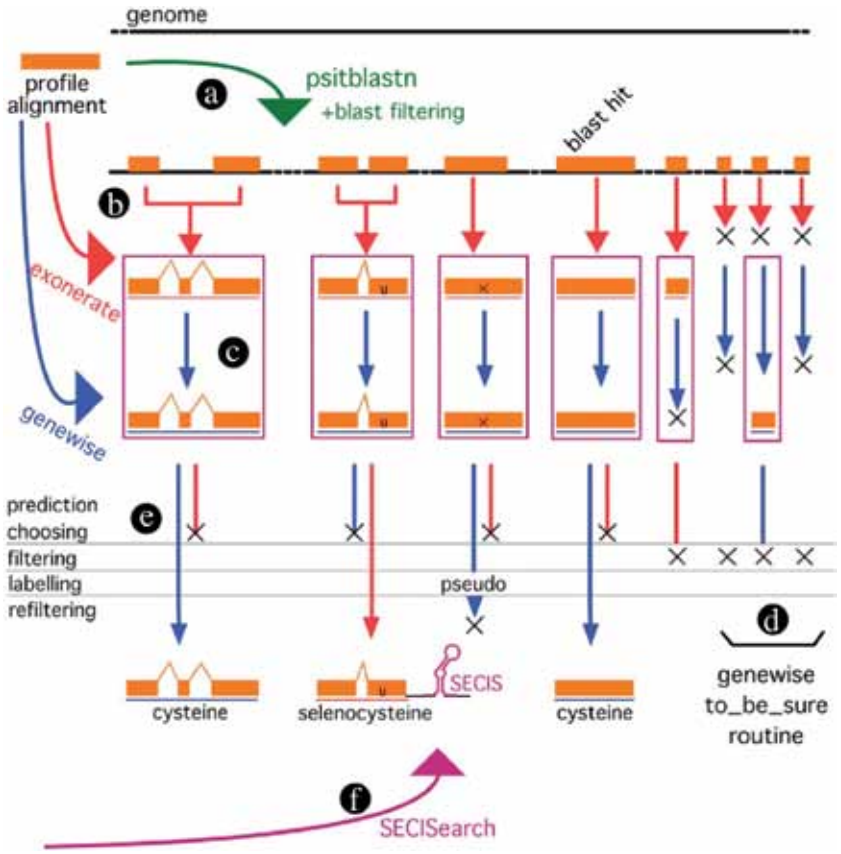


Figure 5.
Schema of the selenoprofiles pipeline. *Initially, a psitblastn search is run using a PSSM built from the profile alignment (a). The resulting genomic intervals are merged into 'superexon' intervals, and cyclic exonerate is run on each of them (b). Then, genewise is run both to refine exonerate predictions (c) and when exonerate failed recovering blast alignments (d—genewise 'to be sure' routine). The exonerate or the genewise prediction is chosen (e), and then results are filtered, labeled and then refiltered with family-specific filters. Lastly, SECISearch is used to detect potential SECIS elements downstream of the genes (f).*

6. ENCODE Project

We are part of the NHGRI funded ENCODE project. An NHGRI funded project which has as a goal the identification of all functional domains in the human genome. We participated in the grants lead by Tom Gingeras (Cold Spring Harbour Laboratory) to characterize the transcriptional activity of the human genome, and by Tim Hubbard (Welcome Trust Sanger Institute), where we contribute to build the reference annotation of the human genome. We lead the ENCODE RNA analysis working group. A paper describing the data produced by the ENCODE project has just been published (The ENCODE project Consortium 2011), and during year 2011 we plan to publish the main biological results of the project.

References:

Guttman, M., I. Amit, et al. (2009).
"Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals."
Nature 458(7235): 223-7.

Guttman, M., M. Garber, et al. (2010).
"Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs."
Nat Biotechnol 28(5): 503-10.

Harrow, J., F. Denoeud, et al. (2006).
"GENCODE: producing a reference annotation for ENCODE."
Genome Biol 7 Suppl 1: S4 1-9.

Jiang, H. and W. H. Wong (2009).
"Statistical inferences for isoform expression in RNA-Seq."
Bioinformatics 25(8): 1026-32.

Li, H. and N. Homer (2010).
"A survey of sequence alignment algorithms for next-generation sequencing."
Brief Bioinform 11(5): 473-83.

Mariotti, M. and R. Guigo (2010).
"Selenoprofiles: profile-based scanning of eukaryotic genome sequences for selenoprotein genes."
Bioinformatics 26(21): 2656-63.

Marques, A. C. and C. P. Ponting (2009).
"Catalogues of mammalian long noncoding RNAs: modest conservation and incompleteness."
Genome Biol 10(11): R124.

Montgomery, S. B., M. Sammeth, et al. (2010).
"Transcriptome genetics using second generation sequencing in a Caucasian population."
Nature 464(7289): 773-7.

Nikolaou, C., S. Althammer, et al. (2010).
"Structural constraints revealed in consistent nucleosome positions in the genome of S. cerevisiae."
Epigenetics Chromatin 3(1): 20.

Orom, U. A., T. Derrien, et al. (2010).
"Long noncoding RNAs with enhancer-like function in human cells."
Cell 143(1): 46-58.

The ENCODE project Consortium (2011).
"A User's Guide to the Encyclopedia of DNA Elements (ENCODE)."
PLoS Biol in press.

Tilgner, H. and R. Guigo (2010).
"From chromatin to splicing: RNA-processing as a total artwork."
Epigenetics 5(3).

Tilgner, H., C. Nikolaou, et al. (2009).
"Nucleosome positioning as a determinant of exon recognition."
Nat Struct Mol Biol 16(9): 996-1001.

Trapnell, C., B. A. Williams, et al. (2010).
"Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation."
Nat Biotechnol 28(5): 511-5.

Zheng, S. and L. Chen (2009).
"A hierarchical Bayesian model for comparing transcriptomes at the individual transcript isoform level."
Nucleic Acids Res 37(10): e75.

PUBLICATIONS

International Aphid Genomics Consortium (including Guigó R and Câmara F).
"Genome sequence of the pea aphid Acyrthosiphon pisum".
PLoS Biol, 8(2):e1000313 (2010).

Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, Guigo R, Dermitzakis ET.
"Transcriptome genetics using second generation sequencing in a Caucasian population".
Nature, 464(7289):773-7 (2010).

International Cancer Genome Consortium (including Guigó R).
"International network of cancer genome projects".
Nature, 464(7291):993-8 (2010).

Tilgner H, Guigó R.
"From chromatin to splicing: RNA-processing as a total artwork".
Epigenetics, 2010 Apr 27;5(3). [Epub ahead of print]

Novoa I, Gallego J, Ferreira PG, Mendez R.
"Mitotic cell-cycle progression is regulated by CPEB1 and CPEB4-dependent translational control".
Nat Cell Bio, 12(5):447-56 (2010).

Vieira J, Ferreira PG, Aguiar B, Fonseca NA, Vieira CP.
"Evolutionary patterns at the RNase based gametophytic self - incompatibility system in two divergent Rosaceae groups (Maloideae and Prunus)".
BMC Evol Biol, 10:200 (2010).

Stajich JE, Wilke SK, Ahrén D, Au CH, Birren BW, Borodovsky M, Burns C, Canbäck B, Casselton LA, Cheng CK, Deng J, Dietrich FS, Fargo DC, Farman ML, Gathman AC, Goldberg J, Guigó R, Hoegger PJ, Hooker JB, Huggins A, James TY, Kamada T, Kilaru S, Kodira C, Kües U, Kupfer D, Kwan HS, Lomsadze A, Li W, Lilly WW, Ma LJ, Mackey AJ, Manning G, Martin F, Muraguchi H, Natvig DO, Palmerini H, Ramesh MA, Rehmeyer CJ, Roe BA, Shenoy N, Stanke M, Ter-Hovhannisyan V, Tunlid A, Velagapudi R, Vision TJ, Zeng Q, Zolan ME, Pukkila PJ.
"Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom Coprinopsis cinerea (Coprinus cinereus)".
Proc Natl Acad Sci U S A, 107(26):11889-94 (2010).

Mariotti M, Guigó R.
"Selenoprofiles: profile-based scanning of eukaryotic genome sequences for selenoprotein genes".
Bioinformatics, 26(21):2656-63 (2010).

Manichanh C, Reeder J, Gibert P, Varela E, Llopis M, Antolin M, Guigo R, Knight R, Guarner F.
"Reshaping the gut microbiome with bacterial transplantation and antibiotic intake".
Genome Res, 20(10):1411-9 (2010).

Ørom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, Guigo R, Shiekhattar R.
"Long noncoding RNAs with enhancer-like function in human cells".
Cell, 143(1):46-58 (2010).

Arensburger P, Megy K, Waterhouse RM, Abrudan J, Amedeo P, Antelo B, Bartholomay L, Bidwell S, Caler E, Camara F, Campbell CL, Campbell KS, Casola C, Castro MT, Chandramouliswaran I, Chapman SB, Christley S, Costas J, Eisenstadt E, Feschotte C, Fraser-Liggett C, Guigo R, Haas B, Hammond M, Hansson BS, Hemingway J, Hill SR, Howarth C, Ignell R, Kennedy RC, Kodira CD, Lobo NF, Mao C, Mayhew G, Michel K, Mori A, Liu N, Naveira H, Nene V, Nguyen N, Pearson MD, Pritham EJ, Puiu D, Qi Y, Ranson H, Ribeiro JM, Roberston HM, Severson DW, Shumway M, Stanke M, Strausberg RL, Sun C, Sutton G, Tu ZJ, Tubio JM, Unger MF, Vanlandingham DL, Vilella AJ, White O, White JR, Wondji CS, Wortman J, Zdobnov EM, Birren B, Christensen BM, Collins FH, Cornel A, Dimopoulos G, Hannick LI, Higgs S, Lanzaro GC, Lawson D, Lee NH, Muskavitch MA, Raikhel AS, Atkinson PW.
"Sequencing of Culex quinquefasciatus establishes a platform for mosquito comparative genomics".
Science, 330(6000):86-8 (2010).

Nikolaou C, Althammer S, Beato M, Guigo R.
"Structural constraints revealed in consistent nucleosome positions in the genome of S. cerevisiae".
Epigenetics Chromatin, 3(1):20 (2010).

BIOINFORMATICS AND GENOMICS

Group: Comparative Bioinformatics

Group structure:

Group Leader: Cédric Notredame

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

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

Technician: 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: *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.[1-9]

RESEARCH PROJECTS

1. Improvement of Multiple Sequence Alignments

Our lab has pioneered the development of a new generation of multiple sequence alignment methods, based on consistency. We keep working actively on this project, addressing the many new issues that are arising from the rapid development of large-scale genomics. Indeed, in order to keep with the current pace of data production, methods will not only need to be accurate but they will also need to be able to scale with the data. Interestingly the recent availability of large CPU capacity ove the cloud means that the resources are there and that available methods need to be modified in order to make the best of these new resources. We have therefore started adapting massively parallelizing T-Coffee so that it can be deployed onto the cloud.

Table 1.

Summary of the CPU/Cost for running RV11 onto the Amazon EC2.

Instance type	Features	ECU	Scale	Speedup (%)	Total time	Total cost ($)
Small	1 CPU – 1.7 GB – 32 bit	1	1	100	2272	0.054
Extra large	4 CPU – 15 GB – 64 bit	8	6.5	81	359	0.068
Double Extra large	4 CPU – 32 GB – 64 bit	13	9.0	88	257	0.086
Quad. Extra large	8 CPU – 64 GB – 64 bit	26	12.4	47	200	0.133
Medium	2 CPU – 1.7 GB – 32 bit	5	3.8	76	607	0.029
Extra large	8 CPU – 7 GB – 64 bit	20	9.2	46	266	0.050

2. Long Non Coding RNA analysis

We are working actively on the development of efficient and accurate methods for the ab-initio discovery and the multiple comparisons of long-non coding RNAs. Over the last years we have applied the methods developed in the lab to help confirming the existence of a functionally active secondary in *Drosophila* mRNA, involved in a new polyadenylation mechanism. We have also contributed to the discovery of a novel class of long non-coding RNAs exhibiting enhancer-like functions and are now focusing our efforts on the improvement of computational methods designed to study the evolution and the conservation of these sequences across genomes.

3. Structure based Protein Comparisons

The relationship between structural and sequence evolution is complex. On the one hand, protein structures vary much less during the evolutionary process than their corresponding sequences. As a consequence, homology detection remains possible through structural comparisons at evolutionary distances where sequence analysis cannot be used to support homology analysis. This should make structure based comparison an ideal complement to standard phylogenetic analysis. Structures however evolve under such a strong pressure of selection that it can be hard to distinguish convergent and divergent evolution. Taking advantage of our capacity to produce highly accurate multiple structural based sequences alignments with the T-Coffee package we have started exploring the computation of structure based sequence clustering and establishing the merits of this type of analysis. Our approach relies on the comparison of intra-molecular distances between aligned proteins, with results suggesting that structure based comparisons can support fine grain classifications. These classification may be used to infer functional relationships, even between distantly related sequences. Because they can deal with distantly related paralogues, these comparison can be much more informative than sequence based clusterings (See below). Our efforts are now focused on determining the phylogenic merits of this type of approach, most notably by assessing how structures may be considered as slow evolving characters allowing the reconstruction of very deep nodes.

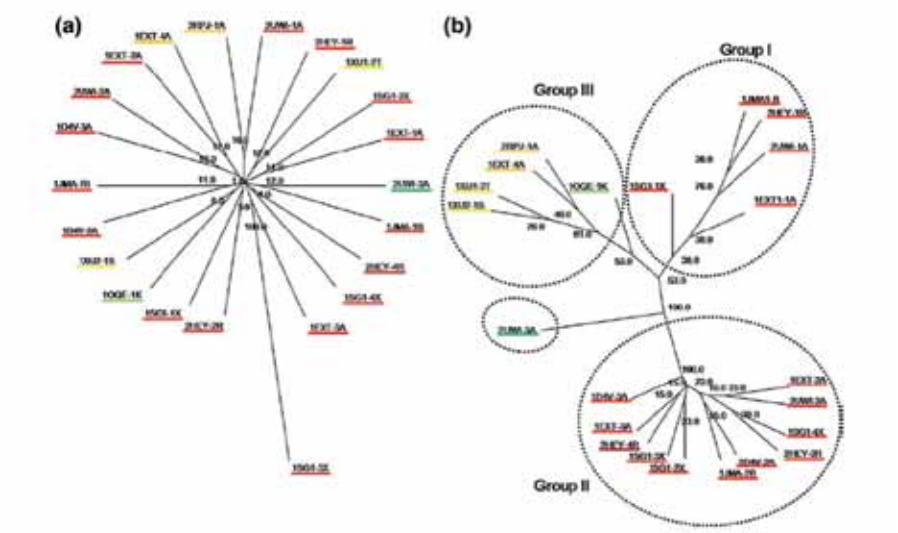


Figure 2.
Tree analysis of current module-based CRD classification compared to a new clustered based classification.
(a) Phylogenetic analysis of the 22 structures of non redundant CRDs (see Methods). The neighbor-joining tree was derived from MSA of CRDs using 3D-Coffee.
(b) T-RMSD structural clustering derived from MSA obtained by 3D-Coffee. For A and B, each CRD is represented by its PDB identifier, followed by its position number within the architecture of its corresponding TNFR and the chain identifier of the corresponding PDB file. The module composition is indicated by colored underlines: A1-B1, dark red; A1-B2, red; A1-C2, orange; A1-D2, yellow; X2-N, light green; and A2-B1, dark-green.
Bootstrap values are given for each node.

PUBLICATIONS

Yau WL, et al.
"Cyclosporin A treatment of Leishmania donovani reveals stage-specific functions of cyclophilins in parasite proliferation and viability."
PLoS Negl Trop Dis, 4(6):e729 (2010).

Orom UA, et al.
"Long noncoding RNAs with enhancer-like function in human cells."
Cell, 143(1):46-58 (2010).

Notredame C.
"Computing multiple sequence/structure alignments with the T-coffee package."
Curr Protoc Bioinformatics, Chapter 3:Unit 3 8 1-25 (2010).

Magis C, et al.
"T-RMSD: a fine-grained, structure-based classification method and its application to the functional characterization of TNF receptors."
J Mol Biol, 400(3):605-17 (2010).

Lingner T, et al.
"Predicting phenotypic traits of prokaryotes from protein domain frequencies."
BMC Bioinformatics, 11:481 (2010).

Gauthier JA, et al.
"Multichannel Sequence Analysis Applied To Social Science Data."
Sociological Methodology, 40(1):1-38 (2010).

Di Tommaso P, et al.
"Cloud-Coffee: implementation of a parallel consistency-based multiple alignment algorithm in the T-Coffee package and its benchmarking on the Amazon Elastic-Cloud."
Bioinformatics, 26(15):1903-4 (2010).

Dalloul RA, et al.
"Multi-platform next-generation sequencing of the domestic turkey (Meleagris gallopavo): genome assembly and analysis."
PLoS Biol, 8(9) (2010).

Coll O, et al.
"A novel, noncanonical mechanism of cytoplasmic polyadenylation operates in Drosophila embryogenesis."
Genes Dev, 24(2):129-34 (2010).

BIOINFORMATICS AND GENOMICS

Group: Comparative Genomics

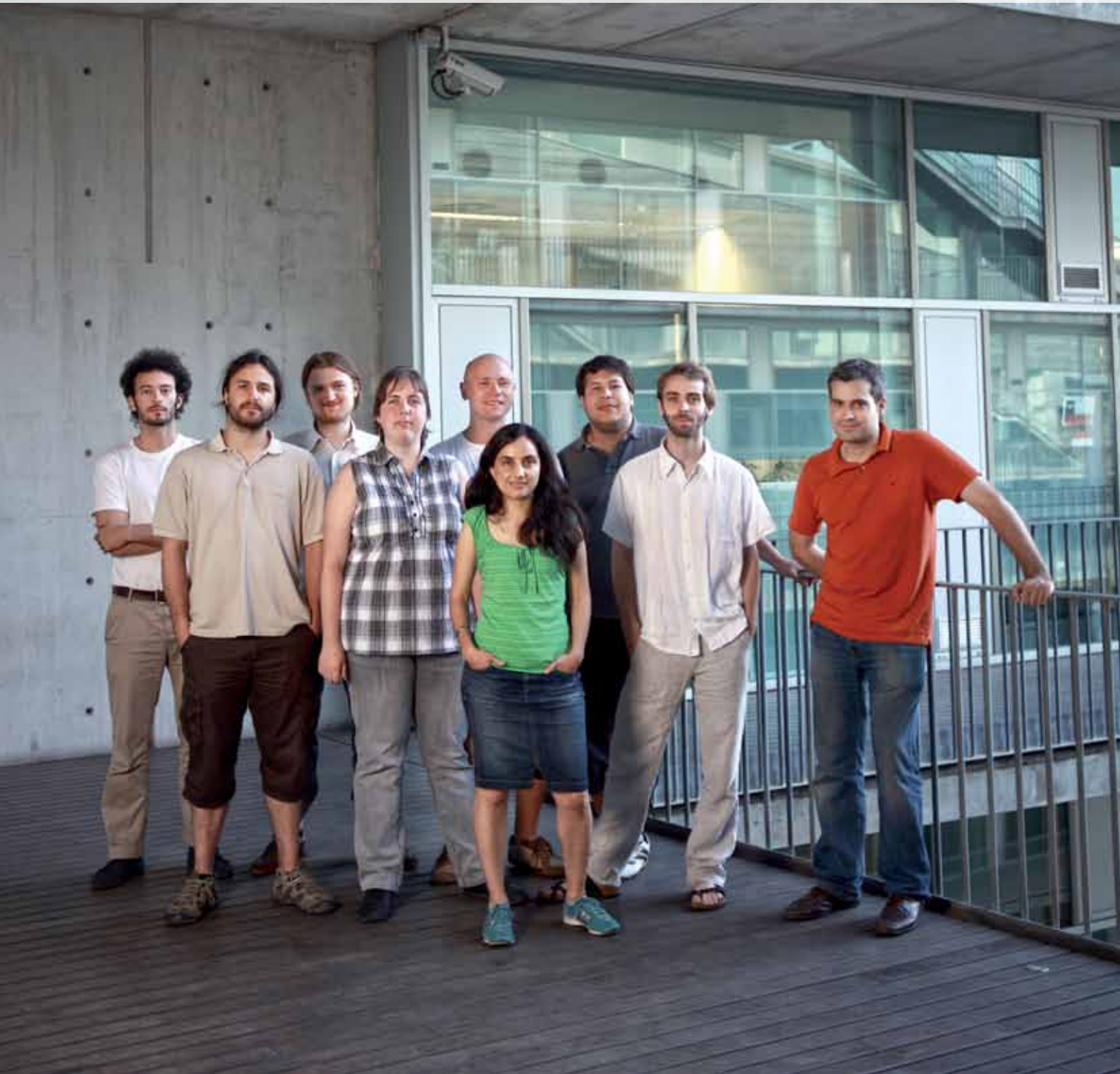
Group structure:

Group Leader: Toni Gabaldón

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

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

Technicians: Marina Marcet-Houben



SUMMARY

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

RESEARCH PROJECTS

1. Discovery of new genes involved in mitochondrial disease

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

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 performed a large-scale analysis of the impact of horizontal transfer of bacterial genes in the evolution of fungi that showed that this mechanism has been frequent in diverse lineages.

3. Comparative genomics of apoptosis and other programmed cell deaths

Programmed cell-death is a central biological process that in eukaryotes it evolved into different complex pathways. It acts as a control mechanism of homeostasis in cell number and is triggered by the onset of a coordinated biochemical cascade of events, a process called apoptosis in metazoan species. Although much progress has been made in the recent years, its origins and evolutions remain to be resolved. The aim of this project is to trace the evolution of the cell-death pathways in eukaryotes and to unravel the evolutionary relationship between caspases and mitochondria, which play a central role in the caspases-dependent apoptotic process. Furthermore we will use genomic-context techniques to discover proteins that evolved coordinately with known apoptotic components. Such predictions will be tested experimentally. This project is performed in collaboration with the lab of Dr. Cristina Muñoz-Pinedo (IDIBELL, Barcelona) We have developed the DeathBase, a database for the structure, function and evolution of apoptotic proteins (*www.deathbase.org*)

4. Phylogenomics and genome evolution

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

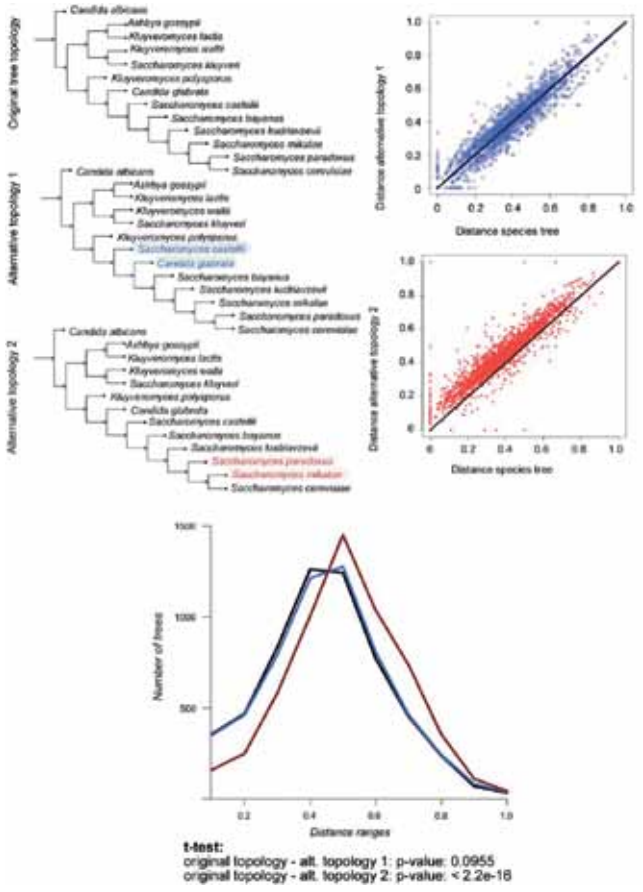


Figure 1.
Example of how our program for the comparison of phylogenetic trees (treeKO, http://treeko.cgenomics.org) allows contrasting different evolutionary hypotheses, based on the distribution of distances to a large collection of phylogenetic trees (in this case all the evolutionary histories of Saccharomyces cerevisiae genes).

5. Evolution of the eukaryotic cell

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

PUBLICATIONS

Pryszcz LP, Huerta-Cepas J, Gabaldón T.
"MetaPhOrs: orthology and paralogy predictions from multiple phylogenetic evidence using a consistency-based confidence score."
Nucleic Acids Res, Epub 2010 Dec 11.

Roetzer A, Klopf E, Gratz N, Marcet-Houben M, Hiller E, Rupp S, Gabaldón T, Kovarik P, Schüller C.
"Regulation of Candida glabrata oxidative stress resistance is adapted to host environment."
FEBS Lett, Epub 2010 Dec 13.

Huerta-Cepas J, Gabaldón T.
"Assigning duplication events to relative temporal scales in genome-wide studies."
Bioinformatics, Epub 2010 Nov 11.

Lingner T, Mühlhausen S, Gabaldón T, Notredame C, Meinicke P.
"Predicting phenotypic traits of prokaryotes from protein domain frequencies."
BMC Bioinformatics, 11:481 (2010).

Gil Borlado MC, Moreno Lastres D, Gonzalez Hoyuela M, Moran M, Blazquez A, Pello R, Marin Buera L, Gabaldon T, Garcia Peñas JJ, Martín MA, Arenas J, Ugalde C.
"Impact of the mitochondrial genetic background in complex III deficiency."
PLoS One, 5(9), pii: e12801 (2010).

Huerta-Cepas J, Capella-Gutierrez S, Pryszcz LP, Denisov I, Kormes D, Marcet-Houben M, Gabaldón T.
"PhylomeDB v3.0: an expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions."
Nucleic Acids Res, Epub 2010 Nov 12.

Benítez-Páez A, Villarroya M, Douthwaite S, Gabaldón T, Armengod ME.
"YibK is the 2'-O-methyltransferase TrmL that modifies the wobble nucleotide in Escherichia coli tRNA(Leu) isoacceptors."
RNA, 16(11):2131-43 (2010).

Gabaldón T, Capella-Gutiérrez S.
"Lack of phylogenetic support for a supposed actinobacterial origin of peroxisomes."
Gene, 465(1-2):61-5 (2010).

Huerta-Cepas J, Marcet-Houben M, Pignatelli M, Moya A, Gabaldón T.
"The pea aphid phylome: a complete catalogue of evolutionary histories and arthropod orthology and paralogy relationships for Acyrthosiphon pisum genes."
Insect Mol Biol, 19 Suppl 2:13-21 (2010).

Díez J, Walter D, Muñoz-Pinedo C, Gabaldón T.
"DeathBase: a database on structure, evolution and function of proteins involved in apoptosis and other forms of cell death."
Cell Death Differ, 17(5):735-6 (2010).

Gómez-Baldó L, Schmidt S, Maxwell CA, Bonifaci N, Gabaldón T, Vidalain PO, Senapedis W, Kletke A, Rosing M, Barnekow A, Rottapel R, Capellá G, Vidal M, Astrinidis A, Piekorz RP, Pujana MA.
"TACC3-TSC2 maintains nuclear envelope structure and controls cell division."
Cell Cycle, 2010 Mar 18;9(6). [Epub ahead of print]

International Aphid Genomics Consortium.
"Genome sequence of the pea aphid Acyrthosiphon pisum."
PLoS Biol, 8(2):e1000313 (2010).

Ahmadinejad N, Dagan T, Gruenheit N, Martin W, Gabaldón T.
"Evolution of spliceosomal introns following endosymbiotic gene transfer."
BMC Evol Biol, 10:57 (2010).

Gerardo NM, Altincicek B, Anselme C, Atamian H, Barribeau SM, de Vos M, Duncan EJ, Evans JD, Gabaldón T, Ghanim M, Heddi A, Kaloshian I, Latorre A, Moya A, Nakabachi A, Parker BJ, Pérez-Brocal V, Pignatelli M, Rahbé Y, Ramsey JS, Spragg CJ, Tamames J, Tamarit D, Tamborindeguy C, Vincent-Monegat C, Vilcinskas A.
"Immunity and other defenses in pea aphids, Acyrthosiphon pisum."
Genome Biol, 11(2):R21 (2010).

Peña A, Teeling H, Huerta-Cepas J, Santos F, Yarza P, Brito-Echeverría J, Lucio M, Schmitt-Kopplin P, Meseguer I, Schenowitz C, Dossat C, Barbe V, Dopazo J, Rosselló-Mora R, Schüler M, Glöckner FO, Amann R, Gabaldón T, Antón J.
"Fine-scale evolution: genomic, phenotypic and ecological differentiation in two coexisting Salinibacter ruber strains."
ISME J, 4(7):882-95 (2010).

Milinkovitch MC, Helaers R, Depiereux E, Tzika AC, Gabaldón T.
"2x genomes--depth does matter."
Genome Biol, 11(2):R16 (2010).

Huerta-Cepas J, Dopazo J, Gabaldón T.
"ETE: a python Environment for Tree Exploration."
BMC Bioinformatics, 11:24 (2010).

Marcet-Houben M, Gabaldón T.
"Acquisition of prokaryotic genes by fungal genomes."
Trends Genet, 26(1):5-8 (2010).

Schlüter A, Real-Chicharro A, Gabaldón T, Sánchez-Jiménez F, Pujol A.
"PeroxisomeDB 2.0: an integrative view of the global peroxisomal metabolome."
Nucleic Acids Res, 38(Database issue):D800-5 (2010).

Reviews

Roetzer A, Gabaldón T, Schüller C.
"From Saccharomyces cerevisiae to Candida glabratain a few easy steps: important adaptations for an opportunistic pathogen."
FEMS Microbiol Lett, doi: 10.1111/j.1574-6968.2010.02102.x. Epub 2010 Sep 16.

Gabaldón T.
"Peroxisome diversity and evolution."
Philos Trans R Soc Lond B Biol Sci, 365(1541):765-73 (2010).

Mihailovich M, Militti C, Gabaldón T, Gebauer F.
"Eukaryotic cold shock domain proteins: highly versatile regulators of gene expression."
Bioessays 32(2):109-18 (2010).

Other Published Work

Huerta-Cepas J, Gabaldón T.
"Del genoma al filoma."
Investigación y Ciencia, 403 (2010).

BIOINFORMATICS AND GENOMICS

Group Evolutionary Genomics

Group structure:

Group Leader: Fyodor Kondrashov

Graduate Students: Margarita Meer, Inna Povolotskaya, Onuralp Soylemez

Postdoctoral Fellows: Peter Vlasov, Romain Derelle, Tobias Warnecke

Research Assistants: Michael Breen, Maria Plyuscheva

SUMMARY

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

Most of the work that we completed in the past year has been focused on the question of the relationship between genotype and phenotype. We have pursued this question in three different forms: compensatory evolution in proteins and tRNAs, the evolution of gene duplications and the population-specificity of human mitochondrial disease mutations. The first issue involves the study of which genotypes can be traversed in the course of evolution. The second issue deals with the expected impact of gene copy number on selectable phenotype. The final issue deals with the question of how common are the polymorphisms in the human population that have an adverse effect on health.

RESEARCH PROJECTS

1. Compensatory evolution

Many genetic phenomena are described from a simple perspective of linking one gene, or one mutation, to a phenotype. We often hear in the news that a gene for cancer, or some other specific disease, has been described. However, this sort of perspective assumes that the function of a particular gene or a mutation does not depend on the function of other genes or the impact of other mutations. When we say that a mutation causes a particular disease, that is to say a genotypic change leads to a certain phenotype, we mean to say that in an individual with a specific set of genes a particular mutation causes a specific disease. However, if we take a different individual with a different set of genes that mutation may not have the same effect. The reason why this approach generally works is because humans are genetically very similar to each other. Because of this high level of similarity, mutations that cause a disease in one individual are also very likely to have the same effect to occur in another person with a different genetic makeup. But even in the human population this is not always the case, as we know that people with different genetic makeup are susceptible to different diseases. In fact, the entire concept of personalized medicine is hinged on the idea that different mutations effect different people differently depending on their genetic makeup. After all, if all genetic changes affected everyone in exactly the same way there would be no need to personalize medical treatments.

Our laboratory is interested in understanding the interactions between different mutations and how these interactions change the impact a specific mutation has on the phenotype. We tend to study this issue on the level of interactions that are observed between different species because, as mentioned previously, individuals from the same species are usually too similar to reveal many examples of such interdependence. On of our studies has been an examination of the interaction of the impact of mutations in the stem-structures of mitochondrial tRNAs. We have been able to show that the switch between complementary pairs (AU or GC) in the stem structures occurs frequently, and when it does occurs though the deleterious AC or GU intermediates. In this study, we have been able to show that the impact of the mutation in mitochondrial tRNA stem depends to a very large degree on its compensatory interaction.

We also investigated the importance of interaction of different mutations in protein structures. To do this we took proteins the ancestry of which could be traced to the common ancestor of all extant life forms, and examined the rate of evolution of these proteins. We have been able to show that despite the extremely long times of divergence of these proteins, ~3.5 billion years of independent evolution, they continue to diverge from the common ancestor. Doing a bit of theory we show that evolution cannot be so slow if the phenotypic impact of a mutation is independent from all other sites in the protein. Alternatively, we can come up with an example of the types of interactions that are necessary to lead to extremely slow evolution (Figure 1).

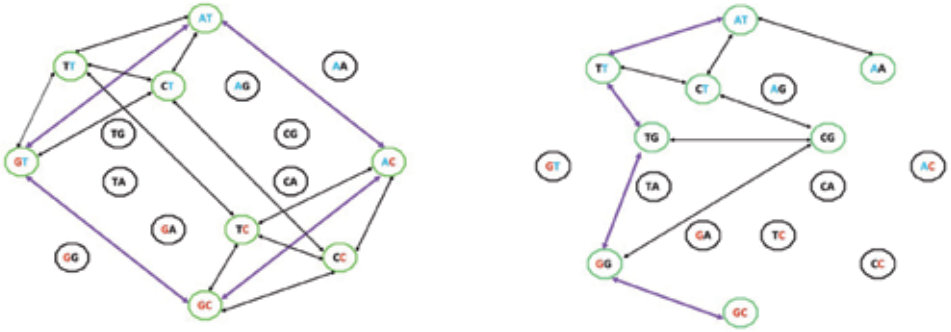


Figure 1.
Two hypothetical maps of genotype to phenotype interactions. Genotypes in green circles correspond to those that produce a viable phenotype, those in black circles represent non-viable genotypes. If A or G nucleotides in the second position are deleterious regardless of the interaction with the nucleotides in the first position then evolution between AT and GC can proceed via two shortest routes each with a length of just two substitutions (see left part of figure). This situation is different if the impact of a mutation in one site depends on the nucleotides occupying the other site, such as demonstrated in the example on the right side of the figure. In the figure, for example, the impact of nucleotide A in the second position is detrimental if the first position is occupied by T, C or G, but is benign if the nucleotide in the first position is A. In such a situation of dependence of mutations on the genetic context in which they occur, evolution between GC and AT genotypes can proceed by only one shortest route that consists of four mutations. Therefore, non-trivial interactions between mutations must slow down the rate of evolution.

2. Gene duplication

Gene duplications occur frequently in evolution and are thought to be the key to functional innovation on the molecular level. The usual verbal argument for the mechanism of the emergence of new functions after a gene duplication is the following. A single gene is present in the genome and performs its function. When this gene is duplicated its new copy is completely redundant because the function is already performed and, therefore, the second copy is unnecessary in that it does not effect the phenotype. Therefore, this second copy is released from the constraints of having to code for the original function and can evolve in random directions with some probability actually acquiring a new function. Even within this simple model there are many parameters that have important implications for the probability of evolving this new function. However, there are alternatively models that deal with how a gene duplication evolved and impacts the phenotype, including those that assume that a gene duplication just by occurring actually impacts the phenotype. One of these possibilities is that the extra gene copy leads to extra gene product being produced and this increase in dosage can have an effect on the phenotype. In the book chapter published this year I have laid out the arguments in favor of the dosage as being an important force in maintaining gene duplications in the genome. In addition, I have made an attempt at bringing some sense of order to the theoretical field of gene duplication by publishing a classification of a wide variety of theories on the emergence, maintenance and evolution of gene copies.

3. Pathogenic mutations in the mitochondria

A substantial fraction of variability in the human population contributes to the development of different pathologies. Some of these variations are found in the mitochondrial genome. One of the unresolved questions on the contribution of existing variability to human disease is whether or not a particular pathology is cause by a small number of common variants or by a large number of different, rare variants. If the latter hypothesis is correct, we would expect that people of different genetic backgrounds but with the same genetic disease share the same causative pathogenic mutations. Alternatively, if the former hypothesis is correct, then people with the same pathology but different genetic backgrounds will have different causative mutations. We have compared the density of segregating polymorphisms in two different human populations, African and outside of Africa population. We then compared the density of different types of polymorphisms in these two populations. As has been reported previously, the African population is more variable in all kinds of polymorphisms, including those that are predicted to be deleterious. This is due to the recent expansion of the human population from Africa, which led to the African population retaining more of the variability compared to the out-of-Africa population. The only exception

was the class of polymorphisms that correspond to mutations that have been clinically described to cause genetic pathologies, in which the non-African population had more polymorphisms. Because all other classes of polymorphisms show the opposite pattern, the larger number of proved disease causing mutations in the non-African population is due to the higher rate of discovery of such polymorphisms in that population. Thus, what is known to be pathogenic to one population does not exist in another population and, therefore, genetic diseases are more likely to be caused by many rare variants, specific to each population, than common variants shared among different human populations.

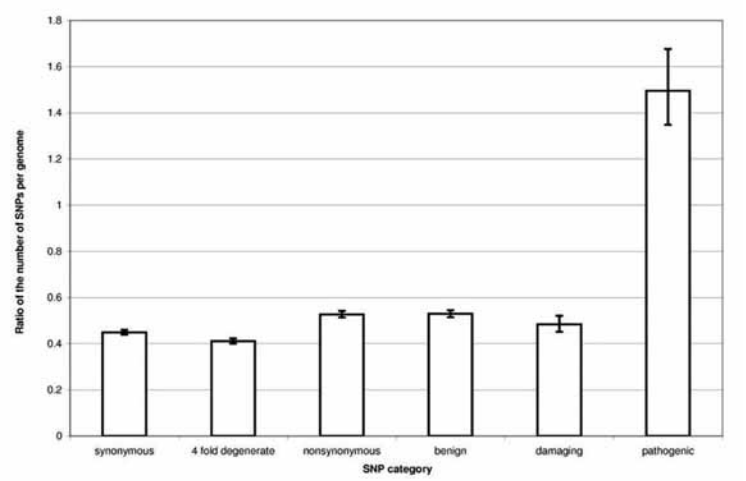


Figure 2.
For each category of variability the number of polymorphism in the non-African divided by the number of polymorphisms in the African population is shown.

PUBLICATIONS

Breen MS, Kondrashov FA.
"Mitochondrial pathogenic mutations are population-specific."
Biol Direct, 5:68 (2010).

Povolotskaya IS, Kondrashov FA.
"Sequence space and the ongoing expansion of the protein universe."
Nature, 465:922-926 (2010).
[Featured on Faculty of 1000]

Meer MV, Kondrashov AS, Artzy-Randrup Y, Kondrashov FA.
"Compensatory evolution in mt-tRNAs navigates valleys of low fitness."
Nature, 464:279-282 (2010).

Kondrashov AS, Povolotskaya IS, Ivankov DN, Kondrashov FA.
"Rate of sequence divergence under constant selection."
Biol Direct, 5:5 (2010).
[Highly Accessed]

Reviews

Kondrashov FA, Kondrashov AS.
"Measuring spontaneous mutation rates in the recent past and the near future."
Philos Trans R Soc Lond B Biol Sci, 365:1169-1176 (2010).

Innan H, Kondrashov FA.
"The evolution of gene duplications: classifying and distinguishing between models."
Nat Rev Genet, 11:97-108 (2010).

Book chapter

Kondrashov FA.
"Gene Dosage and Duplication".
In Evolution after Gene Duplication, Eds. Dittmar K and Liberles D, Wiley-Blackwell, Hoboken, New Jersey (2010).

BIOINFORMATICS AND GENOMICS

Group Evolutionary Genomics

Group structure:

Group Leader: Gian Gaetano Tartaglia

Postdoctoral Fellows: Marianela Masin, Matteo Bellucci

Doctoral Students: Federico Agostini

Pre-doctoral Students: Priscilla De Rosa

SUMMARY

We have recently developed a powerful algorithm to predict protein-RNA associations that we are validating using a series of assays including ribonucleoprotein immunoprecipitation-microarray (Rip-Chip) and cross-linking and immunoprecipitation of RNA-protein complexes (CLIP). We have previously worked on the prediction of interactomes of amyloid fibrils in the cellular context[1] and developed models for calculating the interaction potential with molecular chaperones[2]. Previously, the group leader developed a series of powerful bioinformatics models to predict protein toxicity under a variety of environmental conditions[3,4]. These models are updated and improved by members of the group.

RESEARCH PROJECTS

1. Large-Scale Predictions of RNA-Binding Proteins and Long Non-Coding RNAs

We have recently developed a method to perform proteome-wide predictions of transcriptomic associations using physico-chemical properties of molecules. Our algorithm was validated on a large set of experimental data including known interactomes of long non-coding RNAs. We performed a detailed analysis of the human mitochondrial RNA processing complex and demonstrated that our algorithm predicts regions involved in macro-molecular recognition with very high accuracy.

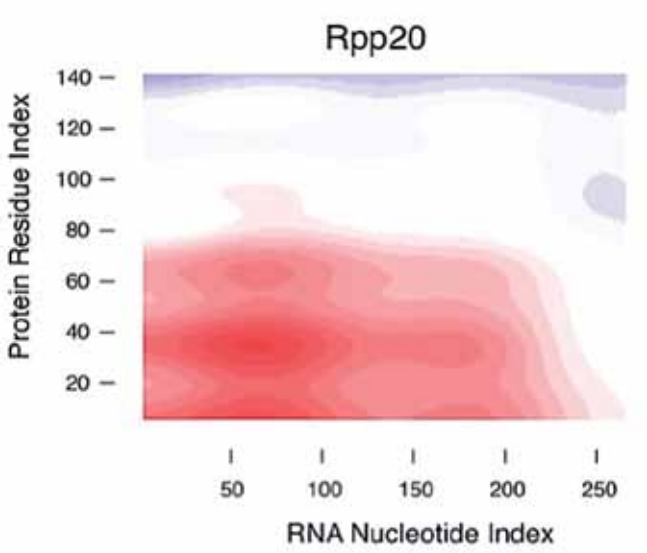


Figure 1.
Example of interaction contacts predicted with our computational methods. We illustrate here the case of the human Mitochondrial RNA Processing complex (MRP). Amino acid regions of protein Rpp20 involved in physical binding with RNA are highlighted in red. Our predictions are in complete agreement with experimental evidence.

2. Interactions Between Long Noncoding RNAs and Proteins in Cancer

For long time RNA has been considered the "working draft" employed by the cell to transfer information from DNA to proteins. Only recently, noncoding RNAs have been linked with a surprisingly wide variety of cellular functions, including epigenetic silencing, transcriptional regulation, and RNA processing and modification. Large numbers of mouse noncoding RNAs were found expressed during embryonic stem cell differentiation, suggesting specific regulatory roles in development.

Long noncoding RNAs (lncRNAs) comprise a large part of the human transcriptome but their functions have not yet been studied extensively and their impact on gene regulation is mostly unknown. Next Generation Sequencing (NGS) technologies allow for the quantification of the human transcriptome at single nucleotide resolution and facilitate detection of lncRNAs at large scale. However the identification of thousands of lncRNA raises a multitude of questions, which we would like to address using a combination of NGS, bioinformatics analysis and novel laboratory methods. These include the definition and classification of lncRNA loci, their tissue and developmental stage specific expression patterns, the interaction of lncRNAs with RNA-binding proteins (RBPs) and the functional role in the human proteome as well as the role of lncRNA in genetic diseases and cancer.

1 *Olzscha H, Schermann SM, Woerner AC, Pinkert S, Hecht MH, Tartaglia GG, Vendruscolo M et al. "Amyloid-like Aggregates Sequester Numerous Metastable Proteins with Essential Cellular Functions." Cell, 144:67-78 (2011).*

2 *Tartaglia GG, Dobson CM, Hartl FU and Vendruscolo M. "Physicochemical determinants of chaperone requirements." J. Mol. Biol, 400:579-588 (2010).*

3 *Tartaglia GG, Pawar AP, Campioni S, Dobson CM, Chiti F and Vendruscolo M. "Prediction of aggregation-prone regions in structured proteins." J Mol Biol, 380:425-36 (2008).*

4 *Luheshi LM, Tartaglia GG, Brorsson A, Pawar AP, Watson IE, Chiti F, Vendruscolo M et al. "Systematic in vivo analysis of the intrinsic determinants of amyloid Beta pathogenicity." PLoS Biol, 5:e290 (2007).*

We have developed a novel method to perform proteome-wide predictions of transcriptome associations using physico-chemical properties of molecules. Our algorithm was validated on a large set of experimental data including interactions with long non-coding RNAs from six model organisms. We are in the process of developing an NGS analysis pipeline to detect and quantify lncRNAs from mRNA-seq experiments. We plan to investigate the functional role of expressed lncRNAs in the human proteome using a series of coordinated experiments that will be performed *in silico*, *in vitro* and *in vivo*. More specifically, our aims are: i) detection, classification and expression quantification of lncRNAs in multiple mRNA samples from chronic lymphocytic leukemia (CLL) tumor and normal tissue; ii) identification of protein partners for expressed lncRNAs; iii) detection of common RNA structures for protein recognition; iv) mapping of RNA-protein interaction networks to pinpoint disease-related targets; v) prediction and functional analysis of genetic variation and chromatin states in lncRNA-protein interaction sites from NGS data (whole genome, exome, transcriptome and epigenom paired-end sequencing of CLL samples); vi) structure-function elucidation of selected ncRNAs-protein complexes.

The role of ncRNAs *in vivo* and their possible involvement in the alteration of cellular pathways could be confirmed making use of a variety of technologies including CLIP and Ribotrap: the former is based on UV cross-linking followed by immunoprecipitation, the latter relies on the co-expression of a reporter RNA and a recombinant RNA-binding protein of interest to form a complex suitable for affinity capture. In particular, a variation of the CLIP approach named PAR-CLIP would be used to determine the binding sites and motifs of cellular RBPs with high resolution using NGS. This protocol relies on the incorporation of photoreactive ribonucleoside analogs and crosslinked sites are revealed by thymidine to cytidine transitions in cDNA immunopurified from 4-thiouridine-treated cells[1]. Collaboration with the group of Dr. Jerney Ule (University of Cambridge) is planned to investigate the dynamic composition of ribonucleoprotein complexes *in vivo* using iCLIP[2]. This work is being carried out in collaboration with Dr. Stephan Ossowski at the CRG (Genomic and Epigenomic Variation in Disease).

1 *Hafner M et al. "Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP." Cell (2010).*

2 *Koenig J et al. "iCLIP Reveals the Function of hnRNP Particles in Splicing at Individual Nucleotide Resolution." Nat. Struct. Mol. Biol. (2010).*

3. Aggregation propensity of Proteomes

We have previously observed a remarkable anti-correlation between expression levels of coding transcripts and aggregation rates of corresponding proteins[1]. We believe that this relationship is generated by an evolutionary pressure acting to reduce the risk of aggregation in highly crowded cellular compartments (we called this the "life on the edge" hypothesis)[2]. Such an evolutionary pressure arises because failure of proteins to fold correctly gives rise to cellular malfunctions and diseases. Using the data available from genome databases such as *Ensembl*, we plan to apply a series of prediction tools developed in our laboratory, including the Zyggregator method[3], to determine aggregation propensities of human proteins. More specifically, we plan to investigate if protein isoforms characterized by a high aggregation potential have lower expression levels because their toxicity can potentially impair cell viability. By comparing mRNA expression levels and protein abundances, we want to investigate if post-translational modifications are able to play a role in regulating the delicate balance between aggregation and expression[4].

An interesting point can be raised with regards to non-coding transcripts and the hypothesis that most of them are untranslated because their corresponding proteins would be highly aggregation prone (Figure 2). If proven on a large scale, this finding will be relevant because it will allow the identification of a class of ncRNAs that lack the ability to generate functional proteins.

1 *Tartaglia GG, Pechmann S, Dobson CM and Vendruscolo M. "Life on the edge: a link between gene expression levels and aggregation rates of human proteins." Trends Biochem. Sci, 32:204-206 (2007).*

2 *Tartaglia GG and Vendruscolo M. "Correlation between mRNA expression levels and protein aggregation propensities in subcellular localisations." Mol Biosyst, 5:1873-1876 (2009).*

3 *Tartaglia GG and Vendruscolo M. "The Zyggregator method for predicting protein aggregation propensities." Chem Soc Rev, 37:1395-1401 (2008).*

4 *Laurent JM, Vogel C, Kwon T, Craig SA, Boutz DR, Huse HK, Nozue K et al. "Protein abundances are more conserved than mRNA abundances across diverse taxa." Proteomics, 10:4209-4212 (2010).*

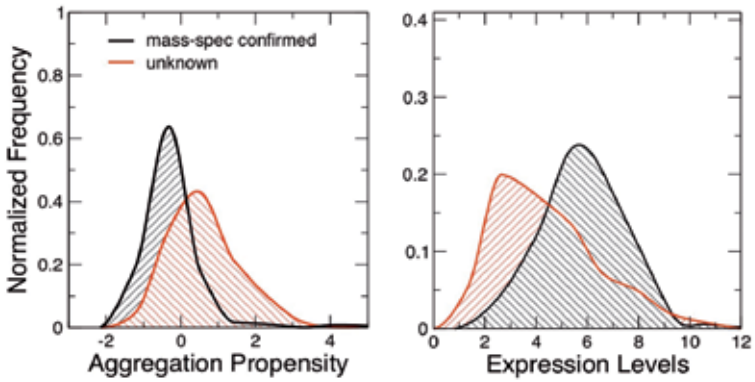


Figure 2.
Exploring the "Life of the Edge" hypothesis.
Coding transcripts confirmed by protein mass-spectrometry are associated with lower aggregation propensities and higher expression levels than transcripts whose coding ability is unknown (preliminary results).

4. Proteomes in Dire Straits: The Role of Dosage Sensitivity in Determining Cell Viability

Protein disorder and aggregation are frequent causes of cancer and neurodegenerative diseases. Disordered regions are prone to make promiscuous interactions if protein concentration is elevated, which leads to severe impairment of cell viability[1]. High protein concentration also favours the formation of amyloid aggregates that play a major role in neurodegenerative disorders such as Alzheimer's and Parkinson's diseases. Given the toxicity of amyloid aggregates in the cellular context, it has been suggested that a strong evolutionary pressure should be present in cells to reduce the concentration of amyloidogenic proteins[2]. Disorder and aggregation are tightly connected, especially if we consider that the toxicity of amyloid fibrils is highly correlated with the capacity of aggregates to promote aberrant interactions with disordered proteins[3].

In this project, we aim to investigate a series of relationships between protein concentration, aggregation and disorder. We plan to use both computational and experimental techniques to study the impact of changes in molecular concentration on protein homeostasis. Our approach is unique because it is based on the hypotheses that aggregation and disorder represent two complementary pressures that shape the range of activity of proteomes.

We will particularly focus on concentration changes that alter cell viability by destabilising protein networks in which disordered and aggregation-prone proteins are involved. In this regards, we would like to test how essential interactions are altered at high concentrations. We speculate that some specific biochemical pathways are enriched in proteins that are either poorly soluble compared to the levels at which they are expressed or highly prone to form promiscuous interactions, which would make them vulnerable targets under stress conditions or ageing. Since the formation of aggregates and aberrant interactions represent a considerable energetic drain for the cell, we would like to investigate if diseases result from a *misfolding cascade* that starts from a series of unregulated events and ends with cellular dysfunction and death.

Predictions will be performed using a series of novel algorithms that estimate the interacting potential of proteins and their aggregation propensities. To better understand how toxicity arises from aberrant interactions with disordered proteins, we will use *E. coli* and *S. cerevisiae* as model systems. More specifically, we will use quantitative proteomics to determine how the interaction profiles of proteins change as their concentration increases, and biophysical techniques to examine their aggregation. We are interested in determining the concentrations at which molecules will change their interactome inducing toxic effects. In particular, we will compare proteins that do and do not cause toxicity when they are overexpressed. In this way we aim to provide a general understanding of how increases in gene dosage cause can cause disease, and how selection acts to prevent detrimental increases in protein concentration. This work is being carried out in collaboration with Prof. Ben Lenher at the CRG (Genetic Systems).

1 *Vavouri T, Semple JI, Garcia-Verdugo R, Lehner B. "Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity." Cell, 138(1):198-208 (2009).*

2 *Tartaglia GG, Pechmann S, Dobson CM, Vendruscolo M. "Life on the edge: a link between gene expression levels and aggregation rates of human proteins." Trends Biochem Sci., 32(5):204-6 (2007).*

3 *Olzscha H, Schermann SM, Woerner AC, Pinkert S, Hecht MH, Tartaglia GG, Vendruscolo M, Hayer-Hartl M, Hartl FU, Vabulas RM. "Amyloid-like Aggregates Sequester Numerous Metastable Proteins with Essential Cellular Functions." Cell (2011).*

5. Protein–protein interactions and a-synuclein degradation in Parkinson's disease

Cells count on surveillance mechanisms that identify non-functional or altered intracellular components and eliminate them from inside cells. Alterations of these systems in neurons have been proposed to be involved in the pathogenesis of different neurodegenerative disorders[1]. Aberrant a-synuclein degradation is implicated in Parkinson's disease (PD) pathogenesis because the protein accumulates in the Lewy inclusion bodies associated with the disease. Little is known, however, about the pathways by which wild type a-synuclein and its pathogenic mutations, A30P and A53T, are normally degraded. Cuervo and colleagues found that wild-type a-synuclein was selectively translocated into lysosomes for degradation by the chaperone-mediated autophagy pathway (CMA). CMA targets specific cytosolic proteins that are recognized by the heat shock cognate protein of 70 kDa (hsc70) that, after interacting with the lysosome- associated membrane protein type 2A (Lamp2a), are translocated into the lysosomal lumen for rapid degradation. The binding of the substrates to Lamp2a at the lysosomal membrane is the rate-limiting step in the degradation process since overexpression of human Lamp2a in cultured cells results in increased activity of this pathway[2]. Although the pathogenic A53T and A30P a-synuclein mutants bind to Lamp2a they appear to act as uptake blockers, inhibiting both their own degradation and that of other substrates[3]. Using a computational sequence alignment, based on the compensated pathogenic deviations theory[4], throughout evolutionary species it is possible to see a strong correlation between changes in the DNA sequence of a-synuclein and Lamp2a. These changes lead to a punctual mutation of a single amino acid in the corresponding proteins and their orthologs throughout evolution. Notably, a-synuclein and Lamp2a are both developmentally regulated in cortical neuron cultures and *in vivo* in the central

1 *Finkbeiner S, Cuervo A, Morimoto R, Muchowski P. J Neurosci, 26:10349–10357 (2006).*

2 *Cuervo AM and Dice JF. Traffic, 1:570–583 (2000).*

3 *Cuervo AM, Stefanis L, Fredenburg R, Lansbury PT and Sulzer D. Science, 305:1292–1295 (2004).*

4 *Kondrashov AS, Sunyaev S and Kondrashov FA. Proc. Natl. Acad. Sci., 99:14878–14883 (2002).*

nervous system, and they physically interact as indicated by co-immunoprecipitation[5]. We hypothesized that the a-synuclein and Lamp2a interaction is possibly a case of ancestral compensation, and that alteration of this compensative substitution is a crucial factor that could have led the PD in humans. In order to validate our hypothesis and to address this fundamental question we are carrying out experiments of co-immunoprecipitation probing both wild type and A53T a-synuclein against the different substitutions in Lamp2a sequences. Furthermore we are planning to test the effect of the different interactions in isolated intact lysosomes and, finally, the outcome of these assays will be used to evaluate the pathologic accumulation of a-synuclein in neuronal cells. This work is being carried out in collaboration with other groups within the CRG, the laboratories of Evolutionary Genomics headed by Dr. Fyodor Kondrashov and the laboratory of Genetic Causes of Disease headed by Prof. Xavier Estivill.

5 *Vogiatzi T, Xilouri M, Vekrellis K and Stefanis LJ. Biol. Chem., 283:23542–23556 (2008).*

PUBLICATIONS

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

Szczepankiewicz O, Cabaleiro-Lago C, Tartaglia GG, Vendruscolo M, Hunter T, Hunter GJ, Nilsson H, Thulin E, Linse S.
"Interactions in the native state of monellin, which play a protective role against aggregation".
Mol Biosyst, Epub 2010 Nov 15.

Lee Y, Zhou T, Tartaglia GG, Vendruscolo M, Wilke CO.
"Translationally optimal codons associate with aggregation-prone sites in proteins".
Proteomics, 10(23):4163-71 (2010).

Tartaglia GG, Vendruscolo M.
"Proteome-level interplay between folding and aggregation propensities of proteins".
J Mol Biol, 402(5):919-28 (2010). (*)

Tartaglia GG, Dobson CM, Hartl FU, Vendruscolo M.
"Physicochemical determinants of chaperone requirements".
J Mol Biol. 400(3):579-88 (2010). (*)

Brorsson AC, Bolognesi B, Tartaglia GG, Shammas SL, Favrin G, Watson I, Lomas DA, Chiti F, Vendruscolo M, Dobson CM, Crowther DC, Luheshi LM.
"Intrinsic determinants of neurotoxic aggregate formation by the amyloid beta peptide".
Biophys J, 98(8) 1677-84 (2010). (*)

Books

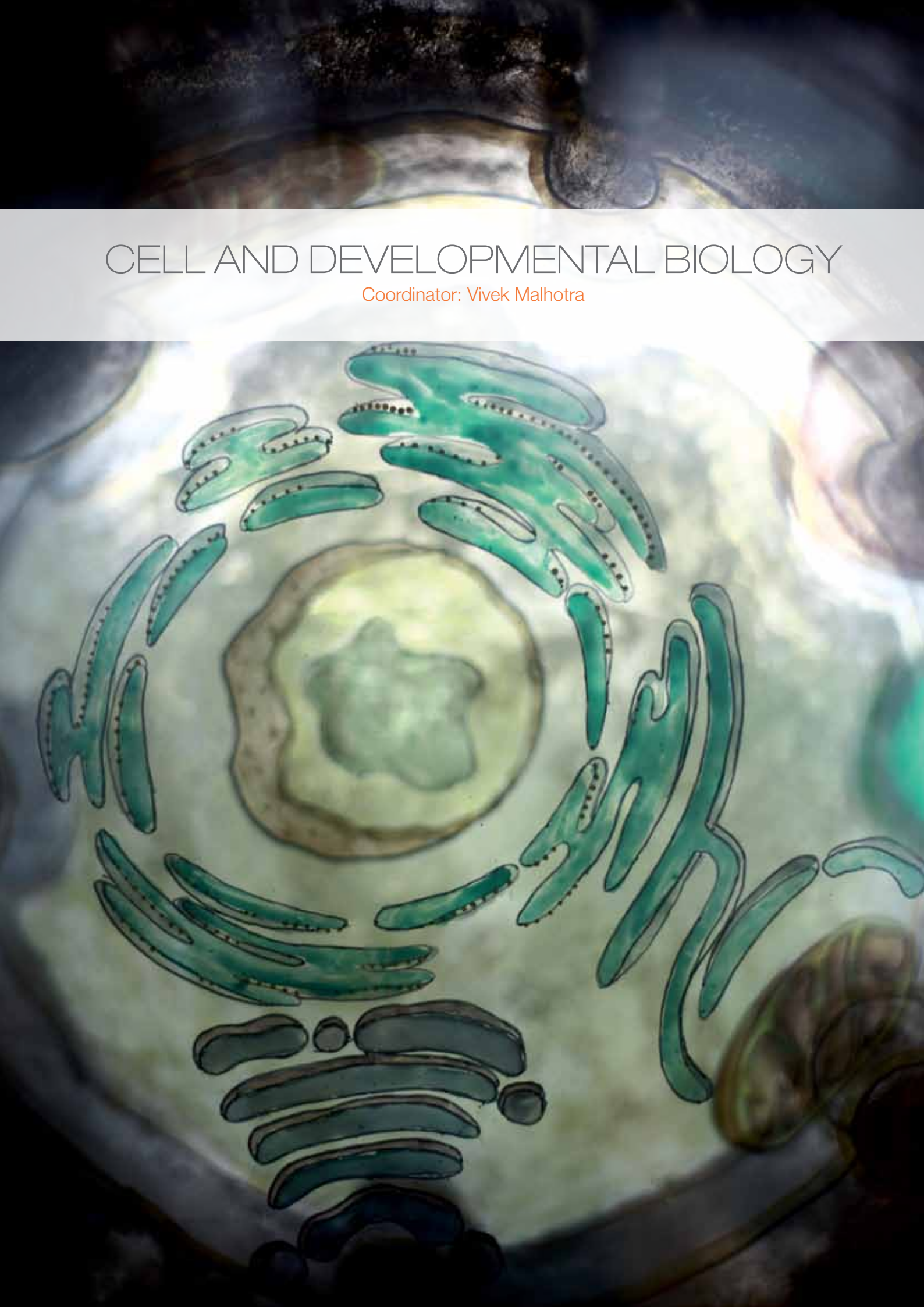
Bellucci, M.
"Protein-Protein Interactions: A tool kit to puzzle out functional networks".
VDM Verlag Dr. Müller (2010). (**)

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

(**) This book results from the work of Dr. Matteo Bellucci at the University of Bologna, Italy.

CELL AND DEVELOPMENTAL BIOLOGY

Coordinator: Vivek Malhotra



Vivek Malhotra and Manuel Mendoza were awarded the Advanced and the Starting Grant from the ERC, respectively, and Isabelle Vernos was elected to serve on the ERC council.

The composition of the program of Cell and Developmental Biology follows:

Coordinator:	Vivek Malhotra
Senior Group Leader:	Isabelle Vernos
Junior Group Leader:	Hernan Lopez-Schier
Junior Group Leader:	Manuel Mendoza
Junior Group Leader:	Jérôme Solon
Junior Group Leader:	Pedro Carvalho

CELL AND DEVELOPMENTAL BIOLOGY

Group: Intracellular Compartmentation

Vivek Malhotra is an ICREA Research Professor and the coordinator of the department.

Group Structure:

Group Leader: Vivek Malhotra

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

PhD student: Caroline Bruns

Technician: Anne-Marie Alleaume, Maria Ortega

SUMMARY

We are interested in the mechanism of protein secretion and biogenesis of Golgi membranes.

RESEARCH PROJECTS

Our main goal is to understand the mechanism of Golgi biogenesis and the biogenesis of cargo filled transport carriers during protein secretion. Key projects of the lab are summarized below.

We performed a genome wide screen in *Drosophila* tissue culture cells to identify new proteins involved in protein secretion. Some of these proteins also had a role in the organization of the Golgi membranes and therefore called TANGO for **T**ransport **AN**d **G**olgi **O**rganization (Bard et al., Nature 2006).

1. Biogenesis of mega transport carriers

Patrik Erlmann

COPII coated vesicles of 60-90 nm average diameter that mediate cargo export at the ER are too small for the trafficking of bulky collagens (collagen VII is a rod like structure of 400 nm that cannot be compressed), how is the bulky cargo like collagen VII exported from the ER? We found that TANGO1 was required for the export of collagen VII but not collagen I from the ER in Hela cells (Saito et al., Cell 2009). Most other cargoes were exported independent of TANGO1. Is TANGO1 a receptor for collagen VII loading at the ER? The mouse knockout of TANGO1 has been generated; the extracellular matrix, development of chondrocytes and bone mineralization is defective in these mutants, which leads to dwarfism and neonatal lethality (Solloway and colleagues, Genentech, USA. In review). The defect has been mapped to a block in collagen export from the ER in chondrocytes, fibroblasts, endothelial cells and mural cells. The investigators report a role of TANGO1 in the export of a large number, but not all, of collagens. The specificity of TANGO1 in collagen export at the ER is surprising and awaits further analysis. Regardless, TANGO1 is an important component of the protein export machinery and provides a means to understand the biogenesis of mega vesicles for the transport of bulky cargoes. We have identified two new binding partners of TANGO1 called TALI (TAngo LIke) and C-TAGE 5 and both are required for collagen VII export from the ER in mammalian cells. Our data suggests that TANGO1 exists as a complex with either TALI or c-TAGE5 (Saito et al., in preparation; Erlmann et al., in preparation). The cytoplasmic domains of these proteins bind the COPII coat proteins Sec23/Sec24 and the lumenal domain of TANGO1 binds collagen. Based on our findings we suggest that binding of the cytoplasmic domains of these proteins inhibits or retards the events leading to the biogenesis of COPII carriers. This allows the carrier to grow in size to accommodate the bulky collagens (Figure 1).

We are reconstituting the biogenesis of collagen containing COPII coated mega carriers *in vitro* from isolated ER membranes. This assay will be used to reveal the function of TANGO1 and its interacting partners in the biogenesis of mega transport carrier. The binding partners of the TANGO1, TALI and c-TAGE are being identified by yeast two-hybrid and co-immunoprecipitations. We will test whether the cytoplasmic domains of these proteins inhibit or retard the GTPase activity of Sar1p, which could also play a significant role in the biogenesis of a collagen containing mega transport carrier at the ER exit site.

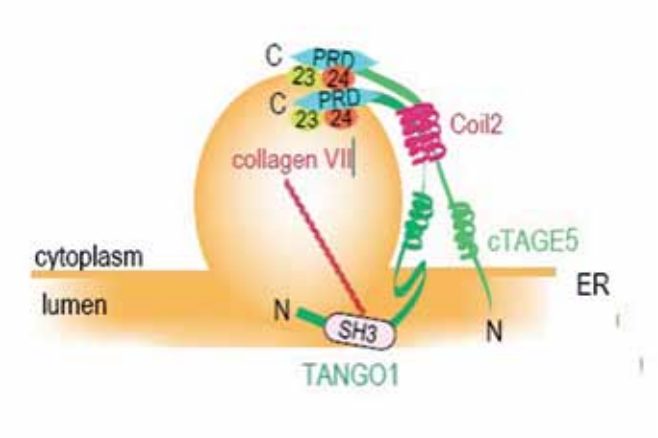


Figure 1.
TANGO1-cTAGE5 dependent loading of collagen into COPII carriers.
TANGO1 binds collagen through its SH3 like domain in the ER. The cytoplasmic coiled-coiled 2 domains of TANGO1 and cTAGE5 bind and their individual PRD's bind Sec24 of the COPII coats. The dimeric PRD's prevent the completion of COPII biogenesis, which, as a result, grows into a mega carrier to accommodate bulky collagen.

2. Cargo sorting at the Trans Golgi Network

Julia von Blume, Amy Curwin, Anne-Marie Alleaume

Our genome wide screen also revealed the involvement of an actin severing protein called twinstar (ADF/cofilin in mammalian cells) in protein secretion. We found that ADF/cofilin were required for the sorting of secretory cargo at the TGN (von Blume et al., 2009). But how do the cytoplasmic proteins ADF/cofilin interact with the soluble secretory cargo in the lumen of the TGN? Our more recent findings indicate that ADF/cofilin bind a Ca2+ pump called SPCA1 in the TGN. This binding is mediated by dynamic actin and required for the activity of SPCA1. Ca2+ pumped into the lumen is necessary for the sorting and export of Ca2+ binding proteins from the TGN (von Blume et al., 2011, in revision).

Our aim is to address the following: how does dynamic actin regulate the activity of SPCA1 and what other proteins are involved in this process? We will design Ca2+ binding proteins that cannot bind Ca2+ to test their trafficking in cells.

3. Isolation of Golgi to cell surface transport carriers

Yuichi Wakana

There are a number of different exit routes for cargo export from the TGN (Bard and Malhotra, 2006; Malhotra and Campelo, 2011). We have found that the serine/threonine kinase PKD is a key regulator for the trafficking of cargo that contains the basolateral sorting signals. PKD binds to the TGN by a DAG and ARF1 dependent process and is subsequently activated by Gßg and PKCh. The activated PKD phosphorylates PI4KIIIß, which generates PI4P from PI at the TGN. PKD also phosphorylates (and dissociates) the ceramide transporter CERT and the sterol sensor OSBP from the TGN. These findings suggest that PKD has a role in lipid homeostasis at the TGN. Compromising the activity of PKD, by inactivation or siRNA based knockdown, results in the accumulation of cargo filled tubules attached to the TGN. PKD therefore appears to be required for the events leading to the separation, by fission, of the TGN to cell surface transport carriers. Does the PKD dependent lipid homeostasis at the TGN play a role in membrane fission? What else is required for the biogenesis of PKD dependent TGN to cell surface transport carriers? To address these issues we have established conditions in permeabilized cells to generate, in a PKD dependent manner, cargo containing transport carriers that are destined to the cell surface. We have used this procedure to isolate a class of carriers that contain a number of proteins (cargo) that are secreted by Hela cells. The polypeptide composition of these carriers (collaboration with Dr. Matthias Mann, MPI Martinsried, Munich) has revealed a number of new molecules including actin and actin binding proteins (Wakana et al., in preparation). These isolated transport carriers and their associated proteins provide a means to understand the PKD dependent trafficking from the TGN to the cell surface.

The polypeptides contained in the isolated vesicles described above will be tested for their role in the biogenesis and trafficking of respective transport carriers. For this we will test whether knockdown of the specific polypeptides affect the trafficking from the TGN to the cell surface in intact Hela cells. The site of accumulation of the cargo will reveal whether the respective polypeptides are involved in the biogenesis, migration, docking or the fusion of the transport carriers. Identification of their interacting partners will increase the repertoire of molecules involved in this step of protein transport and, therefore, give a better understanding of the underlying mechanism.

4. Reconstitution of complete vesiculation of Golgi membranes *in vitro*

Josse van Galen, Felix Campelo

Treatment of intact cells with the drug Ilimaquinone (IQ) converts the entire Golgi complex into small vesicles (See references in Malhotra and Campelo, 2011). This process of Golgi vesiculation requires the trimeric subunit Gßg and PKD. But why is PKD required only for the biogenesis of Golgi to cell surface transport carriers and not in the formation of other carriers at the Golgi that are destined to the endosomes and the ER? Another way to pose this question is how does IQ vesiculate other compartments of the Golgi? To address this question, we have reconstituted the en masse vesiculation of isolated Golgi membranes by IQ, rat liver cytosol and an ATP regenerating system. Electron microscopy has revealed the conversion of Golgi stacks into small vesicles (collaboration with Dr. Bill Brown, Cornell University, USA). We plan to use this assay to understand the mechanism of membrane fission by IQ and PKD, respectively.

We will determine the lipid composition of the isolated Golgi membranes treated with IQ to monitor changes in the levels of DAG, PI4P, PA, LPA and the sterols (collaboration with Dr. Felix Wieland of the University of Heidelberg, Germany). The change in the organization of the Golgi membranes will be confirmed by conventional electron microscopy in collaboration with Dr. Bill Brown. This will provide an understanding of the changes in lipid composition during the events leading to the vesiculation (by membrane fission) of Golgi membranes. We will test the effect of inhibiting PKD, DAG, PI4P, PA and LPA production on the IQ mediated changes in Golgi morphology. We will test the effect of adding constitutively activated PKD, in the absence of IQ, to monitor effects on Golgi morphology and lipid composition. These findings will allow us to design better experiments to test the role of specific lipids during the biogenesis of transport carriers (in a PKD dependent manner) that form at the Golgi and are destined to the cells surface. We hope this information will also help reveal an understanding of how IQ vesiculates the cis-medial and trans Golgi cisternae by a PKD independent process.

5. Identification of Proteins Involved in Mucin Secretion (PIMS)

Sandra Mitrovic, David Cruz

Airways mucin, the gel forming component of mucus, provides protection against pathogens and other forms of environmental abuse and thus essential for the normal physiology of the epithelium. Hyper secretion of mucin is a common feature of chronic obstructive pulmonary disease (COPD) and asthma; hypertrophy of the submucosal glands and hyperplasia of the surface goblet cells increases mucin production, which increases the morbidity and mortality in patients with severe form of the disease. How is mucin secreted by the goblet cells? We established an assay to screen 7343 gene products by siRNA dependent knockdown, to monitor their requirement in the phorbol ester (PMA) dependent secretion of Mucin 5AC from the human goblet cells. This procedure revealed the involvement of 29 proteins in the docking and fusion of Mucin 5AC containing granules with the plasma membranes. We are characterizing their specific roles in the process of mucin secretion.

In addition to revealing the role of PIMS and their interacting partners in mucin secretion we will generate mouse knockouts for three of the genes (PIMS) to understand their physiological role *in vivo*. We expect these knockouts will serve as good model systems to explore the role of mucin secretion in the development of COPD.

6. Unconventional protein secretion.

Juan Duran, Caroline Bruns, Julien Villeneuve

Eukaryotic cells secrete a class of proteins that do not enter the ER. In general, proteins that are secreted unconventionally are involved in processes such as tissue reorganization, immune surveillance and cell survival. While it is not clear how many proteins are secreted unconventionally, a recent report claims that of the proteins identified by sequencing extracellular proteins of the parasitic protozoan Leishmania (Viannia) braziliensis, only 5% contained a signal sequence whereas the vast majority (57%) lacked a signal sequence for entering the conventional secretory pathway (Cuervo et al., 2009). This raises the possibility that eukaryotic cells likely secrete a large number of proteins unconventionally. However, their identification has been difficult because most [known] proteins secreted unconventionally are small in size (10-20 kDa) and lack any obvious domains or features which could be used to identify other such proteins by standard proteomic approaches. We found that a signal sequence lacking Acyl CoA binding protein (ACBP in dictyostelium and its ortholog Acb1 in Yeast) was secreted without entering the ER-

Golgi pathway. Interestingly, autophagy related proteins, protein involved in trafficking to the early endosomes, formation of multivesicular body (MVB), the plasma membrane t-SNARE Sso1 (but not Sso2) and the Golgi attached GRASP (Grh1) were required for Acb1 secretion. Based on our findings, we have suggested the following steps in Acb1 secretion: Acb1 is packed into an autophagosome like vesicle or a secretory autophagosome (step1); Acb1 containing secretory autophagosomes fuse with an early endosome (step2); The endosomes mature into a multivesicular body (step3); Either the endosome (from step 2) or the MVB (from step 3) fuses with the plasma membrane to release vesicles in the extracellular space (step4). The vesicles lyse in the extracellular space and release Acb1 (step5) (Figure 2).

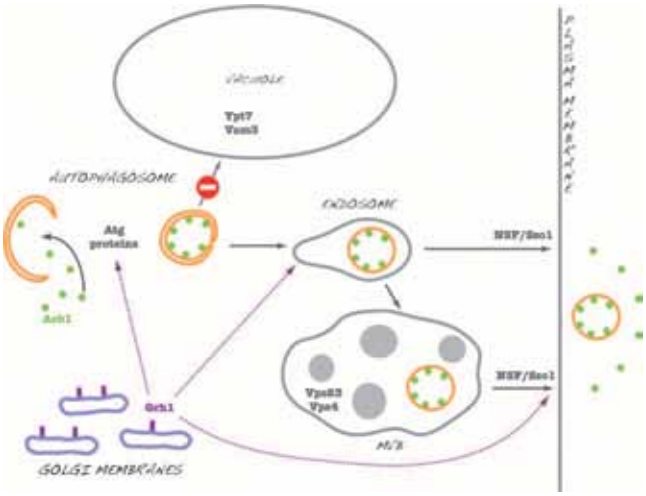


Figure 2.
The Acb1 secretion pathway.
Cytosolic Acb1 is packaged into autophagosomes, which fuse with early endosomes. The early endosomes containing Acb1 either fuse directly with the cells surface, or more likely mature into a multivesicular body (MVB). The MVB fuses with the cell surface to release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

We are developing a simple quantitative assay to monitor Acb1 secretion and screen the entire yeast genome to identify new proteins involved in Acb1 secretion. We want to identify membrane bounded Acb1 containing intermediates during secretion. This will be performed in collaboration with Dr. Michael McCaffrey of Johns Hopkins Medical School, USA. Gene products involved in the formation and consumption of these transport intermediates will be identified and characterized. We will also test whether secretion of signal less proteins such as interleukin -2β, MIF, insulin-degrading enzyme in the mammalian cells is mediated by secretory autophagsomes. The key issues we want to address are: how are the signal sequence lacking proteins packed into an autophagosome and how secretory autophagosomes avoid fusion with the lysosomes?

PUBLICATIONS

Duran JM, Anjard C, Stefan C, Loomis WF and Malhotra V.
"Unconventional secretion of Acb1 is mediated by autophagosome."
J Cell Biol, 188:527-536 (2010).

Pusapati GV, Krndija D, Armacki M, von Wickert G, von Blume J, Malhotra V, Adler G and Seufferlein T.
"Role of the second cysteine-rich domain and Pro 275 in PKD2 interaction with ARF1, TGN recruitment and protein transport."
Mol Biol Cell, 21:1011-1022 (2010).

Guizzunti G, Brady TP, Fischer D, Malhotra V and Theodorakis EA.
"Chemical biology studies on norrisolide."
Bioorg Med Chem, 18: 2115-2122 (2010).

Cabral M, Anjard C, Malhotra V, Loomis W, Kuspa A.
"Unconventional Secretion of AcbA in Dictyostelium discoideum through a vesicular intermediate."
Euk Cell, 9:1009-1017 (2010).

Malhotra V and Campelo F.
"PKD Regulates Membrane Fission to Generate TGN to Cell Surface Transport Carriers."
Cold Spring Harb Perspect Biol, in press.

Book chapter

Malhotra V, Warren G and Mellman I.
"Protein trafficking between membranes."
In: Lewin's Cells, Second edition, Eds Cassimeris, Lingappa and Plopper, Pp 345-390 (2010).

CELL AND DEVELOPMENTAL BIOLOGY

Group: Microtubule Function and Cell Division

Isabelle Vernos is an ICREA Research Professor.

Group structure:

Group Leader: Isabelle Vernos

Postdoctoral Fellows: Teresa Sardon, Sylvain Meunier, Roser Pinyol, Marti Badal (since March 2010)

Students: Vanessa Dos Reis Ferreira, Isabel Peset (until March 2010), Martin Schütz, David Vanneste
Antonios Lioutas, Jacopo Scrofani (since september 2010), Tommaso Cavazza (since October 2010)

Technicians: Leonor Avila, Nuria Mallol, Violeta Beltran (since september 2010)

SUMMARY

Research in my lab is directed at understanding the role of the microtubule network in cell organization and function. To address this question we study various microtubule-associated proteins (molecular motors and MAPs) and their regulators (kinases, phosphatases and the small GTPase Ran during M-phase). One major goal is to unravel how the self-organization of cellular components results in the morphogenesis of dynamic molecular machines. In 2010, we have focused on the morphogenesis of the bipolar spindle in mitosis and meiosis.

Our favourite experimental system is the Xenopus egg extract system for studies on cell cycle progression and regulation, microtubule dynamics, spindle assembly and chromosome behaviour (Karsenti and Vernos, 2001). We combine it with the use of human tissue culture cells in which we validate some of the results obtained in egg extract.

RESEARCH PROJECTS

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

1. The centrosome: a cellular regulatory hub

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

The centrosomal Aurora A kinase has been implicated in several important processes including centrosome maturation during G2, mitotic entry, centrosome separation and bipolar spindle assembly. We have previously shown that Aurora A interacts with TPX2 in a RanGTP dependent manner after nuclear envelope breakdown resulting in kinase activation. Using the Xenopus egg extract system we found last year that Aurora A works though different mechanisms to regulate MT assembly during mitosis, ensuring bipolar spindle formation (Sardon et al, 2008). At the centrosome active Aurora A is required for efficient MT nucleation. In addition, it promotes MT growth by recruiting TACC3 that works in concert with XMAP215 to oppose the destabilizing activity of XKCM1 (Peset I. et al, 2005). To obtain a useful tool to further investigate the function of Aurora A during the cell cycle we have collaborated with the group of Prof Giannis (Leipzig, Germany) to look for a specific inhibitor for this kinase. We have found one compound that shows specificity for Aurora A inhibition in cells suggesting that it could be useful for basic and applied research (Sardon et al, 2009).

This year, in collaboration with the group of P. Aloy (IRB, Barcelona) we have used a biocomputing approach together with experimental validation to predict a set of putative substrates for the centrosomal kinase Aurora A (Sardon et al, 2010).

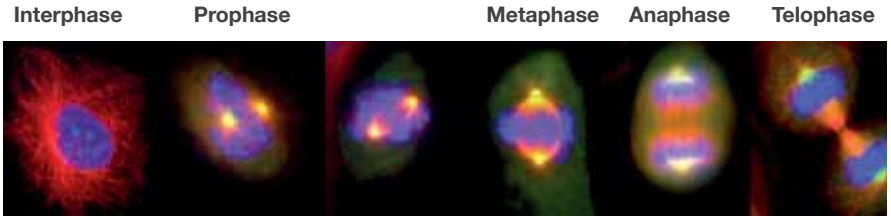


Figure 1.
Immunofluorescence analysis of HeLa cells shows the specific localization of the kinase Aurora A to the centrosome and the spindle poles during mitosis. MTs are labelled in red, DNA in blue and Aurora A in green.

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

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

We have previously identified TPX2 as being essential for this pathway and for spindle assembly both in mitosis and meiosis (Gruss and Vernos, 2004; Gruss et al., 2002; Wittmann et al., 2000). Interestingly, TPX2 is highly conserved in plants that relie exclusively on an acentrosomal pathway for spindle assembly as they do not contain centrosomes like animal cells. Last year in collaboration with two groups working in plants we have shown that TPX2 has conserved functions in animals and plants (Vos et al, 2008; Evrard et al, 2009). We are continuing our efforts to unravel the molecular mechanism underlying the RanGTP dependent pathway for MT assembly in M-phase, focusing on TPX2 and its interaction partners.

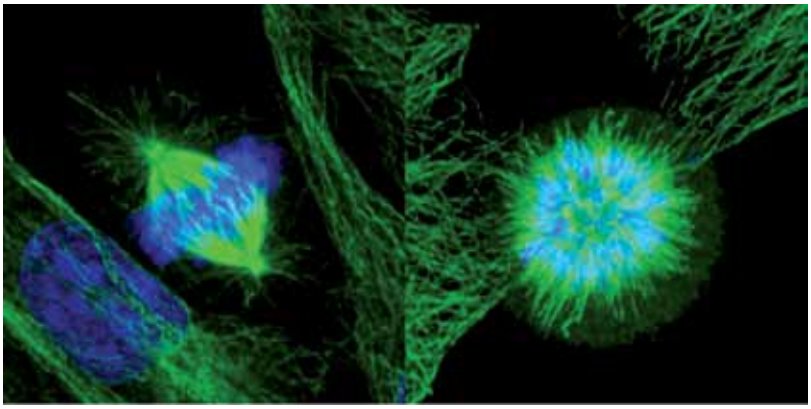
3. Role of molecular motors in spindle assembly and chromosome movements

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

Last year we have characterized another motor, Hklp2, the human homologue of Xklp2 (Boleti et al, 1996; Wittmann et al, 1998) and found that this motor plays a role in bipolar spindle assembly and stability.

Although Eg5 has a predominant and essential role in bipolar spindle establishment, Eg5 inhibition does not compromise the stability of the bipolar spindle at metaphase. We found that Hklp2 becomes essential under these conditions for spindle stability. Hklp2 localizes to the spindle microtubules and the chromosomes in metaphase. This steady state distribution is essential for Hklp2 role in promoting the switch from the monopolar to the bipolar configuration and in stabilizing spindle bipolarity in metaphase. Our data provide an additional understanding of the mechanism driving the initial establishment and subsequent maintenance of the bipolar spindle at metaphase, and reveals the existence of a novel specific mechanism that stabilizes bipolar spindles and may be essential to prevent mitotic defects (Vanneste et al, 2009).

Figure 2.
Confocal images of mitotic HeLa cells. At metaphase, the bipolar spindle has assembled around the condensed chromosomes positioned at an equatorial position ready to be segregated by this molecular machine (left). When some motors like Eg5 and Hklp2 do not function properly the bipolar spindle is unstable and collapse into a monopolar configuration with the subsequent failure of cell division (right). MTs are shown in green, chromosomes in blue.



PUBLICATIONS

Sardon T, Pache RA, Stein A, Molina H, Vernos I, Aloy P.
"Uncovering new substrates for Aurora A kinase."
EMBO Rep, 12:977-84 (2010).

Vernos I.
"Isabelle Vernos: monitoring around the mitotic spindle."
J Cell Biol, 188(5):616-7 (2010).

CELL AND DEVELOPMENTAL BIOLOGY

Group: Sensory Cell Biology and Organogenesis

Group structure:

Group Leader: Dr. Hernán López-Schier (Ramón y Cajal fellow)

Postdoctoral Fellows: Dr. Adele Faucherre (Marie Curie postdoctoral fellow), Dr. Jean Pierre Baudoin (FRM postdoctoral fellow), Dr. Jacobo Cela, Dr. Rachele Allena (shared with James Sharpe lab), Dr. Sabrina Desbordes

Graduate Students: Indra Wibowo, Filipe Pinto Teixeira, Jesús Pujol Martí, Alessandro Mineo, Andrea Zecca, Mayra Eduardoff

Technician: Andrea Durán

SUMMARY

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

RESEARCH PROJECTS

1. Cellular responses to polarity signals, in particular planar cell polarity

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

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

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

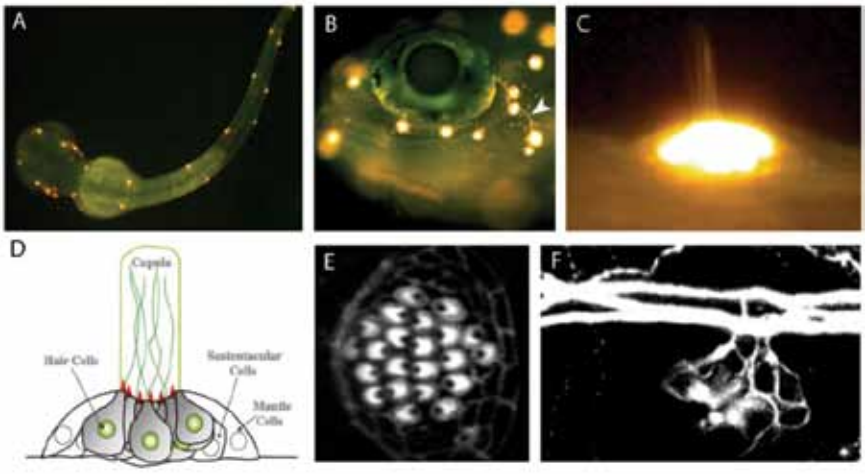


Figure 1.
Low magnification view of a living zebrafish larva whose hair cells in the lateral line were labelled with the fluorescent vital dye DiAsp (bright orange). It shows the superficial and systematic distribution of neuromasts along the anterior (head) and posterior (trunk and tail) lateral-line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

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

2. Sensory organ innervation

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

3. Sensory-organ growth and regeneration, with an emphasis on epithelial remodelling

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

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

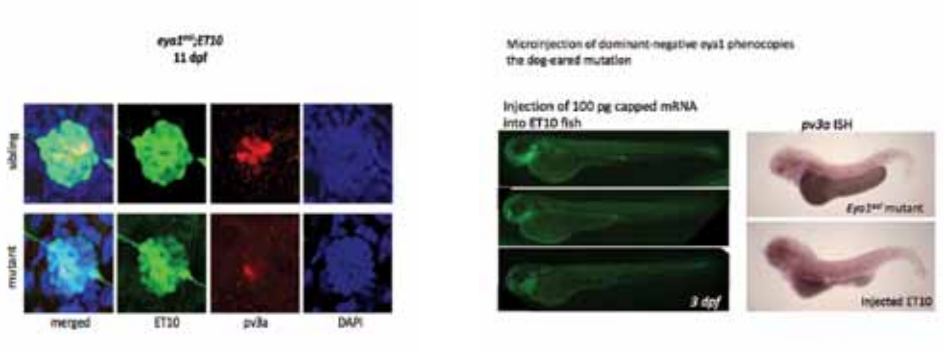


Figure 2:
High magnification view of wild type and eya1/dog eared zebrafish larvae whose hair cells in the lateral line were labelled with PV3a (red). Fish are transgenics for ET10, which expresses the GFP in supporting cells (green). Blue is DAPI staining. It shows that hair cells degenerate in eya1/dog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

4. Tissue mechanics

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

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

PUBLICATIONS

Wibowo I, Pinto-Teixeira F, Satou C, Higashijima S, López-Schier H.
"Compartmentalized Notch signaling sustains epithelial mirror symmetry."
Development, in press.

Faucherre A, López-Schier H.
"Delaying Gal4-driven gene expression in the zebrafish with morpholinos and Gal80."
PLoS One, in press.

Swoger J, Muzzopappa M, López-Schier H, Sharpe J.
"4D retrospective lineage tracing using SPIM for zebrafish organogenesis studies."
J Biophotonics, doi: 10.1002/jbio.201000087. Epub 2010 Oct 5.

Faucherre A, Baudoin JP, Pujol-Martí, López-Schier, H.
"Multispectral four-dimensional imaging reveals that evoked activity modulates peri-pheral arborization and the selection of plane-polarized targets by sensory."
Development, 137(10):1635-43 (2010).

Pujol-Martí J, Baudoin JP, Faucherre A, Kawakami K. López-Schier H.
"Progressive neurogenesis defines lateralis somatotopy."
Dev Dyn, 239(7):1919-30 (2010).

Cobreros-Reguera L, Fernández-Miñán A, Fernández-Espartero CH, López-Schier H, González-Reyes A, Martín-Bermudo MD.
"The Ste20 kinase misshapen is essential for the invasive behaviour of ovarian epithelial cells in Drosophila."
EMBO Rep, 11(12):943-49 (2010).

Review

López-Schier H.
"Fly fishing for collective cell migration."
Curr Opin Genet Dev, 20(4):428-32 (2010).

CELL AND DEVELOPMENTAL BIOLOGY

Group: Coordination of Cytokinesis with Chromosome Segregation

Group structure:

Group leader: Manuel Mendoza

Postdoctoral Fellow: Alexandre Vendrell
Eva Kiermaier (shared with the Solon lab; since October 2010)

PhD students: Gabriel Neurohr, Aina Masgrau, Iris Titos

Technician: Trinidad Sanmartin

Visiting student: Marta Inglés (July-August 2010)

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. These dramatic morphological changes need to be carefully coordinated with each other in space and time. To learn more about the principles underlying this coordination, we focus on the events at the end of the cell cycle: chromosome segregation and cytokinesis, 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. Mechanisms of chromosome segregation sensing by the NoCut checkpoint

Cell division is completed through partition of the cytoplasm by ingression of the cleavage furrow, and the subsequent cleavage of the cell membrane into two during abscission. At the same time, the replicated chromosomes are segregated to opposite ends of the cell. Whereas cleavage furrow ingression usually proceeds concomitantly with poleward movement of the chromosomes, abscission must take place exclusively after the last pair of sister chromatids have been pulled out of the cleavage plane. In budding yeast and human cells, a checkpoint known as NoCut delays completion of cytokinesis when chromosome segregation is impaired. Inactivation of NoCut leads to premature abscission, and late-segregating chromosomes are trapped and cut by the cytokinesis machinery.

The Chromosome Passenger Complex (CPC) component Aurora B kinase (Ipl1 in yeast) plays a central role in the NoCut checkpoint (Fig. 1). The CPC localizes to the spindle midzone during anaphase, and it is activated by the presence of acetylated chromatin around the midzone. Ipl1 activity mediates the translocation of two anillin-related proteins, called Boi1 and Boi2, to the site of cytokinesis during anaphase, where they inhibit cytokinesis until chromosomes are cleared from the cleavage plane.

NoCut relies on a complex network of factors to monitor chromatin segregation away from the spindle midzone. We are systematically characterizing the NoCut mechanism, taking advantage of genetic screens to identify new checkpoint components. A genome-wide screen for non-essential NoCut genes has identified about 50 NoCut candidate genes. We predict that analysis of the identified mutants will identify factors that act upstream of the CPC, enabling its activation by chromatin (*sensors*); downstream of Boi1 and Boi2, inhibiting the abscission machinery at the site of cytokinesis (*effectors*); and mediating the communication between the CPC and Boi1/2, by shuttling between the nucleus and the cytoplasm or affecting the shuttling of other NoCut components (*transducers*). For secondary screens on the initial set of candidates, we have developed robust assays for imaging of mitosis and cytokinesis using live cell microscopy with a variety of fluorescent reporters. These include tools to visualize the plasma membrane, monitoring cell cycle progression, and to image chromosome condensation and segregation in living cells. We are now focusing our main efforts on the analysis of two protein complexes, one a histone acetylase and the second an mRNA deadenylase, in the regulation of cytokinetic abscission.

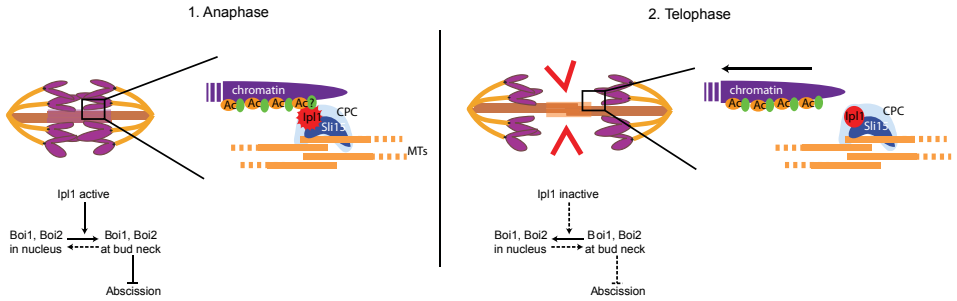


Fig. 1. *Model of the NoCut checkpoint. (A) In early anaphase, chromosomes (depicted in purple) are in contact with the CPC, which binds to spindle midzone microtubules (in orange). Inset: the CPC subunits Ipl1 and Sli15 are depicted in red and blue, respectively. Midzone-bound Ipl1 is kept active, probably through interaction with chromatin-associated factors (in green), which require histone acetylation (orange circles) to activate the CPC. As a result, Boi1 and Boi2 localize to the bud cortex, where they inhibit abscission. (B) When chromosomes are segregated away from the midzone, the CPC is no longer activated by chromatin and the NoCut signal is turned off; Boi1 and Boi2 leave the bud neck, and abscission (represented by red triangles) ensues.*

2. Coordination of chromosome length with spindle elongation

We are also interested in additional mechanisms ensuring robust chromosome segregation. The function of the mitotic spindle is to separate sister chromatids away from each other. When the length of the anaphase spindle reaches twice the length of the longest chromosome arm, chromosome segregation is complete. Are there mechanisms that allow the cell to adjust the size of the anaphase spindle and/or anaphase dynamics, to the length of chromosome arms? We are addressing this question through manipulation of chromosome arm length in yeast cells, and analysis of mitotic and cytokinetic events by live cell microscopy in normal and mutant strains. Using this strategy, we have discovered a novel feedback system, which coordinates axial chromosome compaction with anaphase spindle length. We are characterizing this and other anaphase controls through a multidisciplinary approach, which combines live cell imaging, classical molecular and genetic techniques and state-of-the-art genomics and proteomics.

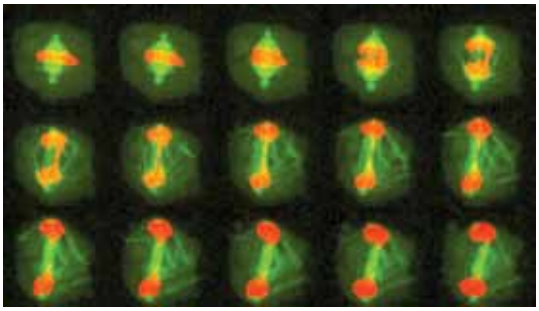


Fig. 2.
Anaphase in a Drosophila S2 cell. Histones are labeled in red (H2B-GFP) and microtubules in green (Tubulin-mCherry). Frames were acquired every minute. Images by Evi Kiermaier.

3. Chromosome Compaction and Segregation in *Drosophila* cells

Unlike in yeast, cell size and spindle length varies considerably within multicellular organisms. As asymmetric divisions produce cells of different fates and sizes, coordination between chromosome compaction and spindle length becomes particularly important. To understand the molecular basis of this coordination and its relevance in maintaining genomic stability, we have started to use the fruit fly *Drosophila* as a model system.

CELL AND DEVELOPMENTAL BIOLOGY

Group: Biomechanics of Morphogenesis

Group structure:

Group leader: Jérôme Solon

Postdoctoral Fellow: Laure Saias

Graduate Students: Natalia Czerniak, Eva Kiermaier (shared with Manuel Mendoza lab)

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.

RESEARCH PROJECTS

1. Mechanisms driving tissue constriction and fusion during *Drosophila* development at the cellular scale

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

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

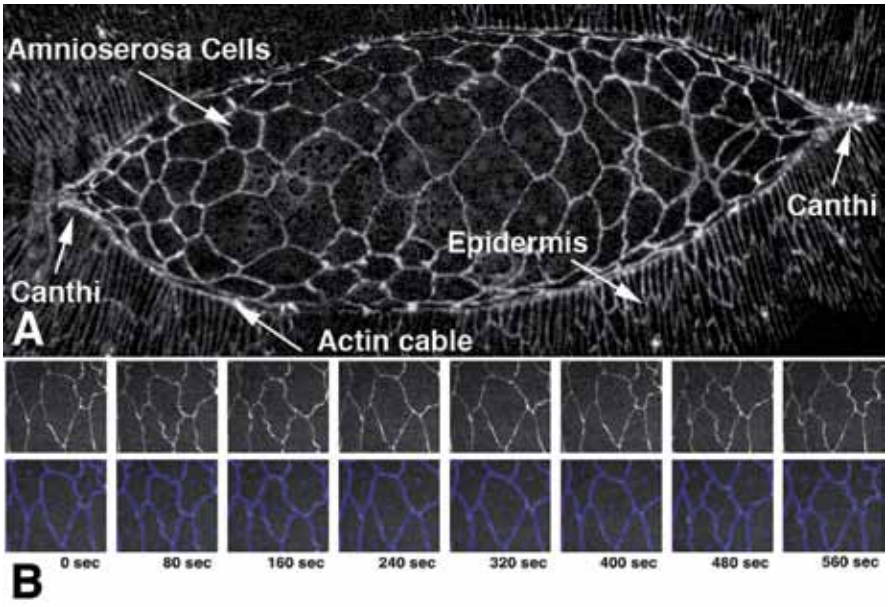


Fig. 1.
(A) Armadillo-GFP highlighting cell membranes during Drosophila dorsal closure. *The epidermal tissues converge from the lateral part (top and bottom of the picture) to the dorsal part (in the center of the picture) of the embryo. The dorsal part is covered with amnioserosa cells, which provide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

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

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

2. Spatial and temporal coordination of the cellular contractions in the amnioserosa tissue

We are also interested in establishing the connection between the intracellular contractions and the tissue organization and constriction. The amnioserosa tissue is composed of about hundred cells; each of them contracts individually. How does this sum of individual contractions generate a global constriction of the tissue and how does the global shape and tension of the tissue influence the individual cell contractions? We are addressing these questions by combining different imaging techniques available at the ALMU (spinning disk, confocal and two-photon microscopy) with automated image analysis and physical modeling.

CELL AND DEVELOPMENTAL BIOLOGY

Group: Organelle biogenesis and homeostasis

Group structure:

Group leader: Pedro Carvalho (from July 2010)

PhD students: Alexandra Grippa (from September 2010), Annamaria Ruggiano (from September 2010)

Technician: Josep Pareja (from September 2010)

SUMMARY

In most eukaryotic cells, roughly a third of the genes encode for secreted and cell surface proteins. The biogenesis of this class of proteins occurs in the endoplasmic reticulum (ER) and demands a considerable amount of cellular resources. In the ER lumen, a myriad of chaperones and enzymes assist protein folding and assembly of multisubunit protein complexes. The ER is also involved in the synthesis of most of the lipids in cells. Perturbations in any of these processes often result in ER stress, a hallmark of many diseases including diabetes, obesity and cancer.

RESEARCH PROJECTS

Our long-term goal is to understand the mechanisms underlying ER homeostasis with emphasis on the following topics:

1. Molecular mechanisms of ER- associated protein degradation (ERAD)

Despite the dedicated activity of the folding factors in the ER, large numbers of misfolded proteins are produced. ER quality control mechanisms ensure that only the folded proteins follow the secretory pathway. Thus, a central aspect of the quality control is the elimination of misfolded proteins from the membrane and lumen of the ER. This occurs mostly by a process called ER-associated protein degradation (or ERAD).

The ERAD pathway involves several different steps. The process begins with the recognition of a substrate as being misfolded. Once selected, the substrate is transported across the ER membrane (or retrotranslocated). On the cytoplasmic side of the membrane, the substrate is ubiquitinated by specific, membrane-bound ubiquitin ligases and released into the cytosol for degradation by the proteasome.

Although most ERAD substrates follow this general scheme, different substrates require distinct components suggesting the existence of multiple routes in ERAD. These alternative routes have been best characterized in *S. cerevisiae*. In this organism, two ubiquitin ligases complexes, the Hrd1 complex and the Doa10 complex, are required for the degradation of different types of misfolded proteins depending on the location of the misfolded domain: proteins with a misfolded domain in the lumen (ERAD-L substrates) or membrane (ERAD-M substrates) of the ER are targeted to the Hrd1 complex, proteins with a misfolded domain on the cytoplasmic side of the ER membrane (ERAD-C substrates) are targeted to the Doa10 complex.

More recently, we have focused on a specific aspect of Hrd1-dependent degradation: How misfolded luminal proteins (ERAD-L) are moved through the ER membrane into the cytosol? It has been postulated that the misfolded proteins are moved through a protein conducting channel however the identity of that channel has remained elusive. We have now found that the ubiquitin-ligase Hrd1p itself is the main membrane component in the retrotranslocation of misfolded ER luminal substrates. Our findings also provided the first mechanistic insight into the early stages of ERAD-L.

Despite this progress, the molecular mechanisms of ERAD are still largely unknown. Key unresolved issues under investigation in our lab include the determinants for substrate selectivity and the mechanism for translocation of misfolded proteins across the ER membrane.

2. Lipid droplets biogenesis and dynamics

Cytoplasmic lipid droplets are depots of neutral lipids (triglycerides, cholesterol esters) enclosed in a monolayer of phospholipids and associated proteins. These organelles appear to be extremely dynamic however they maintain a very intimate relation with the ER. In fact, the prevailing model for lipid droplet biogenesis involves the budding of the nascent lipid droplet from the ER. Lipid droplets are present in most eukaryotic cells where they may serve as reservoirs to store cellular energy and as building blocks for membrane lipids. Excessive lipid accumulation in cells is also the hallmark of obesity and atherosclerosis, but remarkably little is known about lipid-droplet cell biology.

Among the questions we are trying to address are the mechanisms involved in lipid droplet formation and how proteins are targeted to these organelles.

PUBLICATIONS

Carvalho P, Stanley AM, Rapoport TA.
"Retro-translocation of a misfolded luminal ER protein by the ubiquitin-ligase Hrd1p."
Cell, 143(4):579-591 (2010). (*)

(*) This publication results from the work of Dr Pedro Carvalho in the laboratory of Dr. Tom Rapoport at Harvard Medical School, USA.

SYSTEMS BIOLOGY

Coordinator: Luis Serrano

The EMBL/CRG Systems Biology Research Unit encompasses a range of systems biology topics to provide a rich, stimulating environment. During 2010, we underwent a programme review by an international committee whose excellent outcome put us amongst Europe's top Systems Biology programmes. The individual groups in the Programme work on very different systems, yet they approach these systems with a common style of investigation. The panel was especially impressed by the tight integration between experimental and modeling work in all groups, and by the fact that all groups generate the data they need themselves, and are fearless in embracing and developing new technologies to do so. In the sometimes hard to balance the equilibrium between so-called dry and wet biology, this Programme seems to excel at modeling to address fundamental biological questions, rather than searching for problems to which favorite modeling approaches can be applied, which is still so often the case in many Systems Biology laboratories.

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, Raik Grünberg, Eva Yus, Almer Van der Sloot, Andreu Alibés, Maria Lluch, Julia Burnier Marc Guell, Javier Delgado

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

Collaborator: Peter Vanhee

Technicians: Sira Martinez, Tony Ferrar, Claire Portugal

SUMMARY

Our group is interested in the 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 PROJECTS

1. Quantitative understanding of *M. pneumoniae*

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

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

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

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

2. Signal Transduction and Disease

Cancer therapy

The efficacy of current treatments for some types of solid tumours is disappointingly poor. Thus, new therapies using novel tumour-selective anti-cancer agents are necessary. A major aim of anti-tumour therapies is to inhibit proliferation and induce death of tumour cells without affecting normal cells. In this regard, members of TNF ligand/receptor family are of interest since they regulate both apoptosis and cell proliferation. One TNF family member, TRAIL, is of particular interest since it selectively induces death of tumour cells without affecting normal cells. Currently, TRAIL and TRAIL-specific antibodies are being investigated as anti-cancer agents. However, one drawback to their efficacy is that they bind to multiple receptors, not all of which transduce an apoptotic signal. Previously, we developed DR5 receptor-selective TRAIL variants, which are potent inducers of apoptosis in various tumour cells, are more efficacious than native TRAIL, and display synergistic effect in combination with other chemotherapy treatments or radiotherapy.

Structural Systems Biology (Understanding Signal Transduction)

Understanding signal transduction pathways is capital for human health. Current efforts to do so involve knock out experiments and the use of small molecules or antibodies to interfere with selected protein-protein interactions along the pathway. However, signal transduction pathways in higher eukaryotes are characterized by the existence of multiple interactions for any of its particular components. 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. Moreover, there is ample evidence that the majority of the drugs in the market are not as specific as thought and they block more than one activity. Here we want to use a different tool that could allow the specific selection and activation, or inhibition, of specific routes in a pathway and also provide much higher specificity than small molecules. This tool is rational protein design. Structures of macromolecules and especially of molecular machines in combination with protein design could provide quantitative parameters, help to elucidate functional networks, or allow rational designed perturbation experiments for reverse engineering. Computer aided protein design has shown to be able to modify in a rational way protein-protein interactions, tuning specificity, affinity and interestingly kinetics of binding. Protein design can also be used to predict using structural information the partners of a particular domain, or protein.

As a scientific target we have selected 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 (Fig 1), determining the concentrations of all proteins in different cell lines and applying different perturbations.

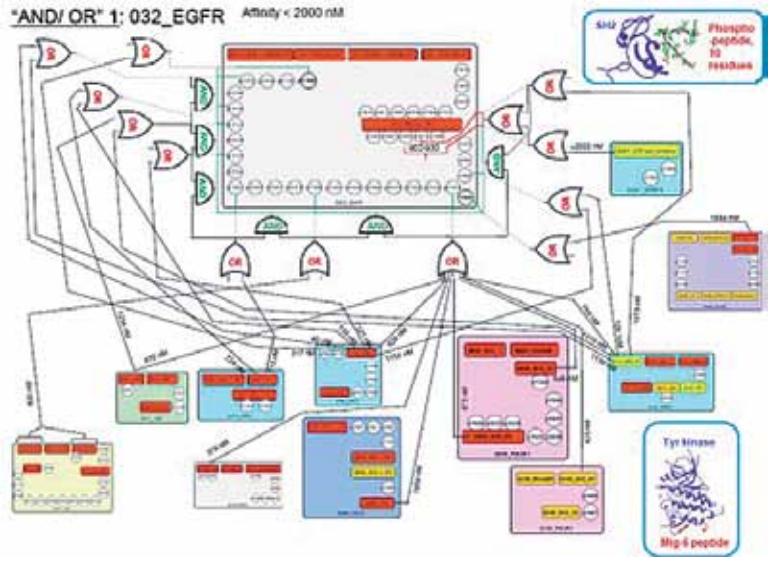


Fig. 1.
Snapshot of part of the EGF-MAPK pathway illustrating the decomposition of the interaction network into competing and non-competing interactions

PUBLICATIONS

Vivancos AP, Guell M, Dohm JC, Serrano L and Himmelbauer H.
"Strand-specific deep sequencing of the transcriptome."
Genome Res, 20:989-99 (2010).

Vanhee P et al.
"PepX: a structural database of non-redundant protein-peptide complexes."
Nucleic Acids Res, 38:D545-51 (2010).

Sanchez-Hidalgo M et al.
"Conformational stability and activity of circular Enterocin AS-48 derivatives."
Protein Pept Lett, 17:708-14 (2010).

Raineri E, Ribeca P, Serrano L and Maier T.
"A more precise characterization of chaperonin substrates."
Bioinformatics, 26:1685-9 (2010).

Maurer-Stroh S et al.
"Exploring the sequence determinants of amyloid structure using position-specific scoring matrices."
Nat Methods, 7:237-42 (2010).

Magis C, Stricher F, van der Sloot AM, Serrano L and Notredame C.
"T-RMSD: a fine-grained, structure-based classification method and its application to the functional characterization of TNF receptors."
J Mol Biol, 400:605-17 (2010).

Kiel C, Yus E and Serrano L.
"Engineering signal transduction pathways."
Cell, 140:33-47 (2010).

Grünberg R and Serrano L.
"Strategies for protein synthetic biology."
Nucleic Acids Res, 38:2663-75 (2010).

Grünberg R, Ferrar TS, van der Sloot AM, Constante M and Serrano L.
"Building blocks for protein interaction devices."
Nucleic Acids Res, 38:2645-62 (2010).

Alibes A, Serrano L and Nadra AD.
"Structure-based DNA-binding prediction and design."
Methods Mol Biol, 649:77-88 (2010).

Alibés A, Nadra AD, De Masi F, Bulyk ML, Serrano L, Stricher F.
"Using protein design algorithms to understand the molecular basis of disease caused by protein-DNA interactions: the Pax6 example."
Nucleic Acids Res, 38(21):7422-31 (2010).

Bertero M, Verschueren E, Paetzold B, Serrrano L.
"Enhanced SnapShot: Macromolecular machines."
Cell, 143(4):652, 652.e1 (2010).

Szegezdi E, Reis CR, Van Der Sloot A, Natoni A, O'Reilly A, Reeve J, Cool RH, O'Dwyer M, Knapper S, Serrano L, Quax WJ, Samali A.
"Targeting AML through DR4 with a novel variant of rhTRAIL."
J Cell Mol Med, 2010 Nov 10. doi: 10.1111/j.1582-4934.2010.01211.x. [Epub ahead of print].

Rakoczy EP, Kiel C, McKeone R, Stricher F, Serrano L.
"Analysis of disease-linked rhodopsin mutations based on structure, function, and protein stability calculations."
J Mol Biol, Epub 2010 Nov 19.

Vanhee P, Verschueren E, Baeten L, Stricher F, Serrano L, Rousseau F, Schymkowitz J.
"BriX: a database of protein building blocks for structural analysis, modeling and design."
Nucleic Acids Res, Epub 2010 Oct 23.

Muñoz IG, Prieto J, Subramanian S, Coloma J, Redondo P, Villate M, Merino N, Marenchino M, D'Abramo M, Gervasio FL, Grizot S, Daboussi F, Smith J, Chion-Sotinel I, Pâques F, Duchateau P, Alibés A, Stricher F, Serrano L, Blanco FJ, Montoya G.
"Molecular basis of engineered meganuclease targeting of the endogenous human RAG1 locus."
Nucleic Acids Res, Epub 2010 Sep 16.

Brooks MA, Gewartowski K, Mitsiki E, Létoquart J, Pache RA, Billier Y, Bertero M, Corréa M, Czarnocki-Cieciura M, Dadlez M, Henriot V, Lazar N, Delbos L, Lebert D, Piwowarski J, Rochaix P, Böttcher B, Serrano L, Séraphin B, van Tilbeurgh H, Aloy P, Perrakis A, Dziembowski A.
"Systematic bioinformatics and experimental validation of yeast complexes reduces the rate of attrition during structural investigations."
Structure, 18(9):1075-82 (2010).

SYSTEMS BIOLOGY

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

Group structure:

Group Leader: James Sharpe

Staff Scientist: Jim Swoger

Postdoctoral Fellows: Jean-François Colas, Henrik Westerberg, James Cotterell, Niamh Nowlan, Rachele Allena

Technicians: Laura Quintana, Martina Niksic

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

SUMMARY

The Sharpe lab has 2 primary goals:

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

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

RESEARCH PROJECTS

1. Development of novel 3D and 4D optical imaging technologies

In addition to OPT, which remains a central technology for the lab, we have also now designed and constructed a SPIM apparatus (selective plane illumination microscope), for which Dr. Jim Swoger was one of the original co-inventors (Huiskin et al, 2004 Science 305:1007). We are currently exploring a wide range of imaging applications for this technology, including samples from mouse, chick, zebrafish and *Drosophila*), and this year published live time-lapse imaging results on Zebrafish, in collaboration with the Lopez-Scheir group (Swoger et al, J. Biophotonics 2010). The exploration of possible applications is also being extended by collaborating with a variety of researchers across Europe.

2. OPT imaging for quantitative assessment of mouse models of diabetes

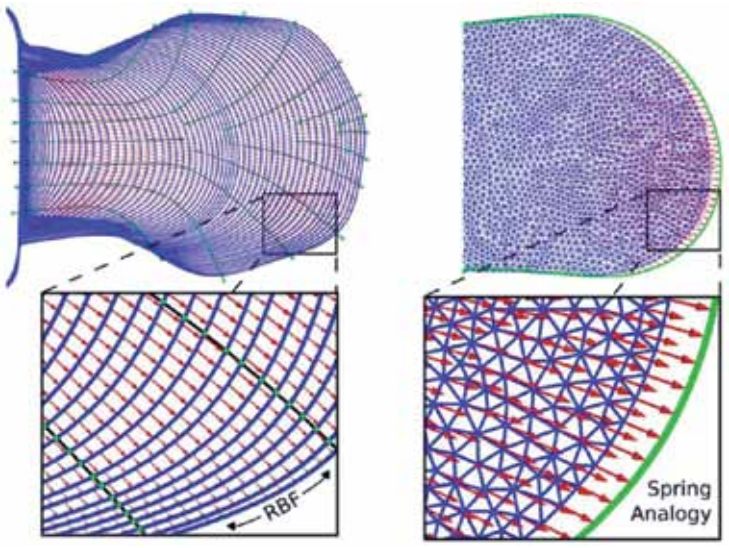
We have explored many new applications for OPT, and one of the most exciting has been the ability to quantify the number of Islets of Langerhans in an intact adult mouse pancreas in a single scan. Until now this has been performed using the time-consuming approach of traditional histology (cutting hundreds of thin paraffin sections for each pancreas). By contrast, we have demonstrated that the speed of OPT makes it feasible to compare many pancreata in a single study, Nature Methods (2007), Diabetes (2010). We are also partners in a 7FP EU integrated project entitled VIBRANT, with the goal of further improving the resolution of OPT and SPIM. In particular, we aim to use 3D optical imaging as a verification technique for new non-optical *in vivo* imaging techniques (such as MRI) being developed within the consortium, to estimate total beta-cell mass within living animals, and ultimately human patients.

3. 4D Time-lapse imaging of limb bud development

Optical projection tomography has proven to be a powerful tool for developmental biologists. But until recently OPT has been performed almost exclusively on fixed specimens which have been optically cleared to increase the quality of 3D images obtained. Recently we have explored various improvements to allow the 4D imaging of the developing mouse limb bud in culture (Nature Methods, 2008; Organogenesis, 2009), and this technique now provides us with data on tissue movements and dynamic gene expression patterns. During 2010 we have also used in-ovo time-lapse imaging of chick limb buds to observe this process at the cellular leve (Boehm et al. PLoS Biology, 2010).

4. Computer model of limb mechanics

It has become clear from early modelling results that explaining phenomena as apparently simple as the dorso-ventral flattening of the limb bud, may be more complicated than previously thought. We have created a finite-element model (FEM) of limb development within which we are exploring different hypotheses. An essential aspect of this project is that we also perform lab-work to generate our own empirical data for the simulation – the model therefore serves as a framework for combining different types of information. While various types of biological material have previously been mechanically modelled over a short time periods (for example stress analysis on bones and cartilage) a mathematical/physical description of 3D embryonic tissue displaying volumetric growth over a period of hours or days has not previously been achieved, and this is therefore one of the general goals of this project. Our first 3D FEM on a short phase of development has been published this year (Boehm et al. PLoS Biology, 2010), and we have also developed a 2D description of the tissue movements over the full period of limb bud growth.



5. Computer models of spatially-patterning gene networks

The interpretation of morphogen gradients is a pivotal concept in developmental biology and several mechanisms have been proposed to explain how gene regulatory networks (GRN's) achieve concentration-dependent responses. However, the number of different mechanisms that may exist for cells to interpret morphogens, and the importance of design features such as feedback or local cell-cell communication, is unclear. A complete understanding of such systems will require going beyond a case-by-case analysis of real morphogen interpretation mechanisms and mapping out a complete GRN "design space". We are therefore performing theoretical projects which search for the network design principles in a 1D simulation of pattern formation. We generated an atlas of design space for GRN's containing 3 genes capable of patterning a homogeneous field of cells into discrete gene expression domains by interpreting a fixed morphogen gradient. We uncovered multiple very distinct mechanisms distributed discretely across the atlas, thereby expanding the repertoire of morphogen interpretation network motifs. Analyzing this diverse collection of mechanisms also allowed us to predict that local cell-cell communication will rarely be responsible for the basic dose-dependent response of morphogen interpretation networks.(Cotterell & Sharpe, Molecular Systems Biology 2010)

PUBLICATIONS

Boehm B, Rautschka M, Quintana L, Raspopovic J, Jan Z, Sharpe J.
"A landmark-free morphometric staging system for the mouse limb bud."
Development, in press.

Marcon L, Arqués CG, Torres MS, Sharpe J.
"A computational clonal analysis of the developing mouse limb bud."
PLoS Comput Biol, in press.

Cotterell J, Sharpe J.
"An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients."
Molecular Systems Biology, 6:425 (2010).

Swoger J, Muzzopappa M, Lopez-Schier H and Sharpe S.
"4D retrospective lineage tracing using SPIM for zebrafish organogenesis studies."
Journal of Biophtonics, doi: 10.1002/jbio.201000087. Epub 2010 Oct 5.

Capellini TD, Vaccari G, Ferretti E, Fantini S, He M, Pellegrini M, Quintana L, Di Giacomo G, Sharpe J, Selleri L, Zappavigna V.
"Scapula development is governed by genetic interactions of Pbx1 with its family members and with Emx2 via their cooperative control of Alx1."
Development, 137(15):2559-69 (2010).

Boehm B, Westerberg H, Lesnicar-Pucko G, Raja S, Rautschka M, Cotterell J, Swoger J, Sharpe J.
"The role of spatially controlled cell proliferation in limb bud morphogenesis."
PLoS Biol, 8(7):e1000420 (2010).

Alanentalo T, Hörnblad A, Mayans S, Karin Nilsson A, Sharpe J, Larefalk A, Ahlgren U, Holmberg D.
"Quantification and three-dimensional imaging of the insulitis-induced destruction of beta-cells in murine type 1 diabetes."
Diabetes, 59(7):1756-64 (2010).

Jimenez-Guri E, Udina F, Colas JF, Sharpe J, Padrón-Barthe L, Torres M, Pujades C.
"Clonal analysis in mice underlines the importance of rhombomeric boundaries in cell movement restriction during hindbrain segmentation."
PLoS One, 5(4):e10112 (2010).

Review

Nowlan NC, Sharpe J, Roddy KA, Prendergast PJ, Murphy P.
"Mechanobiology of embryonic skeletal development: Insights from animal models."
Birth Defects Res C Embryo Today, 90(3):203-213 (2010).

SYSTEMS BIOLOGY

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

Group structure:

Group Leader: Mark Isalan

Postdoctoral Fellows: Mireia Garriga, Phil Sanders, Yolanda Schaerli (Started Oct 2010), Manjunatha Kogenaru (Started Nov 2010), Emmanuel Fajardo (left Jul 2010), Frank Herrmann (left Oct 2010)

Students: Andreia Carvalho, Marco Constante

Technician: Rebecca Baumstark

SUMMARY

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

RESEARCH PROJECTS

1. Shuffling *E. coli* transcription

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

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

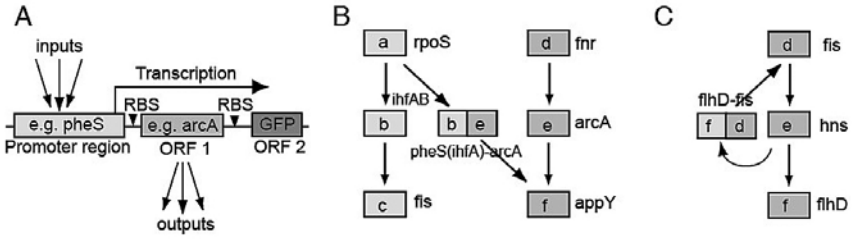


Figure 1.

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

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

Reference:

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

The propagation of perturbations in shuffled networks

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

2. Patterning Gene Networks

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

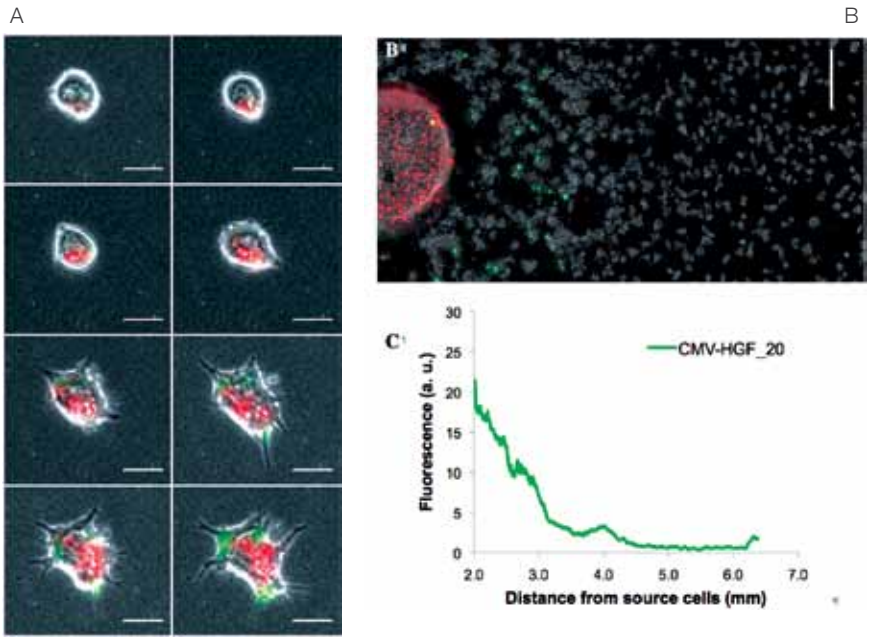


Fig. 2.
Synthetic 3D 'morphogen' patterns in MDCK cysts.
A. *Localized HGF transfection (red), followed by HGF secretion, distal induction of GFP expression and tubule formation, in an MDCK cyst containing a reporter gene network.*

B*. Spatial expression of GFP over fields of cysts, in response to a localized source of HGF (secreting cells; expressing a red fluorescent protein). 24hrs after seeding the HGF-secreting cells, the cysts surrounding the sender region express GFP and undergo tubule formation. Cysts further way from the sender cells do not express GFP and do not tubulate or scatter.*
C. Quantification of the expression of GFP using imageJ (Radial profile plugin) which measures average GFP intensity radially, from a source. Bars, 1 mm.

3. Synthesising zinc fingers for genome engineering and gene repair

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

PUBLICATIONS

Munteanu A, Constante M, Isalan M, Sole RV.
"Avoiding transcription factor competition at promoter level increases the chances of obtaining oscillations."
BMC Systems Biology, 4:66 (2010).

PATENTS

Garriga-Canut M and Isalan M.
"Peptides and Uses."
European patent application no. 10187818.9 (2010).

SYSTEMS BIOLOGY

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

Group structure:

Group Leader: Ben Lehner

Postdoctoral researchers: Jennifer Semple, Tanya Vavouri, Rob Jelier, Olivia Casanueva, Benjamin Schuster-Böckler, Mirko Francesconi

PhD students: Alejandro Burga Ramos, Angela Krüger, Adam Klosin, Kiana Toufighi (co-supervised with L. Serrano)

Technician: Laura Biondini

SUMMARY

Our main interest is in the genetics of individuals. As part of this we are working on methods to predict phenotypic variation from individual genome sequences. This means predicting both the typical outcome of a sequence variant, and the actual outcome in each individual. Mutations often have different outcomes in different individuals, and we aim to understand the genetic, stochastic and environmental causes of this variation. As experimental systems we use the model organisms *C. elegans* and *S. cerevisiae*.

RESEARCH PROJECTS

1. Predicting how mutations combine to cause phenotypic change

Most human diseases are genetically complex, with contributions from mutations in many different genes. Mutations in more than one gene can combine synergistically to cause phenotypic change, and systematic studies in model organisms show that these genetic interactions are pervasive. However, in human association studies such non-additive genetic interactions are very difficult to identify because of a lack of statistical power. One approach to resolve this is to predict candidate modifier interactions between loci, and then to specifically test these for associations with the phenotype. We have tested and evaluated this approach ('network guided modifier screening') in both *S. cerevisiae* and *C. elegans*, finding that a single high-coverage, high-quality functional network can successfully predict genetic modifiers for the majority of genes. Using this network we demonstrated that it is possible to rapidly expand the number of modifier loci known for a gene, predicting and validating new genetic interactions for signal transduction genes.

2. Genes confer correlated robustness to genetic, environmental and stochastic perturbations

Gene inactivation often has little apparent consequence for the phenotype of an organism. This property, genetic (or mutational) robustness, is pervasive, and has important implications for disease and evolution, but is not well understood. Dating back to at least Waddington it has been suggested that mutational robustness may be related to the requirement to withstand environmental or stochastic perturbations. We have shown that global quantitative data from yeast are largely consistent with this idea. Considering the effects of mutations in all non-essential genes shows that genes that confer robustness to environmental or stochastic change also buffer the effects of genetic change, and with similar efficacy. This means that selection during evolution for environmental or stochastic robustness (also referred to as canalization) may frequently have the side effect of increasing genetic robustness. A dynamic environment may therefore promote the evolution of phenotypic complexity. It also means that "hub" genes in genetic interaction (synthetic lethal) networks are generally genes that confer environmental resilience and phenotypic stability.

3. Noise and plasticity in gene expression

Gene expression responds to changes in conditions but also stochastically among individuals. In budding yeast, both expression responsiveness across conditions ("plasticity") and cell-to-cell variation ("noise") have been quantified for thousands of genes and found to correlate across genes. It has been argued therefore that noise and plasticity may be strongly coupled and mechanistically linked, which would have important consequences for the evolution of transcription networks (Lehner, Molecular Systems Biology 2008). This proposal is also consistent with some theoretical ideas, but a strong coupling between noise and plasticity also has the potential to introduce cost-benefit conflicts during evolution. For example, if high plasticity is beneficial (genes need to respond to the environment), but noise is detrimental (fluctuations are harmful), then strong coupling should be disfavoured. By analysing existing datasets, we found evidence that cost-benefit conflicts do occur and these conflicts constrain the evolution of gene expression and promoter architecture. Further, the global quantitative data in yeast suggest that one mechanism that relieves the constraints imposed by noise-plasticity coupling is gene duplication, providing an example of how duplication can facilitate escape from adaptive conflicts.

4. Whole-animal antibiotic selection

C. elegans is one of the mostly widely used model organisms. However the generation of transgenic nematodes is still a rather laborious process. We have developed a selection system using antibiotics that allows the rapid and easy isolation of large populations of transgenic worms. This approach is sufficiently powerful to select single-copy transgenes, does not require any particular genetic background, and works in other species such as *C. briggsae.*

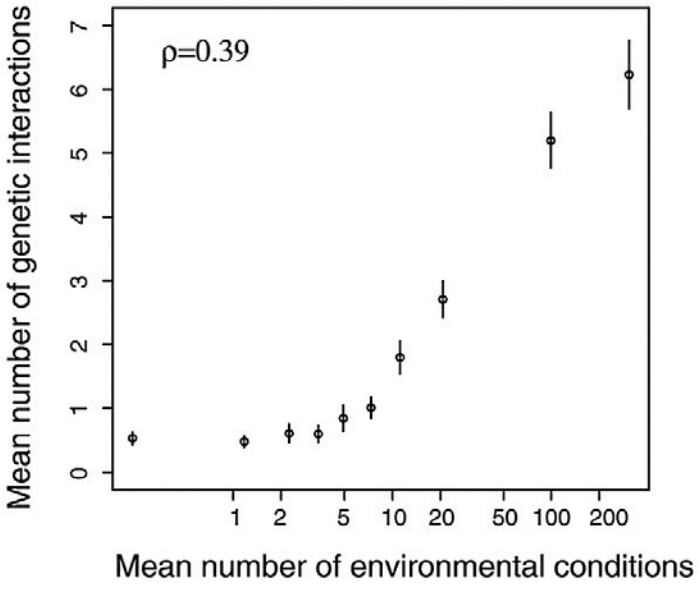


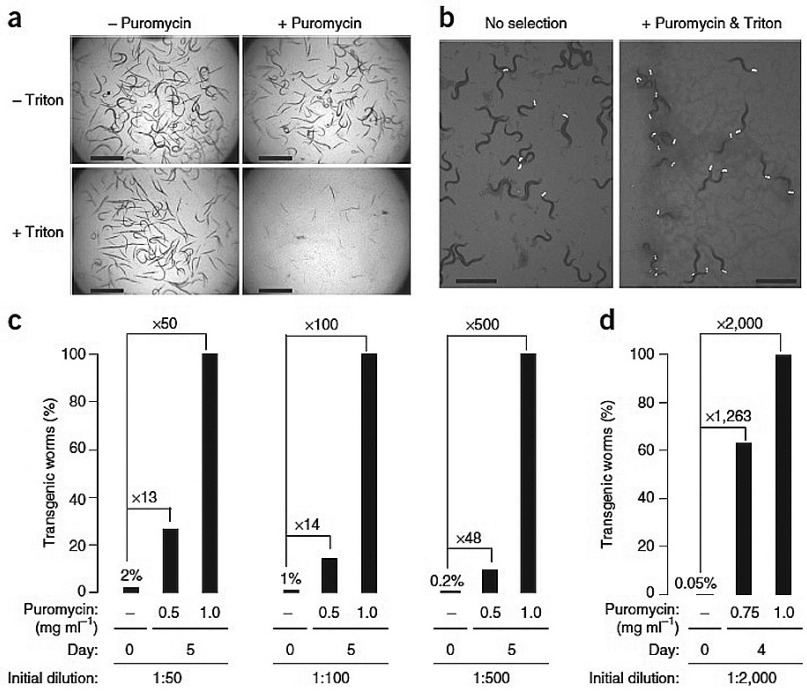
Fig 1.
The correlated effects of gene deletions on genetic and environmental robustness in yeast. *The effects of mutations on environmental robustness (the number of different environmental conditions in which a gene is required for growth) and mutational robustness (the number of synthetic lethal interactions made by a mutation) are compared across 4656 gene deletions in yeast. Data are plotted for ten equally sized bins of genes. Error bars are +/– one standard error. Spearman rank correlation coefficient across all genes (ρ) = 0.39, $p<2.2\times10^{-16}$.*

Fig 2.
Puromycin selection in C. elegans.

(a) *Cultures initiated with wild type L1 larvae (15 worms µl^{-1}) after 4 days of growth with or without 0.5 mg ml^{-1} puromycin and 0.1% Triton X-100, in the presence of DH101 bacteria.*

(b) *Nematodes of the BCN4004 strain carrying an extra-chromosomal array with the puromycin resistance gene and a Pmyo2::mCherry marker. L1 larvae were grown for 4 days in liquid medium with or without selection, transferred to non-selective plates seeded with bacteria, allowed to grow for 3 days and photographed.*

(c&d) *Enrichment of transgenic worms by puromycin selection. Cultures were initiated with various dilutions of BCN2081 L1 larvae (carrying a single copy integrated puromycin resistance transgene) with wild type N2 larvae. Plots show the percentage of BCN2081 transgenic worms after 5 (c) or 4 (d) days of selection under the indicated conditions. The fold-increase in the percentage of transgenic worms after selection is shown for each experiment. Scale bar = 1 mm.*



PUBLICATIONS

Articles

Lehner B.
"Conflict between noise and plasticity in yeast."
PLoS Genetics, 6(11):e1001185 (2010).

Semple JI, Garcia-Verdugo R, Lehner B.
"Rapid selection of transgenic C. elegans using antibiotic resistance."
Nature Methods, 7(9):725-7 (2010).

Lee I, Lehner B, Vavouri T, Shin J, Fraser AG, Marcotte EM.
"Predicting genetic modifier loci using functional gene networks."
Genome Research, 20(8):1143-53 (2010).

Lehner B.
"Genes confer similar robustness to environmental, stochastic, and genetic perturbations in yeast."
PLoS One, 5(2):e9035 (2010).

Review

Lehner B, Kaneko K.
"Fluctuation and response in biology."
Cell Mol Life Sci, 2010 Nov 30.

SYSTEMS BIOLOGY

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

Group structure:

Group Leader: Matthieu Louis

Postdoctoral Fellows: Alex Gomez-Marin, David Jarriault

Students: Julia Riedl, Aljoscha Schulze, Andreas Braun

Visiting Undergrad: Biafra Ahanonu

Technicians: Verena Hotz, Mariana Lopez-Matas, Moraea Philipps

SUMMARY

The mission of our group is to unravel structure-function relationships between neural circuits, sensory coding and adaptive behaviours. We are interested in understanding how odour tracking comes about in terms of circuit computation. This problem is tackled in the olfactory system of the fruit fly *Drosophila melanogaster* larva. Our research combines a variety of experimental and computational techniques aiming to define how naturalistic odorant stimuli are encoded and processed in the olfactory system. In particular, we are developing new computational tools to monitor and classify stereotypical behaviours automatically. We are seeking to identify circuits involved in the processing of olfactory information and the making of decisions underlying chemotaxis.

RESEARCH PROJECTS

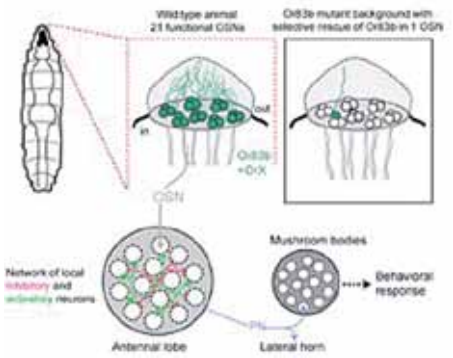
1. Peripheral representation of odour stimuli

David Jarriault, Mariana Lopez-Matas, Moraea Philipps, Julia Riedl, Aljoscha Schulze

The larval 'nose' is composed of 21 olfactory sensory neurons (OSNs) expressing typically one type of odorant receptors along with a ubiquitously expressed co-receptor (Or83b) (Fig 1). Previously, we have demonstrated that individual odorant receptors have overlapping, yet distinct, ligand tuning properties (Asahina, Louis et al., 2009). Accordingly, each OSN can be viewed as distinct information channel to the olfactory system. We have also shown that the information transmitted by a single functional OSN is sufficient for the detection of minute changes in odour intensity. Performing electrophysiological recordings on single OSNs, we are characterizing how odour intensity and quality is represented in the activity of single OSNs. Whether one type of odorant receptor alone is capable of encoding the quality of an odour remains an open question in the larva. We have undertaken to disentangle the contribution of single OSNs to the representation of static and dynamic odour stimuli (Fig 1). In Pavlovian conditioning experiments, we found evidence that a single functional OSN is sufficient to mediate odour quality discrimination. This finding highlights the surprisingly high coding potential of single sensory neurons.

Odours are represented by dynamical patterns of neural activity transmitted from the OSNs to higher brain centres (Fig 1). We have devised a novel preparation to carry out extracellular and intracellular recordings from the peripheral olfactory system. Using this technique, we have begun to compare the activity elicited in identified OSNs. Our current goal is to explain the ability (or inability) of larvae with a single functional OSN to discriminate between distinct odours. By taking advantage of new odour delivery systems that allow us to generate temporal changes in odour concentration during electrophysiological recordings, we are now in position to study the encoding of naturalistic odour stimuli in the peripheral olfactory system.

Figure 1.
Larval 'nose' (dorsal organ) comprising 21 olfactory sensory neurons (OSNs). All OSNs are silenced in the ***Or83b*** *mutant background. Each of the 21 OSNs projects to one glomerulus in the antennal lobe. Every glomerulus is innervated by a single projection neuron (PN). Each projection neuron sends its axon to a subset of the glomeruli in the Mushroom bodies (MB). Expression of Or83b is necessary for the proper trafficking of all ORs to the OSN dendrites. As all OSNs are silenced in the Or83b null background, animals with a single functional OSN are generated by selectively restoring the expression of Or83b in a given OSN.*



We are also interested in characterizing how odour-specific activity patterns evolve across different layers of the olfactory system. Functional imaging is applied to monitor the activity in neuronal ensembles. By expressing genetically encoded calcium sensors (GCaMP) in targeted neurons, we seek to establish precise correlations between sensory input and circuit activity. Our goal is to clarify the role of local inhibitory interneurons in the antennal lobe (Fig 1), a circuit that is thought to achieve gain control and decorrelation. Our long-term goal is to extend this analysis to motor control, and to propose a model for how ethologically relevant sensory input (e.g., increase or decrease in the concentration of a food odour) induce specific behavioural output (e.g., orientation towards the odour source).

2. Behavioral algorithms directing decision-making in chemotaxis

Alex Gomez-Marin

Bacteria and nematodes chemotax according to *indirect* orientation mechanisms, which consists in improved biased random walks. In contrast, *Drosophila* larvae employ a *direct* orientation mechanism where motion is locally aligned with the odour gradient. Previously, we have demonstrated that *Drosophila* larvae do not require bilateral olfactory inputs to perform chemotaxis. Having ruled out a mechanism purely based on stereo sampling (comparisons between the left and the right inputs), we are now investigating how larvae with unilateral olfactory function are capable of extracting directional information from spatially distributed odour cues.

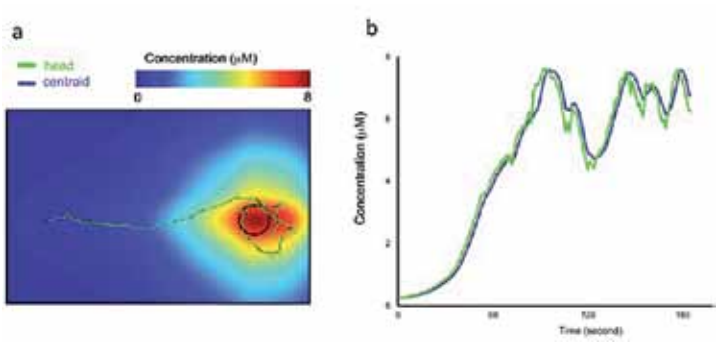


Figure 2.
Correlating sensory input with behavioural output during larval chemotaxis.
(a) High-resolution analysis of the trajectory of larva. The position of the head is coloured in green. The centroid (centre of mass) is coloured in blue. The trajectory is superimposed on a 2-diimensional reconstruction of an odour gradient of ethyl butyrate. A unique source is centred at right of the arena. Absolute concentration of the stimulus in gaseous phase was measured by FT-IR spectroscopy.
(b) Reconstructed time course of the odour concentation experienced by the head and centroid during the trajectory on the left.

We have dissected the sensorimotor algorithm governing larval chemotaxis. Combining high-resolution quantification of olfactory input and behavioural output, we have shown that larval chemotaxis is an active sampling process analogous to sniffing in vertebrates. Larvae orient in odour gradients through a sequential organization of behavioural modes: runs, stops, lateral head sweeps (or head casts) and directed turns. We have found that stereotypical patterns in the history of stimulus control two classes of decisions: when to turn or where to turn. Positive gradients are detected through high-amplitude head casts prior to a turn. Computer-vision analyses permitted us to generate large datasets for individual larvae engaged in chemotactic tasks. By genetically reengineering the peripheral olfactory circuit, we have examined how orientation adapts to losses and gains of function in the olfactory circuit. Our findings suggest that larval chemotaxis represents an intermediate navigation strategy between the biased random walks of *E. coli* and *C. elegans*, and the stereo-olfaction observed in vertebrates such as rats and humans.

3. Remote control via optogenetic stimulation and closed-loop high-resolution tracking

Alex Gomez-Marin and Aljoscha Schulze

To test our current model about the active sampling and decision-making processes controlling navigation in odour gradients, we use optogenetics to interfere with the sensory experience of freely moving larvae. We make use of light-activated ion channels such as channelrhodopsin-2 (ChR2). Ectopic expression of ChR2 in single OSNs provides a powerful technique to control olfactory input at a single spike level. We have developed a system capable of measuring neuronal activity in single OSNs in response to controllable olfactory or light-evoked stimuli. By combining these technical advances, we have undertaken a quantitative characterization of synthetic olfactory stimuli and we are now in a position to reverse engineer naturalistic dynamic odour stimuli.

To monitor larval locomotion at very high temporal and spatial resolutions, we have devised a new tracking system. This closed-loop tracker is based on a moving stage that carries a camera and several LEDs: it allows us to monitor the motion of a single animal (0.4 cm in length) across a large area (over 1 m^2). The posture of the larva is calculated in real time at a rate of 30Hz and the software triggers light flashes based on a stimulation protocol previously specified by the experimenter. Using this tool, our ultimate objective is to combine real-time high-resolution tracking and optogenetics to remote control the orientation behaviour of a larva.

4. Identification of new centres participating to orientation behaviour

Verena Hotz and Julia Riedl

To map the circuits participating in larval chemotaxis, we have initiated a behavioural screen (Fig 3). We take advantage of thousand of existing Gal4-drivers to study the effects of disabling sub-regions of the larval brain by expressing the tetanus toxin light chain (inhibitor of synaptic transmission). While specific neurons are silenced, we test the orientation capabilities of mutants to directional sensory stimuli. The anatomy and function of neuronal subsets associated with interesting phenotypes are inspected further. We use functional calcium imaging and high-resolution behavioural analysis to clarify the specific role of neurons involved in the processing of olfactory signals. In addition, remote controlled activation using specific effectors (e.g. ChR2 and TrapA1) is exploited to establish new circuit-function relationships underlying larval chemotaxis.

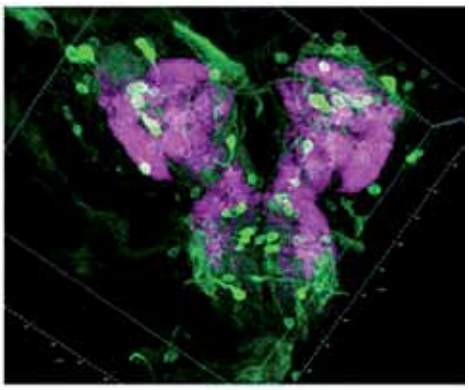


Figure 3.
Newly identified neural circuit potentially involved in chemotaxis behaviour. Immunostaining against mcd8-EGFP (green) and the synaptic marker nc82 (magenta)

PUBLICATIONS

Gomez-Marin A*, Duistermars BJ*, Frye MA and Louis M.
"Mechanisms of odour-tracking: multiple sensors for enhanced perception and behavior."
Frontiers in Cellular Neuroscience, 4:6 (2010). [* equal contribution]

Dushyant M, Louis M and Gerber B.
"Adaptive adjustment of Generalization-Discrimination Balance in Larval Drosophila",
Journal of Neurogenetics, 24:168 (2010).

SYSTEMS BIOLOGY

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

Group structure:

Group Leader:	Johannes Jäger
Technician/Lab Manager:	Hilde Janssens
Technician/Programmer:	Damjan Cicin-Sain
Postdoctoral Fellows:	Anton Crombach, Eva Jiménez Guri, Bárbara Negre, Karl Wotton
PhD Student:	Astrid Hörmann
Fulbright Fellow:	Maria Iliakova

SUMMARY

Natural selection acts on phenotypic variability within populations. However, we still lack a coherent view of how such variability arises during development, and how it reflects molecular variation in the genome. The relationship between genotype and phenotype is complex and non-linear. Therefore, we need a systems-biology approach to address this question. We are carrying out an integrative, comparative analysis of real evolving developmental gene regulatory networks using a novel reverse-engineering approach (the gene circuit method). Gene circuits are computational tools to extract regulatory information from quantitative spatial gene expression data (Fig. 1). We study the evolution of the following networks in dipteran insects (flies, midge, and mosquitoes): the gap gene network involved in pattern formation in the early embryo of dipterans, the gene network underlying muscle and heart development during organogenesis, and the thoracic patterning network responsible for the positioning of mechanosensory bristles on the dorsal cuticle of the animal (Fig. 2A). We are establishing three dipteran species—the fruit fly *Drosophila melanogaster*, the scuttle fly *Megaselia abdita* and the moth midge *Clogmia albipunctata*—as model systems to experimentally and quantitatively test hypotheses derived from systems-biology approaches to evolutionary developmental biology (Fig. 2B). Our models allow us to infer the regulatory interactions necessary and sufficient to explain the observed expression patterns by fitting models to data. Models from different species can be compared to reveal which interactions are conserved and which have diverged during evolution. In addition, we study evolutionary transitions between species using an *in silico* evolution approach. We will test these predictions by using RNA interference (RNAi) in various species and reporter assays in *Drosophila*. Our approach provides an integrative view of network evolution across multiple levels, from the molecular to the phenotypic. To our knowledge, this has not yet been achieved for any real developmental system.

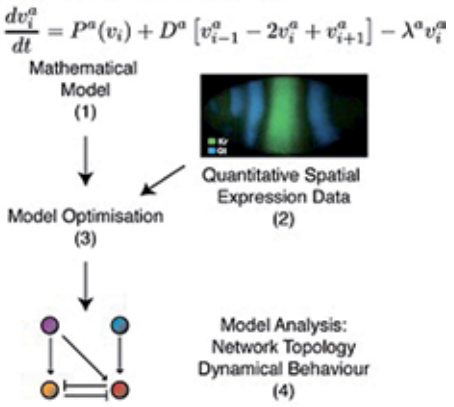


Fig.1: *The Gene Circuit Method*

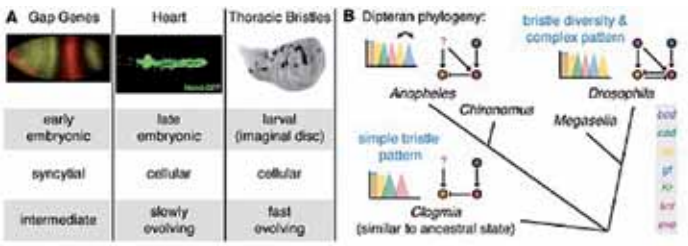


Fig. 2: ***(A) Evolving developmental gene regulatory networks. (B) Trends in dipteran evolution:***
The thoracic bristle patterning network has evolved from simple to complex, diversified patterns . The gap network shows convergent evolution: Gap domains are shown schematically (anterior is to the left).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

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

Astrid Hörmann, Hilde Janssens

Aim. Due to the lack of a quantitative *in vitro* transcription assay, we still do not have a precise mechanistic understanding of eukaryotic gene regulation. For this reason, our current gene network models do not include molecular details, such as specific transcription factor binding sites and *cis*-regulatory elements. As a first step towards resolving this issue, we will look into the regulation of a particular gap gene (*gt*) using our reverse-engineering approach with a 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 will quantify gene expression in each of these strains, and will use the resulting data sets to fit a model of transcriptional regulation. This will allow us to predict the contribution of particular binding sites to the expression of specific *gt* domains. These predictions will then be tested experimentally.

2. A quantitative study of gap gene mutants in *Drosophila melanogaster*

Hilde Janssens

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

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

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

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

Aim. We aim to create gene circuit models of the gap gene network in *Drosophila*, *Megaselia* and *Clogmia*. To achieve this, we require detailed characterization of the early stages of development, as well as quantitative spatial gene expression patterns for all three species. Such data sets can be acquired using immunofluorescence (for protein) or *in situ* hybridization (for mRNA expression patterns). These data will be analyzed to characterize spatial variability of gene expression domains and gene expression dynamics. Gene network models will be obtained by fits to data (Fig. 1). We will characterize and compare the dynamical behavior of these networks. The resulting gene network topologies will be used to predict, which aspects of the network have been conserved, and which ones diverged during evolution. These predictions will be verified using RNAi and reporter assays.

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

patterns for all three species (Fig. 3). These data are now being used to fit gene circuit models, and will serve as a basis for our *in silico* evolution study (see below). We have obtained models from mRNA data in *Drosophila*, and have shown that they are equivalent to models fit to protein data (Fig. 3). We are systematically comparing and improving numerical algorithms required for model fitting, solution and analysis.

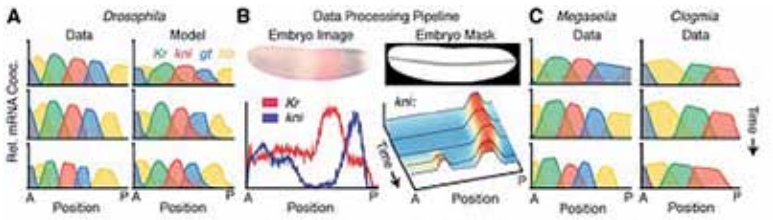


Fig. 3: *(A) Drosophila gap gene circuit models vs. mRNA data. (B) Data processing pipeline: an embryo mask is used to identify a 10% stripe along the midline of the embryo along which expression profiles (lower left) are measured; expression boundaries are approximated by polynomial functions to yield an integrated dataset (lower right) .(C) Megaselia and Clogmia datasets. A = anterior, P = posterior*

4. Modeling the evolutionary dynamics of the gap gene network

Anton Crombach

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

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

5. Modeling development and evolution of the fly heart

Eva Jiménez-Guri, Damjan Cicin-Sain

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

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

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

Bárbara Negre

Aim. Another interesting example of an evolving network is the thoracic bristle patterning network. It is responsible for the positioning of mechanosensory bristles on the thorax of the fly during larval development and pupation. In contrast to both gap and heart networks, it is evolving extremely fast. Thoracic bristle patterns are used as species-defining traits. 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 have initiated 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. This project has only started in November 2010. We have established the experimental protocols required for the expression screen, and have ordered a large set of candidate clones for the screening process.

PUBLICATIONS

Jostins L and Jaeger J.
"Reverse engineering a gene network using an asynchronous parallel evolution strategy."
BMC Syst Biol 4:17 (2010).

García-Solache M, Jaeger J and Akam M.
"A systematic analysis of the gap gene system in the moth midge Clogmia albipunctata."
Dev Biol, 344:308–18 (2010).

Jaeger J.
"A matter of timing and precision."
Mol Syst Biol, 6:427 (2010).
[comment on: Morton de Lachapelle A & Jaeger J & Monk NAM. *"Reverse Engineering of Gene Regulatory Networks."* In: Learning and Inference in Computational Systems Biology, Lawrence ND, Girolami M, Rattray M & Sanguinetti G (eds.), MIT Press, Cambridge, MA.Bergmann S (2010). Mol Syst Biol 6:351]

Jaeger J.
"Neither fish nor fowl: Review of "The algorithmic beauty of sea shells""
4th edn. by Hans Meinhardt.
BioEssays 32: 638–40.

Review

Jaeger J.
"The Gap Gene Network."
Cell Mol Life Sci, Epub 2010 Oct 8.

ANTIBODY PROJECT

Group Leader: Luis Serrano, Doris Meder

Technicians: Katie Broadbent, Martin Gigirey

SUMMARY

The aim of the antibody project is to create a library of human transcription factors which are validated by Elisa, western blot and immunofluresence. Researchers within the CRG have access to these antibodies to aid their research. Also the group provides an antibody production service whereby laboratories can design and order specific antibodies of interest. Finally the antibody project commercializes these various antibodies to external companies.

Transcription Factor collection

We have a collection of approximately 300 polyclonal antibodies and a further 100 polyclonal antibodies (not transcription factors). Of these 400 antibodies, approximately 300 have been purified. We also have a collection of 88 purified monoclonal antibodies (transcription factors).

We have a page on the intranet showing which antibodies we have in our collection and the status of the validation process for each and any available results. This page is updated regularly and researchers can use it to request antibodies from us.

In 2010 we have continued with the validation of these polyclonal antibodies by western blot. We have also screened the 300 polyclonal antibodies by Immunofluresence and tested all purified monoclonal antibodes by Elisa. Validation work is still ongoing for both the polyclonal and monoclonal collections.

Service to groups within the CRG for antibody design and production

We have a service available for laboratories interested in designing a specific antibody which perhaps is not commercially available. In these cases the researcher will use the Goldquest 2 program available on our intranet to design 3 peptides sequences specific to the protein of interest. Once the peptides have been designed, the sequences are forwarded by the antibody project to Eurogentec who then produce the peptides and the antibodies. The antibodies can also be purified if required.

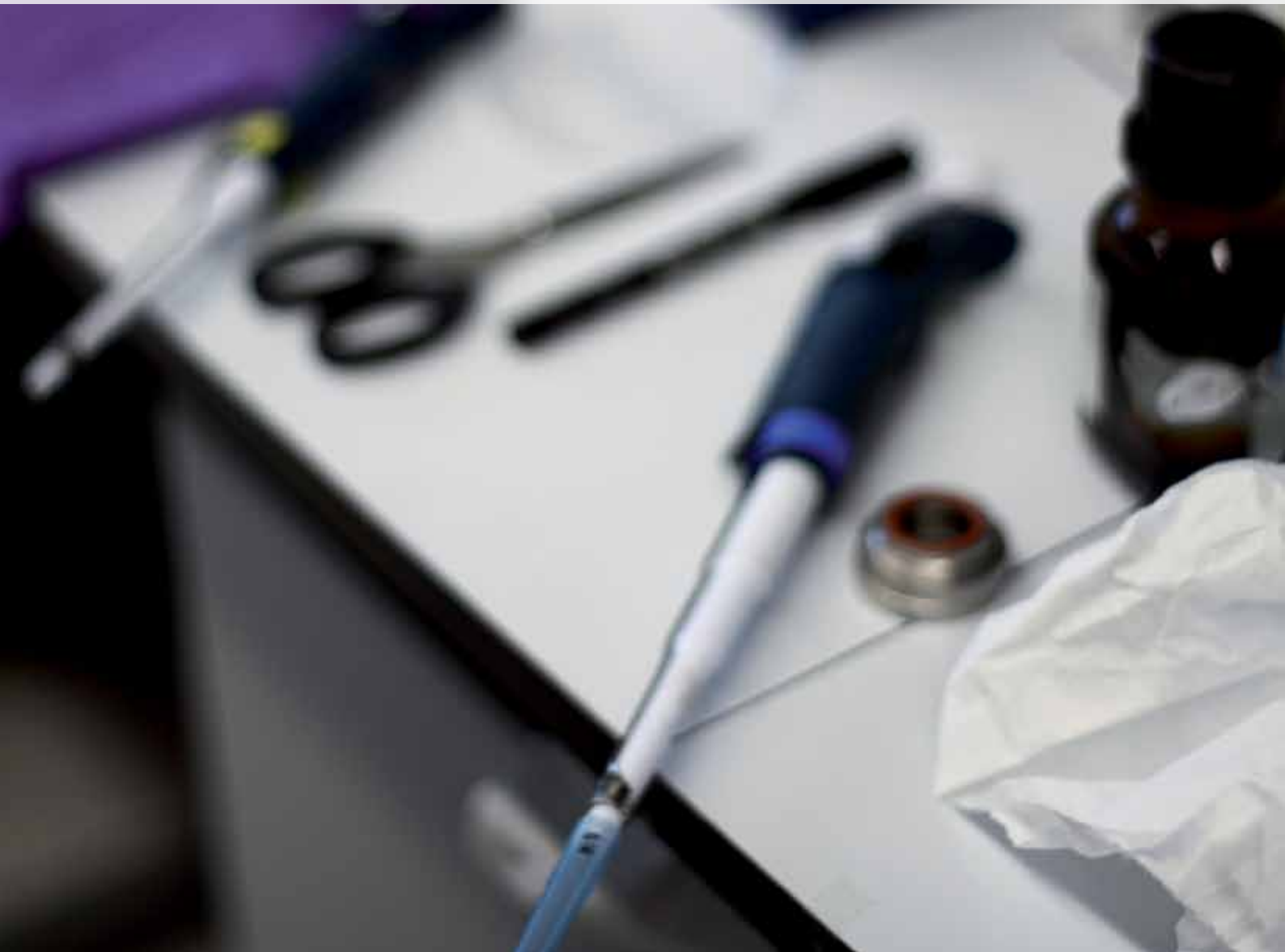
In 2010 this antibody production service has been used to design around 100 antibodies for the different groups within the CRG. These antibodies have been validated in the antibody project by Elisa and further validation by western blot is in process.

Commercialisation

The antibody project is working with the technology transfer group in the commercialization of various antibodies to external companies. On the CRG website, external companies can access which antibodies we have currently available for commercialization, (polyclonal and monoclonal), and which will be available in the near future. The website is updated frequently by us and also we prepare the aliquots and datasheets when commercialized. Currently we are selling our antibodies to Abcam and Millipore.

CORE FACILITIES

Director: Doris Meder



The program currently comprises eight Core Facility Units, Genotyping, Microarrays, Ultrasequencing, Proteomics, Microscopy, High-throughput Screening, FACS and Bioinformatics, as well as two Internal Service Units, the Histology Service and the newly created Protein Service that had been set up within the Systems Biology program.

The core facilities and the technologies they offer continue to be a valued support pillar of the research performed in the CRG. Following the aim to continuously evolve and invest in the latest technologies, some of new developments to be highlighted are the purchase of a super-resolution microscope, the Leica STED system in the Advanced Light Microscopy Unit, as well as the acquisition of a the first last generation DNA sequencer in Spain, the Illumina HiSeq 2000 in the Ultrasequencing Unit, and the acquisition of a high-capacity array reader, the Illumina iScan in the Genotyping Unit. Also the computing infrastructure has been upgraded significantly and the Bioinformatics Core has taken up the challenge to develop pipelines for the analysis of all the data being produced in high-throughput in the other core facilities. Especially the Genomics Units have become increasingly aligned with the Bioinformatics Core in order to offer to the users an integrated service ranging from sample preparation to in-depth data analysis.

CORE FACILITIES

Core Facility: Genomics Unit
Genotyping Unit

Unit Structure:

Head of the Unit: Mònica Bayés (until the end of February)

Manager of the Unit: Magda Montfort (since March)

Project Manager: Magda Montfort (January-February)

Technicians: Anna Puig, Carles Arribas, Sílvia Carbonell, Sebastián Morán, Cinta Pegueroles, Pilar Herruzo

SUMMARY

Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation in the human genome. A fraction of this genetic variation is likely to explain the majority of the differences between individuals, including their predisposition to develop common human disorders, such as cardiovascular disease, diabetes, asthma and cancer, and their differences in response to drugs. Other goals of SNP research include population genetics and trait selection for agricultural, cattle farming or aquaculture applications. Some of the genotyping technologies also enable to accurately characterize copy number variants (CNVs), loss of heterozygousity (LOH) and DNA methylation status.

SERVICES

The Genotyping Unit is one of the nodes of the Plataforma en Red de Genotipado Carlos III (CeGen-ISCIII, *www.cegen.org*). It provides support to PRBB users and external users from public or private institutions for genotyping projects in every aspect of research, from experiment design, DNA extraction, genotyping, methylation profiling, and data interpretation.

At the CeGen-ISCIII CRG Barcelona Node several medium and high throughput genotyping and related services were available in 2010:

1. Automated DNA extraction from blood or other tissues (Chemagen).

2. DNA quantification using Picogreen (Molecular Probes).

3. Whole Genome Amplification using GenomiPhi (Amersham).

4. Custom Genotyping by SNPlex (Applied Biosystems): genotyping of 24-48 SNPs selected by customer.

5. Custom Genotyping by VeraCode GoldenGate (Illumina) that are low plex GoldenGate genotyping arrays compatible with the BeadXpress System: genotyping of 48-384 SNPs selected by customer.

6. Custom Genotyping with GoldenGate technology (Illumina): genotyping of 96-1,536 SNPs selected by the customer.

7. Focused-content SNP Genotyping with GoldenGate technology on all catalogue Illumina products: African American admixture panel, DNA test panel, Cancer SNP panel and MHC panel set, Human Linkage V panel, Mouse LD Linkage panel and Mouse MD Linkage panel.

8. Focused-content SNP Genotyping with Infinium technology on all catalogue Illumina products: HumanOmni2.5-quad (NEW), HumanOmni1-quad, HumanOmni-Express (NEW), Human1M-quad, Human660W-quad, HumanCytoSNP-12, HumanLinkage-24 (NEW), BovineSNP50, CanineSNP20, EquineSNP50, others.

9. Custom DNA methylation profiling using Veracode GoldenGate (Illumina): 96-384 sites selected by costumer.

10. Focused-content DNA methylation profiling using Infinium technology (Illumina): HumanMethylation27 (Figure 1) and HumanMethylation450 (NEW).

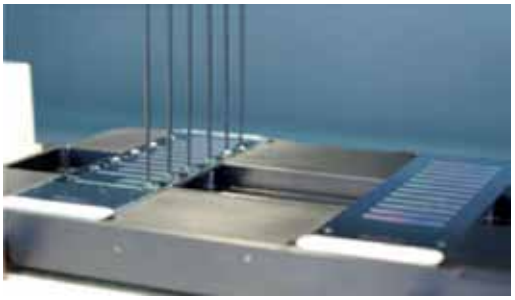


Fig 1.
Loading DNA samples on Ilumina HumanMethylation27 BeadChips using liquid handling robots.

All services are integrated with robust software tools for experimental design, management of data and analyses. Extensive quality control measures let us further refine the quality of data.

During 2010, Illumina iScan system has been acquired (Figure 2). This new scanner is being used with the recently released high density (HD) assays for genotyping and methylation technologies. Liquid handling robots had been upgraded in order to be able to process the new HD and GoldenGate BeadChips.

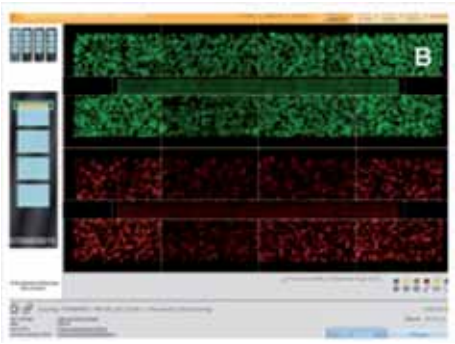


Fig 2.
A. The facility is equipped with state-of-the-art instruments including high resolution scanners. B. Section derived from a HumanOmni1-Quad Beadchip under standard scanning conditions using Illumina iScan system.

During this period, Genotyping Unit has finished 56 genotyping projects, produced more than 650K genotypes with SNPlex, near 4 million genotypes for GoldenGate custom designs and has processed around 200 Beadchips through Infinium technology. It is worth to notice that Veracode GoldenGate and GoldenGate BeadArray services were used at maximum capacity during this period. The number of processed samples and the number of work orders for each service are detailed in the following table.

Service	Samples	Work Orders
Custom Genotyping	24831	32
Infinium Genotyping	600	13
Infinium Methylation	456	11
DNA Extractions	3905	16
DNA Quantifications	22608	56
DNA WGA	2678	7

PUBLICATIONS

Alonso P, Gratacòs M, Segalàs C, Escaramís G, Real E, Bayés M, Labad J, Pertusa A, Vallejo J, Estivill X, Menchón JM.
"Variants in estrogen receptor alpha gene are associated with phenotypical expression of obsessive-compulsive disorder."
Psychoneuroendocrinology, Epub 2010 Sep 17.

Baca-Garcia E, Vaquero-Lorenzo C, Perez-Rodriguez MM, Gratacòs M, Bayés M, Santiago-Mozos R, Leiva-Murillo JM, de Prado-Cumplido M, Artes-Rodriguez A, Ceverino A, Diaz-Sastre C, Fernandez-Navarro P, Costas J, Fernandez-Piqueras J, Diaz-Hernandez M, de Leon J, Baca-Baldomero E, Saiz-Ruiz J, Mann JJ, Parsey RV, Carracedo A, Estivill X, Oquendo MA.
"Nucleotide variation in central nervous system genes among male suicide attempters."
Am J Med Genet B Neuropsychiatr Genet,153B(1):208-13 (2010).

de Cid R, Ramos MD, Aparisi L, García C, Mora J, Estivill X, Farré A, Casals T.
"Independent contribution of common CFTR variants to chronic pancreatitis."
Pancreas, 39(2):209-15 (2010).

Gratacòs M, Escaramís G, Bustamante M, Saus E, Agüera Z, Bayés M, Cellini E, Cid R, Fernández-Aranda F, Forcano L, González JR, Gorwood P, Hebebrand J, Hinney A, Mercader JM, Nacmias B, Ramoz N, Ribasés M, Ricca V, Romo L, Sorbi S, Versini A, Estivill X.
"Role of the neurotrophin network in eating disorders' subphenotypes: Body mass index and age at onset of the disease."
J Psychiatr Res, 44(13):834-40 (2010).

Muiños-Gimeno M, Montfort M, Bayés M, Estivill X, Espinosa-Parrilla Y.
"Design and evaluation of a panel of single-nucleotide polymorphisms in microRNA genomic regions for association studies in human disease."
Eur J Hum Genet, 8(2):218-26 (2010).

Soria V, Martínez-Amorós E, Escaramís G, Valero J, Crespo JM, Gutiérrez-Zotes A, Bayés M, Martorell L, Vilella E, Estivill X, Menchón JM, Gratacòs M, Urretavizcaya M.
"Resequencing and association analysis of arylalkylamine N-acetyltransferase (AANAT) gene and its contribution to major depression susceptibility."
J Pineal Res, 49(1):35-44 (2010).

Johansson S, Halmøy A, Mavroconstanti T, Jacobsen KK, Landaas ET, Reif A, Jacob C, Boreatti-Hümmer A, Kreiker S, Lesch KP, Kan CC, Kooij JJ, Kiemeney LA, Buitelaar JK, Franke B, Ribasés M, Bosch R, Bayés M, Casas M, Ramos-Quiroga JA, Cormand B, Knappskog P, Haavik J.
"Common variants in the TPH1 and TPH2 regions are not associated with persistent ADHD in a combined sample of 1,636 adult cases and 1,923 controls from four European populations."
Am J Med Genet B Neuropsychiatr Genet, 153B(5):1008-15 (2010).

Soria V, Martínez-Amorós E, Escaramís G, Valero J, Pérez-Egea R, García C, Gutiérrez-Zotes A, Puigdemont D, Bayés M, Crespo JM, Martorell L, Vilella E, Labad A, Vallejo J, Pérez V, Menchón JM, Estivill X, Gratacòs M, Urretavizcaya M.
"Differential association of circadian genes with mood disorders: CRY1 and NPAS2 are associated with unipolar major depression and CLOCK and VIP with bipolar disorder."
Neuropsychopharmacology, 35(6):1279-89 (2010).

Franke B, Vasquez AA, Johansson S, Hoogman M, Romanos J, Boreatti-Hümmer A, Heine M, Jacob CP, Lesch KP, Casas M, Ribasés M, Bosch R, Sánchez-Mora C, Gómez-Barros N, Fernàndez-Castillo N, Bayés M, Halmøy A, Halleland H, Landaas ET, Fasmer OB, Knappskog PM, Heister AJ, Kiemeney LA, Kooij JJ, Boonstra AM, Kan CC, Asherson P, Faraone SV, Buitelaar JK, Haavik J, Cormand B, Ramos-Quiroga JA, Reif A.
"Multicenter analysis of the SLC6A3/DAT1 VNTR haplotype in persistent ADHD suggests differential involvement of the gene in childhood and persistent ADHD."
Neuropsychopharmacology, 35(3):656-64 (2010).

PUBLICATIONS BY CRG GENOTYPING UNIT USERS

Arcos-Burgos M, Jain M, Acosta MT, Shively S, Stanescu H, Wallis D, Domené S, Vélez JI, Karkera JD, Balog J, Berg K, Kleta R, Gahl WA, Roessler E, Long R, Lie J, Pineda D, Londoño AC, Palacio JD, Arbelaez A, Lopera F, Elia J, Hakonarson H, Johansson S, Knappskog PM, Haavik J, Ribases M, Cormand B, Bayes M, Casas M, Ramos-Quiroga JA, Hervas A, Maher BS, Faraone SV, Seitz C, Freitag CM, Palmason H, Meyer J, Romanos M, Walitza S, Hemminger U, Warnke A, Romanos J, Renner T, Jacob C, Lesch KP, Swanson J, Vortmeyer A, Bailey-Wilson JE, Castellanos FX, Muenke M.
"A common variant of the latrophilin 3 gene, LPHN3, confers susceptibility to ADHD and predicts effectiveness of stimulant medication."
Mol Psychiatry, 15(11):1053-66 (2010).

Castro-Giner F, de Cid R, Gonzalez JR, Jarvis D, Heinrich J, Janson C, Omenaas ER, Matheson MC, Pin I, Anto JM, Wjst M, Estivill X, Kogevinas M.
"Positionally cloned genes and age specific effects in asthma and atopy: an international population-based cohort study (ECRHS)."
Thorax, 65:124-131 (2010).

Costas J, Gratacòs M, Escaramís G, Martín-Santos R, de Diego Y, Baca-García E, Canellas F, Estivill X, Guillamat R, Guitart M, Gutiérrez-Zotes A, García-Esteve L, Mayoral F, Moltó MD, Phillips C, Roca M, Carracedo A, Vilella E, Sanjuán J.
"Association study of 44 candidate genes with depressive and anxiety symptoms in post-partum women."
J Psychiatr Res, 44:717-24 (2010).

Donner J, Haapakoski R, Ezer S, Melén E, Pirkola S, Gratacòs M, Zucchelli M, Anedda F, Johansson LE, Söderhäll C, Orsmark-Pietras C, Suvisaari J, Martín-Santos R, Torrens M, Silander K, Terwilliger JD, Wickman M, Pershagen G, Lönnqvist J, Peltonen L, Estivill X, D'Amato M, Kere J, Alenius H, Hovatta I.
"Assessment of the neuropeptide S system in anxiety disorders."
Biol Psychiatry, 68(5):474-83 (2010).

Saus E, Soria V, Escaramís G, Vivarelli F, Crespo JM, Kagerbauer B, Menchón JM, Urretavizcaya M, Gratacòs M, Estivill X.
"Genetic variants and abnormal processing of pre-miR-182, a circadian clock modulator, in major depression patients with late insomnia."
Hum Mol Genet, 19(20):4017-25 (2010).

Saus E, Soria V, Escaramís G, Crespo JM, Valero J, Gutiérrez-Zotes A, Martorell L, Vilella E, Menchón JM, Estivill X, Gratacòs M and Urretavizcaya M.
"A haplotype of glycogen synthase kinase 3b is associated with early onset of unipolar major depression."
Genes Brain and Behavior, 9(7):799-807 (2010).

Sunyer J, Basagaña X, González JR, Julvez J, Guerra S, Bustamante M, de Cid R, Antó JM, Torrent M.
"Early life environment, neurodevelopment and the interrelation with atopy."
Environ Res, 110(7):733-8 (2010).

del Río-Espínola A, Fernández-Cadenas I, Rubiera M, Quintana M, Domingues-Montanari S, Mendióroz M, Fernández-Morales J, Giralt D, Molina CA, Alvarez-Sabín J, Montaner J.
"CD40 -1C>T polymorphism (rs1883832) is associated with brain vessel reocclusion after fibrinolytic therapy in ischemic stroke patients."
Pharmacogenomics, 11:763-72 (2010).

Abulí A, Bessa X, González JR, Ruiz-Ponte C, Cáceres A, Muñoz J, Gonzalo V, Balaguer F, Fernández-Rozadilla C, González D, de Castro L, Clofent J, Bujanda L, Cubiella J, Reñé JM, Morillas JD, Lanas A, Rigau J, García AM, Latorre M, Saló J, Fernández Bañares F, Argüello L, Peña E, Vilella A, Riestra S, Carreño R, Paya A, Alenda C, Xicola RM, Doyle BJ, Jover R, Llor X, Carracedo A, Castells A, Castellví-Bel S, Andreu M; Gastrointestinal Oncology Group of the Spanish Gastroenterological Association.
"Susceptibility genetic variants associated with colorectal cancer risk correlate with cancer phenotype."
Gastroenterology, 139(3):788-96, 796.e1-6 (2010).

Corominas R, Sobrido MJ, Ribasés M, Cuenca-León E, Blanco-Arias P, Narberhaus B, Roig M, Leira R, López-González J, Macaya A, Cormand B.
"Association study of the serotoninergic system in migraine in the Spanish population."
Am J Med Genet B Neuropsychiatr Genet, 153B(1):177-84 (2010).

Del Valle J, Feliubadaló L, Nadal M, Teulé A, Miró R, Cuesta R, Tornero E, Menéndez M, Darder E, Brunet J, Capellà G, Blanco I, Lázaro C.
"Identification and comprehensive characterization of large genomic rearrangements in the BRCA1 and BRCA2 genes."
Breast Cancer Res Treat, 122(3):733-43 (2010).

Domingues-Montanari S, Fernández-Cadenas I, Del Río-Espinola A, Corbeto N, Krug T, Manso H, Gouveia L, Sobral J, Mendioroz M, Fernández-Morales J, Alvarez-Sabin J, Ribó M, Rubiera M, Obach V, Martí-Fàbregas J, Freijo M, Serena J, Ferro JM, Vicente AM, Oliveira SA, Montaner J.
"Association of a genetic variant in the ALOX5AP gene with higher risk of ischemic stroke. A case-control, meta-analysis and functional study."
Cerebrovasc Dis, 29:528-537 (2010).

Domingues-Montanari S, Hernandez-Guillamon M, Fernandez-Cadenas I, Mendioroz M, Boada M, Munuera J, Rovira A, Maisterra O, Parés M, Gutierrez M, Alvarez-Sabin J, Chacón P, Delgado P, Montaner J; Stroke Project Cerebrovascular Diseases Study Group, Spanish Society of Neurology.
"ACE variants and risk of intracerebral hemorrhage recurrence in amyloid angiopathy."
Neurobiol Aging, Epub 2010 Apr 9.

Domingues-Montanari S, Fernández-Cadenas I, Del Río-Espinola A, Mendioroz M, Ribo M, Obach V, Marti-Fabregas J, Freijo M, Serena V, Corbeto N, Chacon P, Alvarez-Sabin J, Montaner J.
"The I/D polymorphism of the ACE1 is not associated with ischemic stroke in Spanish individuals."
Eur J Neurol, 17(11):1390-2 (2010).

Fernandez-Cadenas I, Mendioroz M, Domingues-Montanari S, Del Rio-Espinola A, Delgado P, Ruiz A, Hernandez-Guillamon M, Giralt D, Chacon P, Navarro-Sobrino M, Ribo M, Molina CA, Alvarez-Sabin J, Rosell A, Montaner J.
"Leukoaraiosis is associated with genes regulating blood brain barrier homeostasis in ischemic stroke patients."
Eur J Neurol, doi: 10.1111/j.1468-1331.2010.03243.x. Epub 2010 Dec 1.

Jurado S, Nogués X, Agueda L, Garcia-Giralt N, Urreizti R, Yoskovitz G, Pérez-Edo L, Saló G, Carreras R, Mellibovsky L, Balcells S, Grinberg D, Díez-Pérez A.
"Polymorphisms and haplotypes across the osteoprotegerin gene associated with bone mineral density and osteoporotic fractures."
Osteoporos Int, 21(2):287-96 (2010).

Laayouni H, Calafell F, Bertranpetit J.
"A genome-wide survey does not show the genetic distinctiveness of Basques."
Human Genetics, 127:455-458 (2010).

Moreno-Estrada A, Aparicio-Prat E, Sikora M, Engelken J, Ramírez-Soriano A, Calafell F, Bosch E.
"African signatures of recent positive selection in human FOXI1."
BMC Evol Biol, 10:267 (2010).

Pomares E, Riera M, Permanyer J, Méndez P, Castro-Navarro J, Andrés-Gutiérrez A, Marfany G, Gonzàlez-Duarte R.
"Comprehensive SNP-chip for retinitis pigmentosa-Leber congenital amaurosis diagnosis: new mutations and detection of mutational founder effects."
Eur J Hum Genet, 18(1):118-24 (2010).

Ribasés M, Ramos-Quiroga JA, Sánchez-Mora C, Bosch R, Richarte V, Palomar G, Gastaminza X, Bielsa A, Arcos-Burgos M, Muenke M, Castellanos FX, Cormand B, Bayés M, Casas M.
"Contribution of LPHN3 to the genetic susceptibility to ADHD in adulthood: a replication study."
Genes Brain Behav, doi: 10.1111/j.1601-183X.2010.00649.x. Epub 2010 Oct 7.

Sánchez-Mora C, Ribasés M, Ramos-Quiroga JA, Casas M, Bosch R, Boreatti-Hümmer A, Heine M, Jacob CP, Lesch KP, Fasmer OB, Knappskog PM, Kooij JJ, Kan C, Buitelaar JK, Mick E, Asherson P, Faraone SV, Franke B, Johansson S, Haavik J, Reif A, Bayés M, Cormand B.
"Meta-analysis of brain-derived neurotrophic factor p.Val66Met in adult ADHD in four European populations."
Am J Med Genet B Neuropsychiatr Genet, 153B(2):512-23 (2010).

CORE FACILITIES

Core Facility: Genomics Unit
Ultrasequencing Unit

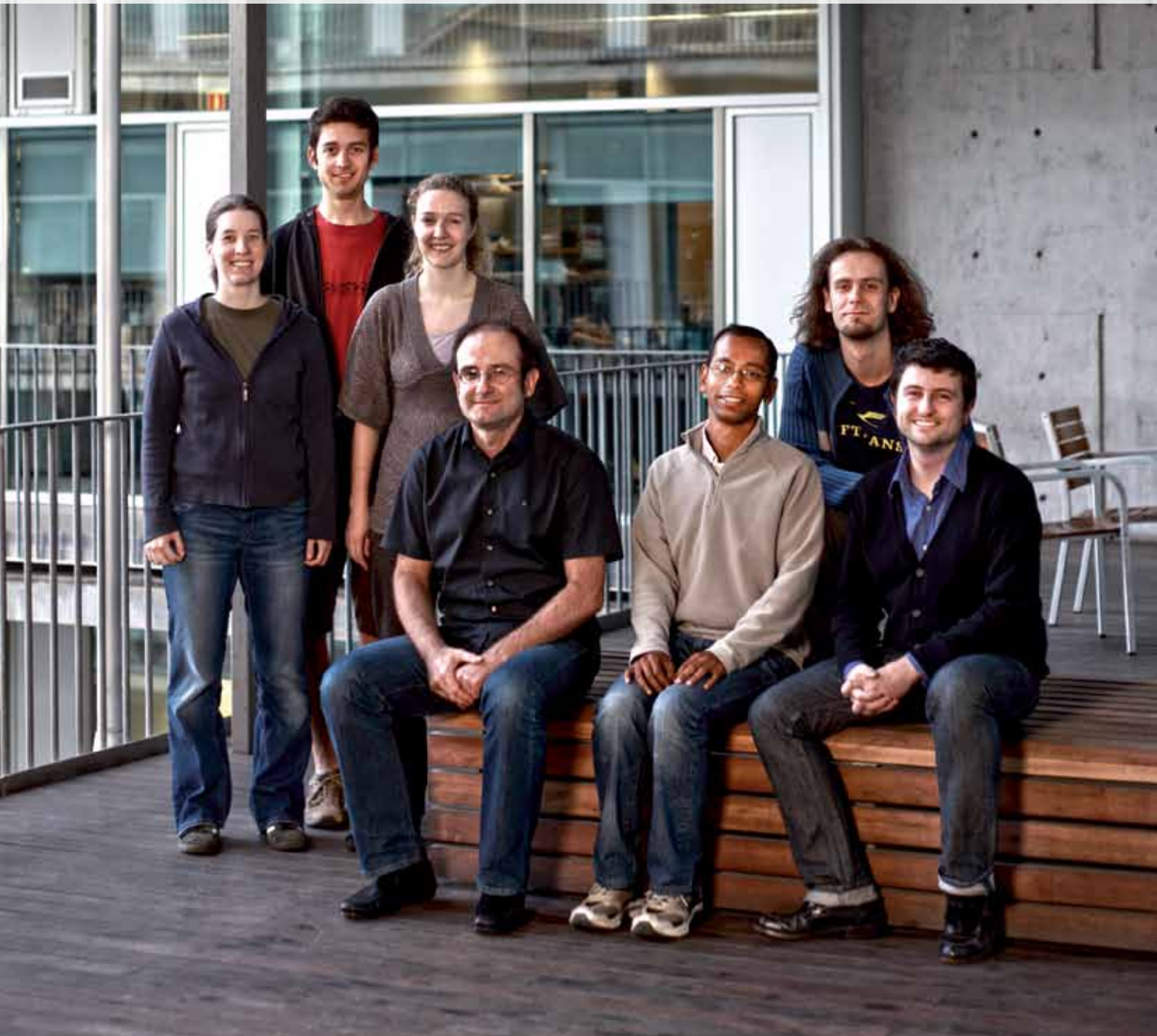
Unit Structure:

Unit Leader: Heinz Himmelbauer

Laboratory Technicians: Anna Menoyo, Ester Castillo (until April 2010), Maria Luisa Campos (until June 2010), Rebecca Curley (from February 2010), Miquel Àngel Adrover (from July 2010), Kadri Reis (from July 2010) Maik Zehnsdorf (GABI-Future), Hui Kang (GABI-Future)

Bioinformaticians: Matthew Ingham (until June 2010), Debayan Datta, Darek Kedra (GABI-Future), Juliane Dohm (GABI-Future)

PhD student: André Minoche (GABI-Future)



SUMMARY

The sequencing field has experienced a major boost by the introduction of innovative, second-generation sequencing technologies. The Ultrasequencing Unit utilizes three different high-throughput sequencing platforms. At present, the Unit is equipped with two Genome Analyzer IIx and one HiSeq2000 sequencer from Illumina, and one Roche 454 FLX sequencing instrument. A wide range of different state-of-the-art sequencing services is provided to the CRG, PRBB, and to external users.

SERVICES

Next generation sequencing has paved the way for answering a large number of different questions in biology and in biomedical sciences. The protocols currently in use in the Unit are genomic sequencing (*de novo* sequencing, re-sequencing, ChIP-Seq, long-range mate-pair sequencing), transcriptome analysis (mRNA-Seq, strand-specific mRNA Seq, small RNA sequencing), and amplicon sequencing. In addition, procedures for indexing samples are in use, allowing the pooling of samples. The Illumina and 454 platforms in the Ultrasequencing Unit complement each other: Illumina sequencing produces large numbers of short reads per run, i.e. up to 700 million reads per HiSeq flowcell with reads up to 100 bases. With the 454 technology, we generate approximately 1 million reads in a single run (read length 500 bases, 300-500 Mb sequence output per run). Thus, Solexa is very well suited for applications such as ChIP-Seq, miRNA profiling, and genome re-sequencing, while long-read 454 sequencing is very appropriate for projects such as *de novo* genomic sequencing and transcriptome characterization in non-model species lacking a reference genome.

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

RESEARCH PROJECTS

The research of the Ultrasequencing Unit focuses on the development of new procedures, both in the lab, and for data analysis, such as the development of DSSS (Direct Strand Specific Sequencing), a protocol for strand-specific sequencing of prokaryotic and eukaryotic RNA samples (Fig. 1), and the analysis of structural variation in complex genomes (Fig. 2). With funding provided by the German Federal Ministry of Education and Research (BMBF) in the context of the GABI-FUTURE program, we have sequenced the genome of the crop plant sugar beet (*Beta vulgaris*) using 454 and Illumina technology, and we are in the process of assembling, annotating and interpreting the genome draft. Working on a genome about one quarter of the size of the human genome, this project enables us to establish and test the workflows required to process and analyze sequence data from complex organisms with no reference genome sequence available.

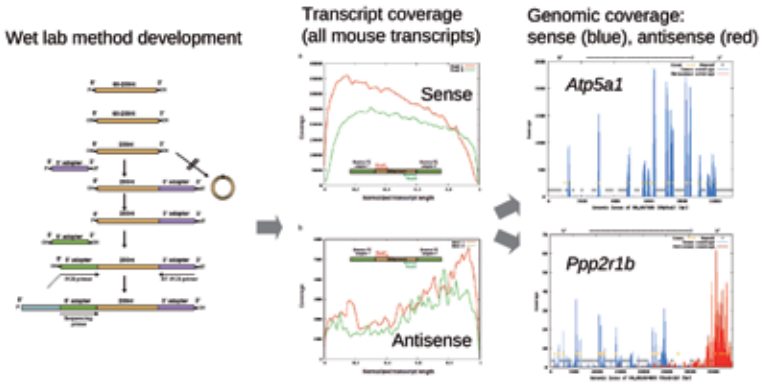


Fig 1.
Direct strand specific sequencing – a new method for generating transcript data.

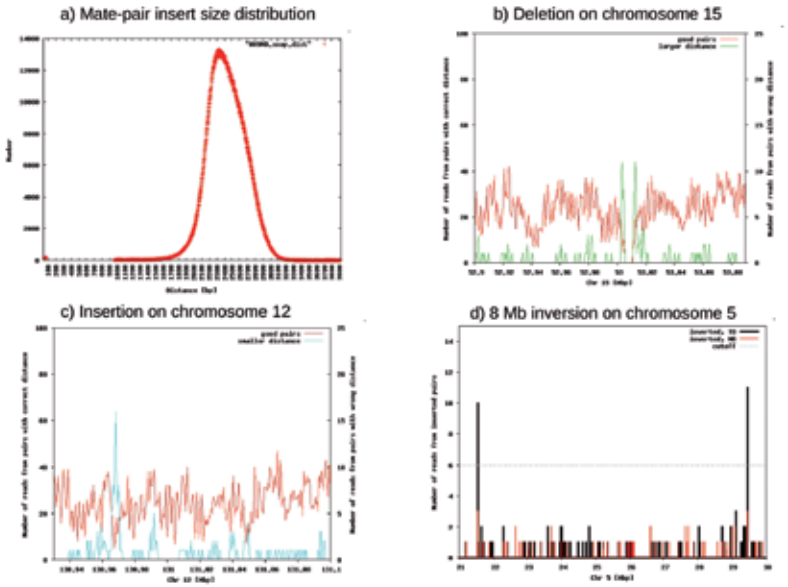


Fig 2.
Discovery of structural variation between human genomes using long-range mate-pair libraries.

PUBLICATIONS

Meunier D, Patra K, Smits R, Hägebarth A, Lüttges A, Jaussi R, Wieduwilt MJ, Quintanilla-Fend L, Himmelbauer H, Fodde R, Fundele R.
"Expression analysis of Proline Rich 15 (Prr15) in mouse and human gstrointestinal tumours."
Mol Carcinog, in press.

Vivancos AP, Güell M, Dohm JC, Serrano L, Himmelbauer H.
"Strand-specific sequencing of the transcriptome."
Genome Research, 20:989-999 (2010).

The International Cancer Genome Consortium (incl. H. Himmelbauer).
"International network of cancer genome projects."
Nature, 464:993-998 (2010).

Lange C, Mittermayr L, Dohm JC, Holtgräwe D, Weisshaar B, Himmelbauer H.
"High-throughput identification of genetic markers using representational oligonucleotide microarray analysis."
Theoretical and Applied Genetics, 121:549-565 (2010).

Doss MX, Wagh V, Schulz H, Kull M, Kolde R, Pfannkuche K, Nolden T, Himmelbauer H, Vilo J, Hescheler H, Sachinidis S.
"Global transcriptomic analysis of murine embryonic stem cell-derived brachyury (T) cells."
Genes to Cells, 15:209-228 (2010). (*)

(*) This publication results from the work of Dr. Heinz Himmelbauer at the Max Planck Institute for Molecular Genetics, Berlin, Germany.

CORE FACILITIES

Core Facility: Genomics Unit
Microarrays Unit

Unit structure:

Head of the Unit: Mònica Bayés (until the end of February)

Manager of the Unit: Anna Ferrer (since March)

Technicians: Heidi Mattlin, Maria Aguilar

Bioinformaticians: Sarah Bonnin, Manuela Hummel

SUMMARY

Microarrays allow us to analyze expression profiles (mRNA and microRNA levels) and structural variation (DNA copy number) on a genome-wide level. The Microarrays Unit is equipped with all the instruments to carry out microarray analysis with the latest version of Agilent and Illumina technologies. For both technologies, the Microarrays Unit provides a full service, including technical advice in experimental design, quality control of starting material, sample and array processing, and data analysis.

SERVICES

The laboratory provides microarray methodologies as a service at established rates to scientists from the CRG, other PRBB institutions and other external public and private institutions. Services offered include: sample quality control, RNA and DNA sample labelling, hybridization of microarrays and data processing and analysis, as well as array design and fabrication by spotting.

The Microarrays Unit offers a range of options suitable for a variety of applications including:

> mRNA expression profiling on Agilent microarrays using the 2-color and 1-color protocols, and Illumina BeadArrays (1-color protocol).
> miRNA expression profiling on Agilent microarrays.
> array based comparative genomic hybridization (aCGH) on spotted BAC and Agilent microarrays.
> chromatin immunopreciptation on array (ChIP-on-chip) on spotted promoter, CpG island and Agilent microarrays.
> DNA capture using the SureSelect Target Enrichment system (new service).

Some space reorganizations had taken place during 2010: the Ultrasequencing and Microarrays Units are sharing the laboratory for better interaction between technicians, and the bioinformaticians from the Microarrays Unit had moved to the Bioinformatics Unit space. This reorganization instigates collaborations between the three units.

We have carried out an evaluation of the microarray technologies the unit has (Agilent 1-color, 2-colors, and Illumina), as an annual internal control for the unit. We used the commercial RNA reference sample (*Universal Human Reference RNA*), our internal control that we include in all microarray projects (figure 1). The results show again, as in the previous year, low technical variability in the intra-platform measurements, as well as high inter-platform concordance.

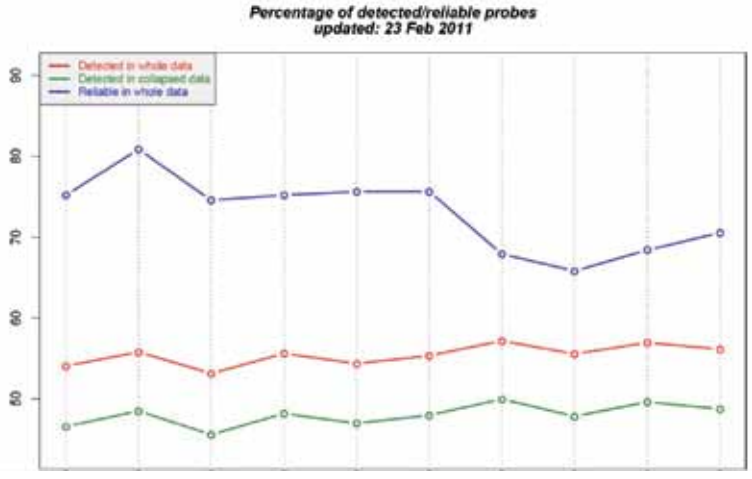


Figure 1.
Bioinformatic tools for monitoring microarrays quality based on the use of a commercial reference RNA sample (UHRR, Stratagene). The percentage of detected (intensity above a certain threshold) microarray probes (red) and genes (blue) and percentage of reliable (intensity sufficiently larger than background) probes (green) from the microarray platform test done at 2009 (green rectangle) and 2010 (red rectangle)

We have also carried out a test for reducing the RNA input for the mRNA profiling projects using the Agilent platform. We tested the use of 25ng and 100ng of total RNA as an input for the microarray gene expression experiments and compared the results to the data obtained with the standard protocol of 500ng of total RNA as a starting material. The percentage of detected genes and signal intensities correlated very well. Now, 100ng of total RNA is the standard amount that we use at the Microarrays Unit for mRNA expression profiling experiments using the Agilent platform.

We have done a microRNA profiling project using FFPE samples obtaining data of similar quality as when using non degraded RNA samples.

We have set up a new service: DNA-capture for sequencing, using the SureSelect Target Enrichment system from Agilent. It allows capturing the genomic regions of your interest prior to high-throughput sequencing, reducing the cost of the experiment while ensuring high gain from the sequencing. For that purpose, the Microarrays Unit has worked closely together with the Ultrasequencing and Bioinformatics Units. The collaboration between the three units has resulted in an R software package for quality control in Target Enrichment experiments (figure 2).

A. **B.**

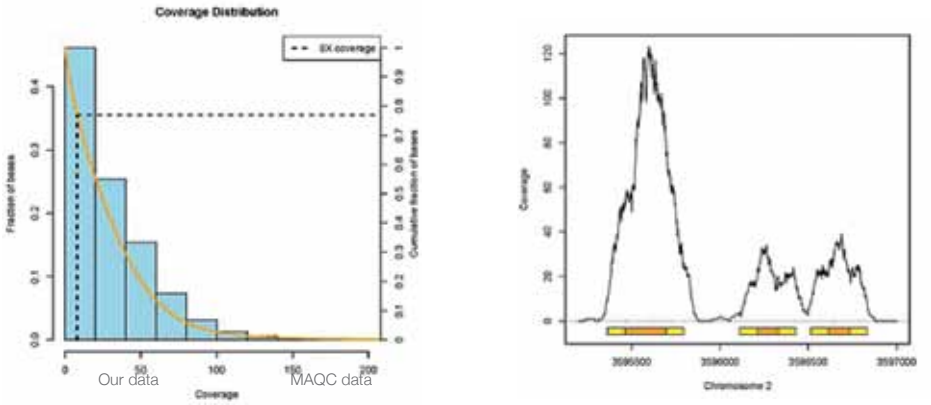


Figure 2.
Diagnostic plots provided within the Bioconductor package for Target Capture Quality Control.

a) On-target coverage histogram. The thick orange line shows the cumulative fraction of target bases (right y-axis) with a read coverage of at least x. Dashed lines highlight the fraction of target bases with ≥8X coverage.

b) Per-base coverage along chromosomal positions. Target regions (plus addition of 100 bases on both sides) are highlighted in orange (yellow).

We have performed a study to compare two different softwares that the unit is using for data extraction from Agilent arrays: GenePix (Axon) and Feature Extraction (Agilent). The results show that both methods are very similar, but that Feature Extraction is able to detect very high signal in probes for which GenePix reaches saturation. The use of the Feature Extraction software is now in the normal protocol for all Agilent gene expression experiments.

We have performed a comparative study between different normalization methods for microRNA profiling. For that, our own data and published data have been compared. The results show that our data is very similar to the published ones and that the normalization method used in the unit as the standard approach was good, but it could be slightly improved. This improvement is used as the standard method now.

During 2010 the unit has finished 53 projects. From those, 35 consisted in mRNA expression profiling, the most common application at the Microarrays Unit. In total, we have processed 628 samples for mRNA expression profiling, 160 samples for microRNA profiling, 64 for aCGH and 18 samples using the SureSelect Target Enrichment system for DNA-capture. In addition, we have fabricated 166 arrays using the Versarray spotter. The unit also provides access and support to real-time qPCR instruments and offers a new service for DNA shearing using the Covaris system.

PUBLICATIONS

Ortega FJ, Moreno-Navarrete JM, Pardo G, Sabater M, Hummel M, Ferrer A, Rodriguez-Hermosa JI, Ruiz B, Ricart W, Peral B, Fernández-Real JM.
"MiRNA expression profile of human subcutaneous adipose and during adipocyte differentiation."
PLoS One, 5(2):e9022 (2010).

PUBLICATIONS BY CRG MICROARRAY UNIT USERS

Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, Guigo R, Shiekhattar R.
"Long Noncoding RNAs with Enhancer-like Function in Human Cells."
Cell, 143(1):46-58 (2010).

Guidi M, Muinos-Gimeno M, Kagerbauer B, Marti E, Estivill X, Espinosa-Parrilla Y.
"Overexpression of miR-128 specifically inhibits the truncated isoform of NTRK3 and upregulates BCL2 in SH-SY5Y neuroblastoma cells."
BMC Mol Biol, 11(1):95 (2010).

Novoa I, Gallego J, Ferreira PG, Mendez R.
"Mitotic cell-cycle progression is regulated by CPEB1 and CPEB4-dependent translational control."
Nat Cell Biol, 12(5):447-56 (2010).

Richly H, Rocha-Viegas L, Ribeiro JD, Demajo S, Gundem G, Lopez-Bigas N, Nakagawa T, Rospert S, Ito T, Di Croce L.
"Transcriptional activation of polycomb-repressed genes by ZRF1."
Nature, 468(7327):1124-8 (2010).

PUBLICATIONS BY EXTERNAL MICROARRAY UNIT USERS

González E, Fernández MR, Marco D, Calam E, Sumoy L, Parés X, Dequin S, Biosca JA.
"Role of Saccharomyces cerevisiae oxidoreductases Bdh1p and Ara1p in the metabolism of acetoin and 2,3-butanediol."
Appl Environ Microbiol, 76(3):670-9 (2010).

Salgado R, Toll A, Alameda F, Baró T, Martín-Ezquerra G, Sanmartín O, Martorell-Calatayud A, Salido M, Almenar S, Solé F, Pujol RM, Espinet B.
"CKS1B amplification is a frequent event in cutaneous squamous cell carcinoma with aggressive clinical behaviour."
Genes Chromosomes Cancer, 49(11):1054-61 (2010).

CORE FACILITIES

Core Facility: CRG/UPF Proteomics Unit

Unit Structure:

Unit Leader: Henrik Molina (CRG)

Scientists: Cristina Chiva (UPF), Eva Borras (UPF), Carolina de la Torre (ProteoRed), Guadalupe Espadas-Garcia (CRG), Francesco Mancuso (CRG)

SUMMARY

Proteome & Proteomics: The word proteome was first conceived by Marc Wilkins in 1984. Proteomics is a contraction of *proteome* and *-omics* and translate into "the study proteins". Generally speaking, the genome is a relative static entity that is in sharp contrast to the very dynamic proteome. The proteome changes over time, between different cellular compartments, or in response to stimulation. Even when focusing a study to only one organism, the term proteome might not make much sense. Because of the ever changing conditions, the plural form of proteome, proteome**s**, is much more appropriate.

Proteomics is the tool used to decipher proteomes and proteomics includes the analysis of:

> How genes are combined to give rise to alternative gene products (splicing) or isoforms
> Where specific proteins are expressed,
> When proteins are expressed,
> How proteins interact (protein-protein interactions),
> How proteins are being modified (post-translational modifications), and
> What are the dynamics of proteins with respect to stimuli or other cellular states?

Mass spectrometry has proven a valuable technology for studying proteomics and it is safe to say that mass spectrometry has been the driving force of proteomics. Today, close to half of all proteomics studies rely on mass spectrometry, which is also the case for the CRG/UPF Joint Proteomics Unit .

More than ninety percent of the unit's operating is based on two modern Orbitrap mass spectrometers equipped with nano-HPLC. Older QTOF and MALDI mass spectrometers are also in unit but are used in-frequent. The MALDI mass spectrometer is mainly dedicated as a walk-up instrument. The unit is currently considering purchasing an additional mass spectrometer that will be dedicated to SRM/MRM type experiments. SRM/MRM is a technique that complements discovery type analysis which is currently the main type of analysis conducted by the Unit.

The Proteomics Unit is working close with many researchers at CRG as well as at PRBB. Examples include proteomics analysis of *'bugs'* (*Mycoplasma* bacteria from the Serrano lab, CRG) and redox controlled pathways (Hidalgo lab, UPF) to tissue samples from mice brains (Ozaita lab, UPF) and global phosphopeptide analysis (Beato lab, CRG).

Typical services provide by the Unit are:

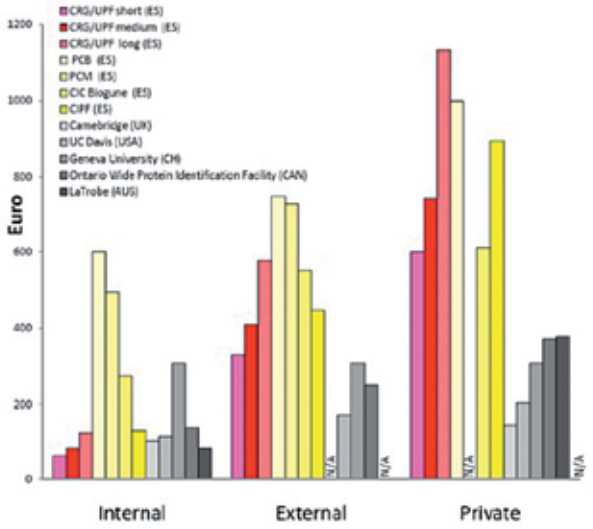
> Identifications of proteins by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS)
> Determination of molecular mass of molecules.
> Phosphorylation enrichments and analysis by LC-MS/MS
> Relative quantitation of proteins based on stable isotopes

FINANCE

In last half of 2010 the Unit has worked on preparing a new cost model for services (to be implemented in 2011). The Unit has priced all process to be offered and when addressing a specific proteomics question, it is now possible to combine the different blocks to offer the best solution and at the same time be able to easy calculate a price for *tailored* proteomics experiments.

Though difficult to compare between proteomics facilities (depends e.g. on instrumentation and charging models) the Proteomics Unit's has conducted a price comparison between similar facilities in Spain and on three continents. Our new internal price for the analysis of a medium complex sample will be lower than the compared Spanish facilities but on par with compared facilities in North America, Australia and Europe. When it comes to external prices, the compared Spanish facilities, including the Proteomics Facility at CRG, will be/is more expensive than the non-Spanish facilities (Figure 1). This is an interesting observation and leads to the question: If a unit has spare capacity, is it better to run idle or try to create income my analysing external samples at discounted external prices? Such an approach could help to cover basic costs, exemplified by warranties for instruments, thereby lowering internal prices.

Price analysis (3 continents /6 countries) for LC-MS/MS based identification of proteins in a sample of medium complexity



Price information's are obtained from the websites of 10 proteomics facilities (CRG: Red-ish, Spanish facilities: yellow-ish and non-Spanish facilities: grey-ish). The chosen non-Spanish facilities are chosen as being the top hits when searching Google ". Where no details were available, it is assumed that the price is for a sample of medium complexity. If not specified we assumed that the price includes digestion with trypsin and simple data analysis.

AFFILIATIONS

The Joint UPF/CRG Proteomics Core Facility is a member of ProteoRed, the Spanish Proteomics Network.

PUBLICATIONS

Bunkenborg J, Garcia GE, Paz MI, Andersen JS, Molina H.
"The Minotaur proteome: avoiding cross-species identifications deriving from bovine serum in cell culture models."
Proteomics, 16:3040-4 (2010).

PUBLICATIONS BY CRG PROTEOMICS UNIT USERS

Sardon T, Pache RA, Stein A, Molina H, Vernos I, Aloy P.
"Uncovering new substrates for Aurora A kinase."
EMBO Rep, 11:977-84 (2010).

Papadopoulos C, Arato K, Lilienthal E, Zerweck J, Schutkowski M, Chatain N, Müller-Newen G, Becker W, de la Luna S.
"Splice Variants of the Dual Specificity Tyrosine Phosphorylation-regulated Kinase 4 (DYRK4) Differ in Their Subcellular Localization and Catalytic Activity."
J Biol Chem, Epub 2010 Dec 2.

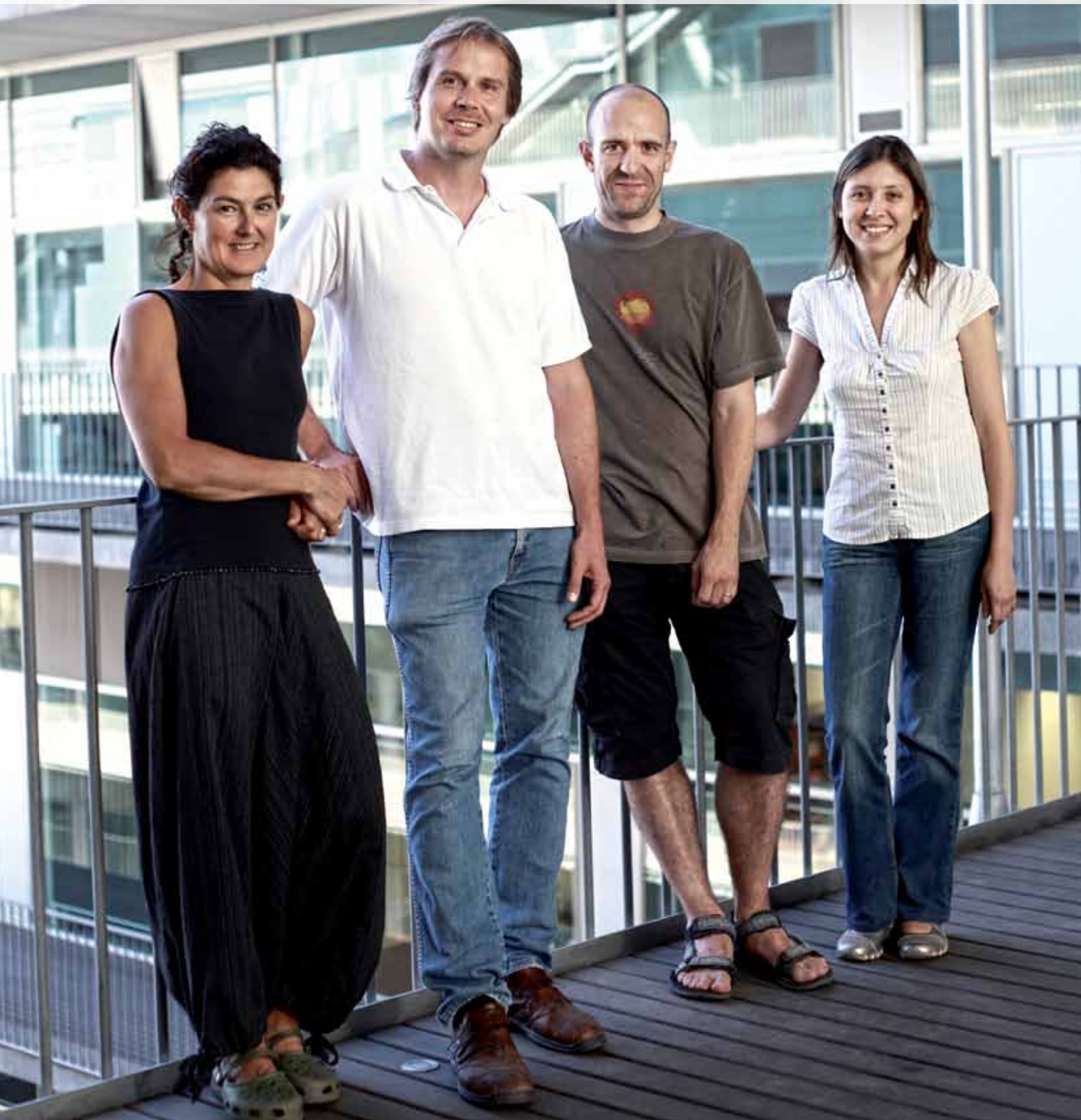
CORE FACILITIES

Core Facility: CRG/UPF Advanced Light Microscopy Unit

Unit structure:

Unit Leader: Timo Zimmermann

Technicians: 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 covers the spectrum of advanced imaging applications from thick tissue reconstruction to fast *in vivo* imaging to the sensitive detection of very faint signals. 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, 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.

In this year the ALMU added a dedicated super-resolution microscope that improves optical resolution to approx. 70nm to its equipment list. This constitutes an improvement in resolution by a factor of two to three. The unit is used regularly by researchers from all CRG programs 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 2010 to provide instrumentation at the forefront of imaging technology through the installation of a super-resolution microscope that was introduced to the field just months before. The new continuous wave stimulated emission depletion (CW-STED) confocal microscope from Leica uses a strong orange depletion laser (596 nm) for imaging of green fluorescent dyes at lateral resolutions around 70 nm. This works with fluorophores like Alexa488 and Oregon Green, fluorescent proteins like YFP and also with specific two-color labels for multichannel super-resolution imaging. The system was co-financed by the Ministry for Science and Innovation (MICINN) from funds available in the “Fondo Europeo de Desarrollo Regional” (FEDER) program of the European Union.

The unit also installed the confocal High Content Screening software HCS A from Leica, which allows to perform highly automated screening runs. The program was acquired for SysteMTB, a FP7 consortium of research groups working on Tuberculosis, which is coordinated by Luis Serrano. It will be used for the microscopic intracellular localization of all mycobacterial proteins.

Among the seven available microscope 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 at least two 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.

In the ongoing effort to streamline core facility infrastructures inside PRBB, CRG and the UPF Department of Experimental and Health Sciences have agreed to unite the confocal microscopy facility of the UPF with the ALMU. Accordingly, the UPF confocal microscope was moved into a location inside CRG and the UPF microscopy specialist, Dr. Xavier Sanjuan, moved into the ALMU office.

The total microscope usage time of the joint unit in 2010 reached 17750 hours (16525 CRG + 1225 UPF), corresponding to more than eight hours of daily usage per microscope system. The usage has grown by 27% compared to the year before, reflecting the continuing high need for light microscopy by CRG and PRBB researchers. During the year, 123 users from 24 of the CRG research groups have used the unit. On average, 53 users from 20 CRG groups worked on the microscopes every month. Additionally the unit was used by 74 users from 15 groups from UPF and IMIM.

During the year, the ALMU staff has participated in teaching masters courses of the Universitat Pompeu Fabra (UPF), as well as in microscopy courses in other Barcelona universities. They have continued to participate as speakers and instructors in courses, workshops and seminars at institutes in Spain and in other European countries.

In October the unit hosted the Leica International Advanced Confocal Microscopy Course, a four-day course for 12 selected participants from several countries. It focused on measuring Fluorescence Resonance Energy Transfer using confocal and lifetime methods and was supplemented by three open lectures of leading researchers in that area.

Together with Doris Meder (head of Core Facilities) and Raul Gomez (Screening/Robotics Unit), the head of the unit, Timo Zimmermann, co-organised in May a technology seminar on High Content Screening Methods that brought together four high level speakers from that field.

As a follow-up on this subject, the ALMU co-organised and hosted a joint workshop of the companies Leica (Spain) and Cytoo (France) that showcased the new confocal screening software HCS A by Leica and the technology of micropatterned glass coverslips that generate reproducible single cell shapes by Cytoo. Both technologies are now available in the facility.

Together with Doris Meder, Timo Zimmermann, co-organised in December a technology symposium on Super-Resolution Light Microscopy that brought together three international high level speakers from that field from Spain, France and Germany.

The ALMU is participating in EuroBioImaging, an initiative of the biomedical imaging field that has been recently accepted for the preparatory phase of the ESFRI roadmap. Timo Zimmermann is currently the national contact person for Spain for biological imaging in this intiative. He organized and hosted a meeting of Spanish light microscopy units in June at the CRG that was attended by 40 microscopy specialists from 29 institutes all over Spain. In this meeting, EuroBioImaging was presented by the visiting project manager Antje Keppler and the formation of a Spanish Light Microscopy Network was discussed.

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.

Fig 1.
Usage statistics 2008-2010 of the Advanced Light Microscopy Unit

CORE FACILITIES

Core Facility: High-Throughput Screening Unit

Unit structure:
Unit Responsible: Raúl Gomez Riera

Technicians: Anja Leimpek

SUMMARY

The High-Throughput Screening (HTS) Unit provides robotic equipment for the automation of complex assays which require processing of multiple samples and high-content screening services for a number of research projects. The facility is equipped with state of the art instrumentation, which allows measurement of all typical screening assay signals such as fluorescence and luminescence. The experienced personnel of the HTS Unit provides also the necessary know-how for efficient assay optimization and execution, training of the researchers on the instruments and assistance with subsequent data analysis.

FACILITY OVERVIEW

In 2010 the Unit has successfully performed two high-throughput RNAi screenings. One of the screens concerned the validation of hits which had previously been identified in the first genome-wide RNAi screen carried out in 2009 as part of a project of Vivek Malhotra's laboratory. The goal of the experiment was to identify genes implicated in mucin synthesis and secretion. The respective assay was based on immunodetection of mucin coupled to a sensitive luminescence readout.

The second genome-wide RNAi screen was performed in collaboration with the laboratory of Juan Valcarcel. The aim of this experiment was the identification of genes influencing alternative splicing using the Fas receptor as model system. The screening consisted of three major steps that were automated: 1) transfection of the siRNA library (ca. 21,000 genes); 2) preparation of cell lysates and RNA extraction using the mRNA Catcher™ PLUS system (Invitrogen), and 3) dispensing reagents for reverse transcription and conventional barcoded PCR amplification for subsequent sequencing quantification using the Illumina/Solexa technology.

Table 1.
Shows the workflow and timelines of the two screenings.

Number	Task	2010							
		January	February	March	April	May	June	July	August
1	**Splicing Project**								
1.1	Programming and settings								
1.2	Assay validation (controls)								
1.3	Pilot screen (library subset)								
1.4	Whole genome siRNA screen								
2	**Mucin Project**								
2.1	Setup automation protocol								
2.2	Confirmation screen								
2.3	Data analysis								

In addition to the two genome-wide screens, an automated ELISA was developed to determine the titer of 200 antibodies developed within the CRG antibody project. The same sera were also characterized by immunofluorescence staining which was performed using one of the liquid handler platforms available in the Unit. A fluorescence-based cell viability assay has been set up as well for this program that can be offered to other projects.

In order to further enhance the capabilities of the HTS platform, a Tecan Freedom EVO liquid handling system was acquired which is a fully automated robotic screening station. This system will be used together with the Protein Services Unit for cloning, expression screening, small-scale protein purification and also for diverse range of applications including primary and secondary screening and library management.

TECAN Freedom EVO liquid handling platform

Finally, in 2010 the vacant Head of Unit position was filled with the recruitment of Renza Roncarati who will assume responsibility in January 2011.

CORE FACILITIES

Core Facility: Bioinformatics Unit

Unit structure:

Unit Leader: Guglielmo Roma

Bioinformaticians: Andreu Alibés*, Antonio Hermoso, Ernesto Lowy*, Francesco Mancuso*, Luca Cozzuto*

SUMMARY

The Bioinformatics core provides expertise in bioinformatics, statistics, data analysis, and scientific software development to support CRG, PRBB, and external research groups. The team periodically organizes courses to train biologists on the use of bioinformatics resources.

FACILITY OVERVIEW AND SERVICES

Established in September 2009, the Bioinformatics Unit was finally up-and-running at the beginning of 2010 after the recruitment of additional members and the set-up of first key equipment - such as a disk server with 360 TB, a cluster with around 500 cores and several servers dedicated to different tasks.

During this year, the team has contributed to several research projects by providing support and expertise in programming and advanced data analysis, focusing primarily on high-throughput genomics (including microarrays, genotyping, and next-generation sequencing) and proteomics technologies. All methodologies developed and implemented in collaboration with the colleagues from other core facilities and research groups are therefore offered as data analysis services to new users.

Due to high request of microarray and methylation arrays analyses, the team developed different pipelines using various packages from R and Bioconductor that support Illumina, Affymetrix, and Agilent platforms. The analysis workflow consists of data preprocessing, normalization, and statistical inference. Then, identified gene lists can be further investigated by applying various downstream methods to interpret the biological significance, including gene set enrichment, over-representation of biological functions using ontology terms, and pathways analysis. Pattern discovery tools (*de novo* and not) are used to look for sequence motifs shared in upstream regions of co-regulated genes.

Regarding next-generation sequencing, the team has already implemented several methods for the analysis of a wide range of applications including: ChIP-seq using chromatin immunoprecipitation to identify binding sites of DNA-associated proteins, such as transcription factors, histone marks, polymerase, among others; RNA-Seq transcriptome analysis to detect novel and rare transcripts (both coding and non-coding), alternative splicing variants, RNA editing, and expression levels; DNA re-sequencing (whole genome and target-enrichment/exome sequencing) to study genetic variation, such as SNPs, copy number changes, and chromosomal rearrangements; Hi-C using genome-wide chromosome conformation capture to detect long-range interactions between chromatin segments; de-novo genome and transcriptome sequencing and relative functional gene annotation analysis.

Additionally, the unit has started a close collaboration with the Proteomics Unit to investigate different extraction tools, not only in respect to a side-by-side comparison but also to understand optimal settings for extraction of tandem MS data.

As further important commitment, the team has been responsible for the development of scientific databases and web interfaces. After the evaluation of several commercially available laboratory management tools, the unit has been working 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. Another important achievement has been the development of a website comprising all the data generated by Prof. Luis Serrano's laboratory during their systems biology studies on *Mycoplasma Pneumoniae*. The huge effort in the integration of genomics, transcriptomics, proteomics, and metabolomics data will be used as basis for the development of future databases for the study of other bacterial genome projects.

Finally, the team has dedicated initial efforts to the organization of tutorials to train biologists on the use of common bioinformatics resources as well as on basic programming skills.

** joined during 2010*

PUBLICATIONS

Sanseverino W*, Roma G*, De Simone M, Faino L, Melito S, Stupka E, Frusciante L, Ercolano MR.
"PRGdb: a bioinformatics platform for plant resistance gene analysis."
Nucleic Acids Res, 38(Database issue): D814-21 (2010). [* joint first authors]

Romito A, Lonardo E, Roma G, Minchiotti G, Ballabio A and Cobellis G.
"Lack of Sik1 in Mouse Embryonic Stem Cells Impairs Cardiomyogenesis by Down-Regulating the Cyclin-Dependent Kinase Inhibitor p57kip2."
Plos One, 5(2):e9029 (2010).

CORE FACILITIES

Core Facility: CRG/UPF Flow Cytometry (FACS) Unit

Unit structure:

Unit Leader: Òscar Fornas (UPF)

Operator: Neus Romo (CRG)

SUMMARY

Flow Cytometry is a technology that measures multiple cell parameters based on fluorescence emission of each cell in suspension at rapid rate (more than 50.000 events/s). This multi-parametric analysis (up to 18 biological parameters simultaneously) gives a high accurate cell phenotype and the possibility to sort up to four different subsets of cells at the same time.

The facility is operated by UPF since 2001 and the CRG/UPF Joint Flow Cytometry Unit starts on September 2009, officially from March 1st 2010.

The provided services and instrumentation to PRBB and external users, covers most of Flow Cytometry applications from simple to multicolour flow analysis (See figure 1A for maximum available colours/fluorescence).

The personnel assist users for experimental, project planning and advisement as well as the required training for self-service instrumentation usage. Data analysis support is also provided to users.

The CRG/UPF joint facility becomes a wide facility in terms of equipment (figure 1B), consisting on 5 analyzers and 2 cell sorters as follows:

> Cytometers (analyzers):

FACScan (UPF)	(1 laser and 3 fluorescences)
FACScalibur (UPF)	(2 laser and 3+2 fluorescences)
LSR 1 (UPF)	(3 laser and 3+2+2 fluorescences)
FACScanto (CRG)	(2 laser and 4+2 fluorescences)
LSR2 (CRG)	(4 laser and 4+2+2+2 fluorescences)

> Cytometers (cell sorters):

FACSvantage (UPF)	(3 laser and 3+2+2 fluorescences)
FACSAria (CRG)	(5 laser and 5+5+3+3+2 fluorescences)

A

B

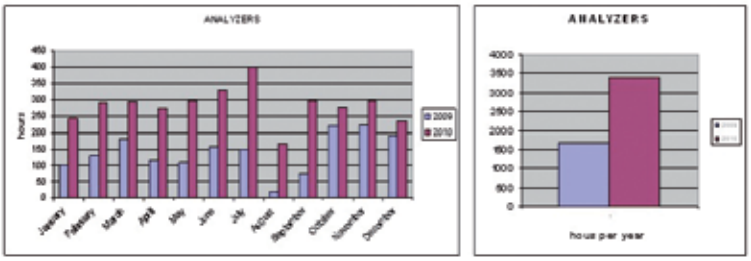
Instrument	Laser Ex line (nm)	Parameters FSC/SSC+FL(n)	System	Application
FACScan	488	2 + 3	analog	analyzer
FACScalibur	488, 633	2 + 4	analog	analyzer
LSR	488, 633, 350	2 + 6	analog	analyzer
FACScanto	488, 633	2 + 6 (9)	digital	analyzer
LSR II	488, 633, 355, 405	2 + 10 (16)	digital	analyzer
FACSDiVa	488, 633, 325	2 + 6/8	digital	cell sorter
FACSAria II SORP	488, 640, 355, 405, 561	2 + 18 (25)	digital	cell sorter

Fig. 1.
Instrument features. (A) General optical configuration summarizing available fluorescence wavelength detection and related parameters/fluorescence by laser type. (B) Principal features of available instrumentation, laser type and fluorescence detector capacity per instrument.

Current equipment offers the possibility to fit most flow applications on any instrument but since Flow Cytometry is continuously in development, new fluorochromes and applications are daily appearing, users demand falls on most advanced instruments. Since the old cell sorter can only fits simple assays with classic flourochromes the incorporation of a last generation cell sorter was acquired on September 2009 increasing technical possibilities, covering most of cell sorting applications.

The Flow Cytometry demand has increased a lot (from 2009 to 2010) duplicating the analyzers usage hours from 1700 to 3500 (figure 2A). Cell sorting was more than three times from 300 hours (2009) to 1100 (2010) (figure 2B).

A



B

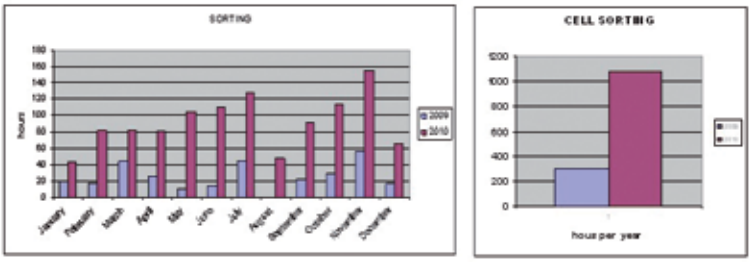


Fig. 2.
Facility statistics. *General statistics comparison between 2009 and 2010. Amount of hours per month and year of analyzers (A) and cell sorting (B).*

During the year the facility staff has participated in teaching masters courses of the Universitat Pompeu Fabra (UPF) as well as in the Spanish BD FACS Users Forum as speaker in basic and clinical Flow Cytometry.

CORE FACILITIES

In-House Services: Histology Service

Service structure:

Service Manager: Salvador Aznar Benitah PhD, ICREA Researcher, PI of the Epithelial and Homeostasis Laboratory, Dept. Differentiation and Cancer (CRG)

Staff Scientists: Alexis Rafols Mitjans, Senior Technician, FP (CRG), Marina Nuñez Alfonso, FP (CRG)

Associates: Juan Martin Caballero PhD, Veterinarian, Head and of the Animal Facility of the PRBB
Juana M. Flores, Professor of Pathologic Anatomy and Director of the Department of Medicine and Surgery, Faculty of Veterinary Medicine, Universidad Complutense de Madrid

SUMMARY

The Histology Unit was established in 2007 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.
> Immunohistochemical staining of paraffin embedded and frozen samples (starting mid 2010).
> 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.

In 2010, we implemented a new service to perform full Postmortem Histopathological analysis of mouse models. This service is provided externally, but in collaboration with the unit, by Dr. Juana M. Flores (Professor of Pathologic Anatomy and Director of the Department of Medicine and Surgery, Faculty of Veterinary Medicine, Universidad Complutense de Madrid) in association with the CRG.

Due to the exponential increase in the demand of services that the unit had during 2008-2010, we succesfully applied for, and obtained, funding to the Ministry of Health (MICINN) to hire a second technician for the unit in 2009, Marina Nuñez Alfonso. In addition, in 2010 we obtained a second technician position financed by the MICINN, to stabilize the contract of Alexis Rafols Mitjans.

In 2010 we also started to implement and set up immunohistochemical staining of OCT embedded frozen sections, paraffin sections and fresh tissue. We expected that in 2010 this service to be fully operational for all users. Although there has been little demand for this service we intend on maintaining this service should any researcher at the CRG require it.

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.

FUTURE SERVICES

We had discussed during 2010 with all the users of the unit the feasibility and possible demand for providing ***in situ* hybridization** services (detection of RNAs on tissue sections). Although there is an internal demand, establishing this service would require a new separate RNAse free laboratory, for which at the moment the CRG has not the capability to provide the space. A second service discussed with the users is ***Laser capture microdissection***. Since this service is already provided with the IMIM, we have discussed with the IMIM Administration (Montserrat Torá) the possibility that their unit provides such service to CRG users.

CORE FACILITIES

In-House Services: Protein Expression Service

Service structure:
Service Manager: Michela Bertero

SUMMARY

The CRG Protein Service team produces and purifies proteins (mostly from *E. coli*) for CRG researchers. The Service performs different types of experiments, from the preparation of the initial expression vectors to screening of diverse expression conditions to protein purification on milligram scale using diverse chromatographic methods. A new multi-functional robotic platform (TECAN EVO200) was implemented in 2010 to automate a wide range of molecular biology applications in high-throughput format, including PCR preparation, cloning methods, bacterial transformation protocols, protein expression screening and small scale purification experiments. Finally, several instruments are available to study biophysical properties of proteins and nucleic acids such as stability, folding, kinetic parameters, and molecular interactions.

Upon request, the CRG Protein Service also offers practical **training** on how to use chromatography systems.

MAJOR PROJECTS AND ACHIEVEMENTS

In 2010, the CRG Protein Service performed more than 30 projects in collaboration with ten CRG laboratories. Most of the experiments aim at determining optimal expression conditions and purifying target proteins for antibody production, functional assays and/or biophysical characterization.

One of these projects is in the frame of the **FP7 European project PROSPECTS** (Proteomics in Time and Space), in which the Serrano lab is partner. The overall aim is to quantitatively characterize the epidermal growth factor receptor (EGFR) signaling pathway, the most important pathway that regulates growth, survival, proliferation, and differentiation in mammalian cells. Most types of tumors have been found associated with mutations in the EGFR system. Currently, 177 proteins have been shown to be involved in the EGF signal transduction pathway. The Protein Service screened the expression of all the 177 proteins in two different bacterial strains and with different expression conditions. 126 proteins were successfully purified and used by the Serrano team to determine absolute protein concentrations in different cell lines.

The Protein Service and the Bioinformatics Core facility have developed and implemented the Protein Wiki. The **Protein Wiki** is online tool that allows management of user requests, experiments, results and data. The tool also provides an easy interface for the user to follow on the experiments and receive an electronic output. The Protein Wiki contains main protocols on protein expression and purification to be shared with the users.

CRG TRANSVERSAL INSTITUTIONAL PROJECT

In efforts to understand the mechanisms of C/EBPa induced reprogramming of pre-B cells into macrophages several ongoing collaborations exist between the Graf group and other groups at the CRG. The first one is based on the murine system developed earlier (Bussmann et al, Cell Stem Cell, 2009) and concerns the question about whether transdifferentiation involves the re-positioning of nucleosomes genome-wide. For this purpose Laura Gaveglia in Miguel Beato's group has digested chromatin of uninduced and induced cells with micrococcus nuclease and isolated mononucleosomes whose DNA was analyzed by deep sequencing. Preliminary data, as analyzed by Giancarlo Castellano, showed major differences. Current experiments are aimed at reproducing these results and to obtain data of higher resolution, using DNAs from different time points.

The second collaboration is with Heinz Himmelbauer and Cedric Notredame and is based on the observation that reprogrammed macrophages can revert into pre-B cells at low frequencies, exhibiting chromosome translocations. This raises the possibility that C/EBPa induces chromosome aberrations during myeloid differentiation. To study this, the DNA of induced cells is being analyzed by paired end sequencing and compared to noninduced cell DNA.

A third project uses the transcriptome data obtained from mature human B cells (a Burkitt's lymphoma line) transdifferentiated by C/EBPa into macrophages, a system recently developed by Francesca Rapino of the Graf lab. Kiana Toufighi in Luis Serranos's group is performing a systems biology analysis of the changes in gene expression during reprogramming, comparing them with changes in gene expression of human epithelial stem cells differentiating into keratinocytes. The aim is to identify genes that are deregulated during differentiation/transdifferentiation in different cell systems.

A fourth collaboration, also using the human transdifferentiation system, involves Rory Johnson, a staff scientist in Roderic Guigo's group. Rory has analyzed the RNAs from the transdifferentiating cells to identify long noncoding RNAS that change in expression, using a customized array made by the group. A number of such RNAs have been identified and experiments are ongoing to determine whether their knockdown or overexpression influences the switching behaviour of the transdifferentiating cells, and perhaps cause changes in the expression of nearby genes.

HIGHLIGHTS

HIGHLIGHTS

GENE REGULATION PROGRAMME

New functions for new genes: non-coding RNAs get into the action

Thanks to the new technological advances in genome sequencing, we know that a major portion of the genome is being transcribed but only a small portion of this transcriptome contains the protein-coding sequences. So, even the long non-coding RNAs constitute a large portion of the mammalian transcriptome, their cellular functions remain elusive.

Scientists have already described the role of some small RNAs in gene regulation but they still know a few about the long ones. Long non-coding RNAs are transcripts longer than 100 nucleotides. For a number of these non-coding RNAs, the authors show that their presence increase the synthesis of nearby proteins involved in crucial processes in development and disease.

The work, published in *Cell*, has been led by Ramin Shiekhattar, group leader at the Centre for Genomic Regulation (CRG) in Barcelona and now at the Wistar Institute and co-authored by Cedric Notredame and Roderic Guigo from the Bioinformatics and Genomics Programme at the Centre for Genomic Regulation (CRG).

Shiekhattar and collaborators have seen a new role for a class of long non-coding RNAs in positive regulation of protein-coding genes. "This seriously challenges our understanding of genes and how the human genetic network is regulated" says Dr. Shiekhattar and adds, "we are happy to contribute another important piece to the puzzle".

"Positive regulation of proteins mediated through RNA is a completely new finding" exclaims Ulf Anderson Ørom, first author of the paper. DNA regions have been known for many years to confer both negative and positive regulation of protein synthesis. "The fact that RNA is much easier to manipulate than DNA could mean that this discovery has a vast therapeutic potential" explains Ørom.

Research, using a combination of computational analysis and laboratory experiments, lead to these intriguing findings. "10 years after the Human Genome Project this work illustrates how new databases like ENCODE and Havana, established with high throughput sequencing techniques, can be used to address key biological questions. It is also a showcase of how these new technologies bring together wet lab and *in silico* biologists." states Cedric Notredame, group leader at the CRG Comparative Bioinformatics group.

Reference:

Ørom UA et al.
"Long Noncoding RNAs with Enhancer-like Function in Human Cells".
Cell, 143(1):46-58 (2010).

HIGHLIGHTS

DIFFERENTIATION AND CANCER PROGRAMME

"To be or not to be, that is the question": The role of the ZRF1 gene in embryonic development and carcinogenesis

"To be or not to be, that is the question"… During embryonic development, cells undergoing multiplication have to decide what type of cells they will become. For example, they have to decide whether they will be neuronal, muscular, or epidermal cells. At this point, it is as important for the cell to understand what it should become as what it should not become and therefore a tight control of gene expression is required. In order to coordinate and direct the destined of the stem cells, certain genes are responsible for activating and deactivating other genes that define their specialty during development and that also may be involved in cellular renewal for tissue and organ maintenance.

As described in the work published in *Nature*, ZRF1 is one of the genes that coordinate this regulation in embryonic development. The work has been done by the "Epigenetic events in cancer" group, lead by Luciano Di Croce at the Center of Genomic Regulation, in collaboration with the Pompeu Fabra University, the Nagasaki University School of Medicine and the Freiburg University.

"We know that ZRF1 is present at high levels in cells during their development as well as in altered cancer cells. Now for the first time we understand that its presence is essential for controlling cellular destiny and memory", commented Di Croce. He also explained, "When ZRF1 is active, the proteins that block transcription of the genes that determine the destiny of the cells are removed. This allows the cells to begin on their path towards specialization".

If the gears that control the cellular destiny no longer function, this unleashes problems in the cell since it is converted to a uncontrolled cell. "Now that we know why ZRF1 is important, we would like to identify the specific genes that are being controlled to see in which part of the genome they are located" explained Holger Richly, the first author of the study. This would permit the researchers to identify further genes that may be involved in the differentiation process. "Understanding in detail the gene that we are studying would open up good perspectives to better understanding the processes of tumor formation and carcinogenesis in the long run. ZRF1 is a highly conserved gene that is found in a large variety of organisms, from yeast to mammals, which makes it a perfect candidate for investigating these processes" added Di Croce.

Reference:
Richly H et al.
"Transcriptional activation of polycomb-repressed genes by ZRF1",
Nature, 468(7327):1124-8 (2010).

HIGHLIGHTS

GENES AND DISEASE PROGRAMME

A new mechanism involved in the regulation of stem cells in the adult brain

Collaboration between a group from the University of Valencia –led by Isabel Fariñas– and one from the Centre for Genomic Regulation (CRG) in Barcelona –headed by Mariona Arbonés- has led to the identification of a new biological mechanism involved in the maintenance of the stems cells found in the adult brain and which are responsible for the continual generation of new neurones. The study was published in the prestigious *Cell Stem Cell* journal and in it is described how the cellular quantity of a kinase protein, known as Dyrk1A, determines the correct replication of neural stem cells.

Stem cells present in adult tissues, such as the brain, skin or bone marrow, contribute to the cell renewal of these tissues throughout an individual's lifetime. One of the basic questions in stem cell research is how these cells divide to produce to two daughter cells, one of which retains the plastic characteristics of the stem cell while the other differentiates. In the study, it was shown how the kinase Dyrk1A present in a neural stem cell is distributed evenly or unevenly between the daughter cells and how the cell which inherits the Dyrk1A protein retains the same properties as the parent cell, while the cell receiving less Dyrk1A changes its response to exogenous signals, loses its pluripotency (the capacity to generate the distinct cell types of the brain) and begins to differentiate.

The work also revealed how a reduction in the amount of gene *DYRK1A* compromises the maintenance of the stem cell populations in the adult brain. The encoding gene for protein Dyrk1A in humans is found on chromosome 21, in the area known as the "Down Syndrome Critical Region" and is considered to be one of the candidate genes for explaining some of the neurological alterations associated with this syndrome. Moreover, individuals which have only one functional copy of gene *DYRK1A* exhibit microcephaly, evidence of the importance of this molecule in brain development. Although the current research is being undertaken on adult mice and does not afford information to explain the alterations produced during the neural development of those affected by the aneuploidy of chromosome 21, it does indicate that the amount of molecule Dyrk1A is fundamental for the correct development of many neurological aspects.

Reference:
Ferrón SR et al.
"Regulated segregation of kinase Dyrk1A during asymmetric neural stem cell division is critical for EGFR-mediated biased signalling",
Cell Stem Cell, 7(3):367-379 (2010).

HIGHLIGHTS

BIOINFORMATICS AND GENOMICS PROGRAMME

Big bang in the protein universe

Almost 100 years ago Edwin Hubble observed that distant galaxies are moving away from Earth faster than those that are closer. This relationship between distance and velocity is widely cited as evidence of the origin of the Universe from a Big Bang. Researchers at the Centre for Genomic Regulation copied his approach to investigate the divergence between protein sequences.

"We wanted to know if the divergent evolution between proteins was still proceeding. Today, we can find proteins that are still similar after almost 3,5 billion years of evolution. Our study showed that their divergence continues with these proteins becoming more and more different despite their incredible level of conservation", said Fyodor Kondrashov, principal investigator of the project and leader of the Evolutionary Genomics group at the CRG.

The work done by Fyodor Kondrashov and Inna Povolotskaya goes beyond similarity studies and discusses the evolution of proteins from the view of evolutionary dynamics, offering a new perspective on how protein structures are maintained in evolution. "In the same way that Hubble's observations led to an understanding of the past and the future of our universe, using his approach at a molecular level we get a similar overview that gives us the ability to analyze evolutionary dynamics and get a broad prediction of the possible changes to the proteins in the future", says Inna Povolotskaya, first author of the work and responsible for obtaining and analyzing all data.

Proteins are formed through combinations of amino acids, with only 20 types of amino acids are available to form a particular protein. To obtain the data for their study, the CRG researchers have compared proteins sequences from different species that were available in GenBank, a public database of genetic information. Comparing these sequences the authors measured the distance of proteins from each other and devised a method for measuring how fast the proteins are accumulating different changes. Thus, they could replicate Hubble's approach by correlating the distance between the proteins with the rate of their divergence. The result indicates that even the most distantly-related proteins are still accumulating differences.

The study shows how new techniques of bioinformatics and computational analysis can also expand knowledge at a molecular level. "Our work is a good example of how we can learn new and very fundamental things just by analyzing a larger volume of data that can be obtained by one experimental laboratory", says Kondrashov.

Most changes in a protein are deleterious because they somehow disrupt its structure or function. The authors observation that even very conservative proteins are still diverging challenges this view, because it implies that most amino acids in a protein can be changes without any ill effects. Their explanation is that amino acid changes that are deleterious in one combination can be benign when occurring in a different one. "Thanks to our study we now have a better understanding of protein structure dynamics" declares Kondrashov. It may provide a new perspective to groups working on protein structure to find new targets for design drugs, etc.

Povolotskaya & Kondrashov study also provides new information on how different interactions between different amino acids in the structure of proteins slows down but does not completely prevent evolution.

Reference:
Inna S. Povolotskaya & Fyodor A. Kondrashov.
"Sequence space and the ongoing expansion of the protein universe"
Nature, 465(7300):922-926 (2010).

A new secretory pathway for proteins

Proteins that contain a signal sequence are targeted to the endoplasmic reticulum (ER). In the ER, the signal sequence is cleaved and, if and when permitted, the secretory proteins are exported to the Golgi apparatus. Within the Golgi, the cargo is sorted and thence transported to various cellular destinations including the extracellular space. The eukaryotic cells, however, utilize another, unconventional, mode of protein secretion. Proteins following this route lack a classical signal sequence for entering the ER and their secretion is independent of entry into the Golgi membranes. How is this achieved?

In this a paper, which is published in the *Journal of Cell Biology*, the CRG researchers reveal the stops in the unconventional secretion of a protein called Acb1 in yeast. This paper reports the identification of genes involved in this pathway of unconventional secretion. Based on their data, they propose that Acb1 is secreted by a specific class of autophagosomes. Autophagosomes are compartments that capture cytosolic material for degradation in the lysosomes. However, unlike all other autophagosomes that fuse with the Vacuole thus degrading the contents, Acb1 containing autophagosomes are delivered to the cell surface. In other words, the Acb1 containing autophagosomes evade fusion with the vacuole. These autophagosomes fuse with an early endosomes and mature into a multivesicular body (MVB). MVB's then fuse with the cell surface and release Acb1 in the extracellular space.

Acb1 is required for the synthesis of steroids in the mammalian cells and the secreted form has a role in controlling a number of neurological conditions such panic disorder, depression, addiction, and to regulate high glucose dependent insulin secretion. Understanding the mechanism of Acb1/DBI secretion and its mechanism of action is therefore of fundamental interest for normal physiology and a variety of human pathologies.

Reference:

Duran JM et al.
"Unconventional secretion of Acb1 is mediated by autophagosome".
Journal Cell Biology, 188(4):527-536 (2010).

HIGHLIGHTS

SYSTEMS BIOLOGY PROGRAMME

New sucess for computational biology: new discoveries about the limb development

The study of extremity development in vertebrates has often been used as a model system for spatial control in cell formation. At the beginning of their development, the extremities of vertebrates are basically protuberances of undifferentiated cells with no determined function. As the development advances, between several hours and some days depending on the species, these cells mature until they become specific types of cell, for example, bone, muscle, nerve or blood cells. At the same time, the cells divide and migrate collectively, changing the size and shape of this protuberance until the extremity is fully developed. Part of this transformation is the elongation of the extremity along the proximodistal axis, which was the object of study of the researchers from the Centre for Genomic Regulation (CRG).

Classical experiments try to explain extremity development based on the observation of the number of cell divisions and cellular density throughout the extremity during its development. The most accepted hypotheses support the idea of the existence of a cellular proliferation gradient over which the distal cells (those furthest from the body) divide faster than the others, provoking the elongation of the extremity. In other words, the theory proposes that the cells must know where they are (further or closer from the body) but they do not have to know which direction they are pointing.

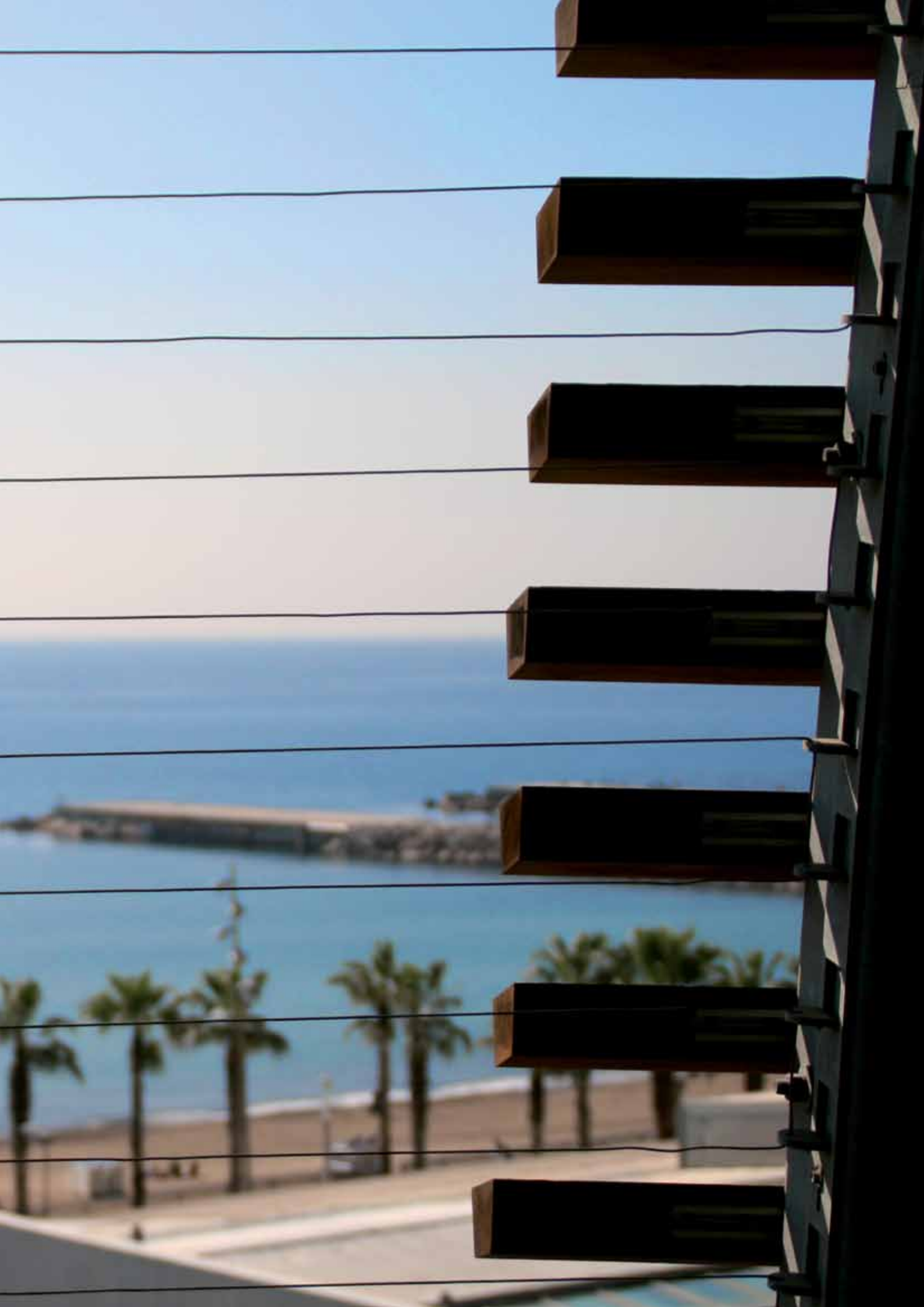
Researchers from the Systems Analysis of Development laboratory of the Centre for Genomic Regulation have studied the development process of the extremities from new angles, which have allowed the observation of important factors in the development of the limb. Specifically, the group lead by James Sharpe has, for the first time, obtained quantitative and comprehensive data about cellular proliferation and density. Instead of a traditional description, the data set is composed of numeric values collected for the whole organ, enabling the construction of a three-dimensional computer model of the development of the extremity. The work, published in the journal *PLOS Biology*, presents the first three-dimensional replica of the growth of an extremity, which allows a realistic simulation of its development.

Thanks to this model, the researchers have observed the existence of new factors to take into account to explain the formation of the extremities. Although previous work suggested that the individual behaviour of the cells did not have to be directional and explained the elongation of the extremities by a cellular proliferation gradient (the distance of each cell to the body determining the speed of division), the work of Sharpe and his collaborators reveals that non-directional factors are insufficient to explain the resulting elongation of the limb. “We have observed that in order to develop properly, the cells must point in the carefully controlled directions” explains Bernd Boehm, the first author of the work.

“Our project has revealed new important factors for extremity development which until now have not been taken into account” affirms the ICREA research professor and chief of the Systems Analysis of Development group of the CRG, James Sharpe. “Experimental data are no longer sufficient and computer models and simulations are here to stay”, he adds. “New microscopic technology and optical projection together with computational analysis of experimental data allow us to create useful models to study processes in much more depth” concludes Sharpe.

Biological research is becoming more and more multidisciplinary and this study is a good example. The work of James Sharpe and his team proposes a model to explain a process, which takes place during the development of the extremities and which is common to the development of other organs, so the data obtained can be extrapolated to these too. To enable this, within the Systems Analysis of Development lab biologists, bioinformaticians, physicists and even engineers work together.

Reference:
Boehm B at al.
“The Role of Spatially Controlled Cell Proliferation in Limb Bud Morphogenesis”
PLOS Biology, 8(7):e1000420 (2010).



Appendix 1

IX CRG ANNUAL SYMPOSIUM

"Medical genome sequencing: understanding the genomes of disease"

The CRG symposia have always been highly successful at bringing together the best scientists in their fields. The ninth edition of the symposium, held on 28th and 29th October 2010, aimed to establish a forum for capturing the latest advances in genomic sequencing technologies and their various medical applications. The two-day meeting "Medical Genome Sequencing: Understanding the Genomes of Disease" featured talks by international keynote speakers outlining the current status of medical sequencing, and setting out the technological challenges ahead.

Revolutionary developments in genome sequencing technology will soon enable us to sequence an entire human genome at an affordable price and within a short time frame. At present, high-throughput sequencing technology is increasingly being applied in basic biomedical science and is being validated for diagnostic sequencing of a small number of established disease genes. When applied more widely, this approach has the potential to elucidate the full genetic background of a given disease, which will be of major importance for the accurate classification of diseases and for the design of therapeutic strategies, bringing personalised genomic medicine a step closer.

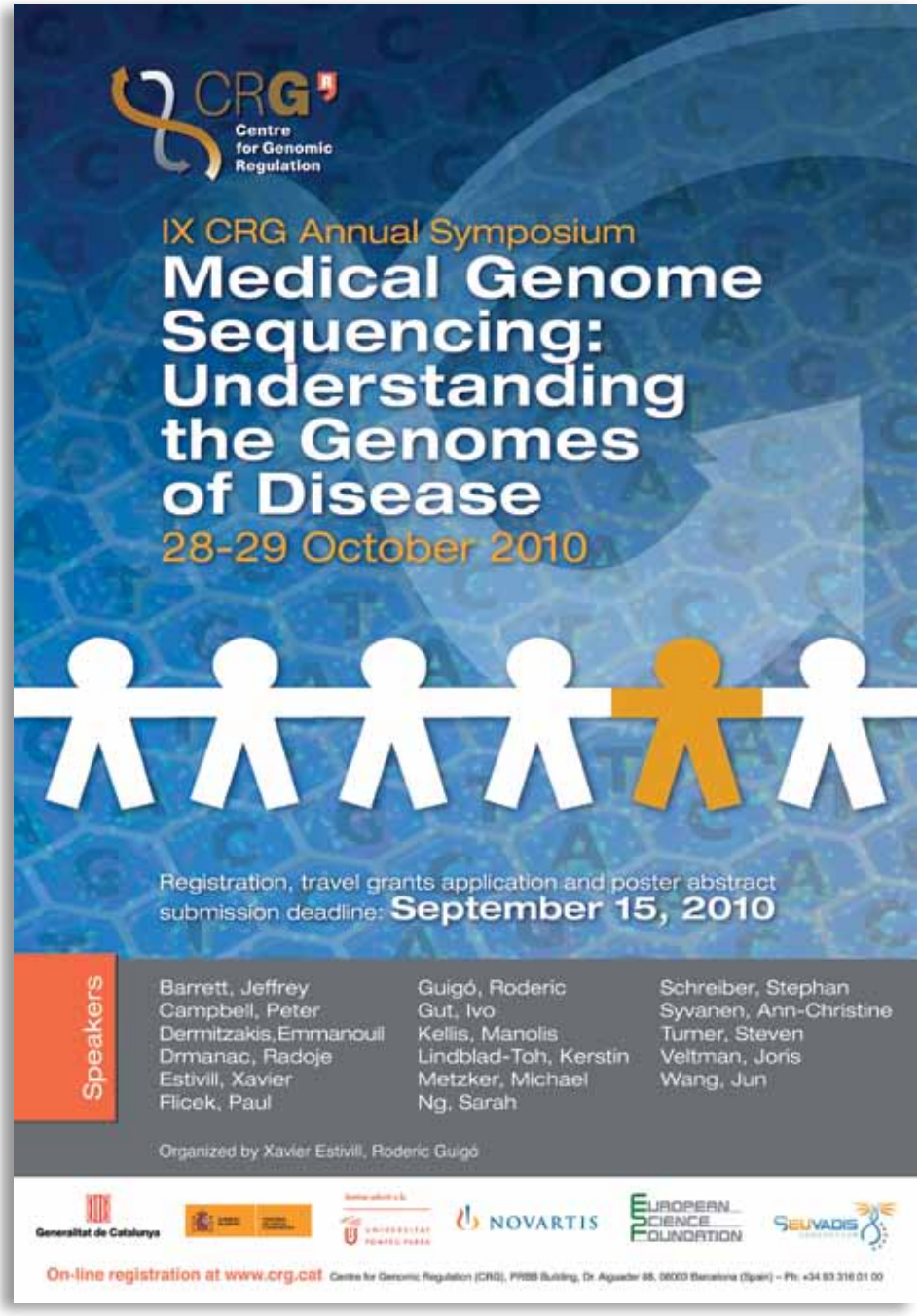
This symposium aimed to provide the basis for future developments in genomics medicine, with a strong focusing on the medical applications of high-throughput genome sequencing. This forum brought together some of the world leaders in the fields concerned and was particularly attractive for both the Spanish and international scientific communities working in these areas. The symposium was open to everyone in the community and included 18 lectures by high-profile speakers, the presentation of 6 papers and a poster session, and was divided into four blocks.

The first day of the symposium focused on the presentation of some of the most novel projects being carried out in the field of sequencing and genome medicine, and on demonstrating their potential for medical application. The second day of the symposium was dedicated to the latest technology in sequencing and its potential application. The symposium closed with a visionary session about the future of personalised medicine. The intention of the organisers of this forum was to find solutions and create synergies at European and international levels which span the full range of actions that could be taken by each participant independently.

All of the attendees agreed that the speakers selected were the world leaders in their respective fields and the best for the different sessions into which the symposium was divided. The topics presented and the discussions afterwards were of great interest for the scientists from the CRG and the entire scientific community in the area of Barcelona in general, as well as for the attendees from the rest of Spain and Europe. It is also very important to highlight that contacts were established between the groups of several speakers, which may, in the future, develop into collaborative projects.

The high number of attendees (more than 250), the level of the guest speakers and the discussions contributed to the internationalisation of the scientific image of Barcelona, Catalonia and Spain.

We believe the final result of this symposium was truly interesting for the participants due to the relevance and prestige of the experts in this area. The content and format of the symposium worked as a forum, clearly suited to attaining these objectives and, for this reason, we consider it to have been a great success.


CRG
Centre for Genomic Regulation
IX CRG Annual Symposium
Medical Genome Sequencing: Understanding the Genomes of Disease
28-29 October 2010
Registration, travel grants application and poster abstract submission deadline: September 15, 2010
Speakers
Barrett, Jeffrey
Campbell, Peter
Dermitzakis, Emmanouil
Drmanac, Radoje
Estivill, Xavier
Flicek, Paul
Guigó, Roderic
Gut, Ivo
Kellis, Manolis
Lindblad-Toh, Kerstin
Metzker, Michael
Ng, Sarah
Schreiber, Stephan
Syvanen, Ann-Christine
Turner, Steven
Veltman, Joris
Wang, Jun
Organized by Xavier Estivill, Roderic Guigó
Generalitat de Catalunya
NOVARTIS
EUROPEAN SCIENCE FOUNDATION
On-line registration at www.crg.cat

Appendix 2

PRBB-CRG SESSIONS 2010

19-11-10 Lluís Armengol
Quantitative Genomic Medicine Laboratories SL, Barcelona, Spain
"Face your next challenge!!! ... market (and sell) your science"

12-11-10 William Schafer
MRC-LBM, Cambridge, United Kingdom
"Pain circuits in worms"

22-10-10 Gerald Rubin
Janelia Farm Research Campus, Ashburn, Virginia, USA
"Genetic Tools for Studying the Anatomy and Function of the Drosophila Nervous System"

08-10-10 Ueli Schibler
Dept. of Molecular Biology, University of Geneva, Switzerland
"The daily rhythms of genes, cells, and organs"

01-10-10 James Hudspeth
Howard Hughes Medical Institute and Laboratory of Sensory Neuroscience, The Rockefeller University, New York, USA
"Making an effort to listen: mechanical amplification by myosin molecules and ion channels in hair cells of the inner ear"

17-09-10 David Searls
University of Pennsylvania, USA
"Molecules, Languages, and Automata"

23-07-10 Tariq Enver
"MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford, United Kingdom"
"Transcriptional programming of normal and leukaemic stem cells"

09-07-10 Bjoern Brembs
FB Biologie, Chemie, Pharmazie, Institut für Biologie - NeurobiologieFreie, Universität Berlin, Germany
"Evolved, not engineered: a systems neurobiology"

02-07-10 Stas Shvartsman
Lewis-Sigler Institute for Integrative Genomics, Carl Icahn Laboratory, Princeton University, USA
"MAPK signaling in equations and embryos"

21-06-10 Thomas Gregor
Laboratory for the Physics of Life, Princeton University, USA
"The onset of collective behavior in eukaryotic cell populations"

18-06-10 Alberto Sánchez Díaz
Paterson Institute for Cancer Research, Manchester, United Kingdom
"Studying the Regulation of Cytokinesis"

11-06-10 Ana Cuervo
Dept. of Developmental and Molecular Biology, Marion Bessin Liver Research Center, Albert Einstein College of Medicine, New York, USA
"Selective autophagy in aging and age-related disorders"

04-06-10 Angela Nieto
Dept. of Developmental Neurobiology, Instituto de Neurociencias de Alicante, Spain
"Ancestral and derived functions of the Snail/Scratch superfamily in health and disease"

21-05-10 Diego di Bernardo
TIGEM, Napoli, Italy
"Synthetic Biology and microfluidics to reverse-engineer, model and control gene regulatory networks"

10-05-10 Gil Ast
Laboratory for the Research of the Genome, Sackler Medical School, Tel Aviv University, Tel Aviv, Israel
"When nucleosome meets mRNA splicing"

07-05-10 Wolfgang Zachariae
Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany
"Masters of Reduction: Control of Chromosome Segregation in Meiosis"

30-04-10 Stefan Hüttelmaier
Core Facility Imaging (CFI), Dept. for Molecular Cell Biology, Martin-Luther-University Halle, Germany
"To be 'caged' or not - Control of mRNA fate by IGF2BPs"

09-04-10 Stefan Thor
Dept. of Clinical & Experimental Medicine, Linkoping University, Sweden
"From neural progenitor to unique neuron: multi-step integration of combinatorial information"

26-03-10 Hans Meinhardt
Max Planck Institute for Developmental Biology, Tuebingen, Germany
"Never reaching a steady state: models for highly dynamic intracellular pattern formation"

12-03-10 Giulio Pavesi
Universitá degli Studi Milano, Dept. Biomolecular Sciences and Biotechnology, Milano, Italy
"In Silico Prediction of Transcription Factor Binding Sites: from Chips to ChIPs"

05-03-10 Massimo Pigluicci
Dept. of Philosophy, City University of New York-Lehman College, USA
"Toward an Extended Evolutionary Synthesis"

19-02-10 Araxi Urrutia
Dept. of Biology & Biochemistry, University of Bath, United Kingdom
"Eukaryotic genome evolution: tales from transcriptomes"

12-02-10 Hyung Don Ryoo
Department of Cell Biology, New York University School of Medicine, New York, USA
"Unfolding pathways of stress response and apoptosis in Drosophila"

29-01-10 Yves Barral
Institute of Biochemistry, Departament of Biology, ETH Zurich, Switzerland
"Role of diffusion barriers in asymmetric cell division"

PROGRAMME & CORE FACILITIES SEMINARS 2010

SYSTEMS BIOLOGY PROGRAMME

28-10-10 Filipa Alves
Instituto Gulbenkian, Lisbon, Portugal
"Mathematical models for the generation and variation of spatial patterns: examples from cell biophysics and embryo gene regulation"

28-10-10 Jaap Kaandorp
University of Amsterdam, The Netherlands
"Modelling morphogenesis of sponges and corals: from genes to organism"

28-10-10 Marten Postma
University of Amsterdam, The Netherlands
"Robustness, sensitivity and dynamical behaviour of a Drosophila gap gene pattern formation model"

15-09-10 Kunihiko Kaneko
Komaba and ERATO Complex Systems Biology, JST, University of Tokyo, Japan
"(1)Evolution of Robustness (2)Pluripotency from itinerant gene expression dynamics"

13-09-10 Shigeru Kondo
Pattern Formation Group, Osaka University, Faculty of Frontier Bioscience, Osaka, Japan
"Animal pigment pattern as a example of the non-linear wave (Turing pattern)"

07-09-10 Kristian Vlahovicek
Bioinformatics group, University of Zagreb, Hungary
"Functional analysis of metagenomes"

02-07-10 Alexander Schmidt
Proteomics Core Facility, Biozentrum, University of Basel, Switzerland
"Global protein profiling of microbes using directed mass spectrometry: Application to the human pathogen L. interrogans"

01-07-10 Rotem Sorek
Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel
"Discovery of antimicrobials using large-scale genomics"

25-06-10 Patrick Cramer
Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität München, Germany
"Towards molecular systems biology of gene transcription and regulation"

16-06-10 Alison Abbott
Nature's Publishing, London, United Kingdom
"Nature and science journalism"

11-06-10 Julio Banga
Instituto de Investigaciones Marinas, CSIC, Vigo, Spain
"Optimization and optimality in computational systems biology"

04-06-10 Michael Smeaton
The Johns Hopkins Bloomberg School of Public Health, Baltimore, USA
"Initiation of Interstrand Cross-link Repair in Mammals"

03-06-10 Cornelia Fritsch
University of Sussex, Brighton, United Kingdom
"Functional evolution of the recently duplicated Bone Morphogenic Proteins Glass bottom boat and Screw"

02-06-10 Dorothy Dankel/Ana Delgado/Roger Strand
University of Bergen, Norway
"Understanding, creating and improving life"

25-05-10 José I. de las Heras
University of Edinburgh, United Kingdom
"A functional antagonism between the pgc germline repressor and torso in the development of somatic cells"

17-05-10 Julián Mensch
Universidad de Buenos Aires, Argentina
"Comparative genomics of evo-devo: signatures of selection and constraint along the embryogenesis of Drosophila"

07-05-10 Victor Neduva
WTCCB, CSBE, School of Biological Sciences, University of Edinburgh, United Kingdom
"Architecture of the budding yeast kinome"

12-03-10 David Irons
Department of Probability & Statistics, School of Mathematics, University of Sheffield, United Kingdom
"Robustness of Positional Specification by the Hedgehog Morphogen Gradient"

05-03-10 Jörg Stülke
Institut für Mikrobiologie und Genetik, University of Göttingen, Germany
"Organisation of metabolism in Bacillus subtilis: Evidence for the presence of protein complexes in central metabolism"

01-02-10 Clifton E. Barry 3rd
Tuberculosis Research Section, NIAID, National Institutes of Health, Bethesda, USA
"Probing the intracellular metabolism of Mycobacterium tuberculosis using qHTS: a systematic approach to target validation"

15-01-10 Mikko Juusola
Department of Biomedical Science, University of Sheffield, United Kingdom
"Bottom-up and top-town processing of information in the Drosophila visual system"

BIOINFORMATICS & GENOMICS PROGRAMME

16-12-10 Javier Herrero
European Bioinformatics Institute (EMBL-EBI) Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom
"Comparative Genomics in Ensembl"

09-12-10 Martin Vabulas
Max Planck Institute of Biochemistry, Department of Cellular Biochemistry, Martinsried, Germany
"Intracellular proteotoxicity: Amyloid-like Aggregates Sequester Numerous Metastable Proteins with Essential Cellular Functions"

25-11-10 Vladimir Saudek
Sanofi-Aventis Research & Development, Lead Generation Compound Realisation, Structure Design & Informatics
"In search of function of FTO, a gene strongly implicated in the obesity phenotype"

18-11-10 Benjamin Audit
Laboratoire Joliot-Curie et Laboratoire de Physique, CNRS, Ecole Normale Supérieure de Lyon & Université de Lyon, France
"Spatio-temporal programme of replication and large-scale organisation of the human genome"

17-11-10 Albert Vilella
European Bioinformatics Institute (EBI), Cambridge, United Kingdom
"Developing trees in Ensembl"

05-11-10 Ángel Rubio
CEIT y TECNUN, Universidad de Navarra, Pamplona, Spain
"Integration of miRNA and mRNA expression data using Lasso regression"

22-10-10 Arnd Benecke
IHES, Paris, France
"The Analysis of Non-pathogenic SIV Infection in Natural Hosts Exemplifies the Need for New Mathematical Genome Representations"

19-10-10 Hernán A. Burbano
Department of Evolutionary Genetics, Max Planck of Evolutionary Anthropology, Leipzig, Germany
"Targeted sequencing of Neanderthal DNA: insights into recent coding and non-coding evolution on the human lineage"

16-09-10 Ariel Fernández
Rice University, Houston, Texas, USA
"Evolutionary insights into the control of drug specificity"

08-09-10 Ralf Jauch
Genome Insitute of Singapore, Singapore
"Where to Pair? – Cooperativity in Transcription"

22-07-10 Federico Agostini
Industrial and Molecular Biotechnology, University of Bologna, Italy
"NMR characterization of BpUreE: a metallo-chaperone from the urease system"

15-07-10 Bárbara Montserrat
Barcelona Supercomputing Center, Life Sciences Department, Barcelona, Spain
"ReLA, a local alignment search tool for the identification of regulatory regions and conserved transcription factor binding sites"

08-07-10 Mark Borodowsky
Bioinformatics and Computational Genomics, Georgia Tech, Atlanta, Georgia, USA
"Where do We Stand with Gene Finding at a Time of Next Generation Sequencing?"

01-07-10 Mario Cáceres
Institut de Biotecnologia i de Biomedicina "Vicent Villar Palasí", UAB, Bellaterra, Barcelona, Spain
"INVFEST: Functional and evolutionary analysis of polymorphic"

28-06-10 Eugene V. Koonin
NCBI, NLM, NIH, Bethesda, USA
"The Tree (or Forest?) of Life in the Age of Genomics"

10-06-10 Ori Kalid
Computational Drug Discovery, EPIX Pharmaceuticals, USA
"Structure-based discovery of small molecule correctors of deltaF508 CFTR"

31-05-10 Matteo Bellucci
Bioinorganic Chemistry Group, Department of Agro-Enviromental Science and Technology, University of Bologna, Italy
"Urease and hydrogenase accessory systems from Helicobacter pylori: the role of the metallo-chaperone UreE"

12-05-10 Masha Plyuscheva
Center for Advanced Studies (CEAB, CSIC), Blanes, Girona
"Bioluminescence and fluorescence in scale-worms (Polychaeta, Polynoidae)"

30-04-10 Evgeny Rogaev
University of Massachusets Medical School, Program in Neuroscience, Worcester, USA
"Molecular Genetic pathway to Alzheimer's disease"

29-04-10 Ana Rojas
Computational Biology and Bioinformatics, IMPPC, Badalona, Spain
"A Computational approach to Dna Damage Response"

15-04-10 Seirian Summer
Behavioural & Population Ecology, Institute of Zoology, Zoological Society of London, United Kingdom
"Phenotypic plasticity in queen and worker castes of eusocial wasps"

09-04-10 Nina Stoletzki
Centre for the Study of Evolution, School of Life Sciences, University of Sussex, Brighton, United Kingdom
"Disentangling selective and non-selective forces @ synonymous sites"

11-03-10 Lars S. Jermiin
CSIRO Entomology, Canberra, Australia
"Comparative Genomics, Phylogenomics, and the Frustrating Role of Model Misspecification"

01-03-10 Dan Tawfik
Department of Biological Chemistry, Weizmann Institute of Science, Rehovot, Israel
"How do proteins evolve?"

25-02-10 Ivan Junier
Institute of Complex Systems, Paris, France
"Genomic organization in the bacterial world"

01-02-10 Federica Batisttini
University of Sheffield, United Kingdom
"Sequence Selectivity of DNA Wrapping in the Nucleosome"

28-01-10 Jack Werren
University of Rochester, New York, USA
"Using Nasonia and Haploid Genetics to Reveal Evolutionary Genomic Processes of Adaptation and Speciation?"

CELL & DEVELOPMENTAL BIOLOGY PROGRAMME

19-11-10 Thomas Surrey
Cell Biology and Biophysics, EMBL, Heidelberg, Germany
"In vitro reconstitution of a minimal anaphase spindle midzone"

26-10-10 Stefan Westermann
IMP, Vienna, Austria
"Regulation of microtubule plus-ends by the Aurora kinase Ipl1p"

21-10-10 Sophie Jamet
University of Paris, France
"How to sort out thousand categories of axons? Odorant Receptors role in the developing olfactory map in mice"

07-10-10 Omar Quintero
Penn State College of Medicine, Pennsylvania, USA
"Cellular powerhouses on the move: a novel form of myosin-based mitochondrial transport"

23-07-10 Benjamin Glick
The University of Chicago, Chicago, Illinois, USA
"A new layer of regulation in the secretory pathway"

01-07-10 Ilya Leventhal
Kai Simons lab, Max Planck Institute, Cell Biology and Genetics, Dresden, Germany
"A new layer of regulation in the secretory pathway"

01-06-10 Ewa Paluch
Max Planck CBG, Dresden, Germany
"Control of bleb and lamellipodia formation during cell migration"

29-04-10 Christopher Burd
University of Pennsylvania, USA
"Regulated sorting in the endosomal system"

27-04-10 Alessio Maiolica
Institute of Molecular Systems Biology, Zurich, Switzerland
"Structural investigation of the Ndc80 complex by chemical cross-linking and mass spectrometry"

26-04-10 Rachele Allena
Ecole Centrale, MSSMat Laboratory, Paris, France
"Drosophila embryo: a multiscale model for cell mechanics"

13-04-10 Yoana Arboleda
Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany
"Mechanical cell properties in germ layer progenitor migration during Zebrafish gastrulation"

01-04-10 Claudio Collinet
Zerial Group, Max Planck Institute for Molecular Cell Biology and Genetics, Dresden, Germany
"Systems Survey of Endocytosis by Functional Genomics and Quantitative Multi-Parametric Image"

08-03-2010 Laure Saias
Lab de Physico Chimie, Institut Curie, Paris, France
"Development of a new Lab-On-Chip device for rare cell sorting. Toward circulating tumor cell capture and cancer treatment orientation"

04-03-10 Eva Kiermaier
IMP - Research Institute of Molecular Pathology, Vienna, Austria
"Construction of an artificial kinetochore in vivo"

GENES & DISEASE PROGRAMME

14-12-10 Ilaria Guella
Department of Biology and Genetics for Medical Sciences, University of Milan, Italy
"Searching for myocardial infarction susceptibility genes".

10-12-10 Georg Auburger
Section Molecular Neurogenetics, Goethe University, Frankfurt am Main, Germany
"Polyglutamine neurodegeneration, Ataxin-2 and RNA processing"

26-10-10 Hyun Hor
Centre for Integrative Genomics, "Le Génopode", University of Lausanne, France
"Molecular genetics of narcolepsy"

25-10-10 Charlotte N. Hor Henrichsen
Group of Prof. A. Reymond, Centre for Integrative Genomics, University of Lausanne, France
"Genome structure and gene expression in human and mouse"

20-09-10 Batsheva Kerem
Department of Genetics, The Life Sciences Institute, The Hebrew University, Jerusalem, Israel
"The molecular basis for genome instability in early stages of cancer development"

14-04-10 Marc Friedländer
Max Delbrück Center for Molecular Medicine, Rajewsky laboratory, Berlin, Germany
"Discovery and profiling of small RNAs using deep sequencing"

11-01-10 Pablo Villoslada
Neuroimmunology Group, Department of Neurosciences, Institute of Biomedical Research August Pi i Sunyer, Hospital Clinic, Barcelona, Spain
"Development of neuroprotective therapies for brain diseases"

DIFFERENTIATION & CANCER PROGRAMME

09-09-10 Renee Beekman
Sep of Hematology, Erasmus University Rotterdam, The Netherlands
"Identification of haplo-insufficient genes (HIGs) in AML using retroviral insertion mutagenesis"

20-07-10 Gioacchino Natoli
Department of Experimental Oncology, European Institute of Oncology (IEO), IFOM-IEO Campus, Milan, Italy
"The epigenome and the control of inflammatory gene expression".

15-06-10 Charo Robles
Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Munich, Germany
"Quantitative proteomics applied to the mammalian circadian clock"

11-06-10 Sofia Francia
Fabrizio d'Adda di Fagagna's lab, IFOM Foundation-The FIRC Institute of Molecular Oncology Foundation, Milano, Italy
"A novel role for the RNA-interference machinery in the regulation of the DNA-damage response"

27-05-10 Jessy Cartier
INSERM U866 "Lipides, Nutrition, Cancer", Faculté de Médecine, Dijon, France
"The influence of cIAP1 on the balance between proliferation and differentiation and on the transcription factor E2F-1"

10-05-10 Cornelis Murre
Division of Biological Sciences, University of California, San Diego, La Jolla, California, USA
"Regulatory network that underpins the development of early hematopoietic progenitors."

27-04-10 Achim Leutz
Humboldt University, Berlin-Brandenburg Center for Regenerative Therapies, Max-Delbrueck-Center for Molecular Medicine (MDC), Berlin, Germany
"Post transcriptional C/EBPbeta modifications and epigenetic switchboard functions"

24-03-10 Constanze Bonífer
Division of Experimental Haematology, Leeds Institute of Molecular Medicine, University of Leeds, United Kingdom
"Dissecting the molecular principles of gene activation in the hematopoietic system using ES cell differentiation"

29-01-10 Jason Doles
Lab of Dr. Michael Hemann, Department of Biology, MIT, Cambridge, USA
"Interrogating chemotherapeutic response in mouse models of cancer"

GENE REGULATION PROGRAMME

05-11-10 Rui Gonçalo Martinho
Instituto Gulbenkian de Ciência, Oeiras, Portugal
"aPKC is required for spindle orientation during symmetric mitosis"

27-09-10 Cristina Ottone
CE.IN.GE Biotecnologie Avanzate, Naples, Italy
"The translational repressor Cup promotes germ plasm accumulation and germ cell development in Drosophila embryos"

07-09-10 Antonio Gentilella
Department of Neuroscience, Temple University School of Medicine, Philadelphia, Pennsylvania, USA
"Role and regulation of the co-chaperone BAG3 in cells of glial and neuronal origin"

12-08-10 Panagiotis Papasaikas
Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, USA
"Computational and Functional Analyses of Splicing Regulation in Drosophilids"

26-03-10 Angela Krämer
Department of Cell Biology, University of Geneva, Switzerland
"Unsuspected functions for splicing factor SF1"

23-02-10 Andreas Ladurner
EMBL, Heidelberg, Germany
"Charting the epigenetic landscape- from chemistry-in-action to genome-environment interactions"

23-02-10 Javier Martinez
IMBA, Vienna, Austria
"The power of siRNAs to discover and validate RNA metabolic enzymes"

15-ene-10 Sylvia Dyballa
Universität Tübingen, Interfakultäres Institut für Biochemie, Germany
"Myosin Dependent ER Inheritance and RNP Localization in Budding Yeast"

CORE FACILITIES

17-12-10 Paolo Ribeca
Algorithm Development Group, Centro Nacional de Análisis Genómico (CNAG), Barcelona, Spain
"An overview of mapping algorithms for high-throughput sequencing"

21-10-10 Luis Gustavo Guedes Corrêa
Fermentas
Expert Seminar on Molecular Cloning -"Get your clone in 3 days!"

07-10-10 Carsten Schulz
EMBL, Heidelberg, Germany
"FRET Constructs and Their Development for Applications in Cells"

07-10-10 Philippe Bastiaens
Max Planck Institute of Molecular Physiology, Dortmund, Germany
"FLIM Applications"

05-10-10 Eike Latz
Institut für Angeborene Immunität, Universitätsklinikum Bonn, Germany
"Molecular mechanisms of NLRP3 and AIM2 inflammasome activation"

08-07-10 Guillem Abril / Jaume Palou
Centre d'Empreses de Noves Tecnologies, Parc Tecnològic del Vallès, Cerdanyola del Vallés, Barcelona
"LabVIEW and DIAdem training course"

02-06-10 Niels Kruize
Covaris
"Covaris system: controlled DNA shearing for next generation sequencing and library construction, cell/tissue disruption for protein targets and biomar"

01-06-10 Cellectis Technical Personnel
Cellectis
"Meganuclease-driven targeted integration: fast and effortless generation of stable cell lines"

26-05-10 Stephanie Urschel / Yohan Royer
Thermo Scientific Genomics
"Overview of current strategies in qPCR design – Introducing Thermo Scientific Solaris™ qPCR Assays"

19-03-10 Denis Puthier
Laboratoire INSERM, TAGC/ERM206, Parc Scientifique de Luminy, Marseille, France
"TranscriptomeBrowser Training Course"

25-01-10 Guillem Abril
National Instruments Spain, Centre d'Empreses de Noves Tecnologies, Cerdanyola del Vallès, Spain
"LabVIEW and DIAdem training course"

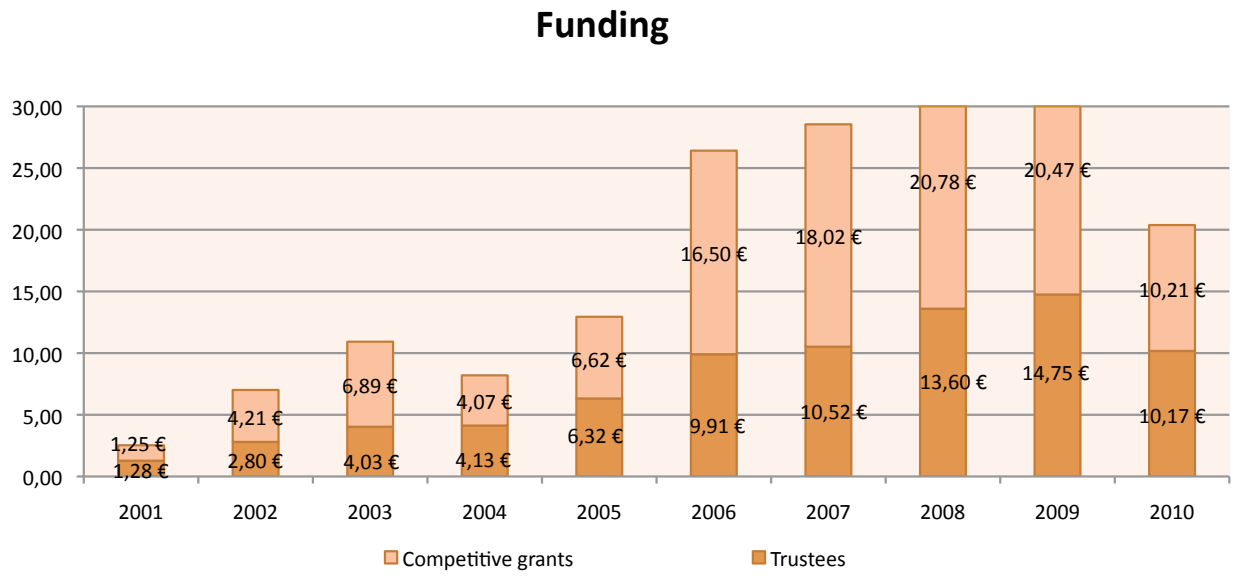
TECHNOLOGY TRANSFER DEPARTMENT

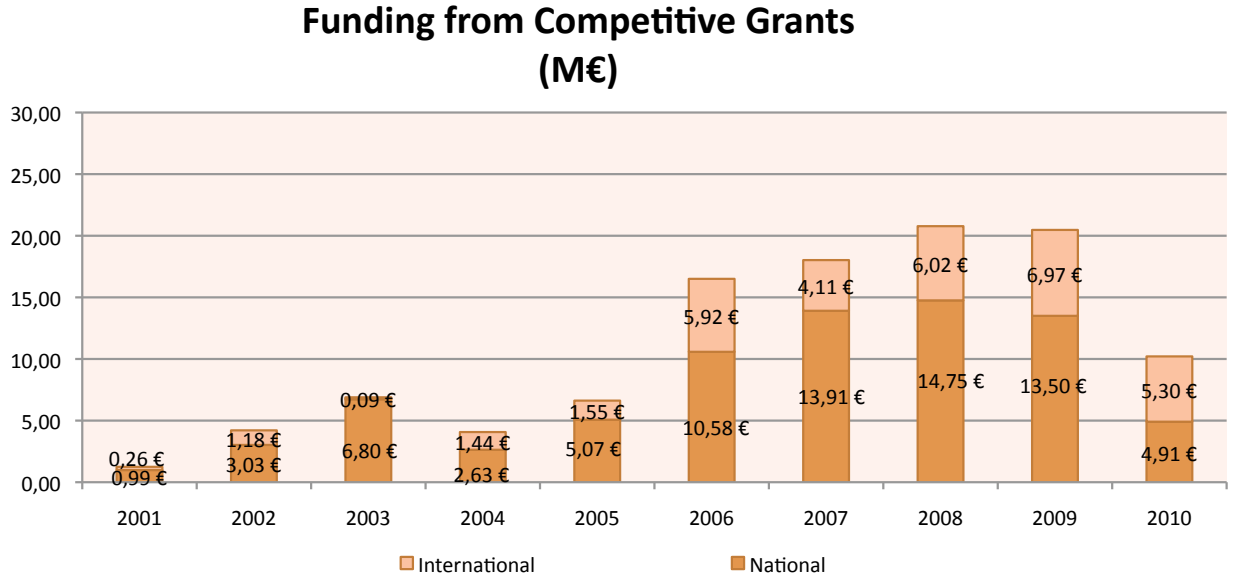
28-07-10 Jordi Quintana
Drug Discovery Platform, Parc Científic de Barcelona, Barcelona, Spain
"Chembiobank: A Chemical Biology Initiative for Drug Discovery"

Appendix 3

FUNDING

FUNDING EVOLUTION





GRANTS & OTHER FUNDING

Table including competitive funds, contracts and sponsorship obtained from 1st January to 31st December 2010 by the CRG.

ORGANISM	AMOUNT (€)
MINISTERIO DE CIENCIA E INNOVACION	4.535.507,33 €
EUROPEAN COMMISSION	3.612.311,19 €
ROCHE DIAGNOSTICS, S.L.	974.138,00 €
HUMAN FRONTIER SCIENCE PROGRAM	207.007,20 €
EUROPEAN MOLECULAR LABORATORY ORGANIZATION	150.337,04 €
FUNDACIO MARATO TV3	125.332,00 €
FONDATION JEROME LEJEUNE	99.999,99 €
NOVARTIS FARMACEUTICA S.A.	84.000,00 €
AGAUR - AGENCIA GESTIO D'AJUTS UNIVERSITARIS	76.769,16 €
MILLIPORE CORPORATION	62.843,71 €
THE LEUKEMIA & LIMPHOMA SOCIETY	53.067,29 €
NATIONAL INSTITUTE OF HEALTH NIH	47.260,00 €
NATIONAL MPS SOCIETY	30.902,70 €
CELLECTIS	30.000,00 €
OTHERS	118.995,31 €
TOTAL	**10.208.470,92 €**

OUTSTANDING RESEARCH PROJECTS

Table containing the most prominent competitive projects granted to CRG researchers from 1st January to 31st December 2010.

PROJECT TITLE	PROJECT TYPE	PI	ROLE	AMOUNT (€) RECEIVED BY CRG
"Rutes de senyalització Neuregulin/ErbB4 i desenvolupament dels circuits inhibitoris corticals."	Marato TV3 Foundation	Maria del Mar Dierssen	Partner	125.332,00 €
"Unveiling the Iberian lynx genome"	CSIC General Foundation	Roderic Guigó	Partner	4.822,00 €
"ESGI_European Sequencing and Genotyping Infrastructure"	PF7 Collaborative projects	Xavier Estivill	Partner	425.000,00 €
"Methods for high-resolution analysis of genetic effects on gene expression"	NIH Research Projects	Roderic Guigó	Partner	47.260,00 €
"Sharing capacity across Europe in high-throughput sequencing technology to explore genetic variation in health and disease: GEUVADIS"	FP7 Coordination Action	Xavier Estivill	Coordinador	340.642,00 €
"Proteomics Research Infrastructure Maximising knowledge EXchange and access (XS): PRIME-XS"	FP7 Collaborative Projects (Large-scale integrating project)	Henrik Molina	Partner	520.534,00 €
"Research infrastructure for imaging technologies in biological and biomedical sciences: EURO-BIOIMAGING"	Collaborative Projects (Large-scale integrating project)	James Sharpe	Partner	36.380,00 €
"Dissecting the molecular mechanisms regulating somatic cell reprogramming"	HFSP Program Grants	Maria Pia Cosma	Coordinador	117.728,00 €
"Odor recognition in natural environments: Bayesian inference from insects to mammals"	HFSP Program Grants	Matthieu Louis	Partner	120.000,00 €
"Mechanism of Unconventional Protein Secretion"	ERC Advanced Grants **	Vivek Malhotra	Individual	2.206.963,00 €
"Chromosome Segregation and Aneuploidy"	ERC Starting Grants (StG) **	Manuel Ernesto Mendoza	Individual	1.453.679,00 €
Quantitative Models of Cellular and Developmental Biology"	ESF Research Networking Programmes	Hernán López-Schier	Coordinador	529.250,00 €
"Multi-objective optimization for modeling developmental gene regulatory networks"	ERANET - Complexity	Yogi Jaeger	Partner	54.500,00 €
			TOTAL	**5.982.091 €**

INSTITUTIONAL AGREEMENTS

These are the agreements signed with companies and other organisations during 2010.

SPONSORS

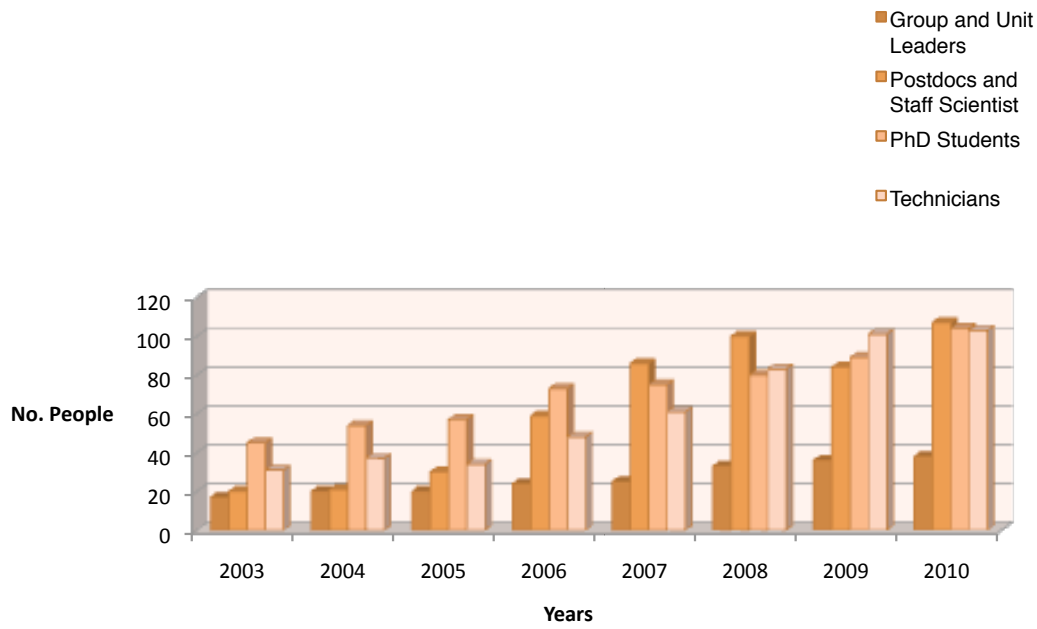
These are the companies that have contributed to the activities of the CRG during 2010.

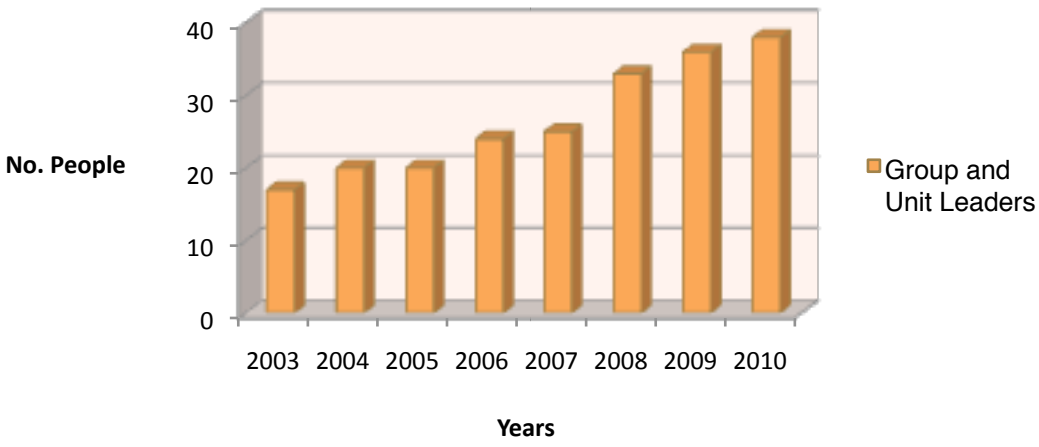
STRATEGIC PARTNERS

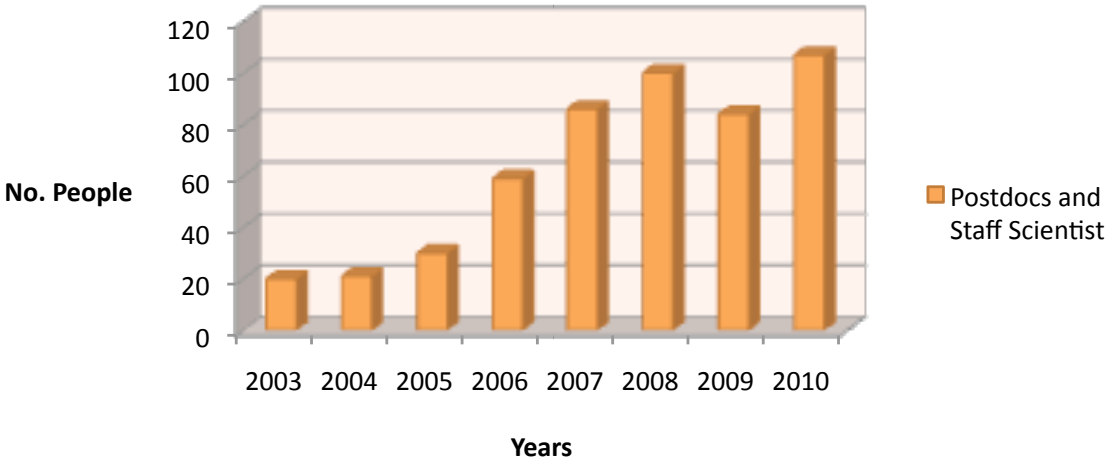
These are CRG partners, who contribute to strategic projects of the institute.

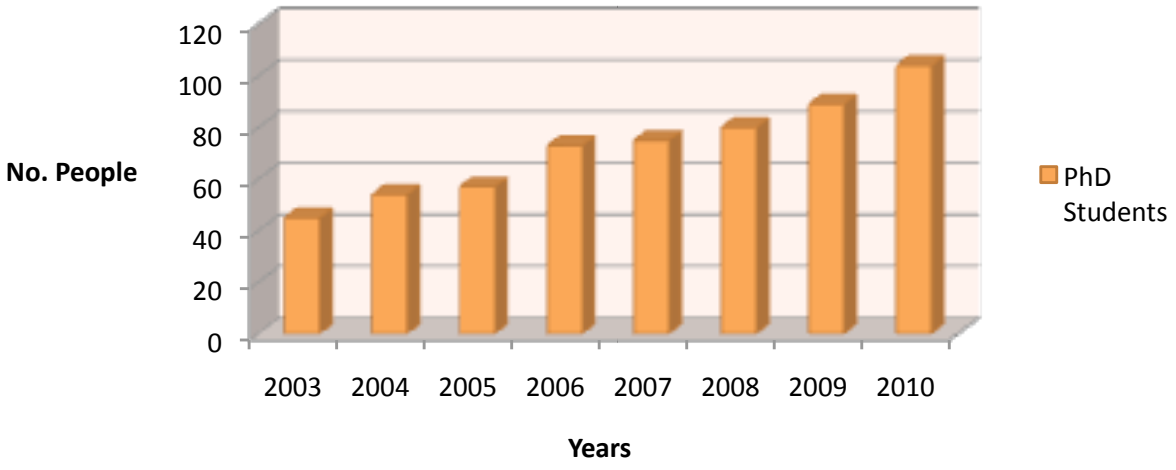
Appendix 4

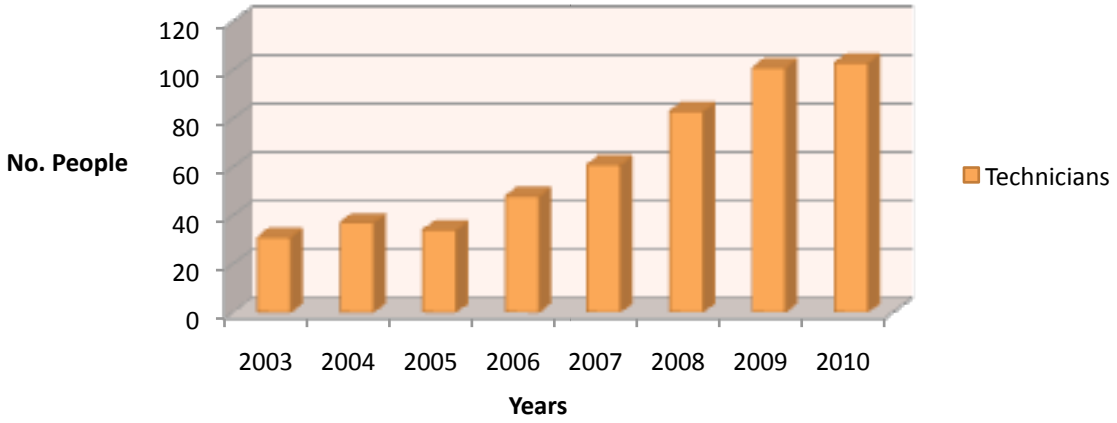
PERSONNEL EVOLUTION

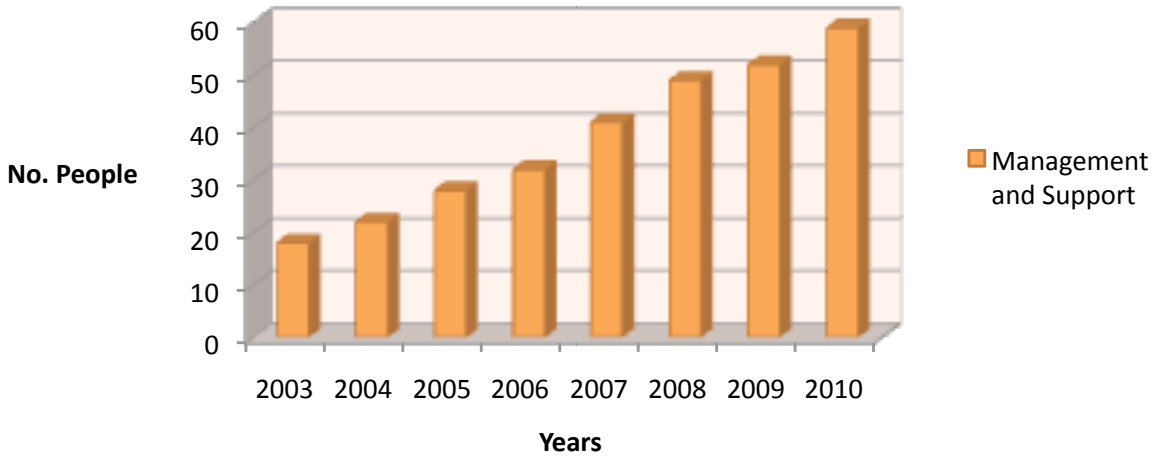





No. People
120
100
80
60
40
20
0
2003
2004
2005
2006
2007
2008
2009
2010
Years
Postdocs and Staff Scientist


No. People
120
100
80
60
40
20
0
2003
2004
2005
2006
2007
2008
2009
2010
Years
PhD Students


No. People
120
100
80
60
40
20
0
2003
2004
2005
2006
2007
2008
2009
2010
Years
Technicians


No. People
60
50
40
30
20
10
0
2003
2004
2005
2006
2007
2008
2009
2010
Years
Management and Support


Premios con aire de cuplé
El Liure acoge la entrega de los Ciutat de Barcelona
ep europa press
Ciencia
Los altos cargos de la ciencia siguen siendo un coto masculino
DIARIO MEDI
Los sindicatos pelearán para que los complementos salariales no se toquen
Los nuevos dioses
«Els polítics han d'entendre que sense ciència no hi ha riquesa»
mujer hoy
CIENTÍFICAS
europapress.es
Roderic Guigó
BIOINFORMÀTIC DEL CRG
Secrets d'alcova de les cèl·lules mare
Arriba a Barcelona un secuenciador de genoma ultra rápido
Barcelona reuneix els millors genetistes per identificar els elements del genoma

Appendix 5

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