SEQUENCE MAPPING AND ASSEMBLY ASSESSMENT PROJECT (SMAAP) RGASP3 / DNGASP

April, 5th-7th, 2011

Museu Colet
c/ Buenos Aires, 56
Barcelona

Organizers:

biocat International Center for Scientific Debate

Initiative fostered by: Welfare Projects "la Caixa" Foundation

CRG Centre for Genomic Regulation

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

SEQUENCE MAPPING AND ASSEMBLY ASSESSMENT PROJECT (SMAAP) RGASP3 / dnGASP

April, 5th-7th, 2011

Museu Colet, c/ Buenos Aires, 56
Barcelona - Spain

Introduction

From 5th to 7th April 2011 Barcelona will host the "**Sequence Mapping and Assembly Assessment Project dnGASP / RGASP3 Workshop**" with the aim of comparing and evaluating methods and strategies for *de novo* genome assembly (dnGASP) and RNASeq read alignment (RGASP3) using data from 2nd generation sequencing.

The workshop will bring together participants of the **RGASP3** and **dnGASP** contests to present their methods, which will be evaluated by the independent Evaluation Committee using standardized metrics to compare the different approaches. Additionally, there will be five keynote presentations by Invited Speakers.

The use of 2nd generation sequences is becoming the "*de facto*" standard for monitoring genomes and genome activity for diverse applications in Medicine, Agriculture, and Biotechnology. The workshop will contribute to identify the most promising computational approaches to efficiently deal with 2nd generation sequencing data.

This workshop is organized by the International Center for Scientific Debate (ICSD), an initiative of Biocat, with the support of Welfare Projects "la Caixa" Foundation, the Centre for Genomic Regulation (CRG) and the National Center for Genomic Analysis (CNAG).

Program

Tuesday, April, 5th

8:45 **Registration**

9:00 **Presentation**

9:30 **Introduction to dnGASP / RGASP**
Ivo Gut – director CNAG. Roderic Guigó – director Bioinformatics and Genomics CRG

9:45 **Report on GAW / Assemblathon.** Tyler Alioto – CNAG . Paolo Ribeca – CNAG

10:15 **Evaluation methods and overview of results.** Tyler Alioto – CNAG . Andre Corvelo - CNAG

11:00 **Coffee break**

11:30 **Lecture:** "**Columbus: templated de novo assembly of short reads".** Daniel Zerbino – University of California. Santa Cruz. California. USA

12:10 **Shaun Jackman** – Group: "Steven Jones" Canada's Michel Smith Genome Science Center. Vancouver. Canada.

12:35 **Zeming Ning** – Group: "Zemin Ning". Wellcome Trust Institute Sanger. Hinxton. United Kingdom.

1:00 **Lunch**

2:00 **Lecture: Assembly: Progresses and Challenges.** Yingrui Li. BGI. Beijing. China

2:40 **Bryan Downie** – Group: "Mathias Platzer" FLI Leibniz Institute for Age Research. Jena, Germany.

3:05 **Victor Solovyev** – Group: "Softberry" Royal Holloway, University of London. United Kingdom

3:30 **Rayan Chikhi** – Group: Symbiose" IRISA/ INRIA. Rennes. France

4:00 **Coffee break**

4:30 **Alexey Sergushichev** – Group "SPbSU ITMO". SPbSU ITMO. Saint Petersburg. Russia

4:55 **Ekaterina Khmrameeva** – Group "Genoanalytica" Genoanalytica. Moscow. Russia

5:20 **Jared Simpson** – Group: "Richard Durbin" Wellcome Trust Sanger Institute. Hinxton. London. United Kingdom

6:15 **Bus to Barcelona Supercomputing Center**

6:45 **Visit to Barcelona Supercomputing Center**

8:00 **Dinner at Restaurant Temporada** (10' walk from the visit venue)

10:00 **Bus to the Hotel**

Wednesday, April, 6th

9:00 **dnGASP Results of Evaluation –** Tyler Alioto. André Corvelo

11:00 **Coffee break**

11:30 **Discussion**

12:30 **The way ahead.** Ivo Gut – Director CNAG. Barcelona. Spain

1:00 **Lunch**

2:00 **Lecture: "Biological and Computational Lessons Learned from Analyses of Human and Fly Transcriptomes"**. Thomas Gingeras. Cold Spring Harbor Laboratory. New York. USA

3:00 **Wellcome RGASP3.** Roderic Guigó – CRG

3:05 **Lecture "Gencode"**. Tim Hubbard. Wellcome Trust Sanger Institute. Hinxton. Cambridge. United Kingdom

3:30 **RGAPS1 and RGASP2** . Jen Harrow. Wellcome Trust Sanger Institute. Hinxton. Cambridge. UK

4:00 **RGASP 3** Paul Bertone. Eurepean Bioinformatics Institute. London. UK

4:30 **Coffee break**

5:00 **Zemin Ning** – Group: "Sequencing Informàtics". Wellcome Trust Sanger Institute. Hinxton. Cambridge. UK

5:15 **Paolo Ribeca** – CNAG

5:30 **Nicola Vitulo** – Group: "CRIBI". University of Padova. Italy

5:45 **Victor Solovyev** - Group: "Softberry" . Royal Holloway, University of London. UK

6:00 **Xi Wang** -Group: "Tsinghua University". Tsinghua University. Beijing. China

6:15 **Jan Prins** – Group: "MapSplice". University of Kentucky and University of North Carolina at Chapel Hill. USA

6:30 **Sheng Li** - Group: "Mason Lab" Weill Cornell. New York. USA

6:45 **Alexander Dobin** – Group: "Gingeras lab". Cold Spring Harbor Laboratory. New York. USA

7:00 **Gunnar Rätsch** – Group:" Rätsch lab". Friedrich Miescher Laboratory of the Max Planck Society. Tübingen. Germany.

7:15 **Thomas Wu** – Group: "Genentech". Genentech. San Francisco. USA

7:30 **Free time**

Thursday, April, 7th

9:00 **Lecture "Simulating RNA-Seq Data for Benchmarking Alignment Algorithms".** Greg Grant – University of Pennsylvania. Philadelphia. USA

9:40 **RGASP 3 Evaluation and predictions.**

11:00 **Coffee break**

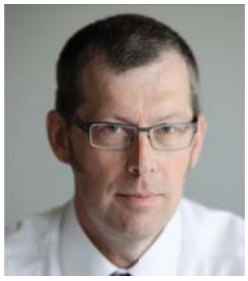
11:30 **RGASP 3 Evaluation and predictions**

1:00 **Lunch**

2:00 **Validation discussion on:** a) Plans for experimental work b) Possible publication c) New editions

	Tuesday, April 5th		**Wednesday, April 6th**		**Thursday, April 7th**
8:45	Registration				
9:00	Presentation	**9:00-11:00**	Results of Evaluation Tyler Alioto/Andre Corvelo	**9:00**	Greg Grant
9:30-11:00	Introduction to dnGASP/RGASP Ivo Gut/Roderic Guigo Report on GAW/Assemblathon Tyler Alioto/Paolo Ribeca Evaluatiion methods and overview of results Tyler Alioto/Andre Corvelo			**9:40-11:00**	Evaluation of predictions
11:00	Coffee break	**11:00**	Coffee break	**11:00**	Coffee break
11:30-1:00	11:30am-12:10pm Daniel Zerbino 12:10pm-12:35pm Shaun Jackman 12:35pm-1:00pm Zemin Ning	**11:30-1:00**	11:30am Discussion 12:30pm pmThe way ahead Ivo Gut (CNAG)	**11:30-1:00**	Evaluation of predictions
1:00	Lunch	**1:00**	Lunch	**1:00**	Lunch
2:00-2:40	Li Yingrui	**2:00-3:00**	Tom Gingeras	**2:00-3:30**	Validation Discussion on: a) Plans for Experimental Work b) Possible Publication c) New Editions
2:40-4:00	2:40 Bryan Downie(Mathias Platzer) 3:05 Victor Solovyev (Softberry) 3:30 Rayan Chikhi (Symbiose)	**3:00-4:30**	3:00pm Welcome by Roderic Guigó (CRG) 3:05pm Tim Hubbard "Gencode" 3:30pm Jen Harrow RGASP 1 and RGASP 2 4:00pm Paul Bertone "RGASP 3"		
4:00	Coffee break	**4:30**	Coffee break		
4:30-6:00	4:30 Alexey Sergushichev (SPbSU ITMO) 4:55 Ekaterina Khmrameeva (Genoanalytica) 5:20 Jared Simpson (Richard Durbin)	**5:00-7:30**	5:00 Zening Ning (Seq informatics) 5:15 Paolo Ribeca (GEM) 5:30 Nicola Vitulo (CRIBI) 5:45 Victor Solovyev (Softberry) 6:00Xi Wang (Tsinghua University) 6:15 Jan Prins (MapSplice) 6:30 Sheng Li (Mason Lab) 6:45 Alexander Dobin (Gingeras lab) 7:00 Gunnar Rätsch (Rätsch lab) 7:15 Thomas Wu (Genentech)		
6:15	Bus to BSC				
6:45	Visit to BSC				
8:00	Dinner at Restaurant Temporada				
10:00	Bus to the hotel				

Scientific Organizers



Ivo Gut

Director of the CNAG since January 2010.He qualified in Chemistry at the University of Basel in 1985 and obtained a PhD in Physical Chemistry from the same university in 1990. Then he joined the research group of Prof. Irene Kochevar at Harvard Medical School as a research fellow.Between 1993 and 1996 he was research fellow with Dr. Stephan Beck at the Imperial Cancer Research Fund in London. Later, at the Max-Planck-Institute for Molecular Genetics he led a group in the Department for Vertebrate Genomics of Prof. Hans Lehrach. In the 11 years before joining the CNAG he worked at the Centre National de Génotypage of the Commissariat à l'Energie Atomique in Evry, France, first as Head of Technology Development and later as Associate Director under Prof. Mark Lathrop.His research interests are high-throughput nucleic acid analysis, sequencing, SNP genotyping, genomics, genetics, nucleic acid and protein analysis methods (molecular biological techniques and chemical modification), implementation of methods, automation and analysis. He is the author of over 100 research papers, inventor of 24 patents and patent applications, founder of 4 biotech companies (Genom Analytik GmbH, Biopsytec GmbH and Epigenomics AG, Integragen SA). He is the coordinator of the 12M€ EU FP7-funded Integrated Project READNA on DNA sequencing technology development.

Roderic Guigó

Program Coordinator of the CRG and Bioinformatics Professor at University Pompeu Fabra. PhD from the Universitat de Barcelona. He worked in the development of mathematical and computer models in Population Genetics and Evolutionary Ecology. He stayed in the Molecular Biology Computer Research Resource at the Dana Farber Cancer Institute--- Harvard University (Division of Biostatistics). Postdoctoral fellow with Dr.Temple F. Smith. BioMolecular Engineering Research Center at Boston University. During these years, he was involved in several projects in the field of sequence analysis: gene identification, automatic knowledge extraction from biosequence databases, protein sequence pattern analysis, and molecular evolution. In Spring 1992, he moved to Los Alamos National Laboratory, where he was a postdoctoral fellow at the Theoretical Biology and Biophysics Group, working essentially on genome analysis related problems: estimation of genome's protein coding density, and characterization of large scale genome structure. Since 1994, researcher at the IMIM. Since 1999 he'salso associated professor with the Universitat Pompeu Fabra. Since year 2005 he is coordinating the Bioinformatics and Genomics program of the Centre de Regulació Genòmica (CRG) in Barcelona.

Invited Speakers

Daniel Zerbino

Postdoctoral researcher at the University of California, Santa Cruz, where he works with David Haussler's group on a variety of sequence analysis problems, such as RNAseq analysis and comparative genomics. Before coming to California he studied engineering at the Ecole Polytechnique (ParisTECH) in Palaisau, France then at the Ecole des Mines de Paris. He then completed PhD studies in Cambridge, UK, under the supervision of Dr Ewan Birney. His doctoral project was centered on the development of the short read de novo assembler Velvet (www.ebi.ac.uk/~zerbino/velvet), and of its RNA sibling, Oases He now maintains the Velvet code and assists developers who wish to develop specific patches or extensions to the program. (www.ebi.ac.uk/~zerbino/oases).

Yingrui Li

Beijing Genomics Institute at Shenzhen. In 2007-2008, he was one of the main bioinformatics analysts in the Yan Huang Project, and the research results "The first Asian genome map" was published on the Nature. Then, Yingrui Li was the first head of scientific systems in BGI and led the team tackling a series of major projects, such as the1000 Genomes Project, China & Denmark diabetes research, in addition to updating of the SOAP package, developing the software of genomic polymorphism, and establishing methods for analyzing whole genome exon sequencing data and Yanhuang 99 and Yanhuang methylation. All of the above research results were published or submitted to Nature, Science and other top international journals. From 2009 to present, he has applied and taken part in 7 projects, including as a director in the National 973 Project. He has published 18 papers with a total SCI Impact Factor of 389.22, nine of which had an Impact Factor greater than 25 and nine for which he was the first author or co-author.

Thomas Gingeras

Professor and Head of Functional Genomics at Cold Spring Harbor Laboratory, New York, USA. Dr. Gingeras joined the faculty of Watson School for Biological Sciences in April 2008 as Professor and Head of Functional Genomics and has been one of two instructors for the Genetics-Genomics course. The general research goals of his laboratory are to investigate the organization of eukaryotic genomes and to study the functional roles of non-protein coding RNAs. Currently his laboratory is mapping the transcriptomes of human and fly genomes. This work is undertaken as part of their participation in the NHGRI funded Encyclopedia of DNA Elements (ENCODE) and model genomes (mod-ENCODE) projects. The aims of each of these projects are to generate comprehensive collection of maps detailing the sites of RNA transcription and to characterize the diverse collection of RNAs produced at these loci providing transcriptional start sites (TSS), transcriptional termination sites (TTS), RNA splices sites, and the possible product-precursor relationships between mapped long (>200 nts) and short (<200 nts) RNA transcripts using RNAseq methodologies.

Tim Hubbard

Graduated with a BA in Biochemistry from University of Cambridge in 1985 and a PhD in Protein Design from the Department of Crystallography, Birkbeck College, London, in 1988. Following a postdoctoral fellowship at the Protein Engineering Research Institute in Osaka under the EU scientific training program in Japan (1989-90) he returned to Cambridge becoming a Zeneca Fellow at the Medical Research Council (MRC), Centre for Protein Engineering. In 1997 he joined the Wellcome Trust Sanger Institute to become Head of Human Genome Analysis. He has been Head of Informatics and a Board of Management member since 2007. Before joining the Sanger Institute Tim worked mainly on protein folding, classification and design. In 1994 he co-founded SCOP (the Structural Classification of Proteins database), thought to be the first biological database designed from the start to take advantage of the then new World Wide Web. He developed algorithms to make protein structure predictions and assess their accuracy and to calibrate the reliability of sequence alignment methods. He was one of the most successful participants in the first CASP (Critical Assessment of Structure Prediction) competition in 1994 and a co-organiser of subsequent CASP competitions (CASP2-CASP7) until 2007.

Greg Grant

Obtained his undergraduate degree in math (1988) and his Ph.D. in math (1995) from the University of Maryland, College Park. His field was Algebraic Number Theory and Arithmetic Geometry. His advisor was Professor David Rohrlich, now at Boston University. Greg spent his last two years as a graduate student visiting the University of Pennsylvania. He worked as a Visiting Assistant Professor in the math department of Johns Hopkins for the academic year 96-97. He has taken a masters in computer science at the University of Pennsylvania (1999), and as of January 1st, 1998 have a 2 and 2/3 year post-doc in the computational biology group at the University of Pennsylvania under Professor Warren Ewens. Starting January 2001 he was funded five years to stay at the University of Pennsylvania Computational Biology and Informatics Laboratory under an NIH career transition grant funded by The National Human Genome Research Institute. Greg Grant is interested in Gene Expression Analysis, Genomic Database Development, Cancer Genomics - particularly aCGH analysis and Next Generation Sequencing Analysis.

Evaluation Committee

- Paul Bertone. EBI. UK
- Nick Goldman. EBI. UK
- Pär Engström. EBI. UK
- Botond Sipos. EBI. UK
- Tamara Steijger. EBI. UK
- Gregory Grant. University of Pennsylvania. Philladelphia. USA
- Tyler Alioto. Centre Nacional d'Anàlisi Genòmica
- Andre Corvelo. Centre Nacional d'Anàlisi Genòmica. Spain
- Emanuele Raineri. Centre Nacional d'Anàlisi Genòmica. Spain
- Simon Heath. Centre Nacional d'Anàlisi Genòmica. Spain

Details of the contents of dnGASP

Participants of dnGASP (*de novo* genome assembly), organized by the National Center for Genome Analysis (CNAG), will have to present analysis methods to generate high quality de novo genome assemblies for large eukaryotic genome from short-read sequence data. Participants will be given a data set (a reference genome) to be processed before 1st of March 2011. This reference genome is an unidentified naturally composed eukaryotic genome of known sequence, with the following characteristics:

Ploidy:	diploid (SNP frequency ~1/1000)
Genome size:	~1.8Gb
Chromosome number:	14
GC content:	~42% (36-50%)
Repeat content:	similar to vertebrate repeat content
Derivation:	Our "genome" sequence is derived from sequence assembled by a traditional combination of WGS and clone-based approaches using Sanger technology, an undisclosed transformation was applied to the sequences to mask their identity, and finally alleles were simulated based on a realistic SNP distribution (SNPs and small indels). The genome additionally contains a minimal amount of purely artificial sequence introduced by the evaluation committee.

Evaluation criteria of the dnGASP

- Use of the standard measures of assembly quality (N50/N90 and the largest/mean/median contig/scaffold size),
- The fact that the reference genome sequence is absolutely known and the position from which each read was simulated is known.
- Assemblies shall be aligned to the reference genome using established alignment methods.
- Level of completeness (e.g. coverage, number of gaps closed, N50 of aligned contigs, etc.)
- Level of correctness (e.g. synteny, accuracy of gap size estimation).
- The ability to bridge different types of repeats (both “naturally” occurring and artificially constructed and embedded) varying in repeat unit size, total length, copy number within the genome and amount of variation.
- Other factors which might impact assembly quality such as variation in GC content or SNP rate.

dnGASP has 9 participating groups from the USA, Canada, China, Russia, UK, Germany and Spain.

List of participants of the dnGASP

1. Shaun Jackman
2. Zeming Ning
3. Yingrui Li
4. Bryan Downie
5. Victor Solovyev
6. Rayan Chikhi
7. Alexey Sergushichev
8. Ekaterina Khamrameeva
9. Jared Simpson

Details of the contents of RGASP3

Participants of RGASP3 (RNASeq read alignment), organised by Paul Bertone (EBI) with input from the Wellcome Trust Sanger Institute and the CRG, will have to present analysis methods to generate high-quality RNASeq read alignments that can be used for efficient transcriptome characterization (transcript discovery and quantitation). Participants will be given a data set to be also processed before 1st of March 2011. The data consists of:

1. Mouse whole brain RNASeq data (David Adams lab, WTSI/UK). Paired-end Illumina 76bp reads, insert sizes 175-225 bp
2. K562 cell line (human chronic myelogenous leukemia) RNASeq data (Tom Gingeras lab, CSHL/USA) a) whole cell, b) nuclear fraction, c) cytosolic fraction Paired-end Illumina 76bp reads
3. Simulated human RNA-seq data (Gregory Grant, University of Pennsylvania) Paired-end Illumina 76bp reads, mean insert size 225 bp. Quality scores are not simulated
4. Mouse whole brain RNASeq data (David Adams lab, WTSI/UK). Paired-end Illumina 76bp reads, insert sizes 175-225 bp
5. K562 cell line (human chronic myelogenous leukemia) RNASeq data (Tom Gingeras lab, CSHL/USA) a) whole cell, b) nuclear fraction, c) cytosolic fraction Paired-end Illumina 76bp reads
6. Simulated human RNA-seq data (Gregory Grant, University of Pennsylvania) Paired-end Illumina 76bp reads, mean insert size 225 bp. Quality scores are not simulated

Evaluation criteria of the RGASP3

- Evaluation will be genome wide using GENCODE annotation.
- Submitted alignments will be processed, in a uniform manner, by RNA-seq analysis tools to assess the impact of mapping performance on detected genes and transcripts.
- Evaluation metrics will be standard gene prediction assessment metrics: sensitivity, specificity and correlation coefficient at the nucleotide, exon, transcript and gene level.
- There will be an evaluation at the level of exon splice boundaries. This will assess the ability of each program to place short reads across splice junctions

RGASP3 has 11 participating groups from United States, United Kingdom, China, Italy, Germany, and Spain

List of participants of the RGASP3

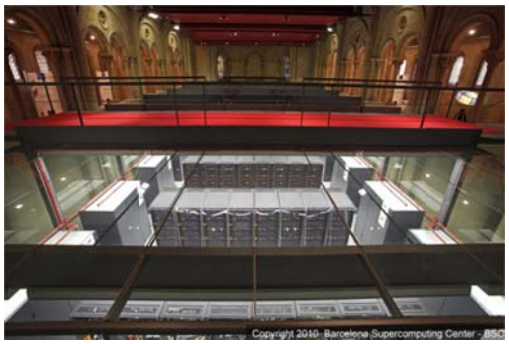
1. Zening NIng
2. Paolo Ribeca
3. Nicola Vitulo
4. Victor Solovyev
5. Xi Wang
6. Jan Prins
7. Sheng Li
8. Alexander Dobin
9. Gunnar Räetsch
10. Thomas Wu
11. Francesca Finotello

Practical information

Venue

Museu Colet
c/ Buenos Aires, 56
08006 Barcelona

Visit to the Barcelona Supercomputing Center



http://www.bsc.es/index.php

c/ Jordi Girona, s/n
08028 Barcelona

HOTEL
BALMORAL
BARCELONA
****

Hotel Balmoral
Via Augusta, 5
08006 Barcelona
Phone: + 34 93 93 217 87 00

Contact person during the event

Laia Arnal
Project Director
International Center for Scientific Debate (ISCD)
larnal@biocat.cat
Phone: + 34 93 310 33 10
Mobile: + 34 662 31 55 29

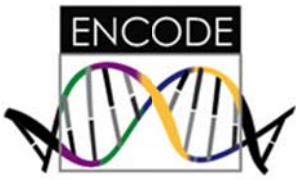
Organizers

The **International Center for Scientific Debate** (ICSD) is an initiative of Biocat, fostered by Welfare Projects "la Caixa" Foundation, which aims to drive top-notch international scientific meetings promoting dialogue, collaboration and open exchange of knowledge among experts of renowned prestige and the Catalan Scientific community. The meetings are global, integrative and multidisciplinary focused helping to tackle social needs in the field of life sciences and health, taking into consideration the complexity and constantly changing conditions of the World. The ISCD also aims to collaborate in the dissemination of knowledge, approaching science to society and contributing to position Barcelona and Catalonia as a city and a country of scientific excellence. More information: http://www.biocat.cat/en/icsd

The **Centre for Genomic Regulation** (CRG) is an innovative centre for basic research created in December 2000 by initiative of the former Department of Universities, Research and Information Society (DURSI) of the Catalan Government. The CRG is legally constituted as a non-profit foundation and has the participation from the Catalan Government through the Innovation, Universities and Enterprise Department (DIUE) and the Health Department (DS), as well as from the Pompeu Fabra University (UPF), and the Spanish Ministry of Science and Innovation (MICINN). It is a unique centre in Spain, based on a non-bureaucratic organization research model, whose objective is to promote basic research in biomedicine and, particularly, in the genomics and proteomics areas. Group leaders at the CRG are recruited internationally and receive support from the centre to set up and run their groups. They are evaluated by an external Scientific Advisory Borad, made up of 10 world leaders in the different areas.

The **Centre Nacional d'Anàlisi Genòmica** (CNAG) was created in late 2009 with support from the Spanish and the Catalan governments. Its vocations is to carry out large scale-projects in genome analysis that will lead to significant improvements in people's health and quality of life, in collaboration with the Catalan, Spanish, European and International Research Community. It is situated in the Barcelona Science Park, one of the largest research clusters in the Life Sciences in Southern Europe.The CNAG was established with the objective of being a center of excellence in genome research with critical mass to support large-scale genomics projects from study design, through data generation, data analysis and data interpretation. It forms the basis for competitive, result-oriented high-throughput genome projects. Current sequencing capacity is in excess of 250 Gbases per day and it is one of the top 5 centers in Europe in terms of 2nd generation sequencing capacity. CNAG's main project is within the framework of the Spanish Contribution to the International Cancer Genome Consortium on Chronic Lymphatic Leukemia.

With the collaboration of:



The National Human Genome Research Institute (NHGRI) launched a public research consortium named **ENCODE**, the Encyclopedia Of DNA Elements (ENCODE), in September 2003, to carry out a project to identify all functional elements in the human genome sequence. The project started with two components - a pilot phase and a technology development phase. The pilot phase tested and compared existing methods to rigorously analyze a defined portion of the human genome sequence.The findings highlighted the success of the project to identify and characterize functional elements in the human genome. The technology development phase also has been a success with the promotion of several new technologies to generate high throughput data on functional elements. With the success of the initial phases of the ENCODE Project, NHGRI funded new awards in September 2007 to scale the ENCODE Project to a production phase on the entire genome along with additional pilot-scale studies. The ENCODE production effort is organized as an open consortium and includes investigators with diverse backgrounds and expertise in the production and analysis of data. This production phase also includes a Data Coordination Center to track, store and display ENCODE data along with a Data Analysis Center to assist in integrated analyses of the data.

The **Wellcome Trust Sanger Institute** is a charitably funded genomic research centre located in Hinxton, nine miles south of Cambridge in the UK. A leader in the Human Genome Project, WTSI is now focused on understanding the role of genetics in health and disease. The WTSI passion for discovery drives its quest to uncover the basis of genetic and infectious disease. The WTSI aims to provide results that can be translated into diagnostics, treatments or therapies that reduce global health burdens. The research at the Wellcome Trust Sanger Institute builds understanding of gene function in health and disease as well as creating resources of lasting value to biomedical research. The WTSI studies diseases that have an impact on health globally by investigating genomes. Building on the WTSI past achievements and based on priorities that exploit the unique expertise of the Faculty of researchers, the WTSI will lead global efforts to understand the biology of genomes. The Wellcome Trust Sanger Institute is convinced of the importance of making this research available and accessible for all audiences.

The **European Bioinformatics Institute** (EBI) is a non-profit academic organisation that forms part of the European Molecular Biology Laboratory (EMBL). The EBI is a centre for research and services in bioinformatics. The Institute manages databases of biological data including nucleic acid, protein sequences and macromolecular structures.As we move towards understanding biology at the systems level, access to large data sets of many different types has become crucial. Technologies such as genome-sequencing, microarrays, proteomics and structural genomics have provided 'parts lists' for many living organisms, and researchers are now focusing on how the individual components fit together to build systems. The hope is that scientists will be able to translate their new insights into improving the quality of life for everyone. However, the high-throughput revolution also threatens to drown us in data. There is an ongoing, and growing, need to collect, store and curate all this information in ways that allow its efficient retrieval and exploitation.

The READNA (REvolutionary Approaches and Devices for Nucleic Acid analysis) consortium includes projects to accelerate new breakthrough DNA sequencing technologies and methods to enhance existing analysis methods. The consortium includes 16 European partners from both academia and industry, and launches with a €12m grant via the European Commission's seventh Framework Programme (FP7). Participants come from a diverse range of scientific disciplines. The interdisciplinary nature of the consortium will allow the exploration of novel concepts of nucleic acid analysis. Via the READNA project, results are expected to improve many aspects of patient care including the development of new personalised medical strategies and treatments. The ultimate aim will be to progress technologies enabling sequencing of an entire human genome for 1000€ in less than one day.

Sequence Mapping and Assembly Assessment Project

RGASP3 / dnGASP Workshop

Venue:

Museu Colet
c/ Buenos Aires, 56 Barcelona – Spain

Organizers:

biocat International Center for Scientific Debate

Initiative fostered by:
Welfare Projects "la Caixa" Foundation

CRG Centre de Regulació Genòmica

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

With collaboration of:

EMBL-EBI

wellcome trust sanger institute

READNA

ENCODE