sequencing for a
better life

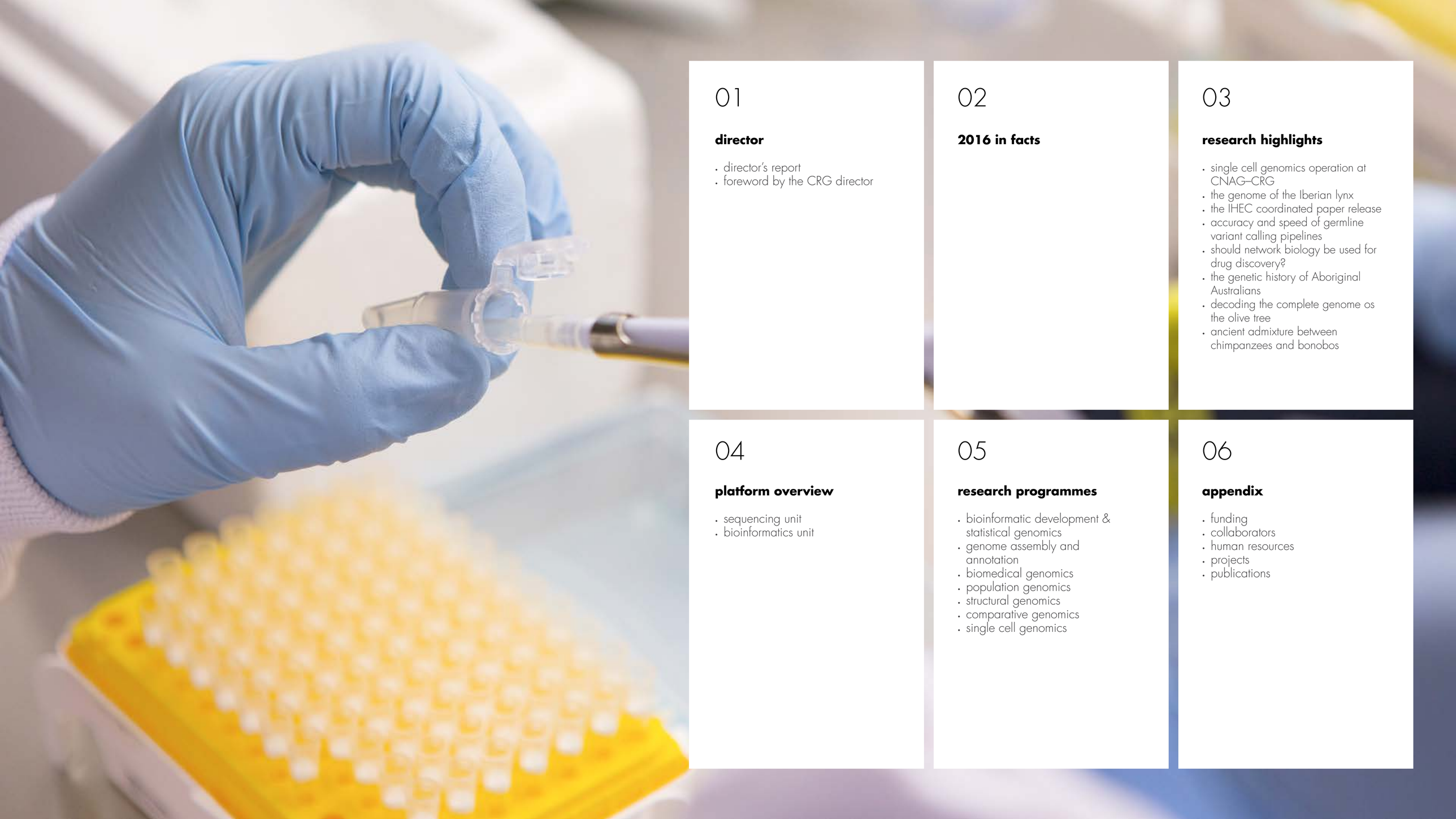
annual report 2016

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG
Centre for Genomic Regulation

Generalitat de Catalunya

MINISTERIO
DE ECONOMÍA
Y COMPETITIVIDAD



cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG
Centre for Genomic Regulation

2016 has been another productive and successful year for CNAG-CRG. We have continued our strategic path to offer the best possible support to our collaborators in their research projects. 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.

This year has seen many highlights. We have taken our quality system to the next level by our ISO17025:2005 accreditation with the scope of DNA/RNA analysis by high throughput sequencing (NGS) by the Spanish national accreditation body ENAC. The ISO17025:2005 accreditation covers the laboratory and data analysis. This puts us in a unique position of having achieved this with such a wide scope. We are one of very few NGS operations to be accredited and to be able to offer it both for academic research projects and clinical application. We have started working with the clinical services of several hospitals for personalized medicine.

In 2016 we started the upgrade process of our sequencer park. The first Illumina HiSeq4000 was received and taken into operation. Our work on the Oxford Nanopore sequencer has reached a level where we can now offer this to our collaborators. The nanopore sequencers provide an orthogonal type of sequence to the Illumina short-read sequencing. Sequences from the nanopore can reach readlengths of several 10s kb. This datatype can be perfectly applied in combination with Illumina short-reads in de novo assembly projects. We are investigating its use in projects where genomes are heavily rearranged, such as for example in cancer.

The EU-funded project BLUEPRINT reached its conclusion in 2016 with the publication of a series of 41 scientific papers within the IHEC framework in journals of the Cell Press group and other high impact journals. These papers report on the relationship of complete epigenetic descriptions of cells of the immune system and put them in context of different diseases. They also describe tools that were developed to capture the entire content of epigenetic profiles. CNAG-CRG played a key role in this effort by sequencing and analysing nearly 200 whole genome methylomes.

The RD-Connect database that was developed at the CNAG-CRG was made available to the European International Rare Disease Research Consortium investigators for βtesting this year. The database received a lot of praise from the βtesters. Towards the end of the year saw the installation of the RD-Connect server.

Several of our de novo assembly and annotation projects arrived at publication this year, notably the iberian lynx, the turbot and the olive tree. The iberian lynx is a critically endangered felid. The annotated genome sequence allows the tuning of conservation efforts of this species with less than 200 remaining animals. Supported by funding from the Emilio Botin Foundation we sequenced an over 1000 year old olive tree, the oldest still living organism ever sequenced. This genetic information will help olive trees in their development and protection against infection.

Personalized medicine is arriving and genome analysis is its major tool, as it provides unprecedented resolution for diagnosing 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.

Ivo G. Gut
Director

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG
Centre for Genomic Regulation

The CRG will be making Integrative Biology the cornerstone of its scientific programme in the coming five years. Addressing the complexity of biological systems, and more specifically of humans, now more than ever requires concerted consensus-based and integrative approaches and biomedical interdisciplinary science. The CRG has generated important scientific insights into our understanding of the organisation, deployment and evolution of genetic information, the internal workings of cells, their differentiation and reprogramming, their collective organisation in tissue formation and alterations in diseases, including cancer. The CRG's strategic fields are medical genetics and personalised medicine. The recent integration of the CNAG within the CRG to become the CNAG-CRG is a key factor in this regard. An infrastructure of excellence such as the CNAG, with its associated technological development research, in combination with fundamental biological research and other resources (e.g. the European Genome-phenome Archive) at the CRG, will position the latter as a leading international player in personalised medicine. This, together with the presence of other prominent research institutes and infrastructures such as the Barcelona Super Computing Centre-Centro Nacional de Supercomputación (BSC-CNS) and hospitals, will make Barcelona a major research hub.

Luis Serrano
Director of the CRG

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG
Centre for Genomic Regulation

2016 in facts – A look back at the year

January

Four of CNAG-CRG PI's (Ivo Gut, Simon Heath, Marc A. Martí-Renom and Tomàs Marquès-Bonet) are present on the top zone of the Ranking of Scientists in Spain published by the Webometrics Ranking of World Universities.

February

Tomàs Marquès-Bonet is coauthor of a study published in Nature that finds first genetic evidence of modern human DNA in a Neanderthal individual.

March

The complete sequencing of the genome of the turbot, conducted by scientists form CNAG-CRG and CRG is published.

April

CNAG-CRG organises the 2nd edition of the workshop Introduction to Genome Analysis to help collaborators get a better understanding and improve interpretation of genome data.

May

Top scientists present their latest research discoveries and ideas at the 5th CNAG Symposium on Genome Research: Single Cell Studies.
CNAG-CRG achieves ISO 17025:2005 accreditation for its Next Generation Sequencing platform.

June

CNAG-CRG participates in the complete genome sequencing of an ancient olive tree that is over 1,000 years old, and that will probably live another 1,300 years.

July

120 students visit CNAG-CRG within the "Summer Scientists" programme, an outreach initiative to boost scientific vocations among high-school students.

September

Oscar Lao, co-authors a study published in Nature that analyses the genome of the Aboriginal Australian population helping to understand how modern humans left the African continent.

October

Researchers from CNAG-CRG benchmark six combinations of state-of-the-art read aligners and variant callers for WES and WGS to improve accuracy and computing costs.

The study led by Tomàs Marquès-Bonet revealing ancient admixture events among bonobos and chimpanzees is published in Science.

November

Cell Reports publishes a study led by CNAG-CRG that sheds light on a specific aspect of the epigenome: DNA methylation. This paper is part of a collection of 41 coordinated studies developed within the International Human Epigenome Consortium (IHEC) published in high-impact journals.

December

Publication of the genome sequencing of the Iberian lynx (Lynx pardinus), currently one of the world's most endangered felines. The sequencing and assembly of the genome was done at CNAG-CRG.

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG
Centre for Genomic Regulation

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG
Centre for Genomic Regulation

research highlights
single cell genomics operation at CNAG-CRG

Single Cell Genomics recently entered the spotlight of basic and translational research areas by providing high-resolution snapshots of complex samples. Profiling the molecular architecture of single cells in heterogeneous and dynamic systems allows their deconvolution in distinct cell types or states. Cell-to-cell variability in genetic, epigenetic and transcriptional profiles describe phenotype differences between cells and allows the hypothesis-free assessment of sample heterogeneity. The Single Cell Genomics team has implemented several techniques to profile genomes and transcriptomes scalable to the characterization of 1000s of individual cells. Specifically, we implemented massively parallel single-cell RNA sequencing (MARS-Seq) allowing the precise gene expression quantification through digital 3`-trancript counting. Full-length mRNA is sequenced using Smart-seq2, providing a technique to sensitively capture and characterize single cell transcriptomes. Genetic alterations and epigenomic variance can be identified by whole-genome amplification (scWGA) and open chromatin profiling (scATACseq) strategies.

The performance of the team's transcriptomics applications was evaluated in a comparative study that included a broad spectrum of single cell RNA sequencing techniques. Teaming up with groups from Germany (LMU, Munich) and Sweden (KI, Stockholm) MARS-Seq and Smart-seq2 were tested for their efficiency and accuracy to assess a single cell's RNA expression profile. A standardized experimental setup allowed comparative analysis across techniques; based on endogenously expressed genes in mouse embryonic stem cells and the analysis of artificial RNAs (ERCC) with defined concentration and length. While both techniques showed high level of accuracy, Smart-seq2 turned out to be the most sensitive approach to determine transcriptional activity of even lowly expressed genes. On the other hand, MARS-Seq represents one of the most cost-effective techniques suitable for large-scale projects and the profiling of 1000s of single cells.

Single cell transcriptome methods rely on the availability of fresh starting material, largely excluding sampling at locations without immediate access to specialized equipment. This has major implications on study designs and is particularly challenging in clinical context or in time course experiments. In order to disconnect time and location of sampling from subsequent processing steps, we established a workflow that allows the transfer of isolated single cells following their lysis in individual wells. Our recent study further increased the flexibility of sample handling by implementing cryopreservation for long-term storage. We showed that the RNA content from cryopreserved cells is indistinguishable from freshly prepared samples and that both conditions are equally suitable to assess biological relevant differences between single cells. The work included clinical samples. Archiving material in cryoprotectants could drastically broaden the scope of single cell application in this context. Moreover, the simultaneous processing of multiple samples largely reduces technical artifacts (batch effects) that occur when single cells are processed at different time points. Together, cryopreservation provides a suitable solution to avoid bottlenecks in sample supply, while increasing the robustness of the results.

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG
Centre for Genomic Regulation

research highlights
the genome of the Iberian lynx

The Iberian lynx (Lynx pardinus) is considered the most endangered felid in the world by the International Union for Conservation of Nature (IUCN). For years, this emblematic species has been the target of several coordinated efforts to avoid its extinction. The CNAG-CRG, playing an active role in a genome project begun in 2010, coordinated by EBD-CSIC and funded by Banco Santander, combined whole genome shotgun sequencing and fosmid pool sequencing strategies in order to generate the first annotated draft assembly of the Iberian lynx genome. Importantly, this is the first mammalian genome entirely sequenced and assembled in Spain. In addition to "Candiles", the individual that was chosen for assembly, the CNAG-CRG re-sequenced the genomes of ten other Iberian lynx coming from the two main populations left on the Iberian Peninsula and one Eurasian lynx as well.

The study detected a series of severe population bottlenecks predating the known demographic decline in the 20th century that have greatly influenced the evolution of the Iberian lynx genome. We observed drastically reduced rates of weak-to-strong substitutions associated with GC-biased gene conversion and increased rates of fixation of transposable elements. The genome of individuals from the two remnant Iberian lynx populations show multiple signatures of genetic erosion, including a high frequency of potentially deleterious variants and substitutions, as well as the lowest genome-wide genetic diversity reported so far in any mammalian species. Noticeably, it has preserved only half of the genetic diversity observed in the cheetah (Acinonyx jubatus).
The extreme genomic erosion we see in the Iberian lynx genome may hamper short- and long-term viability through reduced fitness or adaptive potential. However, the knowledge and resources developed in this study will give a boost to research on felid evolution and conservation genomics, which will in turn benefit the ongoing conservation and management of our national felid.

Work of reference

Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx
Abascal F*, Corvelo A*, Cruz F* et al including Frias L, Ribeca P, Derdak S, Blanc J, Gut M, Gut I, Marques-Bonet T, Alioto T. Genome Biol. 2016 Dec 14;17(1):251.
*contributed equally

The Genome Assembly and Annotation team, together with other scientists managed to read and organize 2.4 billion letters of DNA from Candiles, a male lynx born in the Sierra Morena lynx population, who now forms part of a program for breeding in captivity.(Photo by J.M Perez de Ayala)

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG
Centre for Genomic Regulation

research highlights
the IHEC coordinated paper release

The International Human Epigenome Consortium (IHEC) started in 2010. It is a large international initiative that joins 8 individually funded projects with the objective to generate comprehensive descriptions of the epigenomes of all cell types in the human body. CNAG-CRG is a partner in the EU-funded Project BLUEPRINT and has been responsible for the generation of all of the whole genome bisulphite sequencing and the primary analysis of the data, and contributed to downstream integrative analysis. BLUEPRINT finished its work in 2016 and together with the IHEC summarized its efforts in 41 papers that were concertedly published in journals of the Cell Press group and other high impact journals in November 2016.

The primary goals of IHEC are to coordinate the production of reference maps of human epigenomes for key cellular states relevant to health and diseases, to facilitate rapid distribution of the data to the research community, and to accelerate translation of this new knowledge to improve human health. A critical component of IHEC goals is to coordinate the development of common bioinformatics standards, data models and analytical tools to organize, integrate and display the epigenomic data generated.

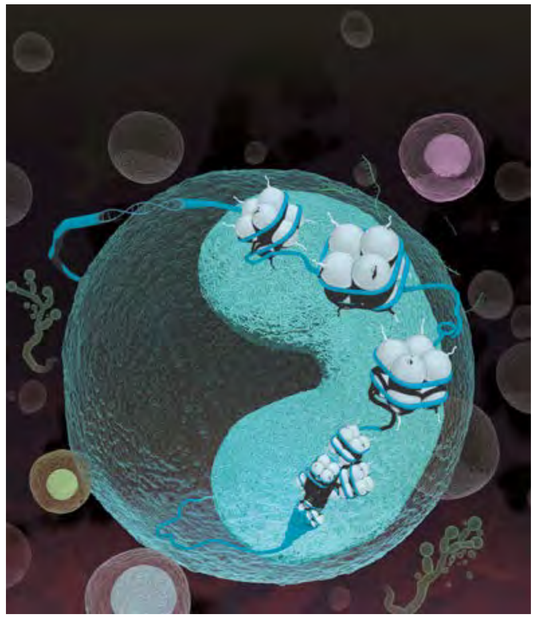
The focus of BLUEPRINT was on the epigenome of blood cell types. Many cell types of the haematological system are involved in immunity. Apart from the reference epigenomes of blood cell types, the relationship of the epigenomes of immune cell types between normal and disease states were investigated. These insights, together with the understanding of how immune cells alter their epigenomes in reaction to or to contribute to a diseased environment, and how the epigenomic changes are established by environmental cues, will likely lead to new biomarkers for better diagnosis and estimation of prognosis. We contributed by generation and analysis of 200 whole genome bisulphite sequences. Apart from our involvement in establishing the reference epigenomes we contributed to studies on haematological cancers and diabetes.

We established compute-efficient pipelines for DNA methylation analysis, defined the data format for DNA methylation sequencing information and contributed to simplified views of genome-wide DNA methylation information.

In particular, the CNAG-CRG led a study that gives insights into mechanisms of normal development and, by comparison, what goes wrong at an epigenetic level in disease. Better understanding of these epigenetic changes in illness will lead to more effective treatment strategies tailored to the genetic profile of each patient.

Work of reference

The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery
Stunnenberg HG; International Human Epigenome Consortium, Hirst M including Gut IG, Heath S as collaborators. Cell. 2016 Nov 17;167(5):1145-1149. doi: 10.1016/j.cell.2016.11.007.



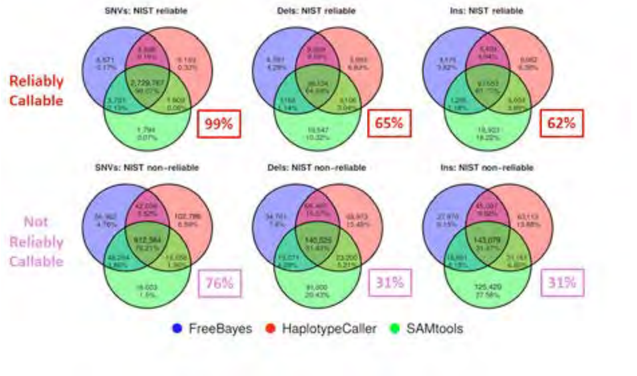
The nucleosome: a complex structure consisting of DNA being wrapped around complexes of proteins (histones). The nucleosome is the major carrier of epigenetic information, and can be affected by a variety of epigenetic mechanisms. (Photo by Spencer-Phillips, EMBL-EBI).

research highlights
accuracy and speed of germline variant calling pipelines

Sequencing a human genome today costs less than 1% of what it did in 2006, and takes hours instead of years. However, computing costs to analyse sequencing data have decreased far less than the cost of whole exome and whole genome sequencing (WES and WGS) and today constitutes a non-negligible fraction of the overall sequence analysis costs. The clinical genetics community is adopting WES and WGS as a standard practice in research and diagnosis and therefore it is essential to choose the most accurate and cost-efficient analysis pipeline.
Researchers from the Data Analysis unit benchmarked six combinations of state-of-the-art read aligners (BWA-MEM, GEM3) and variant callers (FreeBayes, GATK-HC, Samtools) on WES and WGS data from the exhaustively-analysed NA12878 sample. The study, published in Human Mutation, aims to evaluate the robustness of the variant detection process, while taking into account the computing resources required.
The results show that the six variant calling pipelines are consistent in 70% of the genome. Both the agreement between the pipelines and the concordance of the called variants with the NA12878 reference dataset is extremely high for SNVs and somewhat lower for InDels. The latter could be due to poorer performing methods but also to biases in the generation of the reference dataset. Furthermore, we observed a very high concordance between the variants called in the WGS and the WES from the same individual, even if they had been sequenced 3 years appart by two different labs.
However, we confirmed that the concordance of the results is much lower in another 20% of the genome. Finally, we noticed that in the remaining 10% of the genome it is unfeasible to conduct reliable variant calling comparisons with the sequencing technology and tools assayed in this benchmark and remain a big challenge.
Regarding the computing costs, the study found substantial differences between tools. It is notable that GEM3, the alignment tool developed and used at the CNAG-CRG was found to be 4 times faster than the widely used BWA-MEM. While BWA-MEM required almost 300 CPU hours for WGS alignment, GEM3 used less than 60 CPU hours to complete the same task. Further downstream in the pipeline, we also noticed differences in speed of up to 20 times between the fastest caller (FreeBayes) and the rest on WGS samples.

Work of reference

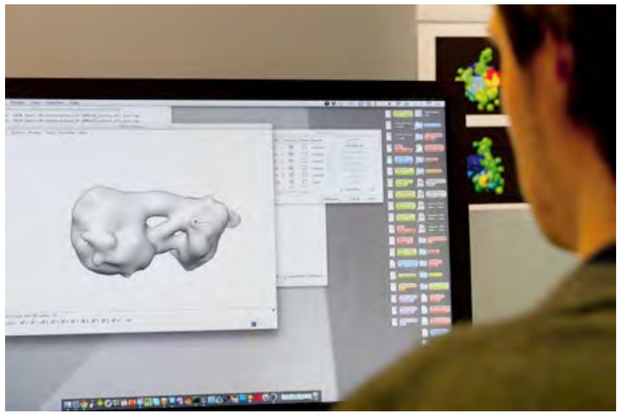
From Wet-Lab to Variations: Concordance and Speed of Bioinformatics Pipelines for Whole Genome and Whole Exome Sequencing
Laurie S, Fernandez-Callejo M, Marco-Sola S, Trotta JR, Camps J, Chacón A, Espinosa A, Gut M, Gut I, Heath S, Beltran S**.
Hum Mutat. 2016 Dec;37(12):1263-1271. doi: 10.1002/humu.23114. Epub 2016 Sep 26.
**Corresponding author



Venn diagrams illustrating concordance of variant identification for GEM3 alignments.

research highlights
should network biology be used for drug discovery?

According to The Tufts Center for the Study of Drug Development (TTCSDD), the development and marketing approval for a New Molecular Entity (NME) takes more than 13 years and costs around US$2.6 billion. Moreover, the cost of putting a new drug into the market has dramatically increased since the 1970s. This raise in drug development cost has led to a dramatic shrinkage of the efficiency, measured in terms of the number of new approved drugs per billion US dollars of research and discovery spending. Factors that have contributed to the raise of drug development costs include increased clinical trial complexity, larger clinical trial size, greater assessment of safety and toxicity drug profiles or evaluation on equivalent drugs to accommodate payer demands for comparative effectiveness data. Simultaneously, the emergence of high throughput technologies such as High Throughput Screenings or Next Generation Sequencing have led to a drug discovery paradigm shift from the traditional single drug perspective towards a more target centric view. The application of these technologies alongside the increasing complexity of the treating diseases and the growing intricacy of the mechanism of action of the drug have also significantly increased the costs of pre-clinical stages. Hence, there is an urgent need for readjusting the drug discovery process to tackle these new problems. More specifically, modern drug discovery programs should be able to deal with the massive amount of data generated in the initial stages of the drug discovery pipeline.
In this scenario, our group has recently published a review arguing that computational methods can play a significant role to reduce the time and costs of pre-clinical stages. Over the last thirty years, computational methods have helped the development of new therapeutics. However, we are still far from extracting all their potential when applied to the drug discovery field. We reviewed how computational methods in general, and network-based methods in particular, could be used to optimize the pre-clinical stages of the drug discovery process.



Work of reference

Should network biology be used for drug discovery?
*Martínez-Jiménez F, Marti-Renom MA***
Expert Opin Drug Discov. 2016 Dec;11(12):1135-1137. Epub 2016 Sep 23.
***corresponding author*

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG
Centre for Genomic Regulation

research highlights
the genetic history of Aboriginal Australians

When and how anatomically modern humans appeared and spread around the world is still not solved. Although consensus among the scientific community supports an African origin, some scientists based on fossil record argue that such spread out of the African continent occurred twice, being the Aboriginal Australians an admixed remnant of such first diaspora whereas others claims that there has been a unique colonizing event.

Within this context, the analysis of the genetic variation of the Aboriginal Australian population is crucial for disentangling among these competing hypotheses. Furthermore, the analysis of the genetic diversity within and between Aboriginal Australian populations can help understanding when and how the Australian continent was colonized and which are the demographic dynamics of the Australian population, both within the continent but also with other human populations.

In this study we had the opportunity, for the first time, to analyze a collection of 84 Aboriginal Australian fully sequenced genomes from nine geographic locations. In this collaboration, the CNAG-CRG Population Genomics team was involved in the analyses of population substructure of the Aboriginal Australian population. We were interested in answering questions such as: what is the amount of population substructure within the Aboriginal Australian population? Can we use this population substructure to predict the geographic ancestry of a given individual? What is the amount of genetic admixture with neighbouring populations such as the Papuans? What was the impact from a genetic point of view of the recent arrival of other populations, such as the North European, to the Australian continent?

Our results suggest that most of current Aboriginal Australian individuals are highly admixed with European, East Asian and Papuan populations. We showed that the obtained percentages of ancestry could be used for predicting when the admixture occurred in time, which we could further corroborate with historical records. An important consequence of such recent admixture was that most of the conducted analyses for understanding the population dynamics within the continent were severely impaired. However, when only the Aboriginal Australian ancestry was considered, we observed a strong correlation between geography and genetic variation, to the extent that the geographic sampling origin could be predicted with a relatively small error for most of the sampled individuals. Overall, our results provided for the first time a picture at high resolution of the complex recent demographic history of the Aboriginal Australians.

Biological anthropologist Dr Michael Westaway from Griffith University obtains a saliva sample from Thanakwith Elder Mr Thomas Wales, Cape York, Australia. (Photo by Tom Cebula, Wall to Wall Media).

Work of reference

A genomic history of Aboriginal Australia
Malaspinas AS, Westaway MC, Muller C et al including Lao O. Nature. 2016 Oct 13;538(7624):207-214. doi: 10.1038/nature18299. Epub 2016 Sep 21.

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG
Centre for Genomic Regulation

research highlights
decoding the complete genome of the olive tree

In 2016 the CNAG-CRG, in a collaborative project including Toni Gabaldón of the Centre for Genomic Regulation (CRG) and Pablo Vargas of the Real Jardín Botánico (CSIC-RJB), finished sequencing and assembling the genome of a nearly 1,300-year-old olive tree belonging to the Spanish Farga variety. The three-year research project was funded by Banco Santander, who also provided the specimen, a monumental olive tree translocated to the Ciudad Financiera in Boadilla del Monte, Spain, in 2005. This was the first time that such an old living being has had its genome sequenced. These "millennial" olive trees are extremely long-lived, with this one likely to live another 1,300 years.

The CNAG-CRG Assembly and Annotation team took a fosmid pool approach to sequencing the genome, a strategy designed to overcome the problem of repetitive sequence and high heterozygosity, resulting in an assembly of its 1.4 billion base-long genome. This reference genome sequence for the olive will provide a valuable resource for studying its domestication history, as well as developmental and physiological processes that might give insight into its longevity, its adaptability to arid conditions, and differences between the varieties, sizes and flavor of olives, for example. Ultimately, it may bolster molecular breeding efforts as well as support new research ways of protecting the olive tree from bacterial and viral infections. In a first step toward these aims, the CNAG-CRG has annotated over 56,000 genes in its genome, more than double that of the human genome, and has started to study their functions.

Work of reference

Genome sequence of the olive tree, Olea europaea
Cruz F, Julca I, Gómez-Garrido J et al including Loska D, Frias L, Ribeca P, Derdak S, Gut M, Gut IG, Alioto TS**. Gigascience. 2016 Jun 27;5:29. doi: 10.1186/s13742-016-0134-5.
**corresponding author

The olive tree sequenced at the CNAG-CRG is a Farga olive tree, named "Santander," from Sierra del Maestrazgo estimated to have been planted around the year 792, and transplanted in 2005 to the Ciudad Financiera del Banco Santander in Boadilla del Monte, Spain.

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

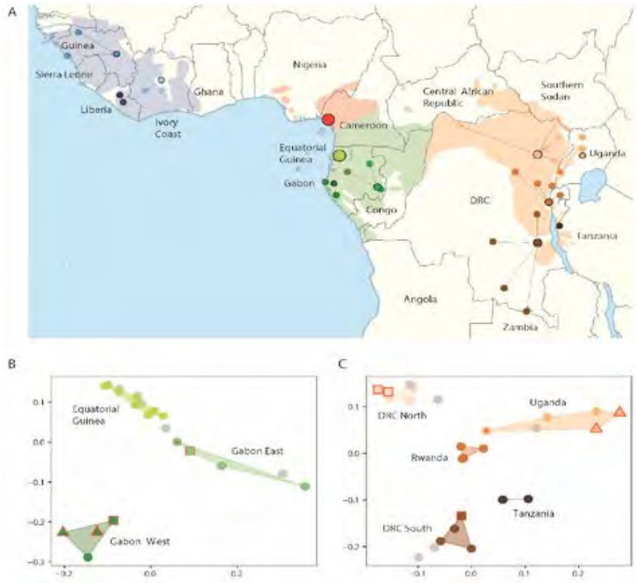
CRG
Centre for Genomic Regulation

research highlights
ancient admixture between chimpanzees and bonobos

Between 1,5 and 2 million years ago chimpanzees (Pan troglodytes) and bonobos (Pan paniscus) split from a common ancestor and evolved important strong physical and behavioural differences. To this day, the existence of gene flow between the species has not been considered due to the Congo River that physically separates the geographical distribution ranges of the species. This study is then, the first study to reveal admixture among the species similar to what has been reported between humans and Neanderthals. This has been possible thanks to the application of recent analytical tools to detect current and ancient admixture among groups.
The studied samples, comprising 75 complete genomes of chimpanzees and bonobos, cover 10 countries in Africa, from the westernmost to the easternmost region of the chimpanzee range. Here, we found that chimpanzees do have very strong geographical stratification of their genome diversity, and thus, the results have a direct application to the conservation of these species because they permit the detection of the origin of chimpanzees confiscated from illegal trafficking.

Work of reference

Chimpanzee genomic diversity reveals ancient admixture with bonobos
de Manuel M, Kuhlwilm M, Frandsen P et al including Lao O, Gut M, Gut I, Marques-Bonet T**.
Science. 2016 Oct 28;354(6311):477-481. Epub 2016 Oct 27.
**corresponding author



Geographic stratification of chimpanzee diversity.
PCA plot of chromosome 21 SNP data for a set of central (left) and eastern (right) chimpanzees. Samples with unknown origin are coloured in gray. The test samples (squares) were found to cluster perfectly with genome data for known samples (dots) (deManuel et al. Science 2016). Triangles correspond to SNP data from fecal samples.

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG
Centre for Genomic Regulation

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG
Centre for Genomic Regulation

sequencing unit

the Sequencing Unit of the CNAG-CRG stands at the forefront of genomic research in Spain, processing thousands of samples every year with a comprehensive palette of state-of-the-art sequencing applications and executing hundreds of sequencing projects from Spain and from international consortia.

Head of the Unit:
Marta Gut

Laboratory Managers:
Julie Blanc, Katja Kahlem, Ana González (until August), Lidia Agueda (from August)

Laboratory Technicians:
Marta Lopez, Ana Gonzalez (from September), Pilar Herruzo, Maite Rico, Caterina Mata, Laetitia Casano, Aurora Padron, Yasmina Mirassou, Giulia Lunazzi (until June), Regina Antoni, Nuria Aventin

Specialized Technicians:
Javier Gutierrez, Esther Lizano (until June), Silvia Speroni (from December)

Quality Mananger:
Lidia Sevilla

Single Cell Genomics Team:
Team leader: Holger Heyn
Postdoctoral Fellows: Amy Lauren Guillaumet, Gustavo Rodríguez
Data Analyst: Elisabetta Mereu (from May)
PhD Students: Atefeh Lafzi (from September)

The Sequencing Unit operates several different types of Illumina sequencers and MinIon sequencers from Oxford Nanopore Technologies (ONT). The ONT MinIons have been taken into data production this year, with their unprecedented length of sequencing reads that allow cheap and portable de novo assembly of prokaryotic and eukaryotic genomes.

The collaborative projects take advantage of the high proficiency and quality standards of the data production in the CNAG-CRG Sequencing Unit with an ISO 9001:2008 and ISO 17025:2005 certified and accredited quality management system. With this system we are able to provide data with quality meeting clinical use regulatory requirements, aiming to bring the highest standards to genome research in Spain and Europe.

Five teams work jointly within the Sequencing Unit – Biorepository, Sample Preparation, Sequencing Production, Support and Single Cell Genomics. The inter-team communication is assured and information tracking guaranteed by the Laboratory Information Management System (LIMS).

The Biorepository team receives, quality controls and stores samples from the collaborators and transfers sample aliquot to the Sample Preparation team which is responsible for preparing sequencing libraries for the Sequencing Production team to load prepared libraries onto the sequencing instruments. The Support team tests and sets up new protocols – manual and automated – and follows the correct functioning of the CNAG-CRG's large and small technical equipment. The Single Cell Genomics team elaborates new applications to increase the portfolio of offerings from ultra-low input samples, down to single cell transcriptomics and genomics.

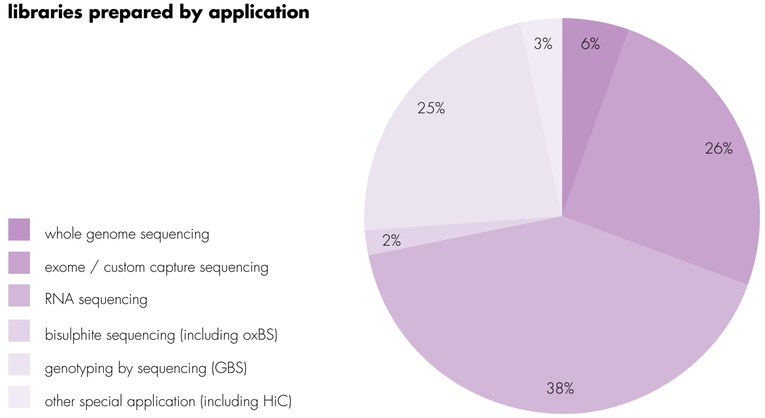
Research Projects

- Implementation of automated protocol for Genotyping-By-Sequencing (GBS)
- Maintaining of the failure rates during sample preparation to less than 5%
- Testing and implementation of Nimblegen SeqCap EZ MedExome Target Enrichment protocol
- Operational use of the mitochondrial DNA capture within the MedExome capture
- Increasing the enrichment and specificity of Nimblegen small captures by implementing a double capture protocol
- Enlarging the protocol portfolio with the Kapa stranded mRNA protocol
- Implementation of a manual Truseq Amplicon low input protocol for the B-CAST project
- Development of automated protocol for B-CAST Truseq Amplicon with several liquid handlers
- Introducing a new Pipetmax robot (Gilson) to the Biorepository pipeline
- Implementation of the management process for the B-CAST project
- Development of sample preparations for MinIon MkI sequencers
- Incorporation of an Illumina HiSeq4000 sequencing instrument
- Implementation of large scale full length mRNA protocols (Smart-seq2) for single cell, low input and degraded samples
- Improvements of 3'-end mRNA count single cell methods – MARS-seq and C1 HT arrays
- Testing and evaluation of multiple displacement amplifications (MDA) and PCR-based methods from single cells
- Preparation and passing the audit of ISO 17025:2005

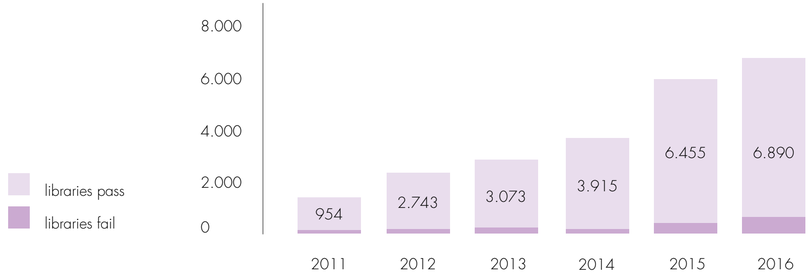
sequencing unit

the Sequencing Unit of the CNAG-CRG stands at the forefront of genomic research in Spain, processing thousands of samples every year with a comprehensive palette of state-of-the-art sequencing applications and executing hundreds of sequencing projects from Spain and from international consortia.

libraries prepared by application



libraries prepared by year



bioinformatics analysis unit

the Bioinformatics Analysis unit develops and operates state-of-the-art pipelines, tools and databases to manage and analyse the sequencing data generated at the CNAG-CRG.

Head of the Unit:
Sergi Beltran

Production Bioinformatics Team:
Matthew Ingham (Manager), Raul Alcántara, Eloi Casals

Data Analysis Team:
Sergi Beltran (Manager), Jordi Camps, Sophia Derdak, Steven Laurie, Inés Martínez, Anastasios Papakonstantinou, Davide Piscia, Joan Protasio, Raul Tonda, Jean-Rémi Trotta

Functional Genomics team:
Simon Heath (Manager), Marc Dabad, Anna Esteve, Marcos Fernández

The unit collaborates closely with internal and external groups and delivers customised high quality user-friendly results and actively participates in its interpretation. The unit is also involved in European projects such as RD-Connect, Elixir-Excelerate and B-CAST.

The activity is carried out by highly skilled data analysts, software engineers and bioinformaticians divided into three teams. The Production Bioinformatics team develops and operates the Laboratory Information Management System (LIMS) and pipelines to process, control the quality and transfer the data generated by the laboratory. The Data Analysis team develops and operates bioinformatics pipelines to analyse sequencing data, mainly related to germinal variant and somatic mutation identification and annotation. Finally, the Functional Genomics team participates in RNA-Seq and Methylation studies. Although most of the activity is related to clinical research (mainly on Mendelian disorders and cancer), the participation in agrigenomics and model organisms projects has been increasing steadily in the past years.

Services

- Collaborative analyses
- Experimental design
- Data processing and quality control
- Germinal variant identification
- Somatic mutation identification
- Variant annotation, filtering and interpretation
- Copy Number Variant identification
- Genotyping by Sequencing (GBS)
- Differential expression and functional analysis
- Identification of isoforms
- Identification of methylation profiles

Research lines

- Bioinformatics for Rare Disease Research and Clinical Genomics
- Development of pipelines and tools
- Benchmarking of data analysis methods
- Variant annotation systems
- Omics data integration
- Agrogenomics (GBS)

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG
Centre for Genomic Regulation

bioinformatics analysis unit

the Bioinformatics Analysis unit develops and operates state-of-the-art pipelines, tools and databases to manage and analyse the sequencing data generated at the CNAG-CRG.

Selected publications

From Wet-Lab to Variations: Concordance and Speed of Bioinformatics Pipelines for Whole Genome and Whole Exome Sequencing. Laurie S, Fernandez-Callejo M, Marco-Sola S, Trotta JR, Camps J, Chacón A, Espinosa A, Gut M, Gut I, Heath S, Beltran S**. Hum Mutat. 2016 Dec;37(12):1263-1271. doi: 10.1002/humu.23114. Epub 2016 Sep 26.
**Corresponding author

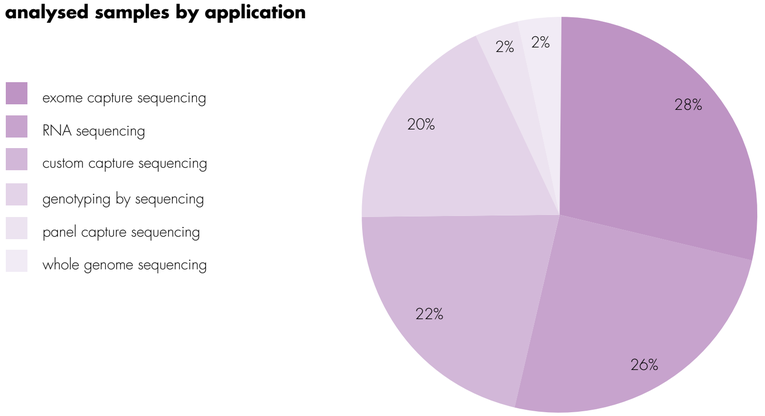
Identification of protein-damaging mutations in 10 swine taste receptors and 191 appetite-reward genes. Clop A, Sharaf A, Castelló A et al including Derdak S, Beltran S. BMC Genomics. 2016 Aug 26;17:685. doi: 10.1186/s12864-016-2972-z.

Novel Candidate Genes and a Wide Spectrum of Structural and Point Mutations Responsible for Inherited Retinal Dystrophies Revealed by Exome Sequencing de Castro-Miró M, Tonda R, Escudero-Ferruz P et al. PLoS One. 2016 Dec 22;11(12):e0168966. doi: 10.1371/journal.pone.0168966. eCollection 2016.

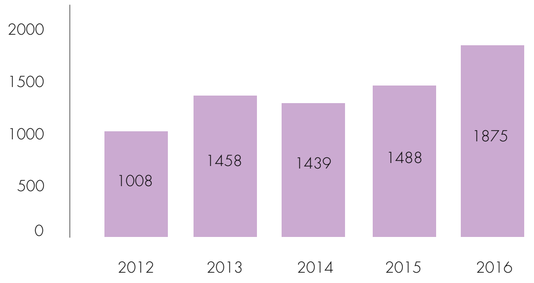
Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx Abascal F*, Corvelo A*, Cruz F* et al including Frias L, Ribeca P, Derdak S, Blanc J, Gut M, Gut I, Marques-Bonet T, Alioto T. Genome Biol. 2016 Dec 14;17(1):251.
*contributed equally

ATRX driver mutation in a composite malignant pheochromocytoma. Comino-Méndez I, Tejera ÁM, Currás-Freixes M, Remacha L, Gonzalvo P, Tonda R, Leton R, Blasco MA, Robledo M, Cascon A. Cancer Genet. 2016 Jun;209(6):272-7. doi: 10.1016/j.cancergen.2016.04.058. Epub 2016 Apr 26.

analysed samples by application



analysed samples by year



cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG
Centre for Genomic Regulation

bioinformatics development and statistical genomics team

Team Leader:
Simon Heath

Staff Scientist:
Emanuele Raineri

Postdoctoral Fellows:
Angelika Merkel, Ron Schuyler

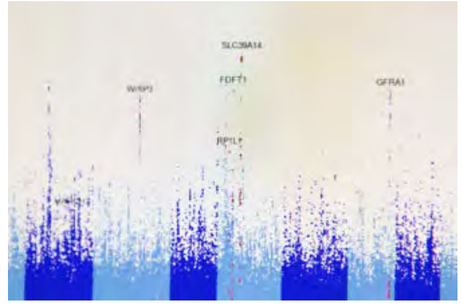
Software Engineer:
Marcos Fernández

PhD Student:
Santiago Marco (until July)

The research focus of the team is on the development and application of efficient methods (from both statistical and computational perspectives) for the large scale processing and integrative analysis of omics datasets. These methods have been developed into analysis pipelines at CNAG-CRG, and have been applied to both in house and collaborative projects. The largest single dataset we are currently working on is the epigenome data from the BLUEPRINT project, which consists of over 200 samples of mostly healthy tissue from different blood cell types, but we are also working on diverse epigenetic datasets from the International Human Epigenome Consortium and the ICGC-TCGA Pan Cancer project.

Research projects

- Efficient methods for large scale DNA methylation analysis from WGBS experiments, generating calls of both genetic and epigenetic variation
- Integrative analysis of diverse epigenetic datasets (RNA-Seq, ChIP-Seq, WGBS) across > 200 healthy and cancer samples from the human haemopoietic system with the aim of uncovering novel epigenetic signatures of genome regulation
- Investigation into the relationship between large scale genome organization predicted from Hi-C experiments and finer scale epigenetic information from DNA methylation and chromatin marks, and to what extent the different epigenetic data types can predict each other
- Characterization of partially methylated domains in different haemopoietic cell types, and their relationship with differentiation and cancer
- A systematic evaluation of all stages of our WGBS analysis pipeline on simulated and real datasets, comparing the ability to call genomic and methylation variants against other commonly used pipelines



Selected Publications

Distinct Trends of DNA Methylation Patterning in the Innate and Adaptive Immune Systems. Schuyler RP, Merkel A, Raineri E, et al including Gut I, Heath S**. Cell Rep. 2016 Nov 15;17(8):2101-2111. doi: 10.1016/j.celrep.2016.10.054.
**corresponding author

Decoding the DNA Methylome of Mantle Cell Lymphoma in the Light of the Entire B Cell Lineage. Queirós AC, Beekman R, Vilarrasa-Blasi R et al including Merkel A, Raineri E, Heath S, Gut IG. Cancer Cell. 2016 Nov 14;30(5):806-821. doi: 10.1016/j.ccell.2016.09.014.

Information recovery from low coverage whole-genome bisulfite sequencing. Libertini E, Heath SC, Hamoudi RA et al including Gut M, Gut IG. Nat Commun. 2016 Jun 27;7:11306. doi: 10.1038/ncomms11306.

Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. Ventham NT, Kennedy NA, Adams AT et al including Heath S, Gut IG. Nat Commun. 2016 Nov 25;7:13507. doi: 10.1038/ncomms13507.

Comprehensive genome and epigenome characterization of CHO cells in response to evolutionary pressures and over time. Feichtinger J, Hernández I*, Fischer C et al including Merkel A, Heath S. Biotechnol Bioeng. 2016 Apr 12. doi: 10.1002/bit.25990. [Epub ahead of print] * Visiting student

genome assembly and annotation team

Team Leader:
Tyler Alioto

Postdoctoral Fellow:
Fernando Cruz

Technician:
Jèssica Gómez

We carry out "genome projects" in the classical sense, i.e. sequencing, assembling and annotating genomes de novo, specializing in large eukaryotic genomes. We also have expertise in assembly and annotation of transcriptomes. Genome assembly is not only difficult due to the sheer size of the data and computational requirements, but also because the biology of genomes is confounded by repetitive elements, polyploidy and variation (single-nucleotide, insertions/deletions, and larger structural variants). We aim to meet and overcome these challenges, developing new computational protocols as each project demands. In particular, we are focusing our efforts on the adoption of single molecule long read technology. Annotation of the gene content of the newly assembled genome is key to understanding the genome, once finished. Our rapid robust annotation pipeline, which we are constantly improving, helps us to accomplish this task.

Research lines

- Genome sequence assembly
- Gene prediction/Genome annotation
- Transcriptome reconstruction
- Functional annotation

Selected publication

Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes. Vlasova A, Capella-Gutiérrez S, Rendón-Anaya M et al including Gómez-Garrido J, Alioto T. Genome Biol. 2016 Feb 25;17(1):32. doi: 10.1186/s13059-016-0883-6.

Whole Genome Sequencing of Turbot (Scophthalmus maximus; Pleuronectiformes): A Fish Adapted to Demersal Life. Figueras A, Robledo D, Corvelo A et al including Gómez-Garrido J, Gut M, Gut IG, Alioto T. DNA Res. 2016 Mar 6. pii: dsw007. [Epub ahead of print]

Genome sequence of the olive tree, Olea europaea. Cruz F, Julca I, Gómez-Garrido J et al including Loska D, Frias L, Ribeca P, Derdak S, Gut M, Gut IG, Alioto TS**. Gigascience. 2016 Jun 27;5:29. doi: 10.1186/s13742-016-0134-5.
**corresponding author

Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. Abascal F*, Corvelo A*, Cruz F* et al including Frias L, Ribeca P, Derdak S, Blanc J, Gut M, Gut I, Marques-Bonet T, Alioto T. Genome Biol. 2016 Dec 14;17(1):251.
*contributed equally

biomedical genomics group

Group Leader:
Ivo G. Gut

Postdoctoral fellows:
Justin Whalley, Gian-Andri Thun, Miranda Stobbe, Darek Kedra (until July), Ryoji Takahashi (from June)

PhD students:
Lukasz Roguski

The Biomedical Genomics Group works on deepening the understanding of the function of the human genome using sequencing data from disease studies. We apply computational methods to determine genetic and genomic causes of disease and reversely also study the effects of the disease on the genome. For our studies we use data that is generated at the CNAG-CRG and combine it with data that we retrieve from other sources. This allows us to increase the power of our studies and ask the data questions that were not necessarily at the base of the initial study design. We have been working on three classes of diseases:
1. Cancer
2. Rare diseases
3. Respiratory disorders

Research projects

- Cancer: Our main effort here goes to the PanCancer study, where we have been investigating the origin of non-random, recurring somatic insertion and deletion mutations in cancer genomes and have been coordinating the Quality Control working group. We have defined five non-redundant measures that can be used to describe the quality of tumour genomes. We have also gained important insight into the dynamics of mutational processes. We concluded the EU-funded project BLUEPRINT, for which we contributed analyses of epigenomes of haematological cancers.
- Rare Disease: In rare diseases we have been concentrating on the RD-Connect project and data with the objective to mine the data for phenotype-specific modifier variants.
- Respiratory Disorders: In respiratory disorders we have been investigating the manifestation of genomic alterations and their relation to gene expression and clinical phenotypes with a particular focus on the presentation of early signs of known and potentially causal genomic events.

Selected publications

New technologies for DNA analysis - a review of the READNA Project. McGinn S., et al., N Biotechnol. 2016 May 25;33(3):311-30. doi: 10.1016/j.nbt.2015.10.003. Epub 2015 Oct 26.

A whole-genome sequence and transcriptome perspective on HER2-positive breast cancers. Ferrari A., et al., Nat Commun. 2016 Jul 13;7:12222. doi: 10.1038/ncomms12222.

Decoding the DNA Methylome of Mantle Cell Lymphoma in the Light of the Entire B Cell Lineage. Queirós AC., et al., Cancer Cell. 2016 Nov 14;30(5):806-821. doi: 10.1016/j.ccell.2016.09.014.

Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells. Dieuleveult M., et al., Nature. 2016 Feb 4;530(7588):113-6. doi: 10.1038/nature16505. Epub 2016 Jan 27.

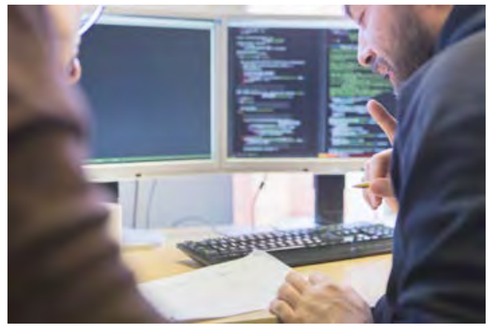
Emphysema- and airway-dominant COPD phenotypes defined by standardised quantitative computed tomography. Subramanian D.R., et al., Eur Respir J. 2016 Jul;48(1):92-103. doi: 10.1183/13993003.01878-2015. Epub 2016 May 26.

population genomics team

Team Leader:
Oscar Lao

PhD Students:
Iago Maceda (from March),
Amrinder Singh (from May)

The Population Genomics team started in 2015. Our team focuses on describing and quantifying the genetic variation present in current populations in order to understand the micro-evolution of the given species and asses the phenotypic consequences of such genetic diversity. In particular, we address questions related to which is the genetic origin from a population point of view of a given individual, which are the demographic and selective factors that shaped the genetic variation present in a population and how ultimately this variation influences and allows us to detect the individual risk in phenotypes of interest. In order to achieve these goals, we are actively working on developing new tools and algorithms for describing population substructure in the genome and understanding the biological implications of such structure, identifying the fingerprint of polygenic adaptation in complex phenotypes and evaluating the impact of archaic introgression in phenotypes of interest. Our team focuses on human species but the universality of the proposed methods allows us to apply them to other model organisms.



Research projects

- Development of new algorithms for predicting genetic ancestry prediction
- Identification of the fingerprint of polygenic adaptation in complex phenotypes such as body mass index or bone mineral density
- Analysis of the impact of archaic introgression in the evolution of complex phenotypes such as bone mineral density or aggression
- Development of new algorithms for statistically inferring demographic parameters and distinguishing amongcompeting models
- Analysis of the genetic variants associated to attention deficit hyperactivity disorder (ADHD) phenotype

Selected publications

Chimpanzee genomic diversity reveals ancient admixture with bonobos. de Manuel M, Kuhlwilm M, Frandsen P et al including Lao O, Gut M, Gut I, Marques-Bonet T**. Science. 2016 Oct 28;354(6311):477-481. Epub 2016 Oct 27.
**corresponding author

A genomic history of Aboriginal Australia.
Malaspinas AS, Westaway MC, Muller C et al including Lao O. Nature. 2016 Oct 13;538(7624):207-214. doi: 10.1038/nature18299. Epub 2016 Sep 21.

structural genomics group

Group Leader:
Marc A. Marti-Renom

Staff Scientist:
Davide Baù

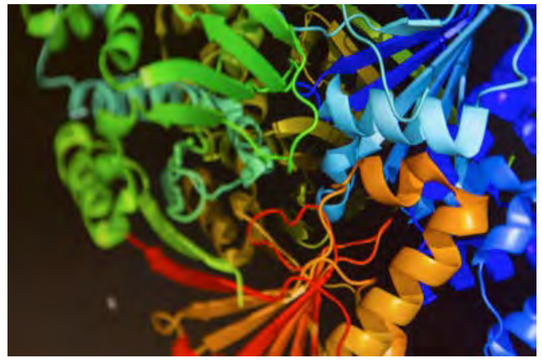
Posdoctoral Fellows:
François Serra, Yannick Spill, Marco di Stefano, Irene Farabella

PhD Students:
David Dufour, Francisco Martínez-Jiménez, Gireesh K. Bogu, Silvia Galan, Paula Soler

Programmer, Technicians:
Michael Goodstadt

Laboratory Technician:
Yasmina Cuartero

How biomolecules fold and function in a three-dimensional space is one of the most challenging questions in biology. For example, we have limited knowledge on how the 2-meter-long DNA molecule folds in the micro-sized nucleus or how RNA, proteins and small chemical compounds fold and interact to perform the most basic functions of the cell. Our research group employ the laws of physics and the rules of evolution to develop and apply computational methods for predicting the 3D structures of macromolecules and their complexes. By doing so, we have recently developed new computational tools to predict the insertion of proteins in the lipid transmembrane of the cell as well as predict drug compounds using network biology.



Research projects / Research lines

- Structure determination of genomes. We develop methods for determining the 3D organization of the chromatin
- Comparative RNA structure prediction. We develop a series of tools for the alignment of RNA structures and the prediction of their structures and functions.
- Protein-Ligand interactions. We develop methods for comparative docking of small chemical compounds and their target proteins.

Selected publications

Should network biology be used for drug discovery? *Martínez-Jiménez F, Marti-Renom MA*** *Expert Opin Drug Discov. 2016 Dec;11(12):1135-1137. Epub 2016 Sep 23.* ***corresponding author*

Structural features of the fly chromatin colors revealed by automatic three-dimensional modeling. Serra F, Baù D, Filion G, Marti-Renom MA** bioRxiv 2016 **corresponding author

Biological insertion of computationally designed short transmembrane segments. Baeza-Delgado C, von Heijne G, Marti-Renom MA, Mingarro I. Sci Rep. 2016 Mar 18;6:23397. doi: 10.1038/srep23397.

comparative genomics group

Group Leader:
Tomas Marques-Bonet (ICREA Research Professor)

Postdoctoral Fellows:
Inna Povolotskaya, Martin Kuhlwilm, David de Juan

PhD Students:
Raquel Garcia, Jessica Hernández, Marc de Manuel, Irene Lobon, Lukas Kuderna, Claudia Fontsere, Aitor Serres, Sojung Han, Manuel Solís, Luis Ferrández

Our main line of research is centered in the discovery of the extent of all kinds of genome variation within different phenotypically genomes. Specifically, we study genome variation (centered on CNVs), gene expression and epigenetic differences in the human species in the context of great ape evolution and other mammalian genomes such as canids. The goal is to create an integrated view of genome evolution by studying changes in the composition, frequency, size and location at every major branch point of recent human evolution.

Research lines

- Genomic variation in ape genomes
- Evolution of Gene Regulation
- Canid evolution

Selected publications

Chimpanzee genomic diversity reveals ancient admixture with bonobos. Manuel M., et al., Science. 2016 Oct 28;354(6311):477-481. Epub 2016 Oct 27.

Evolution and demography of the great apes. Kuhlwilm M., et al., Curr Opin Genet Dev. 2016 Dec;41:124-129. doi: 10.1016/j.gde.2016.09.005. Epub 2016 Oct 4.

Demographic History of the Genus Pan Inferred from Whole Mitochondrial Genome Reconstructions. Lobon I., et al., Genome Biol Evol. 2016 Jul 3;8(6):2020-30. doi: 10.1093/gbe/evw124.

Ancient gene flow from early modern humans into Eastern Neanderthals. Kuhlwilm M., et al., Nature. 2016 Feb 25;530(7591):429-33. doi: 10.1038/nature16544. Epub 2016 Feb 17.

Worldwide patterns of genomic variation and admixture in gray wolves. Fan Z., et al., Genome Res. 2016 Feb;26(2):163-73. doi: 10.1101/gr.197517.115. Epub 2015 Dec 17.

single cell genomics team

Team Leader:
Holger Heyn

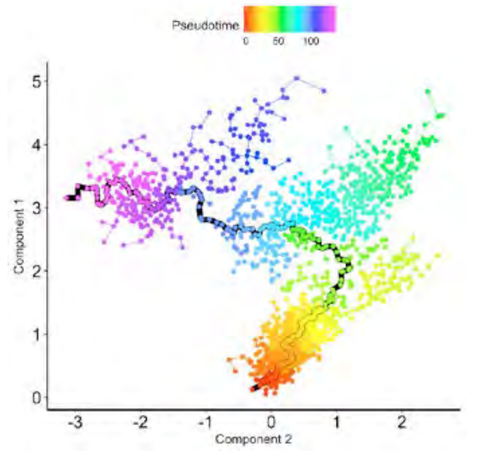
Postdoctoral Fellows:
Amy Guillaumet, Gustavo Rodriguez

Data Analyst:
Elisabetta Mereu (from May)

PhD Students:
Atefeh Lafzi (from September)

The Single Cell Genomics team is dedicated to advance our understanding of genome activity in health and disease. The mission of the team is the implementation of novel single cell genomics technologies and their application in a research context, with a focus on translational research. New computational strategies include methods to deconvolute tissue composition, determine cell type heterogeneity, identify cell type defining markers or track transcriptional dynamics. The team's experience in single cell genomics is unique in Spain and suits research on virtual every species, tissue or disease context.

The Single Cell Genomics team is equipped for and specialized on the large-scale production of transcriptomic, genomic and epigenomic profiles from single cells. Several techniques, such as the single cell RNA and DNA sequencing methods MARS-seq, Smart-seq2 and G&Tseq, were implemented into standard operation procedures, to allow a reliable and reproducible preparation of samples. Single cell projects are conducted with specialized equipment, enabling an immediate sample processing and the monitoring of high quality data production throughout the experiments. Specialized technical equipment includes automated liquid handling robotics and microfluidic single cell devices (C1, Fluidigm). In 2016, the Single Cell Genomics team processed and analyzed >20,000 single cell genomes and transcriptomes.



Single cells ordered along a predicted developmental trajectory. The colours indicate their progression in pseudotime inferred by their transcriptomic information.

Research lines

- Cancer Heterogeneity
- Developmental Dynamics
- Complex Tissue Composition
- Single Cell Genomics Technologies

Selected publications

RNA sequencing validation of the Complexity INdex in SARComas prognostic signature. Lesluyes T., et al., Eur J Cancer. 2016 Apr;57:104-11. doi: 10.1016/j.ejca.2015.12.027. Epub 2016 Feb 23.

Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer. Heyn H., et al., Genome Biol. 2016 Jan 26;17(1):11. doi: 10.1186/s13059-016-0879-2.

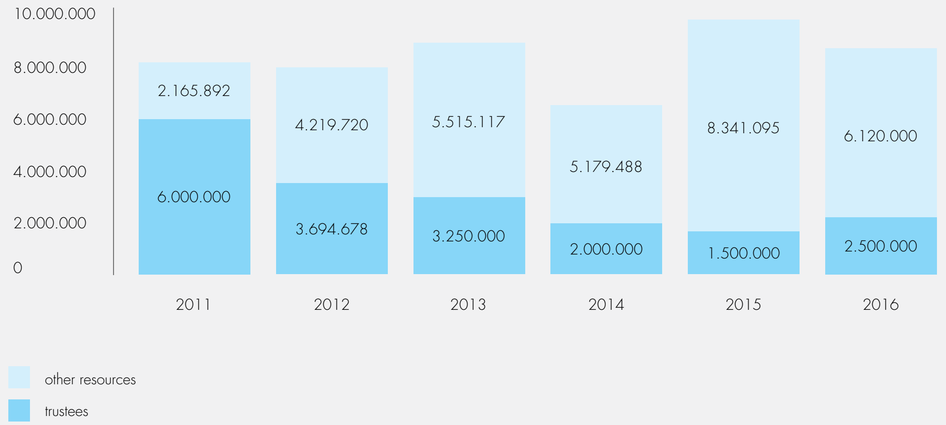
annual report 2016

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

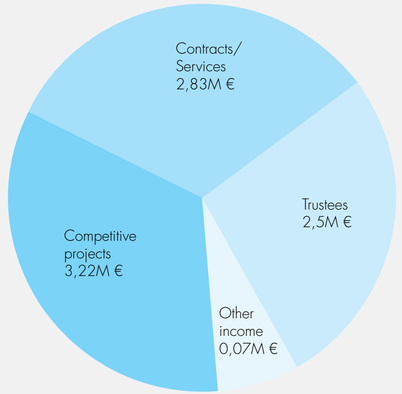
CRG
Centre for Genomic Regulation

funding

funding evolution

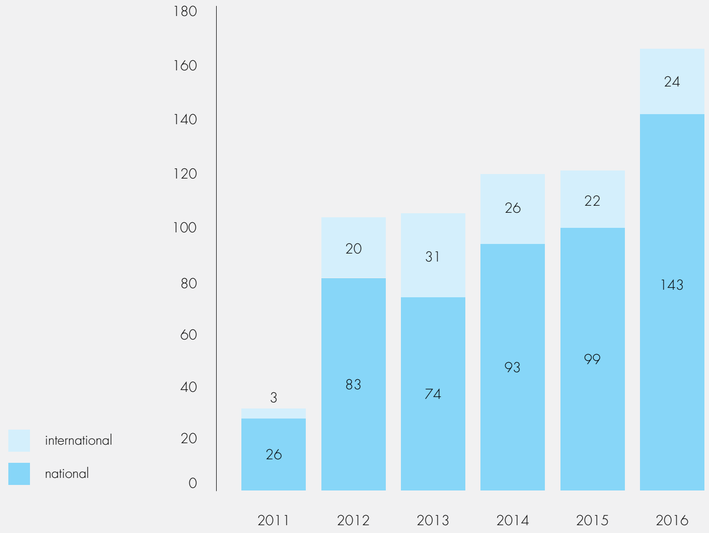


funding by sources 2016

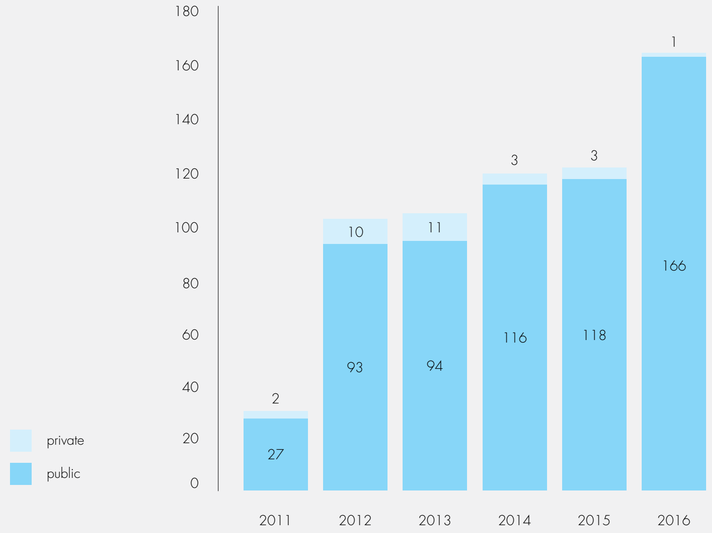


collaborators

collaborators by origin

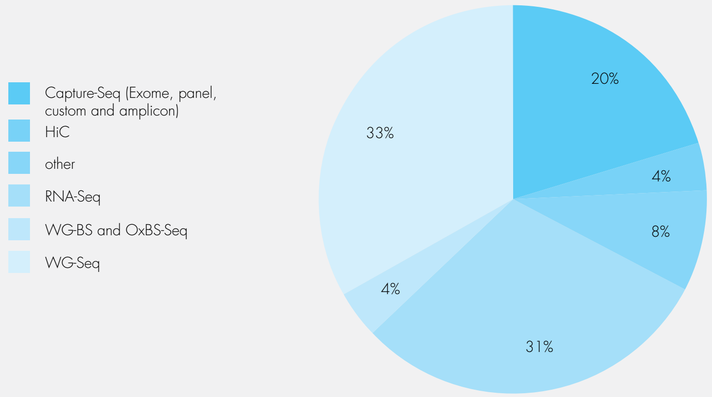


collaborators by sector

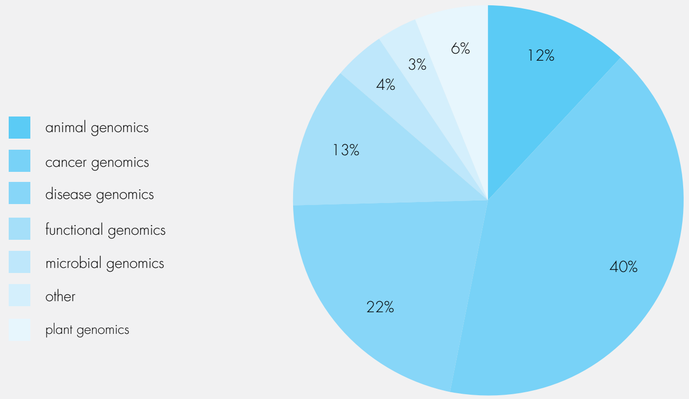


collaborators

2016 activity by sequencing application



2016 activity by research area



cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG
Centre for Genomic Regulation

human resources
staff evolution by area and position

	2016	2015	2014	2013	2012	2011
SEQUENCING (LAB)	25	21	22	19	17	13
Unit heads	2	2	1	1	1	1
Postdocs and staff scientists	2	2	1	-	-	-
Technicians	20	17	20	18	16	12
Predocs	1	0	0	0	0	0
BIOINFORMATICS (IT)	46	42	37	29	25	16
Group / Team / Unit leaders	6	7	5	6	7	6
Postdocs and staff scientists	13	13	14	12	9	5
Technicians	21	18	15	7	4	4
Predocs	6	4	3	4	5	1
MANAGEMENT (ADMIN)	4	5	5	4	5	5
TOTAL	75	68	64	52	47	34

human resources
staff evolution by area and gender

	2016	2015	2014	2013	2012	2011
SEQUENCING (LAB)	25	21	22	19	17	13
Male	3	2	2	2	2	1
Female	22	19	20	17	15	12
BIOINFORMATICS (IT)	46	42	37	29	25	16
Male	36	32	29	25	21	14
Female	10	10	8	4	4	2
MANAGEMENT (ADMIN)	4	5	5	4	5	5
Male	0	0	1	1	2	2
Female	4	5	4	3	3	3
RATIO FEMALES VS TOTAL	48%	50%	50%	46,15%	46,81%	50%

*In both tables data has been updated according to the CRG categories

projects
international competitive projects in force in 2016

acronym	project type	timeline	funding
AIRPROM	FP7 Collaborative Project	2011-2016	459,400€
BLUEPRINT	FP7 Collaborative Project	2011-2016	2,718,950€
RD-CONNECT	FP7 Collaborative Project	2012-2018	1,497,365€
IBD-CHARAC	FP7 Collaborative Project	2012-2017	994,600€
BBMRI-LPC	FP7 Collaborative Project & Coordination and Support Action	2013-2017	494,426€
4DGENOME	ERC Synergy	2014-2019	1,752,612€
PHYLOCANCER	ERC Consolidator	2014-2019	320,004€
BCAST	H2020 Collaborative Project	2015-2020	2,380,125€
MUG	H2020 EINFRA	2015-2018	510,695€
ELIXIR-EXCELERATE	H2020 INFRADEV	2015-2019	200,000€
MIND	H2020 MSCA-ITN	2016-2018	179,155€
SINGEK	H2020 MSCA-ITN	2016-2019	285,753€
TOTAL			11,793,085€

projects
national competitive projects in force in 2016

reference	project type	timeline	funding
REDBIO2014 PT13/0001/0044	FIS/ ISCIII	2014-2017	359,150€
2014 SGR 615	Suport a Grups de Recerca	2014-2017	70,000€
CIN-2015-244-BUSPONTHEAL	ERA-NET	2016-2019	75,000€
BIO2015-71969-REDI	Redes de Excelencia	2015-2017	178,000€
BFU2013-047736-P	Proyectos de Excelencia	2014-2016	205,000€
BIO2015-71792-P	Proyectos de Excelencia	2016-2018	296,450€
BFU2015-68759-P	Proyectos de Excelencia	2016-2018	142,296€
RYC-2013-14797	Ramón y Cajal Contracts	2015-2019	208,600€
BES-2014-070327	Predoc training contract	2015-2019	88,400€
PTA2014-09515-I	Technicians contract	2015-2018	39,000€
CP14/00229	Miguel Servet	2016-2019	232,480€
TOTAL			1,894,376€

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG
Centre for Genomic Regulation

publications

1. *Novel Candidate Genes and a Wide Spectrum of Structural and Point Mutations Responsible for Inherited Retinal Dystrophies Revealed by Exome Sequencing.* de Castro-Miró M, Tonda R, Escudero-Ferruz P et al. PLoS One. 2016 Dec 22;11(12):e0168966. doi: 10.1371/journal.pone.0168966. eCollection 2016.
2. *Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx* Abascal F*, Corvelo A*, Cruz F* et al including Frias L, Ribeca P, Derdak S, Blanc J, Gut M, Gut I, Marques-Bonet T, Alioto T. Genome Biol. 2016 Dec 14;17(1):251. *contributed equally
3. *Genetic Diversity and Population Structure of Rice Varieties Cultivated in Temperate Regions* Reig-Valiente JL, Viruel J, Sales E et al including Gut M, Deardak S. Rice (N Y). 2016 Dec;9(1):58. Epub 2016 Oct 20.
4. *Evolution and demography of the great apes* Kuhlwilm M, de Manuel M, Nater A, Greminger MP, Krützen M, Marques-Bonet T*. Curr Opin Genet Dev. 2016 Dec;41:124-129. doi: 10.1016/j.gde.2016.09.005. Epub 2016 Oct 4. *Corresponding author
5. *From Wet-Lab to Variations: Concordance and Speed of Bioinformatics Pipelines for Whole Genome and Whole Exome Sequencing* Laurie S, Fernandez-Callejo M, Marco-Sola S, Trotta JR, Camps J, Chacón A, Espinosa A, Gut M, Gut I, Heath S, Beltran S** Hum Mutat. 2016 Dec;37(12):1263-1271. doi: 10.1002/humu.23114. Epub 2016 Sep 26. **Corresponding author
6. *Natural Selection in the Great Apes* Cagan A, Theunert C, Laayouni H et al including Marques-Bonet T. Mol Biol Evol. 2016 Dec;33(12):3268-3283. Epub 2016 Oct 30.
7. *Whole exome sequencing analysis reveals TRPV3 as a risk factor for cardioembolic stroke* Carrera C, Jiménez-Conde J, Derdak S et al including Beltran S. Thromb Haemost. 2016 Nov 30;116(6):1165-1171. Epub 2016 Sep 8.
8. *Increased DNA methylation variability in type 1 diabetes across three immune effector cell types* Paul DS, Teschendorff AE, Dang MA et al including Heath S, Gut M, Gut IG. Nat Commun. 2016 Nov 29;7:13555. doi: 10.1038/ncomms13555.
9. *Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease* Ventham NT, Kennedy NA, Adams AT et al including Heath S, Gut IG. Nat Commun. 2016 Nov 25;7:13507. doi: 10.1038/ncomms13507.
10. *The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery* Stunnenberg HG; International Human Epigenome Consortium, Hirst M Including Gut IG, Heath S as collaborators. Cell. 2016 Nov 17;167(5):1145-1149. doi: 10.1016/j.cell.2016.11.007.
11. *β-Glucan Reverses the Epigenetic State of LPS-Induced Immunological Tolerance* Novakovic B, Habibi E, Wang SY et al including Heath S, Gut I. Cell. 2016 Nov 17;167(5):1354-1368.e14. doi: 10.1016/j.cell.2016.09.034.
12. *Distinct Trends of DNA Methylation Patterning in the Innate and Adaptive Immune Systems* Schuyler RP, Merkel A, Raineri E, et al including Gut I, Heath S. Cell Rep. 2016 Nov 15;17(8):2101-2111. doi: 10.1016/j.celrep.2016.10.054.
13. *Decoding the DNA Methylome of Mantle Cell Lymphoma in the Light of the Entire B Cell Lineage* Queirós AC, Beekman R, Vilarrasa-Blasi R et al including Merkel A, Raineri E, Heath S, Gut IG. Cancer Cell. 2016 Nov 14;30(5):806-821. doi: 10.1016/j.ccell.2016.09.014.
14. *Human Oocyte-Derived Methylation Differences Persist in the Placenta Revealing Widespread Transient Imprinting* Sanchez-Delgado M, Court F, Vidal E et al including Marques-Bonet T. PLoS Genet. 2016 Nov 11;12(11):e1006427. doi: 10.1371/journal.pgen.1006427. eCollection 2016.
15. *Chimpanzee genomic diversity reveals ancient admixture with bonobos* de Manuel M, Kuhlwilm M, Frandsen P et al including Lao O, Gut M, Gut I, Marques-Bonet T**. Science. 2016 Oct 28;354(6311):477-481. Epub 2016 Oct 27. **corresponding author
16. *A genomic history of Aboriginal Australia* Malaspinas AS, Westaway MC, Muller C et al including Lao O. Nature. 2016 Oct 13;538(7624):207-214. doi: 10.1038/nature18299. Epub 2016 Sep 21.
17. *Mutations in GLDN, Encoding Gliomedin, a Critical Component of the Nodes of Ranvier, Are Responsible for Lethal Arthrogryposis* Maluenda J, Manso C, Quevarec L et al including Gut M, Gut I. Am J Hum Genet. 2016 Oct 6;99(4):928-933. doi: 10.1016/j.ajhg.2016.07.021. Epub 2016 Sep 8.
18. *The Fanconi anemia DNA damage repair pathway in the spotlight for germline predisposition to colorectal cancer* Esteban-Jurado C, Franch-Expósito S, Muñoz J et al including Serra E, Beltran S. Eur J Hum Genet. 2016 Oct;24(10):1501-5doi: 10.1038/ejhg.2016.44. Epub 2016 May 11.

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG
Centre for Genomic Regulation

publications

19. *Should network biology be used for drug discovery?* Martínez-Jiménez F, Marti-Renom MA** Expert Opin Drug Discov. 2016 Dec;11(12):1135-1137. Epub 2016 Sep 23. **corresponding author
20. *Sequence variation between 462 huma individuals fine-tunes functional sites of RNA processing* Ferreira PG, Oti M, Barann M et al including Esteve-Codina A. Sci Rep. 2016 Sep12;6:32406. doi: 10.1038/srep32406.
21. *PRKG1 and genetic diagnosis of early-onset thoracic aortic disease* Gago-Díaz M, Blanco-Verea A, Teixidó G, Huguet F, Gut M, Laurie S, Gut I, Carracedo A, Evangelista A, Brion M. Eur J Clin Invest. 2016 Sep;46(9):787-94. doi: 10.1111/eci.12662. Epub 2016 Aug 18.
22. *Identification of protein-damaging mutations in 10 swine taste receptors and 191 appetite-reward genes* Clop A, Sharaf A, Castelló A et al including Derdak S, Beltran S. BMC Genomics. 2016 Aug 26;17:685. doi: 10.1186/s12864-016-2972-z.
23. *Atad2 is a generalist facilitator of chromatin dynamics in embryonic stem cells* Morozumi Y, Boussouar F, Tan M et al including Gut M. J Mol Cell Biol. 2016 Aug;8(4):349-62. doi: 10.1093/jmcb/mjv060. Epub 2015 Oct 12.
24. *Somatic Embryonic FGFR2 Mutations in Keratinocytic Epidermal Nevi* Toll A, Fernández LC, Pons T et al including Beltran S, Gut M, Gut I. J Invest Dermatol. 2016 Aug;136(8):1718-21. doi: 10.1016/j.jid.2016.03.040. Epub 2016 Apr 19.
25. *Mitochondrial Complex I Is a Global Regulator of Secondary Metabolism, Virulence and Azole Sensitivity in Fungi* Bromley M, Johns A, Davies E et al including Gut M, Gut I. PLoS One. 2016 Jul 20;11(7):e0158724. doi: 10.1371/journal.pone.0158724. eCollection 2016.
26. *A whole-genome sequence and transcriptome perspective on HER2-positive breast cancers* Ferrari A, Vincent-Salomon A, Pivot X et al including Gut IG, Gut M. Nat Commun. 2016 Jul 13;7:12222. doi: 10.1038/ncomms12222.
27. *CARGO: effective format-free compressed storage of genomic information* Roguski Ł, Ribeca P**. Nucleic Acids Res. 2016 Jul 8;44(12):e114. doi: 10.1093/nar/gkw318. Epub 2016 Apr 29. **corresponding author
28. *Demographic History of the Genus Pan Inferred from Whole Mitochondrial Genome Reconstructions* Lobon I, Tucci S, de Manuel M et al including Dabad M, Marques-Bonet T**. Genome Biol Evol. 2016 Jul 3;8(6):2020-30. doi: 10.1093/gbe/evw124. **corresponding author
29. *Emphysema and airway- dominant COPD phenotypes defined by standardised quantitative computed tomography* Subramanian DR, Gupta S, Burggraf D et al including Gut I. Eur Respir J. 2016 Jul;48(1):92-103. doi: 10.1183/13993003.01878-2015. Epub 2016 May 26.
30. *Whole-genome single nucleotide polymorphism-based linkage analysis in spondyloarthritis multiplex families reveals a new susceptibility locus in 13q13* Costantino F, Chaplais E, Leturcq T et al including Gut I. Ann Rheum Dis. 2016 Jul;75(7):1380-5. doi: 10.1136/annrheumdis-2015-207720. Epub 2015 Aug 14.
31. *Information recovery from low coverage whole-genome bisulfite sequencing* Libertini E, Heath SC, Hamoudi RA et al including Gut M, Gut IG. Nat Commun. 2016 Jun 27;7:11306. doi: 10.1038/ ncomms11306.
32. *Saturation analysis for whole-genome bisulfite sequencing data* Libertini E, Heath SC, Hamoudi RA et al including Gut M, Gut IG. Nat Biotechnol. 2016 Jun 27. doi: 10.1038/nbt.3524. [Epub ahead of print]
33. *Genome sequence of the olive tree, Olea europaea* Cruz F, Julca I, Gómez-Garrido J et al including Loska D, Frias L, Ribeca P, Derdak S, Gut M, Gut IG, Alioto TS**. Gigascience. 2016 Jun 27;5:29. doi: 10.1186/s13742-016-0134-5. **corresponding author
34. *Making sense of big data in health research: Towards an EU action plan* Auffray C, Balling R, Barroso I et al including Gut IG. Genome Med. 2016 Jun 23;8(1):71. doi: 10.1186/s13073-016-0323-y.
35. *ATRX driver mutation in a composite malignant pheochromocytoma* Comino-Méndez I, Tejera ÁM, Currás-Freixes M, Remacha L, Gonzalvo P, Tonda R, Leton R, Blasco MA, Robledo M, Cascon A. Cancer Genet. 2016 Jun;209(6):272-7. doi: 10.1016/j.cancergen.2016.04.058. Epub 2016 Apr 26.
36. *New technologies for DNA analysis - a review of the READNA Project* McGinn S, Bauer D, Brefort T et al including Gut M, Heath S, Gut IG**. N Biotechnol. 2016 May 25;33(3):311-30. doi: 10.1016/j.nbt.2015.10.003. Epub 2015 Oct 26. **corresponding author
37. *MetaTrans: an open-source pipeline for metatranscriptomics* Martinez X, Pozuelo M, Pascal V, Campos D, Gut I, Gut M, Azpiroz F, Guarner F, Manichanh C. Sci Rep. 2016 May 23;6:26447. doi: 10.1038/srep26447.

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG
Centre for Genomic Regulation

publications

38. *Dynamic recruitment of Ets1 to both nucleosome-occupied and -depleted enhancer regions mediates a transcriptional program switch during early T-cell differentiation* Cauchy P, Maqbool MA, Zacarias-Cabeza J et al including Gut M, Gut I. Nucleic Acids Res. 2016 May 5;44(8):3567-85. doi: 10.1093/nar/gkv1475. Epub 2015 Dec 15.
39. *Functional Implications of Human-Specific Changes in Great Ape microRNAs* Gallego A, Melé M, Balcells I et al including Marques-Bonet T. PLoS One. 2016 Apr 22;11(4):e0154194. doi: 10.1371/journal.pone.0154194. eCollection 2016.
40. *Comprehensive genome and epigenome characterization of CHO cells in response to evolutionary pressures and over time* Feichtinger J, Hernández I, Fischer C et al including Merkel A, Heath S. Biotechnol Bioeng. 2016 Apr 12. doi: 10.1002/bit.25990. [Epub ahead of print]
41. *RNA sequencing validation of the Complexity INdex in SARComas prognostic signature* Lesluyes T, Pérot G, Largeau MR et al including Mendez-Lago M, Gut M, Gut I. Eur J Cancer. 2016 Apr;57:104-11. doi: 10.1016/j.ejca.2015.12.027. Epub 2016 Feb 23.
42. *Identification of genetic variation in the swine toll-like receptors and development of a porcine TLR genotyping array* Clop A, Huisman A, van As P, Sharaf A, Derdak S, Sanchez A Genet Sel Evol. 2016 Mar 31;48:28. doi: 10.1186/s12711-016-0206-0.
43. *The Fungus Candida albicans Tolerates Ambiguity at Multiple Codons* Simões J, Bezerra AR, Moura GR, Araújo H, Gut I, Bayes M, Santos MA. Front Microbiol. 2016 Mar 31;7:401. doi: 10.3389 fmicb.2016.00401. eCollection 2016.
44. *Biological insertion of computationally designed short transmembrane segments* Baeza-Delgado C, von Heijne G, Marti-Renom MA, Mingarro I. Sci Rep. 2016 Mar 18;6:23397. doi: 10.1038/srep23397.
45. *Whole Genome Sequencing of Turbot (Scophthalmus maximus; Pleuronectiformes): A Fish Adapted to Demersal Life* Figueras A, Robledo D, Corvelo A et al including Gómez-Garrido J, Gut M, Gut IG, Alioto T. DNA Res. 2016 Jun;23(3):181-92. doi: 10.1093/dnares/dsw007. Epub 2016 Mar 6.
46. *Demographically-Based Evaluation of Genomic Regions under Selection in Domestic Dogs* Freedman AH, Schweizer RM, Ortega-Del Vecchyo D et al including Marques-Bonet T. PLoS Genet. 2016 Mar 4;12(3):e1005851. doi: 10.1371/journal.pgen.1005851. eCollection 2016.
47. *Specific small-RNA signatures in the amygdala at premotor and motor stages of Parkinson's disease revealed by deep sequencing analysis* Pantano L, Friedländer MR, Escaramís G, Lizano E et al. Bioinformatics. 2016 Mar 1;32(5):673-81. doi: 10.1093/bioinformatics/btv632. Epub 2015 Nov 2.
48. *Evaluation of mRNA markers for estimating blood deposition time: Towards alibi testing from human forensic stains with rhythmic biomarkers* Lech K, Liu F, Ackermann K, Revell VL, Lao O, Skene DJ, Kayser M. Forensic Sci Int Genet. 2016 Mar;21:119-25. doi: 10.1016/j.fsigen.2015.12.008. Epub 2015 Dec 28
49. *Worldwide patterns of genomic variation and admixture in gray wolves* Fan Z, Silva P, Gronau I et al including Marques-Bonet T. Genome Res. 2016 Feb;26(2):163-73. doi: 10.1101/gr.197517.115 Epub 2015 Dec 17.
50. *ETE 3: Reconstruction, analysis and visualization of phylogenomic data* Huerta-Cepas J, Serra F, Bork P. Mol Biol Evol. 2016 Jun;33(6):1635-8. doi: 10.1093/molbev/msw046. Epub 2016 Feb 26.
51. *Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes* Vlasova A, Capella-Gutiérrez S, Rendón-Anaya M et al including Gómez-Garrido J, Alioto T. Genome Biol. 2016 Feb 25;17(1):32. doi: 10.1186/s13059-016-0883-6.
52. *Ancient gene flow from early modern humans into Eastern Neanderthals* Kuhlwilm M, Gronau I, Hubisz MJ et al including Marques-Bonet T. Nature. 2016 Feb 25;530(7591):429-33. doi: 10.1038/nature16544. Epub 2016 Feb 17.
53. *Genetic load of loss-of-function polymorphic variants in great apes* de Valles-Ibáñez G, Hernandez-Rodriguez J, Prado-Martinez J, Luisi P, Marquès-Bonet T, Casals F. Genome Biol Evol. 2016 Mar 26;8(3):871-7. doi: 10.1093/gbe/evw040.
54. *Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells* de Dieuleveult M, Yen K, Hmitou I et al including Gut M, Gut I. Nature. 2016 Feb 4;530(7588):113-6. doi: 10.1038/nature16505. Epub 2016 Jan 27.
55. *Dissecting Daily and Circadian Expression Rhythms of Clock-Controlled Genes in Human Blood* Lech K, Ackermann K, Revell VL, Lao O, Skene DJ, Kayser M. J Biol Rhythms. 2016 Feb;31(1):68-81. doi: 10.1177/0748730415611761. Epub 2015 Nov 2.

publications

56. Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer Heyn H*, Vidal E, Ferreira HJ et al including Gut M, Gut I. Genome Biol. 2016 Jan 26;17(1):11. doi: 10.1186/s13059-016-0879-2. *first author
57. *Selective constraints on protamine 2 in primates and rodents* Lüke L, Tourmente M, Dopazo H, Serra F, Roldan ER. BMC Evol Biol. 2016 Jan 22;16(1):21. doi: 10.1186/s12862-016-0588-1.
58. *Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function* Pattaro C, Teumer A, Gorski M et al including Thun GA. Nat Commun. 2016 Jan 21;7:10023. doi: 10.1038/ncomms10023.
59. *Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs* Marsden CD, Ortega-Del Vecchyo D, O'Brien DP, Taylor JF, Ramirez O, Vila C, Marques-Bonet T, Schnabel RD, Wayne RK, Lohmueller KE. Proc Natl Acad Sci U S A. 2016 Jan 5;113(1):152-7. doi: 10.1073/pnas.1512501113. Epub 2015 Dec 22.
60. *Chromatin and RNA Maps Reveal Regulatory Long Noncoding RNAs in Mouse Bogu GK*, Vizán P, Stanton LW, Beato M, Di Croce L, Marti-Renom MA**.Mol Cell Biol. 2015 Dec 28;36(5):809-19. doi: 10.1128/MCB.00955-15. **corresponding author