CRG
Centre for Genomic Regulation

2017 Annual Report

EXECUTIVE SUMMARY

EXCELENCIA SEVERO OCHOA

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

TEXT AND GRAPHICS: CRG scientists, CRG members of the administration and research support team, Kat Arney, Department of Communication & Public Relations

GRAPHIC DESIGN: Ondeuev Comunicació S.L.

PHOTOGRAPHY: Ivan Marti, CRG scientists

VIDEO: Adrià Sunyol

LEGAL DEPOSIT: B 17126-2018

Contents

Foreword

This year (10th April), the official signing of agreements for the establishment of the EMBL outstation in Barcelona finally took place. After more than 20 years, the EMBL opened a new site at the PRBB in Barcelona adjacent to the CRG. This will enhance the PRBB's international visibility and will create, together with the CRG light microscope facility, a leading imaging facility in Europe.

One of the most important events of 2017 was the creation of a network comprising all the Spanish institutes and university units distinguished with the Severo Ochoa and Maria de Maeztu excellence grant award. This alliance (SOMMa, http://somma.es/) has been created to promote Spanish Excellence in research and to enhance its social impact at national and international levels. The CRG currently chairs the alliance.

In the area of science, in the course of the year, our Bioinformatics and Genomics programme was evaluated. The feedback was excellent and endorsed the continuation and reinforcement of Bioinformatics at CRG.

In 2017, three of our young PIs found senior positions at other institutes. Matthieu Louis is now at the Molecular, Cellular and Developmental Biology Department of the University of California, Santa Barbara in the USA; Fyodor Kondrashov has moved to the Evolutionary Genomics Department of the Institute of Science and Technology (IST Austria), in Klosterneuburg, Austria; and Manuel Mendoza is now Team Leader at the Institute of Genetics, Molecular and Cellular Biology (IGBMC) in Strasbourg, France. Finally, James Sharpe moved from CRG to become the Head of the new EMBL Barcelona site. We are delighted that all of them have found excellent positions, thus attesting to the CRG's success in training and promoting young scientists. On the other hand, Elvan Boke and Nicholas Stroustrup, formerly of the Harvard Medical School in the United States, joined the CRG at the beginning of 2017.

Also during the year, and in accordance with our strategic plan, we changed the way we recruit new PIs. For the first time ever, we advertised positions not attached to any programme on the main research topics of our plan. We received more than 150 applications and, after the short-listing and interview phase, we hired two new PIs, Eva Novoa, from the Garvan Institute of Medical Research in Australia, and Arnau Sebe-Pedros, from the Weizmann Institute of Science in Israel, who will be joining us at CRG in 2018 and 2019, respectively.

In the realm of finance, the CRG continues to attract competitive funds. It is important to mention that we were awarded our second four-year 'Centres of Excellence Severo Ochoa Award' by the Spanish Ministry of Economy, Industry and Competitiveness. Additionally, four groups entered two key national technological platforms, for bioinformatics (Guigó, Gabaldón, Gut) and proteomics (Sabidó), supported by the Institute of Health Carlos III (ISCIII). At European level, and besides numerous collaborative projects, particular mention should be made of an ERC Starting Grant (Boke) and an ERC Proof-of-Concept (Valcárcel), the latter aimed at testing the therapeutic potential of new reagents for lung cancer therapy. In the international arena, Roderic Guigó secured two new research awards, from the National Institutes of Health (NIH).

CRG supports Open Science, and in this regard, it is important to mention the EU project led by CRG: ORION. This is a new initiative to promote institutional changes in research funding and performing organisations to make them more receptive to societal needs and to embrace the principles of Open Science.

It is important to emphasize that CRG would not be such an attractive place without the professional and excellent support of its different administration departments which, despite a heavy workload, have managed to create an excellent working environment. More particularly, one of the key recommendations from the external panel in 2016, the ¨business process management¨ project, was implemented in the course of last year with the goal of evaluating and enhancing the administration's performance and effectiveness, thus enabling department-wide continuous improvement.

In a year full of challenges, for the first time we had to cope with the departure of many groups at the same time, as well as with a regional and national political crisis. Moreover, our VAT deductibility is under debate, entailing a significant financial impact and major uncertainty. The support of all our personnel enables us to minimize the impact of these challenges and to guarantee the support and success of all our PIs. CRG continues to be a great place to engage in science, and exciting opportunities and projects await us in 2018.

Luis Serrano
Director

Institutional Highlights

Administration

The administrative team (including CNAG-CRG) is comprised of members from 10 different countries, and provided critical and dedicated support to the CRG community in 2017 in accordance with the priorities defined in the institute's 2017-2021 strategic plan and annual action plan.

One of the key recommendations from the external panel, the "business process management" project, was implemented in the course of last year with the goal of evaluating and enhancing the administration's performance and effectiveness, thus enabling continuous improvement across departments. A number of team-building, team development activities and work-life balance initiatives were continued or launched, including the administrative get-together involving group work around the topic of innovation (drawings, short presentations, joint brainstorming and science activities), the teleworking programme, annual performance evaluations, and easy-science activities.

The administrative team contributed to a number of CRG-wide initiatives: the implementation of the LIBRA gender balance action plan, the HR Excellence in Research project, the first gathering of the alumni board in Barcelona, the recruitment of new PIs, and the seamless departure of several research groups, which coincided with a significant peak of junior PIs leaving the institute, an eventuality provided for in the scientific turnover policy. Within the context of the establishment of the local partnership between the CRG and EMBL and with the financial support of the Spanish Ministry for Economy, Industry, and Competitiveness, the administrative team provided advice and support to the recently-created EMBL site in Barcelona in matters relating to IT, health and safety, facility management and purchasing. The new health and safety initiatives achieved in 2017 include the digitisation of a centralised biosecurity register for the use of biological agents, as well as the appointment of lab safety coordinators as local contacts to guarantee the proper use of health and safety regulations in the lab.

With the goal of reaching beyond the CRG, the members of the administration continue to share benchmarks and engage with other institutes at home and abroad, particularly with EU-LIFE institutes. Roundtables were held locally with CERCA institutes to exchange knowledge and best practices. More specifically, the managing director was invited speaker on two occasions, presenting the evaluation of the administration performed in 2016 by an international external expert panel, thus serving as a model for other national institutes.

Advanced Training

Advanced training at different career stages lies at the core of the CRG's activities. From Master and PhD students to postdoctoral fellows and junior group leaders, the CRG offers a stimulating and enriching environment, a comprehensive training programme and tailored courses on scientific and transferable skills.

ACADEMIC PROGRAMMES

The CRG International PhD Programme is the flagship of the CRG's commitment to training, and is characterised by a combination of dedicated mentoring and scientific training to support early independence and creativity. The programme has been running for more than 10 years and has remained very attractive and competitive, receiving around **400 applications** from over **50 different countries** and admitting **20-30 students every year**.

In 2017, the International PhD Programme continued to attract up-and-coming talent from all over the world through the support of internal and external competitive funds. In addition to the annual **CRG PhD Call,** and together with the IRB Barcelona, the VHIR and the IDIBAPS, we have been running a training programme on research for medical doctors for four years now, called the **PhD4MD programme.** The programme's long-term goal is to train the next generation of physician-scientists who will drive research and bring about a greater impact on patients. A total of 4 MDs joined the CRG between 2014 and 2017.

The CRG Postdoc Programme enables young researchers to work on highly collaborative and interdisciplinary projects across different groups and units at the CRG. After two rounds of successful funding in 2009 (**INTERPOD) and** 2014 (**ImPuLSe**), the Postdoc Programme has been awarded a third grant from the EC through its Marie Skłodowska-Curie H2020 COFUND Actions for the 2017–2022 period (INTREPiD). The third COFUND Programme enables collaborations with other academic institutions and the industry, thus allowing postdocs to establish and maintain connections that will prove useful in building their careers beyond the CRG. Several research institutes, including the Institute Curie, the Research Institute of Vall d'Hebron (VHIR) and companies such as Novartis and Esteve, joined the INTREPiD Programme as external partners.

The **INTREPiD** programme includes a training curriculum targeting non-scientific researcher skills – for example, research ethics, outreach and gender in science – and offering dedicated career development support. The first call opened in September 2017, and 6 new fellows were selected.

Finally, the 5th edition of the **International CRG Summer Internship Programme** for undergraduate students was launched in February 2017: 5 internships were awarded to provide undergraduate students with the opportunity to conduct research at the CRG over the summer (countries of origin: Russia, Denmark, the Ukraine, Italy and the Czech Republic).

The CRG has also several active partnerships with Master programmes to host students for a research internship: UPF, UB, UPC, Faculty of Sciences of the Saint-Joseph University (Beirut), MIT (USA), etc. This year, we agreed to initiate a new alliance with the University of Richmond (USA) to host their Undergraduate students for a research-intensive internship over the summer next year. We will continue to explore other alliances to attract top junior researchers.

The CRG also contributed to the development of the new **BIST Master of Multidisciplinary Research in Experimental Sciences** launched last October 2017. Two students were awarded a fellowship to do Masters at the CRG. A new edition of the Masters is currently open.

TRAINING PROGRAMME

In 2017, the **CRG Advanced Training Programme** was enhanced with several new high-level international courses within the framework of Courses@CRG, as well as a more extensive offering of internal scientific courses and workshops. As in previous years, the courses organised by the CRG were open to both local and international scientific communities and delivered high-quality training in the latest scientific breakthroughs and technologies.

In 2017, the CRG Training Unit co-organised and delivered seven high-quality **Courses@CRG**, one EMBO course and a joint course with FIMM, which were attended by 170 participants from research institutes in 34 different countries, including Argentina, Australia, Israel, Mauritius, Singapore, South Africa and the USA.

- Chromosomal conformation (F. Le Dily)
- Advanced Proteomics (E. Sabidó)
- Tissue Engineering: From stem cells to organoids (L. Batlle)
- Whole-Cell Modelling (L. Serrano, M. Lluch)
- Nextflow: Reproducible in silico genomics (P. Di Tommaso, C. Notredame)
- C. elegans: CRISPR, RNAi and Genetics (B. Lehner)
- CRISPR/Cas9: Genome Editing tool (C. Carolis, L. Di Croce)
- EMBO-Targeted proteomics: Experimental design and data analysis (E. Sabidó)
- Epigenetics in Clinical and Translational Research (L. Di Croce) in collaboration with FIMM and held in Helsinki

The Courses@CRG attracted sponsorship/partnership from Amazon, Bio-Cat, Biogen, IDT, Izasa Scientific, Leica, Nema Metrix, Nikon, Sigma Aldrich, Sonidel, Stem Cell Technologies and Termo Fisher.

A key highlight in 2017 was the development of a rich portfolio of **internal courses focusing on scientific and technology skills** and **transferable skills**, research integrity and career coaching. Internal training has opened up new opportunities for trainers and trainees alike. Technicians, PhD students and Post-doctoral researchers gained experience in teaching and sharing their knowledge and expertise with the rest of the CRG community, while participants benefited from the wide range of training activities offered throughout the year (see below). The 23 workshops straddled all levels of scientific careers, from undergraduate students to group leaders, and were attended by some 450 participants from all CRG and CNAG-CRG programmes and fields.

Eleven Scientific & Technical workshops: Linux, Cluster, Bioinformatics, programming in Python and R, Software in the cluster) were attended by almost 150 CRG and CNAG-CRG attendees.

In 2017, Electronic Lab Book (ELN) training was implemented for all 1st year PhD students at the institute as well as a Research integrity and Open Science starter course.

Since May 2017, an online Research Integrity course has been mandatory for all new-comers at the CRG. It consists of interactive modules, quizzes and short video tutorials on ethics and research integrity.

Almost **300 CRG and CNAG-CRG students, post-docs, technicians and PIs** attended 12 courses strengthening their skills in science communication, management, leadership and entrepreneurship.

Learning by doing is an initiative to enable CRG researchers to do an internship in different management departments of the CRG in order to glean practical experience in managing scientific projects, science communication and in organizing events and trainings for scientists, high school teachers and a general audience. We recruited 10 researchers in the following departments: ISA, Communication and Grants.

Communications, Public Engagement & Science Education

The mission of the CRG Communications is to promote and protect the CRG's reputation of excellence. We do this by developing and implementing clear, consistent and engaging strategies to enhance public understanding of our science and our institute, its researchers and its value to society.



The second edition of the citizen science project **'Stick Out Your Tongue'** started at the end of 2016. This project, within the scope of open science, is the first study into the mouth's microbiome that aims to study the genetic footprint of different microbial communities and explore a possible relationship with environmental characteristics or lifestyle. Funded mainly by the "la Caixa" Bank Foundation and the CRG, the second run of the project aimed, in addition, to incorporate new scientific variables and focal points, tackle new challenges, reach new target audiences and involve and collect samples from different populations and patients suffering from different diseases. This was done by organising a new tour across Spain and by offering new activities based on innovative and participative formats that allowed us to engage with these new target groups, communicate our research and objectives and provide feedback about the project to the participating citizens and the general public. Thanks to this improved strategy, we reached 3,700 people, 34 schools, 13 patient associations and 21 cultural centres.

Throughout the year, the CRG continued to offer public engagement and science education activities, although some initiatives merit particular mention. In the framework of a collaboration with the miniPCR company and the 'Stick Out Your Tongue' project, we were invited to participate in the **Cambridge Science Festival** in Boston in April. This was a truly enriching and successful experience and brought us positive inputs. We also launched a new and successful workshop addressed to parents entitled **'Your children can become scientists'**, aimed at giving tips and advice to parents with children that need to decide their career path. We also engaged in a new pilot education project with **mSchools**, entitled **'The Complexity of Life'**, which consists of a set of materials and resources to introduce students into the field of life sciences through mobile technologies. mSchools is a multi-faceted mEducation programme of Mobile World Capital Barcelona in partnership with the Government of Catalonia, the Barcelona City Council and GSMA, empowering students and teachers to integrate mobile technologies into the classroom, opening up new ways of teaching and learning that improve achievement and employability. The CRG was also one of the institutes hosting students from the **Barcelona International Youth Challenge (BYSC)**, organised by the Catalunya La Pedrera Foundation. The BYSC is a two-week international excellence programme that seeks to stimulate scientific talent among young people from all over the world and to encourage their enthusiasm for pursuing scientific research and a career in science.

From the training standpoint, the department hosted 5 interns in the **'Learning by Doing'** internal internship programme, who carried out small projects related to public engagement, science education and media relations. During the year, we also contributed to two new important institutional initiatives: **ORION** and **SOMMa**. **ORION** is a CRG-coordinated H2020 project launched in 2017. The project's main topic is open science and it explores ways in which research and funding organisations in life sciences and biomedicine can open up the way they fund, organise and engage in research. The

Communications & PR team contributes to several tasks in ORION, particularly the public dialogue and citizen science experiments. **SOMMa** is the alliance of Severo Ochoa Centres and María de Maeztu Units to promote Spanish Excellence in research and to enhance its social impact at home and abroad. The involvement of Communications & PR in SOMMa targets the promotion, visibility and outreach of this alliance.

CRG scientists were also deeply engaged in media activities throughout the year in connection with the new findings published in high-profile journals and institutional and outreach activities that yield articles in written and online media, as well as participation by scientists in radio or TV programmes. The economic value of these media appearances surpassed 13MEUR.

2017 was also a busy year in terms of the organisation of seminars, sessions and scientific meetings, including the international scientific meetings **"Barcelona NGS'17: Structural Variation & Population Genomics"**, co-organised by the International Society for Computational Biology (ISCB); **"Enhancing the Usage of Human Genomics data for the benefit of all. Genotype Tissue Expression (GTEx) project meeting"**; **"Young Scientist Networking Meeting 2017"**, organised together with EMBO and India BioScience; the **"V Core Management Workshop: Training the Trainers"**; the **"16th CRG Symposium: Seventh International Workshop on Genomic Epidemiology"**; the "**15th RECOMB Comparative Genomics Satellite Workshop"** and the **"CRG Annual Proteomics Symposium 2017: Unveiling the complexity of the cell proteome"**.

External funding

In line with the results of the previous year, research groups across the Institute secured the appropriate level of funding required to strengthen their position nationally and internationally.

Once again in 2017, the funding secured at national level spiked considerably (>€10 million). The Institute's scientific leadership was recognised again with a new four-year "Centres of Excellence Severo Ochoa Award" by the Spanish Ministry of Economy, Industry and Competitiveness. In addition, four groups entered two key national technological platforms for bioinformatics (Guigó, Gabaldón, Gut) and proteomics (Sabidó), supported by the Institute of Health Carlos III-ISCIII.

Within the framework of the new regional Strategic Plan for Health Research and Innovation (PERIS 2016-2020), two collaborative projects were launched involving CNAG-CRG groups in personalised medicine, based on genomics analysis for clinical decision-making in oncology (Gut) and undiagnosed rare neurological diseases (Beltrán).

Personalised medicine is also the focus of new collaborative grants awarded under the H2020, revolving around the integration of heterogeneous big data (IASIS/Tartaglia, EGA) and rare diseases (SOLVE-RD/Beltrán), as well as a pilot on improving data reproducibility, reusability and interoperability (European Open Science Cloud-EOSC for Research Pilot).

At European level, highlights include an ERC Starting Grant (Boke) and an ERC ProofofConcept (Valcárcel), the latter aimed at testing the therapeutic potential of new reagents for lung cancer therapy. Moreover, five post-doctoral fellows were awarded a Marie S. Curie fellowship (Fernández, Audergon, Stik, Rogalska, Schmiedel), representing a success rate of more than 50% for the Institute.

In 2017, several research groups extended their source of external support at international level and leveraged private funding, including, among others: two new research awards from the National Institutes of Health-NIH (Guigó) and from the Department of Defense-DoD (Serrano/Gill), respectively; and a research grant awarded by FUNDELA focused on identifying new paths for treatment of amyotrophic lateral sclerosis-ALS (Di Croce).

Innovation and Technology Transfer

The mission of the Technology and Business Development Office at the CRG is to leverage the institution's research for the public good and to further the economic growth of the life sciences sector in our region. Our aim is for the scientific results generated at the CRG to develop into novel therapeutic, diagnostic and other types of products that will contribute to our society's wellbeing.

Irrespective of the stage of development or whether we perform basic or applied science at the CRG, the TBDO is firmly convinced that the most beneficial products and services stem from disruptive scientific achievements. We expect these results to generate a much higher impact in and on society if they receive the right support and are channelled through an appropriate commercialisation strategy.

Consequently, in the course of 2017, TBDO boosted its activity through regular meetings with researchers, holding more than 450 meetings, as well as by guiding CRG innovators into the technology transfer process through different training courses such as: the 4th edition of the **Bio-Business School**, the 2nd edition of the **"From Science to Business"** course, organised by the BIST institutes in collaboration with the ESADE Business School and through different business seminars and round tables.

Driven by its commitment to promote entrepreneurship, the TBDO team actively supported the 2016 awardee of the **S2B Concept** challenge 2nd edition. As a result of team work, the spin-off Microomics S.L. was successfully incorporated in September 2017 and is now actively developing and delivering high-quality metagenomics solutions to companies and research institutes.

microomics®
SMALL THINGS THAT MATTER

In addition, in 2017 TBDO managed 7 projects under its **Commercialization Gap Fund** programme. It should be mentioned that one of these projects, by Jordi Hernandez, a researcher at Juan Valcárcel's laboratory and a young entrepreneur participating in the CRG **Entrepreneur-in-Residence** programme, conducted this innovative project with the support of the CaixaImpulse programme as well.

Also noteworthy is the record number of inventions received and assessed (30), the new priority patents filed (2), increasing the number of patent families active during 2017 to 9, and the number of agreements successfully negotiated with companies, which was 89% up on the previous year's figure. In terms of economic impact for the CRG, the TBDO negotiated and secured around €279,000 (39% increase on the previous year).

In the course of 2017, the TBDO team furthered its engagement in different national and international Biotech/Pharma-related events, such as the **Bio International Conference**, **Bio Europe**, the **2nd Annual Translational Microbiome Conference**, etc.; and held 153 meetings with companies and investors to discuss CRG's technology portfolio and potential future collaborations.

Moreover, Pablo Cironi was guest lecturer at the IESE Business School and participated in different round tables to discuss strategies for marketing early-stage innovations.

Overall, 2017 was a very good year, and we envisage a positive entrepreneurial spirit in the coming years at the CRG.

National and International Dimension

SPAIN

SOMMa

The new alliance of Severo Ochoa centres and María de Maeztu units (SOMMa; www.somma.es) was officially launched in October 2017. SOMMa brings together 25 centres and 16 units accredited through these excellence awards and aims to raise the national and international profile of science in Spain, promote the exchange of knowledge, technology and best practices among its members, the international scientific community and main stakeholders, cooperate with other research centres in Spain to strengthen the R&D+i system and have a voice in Spanish and European science policy. This initiative was led by the CRG's Director, who became the first SOMMa Chair, and was implemented by the ISA team. The alliance's starting activities comprised the establishment of its own governance, the launching of the website and the organisation of task forces to address the different objectives.

EUROPE

Leading collaborative research and training networks

The CRG is leading several collaborative research and training networks funded by the European Commission Framework Programmes thanks to its excellent research effort and the dedicated support of its International and Scientific Affairs team. This ongoing leadership results in high visibility, a strong reputation and a relevant scientific and innovative output.

FP7/H2020-coordinated projects ongoing or initiated in 2017 include:

- **MycoSynVac**: Coordinated by Luis Serrano.
- **OPATHY**: Coordinated by Toni Gabaldón. This is a training network involving 13 Early-Stage Researchers to be hosted at 11 institutions in 7 European countries.
- **LIBRA**: Coordinated by Isabelle Vernos. This is the milestone project to implement gender-equality actions in EU-LIFE research institutes.
- **MiniCell**: Coordinated by Luis Serrano.
- **CellViewer**: Coordinated by Pia Cosma.
- **DivIDE**: Coordinated by Isabelle Vernos. This is another training network involving 11 Early-Stage Researchers to be hosted at 9 institutions in 7 European countries.
- **ORION**: Coordinated by Michela Bertero. This is a new initiative to promote institutional changes in research funding and performing organisations to make them more receptive to societal needs and to embrace the principles of Open Science.

Additional information on the ongoing collaborative projects led by the CRG is available at: http://www.crg.eu/content/research/projects/ec-coordinated.

In addition, the CRG continues to command a strong position in several relevant pan-European infrastructure networks, such as ELIXIR (as part of the Spanish node of the infrastructure), EuroBioImaging (leading one work package in the second preparatory phase and sitting on the interim board), the CORBEL, EXCELERATE and MuG H2020 projects.

European integration

EU-LIFE, the European Life Sciences Institutes for Excellence, is a key initiative chaired by the CRG to promote excellence in research, strengthen integration among European research institutes in life sciences and develop and share best practices in research, research management and training. Several CRG

members are actively participating in EU-LIFE working groups, and this participation yielded the following important outputs (among others):

- EU-LIFE active in the 2 existing EC stakeholders' platforms (ERA and Open Science)
- 3 EU-LIFE position papers: (FP9 statement, H2020, FP9 position paper)
- ERC 10 years' EU-LIFE campaign
- Contribution/Feedback to 8 EC reports
- 3 EC Consultations: H2020 mid-term review; Foresight | Delphi study for FP9 (by invitation); FET FLAGSHIP
- ORION: launching of Open Science H2020 project
- LIBRA Career Development Compass for post-docs implemented (2 workshops)
- LIBRA Recruitment Handbook and workshop; work-life balance workshop
- TedEx Tech Transfer initiative (6 pitches to VC professionals and 4 trainings)
- 3 events to build institutional capacity (Train the Trainer workshop for MSCA applications; making videos in-house for science communication; crisis communication workshop)
- Launching EU-LIFE visiting programme
- EU-LIFE scientific workshop "Principles of homeostasis"
- EU-LIFE signature courses poster and one EU-LIFE symposium (FIMM+CRG)
- Website & newsletter: improved tools for researchers (jobs, funding, science news)

The Core Facilities are members of the Core Facilities Excellence Alliance **"Core For Life"** (www.coreforlife.eu), which also includes EMBL (Heidelberg, Germany), VIB (Gent/Leuven, Belgium), MPI-CBG (Dresden, Germany), VBCF (Vienna, Austria) and the FGCZ (Zurich, Switzerland). Core For Life aims at sharing and consolidating procedures, joining efforts in personnel training and technology validation and sharing access to facilities across institutes.

WORLDWIDE

Through its International and Scientific Affairs team, the CRG explores new global opportunities to attract the most talented researchers, establish scientific collaborations and provide greater exposure for its research. A list of some of the most relevant international actions carried out in 2017 is provided below.

- **Coordination of CRG-Novartis-Africa mobility programme**: we hosted three new students from South Africa to develop their projects at the CRG's facilities and laboratories. We are now collaborating with Novartis to guarantee the continuation of the mobility programme.
- **Collaboration with the "Mujeres por Africa" Foundation**: we hosted the first senior woman scientist supported by the programme "Ellas investigan". Liz Kizito, head of the Agriculture Department at the Christian University in Uganda, was hosted by the Bioinformatics Core Facility to study the genetic variability of *Solanum aethiopicum* in Africa. Jointly with the Mujeres por Africa Foundation, we also organised the "Challenges and Opportunities for research in Africa" workshop that stimulated an interesting discussion among the speakers and the numerous attendees.

Women in Science

At the Gender Balance Committee (GBC) chaired by Isabelle Vernos, all the scientific communities are represented, as well as HR (see members at http://www.crg.eu/en/content/about-us-women-science/gender-balance-committee). The committee's goals are to eliminate gender bias in recruitment processes, attract and recruit female scientists, improve work-life balance, promote career development and establish and disseminate gender-sensitive practices. These goals are in line with the EU-funded LIBRA (Leading Innovative Measures to Reach Gender Balance in Research) project coordinated by the CRG.

In 2017, the CRG implemented actions included in the Gender Equality Plan (GEP) which was launched in August 2016 as part of LIBRA. GEP actions focus on four main areas: Recruitment, Career Development, Work-Life Balance and Gender/Sex dimensions of research.

Last year, the GBC's major focus was Career Development and Recruitment. Some highlights are presented below:

CAREER DEVELOPMENT

- Two female post-docs from the CRG were selected to participate to the **LIBRA Career Development Compass** programme that aims to support female researchers on their way to becoming independent researchers and pursuing a career in academia (http://www.eu-libra.eu/news/20-eu-life-post-docs-their-way-become-pi-first-milestone-knowing-how-they-want-their-career-be). The programme was very well received, and LIBRA is currently investigating opportunities to sustain it for future editions and to allow more post-docs to participate.

- Following the success of the **WOSS: Women Scientists Support Grant** implemented by the GBC last year, two calls were opened again in 2017. The WOSS grant scheme aims to provide support to women scientists who have the ambition and the potential to hold a leading position in research but have to face the challenges associated with maternity. In 2017, four women scientists obtained support from WOSS.
- As in the previous year, the CRG celebrated **International Women's Day** on the 8th of March. Flyers highlighting the career tracks of four female CRG alumni after leaving the CRG were distributed to PRBB personnel. A half-day event, including the showing of a movie followed by a debate with a large audience, was also organised on March 8th. (http://www.eu-libra.eu/news/celebrating-international-womens-day-2017-crg)
- A one-day teaching module on computer programming skills was offered to teenager students from the Barcelona area at the CRG to celebrate **Ada Lovelace day** (this was a joint initiative of the Communications & PR and women researchers at the institute).

RECRUITMENT

Several actions and measures were implemented to ensure a gender bias-free approach during recruitment processes. These include:

- New Job advert template, including information about Gender Equality, Work-Life Balance at the CRG and depicting a diverse work environment.
- A Gender bias video commissioned by the CERCA Institution is now shown to all Group Leader selection panel members to raise awareness of inappropriate procedures: https://www.youtube.com/watch?v=g978T58gELo
- A flyer has also been designed to list the benefits that the CRG provides to support a healthy work-life balance for CRG employees. Such benefits currently include access to childcare facilities, social benefits, dual career opportunities, teleworking etc. The flyer is shared digitally with every candidate that interviews with the CRG.

Scientific Highlights

Home sweet home

Picking the perfect place to live is important, especially if you're a retrovirus.

Unlike viruses such as the common cold, which infect cells and completely destroy them in the process, HIV and other retroviruses convert their genetic information into DNA and insert it into the genome of their host – a human immune cell known as a T cell.

Some infected cells use this viral DNA to make many copies of the virus, which go and infect new hosts. But others go into a resting state: the viral DNA still lurks within the cell's genome but it doesn't produce new viruses. This is known as a latent infection. These latent, hidden viruses then get reactivated at a later date, starting up an active infection that can develop into full-blown AIDS.

Up to 10 in every million immune cells in an HIV patient may harbour a latent infection. Current HIV treatments only work on active viruses, so researchers have been designing drugs that can 'evict' the hidden viruses from their DNA home and make them vulnerable to therapy. In theory, HIV can settle into any location within the genome, and it should be equally easy to remove it from any place. But none of these treatments can shift all the latent viruses, so HIV infection is still impossible to fully cure.

Now Guillaume Filion and his team at the CRG have developed a way of tracking down the exact location of latent HIV in the genome, revealing important clues about why some are harder to reactivate than others.

"We have known since the 1930s that some genes are more active than others, depending on their location in the genome," Filion explains. "So it makes sense that the same would be true for HIV – that it would be easier to integrate or reactivate from certain locations compared with others."

When it comes to deciding where to live, location is everything. If you're moving to an exciting, vibrant city you'd probably choose an area with bustling cafes, bars and shops, rather than a dead zone where everything is closed all the time. And, as Filion and his team have found, HIV has the same idea when it moves into human cells

CRACKING THE BARCODE

One major problem with studying latent HIV infection is that a single cell can have multiple viruses embedded in many places within the genome. This makes it hard to identify the location of each specific virus and work out which ones are reactivated or remain hidden after treatment.

The trick to Filion's new method lies in a genetic 'barcode' – a unique sequence of DNA that the researchers embed in the genetic code of individual virus particles. Once the viruses have inserted themselves into the genome of the host cell (in this case human immune cells grown in the lab) the researchers then extract the cellular DNA and sequence it to find the locations of the latent viruses. And because each virus has its own unique barcode, it's easy to see where each one has embedded. The barcode can also be used to see whether a particular virus has become active and is being 'read', in a similar way to the host cell's regular genes.

"We have been able to use our technique, which we call B-HIVE, to make a map of where HIV likes to go," Filion says. "It is very biased and prefers active regions of the genome – it wants to go where the action is!"

Even amongst the active regions of the genome, the researchers found that some particular areas in the genome are much more desirable than others. For example, some 'addresses' are 100 times more likely to be home to a virus than another, although why they are quite so popular remains a mystery.

As might be expected, the exact location of HIV in the genome affected whether the virus was likely to become active again or not. Viruses that settled down near the 'control switches' responsible for turning on genes were much more active than those in other places. Even so, as Filion points out, there is still a lot of unexplained variability.

"We assumed that the insertion site in the genome would be the main determinant of HIV activity, but we have seen the virus go into the same place in different cells and then behave differently. So there is something else going on that we don't yet understand."

MOVING ON

As well as spotting patterns in the places that HIV likes to live in the genome, Filion and his colleagues also noticed key differences in how viruses in specific locations responded to two reactivation drugs, vorinostat (VOR) and phytohaemagglutinin (PHA), which each have different modes of action. Intriguingly, they found that viruses in one set of genomic 'addresses' were more likely to respond to VOR, while those that had inserted in other locations were preferentially reactivated with PHA.

This suggests that any treatment aimed at treating latent infections might need to combine a cocktail of several different drugs with separate actions in order to flush out all the dormant viruses. And as genetic techniques improve, it might be possible in the future to work out the best combination of reactivation drugs based the different types of viral locations within an individual patient's immune cells, ensuring that all the latent viruses are evicted from every part of the genome.

For Filion and his team, the next step is to see whether they can apply their barcoding technique to label latent HIV DNA within immune cells in animals or even taken directly from infected patients, rather than cells grown in the lab. There are hopes that new gene editing technology, known as CRISPR, might make this possible, although there's a lot more work still to be done.

"The endgame here is not that there will be one drug that can cure HIV, and it may never be possible to reactivate and get rid of every single virus in the body," says Filion. "More importantly, our work helps us to understand the complex relationship between the virus and the host genome – each one is a very small needle in a very big haystack, but we can now find exactly where they are hiding."

REFERENCE WORK:
Chen HC, Martinez JP, Zorita E, Meyerhans A, Filion GJ.
"Position effects influence HIV latency reversal."
Nat Struct Mol Biol, 24(1):47-54. doi: 10.1038/nsmb.3328. Epub 2016 Nov 21.

Swimming in a sea of viruses

A new technique for sifting tiny viruses from seawater could help the hunt for human pathogens.

Looking out of his laboratory window at the CRG, Òscar Fornas can see the sea. Far out on the blue waves, sailors are busily hauling their catch onto the decks of fishing boats, packed full of fish that are too large to slip through the holes in their nets. The smaller the holes in the nets, the smaller the creatures that can be trapped inside. But how small a net do you need to trap something as tiny as a microscopic virus?

In order to understand the world around us, biologists are interested in finding out how animals, plants and microbes – including bacteria and viruses – interact to create an ecosystem. And the oceans are the most fascinating and important ecosystems on the planet, with more significance for the earth's climate than the rainforests.

"Most of the biological diversity on the planet lives in the sea," says Fornas, who leads the Flow Cytometry Unit shared between the CRG and Pompeu Fabra University. "Seawater organisms such as plankton and bacteria fix around half the carbon dioxide in the atmosphere and generate huge amounts of oxygen. But many viruses infect them and could potentially destroy them."

In fact, the ocean is teeming with viruses - just 1 millilitre of seawater can contain about 10 million viruses - but estimates suggest we only know the identity of around one per cent of them.

HUNT THE VIRUS

As DNA sequencing has become faster and cheaper, researchers are increasingly capturing genetic data to understand more about the diversity of species living in an ecosystem. But while it's relatively easy to collect DNA from individual animals or plants, it's much more difficult to pick out single microscopic bacteria or even smaller viruses.

The most popular technique for analysing microbial genomes in the wild is known as 'meta-genomics'. This involves taking a sample from the environment, such as a scoop of soil or a cup of water, and purify-

ing and sequencing a mixture of DNA from all the microbes living there. Clever computational techniques are then used to separate out the genomes of individual species, but virus genomes are extremely small and are likely to be lost amongst all the rest of the data from larger organisms.

Working together with colleagues in Barcelona, Alicante and the US, Fornas wanted to find a way of separating single viruses from this mixture of microbes so each one could be sequenced individually. This approach has been successful with single cells from animal tissues or tumours and bacteria from many different environments. But it had never been done with something as tiny as a virus – around a thousand times smaller than a typical human cell.

SCALING DOWN

To sift the viruses from the saltwater Fornas used a popular lab technique called fluorescence activated cell sorting (FACS), which is often used to separate individual cells from a mixed population. Cells are usually labelled with a fluorescent dye and fed into a machine where carefully designed liquid currents send them one by one through a laser beam for analysis. The cells are then automatically sorted into different multi-well plates for subsequent testing.

Adjusting the FACS machine to cope with something as small as a virus was a big challenge. Although Fornas and his team could label the viruses using a dye that stains viral DNA, the fluorescence is very faint because the genome is so small. There are also many other tiny particles in seawater, from specks of trash to microscopic blobs extruded from larger cells, which can confuse the sorting machine. And the whole process has to be kept free from contaminants that might damage DNA and make it impossible to sequence.

Finding the right settings took many months of testing and troubleshooting. And there was also the arduous day-long cleaning process that had to be performed before every experiment, making sure that every scrap of potentially contaminating DNA was gone. Luckily, Fornas and his team had access to a handy source of ocean viruses, using samples from the Mediterranean lapping outside the lab to optimise the technique.

"Seawater is like soup," says Fornas. "We were working close to the limits of our technology. But we adjusted the lasers, slowed down the flow in the machine, and kept on running tests until we were able to successfully separate individual viruses."

FROM SEA TO SALIVA

As well as his local samples, Fornas also isolated viruses from seawater collected from the sea surface and as far as 4 km down in the depths of the Mediterranean and Atlantic. In total, the team sifted more than 2,000 virus-sized particles yielding 392 rough viral genome sequences. 44 were sent for further more detailed sequencing, and all of them turned out to be viruses that were previously completely unknown to science.

Furthermore, viruses that turn up many times in the collection process are likely to be more common, providing a rough readout of the relative abundance of different species in that part of the ocean. Although finding so many new viruses is impressive, Fornas sees this project as a proof of concept.

"This project showed we can pull together FACS and genomics for viruses," he explains. "We now have a tool that we use to identify new viruses in other ecosystems such as swimming pools, lakes, drinking water and even body fluids – we have already shown that we can use our method to find viruses in saliva. We don't know what viruses are out there that could be harmful to humans, but now we have a tool to find them."

REFERENCE WORK:
Francisco Martinez-Hernandez, Oscar Fornas, Monica Lluesma Gomez, Benjamin Bolduc, Maria Jose de la Cruz Peña, Joaquín Martínez Martínez, Josefa Anton, Josep M. Gasol, Riccardo Rosselli, Francisco Rodriguez-Valera, Matthew B. Sullivan, Silvia G. Acinas & Manuel Martinez-Garcia
"Single-virus genomics reveals hidden cosmopolitan and abundant viruses"
Nature Communications, 8:15892 (2017), doi:10.1038/ncomms15892

Deep freeze

A new freezing method, which can be used anywhere in the world, preserves single cells for scientific analysis.

Here's a challenge: grab a handful of brightly coloured jelly beans, put them all in your mouth and chew. Then try and figure out all the individual flavours. Although you might be able to spot distinctive cherry or tangy lemon, you would probably struggle to identify every single taste. But pop them into your mouth one at a time and each flavour is easy to distinguish.

This scenario is very similar to the problem experienced by researchers investigating the changes in gene activity that happen in individual cells during the development of diseases such as Alzheimer's and cancer. Previously, scientists had to mash up tissue samples containing many thousands or millions of cells and look at the average overall result – just like eating a whole handful of jelly beans at once.

Thanks to improvements in technology, researchers can now look at gene activity patterns in single cells taken from healthy or abnormal tissue and get a true readout of its individual 'flavour'. But getting the right kind of samples for single cell analysis isn't easy.

"We need to start with living cells extracted from fresh material," explains Holger Heyn, leader of the Single Cell Genomics team at the CNAG-CRG. "Here in Barcelona we are right next to the hospital and have all the machines we need to separate the cells. But this simply isn't possible for many medical or research facilities."

Instead, tissue samples taken from a patient are often preserved with formaldehyde so they can be sent off for analysis elsewhere. But this preservative effectively glues all the cells together so they can't be separated. Alternatively, samples can be snap-frozen with dry ice or liquid nitrogen, although this damages the cells so much that they disintegrate upon thawing. And without single cells, researchers can't do single cell analysis.

So Heyn decided to develop an alternative method that could be used to preserve samples gathered from anywhere in the world while still allowing single cells to be separated out at a later date.

SAMPLE AND CHILL

To develop their new method, Heyn and his team took his inspiration from cryopreservation. This technique is usually used to preserve living cells and tissues such as human eggs or embryos in IVF clinics, although some adventurous people choose to cryopreserve their whole body or brain after death (but although there is a good success rate from thawing IVF embryos there are no examples of cryopreserved humans being brought back to life!)

During cryopreservation, the tissue sample is mixed with a special solution containing a gentle preservative chemical called DMSO, along with a protein-rich serum derived from blood, and is then slowly cooled down to 80-˚C in a laboratory freezer or 200- ˚C in liquid nitrogen. The first cooling steps can even be done in a coolbox or portable chiller, making it possible to gather samples from far-flung locations outside a hospital setting.

Once safely frozen, the cryopreserved samples can be stored for at least six months, waiting to be thawed, minced and broken down with enzymes to release single cells for analysis. After testing the technique with cells grown in the lab, blood, bowel and cancer samples, the researchers found that although some of the cells are damaged and lost, a significant proportion survive intact.

After perfecting their technique, Heyn and his team can now recover around 90 per cent of cryopreserved blood cells, although this figure is lower for solid tissues such as tumours. Importantly, they have proved that the freezing process doesn't affect the patterns of gene activity in cryopreserved cells compared with fresh tissue, opening up the possibilities of single cell analysis to research teams who don't have direct access to such complex facilities.

"Our method is very cheap and easy, and you don't need anything fancy," explains Heyn. "We didn't expect that it would work this well – there are no signs of 'shock' from the freezing and we are pretty confident we are getting a representative sample of the tissue back as it was in life."

Because the cryopreservation technique is so simple, many hospitals are now switching to collect samples in this way in the hope that they can set up a collaboration with Heyn or another single cell genomics lab in the future. For example, doctors can gather cells from a tumour at the point of diagnosis, then take more samples as the disease responds to treatment or develops resistance and relapses, using single cell analysis to map the detailed changes across all the genetically distinct pockets of cells that make up a tumour.

Furthermore, researchers working in developing countries now have a way of gathering samples for research projects involving single cell analysis, which would have been impossible in the past.

"All samples are complex and composed of different cell types," Heyn says. "But this fine-grained analysis on a single cell level enables us to generate new knowledge about what's really going on inside tissues, how they are composed and how they are functioning. Our cryopreservation method is going to open up a whole world of single cell samples – it's a real gamechanger."

REFERENCE WORK:
Amy Guillaumet-Adkins†, Gustavo Rodríguez-Esteban†, Elisabetta Mereu†, Maria Mendez-Lago, Diego A. Jaitin, Alberto Villanueva, August Vidal, Alex Martinez-Marti, Enriqueta Felip, Ana Vivancos, Hadas Keren-Shaul, Simon Heath, Marta Gut, Ido Amit, Ivo Gut and Holger Heyn
"Single-cell transcriptome conservation in cryopreserved cells and tissues"
Genome Biology, 18:45, (2017), https://doi.org/10.1186/s13059-017-1171-9

Pass it on

Researchers have discovered an unusual epigenetic heirloom that can be passed down more than 14 generations.

Many families have heirlooms – special items that are passed down the generations, transferring precious memories of bygone times. For tiny nematode worms, this treasure takes the form of chemical marks in the genome, transmitting information about what life was like in the past. Impressively, these cellular memories can be passed down for at least 14 generations – although because worms live, breed and die in the space of a few days, that's still only a few months.

The discovery, published by CRG group leader Ben Lehner in collaboration with researchers from the Josep Carreras Leukaemia Research Institute (IJC) and the Institute for Health Science Research Germans Trias i Pujol (IGTP), was initially made by accident. His PhD student Adam Klosin was studying *C. elegans* worms carrying a transgene array – a long string of repeated copies of a gene encoding a red fluorescent protein – when he noticed something strange.

If the worms were kept at 20˚C, the array of transgenes was less active, creating only a small amount of fluorescent protein. But shifting the animals to a warmer climate of 25˚C significantly increased the activity of the transgenes, making the animals glow bright red under ultraviolet light.

THEN THINGS GOT REALLY WEIRD.

When these worms were moved back to the cooler temperature, their transgenes were still highly active, suggesting they were somehow retaining the 'memory' of their warmer youth. Intriguingly, the bright fluorescence was passed on to their offspring and onwards for another seven generations living at the cooler temperature, even though the original animals only experienced the higher temperature for a brief time. Amazingly, keeping worms at 25 degrees for five generations led to the increased transgene activity being maintained for at least 14 generations once the animals went back to a colder life.

"It's super cool!" Lehner jokes. "This is an artificial system, but the effect is really pronounced. We had to find out what was causing it so Adam abandoned his original PhD project and started working on this instead."

MAKING A MARK

To find out what was causing the strange inheritance pattern, Lehner and his team took a closer look at the transgene array itself, homing in on the ball-shaped proteins (histones) that package DNA inside the cell.

Histones can be modified with chemical 'tags' (epigenetic marks) in a number of different ways. Some epigenetic marks are associated with active genes, while others are linked to gene silencing. In particular, Lehner focused on a histone modification known as H3K9 trimethylation, which helps to shut down gene activity.

As might be expected, the researchers found that the transgenes in animals that had only ever been kept at 20 degrees had high levels of H3K9 trimethylation. Correspondingly, their transgenes were less active and they didn't fluoresce very much. Worms that were then moved to 25 degrees lost the tags, switched on their transgene array and began to glow.

Surprisingly, these brightly fluorescent animals raised in the hotter climate still maintained this reduced histone methylation when they were moved back to the cooler temperature, suggesting that it is playing an important role in locking the memory about the environmental temperature into the genome.

Digging deeper, Lehner and his team found that a protein called SET-25 is responsible for maintaining the histone methylation patterns on the transgene arrays. But they still don't know for sure exactly how the increase in temperature leads to the loss of histone methylation marks. And they also don't know whether the histone methylation patterns themselves are responsible for transmitting the temperature memory down the generations, although they can be seen in eggs and sperm and are present at the earliest stages of worm development.

DOWN THE GENERATIONS

The fluorescent transgene array was put into the worms using genetic engineering techniques, so it might be expected that it acts strangely. But Lehner and his team also found that repeated parts of the normal worm genome that look similar to transgene arrays also behave in a similar way, suggesting that this is a potentially widespread phenomenon and not just restricted to artificially engineered genes.

"This isn't entirely surprising," Lehner says. "There are other repetitive elements in the worm germline that change their activity depending on the temperature, and we do seem to detect a whisper of inheritance a few generations later. But so far we haven't found any 'regular' genes that behave like this."

Although this phenomenon of epigenetic inheritance has been seen in a range of animals, including mammals, the evidence for long-term effects is lacking. Even the best examples, such as the impact of starvation during pregnancy, fade after a couple of generations. This makes Lehner's worms the longest-running example of transgenerational environmental 'memory' ever observed in animals to date. But while it's an intriguing result, it's still not clear exactly how it might be useful to the worms themselves.

"We don't know exactly why this happens, but it might be a form of biological forward-planning," he explains. "Worms are very short-lived, so perhaps they are transmitting memories of past conditions to help their descendants predict what their environment might be like in the future."

Ten generations for a worm is still less than a couple of months. We can predict fairly accurately what the temperature might be like next fortnight, so it makes sense for the worms to try and encode that information to help their great-great-great-great-great-great-great-great grandchildren (or, more accurately, grandworms) prepare for the environment they might hatch into. But it's almost impossible to predict what the climate will be like after that many human lifetimes, so this kind of mechanism probably wouldn't be useful for longer-lived species.

"At the moment this is all speculation," Lehner says. "But biology is so weird that if something like this happens, it probably has been exploited for a purpose somewhere out there in nature."

REFERENCE WORK:

Adam Klosin, Eduard Casas, Cristina Hidalgo-Carcedo, Tanya Vavouri, Ben Lehner

"Transgenerational transmission of environmental information in C. elegans"

Science, 356(6335):320-323 (2017), doi: 10.1126/science.aah6412

Sorting it out

It's your friend's birthday and you want to send them a gift. You wrap it carefully, write their address on the outside, stick on a stamp and put it in the mail. Then, as if by magic, it arrives at their house a couple of days later.

The magic happens in the postal service sorting office as mail workers recognise the stamp and the address, then send the package out with the delivery vans to the correct location. And almost exactly the same thing happens inside our cells on a much smaller scale, as they package up and send out parcels of proteins such as enzymes or hormones out into the body.

This secretory pathway, as it's known, has been studied in great detail over the years. We now know that these secreted proteins are processed in a kind of cellular 'factory' known as the endoplasmic reticulum (ER) then sent to a structure called the Golgi body, where they are modified and packaged. The proteins are sent the right way thanks to molecular 'stamps' and 'addresses' – short regions within a secreted protein that mark it for export.

"Most proteins go through the endoplasmic reticulum and Golgi body for secretion – it's a route we know very well," explains CRG group leader Vivek Malhotra "But there are other secreted proteins that don't make this journey, and we don't know the 'stamps' and the 'sorting offices' that send them on their way."

AN UNCONVENTIONAL JOURNEY

The story starts back in 2007, when Malhotra and his team noticed that yeast and slime mould cells secrete a protein called Acb1 when they are starving. But, strangely, Acb1 lacks any of the characteristic signals that should send it through the usual secretory pathway.

Instead it is gripped by special carrier proteins and taken to a temporary 'sorting office'. This is the Compartment for Unconventional Protein Secretion (CUPS), which is built from components borrowed from the ER, Golgi body and small packages known as endosomes, and only appears under stressful conditions.

Malhotra and his team were curious about whether there were any other proteins that took this unorthodox route out of the cell, and decided to take a closer look at superoxide dismutase 1 (SOD1) – a protein that usually protects us by mopping up toxic chemicals in the body. Like Acb1, SOD1 doesn't have the usual export 'stamp' that directs it through the ER and Golgi body, yet it is still secreted from cells.

SOD1 has been implicated in the neurodegenerative condition amyotrophic lateral sclerosis (ALS, also known as Lou Gehrig's or motor neuron disease). This is a rapidly progressing and incurable disease which kills nerve cells that trigger movement (motor neurons) and eventually leads to paralysis and death, and around a fifth of people with ALS have an inherited fault in the gene encoding the SOD1 protein. Researchers think that faulty SOD1 is secreted by neighbouring cells and taken up by motor neurons, destroying the precious nerve cells instead of protecting them.

Because of this crucial role for SOD1 in ALS, Malhotra and his team wanted to find out whether it also uses the CUPS route to get out of cells, and also to discover the identity of the biological 'stamp' that sends it there.

STARVING AND SORTING

To keep things simple, the scientists started their search in yeast cells, which have very similar secretory pathways to human cells but are easier to grow and study in the lab. They noticed that when they grew the yeast under nutrient-rich conditions the cells secreted a little bit of SOD1. But when the cells were starved of nutrients, they exported nine times the amount.

Next, Malhotra and his team used genetic engineering to change certain molecular building blocks (amino acids) in SOD1, focusing on a region that is the same in both the yeast and the human versions of the protein. They discovered that a sequence of just two amino acids was enough to act as a 'stamp' to send the protein to the CUPS pathway. And, crucially, the same two amino acids are also found in the other unconventionally secreted protein Acb1, suggesting that this might be a universal signal for CUPS.

Finally, to see whether this same pathway might be at work during the development of ALS, the researchers tested healthy and ALS versions of SOD1 in human cells and found that it also is exported through the CUPS pathway when the cells are starved of nutrients.

Putting this all together, Malhotra believes that this work proves that both healthy and faulty versions of SOD1 are secreted from starving cells through the CUPS pathway, and that the little two amino-acid 'stamp' is enough to send them there. But there is still a mystery that needs to be solved.

"Many proteins have the same two amino-acid motif – in fact, it is extremely common," he says. "We still need to find out how SOD1 and proteins like it are specifically recognised and sent to the CUPS, while other proteins are not."

Malhotra thinks that the two-piece 'stamp' is normally hidden in proteins like SOD1 and Acb1 under normal conditions. But when something changes – for example, the protein is faulty or the cell is starving, which affects the shape of proteins – then it becomes exposed. Molecular 'chaperones' then step in to prevent any further unravelling, and instead send the protein off to the CUPS to be secreted out of the cell.

The identities of these chaperones and the exact ways in which they shuttle proteins into the CUPS are currently unknown, but Malhotra and his team are busy tracking them down. They are particularly interested in finding out what triggers harmful SOD1 secretion in ALS patients and – more importantly – working out if they can stop it.

"The discovery of this unconventional 'stamp' directing the secretion of SOD1 is very exciting," he says. "At the moment there are no effective treatments for ALS, so I hope that our findings might provide the basis for the development of much-needed therapies in the future."

REFERENCE WORK:
Cruz-Garcia D, Brouwers N, Duran JM, Mora G, Curwin AJ, Malhotra V.
"A diacidic motif determines unconventional secretion of wild-type and ALS-linked mutant SOD1."
J Cell Biol, pii: jcb.201704056. doi: 10.1083/jcb.201704056. [Epub ahead of print] (2017).

OK Computer

A new 'virtual computer' makes biological data-crunching more reliable.

Imagine you're standing in front of a huge box containing hundreds of different calculators. You take the first one and type in "2+2=" and you get the correct answer: 4. You do the same thing with the next machine and get the same answer, and the next. You carry on testing all the devices, getting the same result every time. But then you get an answer you weren't expecting: 5. With such a simple calculation it's easy to see that something has gone wrong. If the input was correct, then the processor inside the machine must have made a mistake.

Now imagine you're doing the same calculation with multiple enormous sets of genomic data, crunching millions of bits of information together with a computer. You get one answer from a huge Linux supercomputer in the basement of a research institute, but a slightly different one from a cloud-based server, and yet another solution from a Mac. So why are they different, and how do you know which is correct?

BIG DATA, BIG PROBLEM

Scientists who work with this kind of big data are searching for clues to could prevent or treat human diseases or shed light on fundamental biological processes, so they need answers that are reliable and reproducible. This is particularly important in the new era of precision medicine, where doctors make decisions about what treatment a patient should receive based on genetic information.

"Biology is getting more and more computational," says Cedric Notredame, group leader at the CRG. "Twenty years ago it was very expensive to do DNA sequencing, so there was very little data. You could look at a sequence on a piece of paper and analyse it by hand. Now it is so much cheaper and faster – there is much more data so we have to use computers to analyse it."

But there are different software programmes running on different computers with different operating systems, and they do not always give the same answers from the same data. And because there are so many operations and data points involved in these large-scale calculations, it's impossible to figure

out what's gone wrong and how to fix it. What's more, says Notredame, many people have not even realised that this so-called computational instability is even a problem.

"It was an epiphany when we realised there was so much computational instability – this was previously not known at all," says Notredame. "It is a problem because we are moving to an era where drugs and diagnostics are based on genetic data – a computer will spit out a number and tell you your risk of a disease or which drug to take. But it is all based on ranking and probability, so even a tiny variation in output could have a dramatic impact on patients."

Some tech companies have tried to solve this reproducibility problem by building expensive bespoke data pipelines, which lock users into that particular software platform. But Notredame and his team took a simpler approach.

"We built an analysis platform using a technique called virtualisation, which effectively creates a simulated, identical virtual computer inside any machine," he explains. "It's exactly the same idea as the old 80s arcade game simulators that you can run in your PC, but on a much bigger scale."

This 'computer within a computer' means that researchers can run any piece of software inside the virtual environment and get the same result, because their data will always be processed in the same way regardless of the physical machine that they are using.

"We're a small research group, so we needed to build a simple solution that could be easily used by everyone. And we couldn't redesign all the software tools that we have – we wanted to keep running the programmes we're used to using," says Notredame. "Our solution is simple and cost-effective because we did it to solve our own needs."

NEXTFLOW TO THE RESCUE

Following the publication of the paper describing the new virtual platform, known as NextFlow, Notredame and his team decided to make it freely available for others to use. Thousands of researchers are downloading the system every month and many research organisations have adopted it, including the Pasteur Institute in France, the Sanger Institute in the UK, Sweden's National Genomics Infrastructure, the Genome Institute of Singapore and the US National Institutes of Health.

A large international online community has also sprung up to pool ideas and share tools, supported by training workshops and hackathons held at the CRG, pushing the boundaries of what NextFlow can do.

"I love this technology because it is useful, but it's more important that it solves a problem," says Notredame, reflecting on NextFlow's success. "Computational instability is a widespread issue, but there was no solution and you can't correct for it. It's very exciting to know that we have solved a problem that people had not even realised existed yet, but which could have become huge as we enter the Big Data era."

REFERENCE WORK:
Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. *"Nextflow enables reproducible computational workflows."* Nat Biotechnol, 35(4):316-319 (2017). doi: 10.1038/nbt.3820. No abstract available.

Unpacking the genome

Every human cell contains more than two metres of twisted, tightly packed DNA, so switching on the right genes at the right time is a major challenge.

Virtually all the cells in your body share the same set of genetic instructions – around 20,000 genes, encoded in long strands of DNA called chromosomes. But all your cells are not the same. Different types of cells need to use specific sets of genes so they can carry out their particular functions in the body. For example, a liver cell needs to activate genes encoding digestive enzymes and switch off the instructions for making neurotransmitters, while a brain cell has to do the opposite.

What's more, the DNA in every human cell is more than two metres long. It is coiled, twisted and stuffed into the nucleus – a structure smaller than the point of a pin – along with a multitude of proteins. Somehow, in amongst all this molecular confusion, the cell must find and activate the right genes at the right time.

This arrangement of DNA in the nucleus is similar to a tangled ball of knitting yarn. Some parts are tightly squeezed together, while others are loosely packed. Finding and activating a specific gene is like hunting for a specific short stretch of yarn in amongst the tangled mess, releasing it from any tight clusters and loosening the thread so it can be used.

It's already been shown that active genes tend to be in more loosely packed, 'open' compartments of the nucleus compared with inactive genes, but little is known about how genes are organised into these different regions or how their location changes when they are switched on and off.

Understanding how this works at a molecular level is one of the most important challenges in biology, and it's one that CRG senior group leader Thomas Graf wants to solve.

INTO THE FOURTH DIMENSION

The story starts in 2014 when Graf and his colleagues at the CRG – Miguel Beato, Guillaume Filion and Marc A. Marti-Renom – began a major collaboration known as the 4D Genome (ERC Synergy Grant project), investigating how the organisation of DNA changes as genes are switched on or off.

Not only has the team been mapping the organisation of DNA in the nucleus of 'resting' cells, the researchers have also been developing ground-breaking new techniques to track changes in the three-dimensional structure of chromosomes in the nucleus over time (the fourth dimension) as cells shift from one type to another, whether temporarily or permanently.

This kind of transition is seen in development, as multipurpose embryonic stem cells gradually become specialised into particular tissues in the developing embryo or fetus. But this time Graf was particularly curious to see what happens when specialised cells 'reverse' back into stem cells – a process known as reprogramming.

"People had already compared nuclear organisation in specialised cells and stem cells, but they did not know how these changes occur over time," says Graf, "We wanted to catch them in action, asking whether the organisation of the genome changes before or after genes are switched on during reprogramming."

To wind the clock back, the CRG researchers used a variation on a technique developed by Nobel prize-winning Japanese scientists Shinya Yamanaka, who discovered that a cocktail of four proteins (OCT4, SOX2, KLF4 and MYC) could turn specialised cells back into stem cells. These impressive molecules are transcription factors, which bind to particular sites in the DNA close to the start of stem cell-specific genes and switch them on, reprogramming the cell back into a stem cell state.

Unfortunately, the method isn't very efficient for many cell types. For example, only a very small fraction of immune B cells can be reprogrammed with these so-called Yamanaka factors. However, Graf and his team discovered that adding in another protein, known as C/EBP alpha, before the Yamanaka factors led to at least 95 per cent of B cells being converted back into stem cells over the course of eight days.

By taking samples of these cells every two days, the researchers could use their 4-D techniques to follow the changing organisation of DNA in the cells' nuclei as they converted from B cells into stem cells.

UNTANGLING THE DATA

To find out how genes are re-arranged within the nucleus of the cells during reprogramming, Graf and his team used a method called Hi-C. This reveals whether specific regions of DNA are touching each other and reflects how loosely or tightly packed they are.

The team also gathered data on whether certain genes were switched on or off, as well as cataloguing the molecular marks (known as epigenetic modifications) that are associated with active or inactive genes. Much of the practical work was done by postdoctoral fellow Ralph Stadhouders, together with computational biologist Enrique Vidal.

The key to the project's success was a new piece of software developed by Marti-Renom and his team, known as TADbit. It's a bit like a 'Google Earth' for the nucleus, bringing together all the data to build a detailed map of how the DNA is organised in any part of the genome.

"In some ways generating the data is trivial – analysing it is the hard part and takes lots of time and computing power," Marti-Renom says. "These experiments generate billions of pieces of data and need hundreds of thousands of hours of computing time, so our new software was absolutely key to make the analysis automatic and user-friendly."

As might be expected, the researchers discovered that most of the genes that are turned on as the B cells become stem cells appear to move into more active compartments of the nucleus. Intriguingly, they found that this happens several days before the genes are actually switched on.

"The prevailing idea was that genes are switched on by the binding of transcription factors, such as the Yamanaka factors, and then they move into an active region of the nucleus," explains Graf. "But we found that many genes moved first then were activated later. This was an unexpected but very exciting finding."

BACK TO THE BEGINNING

Graf believes that these findings reveal a more important role for changing organisation in the nucleus than was previously thought, and also a potentially new function for transcription factors. Not only do they bind to DNA and switch genes on, he explains, but he thinks they may also have a separate, earlier part to play in unpacking the genome and moving genes into active regions of the nucleus.

"Once the transcription factors untangle the DNA and expose the genes, then it is easy to switch them on," Graf says. "But now the big questions are how do they do it, who do they work with, and what is the engine that drives the reorganisation?"

He and his colleagues in the 4D Genome team are now searching for the molecules that work together with transcription factors to untangle and rearrange DNA. And the they also want to try and manipulate these interactions – whether by altering the DNA or by changing the proteins – to unpick the precise relationship between the four-dimensional changes they see in the nucleus and the resulting patterns of gene activity.

"We are learning the principles of cell fate decisions, and what we are seeing in our reprogramming system is a model for the processes that happen in an embryo," Graf says. "I can't wait to find out what is happening during the earliest days of life when the first pluripotent stem cells are born."

REFERENCE WORK:
Ralph Stadhouders, Enrique Vidal, François Serra, Bruno Di Stefano, François Le Dily, Javier Quilez, Antonio Gomez, Samuel Collombet, Clara Berenguer, Yasmina Cuartero, Jochen Hecht, Guillaume J. Filion, Miguel Beato, Marc A. Marti-Renom & Thomas Graf
"Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming"
Nature Genetics, 50:238–249 (2018), doi:10.1038/s41588-017-0030-7

Research and Scientific Services

The breadth of topics, approaches and technologies at the CRG allows us to ask a wide range of fundamental questions in life sciences and biomedicine. Research at the CRG falls into four main areas: gene regulation, stem cells and cancer; cell and developmental biology; bioinformatics and genomics; and systems biology. As of July 1, 2015, the National Centre for Genome Analysis (CNAG-CRG) is also part of this research structure.

BIOINFORMATICS AND GENOMICS

Roderic Guigó
Coordinator

The programme's scientific highlights in 2017 included the development of the Capture Long Seq (CLS) methodology to exhaustively characterise the transcript diversity of long non-coding RNAs, the development of nextflow, a domain-specific language that enables scalable and reproducible scientific workflows using software containers, the uncovering of a sexual cycle and a recent association with the human host in the emerging fungal pathogen *Candida glabrata* and the Global Score algorithm to predict protein interactions with large transcripts.

Several groups in the programme are participating in a number of large-scale genomic projects, such as ENCODE, GTEx, PanCancer, I5K, F1K, WebOfLife and IASIS (first grant in Catalonia for personalised medicine).

The programme has continued to deploy and support the European Genome-phenome Archive (EGA) in collaboration with the European Bioinformatics Institute (EBI). EGA has been selected as an ELIXIR Core Data Resource and as an ELIXIR Recommended Deposition Database. It has also been selected as one of the Global Alliance for Genomics and Health (GA4GH) Driver Projects. EGA is also one of the European Open Science Cloud (EOSC) Science pilot demonstrators.

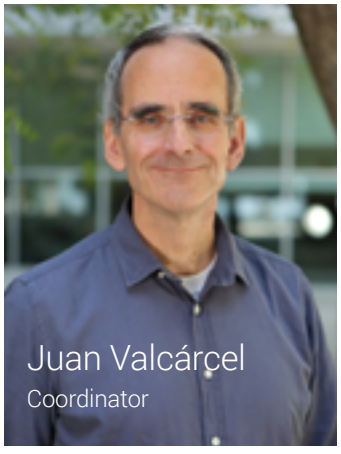
CELL AND DEVELOPMENTAL BIOLOGY

In 2017, Malhotra's lab discovered a di-acidic motif necessary for the secretion of superoxide dismutase (SOD)1, which is secreted without the conventional endoplasmic reticulum-Golgi pathway of protein export. Modulation of this sequence could help in understanding neurotoxicity associated with mutant SOD1 in amyotrophic lateral sclerosis. This lab also discovered that a protein called TANGO1, which is required for the export of bulky collagens, assembles into a ring at ER exit sites, thereby compartmentalizing ER into specific regions for folding and export of bulky collagens from the rest of the ER. Another finding of considerable interest is the demonstration of sphingomyelin in organizing the shape and the function of Golgi membranes. Isabelle Vernos's lab continued to dissect the mechanism of spindle dynamics and discovered a new mitotic partner of a specific kinesin motor and their role in microtubule and chromosome alignment. Sebastian Maurer succeeded in reconstituting RNA migration on microtubules in vitro. This tour de force approach is the first of its kind and a major achievement for understanding the mechanism of kinesin-dependent cargo transport on microtubules. Manuel Mendoza discovered a new mechanism of cell cycle progression via asymmetric division. Solon's lab provided a first description of the evolutionary modulations in the process of epithelial sealing between different fly species. They found an evolutionary transition leading to a simplification of the sealing process in dorsal closure and identified a conserved essential role of the microtubule cytoskeleton in epithelial sealing.

These findings attest to this department's overall interest in addressing the mechanisms underlying cell compartmentation, cell division and tissue organisation.

2017 witnessed the departure of a junior group leader, Manuel Mendoza, who moved to the Institute of Genetics and Molecular and Cellular Biology-IGBMC (Strasbourg, France) as group leader. During his spell at the CRG, Manuel's lab revealed a fascinating mechanism for the modulation of nuclear pore complexes during cell division and the role of this process in controlling differential cell cycle progression in mother and daughter cells. His findings also revealed a challenging connection of exocytosis and abscission and the mechanism by which cells control abscission to ensure normal chromosomal segregation. We are proud of Manuel's achievements and wish him continued success in his new position in France. His collegiality, generosity and scholarly discussions will be sorely missed.

Finally, we would like to welcome two new group leaders. Elvan Boke has joined the department as a Junior Group Leader. She is leading the Oocyte Biology & Cellular Dormancy lab. During her post-doc at Harvard, Elvan discovered that oocytes cluster and segregate mRNA, mitochondria, ER, Golgi, endosomes and lysosomes through a large disordered protein called Xvelo. The Xvelo-mediated creation of this mega sub-compartment called the Balbiani body is a major breakthrough and holds great promise for understanding the basic principles in oocyte development and the overall process of fertilisation. Boke's leading discovery and objectives are recognised by the award of a starting ERC grant. Verena Ruprecht was recruited following her brilliant research at IST in Austria. As a post-doc at IST, Verena discovered how cortical contractile activity leads to rapid amoeboid cell motility. This process of amoeboid migration is witnessed in numerous developmental and cancer metastases, although the mechanism is unclear. Verena's findings opened a brand-new chapter in cell migration and have made this process of fundamental importance amenable to molecular analysis. Verena's expertise in physics and imaging brings new approaches to addressing cell migration, and we look forward to many new discoveries.



GENE REGULATION, STEM CELLS AND CANCER

Research highlights in the programme in 2017 include important insights into the role of the three-dimensional arrangement of chromatin in regulating gene expression. Members of the ERC 4D Genome Synergy project reported the fundamental roles of local genomic context in the latency of HIV virus, the expression of genes in Drosophila cells and how genome topology influences – and is influenced by – the activity of transcription factors during cell reprogramming. Additional work by other groups in the programme unveiled transcriptional networks with key roles in the regulation of embryonic stem cell identity, somatic cell reprogramming and organ regeneration by cell fusion. These processes are also being characterised using super-resolution microscopy, supported by FET grants. Progress was also made in understanding molecular mechanisms of anti-tumour drugs targeting the splicing machinery and efforts to develop novel modulators of this process have been supported by laCaixa Impulse and ERC PoC grants. Finally, Bernhard Payer's group initiated a collaboration with the Eugin fertility clinic, also involving CNAG-CRG, to explore the molecular basis of human oocyte ageing.

SYSTEMS BIOLOGY

2017 witnessed the departure of two of our group leaders. Matthieu Louis, a junior group leader, moved to the University of California at Santa Barbara, and James Sharpe, a senior group leader, became the first director of EMBL Barcelona. During their time at the CRG, Matthieu's lab completed an absolute tour de force of technology development to establish the Drosophila larva as a premier system in which to quantitatively study sensory perception and animal behaviour. The lab discovered new behaviours, developed methods to study them, elucidated the wiring diagram of the larva olfactory system, studied its evolution, and, through the use of a closed-loop optogenetic tracker, developed and tested a multi-level model for how the animal integrates dynamic olfactory signals.

James' lab has been part of the program since it started in 2006 and James was coordinator of the program from 2011 to 2017 and so contributed enormously to the culture of the program and the style of the science we do. The lab's major achievement whilst at CRG was the demonstration that the mechanism that specifies mammalian digits is a molecular Turing system. A vast amount of technology development underlies this major achievement, and the lab continued to develop microscopes and modelling approaches throughout its time at CRG. They also published a fascinating body of work exploring the design space of dynamic gene networks, and, in collaboration with a former program member, Mark Isalan, built some of these patterning networks in bacteria. We may have five digits, but it turns out there are six mechanisms for three genes to interpret a morphogen gradient to build a stripe! We are very excited that James will direct the new site EMBL Barcelona. The institute will focus on the biology of tissues and organs and will bring an additional 6-7 systems biology labs to the PRBB, giving the building an unrivalled concentration of quantitative and integrative biology labs in Europe.

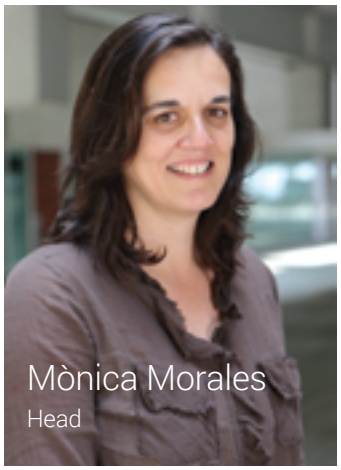
We are very proud of the achievements of both Matthieu's and James' labs - theirs is exactly the kind of original, ambitious, difficult, and long-term quantitative science that we aim to develop in the program. We wish them luck in their new homes and expect to hear great things coming out of their labs for many years to come. We will miss their scientific vision, friendship, and support.

2017 has been quite a productive year for the program. Ben Lehner's lab published their discovery of a long-lasting and chromatin-associated transgenerational epigenetic memory of the environment in *C. elegans*, as well as their discovery that maternal age is a major influence on phenotypic variation in this species. These two studies continue the lab's long-running interest in understanding the causes of phenotypic variation amongst genetically identical individuals. The lab also showed that the signatures of clustered mutations in >1,000 human tumours can be used to identify the molecular mechanisms that cause mutations, including the discovery of a new mutation process that targets mutations to active genes in tumours associated with carcinogen exposure, including alcohol consumption. Luis Serrano's lab continued to develop *Mycoplasma pneumonia* as a 'therapeutic chassis'. They also published the structure of the *Mycoplasma* chromosome at 10 kb resolution and used random mutagenesis and deep-sequencing to determine key sequences of promoter and untranslated regions that influence transcription and translation efficiency in this bacterium. Mara Dierssen's lab continued their work on understanding the changes in neuronal architecture and connectivity that disrupt cortical and hippocampal function in genetic cognitive disorders. They also showed that Neurotrophin-3 infusion rescues fear extinction impairment in a mouse model of pathological fear. Manuel Irimia's lab published the most comprehensive database of alternative splicing events released to date. They also elucidated the role and evolution of Esrp-dependent splicing programs in morphogenesis. In addition, by molecularly characterising the development of the amphioxus neural tube, they presented an important new model for vertebrate brain organization and evolution.

In recognition of their achievements, Manu Irimia was elected as an EMBO Young Investigator, Ben Lehner was elected as an EMBO Member, and Mara Dierssen received both the BigVang Medal and the Trifermed Social Impact of Healthcare Award.

Finally, Nick Stroustrup joined the program as a junior group leader. Nick's Dynamics of Living Systems Lab will develop experimental and computational methods to characterize where, when, and why aging occurs, and how we might effectively intervene in its progression. Whilst at Harvard, Nick

developed the 'Lifespan Machine', which allows researchers to track tens of thousands of nematodes throughout their entire lifespan and he used this to discover a universal scaling law for how interventions alter lifespan. Nick continues the program traditions of hosting groups with an engineering-driven approach and groups tackling a well-established question from a very original (orthogonal) angle. Welcome Nick!



CORE FACILITIES

The Core Facilities programme currently comprises seven Core Units: Genomics, Proteomics, Advanced Light Microscopy, Biomolecular Screening & Protein Technologies, FACS, Bioinformatics, and Tissue Engineering. The programme also comprises the Histology Service and the Storage and Computing Unit that are only accessible to PRBB users, or internal users, respectively.

All units work towards implementing new technologies and applications in response to both our user needs and the future directions in their respective fields. The most prominent new technologies set up in 2017 include:

- Isolation of single virus by flow cytometry (for single-virus genomics and the study of the marine virosphere)
- Identification and isolation of extracellular vesicles by flow cytometry for the study of the vesicle cargo
- Generation of Pseudostratified mucociliary epithelium from normal human bronchial epithelium and of Retina pigmented epithelium from human ES cells
- Derivation and culture of intestine organoids
- New workflow for the elucidation of protein complexes and structural features using chemical crosslinkers
- Protocol for error correction of sequencing reads to allow high sensitivity of mutation detection
- Globin depletion protocol from blood RNA samples during a normal mRNA library prep
- Implementation of most standard bioinformatics pipelines in NextFlow

In order to anticipate future needs in life science research, we are also working towards a further integration of facilities by implementing new cross-facility methodologies that require the collaboration of several units. More particularly, we are setting up genome engineering by CRISPR/Cas9 directly in embryos, deciphering the proteome of extracellular vesicles or generating a full set of *in-house*-produced enzymes for NGS library preparation.

Additionally, we are focusing our efforts on collaboration with the Industry for technology scouting. In 2017, we performed the application-testing of the latest Leica STED objective and organised several scouting events to assess the latest technologies on the market.

The CRG core facilities are not only well-established locally, with users from different institutions in Spain and abroad, but are also recognised **partners in European initiatives**. The Advanced Light Microscopy Unit is a partner in the ESFRI initiative EuroBioimaging (EuBI), and its head, Timo Zimmerman, the national coordinator for biological imaging. The Genomics and Proteomics Units are members of MERIL, the European Research Infrastructure portal listing facilities with more-than-national relevance (CRG being the only Spanish Proteomics Facility).

The Core Facilities are members of the Core Facilities Excellence Alliance **"Core For Life"** (www.coreforlife.eu), as described in the 'National and International Dimension' section.

CNAG-CRG


Ivo Gut
Director

In 2017, CNAG-CRG has further consolidated its position. The inclusion of the CRG Genomics Core Facility has allowed us to reorganise activities and instrumentation with CNAG-CRG focussing on the high-end, high-throughput applications, while the CRG Genomics Core Facility focuses more on applications such as miRNA and ChIP sequencing. Together we have continued our strategic path to offer the best-in-class support to our collaborators for their research projects. Still of particular focus are areas of patient-near research, such as in rare diseases and cancer. From an applications points of view we have extended our expertise in single cell analysis, epigenomics, translational techniques and the integration of population information. We have further developed our advanced analytical capabilities.

This year has seen many highlights. Our quality system running under ISO17025:2005 accreditation with the scope of DNA/RNA analysis by high throughput sequencing (NGS) and ISO9001:2015 certification by the Spanish national accreditation body ENAC has passed its re-accreditation and re-certification with flying colours. Additionally and after a rigorous training program, the CNAG-CRG has become the first European center to obtain Roche's Certified Service Provider program for SeqCap EZ Target Enrichment Systems. We are in the process of adding further accreditations that will facilitate working with the clinical services in particular with a view to personalized medicine.

In 2017, we incorporated a second Illumina HiSeq4000 and retired several of the older sequencing instruments that had reached their end of life. Much effort has gone into the resolution of technical problems inherent to the newer Illumina sequencing instruments that operate with patterned flowcells. An innovative modification has been included in the sequencing protocols that allows handling this problem. On the Oxford Nanopore sequencers we have implemented direct RNA sequencing which opens several new exciting avenues for research projects. The Oxford Nanopore sequencers have turned into a major workhorse in *de novo* assembly projects.

The RD-Connect Genome-Phenome Analysis Platform (GPAP), developed at the CNAG-CRG, was made available to the European International Rare Disease Research Consortium investigators. It already has over 400 users and has been used in many rare disease research projects. It was crucial for the BBMRI-LPC call in which 900 exomes were sequenced. Data was analysed in the RD-Connect GPAP and biological specimens transferred to Eurobiobank. Although RD-Connect is coming to an end, the CNAG-CRG is strengthening its role in the field of rare diseases with upcoming projects such as Solve-RD. In 2017, more than half of the patient samples analysed resulted in disease gene identifications.

The Single Cell Genomics team found a way to cryopreserve biological samples without compromising gene expression profiles compared to freshly processed samples at single cell level. A game-changer result since the use of cryopreservation drastically influences sample accessibility. In addition, the team is taking an active role in the Human Cell Atlas, aimed to build a collection of maps that will describe and define the cellular basis of health and disease.

Personalized medicine is arriving and genome analysis is its major tool, as it provides unprecedented resolution for diagnosing patients. NAGEN1000, a pilot project together with Navarrabiomed funded by the Servicio Navarro de Salud-Osasunbidea and two pilot projects (MedPerCan and URDCat) together with several hospitals in Barcelona funded by the Catalan Department of Health (PERIS) started in 2017. These projects aim to bring genomic analysis into the clinic for the immediate benefit of patients. Moving forward it is clear that CNAG-CRG will play a key role in the implementation of personalized medicine into healthcare. With our sequencing platform, our sophistication in data analysis and our databases to make genomic data more user-friendly we are in a prime position to support this monumental task.

EUROPEAN GENOME-PHENOME ARCHIVE (EGA)

The EGA is a service for permanent archiving and sharing of all types of personally identifiable genetic and phenotypic data resulting from biomedical research projects. The data at EGA was collected from individuals whose consent agreements authorize data release only for specific research use or to bona fide researchers. Strict protocols govern how information is managed, stored and distributed by the EGA project.

Since its launch, researchers from around the world have deposited and accessed data from more than 1000 studies in the EGA of various types. These studies vary from large-scale array-based genotyping experiments on thousands of samples in case-control designs or population based studies, to sequencing-based studies designed to understand changes in the genome, transcriptome or epigenome in both normal tissue and in various diseases such as cancer. As a result, the EGA has grown from about 50 TB to 3500 TB during the last five years.

The EGA team engages in several initiatives and projects, such as securing storage and distribution of human personal identifying data; analysing relationships between genomes and phenomes; evolutionary medicine; it is the driver project of the Global Alliance for Genomics and Health (GA4GH); it is the ELIXIR Core Data Resource of fundamental importance to the life science community and the long-term preservation of biological data, and it is also a major player in EXCELERATE (HORIZON 2020) projects.

ERC Researchers at CRG

STARTING GRANTS


Toni Gabaldón


Manuel Irimia


Fyodor Kondrashov


Manuel Mendoza


Gian Gaetano Tartaglia


Elvan Boke

CONSOLIDATOR GRANTS


Ben Lehner


Toni Gabaldón

ADVANCED GRANTS


Roderic Guigó


Luis Serrano


James Sharpe


Juan Valcárcel

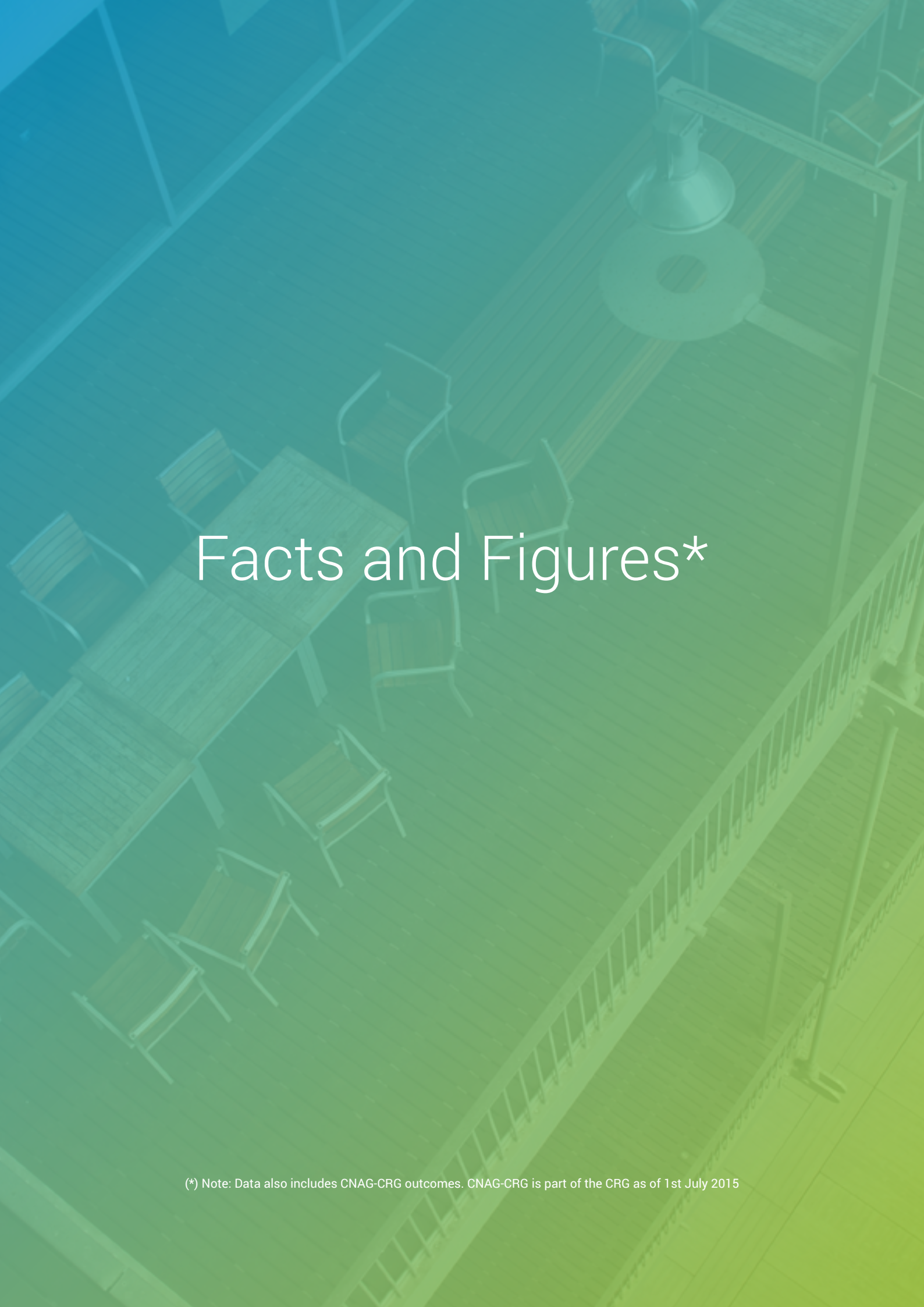
SYNERGY GRANT


Miguel Beato


Thomas Graf


Guillaume Filion


Marc Marti-Renom (CNAG-CRG)



Facts and Figures*

(*) Note: Data also includes CNAG-CRG outcomes. CNAG-CRG is part of the CRG as of 1st July 2015

Publications

225
Total Publications

70%
Open Access Publications

89%
1st Quartile Publications

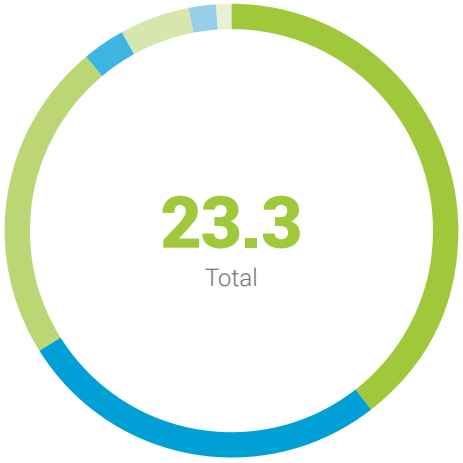
8.8
Average Impact Factor

Funding (M€)

42.6
Total Funding

34.4
CRG

8.2
CNAG



External Funding awarded in 2017

- Public-National: 40% - 9.3
- Public-Europe: 27% - 6.3
- External Sales: 23% - 5.3
- Private-Philanthropic & Sponsorship: 3% - 0.6
- Public-Regional: 5% - 1.1
- Public-International: 2% - 0.4
- Private TBDO: 1% - 0.3

Note: The above graph includes competitive funds obtained during 2017 and pending for final notice of award or grant agreement as of 31/12/2017

Projects

124
Total Ongoing Research Projects and Networks

12
are Ongoing ERC Projects (16 ERC Grantees)

18
are Other Ongoing H2020 Research Projects and Networks

7
are Ongoing EU Coordinated Projects

11
are International Ongoing Research Projects (non-EC)

19
Total Ongoing Postdoctoral Fellowships

7
Total Ongoing EU Coordinated Projects

Staff

TOTAL STAFF

498*
Total
(*FTE, full-time equivalent: 482)

419
CRG

79
CNAG-CRG

RESEARCH STAFF

324
Total

296
CRG

28
CNAG-CRG

SCIENTIFIC SERVICES

98
Total

51
CRG

47
CNAG-CRG

ADMINISTRATION & SCIENTIFIC SUPPORT

76
Total

72
CRG

4
CNAG-CRG

RESEARCH GROUPS
by 31st Dec 2017

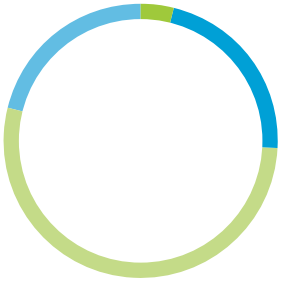
28
Total

26
CRG

2
CNAG-CRG

Research Categories

- Staff Scientists 3%
- Postdocs 41%
- Group Leaders 11%
- Heads of Unit 5%
- PhD Students 40%



Age

- >50 years old 20 (4%)
- 40-50 years old 111 (22%)
- 30-40 years old 263 (53%)
- <30 years old 104 (21%)

INTERNATIONALITY

41
nationalities represented

60%
Group Leaders+Heads of Unit

70%
Postdoctoral Researchers

50%
PhD Students

60%
Total Research Staff

Gender

FEMALE BY PROFESSIONAL CATEGORIES

APPLICANTS TO OUR SELECTION PROCESSES

947 49.5% Women

965 50.5% Men

SELECTED / HIRED CANDIDATES

34 54% Women

29 46% Men

FEMALE INVITED SPEAKERS

36%

an improvement of almost 4% compared to 2016

Events

10
International Events

151
High-profile Seminars

Advanced Training

THESES

23
PhD Theses defended

SCIENTIFIC TRAININGS

11
Internal courses

5
External courses

300
Participants

COURSES@CRG

7
International courses

147
Participants*
*32 different countries

TRANSFERABLE SKILLS & CAREER DEVELOPMENT TRAININGS

11
Internal courses

276
Participants

INNOVATION COURSES

2
Courses

22
Participants

Technology & Business Development

30
New Technologies Identified under Evaluation

PATENTS

2
Priority

1
PCT

3
National Phases

34
Agreements Signed with Industry & Investors

7
Projects Managed under Valorisation

microomics®

1
Spin-offs (Microomics)

Communications, Public Engagement & Science Education

MEDIA RELATIONS

2,567
Media Appearances

272
Written Media

1,985
Online Media

276
Blogs

19
Radio

15
TV

SOCIAL MEDIA

TWITTER FOLLOWERS

8,761
@CRGenomica

1,653
@cnag_eu

FACEBOOK

3,299
Page likes

LINKEDIN FOLLOWERS

3,025
CRG

1,393
CNAG-CRG

YOUTUBE

142,535
Channel views

PUBLIC ENGAGEMENT AND SCIENCE EDUCATION

237
Activities Organised
(39 different categories
of activities)

13,807
Audience Reached

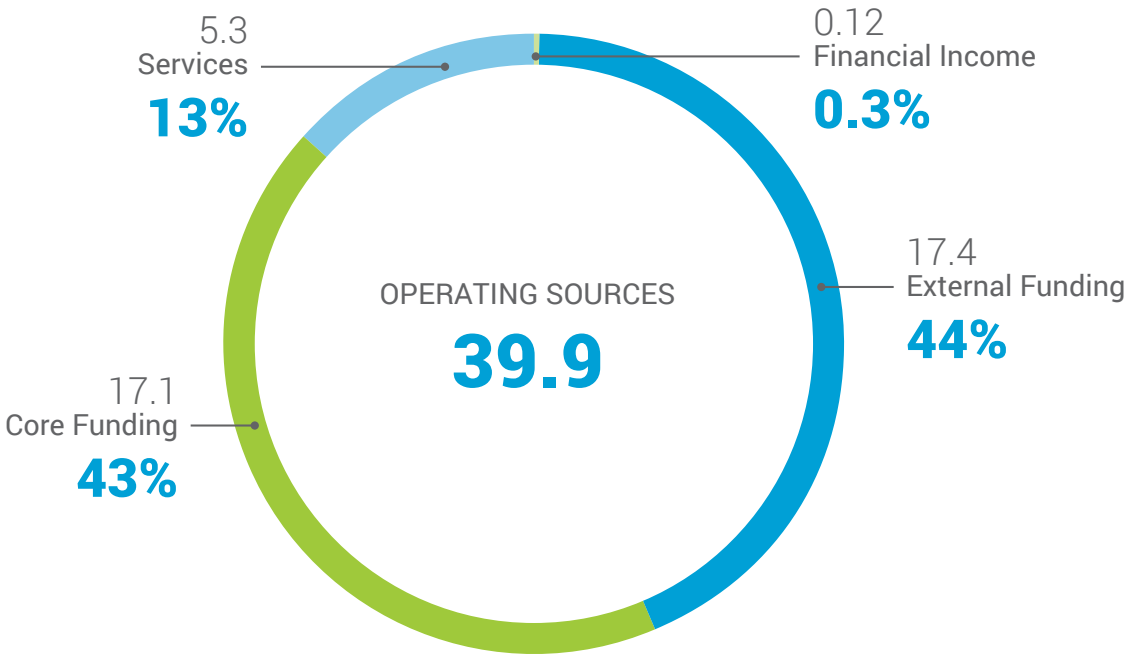
9,176
Schools and Students

4,631
General Public

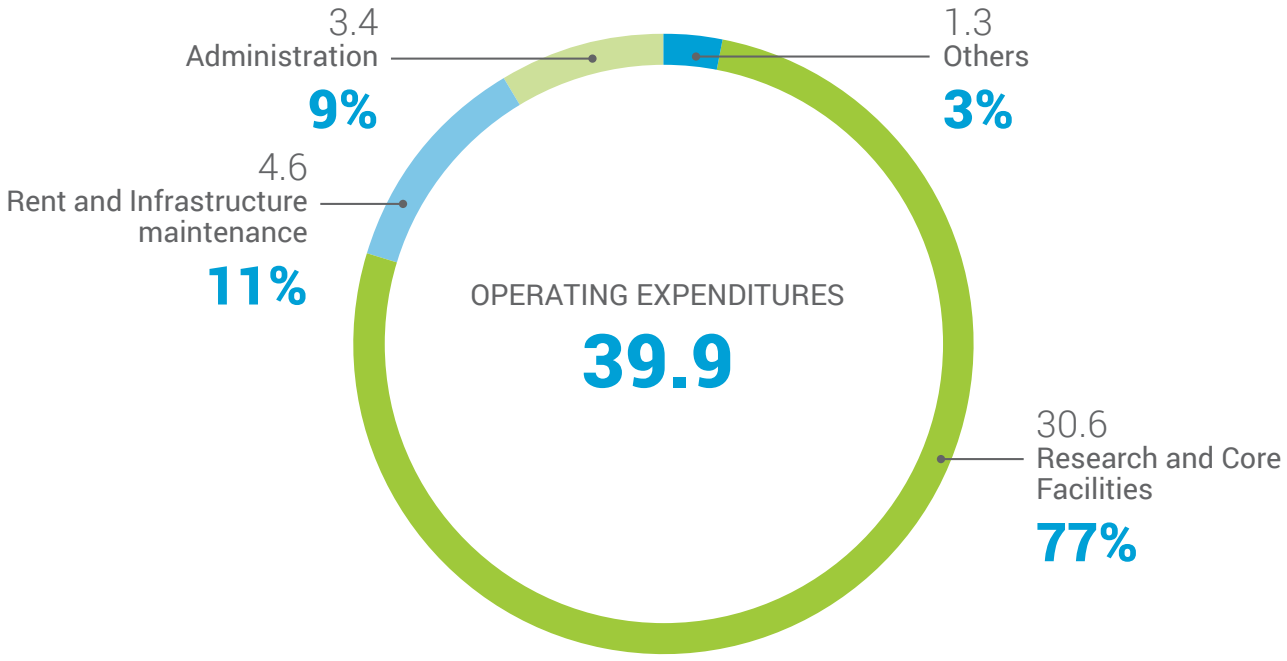
Financial Report

Sources & Uses Managed

OPERATING SOURCES IN M€



OPERATING EXPENDITURES IN M€



Acknowledgements

Support from our trustees, public and private funders and sponsors is key to accomplishing the CRG's mission of discovering and driving knowledge for the benefit of society, public health and economic prosperity.

Trustees

Public Funders

Note: ERDF and ESF funds have been instrumental over the years through different funding schemes and in a variety of activities in supporting our research and keeping our infrastructures state-of-the-art. Further details on the projects co-funded by these funds can be found in the ERDF AND ESF FUNDS AT THE CRG (http://www.crg.eu/en/content/erdf-and-esf-funds-crg) section (CRG website).

Private Funders

OBRA SOCIAL "LA CAIXA"

The "la Caixa" Bank Foundation has supported several key initiatives at the CRG, such as its International PhD Programme, since 2008 and additional scientific and outreach activities since 2014: the partnership between the CRG and the European Bioinformatics Institute (EMBL-EBI) to run the European Genome-phenome Archive (EGA) jointly, and the CRG's first citizen science initiative 'Saca la Lengua' (Stick out your tongue). In the first half of 2016, the Foundation generously decided to fund the second edition of 'Saca la Lengua', which started in October 2016. During 2017, the project had its second tour of Spain, successfully tackling new challenges, reaching new target audiences and collecting samples from different populations and patients suffering from different diseases.

AXA RESEARCH FUND

The "AXA Chair in risk prediction in age-related diseases" was created in 2014 for a 15-year period with a 1-million euro endowment. Dr. Ben Lehner was appointed first chair holder to further his work in the development of personalised medicine to provide people with better protection from the unique risks they face in diseases such as cancer. In 2017, Dr. Bernhard Payer was appointed the second chair holder for a term of 3 years.

NOVARTIS

Novartis engages in extensive collaboration work with the CRG. From 2003 to 2016, the company supported the organisation of the CRG's Annual Symposia, and it also backed an annual fellowship for postdoctoral researchers in the field of genomics between 2004 and 2012. In 2012, a new CRG-Novartis-Africa mobility programme was set up to advance bioinformatics, genetics and genomics research in Africa by mentoring young and promising African scientists. Every year, the programme allows up to four early career researchers from African Universities to spend a 6-month internship at the CRG to carry out their research project under the supervision of a CRG principal investigator.

FUNDACIÓN BOTÍN

The Fundación Botín, through its Science area, and in collaboration with the CRG's Technology and Business Development office, promotes the translation of research results produced in the labs of Dr. Juan Valcárcel (currently) and Dr. Luis Serrano (2007-2013) into the market. They accomplish this by providing economic and management resources to identify promising ideas and results at an early stage, assessing their potential and the best form of protection through intellectual and industrial property rights, while also sourcing the necessary technology and industry partners or investors to help technologies or products to enter the market to the ultimate benefit of society.

FUNDACIÓN RAMÓN ARECES

The Ramón Areces Foundation provided three-year funding for a highly-talented young postdoctoral student to carry out research at the CRG. The successful post-doc, selected from a competitive call, was Xianghua Li, who worked in Dr. Ben Lehner's lab until the first quarter of 2017.

FUNDACIÓ CATALUNYA-LA PEDRERA

The Fundació Catalunya-La Pedrera supports vocational training activities for young and talented students to nurture their interest in science and to pursue a scientific career. Key activities include scientific summer stays at MónNatura Pirineus and at the CRG, where students take part in sessions and events focused on scientific topics with the aim of eventually proposing and developing their own project idea. Since 2016, the CRG has also been one of the institutes hosting students from the Barcelona International Youth Science Challenge (BIYSC), a two-week international excellence summer programme that seeks to stimulate scientific talent among young people from all over the world and to encourage their enthusiasm for pursuing scientific research and a career in science.

FUNDACIÓ MARATO TV3

The Fundació Marató TV3 funds several research projects led by CRG investigators related to different editions of the telethon: three projects from the 2012 edition on 'Cancer' (Thomas Graf, Pia Cosma and Susana de la Luna), two projects from the 2013 edition on 'Neurodegenerative diseases' (Fátima Gebauer and Luciano Di Croce), one project from the 2014 edition on 'Heart disease' (Gian G. Tartaglia) and two projects from the 2016 edition on 'Strokes and traumatic spinal cord and brain injury' (Marc Marti-Renom and Mara Dierssen).

FONDATION JEROME LEJEUNE

The relationship between the CRG and the Jerome Lejeune Foundation began many years ago. They provided support to several of Mara Dierssen's research initiatives linked to the identification of molecular and genetic bases in several pathologies accompanied by mental retardation: Rett Syndrome, Fragile-X Syndrome, William-Beuren Syndrome and Down Syndrome. Dierssen also received the first international Sisley-Jerome Lejeune Award in 2010. In 2016, they awarded a grant to Eduard Sabidó's project on the elucidation of the mechanism of action of epigallocatechin-3-gallate as a therapeutic agent on the cognitive phenotype in Down Syndrome mice models (2015-2017). More recently, in 2017, a new project was awarded to Mara Dierssen, entitled 'EpiGenetic Change Generator in Down Syndrome (2017-2019).

AECC

The Spanish Association Against Cancer (AECC) has supported a number of research projects and initiatives by CRG scientists over the years. In 2015, Pedro Vizán (in Luciano Di Croce's lab) was awarded the AECC Oncologic Research Fellowship for a project that seeks to identify and "attack" stem cells involved in cancer, which will end in 2019.

ZIMIN FOUNDATION

Thanks to the Zimin Foundation, the School of Molecular and Theoretical Biology (SMTB), organised by our researcher, Fyodor Kondrashov, was held in Barcelona for two years in a row (2016 and 2017). The STMB brought together 80 intellectually inquisitive and talented secondary school students and outstanding scientists from all over the world for three weeks in August, all of them working together on real scientific experiments that might yield novel results. They spent the first three days simply discovering the different labs participating in the summer school so that they could subsequently choose the scientific project they were interested in. As a closing event, the students prepared a poster session to present the outcomes of the projects developed over the previous weeks.

FUNDACIÓ BBVA

In the 2016 call of the BBVA Foundation Grants to Researchers and Cultural Creators, Neus Martínez, from James Sharpe's group, was awarded a grant for her research project titled 'Non-Invasive Facial Biomarkers of Mental Diseases'. The aim of the project was to create a facial analysis and modelling application with diagnostic and prognostic value for mental diseases related to genetic alterations of DYRK1A and also translatable to other disorders.

THE VELUX FOUNDATIONS

The Velux Foundations are funding the research project titled 'Regenerating Photoreceptors in Retinitis Pigmentosa', by our PI Pia Cosma. Retinitis pigmentosa (RP) is a severe disease that affects 1 in every 3,500 individuals, who undergo progressive loss of vision and for which there is as yet no cure. We intend to test cell fusion-mediated reprogramming as therapy in rd10 mice, an RP mouse model, with the ultimate goal of regenerating photoreceptors and achieving functional rescue of vision.

SWISS NATIONAL SCIENCE FOUNDATION

The SNSF is currently funding a research project by our PI, James Sharpe, entitled 'Reaction-diffusion networks underlying pattern formation of lymphoid tissue'. The project explores the various possible scenarios of pattern formation in lymphoid tissue.

FUNDACIÓN ESPAÑOLA PARA EL FOMENTO DE LA INVESTIGACIÓN DE LA ESCLEROSIS LATERAL AMIOTRÓFICA (FUNDELA)

Our PI Luciano Di Croce was awarded a grant from FUNDELA on November 2017 to address the identification of new therapeutic targets for ALS treatment by using an epigenetic factors screening.

GLENN FOUNDATION FOR MEDICAL RESEARCH

The Glenn Foundation is currently funding the project 'Temporal scaling in *C. elegans* aging', by our PI Nicholas Stroustrup until October 2018.

THE BARCELONA INSTITUTE OF SCIENCE AND TECHNOLOGY (BIST)

BIST is contributing to several ongoing initiatives at the CRG. First, it is co-funding 2 FI PhD Fellowships from AGAUR in the labs of our PIs Pia Cosma and Roderic Guigó for four years. On the other hand, 2 projects from the first Ignite Call from BIST were awarded to CRG researchers. The first one is by Victoire Neguembor (Pia Cosma lab) and is entitled 'GenStorm: an integrated approach to visualize and model the spatial conformation of genes at the nanoscale level' (March-Nov 2017). The second was awarded to Ishier Raote (Vivek Malhotra lab), to the project 'Enlightening TANGO' (March-Nov 2017).

Sponsors

Read the full version here:
annualreport2017.crg.eu

Centre for Genomic Regulation

Dr. Aiguader, 88
PRBB Building
08003 Barcelona, Spain

Tel.: +34 93 316 01 00
Fax +34 93 316 00 99

comunicacio@crg.eu
http://www.crg.eu

Read the full version of our Annual Report here:
annualreport2017.crg.eu

Members of the Board of Trustees:

Member of: