NGS2009

Conference on Next Generation Sequencing: Challenges and Opportunities

http://ngs2009.uab.es/

SCIENTIFIC COMMITTEE

Miguel Pérez-Enciso (President)
ICREA - Universitat Autònoma de Barcelona.
Bellaterra, Spain

Gabor Marth
Boston College, MA, USA

Gil McVean
University of Oxford, UK

Arcadi Navarro
ICREA - Institut de Biologia Evolutiva (UPF-CSIC)
Barcelona, Spain

Sebastián E. Ramos-Onsins
Centre for Research in Agricultural Genomics (CRAG)
Bellaterra, Spain

Carles Pedrós-Alió
Institut de Ciències del Mar, CSIC
Barcelona, Spain

David Torrents
ICREA - Barcelona Supercomputing Center (BSC).
Barcelona, Spain

Larry Schook
University of Illinois
Urbana - Champaign IL, USA

UAB Esdeveniments

ORGANIZING COMMITTEE

Miguel Pérez-Enciso (President)
ICREA - Universitat Autònoma de Barcelona (UAB).
Bellaterra, Spain

Sebastian Ramos-Onsins
Centre for Research in Agricultural Genomics (CRAG)
Bellaterra, Spain

Arcadi Navarro
ICREA - Institut de Biologia Evolutiva (UPF-CSIC)
Barcelona, Spain

David Torrents
ICREA - Barcelona Supercomputing Center (BSC).
Barcelona, Spain

CONGRESS SECRETARIAT

Teresa Ibáñez
Ultramar Express Event Management
Barcelona, Spain

AGENCIA DE PROMOCIO D'ACTIVITATS I DE CONGRESSOS

Alexandra Garcia
Universitat Autònoma de Barcelona.
Bellaterra, Spain

Why are we here?

The impact of massive parallel sequencing in Biology is difficult to overstate. It is a revolution comparable to the one that followed Sanger sequencing forty years ago. It is not just a dramatic increase in sequencing speed; it means a change in paradigm that obliges researchers, institutions and funding agencies alike. It is also a tremendous and passionate technological race worth millions. No one can now be sure of whether any of the extant technology will prevail in the future or will be replaced by new technologies. With the NGS market growing at an exponential rate by now, fall 2009 is an excellent timing to hold a meeting like this. Next generation sequencing is *this* generation sequence, and has only just begun.

Every single corner of Genetics is being affected by these technologies. To name just two relevant issues: What is the future of GenBank? Unless storage and internet access technology is dramatically scaled up, it seems difficult to believe that all sequences will continue to be deposited in such a 'classical' way. Despite ongoing efforts like the Short Read Archive (SRA) at NCBI, it seems clear that some sort of distributed storage around the globe must be envisaged and implemented following agreed protocols. A second equally relevant issue: What is the future of large sequencing centers? Small consortia or individual labs can now deliver amounts of data that only large centers could afford just a few months ago. Therefore, should funding agencies continue to support them or are better ways to optimize how money is allocated? Large scale sequencing *per se* does not seem to justify any more the existence of dedicated large scale genomic centers. All in all, NGS technology will foster a multipolar research, where dynamic nets should replace a hierarchical world.

This meeting provides an overview of some of the most pressing and rapidly developing areas that utilize NGS. The first talks will provide a general overview of how ultrasequencing has revolutionized Genomics, from cancer research to the 1000 thousand genome initiative together with less studied species like Giant Panda or the sugar beet. Next, crucial bioinformatics and population genomics topics will be reviewed.

Bioinformatics and analytical tools have been recognized, for a reason, as the main bottleneck of current and forthcoming technologies. Therefore, the reduction in costs can be misleading: computer and bioinformatics analyses will be much more expensive in the future than they are now. This leads to another change in paradigm, every PhD student in genomics should have a strong training in bioinformatics. Programming skills are more needed than ever at a time where available software can not fulfill all specific needs for every problem. Although many efficient software and algorithms are published on a weekly basis, daunting challenges await us. Some of the most pressing ones may be to perform complex genome de novo assembly with short reads and identification of all kind of polymorphisms, not only SNPs but also structural variants. How to assess meaningful quality scores that can be applied across platforms and the reliability of polymorphisms uncovered is also fundamental.

The final sections of the meeting review three more topics that have been dramatically reshaped by ultrasequencing: functional, meta- and epigenomics. NGS has promised to substitute microarray expression technology soon and has allowed to uncover an intriguing variety of non coding transcripts, making the understanding of transcription regulation a much more difficult topic than anticipated. Epigenetic changes to DNA that are responsible for transcription regulation can now be studied in detail but genome wide using ChIP-Seq. In the future, one should be able to follow the coevolution of all species making up a whole ecosystems and not only of single individual species. In this meeting we will learn two of the most ambitious initiatives for marine ecosystems and how NGS has made them possible.

But after all the fuss, do we *really* need so much more sequence? Many studies will certainly trade less sequence for a more targeted approach. For some population genomics studies it is far more relevant to have, say, 10 Mb from 100 individuals than a 1 Gb from a single individual. Although several commercial approaches are already available for sequence capture, its performance is not still contrasted across species and genome regions. They are also platform dependent. More importantly sequence capture increases costs significantly. This is one of the areas where we should witness rapid improvements soon. Provided sequencing a whole genome remains more expensive than targeted resequencing ...

To finish, we are extremely grateful to all invited speakers who agreed to participate and thus making this meeting an unrivaled scientific event. A. Navarro, D. Torrents, C. Pedròs, G. Marth and L.B. Schook helped to arrange the scientific programme. We are also profoundly indebted to all public agencies and companies who supported this meeting. Special thanks are due to *Genoma España* and to the three hardware vendors, Roche, Illumina and Applied, who trusted this risky initiative from the very beginning. Thanks also to *Ministerio de Ciencia e Innovación* (MICINN), *Universitat Autònoma de Barcelona* (UAB), *Red Española de Diversidad Biológica, Evolución y Sistemática* (REDES), K-Biosciences, Agilent, *Sistemas Genómicos* and BaseClear.

Enjoy!

Miguel Pérez-Enciso (ICREA-UAB)
Sebastián Ramos-Onsins (CRAG-UAB)
Barcelona, Spain

PROGRAM

THURSDAY 1st OCTOBER

Page

19h00 – GROUP PICTURE

19h15 – WELCOME COCKTAIL AND POSTER VIEWING

FRIDAY 2nd OCTOBER MORNING

Page

9h30 – 11h00: Population Genomics (Chair: A. Navarro)

11h00-11h30 Coffee break

11h30 – 13h00: Population genomics (Chair: S. Ramos-Onsins/DeLorenzo)

13h00 – 15h00 LUNCH AND POSTER VIEWING

15h00 – 16h30: Functional genomics (Chair: J. Betranpetit)

17h00 – 18h30: General and functional Genomics - II (Chair: M. Bink)

20h30 – CONGRESS DINNER

SATURDAY 3rd OCTOBER

POSTER LIST

(Only shown the Corresponding Author)
Please place your poster in the panel with assigned number

Page

General genomics

Bioinformatics and Population Genomics

Transcriptomics and metagenomics

ORAL COMMUNICATIONS

INAUGURAL SPEECH

Sequencing the Cancer Genome (Inaugural speech)

Richard K. Wilson
Washington University, MO, USA

New technology recently has facilitated the complete sequencing of individual human genomes. As the cost and efficiency of this approach continues to improve, we can envision a powerful new means for the study of genes and other genome elements and mechanisms that underlie cancer and other human diseases. I will discuss some of the discoveries made to date with emerging genome sequencing technologies, and how these methods will allow us to better understand both basic biology and human disease. I will focus my remarks on our recent work in acute myeloid leukemia and breast cancer.

GENERAL GENOMICS

Informatics tools for next-generation sequence analysis

Gabor T. Marth
Boston College Biology Department Chestnut Hill, MA, USA

Next-generation DNA sequencing technologies can now produce tens of gigabases of useful data per single run, allowing the discovery of single-nucleotide polymorphisms (SNPs), short insertion-deletions (INDELs), and structural variations (SVs) e.g. translocations, inversions, deletions, and duplications.

The vast throughput of the new sequencing machines already enabled large human sequencing projects, most notably the 1000 Genomes Project. Because of the swift evolution of sequencing technologies, and the rapid scale-up in data throughput software tools for next-generation sequence analysis are still in a state of flux. The 1000 Genomes Project has become a driver for the construction of efficient software pipelines, and the development of appropriate data formats.

This presentation will focus on the informatics aspects of high-throughput human sequencing projects: read mapping, base quality score calibration, SNP / short-INDEL calling, and structural variation (SV) discovery. We have recently developed a complete informatics pipeline to address these questions (MOSAIK read mapper, GIGABAYES SNP/short-INDEL caller, SPANNER SV detection program, GAMBIT alignment viewer). Using large subsets from the 1000 Genomes Project pilot data we will demonstrate how such algorithms are being applied to analyze large human datasets of many hundred samples, extracting high-confidence genetic variants ranging from single base pair to larger genomic events.

Invited Speaker

The 1000 Genomes Project

Gil McVean
Department of Statistics, Oxford, UK

The 1000 Genomes Project is an ambitious programme to provide the genome sequence of over 1000 individuals from across the world, cataloguing all types of genetic variation down to a frequency of 1%. Combining the strengths of multiple next-generation sequencing technologies, with statistical methods that exploit the structure of human genetic variation, we aim to provide a resource that will enhance the power of genome-wide association studies and provide novel insight into our evolutionary history. The project has completed three pilot studies and aims to have data collection complete in 2009. I will describe what we have learnt from the project about population-scale genome sequencing, the use of the project data in medical genetics and what these data have told us about fundamental evolutionary processes including recombination, mutation and natural selection.

Invited Speaker

Plant genomics in the era of high-throughput sequencing: The case of the sugar beet

Heinz Himmelbauer[1,2], Juliane C. Dohm [1,2], Cornelia Lange [2], Ana P. Vivancos [1], Ester Castillo [1], Maik Zehnsdorf [1,2], André Minoche [1,2], Thomas Kraft [3], Markus Wolf [4], Britta Schulz [5], Daniela Holtgräwe [6], Bernd Weisshaar [6]

[1]Centre for Genomic Regulation (CRG), UPF, Barcelona, Spain; [2]Max Planck Institute for Molecular Genetics, Berlin, Germany; [3]Syngenta Seeds GmbH, Bad Salzuflen, Germany; [4]Strube Saatzucht, Söllingen, Germany; [5]KWS Saat AG, Einbeck, Germany; [6]University of Bielefeld, Bielefeld, Germany

Sugar beet (Beta vulgaris) is a diploid species encompassing 18 chromosomes (2n=18) and a haploid genome size of ~800 Mbp. The genome sequence for sugar beet is needed to fully exploit the species' value for evolutionary genomics and as a crop plant. Presently, the completed genomes of representatives of six different genera of flowering plants are available, i.e. from Arabidopsis, poplar, grapevine, papaya, sorghum and rice. Sequencing of other plant genomes is underway, including maize and other cereals, Solanum (potato, tomato), Lotus, and evolutionary model species such as Mimulus and Aquilegia. Since sugar beet is not a close relative to one of the mentioned taxa, its genome sequence will provide essential information on plant genome evolution. The sequencing strategy of the BeetSeq project is chiefly based on a whole-genome shotgun approach, using next-generation sequencing (NGS) technologies. Long-range continuity of scaffolds is established by the integration of Sanger end sequences from bacterial artificial chromosome (BAC) and fosmid clones, as well as paired-end reads with different insert spans (2.5 kb; 4.5kb; 20 kb) generated on the Solexa and 454 NGS platforms. The sequencing effort focuses on the doubled haploid line KWS2320. Many resources are already available for this genotype, including a sugarbeet physical map in BAC clones, BAC end sequences, cDNA sequences, and linkage maps with KWS2320 as one of the parents. Second, the intraspecific variation between sugar beet accessions is rather high. In a pilot genomic sequencing project, sugar beet BAC sequences from two haplotypes revealed extensive unalignable sequence tracts (regions of high divergence; indels) that comprised 10% of either sequence. A large proportion of such indels could be attributed to haplotype-specific integration of transposable elements.

Whole Genome profiling: a new method for Sequence Based Whole Genome Physical Mapping

van Orsouw, NJ [1], van Oeveren, J [1], de Ruiter, M [1], van der Poel, H [1], Kelder, M. [1], Stormo, K [2], Bogden, R [2], van Eijk, MJT [1] Prins, M [1]
[1]*Keygene N.V., Agro Business Park 90, Wageningen, The Netherlands;*
[2]*Amplicon Express, 2345 NE Hopkins Ct., Pullman, WA, USA*

Whole genome sequences are a very important tool to identify the genes that are responsible for important traits. However, the investments necessary to develop a comprehensive whole genome physical map and corresponding sequence assembly are economically unfeasible for many of these organisms. Therefore we have developed Whole Genome profiling (WGP) a new cost effective method to construct high quality sequence-based physical maps. Using the Illumina Genome Analyzer II, such maps are constructed by sequence-based fingerprinting of a 10 GE BAC library. These clones are pooled in a multi-dimensional format, followed by sequencing of 30 bp tags spaced 2-3 kb across each BAC clone. Subsequently the BAC clones are ordered into contigs using these sequence-based anchor points. The availability of a sequence-based map dramatically increases the efficiency of Whole Genome Sequencing (WGS) of the organism of interest. Following proof of principle in Arabidopsis (125 Mb), we have successfully applied WGP in melon (450 Mb) in combination with WGS, and are currently constructing BAC maps for a growing list of plants with varying genome sizes and ploïdy levels. Initial results indicate that WGP is also applicable to larger genomes (e.g. various 2.6 Gb crops) and can be used for comparative mapping between species. Clearly WGP offers an array of applications towards identifying and characterizing economically important genomic regions or genes in a wide range of plant and animal genomes.

Keygene N.V. owns patents and patent applications covering its whole genome technologies.

Comparative Genomics of Tropical Evergreen Fagaceae

Kua CS [1] and Cannon CH [1,2]
[1]Key Lab for Tropical Ecology, Xishuangbanna Tropical Botanic Garden, Chinese Academy of Sciences, Menglun, China; [2]Department of Biological Sciences, Texas Tech University, Lubbock, TX, USA

Evergreen Fagaceae is a diverse and ecological dominant group in tropical Southeast Asia. They are often used in the pollen record to indicate cooler wet climates in the past and their current distribution indicates that they cannot tolerate seasonal climates. For these reasons, the family is a good model group for the historical distribution of evergreen tropical forests in Southeast Asia and the future response of these trees in global climate change. Additionally, the evolution of fruit morphology is under strong selection pressure from seed predation and dispersal, including the repeated convergent evolution of a novel fruit type. Hence, to study the evolution and population dynamics of tropical forest trees, as well as elucidating clues of historical biogeography and the resulted climate change in Southeast Asia and Indochina, we have performed whole genomic shotgun sequencing of 11 species across the Fagaceae family (8 species of Lithocarpus, 2 species of Castanopsisand 1 species of Trigonobalanus) on the Solexa Illumina platform. We evaluated several de novoassembly strategies, given different amounts of data. We explored the classification of the resulting de novocontigs in terms of functional, repetitive elements, and informativeness for comparative ecological and evolutionary analysis.

COMPUTATIONAL CHALLENGES

Invited Speaker

Upcoming Challenges for Multiple Sequence Alignment Methods

Cedric Notredame
Center for Genomic Regulation. Barcelona. Spain

The assembly of multiple sequence alignments is one of the most common tasks in biology. Interestingly, the methods used for reconstructing these alignments are merely heuristics, and more or less well characterized from a biological or a mathematical point of view. The recent increase in the volume of available data is stressing these methods to their limits and the time has come to explore new strategies for coping with the outcome of high-throughput projects (structure, genomics, mass-spec, etc…). During this talk, I will review the current challenges and limitations of existing methods. I will discuss in detail consistency based methods and show why they could constitute the framework of data integration, at all possible levels. Among other things, I will discuss the integration of alternative methods (mcoffee), the integration of sequence and structure based methods (Expresso) and the alignment of RNA sequences using secondary structure information. All the methods presented here are freeware open source that can be either directly accessed or downloaded from the T-Coffee web server (www.tcoffee.org).

Invited Speaker

Next-generation data analysis

Philip Green
Department of Genome Sciences. University of Washington, WA, USA

I will describe the analysis pipeline we are developing for the Illumina GAII sequencer, including a new software package 'next_phred' for image analysis, basecalling & quality assignment, a new program 'phaster' for ultrafast quality-aware alignment of reads to a reference genome, and new file formats and software for efficient storage and manipulation of images and processed data. Next_phred's methods are quite different from Illumina's, and in our experience yield ~80-90% more alignable reads for data generated using Illumina's experimental protocols, with about half the error rate and significantly more discriminating quality values. (These figures are based on comparisons to version 1.3 of the Illumina pipeline software; preliminary analyses suggest next_phred yields ~40% more alignable reads than Illumina's recently released version 1.4.) Because next_phred is designed to be robust to variation in cluster density and size, we believe it likely can be used in conjunction with revised cluster amplification protocols to increase yields even further, and are actively exploring this possibility.

Phaster uses a simple word-frequency based strategy to efficiently search reads against an indexed reference genome. It reports mapping quality scores analogous to those in Maq (Li et al. 2008), and our preliminary tests suggest it is comparable in speed to Bowtie (Langmead et al. 2009) on large-memory machines.

Invited Speaker

Discovering INDEL and Copy Number genomic variation from short reads

Michael Brudno
Dept of Computer Science, University of Toronto, Toronto, Canada

High throughput sequencing (HTS) technologies have enabled the inexpensive sequencing of human genomes, and the discovery of some genomic variants from the resulting short read datasets is well underway. In this talk I will present algorithms for discovery of two types of variants from HTS data: smaller indels (<50bp) and copy number variants (CNVs). First, I will describe MoDIL: Mixture of Distributions Indel Locator, a novel method for finding insertion/deletion polymorphisms from paired short reads. We explicitly model each genomic locus as a mixture of two haplotypes, and our method takes advantage of the high clone coverage to identify both homozygous and heterozygous variation, even if the individual clone sizes are unreliable. Analysis of a recently sequenced genome demonstrates that MoDIL accurately identifies indels >= 20 nucleotides. I will then describe a method to predict CNVs from paired short reads. Our method combines information from paired short reads to identify variable regions and depth-of-coverage to predict the true copy count in the donor genome. Together, the two datasets help overcome both sequencing biases of HTS platforms and spurious read mappings. Our method allows forth detection of CNVs within segmental duplications. We use our method to detect CNVs within the same dataset, and make a total of 9909 calls that show high concordance with previously known CNVs in this individual.

Model-Based Quality Assessment and Base-Calling for Second-Generation Sequencing Data

Corrada Bravo, H. and Irizarry, R. A.
Dept. of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Second-generation sequencing technology sequences millions of short fragments of DNA in parallel and can be used to assemble complex genomes for a fraction of the price and time of previous technologies. However, these data present unprecedented challenges in statistical analysis. For instance, analysis operates on millions of short nucleotide sequences, or reads, which are the result of complex processing of noisy continuous fluorescence intensity data. Large variation in the processing quality of the reads results in infrequent but systematic errors that we have found to mislead downstream analysis in some applications. For instance, a central goal of the 1000 Genomes Project is to quantify variation at the single nucleotide level. At this resolution, small error rates in sequencing are significant. Modeling and quantifying the uncertainty inherent in the generation of sequence reads is essential. We present a simple model to capture uncertainty arising in the base-calling procedure of the Illumina platform. Model parameters have a straightforward interpretation allowing for informative and easily interpretable metrics that capture the variability in sequencing quality. In contrast to other proposed methods for improved base-calling in the Illumina platform, our model provides informative estimates readily usable in quality assessment tools while retaining base-calling performance. We present result on a few applications where our methods improves the robustness of statistics required for downstream analysis.

Sequencing, Sequencing and Sequencing

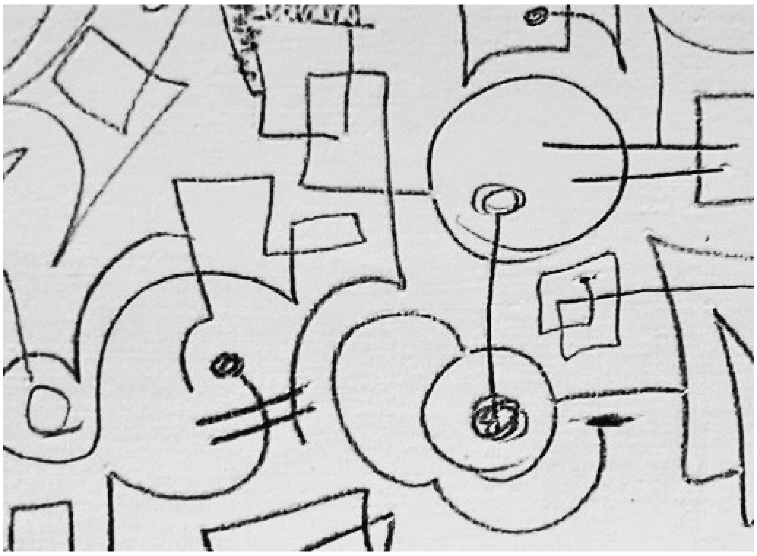
Bicheng Yang and Jun Wang
Beijing Genomics Institute-Shenzhen, Shenzhen, China

Breathtaking progress in DNA sequencing technology has made the costs dropping and throughput increasing in a lighting speed. With more organisms including human sequenced, flood of genetic data is being generated worldwide every day. Progress in genomics has been moving incrementally due to this revolution in sequencing technology. Having a reference genome for certain organism becomes mundane and more different strains are now being sequenced to discover the variations related to certain trait in that species. At the same time, large scale studies in exomics, metagenomics, epigenomics, and transcriptomics all become realistic suddenly. Not only do these studies provide the knowledge to basic research, but also immediate benefits to application. Scientists across many fields are utilizing these data for the development of better crops and livestock; for diagnostics, prognostics and therapies for cancer, neurological disorders and other complex diseases.

BGI is on the cutting edge of translating genomics research into molecular breeding and disease association studies with belief that agriculture, medicine, drug development and clinical treatment would eventually enter a new stage with the understanding of genetic components of all the organisms.

We are dedicating to two projects: “Tree of life” aims to sequence all economically and scientifically important plants/animals and model organisms. The project is best represented by the newly sequenced Giant panda and cucumber. The other one, “gene and health” is focusing on large scale population studies and association studies such as 1000 genomes project and Sino-Danish diabetes project, using whole genome or whole exome sequencing strategies. Working with collaborators in the scientific, agricultural and medical communities, we believe tremendous contributions could be made to quickly and effectively move the translational process.

POPULATION GENOMICS



Invited Speaker

Maximum-likelihood estimation of population-genetic parameters from high-throughput sequencing data

Michael Lynch
Indiana University, Bloomington, IN, USA

High-throughput sequencing strategies will soon lead to the acquisition of high-coverage genomic profiles of hundreds to thousands of individuals within species, generating unprecedented levels of information on patterns of nucleotide heterozygosity and linkage disequilibrium and on the frequencies of nucleotides segregating at individual sites. While offering unprecedented power for the acquisition of population-genetic parameters, these new methods also introduce a number of challenges, most notably a need to account for the sampling of alternative parental alleles at individual nucleotide sites and the introduction of spurious variation by read errors. To minimize the effects of both problems and to avoid ad hoc decisions on data utilization and error rates, we have begun to develop maximum-likelihood methods that generate unbiased and nearly minimum-variance estimates of a number of key parameters, including average nucleotide heterozygosity and its variance among sites, the pattern of decomposition of linkage disequilibrium with physical distance, the rate and molecular spectrum of spontaneously arising mutations, and the allele-frequency spectrum for segregating polymorphisms. These methods define the limits to our ability to estimate population-genetic parameters, while also serving as a platform for identifying optimally efficient experimental designs, e.g., the tradeoff between numbers of individuals sampled and depth of sequence coverage per individual. A general description of the methodology will be presented, along with numerous applications to fully sequenced genomes, including those of several humans.

Invited Speaker

SNP discovery and analysis of selective sweeps using massive parallel short-read sequencing

Martien AM. Groenen [1], H-J Megens [1], RPMA Crooijmans [1], AJ Amaral [1], L Ferretti [2] and LB Schook [3]
[1]Wageningen University, Animal Breeding and Genomics Centre, Wageningen, The Netherlands; [2]Universitat Autonoma Barcelona, Bellaterra, Spain; [3]University of Illinois, Institute for Genomic Biology, Urbana, IL, USA

The Illumina Genome Analyzer (GA) and Roche 454 FLX 'next generation' sequencing platforms were used for SNP identification in a variety of species. Different approaches were used for species whose genome is available (pig, chicken) and for species currently lacking a reference genome (turkey, great tit, tilapia, duck). The high sequence depth allowed accurate SNP identification and estimation of minor allele frequencies resulting in the design of informative iSelect 60K SNP Beadchips for pigs and chicken. E.g. in pigs the number of new SNPs identified exceeded 375,000, which demonstrated that the identification of large numbers of novel SNPs is now feasible in a highly efficient manner. The overall confidence of the SNPs identified by this approach using the porcine SNP60 beadchip demonstrated > 95% of the predicted SNPs were validated. We also used the GA sequences to identify footprints of selection in the porcine and chicken genome. In the pig, sequences generated from pooled Reduced Representation Libraries and covering approximately 2% of the genome of different breeds, were used to estimate nucleotide diversity across chromosomes. Signals of positive selection were identified and a GO term/ KEGG pathway enrichment analysis clearly showed the different traits under selection in domesticated and feral pigs.

Population genomics from individual and pool sequencing

Ferretti Luca[1], Ramos-Onsins Sebastian E.[1,2] and Pérez-Enciso Miguel [1,3]
[1]Departamento de Ciencia Animal y de los Alimentos, Facultad de Veterinaria. Universidad Autónoma de Barcelona, Bellaterra, Spain;
[2]Centre de Recerca en Agro-Genómica (CRAG), Bellaterra, Spain;
[3]Institut Catalá de Recrea i Estudis Avançats (ICREA), Barcelona, Spain

Next generation sequencing technologies allow us to obtain genome-wide data from samples of natural and domesticated populations. However, the cost of high-coverage individual sequencing is still relevant for studies on large samples. The alternative strategies are individual sequencing at low coverage or pool sequencing. Our focus will be on pool sequencing of reduced representation libraries, which are an economic and effective way to detect polymorphism and selection in populations. In this communication we will show how the classical estimators of population genetics have to be modified in order to deal with the uncertainty in the number of alleles that are actually sequenced for a given locus, both for individual sequencing of diploid and polyploid species and for pool sequencing. We will also discuss the challenges that this kind of data represents in terms of data quality and singleton detection and we will show how to correct for these biases. The power of simple tests to detect selection and demographic parameters in these data will be addressed through simulations and compared with the analysis of data from selected regions of the human genome produced by the pilot studies of the 1000 Genomes Project. In the end we will discuss the best choices for the experimental setup of population studies in terms of strategy, coverage, read depth and aim of the project.

A pipeline for studying minor variants in complex genetic populations using long reads from high-throughput sequencing technologies.

Ortega-Serrano, I. [1], Quer, J. [2,3], Rodriguez-Frias F. [3,4], Sánchez, A. [1,5],
[1]Bioinformatics and Statistics Unit, Institut Recerca Hospital Vall d'Hebron, Barcelona, Spain; [2]Liver Unit, Internal Medicine, Hospital Vall d'Hebron, Barcelona, Spain; [3]CIBER enfermedades hepáticas y digestivas del Instituto de Salud Carlos III (CIBERehd), Madrid, Spain; [4]Biochemistry, Hospital Vall d'Hebron, Barcelona, Spain; [5]Dept. of Statistics, Universitat de Barcelona, Barcelona, Spain

High-throughput sequencing technologies have dramatically increased the volume of data obtained at a single experiment, allowing massive analysis in parallel of thousands of sequences and posing new and exciting computational and statistical challenges. Particularly, ultra-deep pyrosequencing performed by the 454 Life Sciences/Roche sequencer has been proposed as a good candidate for detecting clinically relevant minority variants from a complex genetic population. Since it generates longer reads than other platforms (250-400 bases per read) it makes possible the study of wider regions when performing amplicon analysis, providing the chance of observation of interactions among these variants.

A critical issue when studying minor variants in a population is the detection of artifactual changes (errors produced during the amplification and sequenciation processes). We have developed a pipeline to filter some important error sources in order to obtain reliable mutation rate estimates. We have applied this methodology to study the complex viral population obtained from natural isolates from hepatitis B or C chronically infected patients.

Invited Speaker

Population Genomics in the Personal Genome Era

Carlos Bustamante
Dept. of biological Statistics and Computational Biology. Cornell University, (Ithaca) NY, USA.

Invited Speaker

Understanding human genetic variation at the personal and population level through massively-parallel whole-genome sequencing

F. M. De La Vega [1], F. C. L. Hyland [1], S. McLaughlin [2], A.R. MacBride [3], E.F. Tsung [1], H. Peckham [2], C. Scafe [1], C. Lee [2], G. Costa [2], K. Bryc [4], A. Auton [4], C. D. Bustamante [4], M. G. Reese [3], and K. McKernan [2]
[1]Applied Biosystems, Foster City, CA; [2]Applied Biosystems, Beverly, MA,; [3]Omicia, Inc., Emerville, CA, USA; [4]Cornell University, Ithaca, NY, USA

Ultra-high throughput sequencing is becoming a cost-effective method for the analysis of human genomes to discover genetic variation that could have implications in health and disease. We analyzed the SNPs and structural variants from the genomes of five diverse individuals of the HapMap panels; an African-American, a CEPH European, a Mexican, and two Yoruba individuals. Whole-genome sequencing was performed with the Applied Biosystems SOLiD™ System using mate-pair libraries. We identified over 3 million SNPs per individual genome: ~80% are present in dbSNP and the remaining are either novel or personal SNPs. SNPs are under-represented in exons as compared to introns/intergenic regions. Of the coding SNPs, 54% are silent, 45% are missense, and 0.6% are nonsense. We categorized the functions of genes using the Panther ontology, and annotated the damaging potential of non-synonymous SNPs (nsSNPs) using predictions from PolyPhen. About 20% of nsSNPs in this sample are predicted to be damaging. There are fewer damaging SNPs in homozygote than heterozygote state, consistent with the role of purifying selection, this reduction being statistically significant as compared with benign SNP zygosity. We discovered that genes encoding transcription factors, ligases, growth factors, receptors, and RNA helicases are under-represented in the genes with damaging mutations. Further, GPCR genes involved in olfaction, and genes involved in immunity and defense are highly over-represented in the genes with damaging mutations. We identified nsSNP alleles previously associated with human disease (OMIN database), and found very few in homozygous state and none of highly penetrant Mendelian diseases. We report the first individual genomes of admix individuals – this data is likely to be important for future genome-wide association studies in these populations. Admixture change points are derived from SNP allele data leveraging previous genotyping information of the parents of these samples. Recent studies of genetic variation of functional significance in individual genomes have so far mostly focused on SNPs. However, it is becoming clear that structural variation can have functional implication in gene integrity and function. We studied the impact of large indels on gene integrity, by looking for breakpoints regions that overlapped with gene regions. Of the disrupted genes we identified, ~15% are contained in a curated collection of 3,600 human disease genes; the functional or disease impact of these events is currently unknown. Our results suggest that much more genetic variation remains to be uncovered in human populations, in particular structural, which must be considered to obtain a complete picture of their functional impact in individual genome sequences.

Combining Reduced Representation Libraries and Short-Read Sequencing for High-Throughput SNP Discovery in the Absence of Sequenced Genomes.

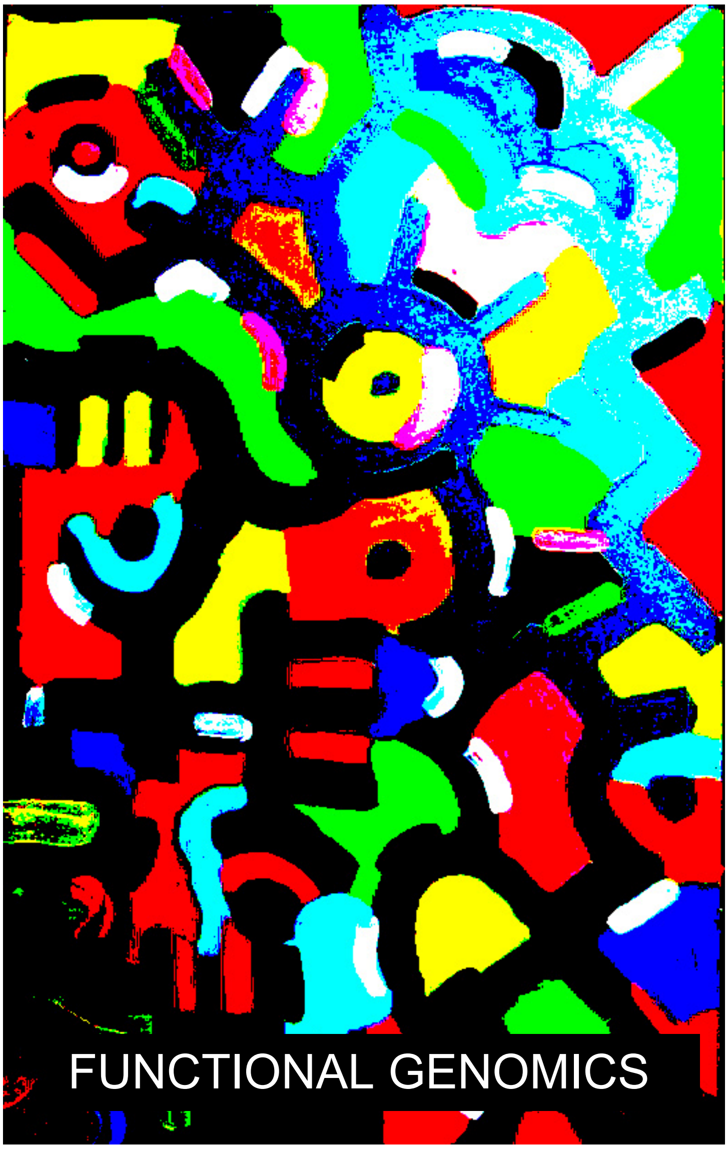
Satkoski, J. [1], Malhi, R. [2], Kanthaswamy, S. [3], Smith, D.G. [1,3]
[1]Dept. of Anthropology, University of California, Davis, CA, USA; [2]Dept. of Anthropology, University of Illinois, Urbana, IL, USA; [3]California National Primate Research Center, University of California, Davis, CA, USA

The completion of the rhesus macaque genome in 2007 led to an explosion of marker discovery and the development of bioinfomatic tools for genetic investigation and the evaluation of human disease models. While the rhesus macaque (Macaca mulatta) is the most common primate model for human disease research, other non-human primates such as cynomolgous macaques (Macaca fasicularis) or baboons (Papio hamadryas ssp.) are increasing in popularity as subjects of biomedical research. However, the lack of genomic information in other non-human primates has delayed comparable methods of research in these species. To this end, we created reduced representation libraries from geographically variable populations of captive cynomolgous macaques and submitted them for short-read sequencing. By reducing genomic complexity in this way, we were able to align the fragments and identify large numbers of SNPs distributed semi-randomly throughout the genome of this species without the need for a complete genome. Using RRLs for each of the five regional populations, we discovered ancestry informative markers (AIMs) for identifying country of origin, markers with high minor allele frequencies in all populations that will be useful for genetic management and markers with low minor allele frequencies useful for disease association studies.

High-Resolution Genome-Wide Mapping of Hermes Transposon Insertion Sites in S. cerevisiae

Mularoni, L [1], Gangadharan, S. [2], Wheelan, S. [1], Craig, N [2]
[1]Oncology Biostatistics and Bioinformatics, Johns Hopkins University School of Medicine, Baltimore, MD, USA; [2]Dept. of Molecular Biology & Genetics, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Analysis of large datasets of transposon insertion sites is useful for understanding the molecular mechanisms that underlie target selection by eukaryotic transposases. We have applied Next Generation Sequencing methods, such as 454 and Solexa, and a ligation-mediated PCR approach to map the sites of insertion of the eukaryotic DNA transposon HERMES in the genome of the bakers' yeast S. cerevisiae. We found that insertions are not uniformly or randomly distributed across the yeast genome but are clustered in specific regions and in certain locations. Here we illustrate the experimental strategy as well as bioinformatics methods used to study the behavior of the autonomous Hermes transposable element in S. cerevisiae; we have extended the consensus target site sequence and analyzed insertion preferences for specific genomic features such as ORFs, TSSs, and nucleosome free regions.



FUNCTIONAL GENOMICS

Invited Speaker

The transcriptional complexity of the human genome: Insights from Next Generation Technologies

Roderic Guigó
Grup de Recerca en Informatica Biomedica (IMIM and UPF), Centre de Regulacio Genomic, Barcelona, Spain.

Transcribed regions have been long been regarded as a distinguishing characteristic of functional portions of the human genome. As part of the Encyclopedia of DNA Elements (ENCODE) project, the sites of transcription in the non-repeat sequences across a representative 1% of the human genome has been determined in a large number of different cell line/tissue samples using of high throughput transcription interrogation technique. In addition, a detailed annotation of the protein coding content of the ENCODE regions has been obtained through a combination of computational, experimental and manual methods. Overall, at least 90% of the ENCODE regions appear to transcribed as primary nuclear transcripts, and about 15% are present as mature processed polyadenylated transcripts. Interestingly up to 30% of these sites of transcription have not been previously identified.

In addition, using a combination of 5'Rapid Amplification of cDNA Ends (RACEs) and high-density resolution tiling arrays, we have systematically explored the transcriptional diversity of protein coding loci. RACE allows detection of low copy number transcripts/isoforms and a high-resolution analysis of genes individually, while pooling strategies and array hybridization permit to reach high-throughput readout. We identified previously unannotated and often tissue/cell line specific transcribed fragments (RACEfrags), both 5' distal to the annotated 5' terminus and internal to the annotated gene bounds for the vast majority (81.5%) of the tested genes. Half of the distal RACEfrags span large segments of genomic sequences away from the main portion of the coding transcript and often overlap with the upstream-annotated gene. 5' most novel detected exons are significantly associated to independently derived evidence of transcription initiation. Notably, more than 50% of the novel transcripts resulting from inclusion of novel exons have changes in their open reading frames. A significant fraction of distal RACEfrags show expression levels comparable to those of known exons of the same locus, suggesting that they are not part of very minority splice forms. These results might revise our current understanding of the architecture of protein-coding genes. They have significant implications for our views on locations of regulatory regions in the genome and for the interpretation of sequence polymorphisms mapping to regions hitherto considered to be "non-coding" ¬†ultimately relating to the identification of disease-related sequence alterations.

Recently, the development of the so-called Next Generation Sequencing Instruments has resulted in an ever increased capacity to interrogate the transcriptional activity of the genome, and in particular to characterize the pattern of alternative splicing characteristic of a given cell type.

Invited Speaker

Using short-read sequencing to dissect allele-specific expression

Andrew G. Clark
Department of Molecular Biology and Genetics. Cornell University, (Ithaca) NY, USA

The large number of independent sequence reads that are obtained from short-read sequencing technologies makes them ideal for problems of tabulating counts of genomic features, and tabulating counts of transcripts of alternative alleles is one with many applications. Sources of distortions of allelic counts, including cis-acting regulatory polymorphism, genomic imprinting, and X chromosome inactivation all can be quantified and distinguished from site-specific cluster biases inherent in the technology. We have applied Illumina sequencing of cDNA in progeny of reciprocal crosses of mice to identify novel cases of genomic imprinting in the mouse fetal and neonatal brain, as well as the placenta. While relatively few novel cases of imprinting were discovered in the brain, the placenta presents a different story, with many instances of both paternal and maternal imprinting. Disruption of imprinting status by interspecific hybridization will be examined by scoring the imprinting status of the equine placenta, as well as that of the F1 hybrids with donkey. Differential allelic expression of X-linked genes in females due to biased X-inactivation is confounded with imprinting and cis-regulatory effects, and a careful decomposition of these effects reveals a significant bias toward inactivation of the paternal X chromosome.

Quantification of Allele-specific Expression Patterns bt GS FLX 454 Technology

Naïra Naouar [1], Remco M. P. Van Poecke [2], Harrie Schneiders [2], Jifeng Tang [2], Antoine Janssen [2], Marcel Prins [2], Michiel J. T. van Eijk [2], Juliette de Meaux [3], Marnik Vuylsteke [1]

[1]VIB Dept. of Plant Systems Biology Ghent University; [2]Keygene NV, Wageningen, The Netherlands; [3]Max Planck Institute for Plant Breeding Research, Cologne, Germany

Differences in gene expression, arising from cis- or trans-regulatory changes, are central in evolution. In order to assess the relative contribution of cis- and trans-regulatory variations to expression differences between the closely related Arabidopsis species, A. thaliana and A. lyrata, we need to differentiate between cis- and trans-changes. One approach involves the quantification of allele specific expression (ASE) in an interspecific cross, such as an F1 hybrid, and subsequent comparison of the allelic with the parental expression ratio. The basic requirement to quantify allele-specific transcripts in a F1 hybrid is a means of identifying the allelic source of the transcript. Single Nucleotide Polymorphisms (SNPs) in the transcripts lend themselves to easy quantify and differentiate the two allele specific transcripts in the hybrid. Here, we provide a first show case that with massive parallel sequencing technology such as 454 GS-FLX sequencing, quantification of allele-specific transcripts in a genome-wide manner comes within reach.

Peak selection coupled with de novo motif identification improves the accuracy of transcription factor binding site prediction in ChIP-Seq data analysis

Boeva, V. [1,2], Surdez,D. [2], Guillon, N. [2], Tirode, F. [2], Fejes, A.P. [3], Delattre, O. [2], Barillot, E. [1]

[1]Lab. of Bioinformatics, Biostatistics, Epidemiology and Computational Systems Biology of Cancer, INSERM Unit 900/Ecoles des Mines/Institut Curie, Paris, France; [2]Lab. of Genetics and Biology of Cancers, INSERM Unit 830/Institut Curie, Paris, France; [3]Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada, USA

ChIP-Seq is an application of next-generation sequencing technologies which makes genome-wide identification of binding sites of DNA-associated proteins possible. This technique is widely used for the identification of transcription factor (TF) binding sites. A number of algorithms for extraction of enrichment regions (peaks) from ChIP-Seq data have been published. However, none of them take into account DNA sequence information, even though, as it was recently shown, direct sites of TF binding always contain binding motifs. Here we present a new motif based algorithm to identify putative binding sites from ChIP-Seq data. The algorithm was implemented as a Java/Perl package called MICSA (Motif Identification for ChIP-Seq Analysis) and is available at http://bioinfo-out.curie.fr/projects/micsa/.

The algorithm, starting from aligned reads, (i) identifies all candidate peaks using the program FindPeaks, (ii) retrieves peaks' DNA sequences, (iii) extracts overrepresented motifs from sequences of a subset of the most significantly enriched peaks, (iv) checks if motifs are present in the remaining peaks and calculates motif p-values, (v) reports peaks without exceeding the user-specified number of false positives.

We compared MICSA to ten published tools for ChIP-Seq data analysis on a dataset for the neuron-restrictive silencer factor and found that MICSA was more accurate in predicting TFBSs.

Invited Speaker

Advanced data Analysis in Targeted Resequencing Projects

Bernd Timmermann[1], Martin Kerick[1], Johannes Röhr[2], Christina Röhr[1], Lars Bertram[1], Kurt Zatloukal[3], Bernhard Herrmann[1], Hans Lehrach[1], Michal-Ruth Schweiger[1,3]
[1]Max-Planck Institute for Molecular Genetics, Berlin, Germany; [2]Dept. of Bioinformatics, Free-University, Berlin, Germany; [3]Dept. of Pathology, Medical University of Graz, Austria

Second generation sequencing techniques allow to sequence large amounts of DNA within a reasonable time-frame. Selective DNA enrichment techniques have become the tools of choice to lower the burden of time and cost even further. This approach is especially useful for the analysis of complex disease such as cancer or as general tool for the analysis of distinct candidate genes for monogenetic diseases.

In the case of cancer, previous studies highlight a number of potential cancer susceptibility genes with consistent risk effects across all published datasets. However more comprehensive experimental approaches are needed to identify novel colon cancer genes, in particular those which contain rare, potentially disease-causing variants.

To this end, we have begun "whole exome" sequencing of colon cancer patients using microarray-based sequence capture of ~180,000 coding regions followed by massively parallel sequencing using the Roche/454 FLX Genome Sequencer. We sequenced to an coverage of 20-fold of the exonic regions. Comparison of the sequence to the reference genome identified on average ~60,000 single nucleotide polymorphisms, approximately 12% of which were not listed in dbSNP. Of all mutations detected about 10% represented non-synonymous substitutions within the coding sequence, while another ~70% were located in adjacent non-coding sequences or micro-RNAs. In addition, we accurately identified ~2,400 small-scale (1-49 base pair (bp)) insertion and deletion polymorphisms.

The main challenge after sequencing is the data analysis and especially the functional annotation and filtering of variants. For this purpose we have developed an automated analysis pipeline for resequencing data. This consists of different functional annotation steps including balance of base specific (phyloP) and multi species conservation. Additionally different project specific filtering steps with theme specific databases are included.

This project represents one of the first systematic assessments of the "whole exome" using next-generation technologies in an attempt to identify novel disease-causing variants in a genetically complex disease.

High throughput sequencing analysis of linkage assay-identified candidate regions in familial breast cancer: methods, analysis pipeline and troubleshooting.

Juan Manuel Rosa-Rosa[1], FJ Gracia-Aznárez[1], Emily Hodges[2], Guillermo Pita[3], Michelle Rooks[2], Greg Hannon[2] and Javier Benitez[1,3]

[1]Human Genetics Group, Spanish National Cancer Research Centre (CNIO), Madrid, Spain; [2]Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, New York, USA; [3]Genotyping Unit (CEGEN), Spanish National Cancer Research Centre, Madrid, Spain

Massive parallel sequencing technology allows the development of studies unaffordable few years ago. However, the analysis protocols have not reached the level of development to extract all the information from the huge amount of data obtained. In this study, we performed high throughput sequencing in two regions located on both chromosomes 3 and 6, recently identified by our group as candidate regions to harbour breast cancer susceptibility genes (Rosa-Rosa et al., Am J Hum Genet, 2009).

In order to enrich the coding region of the 128 described genes located on both candidate regions, a hybrid-selection method was performed as previously described (Hodges et al., Nat Protoc, 2009). We used the DNA from a total of 20 individuals belonging to 9 families (4 families putatively linked to the candidate region on chromosome 3 and 5 families on chromosome 6), and DNA from 4 other unrelated individuals used as controls.

An average of 5.4 million read was obtained per sample, with an average of 37% of the sequences aligned to the target regions. We developed an analysis pipeline based on SOAP aligner to identify candidate variants. An average of 1.800 SNPs were obtained per individual, but only ~1.5% of those passed all the filters to be considered candidates.

Surfing on the surface: mutation detection in human genes coding for cell surface trans-membrane proteins

Parmigiani, RB. [1], Galante, PAF. [1], da Cunha, JPC. [1], Perez, RO. [2], Habr-Gama, A. [2], Gama-Rodrigues, J. [2], de Souza, SJ. [1], Camargo, AA. [1]
[1]Ludwig Institute for Cancer Research, São Paulo Branch, SP, Brazil;
[2]Hospital Alemão Oswaldo Cruz, SP, Brazil

Cell surface proteins are excellent targets for therapeutic and diagnostic interventions. In cancer, both over-expressed and mutated proteins constitute potential targets and therefore different approaches have been applied to find new candidates. Recently, using bioinformatics tools, the whole set of known human genes was searched for trans-membrane (TM) domains, filtering out proteins from other cell membranes and secreted proteins. A catalog of more than 3700 human genes that code for proteins located at the cell surface (called the Human Surfaceome) was generated (Cunha et al, to be published). Here, we used this catalog in a search to identify mutations in samples from metastatic and non-metastatic colorectal tumors. In order to concentrate our sequencing efforts on the catalog of 3700 genes, we took the corresponding genomic coordinates of the coding exons and designed customized capture arrays. The total sequence included on the arrays was 9.2Mb (4.1Mb each). Genomic DNA from eleven paired-samples, consisting of: 1 pool of normal tissues; 4 non-metastatic tumors; 4 metastatic tumors; and 2 metastasis, was captured and used for preparation of 454 shotgun libraries. Sequences were then applied to a bioinformatics pipeline for mutation detection. Our findings will be presented at the conference.

Identification of EMS-Induced Mutations by Whole-Genome Sequencing

Blumenstiel, J. P. [1,2], Gilliland, W. D. [2], Griffiths, J. A. [2], Hawley, R. S. [2,3], Noll, A. C. [2], Perera, A. G. [2], Staehling-Hampton, K. [2], Walton, K. N. [2]
[1]Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas, USA; [2]Stowers Institute for Medical Research, Kansas City, Missouri, USA; [3]Department of Physiology, Kansas University Medical Center, Kansas City, Kansas, USA

The use of chemical mutagens, especially ethyl methanesulfonate (EMS), has become a standard approach for mutagenesis and forward genetic screens. Mapping EMS mutations using standard approaches are difficult and labor intensive. Traditional genetic mapping typically involves three steps. First, identify the chromosome bearing the mutation to be mapped. Second, pinpoint the location of the mutation by genotyping recombinant chromosomes that do or do not carry the mutation of interest and by identifying an association between polymorphisms and the mutation. Finally, sequence the candidate genes to find the causative mutation. This process becomes even more tedious in situations where there are few polymorphic markers or candidate genes. New genome sequencing technologies show tremendous promise in reducing the time needed to identify causative mutations. At Stowers Institute, we have developed a whole-genome sequencing (WGS) approach using Next Generation Sequencing (NGS) technology to identify mutations which eliminates the need for traditional mapping methods. Here, we present our approach from sample preparation to mutation discovery using the Illumina Genome Analyzer.

IntOGen: A novel framework for integration and data-mining of multidimensional oncogenomic data

Gunes Gundem [1], Christian Perez-Llamas [1], Alba Jené [1], Anna Kedzierska [2], Khademul Islam [1], Jordi Deu-Pons [1], Simon J. Furney [1] and Nuria Lopez-Bigas [1]

[1]Research Unit on Biomedical Informatics, Experimental and Health Science Department, Universitat Pompeu Fabra, Barcelona Biomedical Research Park, Spain; [2]Bioinformatics and Genomics program, Centre for Genomic Regulation, Barcelona Biomedical Research Park, Spain

The availability of data from a growing number of oncogenomic studies provides an unprecedented opportunity to understand tumor development from a genomic perspective. However, new integrative methodologies are necessary in order to take full advantage of these valuable data. IntOGen is a novel framework that addresses this by collecting, organizing, analyzing and integrating genome-wide experiments that study several forms of alterations in numerous cancer types. IntOGen explores the data at different levels, from individual experiments to combinations of experiments that analyze the same tumour type, and from individual genes to biological modules, pathways or gene sets.

IntOGen is designed to be an updatable, flexible, extensible and efficient system for integrative oncogenomics analysis and visualization. The system consists of three main components: 1) Data, including oncogenomics experiments and biological modules, coming from next generation sequencing technologies among others. 2) Statistical methods for analysis and integration of the data. A specifically designed statistical framework has been implemented with the objective of identifying driver genes and modules significantly altered in different tumour types. 3) Visualization methods to explore the results in intuitive and efficient ways. We have developed two complementary visualization systems, i) A publicly accessible web system (www.intogen.org) which allows an easy and efficient access to IntOGen results and ii) A standalone Java application, GiTools, which permits a more flexible and sophisticated navigation of IntOGen data and results.

METAGENOMICS

Invited Speaker

Next generation sequencing in marine ecological genomics: tools and applications

Frank Oliver Glöckner

Max Planck Institute for Marine Microbiology, Microbial Genomics and Bioinformatics Group, Bremen, Germany, and Jacobs University Bremen.

Nico M. van Straalen defined ecological genomics as "a scientific discipline that studies the structure and functioning of a genome with the aim of understanding the relationship between the organism and its biotic and abiotic environments". Relative to earlier approaches, application of next generation sequencing (NGS) technologies to ecological genomics produces far more sequence data from extended sample sets. This weaves a much denser network of data promising better foundations for the understanding of ecosystem functioning. Superficially, NGS seems a fast and cheap basis for ecological genomics; however, processing, analyzing and interpreting the flood of data generated is no small burden. Especially true in ocean waters and other high diversity environments, targeted and integrated approaches are needed to turn these data into biological knowledge. This talk will introduce tools and methods which are under development to integrate and interpret phylogenetic and functional information describing marine microbial communities contextualized by environmental parameters. Finally, the MIMAS (Microbial Interactions in MArine Systems) project will serve as an example of this integrative approach: MIMAS will combine diversity analysis with metagenomic, metatranscriptomic and metaproteomic investigations on a sampling series from the Long Term Ecological Research (LTER) site 'Helgoland Roads' to explore the native microbial ecology.

Invited Speaker

Metagenomics versus Next Generation Sequencing Technologies

Douglas B. Rusch
J. Craig Venter Institute, Rockville, MD, USA

The virtual flood of sequence data made possible with the development of new sequencing technologies and the continual improvement of existing technologies poses a number of challenges for the bioinformatics community. These problems are especially acute in the metagenomics community where reference genomes are often unavailable or uninformative, where de novo assembly is problematic, and short reads coupled with high sequence diversity make annotation challenging and computationally expensive. The Global Ocean Survey is the largest metagenomic dataset to date and includes data from all the current major sequencing platforms. The arrival of these new more efficient sequencing technologies have made it possible for us to start exploring in depth the viral, bacterial, and mixed bacterial and eukaryotic samples introducing new challenges to the assembly, classification, and annotation of the data. A variety of strategies are being implemented to accelerate as well as improve our analysis, assembly, and annotation pipelines. These approaches include hardening of our annotation pipelines, the use of more efficient non-sequence similarity based classification algorithms, and redesigned workflows to increase the efficiency of our techniques. We will present the lessons learned and the opportunities for future improvement as metagenomics begins to take advantage of the next generation sequencing technologies.

Large-scale biodiversity analysis through next-generation sequencing

Hajibabaei, M., Shokralla, S., Singer GAC.
Biodiversity Institute of Ontario, University of Guelph, Guelph, Ontario, Canada

Biodiversity analysis is central to understanding ecosystems and monitoring environmental change. Sequencing-based approaches have revolutionized our understanding of species diversity. However, Sanger sequencing is not feasible for large-scale studies of mixed environmental samples, where provisioning the sequencer one sample at a time is not practical. Massive parallel sequencing capability of next-generation sequencers has been used in microbial biodiversity studies through the analysis of ribosomal sequences (i.e. 16S rDNA). Because longer read length is critical for taxonomic assignment of sequences, Roche-454 sequencer, capable of producing ~250 bases (in amplicon workflow) has been preferred for biodiversity studies. We optimized the Roche-454 platform for obtaining standard species-specific mitochondrial sequences, "DNA barcodes" from eukaryotic organisms, especially organisms used for standard biomonitoring. Using this standard marker is important, as reference libraries of DNA barcodes are under construction with 600,000 sequences already available. To facilitate the integration of next-generation sequences in real-world applications such as environmental monitoring, we have developed a bioinformatics pipeline for standard data analysis and visualization. In addition, by developing a comparative phylogenetic profiling approach we have been able to compare biodiversity profiles of various sites using different genetic markers. This work will aid the adoption of next-generation sequencing in large-scale biodiversity studies.

EPIGENOMICS

Invited Speaker

A systems biology view at transcription regulation networks

Henk Stunnenberg
Nijmegen Center for Molecular Life Sciences, Holland

The regulation of gene expression is paramount in growth, development, differentiation, signaling, adaptation to the environment and many other processes. Gene expression is regulated at many levels, but primarily by binding of specific transcription factors to regulatory regions, resulting in the recruitment of activating or repressive factors and subsequent changes in mRNA levels and gene activity. Identification of the target gene and binding site networks of transcription factors is vital to understand its role. The presence or absence of a protein (or histone modification) at a specific genomic location is typically determined using chromatin immunoprecipitation (ChIP). The application of massively parallel sequencing to ChIP (ChIP-Seq) has opened up new avenues at the genome-wide scale to elucidate entire regulatory networks and pathways. So far mostly static views of transcription factor binding has been described, usually restricted to one cell line. The increasing sequence capacity enables for the first time the genome wide identification of transcription factor binding sites, histone marks, DNA methylation as well as RNA polymerase II occupancy and quantitative transcriptome sequencing (RNA-seq) at different time points, conditions and cell lines.

Invited Speaker

Reverse Phenotyping: Towards an integrated (epi)genomic approach to complex phenotypes and common disease

Stephan Beck
UCL Cancer Institute, University College London, London, UK

What determines a phenotype is one of the fundamental questions in biology and medicine. In addition to genetic factors, non-genetic factors such as epigenetic and environmental factors have been shown to play important roles. Of the epigenetic factors, methylation at CpG dinucleotides is the only known biologically relevant epigenetic modification at the DNA level in humans. Knowledge of the methylation status of each of the ~28 million CpG sites (methylome) in the haploid human genome is therefore of great importance for cellular identity, differentiation, development and, if perturbed, for disease aetiology. To understand the rules governing DNA methylation and the consequences if DNA methylation is perturbed requires genome-wide analysis of its temporal and spatial plasticity. Almost 60 years after the discovery of 5-methyl cytosine and about 25 years since the discovery that altered DNA methylation plays a role in disease (particularly in cancer), technologies for methylome analysis have finally become available.

I will present data from our efforts using array- and sequencing-based platforms for high-throughput DNA methylation analysis, discuss some of the lessons learnt and give an outlook on how the data may be used in an integrated approach – termed ‘reverse phenotyping’ – to analyse and better understand the (epi)genomics of phenotypic plasticity in health and disease.

POSTER COMMUNICATIONS

Section General Genomics

P1

Towards the whole sequence of the melon genome

Benjak, A. [1], González, V. [2], Mir, G. [1], Arús, P. [1], Aranda, M. [3], Álvarez-Tejado, M. [4], Droege, M. [5], Du, L. [6], Puigdomènech, P. [2], Garcia-Mas, J. [1]
[1]IRTA, Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB, Departament de Genètica Vegetal, Cabrils, Spain; [2]IBMB Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB, Departament de Genètica Molecular, Barcelona, Spain; [3]CEBAS-CSIC, Murcia, Spain; [4]Roche Applied Science, Barcelona, Spain; [5]Roche Applied Science, Penzberg, Germany; [6]454 Life Sciences, Branford, USA

In the framework of a Spanish Genomics Initiative (MELONOMICS), the 480 Mbp genome of melon (Cucumis melo L.) is being sequenced using a whole genome shotgun strategy with the Roche 454 GS FLX Titanium system. A combination of single reads and paired-end reads of 3, 8 and 20 kb have been performed. In total, 6 X of the melon genome has been sequenced with a final objective of reaching a genome coverage of 20 X. A melon physical map has been constructed after fingerprinting a melon BAC library, and 200 genetic markers have been placed in BACs as anchor points between the physical and the genetic map. More than 40,000 BAC end sequences have also been produced. The first attempts of assembling the melon genome sequence will be presented.

P2

Using 454 sequencing of ESTs for linkage analyses in a dioecious plant species

Bergero, R., Suo, Q., Charlesworth, D

Institute of Evolutionary Biology, Ashworth Laboratories, University of Edinburgh, Kings Buildings, Edinburgh, United Kingdom

The use of dominant, male- and female-meiosis informative markers for linkage studies is often restricted by the inability to combine the derived male and female linkage maps in a single integrated map, for which a number of codominant markers are also required. We constructed a genetic linkage map in the dioecious plant Silene latifolia, first using segregating transposon (MITE) insertions from a single F2 family as dominant markers. To add codominant markers, data from 454 sequencing of the S. latifolia transcriptome were used, and codominant cDNA-based ISVS and SNP markers were developed. A total of 155 codominant markers were obtained. The sex-averaged map integrated with the linkage groups based on MITES spanned a total of 1054 cM, in 12 linkage groups (corresponding to the known number of autosomes, plus the X and the Y chromosome). The average intermarker distance was 2.89 cM. The current X chromosome map spans a total of 79.5 cM, and contains a total of 6 segregating MITE insertions, and 11 genic markers. Two X-linked genic markers, the putative orthologues of A. thaliana AT4G27700 and AT5G52440 loci, appear to cross over with the Y chromosome during male meiosis, and are therefore located in the pseudoautosomal region. Our markers permit comparative linkage analyses by mapping selected orthologous genic markers in the related gynodioecious species S. vulgaris, which lacks a sex chromosome system. This should reveal any major rearrangements during the evolution of the sex chromosomes, allows us to compare recombination rates in X chromosome vs the homologues autosome from a non-dioecious species.

P3

Using a 'framework species' concept for ecological and evolutionary studies in comparative genomics

Cannon. CH [1,2] and Kua, CS [1]
[1]Key Lab in Tropical Ecology, Xishuangbanna Tropical Botanic Garden, Chinese Academy of Sciences, Menglun, China; [2]Department of Biological Sciences, Texas Tech University, Lubbock, TX, USA

Next-gen sequencing provides a revolutionary opportunity to address fundamental ecological and evolutionary questions in natural ecosystems using previously unstudied species. This revolution will be particularly meaningful in the tropics, where species diversity is negatively correlated with scientific knowledge, particularly in regards to genomics. Here, we outline an approach to maximize the impact of next-gen sequencing for both basic and applied issues. This approach uses whole genomic DNA sequence data generated on the Illumina platform, with relatively shallow sequencing (~5x) of several 'framework' species that encompass the major dimensions of phenotypic and geographic variation defined by a particular ecological, evolutionary or management question. Comparative genomics among these framework species can then reveal the most informative genetic elements in the context of the question at hand. In addition to standard assembly-based (reference and de novo) analytic techniques, we have developed novel and conceptually simple assembly-free analyses. This approach is illustrated with three examples. At the population level, we have identified numerous potential markers for use as a DNA fingerprint to determine the geographic origin of wood from an endangered species of tropical timber. At the genus level, by comparing fig species with different sexual systems and growth forms, a remarkable level of genomic divergence is observed, potentially due to the tightly symbiotic nature of their pollination system. At the family level, complimentary sets of putative markers for historical biogeographic and phenotypic adaptation were discovered. We feel that avoiding the 'finished genome' perspective and instead pursuing the shallow sequencing and targeted assembly of informative regions of framework species will allow rapid advance in ecological and evolutionary genomics in the tropics, where these tools are desperately needed.

P4

Partial short-read resequencing of a highly inbred Iberian pig

Esteve A.[1], Kofler R.[2], Vivancos A.P.[2], Himmelbauer H.[2], Groenen MAM [3], Folch JM.[1], Rodríguez MC.[4], Pérez-Enciso M. [1,5]
[1]Departament de Ciència dels Animals i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, Bellaterra, Spain; [2]Ultrasequencing Unit, Centre de Regulació Genòmica, Barcelona, Spain; [3]Wageningen University, Animal Breeding and Genomics Centre, Wageningen, The Netherlands; [4]Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Madrid, Spain; [5]Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain

We have partially resequenced a highly inbred (F = 0.40) Iberian pig sow pertaining to the black hairless Guadyerbas strain. We have employed a reduced representation library strategy. We digested genomic DNA with HaeIII, selected the 160-200 bp range, and generated Solexa fragment libraries omitting pre-amplification. Three lanes were run in the Genome Analyzer II (Illumina) that produced a total gross of 25.3 million 40 nt long reads; 14.1 million reads were retained for further analysis after filtering. We aligned them against the newly released assembly9 using GEM (P. Ribeca, unpublished), MAQ and Mosaik. A total of 5 million of reads matched the assembly unambiguously, spanning 83.1 Mb with at least one read, and 25.1 Mb with at least 3 reads. The average coverage was 4x and was rather uniform across all chromosomes. GEM identified a total of 172,236 SNPs corresponding to 139,853 homozygous SNPs and 32,583 heterozygous SNPs. MAQ detected a total of 65,489 SNPs which 53,947 were homozygous. Gigabayes identified 138.321 SNPs which 115.820 were homozygous. Moreover, it detected 20.315 indels of one base pair. 41,526 SNPs were identified by all three softwares, and 68,778 by at least two. The next immediate goal is to annotate SNPs and identify regions enriched for heterozygous positions in Guadyerbas to look for potential signals of balancing selection.

P5

Genome sequence of the recombinant protein production host Pichia pastoris

De Schutter K. [1,2,7], Lin Y.-C. [3,4,7], Tiels P. [1,5,7], Van Hecke A.[1,5], Glinka S.[6], Weber-Lehmann J. [6], Rouzé P. [3,4], Van de Peer Y.[3,4], Callewaert N. [1,5]

[1]Unit for Molecular Glycobiology, Department for Molecular Biomedical Research, VIB, Ghent-Zwijnaarde, Belgium; [2]Department for Biomedical Molecular Biology, Ghent University, Ghent-Zwijnaarde, Belgium; [3]Department of Plant Systems Biology, VIB, Ghent-Zwijnaarde, Belgium; [4]Department of Plant Biotechnology and Genetics, Ghent University, Ghent, Belgium; [5]Unit for Molecular Glycobiology, L-ProBE, Department of Biochemistry and Microbiology, Ghent University, Ghent- Zwijnaarde, Belgium; [6]Eurofins MWG Operon, Ebersberg, Germany; [7]These authors contributed equally to this work

The methylotrophic yeast Pichia pastoris is widely used for the production of proteins and as a model organism for studying peroxisomal biogenesis and methanol assimilation. P. pastoris strains capable of human-type N-glycosylation are now available, which increases the utility of this organism for biopharmaceutical production. Despite its biotechnological importance, relatively few genetic tools or engineered strains have been generated for P. pastoris. To facilitate progress in these areas, we present the 9.43 Mbp genomic sequence of the GS115 strain of P. pastoris. We also provide manually curated annotation for its 5,313 protein-coding genes..

P6

Mutant HIV minority variants detected by ultradeep sequencing do not condition virological failure in patients starting ARV therapy including low genetic barrier drugs.

Hernández-Novoa B[1], Page C[1], Gutiérrez C[1], Manrique M[2], Pareja-Tobes E[2], Tobes R[2], Moreno S[1]

[1]Hospital Ramón y Cajal, Madrid, Spain; [2]Era7 Information Technologies SLU, Granada, Spain

Background: Clinical impact of mutant HIV minority variants has not been fully established. This study evaluates the participation of mutant-HIV minority variants in response to initial ARV therapy with low genetic barrier drugs (LGBD)..

Methods: Fifteen patients (baseline wild-type standard genotype) initiating ARV therapy with LGBD were selected. A PCR fragment comprising K103N/Y181C/M184V was designed. Primers were tagged per patient allowing ultradeep sequencing in one PicoTiterPlate (454 LifeSciences-Roche). .

Results: All patients presented mutant minority variants to some extent. The mean proportion in which mutations were present was 3.63, 0.00 and 3.63% for K103N, Y181C and M184V, respectively. Undetectable VL was achieved in all cases but one.

Initial ARV therapy	Baseline VL (log)	K103N	Y181C	M184V	Months to VL <1,7 log.
RTV/EFV	4.6	4.96	0.05	5.49	7.0.
AZT/3TC/EFV	5.5	0.91	0.00	1.25	11.0.
ddI/3TC/EFV	5.1	1.45	0.00	1.50	11.0.
ddI/3TC/EFV	5.4	2.82	0.00	3.19	11.0.
AZT/3TC/EFV	3.8	1.14	0.00	0.86	9.0.
3TC/TDF/EFV	5.0	2.38	0.00	2.46	11.0.
AZT/3TC/EFV	5.4	7.37	0.00	5.74	12.0.
3TC/D4T/EFV	4.9	5.97	0.11	5.44	8.0.
ddI/3TC/EFV	5.0	4.43	0.05	4.18	9.0.
AZT/3TC/EFV	5.0	6.41	0.10	6.89	8.0.
ddI/3TC/EFV	3.8	1.64	0.04	2.35	Not achieved.
ddI/3TC/EFV	4.9	0.05	0.00	0.96	4.0.
3TC/ABC/ATV	5.3	2.00	0.00	1.79	6.0.
3TC/TDF/EFV	5.7	0.17	0.00	0.45	10.0.
AZT/3TC/NVP	5.2	12.70	0.04	11.93	10.0.

Conclusions: Although mutant HIV minority variants associated with resistance to LGBD in the initial regimen were detected, virologic failure occurred in only one case. Achievement of undetectable VL was observed in the rest, but with certain delay.

Section Bioinformatics and Population Genomics

P7

Finding selection footprints in the swine genome using massive parallel sequencing

Amaral-AJ.[1], Ferretti L.[2], Megens H-J[1], Crooijmans RPMA[1], Nie H [1], Ramos-Onsins S.E.[2], Perez-Enciso M. [2], Schook L. [3], Groenen MAM [1]
[1]Wageningen University, Animal Breeding and Genomics Centre, Wageningen, The Netherlands; [2]Universitat Autonoma Barcelona – ICREA, Bellaterra, Spain; [3]University of Illinois, Institute for Genomic Biology, Urbana, IL, USA

We investigated whether selection footprints can be identified from GA (Genome analyzer) sequences generated from pooled Reduced Representation Libraries and covering approximately 2% of the genome of Large White, Landrace, Pietrain, Duroc and Wild Boar. Methods were developed to estimate Nucleotide Diversity (ND) considering that, GA sequences were obtained from pooled DNA, singletons were removed and the sequencing error rate. The average ND ranged from 0.0008 to 0.002 depending on chromosome and breed. Genomic locations that have been, putatively, under selection were identified. We found signals of positive selection on SSC8 in the region containing the KIT gene, for white breeds but not for Duroc and Wild Boar. Signals of balancing selection were found for regions on SSC7 containing genes from the MHC complex and from the olfactory receptors complex. Enrichment analysis of KEGG-pathways showed that for regions under positive selection, swine breeds showed higher enrichment of pathways related to growth whereas Wild Boar showed higher enrichment of pathways related to immunity and robustness. Balancing selection resulted in the significant enrichment of pathways related to the olfactory receptors activities in all swine breeds and Wild Boar. Our results raise the possibility of using GA sequencing of pools for identification of selection footprints and present the first global map of regions under selection in the swine genome.

P8

Novel tools and methods for exploring pyrosequencing data including quality assessment and simulation

Balzer, S. [1], Jonassen, I. [2], Malde, K. [1]
[1]Institute of Marine Research, Nordnes, Bergen, Norway; [2]Department of Informatics and Computational Biology Unit, BCCS, University of Bergen, Bergen, Norway

454 pyrosequencing produces huge amounts of data. With average read lengths of approximately 500 base pairs and one million reads per run, the output files used for assembly reach a size of about two gigabytes each.

The primary output is in the form of flowgram data (number of bases incorporated in a homopolymer run), and secondary data like the sequence of bases and corresponding quality values are derived from this. To avoid information loss, the analysis of raw data is essential.

We present a suite of software tools to assist in analysis of flowgram data. Our information extraction tool enables us to aggregate data and calculate statistics. A filtering tool allows ranking of data after different quality criteria and thus provides a more straightforward assembly. Finally, we present a flowgram simulator that generates realistic flowgram data for any sequence input by the user, and demonstrate its use in planning of sequencing projects, benchmarking of assembly methods etc.

The direct use of flowgram data might improve sequencing results, as it has been demonstrated for analysis of SNPs. We explore novel methods for analysis of flowgram data to enable quality control, filtering and improved base calling in context of sequence assembly.

P9

StatSeq : Statistical challenges on the 1000 genome sequences in plants (EU COST Action TD0801)

Bink, Marco [1], Schiex T [2]
[1]Biometris - Plant Research International, Wageningen, The Netherlands;
[2]MIA INRA, Chemin de Borde Rouge, Castanet-Tolosan, Cedex, France

New sequencing technologies either currently available or under development will eventually enable eukaryotic genomes to be sequenced for less than 1000 euros. This technology-push will have a major impact on plant genomics and biological research and lead to a dramatic expansion in both the availability of sequence data and the range of sequence based applications. New innovative techniques are required to unlock the information contained in the sequence data and to apply the acquired knowledge for plant science and crop improvement. The wide variety and often unique characteristics of plant genomes pose additional challenges and opportunities.

The need for and the dissemination of efficient strategies for handling and analyzing high throughput sequence data in plants requires cooperation at the international level to develop new approaches & analytical tools and share best practice. This COST Action will establish a network of researchers that coordinate, focus and strengthen national and pan-European statistical genomics and bioinformatics. It will be built on close interactions with other disciplines such as genetics, genomics and breeding. The Working Groups will arrange workshops, Short Term Scientific Missions, training courses, and publications to disseminate aims and achievements. Further information is now available on www.statseq.eu.

P10

Genome-wide assessment of nucleotide diversity and signatures of selection in chicken using massive parallel sequencing

Megens, H.-J.[1], Zare Y.[1], Amaral A.J.[1], Crooijmans R.P.M.A.[1], Ferretti L.[2], Ramos-Onsins S.E.[2], Perez-Enciso M.[2], Groenen M.A.M.[1]
[1]Wageningen University, Animal Breeding and Genomics Centre, Wageningen, The Netherlands; [2]Universitat Autonoma Barcelona - ICREA, Bellaterra, Spain

Genome-wide estimates of Nucleotide Diversity (ND) were obtained for four commercial chicken populations by massive parallel (GA) sequencing of Reduced Representation Libraries covering around 1.5% of the genome of the two layers, and around 4% of the genomes of broilers (meat chicken) at a minimum read depth of 6, and average read depth of 18 to 45 depending on the population. ND estimations ranged between 0.001 for the layers to 0.002 for the broilers, slightly lower than previously reported using low-depth Sanger sequencing, likely due to stringent filtering of low frequency alleles. As expected, ND was higher in the microchromosomes compared to the macrochromosomes, and was lowest in chromosome Z (the bird equivalent of the X chromosome). Significantly aberrant chromosomal regions were detected that could be interpreted as signatures of selection. One example is a region containing the IGF1 gene, previously implicated in fast growth in broilers, that had a significant signature of selection in both broiler populations, but not in the layers. Conversely, for layers a number of selected regions coincided with QTL for egg production. With ~200K (broilers) to ~50K (layers) of segregating sites covered, GA sequencing provides a powerful tool for detecting signatures of selection in chicken.

P11

Cloud computing and NGS: massively parallel computing for massively parallel sequencing

Pareja-Tobes, E., Manrique, M., Pareja-Tobes, P., Pareja, E., Tobes, R.
Era7 Information Technologies SLU, BIC Granada Avda. de la Innovación 1. Armilla, Granada, Spain

Research using new massively parallel sequencing technologies is continuously generating huge amount of data that are especially suited to be managed by means of new massively parallel computing and storage models.

Next generation sequencing and other recent technologies have improved the accessibility of sequencing but in many cases the bottleneck has now moved to the bioinformatics analysis, tightly related to the availability of high-throughput computation infrastructure. Cloud computing fits perfectly with these new bioinformatics needs and, improving the accessibility of high-throughput computation, specifically with "infrastructure as a service" cloud services.

Key factors of cloud computing are automatic provisioning of infrastructure, on-demand scalability, capacity to adapt programmatically your computing infrastructure in real-time by means of API calls, security and reliability.

Paradigms such as Map/Reduce, XML-based queryable data storage, SOAP/REST SOA architectures, messaging frameworks (AWS SQS, cloudMQ, ...), concurrent programming models such as Actors (Scala, Erlang) are especially adapted both to cloud computing and bioinformatics needs.

NGS data management using cloud computing allows you to maintain the focus on knowledge-driven analysis not on infrastructure provisioning. For example we have developed an annotation pipeline that using cloud computing allows us to annotate a complete bacterial genome in 24 hours.

P12

Genome browser of genetic diversity in Drosophila

Ràmia, M., Casillas, S., Barrón, MG., Egea, R., Barbadilla, A.
Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain

Population genetics studies have so far been based on fragmentary and non-random samples of the genome, providing a partial and often biased view of the population processes. The sequencing of complete genome sequences for 192 individuals of a species which has a high quality reference genome, as is the case of D. melanogaster, will allow both a global view of genomic variation and an understanding of the forces that are responsible for the patterns of polymorphism and divergence along the genome (the DGRP project). We present a novel genome browser for the representation of genetic diversity in this species that automates the estimation of genetic variation along each chromosome and graphically represents all the information extracted from the data (SNP/CNV frequencies, nucleotide diversity, recombination rates, etc.), as well as other features gathered from external sources (TE insertions, mapped QTLs, etc.). The primary analysis of this dataset represents an unprecedented opportunity to describe and explain, for the first time, the evolutionary processes that shape the genome of such a paramount model organism as D. melanogaster. Interestingly, this genome browser can be easily used to create analogous resources for any other species for which polymorphic sequences at the genome-scale are being obtained.

P13

Including dominance effects in genomic selection

Toro, M.A. [1], Varona, L. [2]

[1]Dpto. Producción Animal, ETS Ingenieros Agronomos, Universidad Politécnica de Madrid. [2]Dpto. Producción Animal, Facultad de Veterinaria, Universidad de Zaragoza

A population was simulated for 1000 generations at an effective size of 100 and expanded afterwards to a size of 3000. The genome was assumed to consist 9000 SNPs and 1000 QTLs located at random map positions. Both SNPs and QTLs were biallelic. Mutations were generated at a rate of 2.5x10-3 per locus per generation at a marker loci and at a rate of 2.5 x 10-5 at a QTL loci. Both the additive and the dominance effects were sampled from a standard normal distribution and scaled to get the desired values of h2 (VA/VP) and d2 (VD/VP) being VA, VD and VP the additive, dominance and phenotypic variance. Estimation of marker effects was carried out using Bayes A method with and without including dominance effects. In generation 1003, 25 males and 250 females were selected from 500 males and 500 females based in the estimation of markers effects and they were mated randomly. The advantages of including dominance effects in the model were evaluated in a multigeneration context.

P14

Using Next Generation Sequencing on ancient DNA - preamplified via a new Multiplex approach - to detect migration and population structure

M. Unterländer (1), S. Wilde (1), Ch. Schuh (2), C. Gerling (2), I. Popov (2), M. Woidich (2), Z. Samashev (3), E. Kaiser (2), W. Schier (2), H. Parzinger (4), J. Burger (1)

1. Institute of Anthropology, University of Mainz, Germany. 2. Institute of prehistoric Archaeology, FU Berlin, Germany. 3. Margulan Institute of Archaeology, Academy of Science Kazakhstan, Almaty. 4. Foundation of Prussian Cultural Heritage, Berlin, Germany

Due to the degraded nature of ancient archaeological DNA, de novo NG-sequencing works on only for extremely well preserved samples. We therefore developed a multiplex-PCR based protocol including subsequent NGS of individually tagged PCR products for clonal sequencing on the FLX platform. This includes a set of 31 mitochondrial SNPs plus the HVRI, 36 Y-chromosomal SNPs and 15 loci known to be under high positive selective pressure in humans. We think that this approach in combination with capture based assays and de novo NG-sequencing provides a tool for effective data collection from prehistoric, degraded specimens. We apply this technique to various skeletal remains of differing environmental conditions, mainly from Iron Age (1st Mill B.C.) kurgans of the Central Asian steppe.

P15

Pyrosequencing of non-model sentinel species for gene transcription profiling studies in environmental pollution monitoring

Díaz de Cerio O., Fernández-Lanza V., Cancio I.
Dept. of Zoology and Animal Cell Biology. School of Science and Technology. University of the Basque Country, Bilbao, Spain

Key sentinel species are employed to monitor biological effects of pollution and assess the quality of ecosystems. High throughput "omic" techniques such as microarrays employed in transcriptomics allow identification of single genes and/or gene pathways that could be assessed as biomarkers of exposure to specific chemical compounds. Nevertheless, ecotoxicogenomic studies have to face the problem of the lack of genetic information available for toxicologically relevant species. In this sense, pyrosequencing allows obtaining deep de novo sequence information cost-effectively. We are thus employing the GS-FLX platform in order to sequence the multitissue transcriptome, normalised cDNA (Evrogen), of widely distributed and commonly used aquatic sentinel species (fish and molluscs). We have now completed 1 full plate sequencing-run of the thicklip grey mullet (Chelon labrosus), a suitable sentinel of pollution able to survive heavily polluted marine/estuarine waters in southern Europe. 126 Mb of sequence information have been obtained that upon assembly have resulted in 35,605 contigs and 235 singletons that have been annotated through BasIt2Go with the help of the Spanish Institute of Bioinformatics. Finally, a custom Agilent microarray will be generated to study gene transcription profiles in laboratory exposure experiments and in pollution monitoring programs.

Funded: Basque Government (SAIOTEK-IKERKOSTA S-PC08UN03; Special-Research-Actions AE-2008-1-1).

P16

High-throughput sequencing technologies applied to human gut microbiota research and genomics of pathogens

Gosalbes M.J. [1,2,3], Jiménez N. [2,3], D Auria G. [2,3], Durbán A. [2,3], Peris-Bondia F [1,2], Pérez-Cobas A.E.[3], Latorre A.[1,2,3], Moya A. [1,2,3]

[1]Instituto Cavanilles de Biodiversidad y Biología Evolutiva, Universidad de Valencia, Valencia, Spain; [2]CIBER Epidemiología y Salud Pública
[3]Centro Superior de Investigación en Salud Pública, Avenida Cataluña, Valencia, Spain

Next-generation sequencing technologies have been used for standard sequencing applications, and for novel applications previously unexplored by Sanger sequencing. Here we describe several applications that we are carrying out in our laboratory, using Roche/454 FLX for massively parallel DNA sequencing .

1. Resequencing: We obtained the complete genome of Legionella pneumophila (strain Alcoy 2300/99) related with an important outbreak in Alcoy (Spain) in 1999, by an hybrid Sanger/454-FLX assembly.
2. De novo Sequencing: We are finishing the genome of Acidaminococcus intestinalis, potential human pathogen. We used hybrid FLX standard and FLX Paired Ends to obtain an oriented scaffolding of the genome .
3. Amplicon Sequencing: We used DNA bar coding and pyrosequencing to characterize 84,000 sequences of 16S rRNA genes obtained from gastrointestinal samples, allowing quantitative analysis of community composition in health and disease.
4. Transcriptome: We are analysing the transcriptome from full-length mRNA, from gastrointestinal samples in healthy controls and patients.

P17

Analysis of the gonadal transcriptome during sex determination, sex differentiation and gonadal maturation in the sea bass (Dicentrarchus labrax) and turbot (Scophthalmus maximus) by 454 sequencing and two specific oligo-based microarrays.

Ribas, L. [1], Crespo, B. [2], Díaz, N. [1], Gómez, A. [2], Pardo, B.G. [3], Reinhardt, R. [4], MacKenzie, S. [5], Martínez, P. [3], S. Zanuy, S. [2] and Piferrer, F. [1]

[1]Institut de Ciències del Mar, CSIC. Passeig Marítim, Barcelona, Spain; [2]Instituto de Acuicultura de Torre la Sal, CSIC. Ribera de Cabanes, Castelló, Spain; [3]Depto. de Genética, Facutad de Veterinaria, Campus de Lugo, Lugo, Spain; [4]Max Planck Institute for Molecular Genetics - Ihnestraße, Berlin, Germany; [5]Dept. Bio. Cel., Fisio. i Immunologia, Facultat de Ciències, Bellaterra, Spain

Fish represent unique models to study vertebrate sex determination and differentiation, as well as gonad maturation for several reasons. First, fish are by far the most abundant type of vertebrates and exhibit all reproduction types known in vertebrates (gonochorism, hermaphroditism and unisexuality). Furthermore, the processes of sex determination, differentiation and gonad maturation are the result of complex genetic, environmental and social interactions. Thus, for example, temperature or population density can easily influence the course of sex differentiation, a phenomenon that represents one of the most dramatic examples of phenotypic plasticity.

The gonads are then the only organ in vertebrates that starting from common undifferentiated rudiment can undergo two completely different developmental pathways, resulting in a testis or ovary, and the environmental conditions imposed by modern farming provide an excellent framework where to study these essential biological processes.

We are applying next generation sequencing technologies, specifically 454 sequencing, to obtain a detailed overview of both the European sea bass (Dicentrarchus labrax) and turbot (Scophthalmus maximus) gonadal transcriptome under a variety of developmental and experimental conditions. In addition, oligo-based microarrays are being built enriched with genes expressed in the reproductive axis with the purpose of discovering new genes and signaling routes.

P18

Rhopalodia gibba and its spheroid body - sequencing endosymbiosis

Schönfeld, B I K
Allan Wilson Centre for Molecular Ecology and Evolution, Institute of Molecular BioSciences, Massey University, Palmerston North, New Zealand

Symbiotic relationships have always played an essential role in eukaryote evolution, especially with the symbiogenesis of mitochondria and plastids. These milestones in the evolution of eukaryotic life lie too deep in time to allow detailed reconstruction of the processes leading to organelle formation based on modern organelles, and instead modern symbioses are studied to shed light on these ancient events. One example is the diatom Rhopalodia gibba which acquired the capability to fix nitrogen with its cyanobacterial endosymbiont. New sequencing technology for the first time makes it feasible to create a comprehensive picture of the genetic aspects of this symbiotic relationship. As a starting point of this project, data from the Illumina Genome Analyser has been used to sequence the endosymbiont genome and host cDNA. Preliminary results assessing gene loss and genome changes confirm the advanced level of this symbiosis. The de novo sequencing of non-model organisms with regard to evolutionary changes poses unique challenges. In the framework of my project I present problems involved and novel approaches to next generation data analysis that have been developed at the Allan Wilson Centre to solve them.

P19

Investigating molecular basis of response to selection in bank vole with next generation sequencing

Stuglik, M. [1], Babik, W. [1], Qi, W. [2], Kuenzli, M. [2], Gac, K. [1], Koteja, P. [1], Radwan, J. [1], [1]Institute of Environmental Sciences, Jagiellonian *Universtiy, Krakow, POLAND; [2]Functional Genomics Center Zurich, Winterthurerstr, Zuerich, Switzerland*

Molecular basis of response to selection on complex traits is poorly understood. Both changes in regulatory regions affecting the expression level and changes in coding regions themselves may contribute to adaptation. In the long run, our aim is to investigate both these aspects of response to selection on high metabolic rate in the bank vole. Here, using 454 Titanium sequencing of the normalized cDNA we characterised the transcriptome of heart tissue, which is likely to affect metabolic performance. We used mRNA obtained from four replicate lines of the bank voles selected for high metabolic rate for six generations, as well as from four control lines. We obtained ca. 330 Mb of usable sequences which were assembled into 79829 contigs of the length 96 - 6943 with the median length 371. BLAST searches against the mouse RefSeq database detected 10079 non-redundant genes which were assigned the gene ontology categories. We tested the utility of 454 Titanium to trace the changes in the frequency of SNPs in lines selected for high metabolic rate. Bioinformatical analysis detected > 8000 high quality SNPs. We empirically validated a subset of 22 such detected SNPs using SnapShot method and sequencing, and confirmed the majority of polymorphisms.

P20

Selection of Cancer-Related Gene Exons for Targeted Resequencing with a Flexible and Fully Automated Microarray Platform

Summerer, D. [1], Schracke, N. [1], Wu, H.[2], Cheng, Y.[1], Haase, B.[1], Stähler, C.F.[1], Stähler, P.F.[1]; Beier, M.[1]
[1]febit biomed gmbh, Im Neuenheimer Feld 519, Heidelberg, Germany;
[2]febit inc., 99 Hayden Ave, Suite 620, Lexington, MA, USA

Next-generation sequencing studies are currently limited by an inability to enrich genomic DNA samples for specific regions of interest easily and effectively. This keeps sample throughput at the low-end and calls for focused and more inexpensive methods to analyze complex, eukaryotic genomes. Allowing for larger numbers of samples, the selection of relevant subsets for targeted resequencing greatly increases statistical power, while keeping datasets manageable. Here we present HybSelect®, a method to select human genomic loci of interest by hybridization on scalable, microfluidic DNA microarrays using the Geniom® RT Analyzer. We captured the complete coding sequence (1819 exons) of 115 cancer-related genes of a recently fully sequenced Yoruba individual representing a total region of interest (ROI) of 9.3 Mb on part of a biochip. This corresponds to > 18.4 Mb capacity per biochip. Sequencing using Illumina technology revealed an average depth of coverage of 175.7-fold for all exons, and 97 % of targeted bases were covered at least once. Uniformity was such that 94 % of genes were in a range of < 1 log which indicates applicability to various sequence contexts. A comprehensive analysis of HapMap reference SNPs revealed a concordance of 99.4 %. Using a fully automated microarray processing platform minimizes contamination risk, enhances reproducibility, and accelerates the experimental workflow.

ATTENDEE LIST

Álvarez, Miguel
Roche Diagnostics SL.
Sant Cugat,Spain

Amador Catalán, Amaya
Universitat de Barcelona
Barcelona, Spain

Amaral, Andreia
WUR
Wageningen,The Netherlands

Ameur, Adam
Uppsala University
Uppsala, Sweden

Angelini, Claudia
Istituto per le Applicazioni del Calcolo
Naples, Italy

Anglada, Roger
Universitat Pompeu Fabra
Barcelona, Spain

Aranzana Civit, María José
CRAG (Centre de Recerca en Agrigenòmica)
Cabrils, Spain

Arjona, Rosa
Institut de recerca Hospital Universitari Vall d´Hebron
Barcelona, Spain

Balzer, Susanne
Institute Of Marine Research
Bergen, Norway

Barbadilla Prados, Antonio
Universitat Autònoma Barcelona
Bellaterra, Spain

Barcelo, Anna
Universitat Autònoma Barcelona
Cerdanyola, Spain

Barillot, Emmanuel
Institut Curie - U900
Paris, France

Bayer, Micha
Scottish Crop Research Institute
Dundee, United Kingdom

Bayes, Monica
Center for Genomic Regulation - CRG
Barcelona, Spain

Beck, Stephan
University College From London
London, United Kingdom

Bergero, Roberta
University of Edinburgh
Edinburgh, United Kingdom

Bernd Timmermmann
Max-Planck Institute for Molecular Genetics
Berlin, Germany

Bertranpetit Busquets, Jaume
Universitat Pompeu Fabra
Barcelona, Spain

Bessa Parmigiani, Raphael
Ludwig Institute For Cancer Research
Sao Paulo, Brasil

Bin, Yang
Universitat Autònoma Barcelona
Bellaterra, Spain

Bink, Marco
Plant Research International
Wageningen, The Netherlands

Blanca Postigo, José Miguel
Universidad Poltécnica de Valencia
Valencia, Spain

Blicher, Thomas
The Technical University of Denmark
Lyngby. Denmark

Boeva, Valentina
Institut Curie - U900/ U830
Paris, France

Bosch Fusté, Elena
Universitat Pompeu Fabra
Barcelona, Spain

Brudno, Michael
University Of Toronto
Toronto, Canada

Bustamante, Carlos
Cornell University
Ithaca, USA

Butler, Derek
BaseClear B.V.
Leiden, The Netherlands

Calafell Majo, Francesc
Universitat Pompeu Fabra
Barcelona, Spain

Campos, José Luis
Universitat de Barcelona
Barcelona, Spain

Cancio Uriarte, Ibon
University of The Basque Country
Bakio, Spain

Cannon, Charles
Chinese Academy of Sciences
Menglun, China

Cañizares Sales, Joaquin
Universidad Poltécnica de Valencia
Valencia, Spain

Carrasco Ramiro, Fernando
Centro de biologia Molecular "Severo Ochoa" (CSIC-UAM)
Madrid, Spain

Casillas Viladerrams, Sònia
Universitat Autònoma Barcelona
Bellaterra, Spain

Castellví Bel, Sergi
Hospital Clínic / Fundació Clínic / CIBERehd
Barcelona, Spain

Castelo Valdueza, Robert
Universitat Pompeu Fabra
Barcelona, Spain

Cigudosa García, Juan Cruz
Spanish National Cancer Research Center
Madrid, Spain

Clark, Andrew
Cornell University
Ithaca, USA

Collinet, Kathryn
IRB Barcelona
Barcelona, Spain

Corrada Bravo, Héctor
Johns Hopkins Bloomberg School of Public Health
Baltimore, USA

Cruz Cuevas, Juan
CRAG (CSIC - IRTA - UAB)
Barcelona, Spain

Davis, Gerard
Pfizer Animal Health
Albion, Australia

De feis, Italia
Istituto per le Applicazioni del Calcolo "Mauro Picone" sede di Napoli
Napoli, Italy

De La Cruz, Xavier
IBMB - CSIC
Barcelona, Spain

De La Vega, Francisco M.
Applied Biosystems
Foster City, CA, USA

De Lorenzo, David
Centre Nutren-Nutrigenomics
Lleida, Spain

Deniz, Ozgen
Institute for Research in Biomedicine
Barcelona, USA

Dixon, Richard
Applied Biosystems Europe.
Europe

Dopazo, Ana
CNIC
Madrid, Spain

Drablos, Finn
Norwegian University of Science and Technology
Trondheim, Norway

Durán Moliner, Elisa
Barcelona Supercomputing Center
Barcelona, Spain

Egea Cortines, Marcos
Universidad Politécnica de Cartagena
Cartagena, Spain

Egea Sánchez , Raquel
Universitat Autònoma Barcelona
Bellaterra, Spain

Espinosa, Toni
Universitat Autonoma de Barcelona
Bellaterr, Spain

Estellé Fabrellas, Jordi
INRA
Jouy en josas, France

Esteve, Anna
Universitat Autònoma Barcelona
Bellaterra, Spain

Faria, Rui Miguel
Bioevo, Ebi (upf-csic)
Barcelona, Portugal

Feliubadalo, Lidia
Fundació IDIBELL
Hospitalet de llobregat, Spain

Fernandez Avila, Ana i
INIA
Madrid, Spain

Fernandez Cadenas, Israel
Institut de recerca, Hospital Vall d'Hebron
Barcelona, Spain

Fernández Vidal, Leyden
Barcelona Supercomputing Center
Barcelona, Spain

Ferretti, Luca
Universitat Autònoma Barcelona
Bellaterra, Spain

Foissac, Sylvain
Integromics
Tres Cantos, Spain

Galaverni, Marco
University of Bologna AND National Wildlife Institute (ISPRA-ex INFS)
Bologna, Italy

Gallego Valadés, Paqui
Institut de Recerca Hospital Universitari Vall d´Hebron
Barcelona, Spain

Garcia Garcera, Marc
Universitat Pompeu Fabra
Barcelona, Spain

García Mas, Jordi
IRTA, Centre de Recerca en Agrigenòmica CSIC-IRTA-UAB
Cabrils, Spain

Gironella, Meritxell
CIBERehd/Hospital Clinic
Barcelona, Spain

Gloeckner, Frank Oliver
Max Planck I
Germany

González Miguel, Víctor Manuel
CRAG (CSIC-IRTA-UAB)
Barcelona, Spain

Gonzalez Rivera, Milagros
Hospital Gu Gregorio Marañon
Madrid, Spain

Gonzalez Roca, Eva
Institute for Research in Biomedicine
Barcelona, Spain

González Rosado, Santiago
Barcelona Supercomputing Center
Barcelona, Spain

Gosalbes Soler, Maria José
Instituto Cavanilles - Univ. De Valencia
Valencia, Spain

Gracia Aznarez, Francisco Javier
Spanish National Cancer Research Center
Madrid, Spain

Green, Philip
University Of Washington
Washington, USA

Groenen, Martien
Wageningen Agricultural University
Wageningen, The Netherlands

Guigó, Roderic
Center For Genomic Regulation
Barcelona, Spain

Gyllensten, Ulf
Uppsala University
Uppsala, Sweden

Haase, Bettina
Febit Biomed Gmbh
Heidelberg, Germany

Hajibabaei, Mehrdad
Biodiversity Institute Of Ontario -
University Of Canada
Guelf, Canada

Heckel, Gerald
University of Bern
Bern, Switzerland

Himmelbauer, Heinz
Center For Genomic Regulation
Barcelona, Spain

Hoffmann, Steve
University Leipzig
Leipzig, Germany

Hupé, Philippe
Institut Curie
Paris, France

Ibáñez Garcia, Teresa
Technical Secretariat NGS 2009
Barcelona, Spain

Igartua Arregui, Ernesto
CSIC
Zaragoza, Spain

Iso-touru, Terhi
MTT Agrifood Research Finland
Jokioinen, Finland

Jiménez Hernández, Nuria
Centro Superior de Investigación en
Salud Pública
Valencia, Spain

Jun, Wang
Beijing Genomics Institute
Shenzhen, China

Khatkar, Mehar
University of Sydney
Camden, Australia

Knight, James
454 Life Sciences Corporation. Roche
Branford, USA

Korhonen, Kati
MTT Agrifood Research Finland
Jokioinen, Finland

Kua, Chai-shian
Chinese Academy of Sciences,
Xishuangbanna Tropical Botanical Garden
Menglun town, China

Laayouni, Hafid
Universitat Pompeu Fabra
Barcelona, Spain

Langenberger, David
University Leipzig
Leipzig, Germany

Laugraud, Aurélie
PRABI
Villeurbanne, France

Laurie, Steven
Fundació Institut Municipal d'Investigació
Mèdica/Universitat Pompeu Fabra
Barcelona, Spain

Lazaro, Concepción
Fundació IDIBELL
Hospitalet de llobregat, Spain

Ledda, Alice
Fundació Institut Municipal d'Investigació
Mèdica/Universitat Pompeu Fabra
Barcelona, Spain

Lin, Yao-cheng
VIB Department of Plant Systems
Biology, UGent
Ghent, Belgium

Lok, Si
The University of Hong Kong
Pokfulam, China

López Alonso, Victoria
Instituto de Salud Carlos III
Majadahonda, Spain

Lynch, Michael
Indiana University
Bloomington, USA

Maiques Díaz, Alba
Spanish National Cancer Research Center
Madrid, Spain

Manrique, Marina
Era7 Information Technologies
Granada, Spain

Marth, Gabor
Boston College
Massachussets, USA

Martínez Izquierdo, José Antonio
CRAG-CSIC-IRTA-UAB
Barcelona, Spain

Martínez Portela, Paulino
Universidad De Santiago De Compostela
Lugo, Spain

Mcvean, Gil
Oxford University
Oxford, United Kingdom

Megens, Hendrik-jan
Animal Breeding and Genomics Centre,
Wageningen University
Wageningen, The Netherlands

Mercadé Roca, Jaume
Institute of Predictive and Personalized
Medicine of Cancer
Badalona, Spain

Mercader Bigas, Josep M.
Barcelona Supercomputing Center
Barcelona, Spain

Miñarro Alonso, Antonio
University of Barcelona
Barcelona, Spain

Mir Arnau, Gisela
IRTA/CRAG
Cabrils, Spain

Mohellibi, Nacer
National Research Institute for
agronomics
Versailles, France

Montfort, Amparo
CRAG (Centre de Recerca en
Agrigenòmica)
Cabrils, Spain

Montserrat Sentís, Bàrbara
Barcelona Supercomputing Center
Barcelona, Spain

Mosquera Mayo, José Luis
Institut de Recerca Hospital Universitari
Vall d'Hebron
Barcelona, Spain

Mularoni, Loris
Johns Hopkins University
Baltimore, USA

Naira, Naouar
VIB - Ghent University
Ghent, Belgium

Navarro, Arcadi
Universitat Pompeu Fabra
Barcelona, Spain

Negredo Antón, Ana Isabel
Instituto de Salud Carlos III
Majadahonda, Spain

Noguera Julian, Marc
Institut de Medicina Preventiva i
Personalitzada del Cancer
Badalona, Spain

Notredame, Cedric
Center For Genomic Regulation
Barcelona, Spain

Nuñez Mangado, Fatima
Institut Recerca Hospital Vall d'Hebron
Barcelona, Spain

Oliva Virgili, Rafael
Human Genetics Laboratory, Faculty of
Medicine and Hospital Clinic
Barcelona, Spain

Ortega Serrano, Israel
Institut Recerca - Vall Hebron
Barcelona, Spain

Papadakis, Emmanouil
Technical University Of Denmark
Lyngby, Denmark

Pasquali, Lorenzo
IDIBAPS/Hospital Clinic de Barcelona
Barcelona, Spain

Pastor Hostench, Xavier
Barcelona Supercomputing Center
Barcelona, Spain

Pere, Arus
IRTA-Center for Research in Agricultural
Genomics CSIC-IRTA-UAB
Cabrils, Spain

Perera, Anoja
stowers institute
Kansas city, USA

Pérez Iglesias, Rocio
Hospital universitario Marqués de
Valdecilla
Lierganes, Spain

Pérez Lezaun, Anna
Universitat Pompeu Fabra
Barcelona, Spain

Pérez Llamas, Christian
Universitat Pompeu Fabra
Barcelona, Spain

Pérez-Enciso, Miguel
Universitat Autònoma Barcelona
Bellaterra, Spain

Piferrer, Francesc
Institute of Marine Sciences, CSIC
Barcelona, Spain

Pita Mcpherson, Guillermo
Centro Nacional de Investigaciones
Oncológicas (CNIO)
Madrid, Spain

Plaza, Stéphanie
Universitat Pompeu Fabra
Barcelona, Spain

Pluvinet Ortega, Raquel
Institute of Predictive and Personalized
Medicine of Cancer
Badalona, Spain

Ramayo, Yuliaxis
Universitat Autònoma Barcelona
Bellaterra, Spain

Ramos-Onsins, Sebastian
CRAG - Universitat Autònoma Barcelona
Bellaterra, Spain

Ribas Cabezas, Laia
Institut de Ciències del Mar (CSIC)
Barcelona, Spain

Riechmann, José Luis
Centre De Recerca en Agrigenòmica
(CRAG)
Barcelona, Spain

Rodriguez Mulero, Silvia
Institute for Research in Biomedicine,
IRB Barcelona
Barcelona, Spain

Rosa Rosa, Juan Manuel
Spanish National Cancer Research Center
Madrid, Spain

Ross, Mark T.
Illumina Cambridge Ltd.
Essex, UK

Ruiz De Villa, M. Carme
Universitat de Barcelona
Barcelona, Spain

Ruiz, Alfredo
Universitat Autònoma Barcelona
Bellaterra, Spain

Rusch, Douglas B.
J. Craig Venter Institute
Rockville, MD, USA

Sanchez Pla, Alex
Institut de Recerca. Hospital Universitari
Vall d'Hebron
Barcelona, Spain

Satkoski Trask, Jessica Ann
University Of California Davis
Davis, USA

Schönfeld, Barbara
Allan Wilson Centre - Massey University
Palmerston North, New Zealand

Serra Arenas, Eduard
Institute of Predictive and Personalized
Medicine of Cancer
Badalona, Spain

Servant, Nicolas
Institut Curie
Paris, France

Sevilla, Lidia
Parc Científic De Barcelona
Barcelona, Spain

Sikora, Martin
Universitat Pompeu Fabra
Barcelona, Spain

Sironen, Anu
MTT, Agrifood Research Finland
Jokioinen, Finland

Stuglik, Michal
Jagiellonian University
Krakow, Poland

Stunnenberg, Henk
Nijmegen Center For Molecular Life
Sciences
Nijmegen, The Netherlands

Sumoy Van Dyck, Lauro
IMPPC
Badalona, Spain

Tellgren-roth, Christian
Uppsala University
Uppsala, Sweden

Tobes, Raquel
Era7 Information Technologies
Granada, Spain

Tong, Amy
The University of Hong Kong
Pokfulam, China

Toro Ibáñez, Miguel Angel
Universidad Politécnica de Madrid
Madrid, Spain

Torrents, David
Barcelona Supercomputing Center
Barcelona, Spain

Turon Barrera, Francesc Xavier
CSIC - Centro De Estudios Avanzados De
Blanes
Blanes, Spain

Unterlaender, Martina
Insitut für Anthropologie / Johannes Gutenberg-Uni Mainz
Mainz, Germany

Van Leeuwen, Hans
Nunhems Netherlands BV
Haelen, The Netherlands

Van Orsouw, Nathalie
Keygene N.V.
Wageningen, The Netherlands

Vera Rodríguez, Manuel
Universidad De Santiago De Compostela
Lugo, Spain

Vera, Gonzalo
Universitat Autònoma Barcelona
Bellaterra, Spain

Vuylsteke, Marnik
VIB
Gent, Belgium

Wahlberg, Per
INRA
Jouy en josas, France

Wheelan, Sarah
The Johns Hopkins University School of Medicine
Baltimore, USA

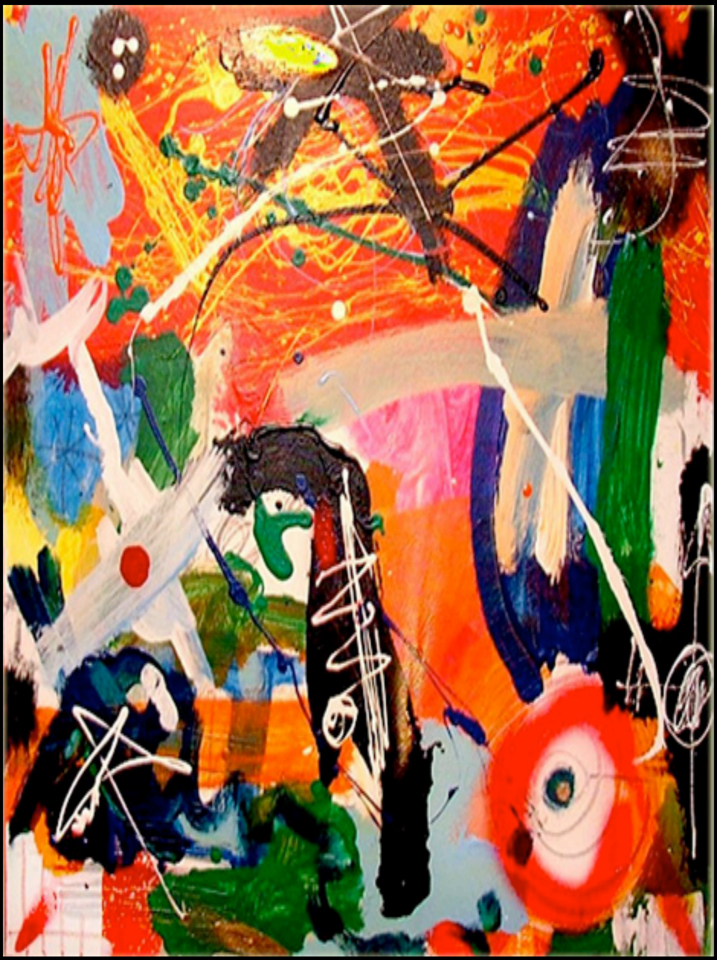
Wilson, Richard K.
Washington University
Sant Louis, MO, USA

Yankilevich, Patricio
Integromics
Tres cantos, Spain

Zeitouni, Bruno
Institut Curie
Paris, France

Zytnicki, Matthias
National Research Institute for Agronomics
Versailles, France

About the logo



Antonio Moreno was born in Melilla, 1963. He started doing video and photography but soon was attracted to painting. Initially selftaught, later takes courses at the Universidad Popular Pablo Iglesias in Madrid, and attend a workshop by Jordi Teixidor. From 2003 onwards he had had exhibitions all over Spain. He now alternates painting, photography, video, digital art and sculpture.

His work makes part of the following collections: *Museo de Arte Contemporáneo de Calvià* (Mallorca), *Museo de Arte Contemporáneo de Marmolejo* (Jaén), *Museo de Arte Contemporáneo de Azuaga* (Badajoz), *Fundación Universidad Rey Juan Carlos I* (Madrid), *Colegio de Arquitectos Técnicos de Almería, Real Club Náutico* (Las Palmas de Gran Canaria).

http://antoniomoreno.wordpress.com/

UAB
Universitat Autònoma
de Barcelona

icrea
Institució Catalana de Recerca i Estudis Avançats

crag CENTRE FOR RESEARCH IN AGRICULTURAL GENOMICS
CSIC IRTA UAB

UNIVERSITAT POMPEU FABRA
UPF

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Genoma España

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN

REDES moleculares

Applied Biosystems

illumina

Roche

Agilent Technologies

K Biosciences
www.kbioscience.co.uk

sistemas genómicos
servicios de análisis genético

BASECLEAR
FOR 100% DNA RESULTS

Thanks so much!