CNIO FRONTIERS Meetings 2011

Centro Nacional de Investigaciones Oncológicas (CNIO)
Melchor Fernández Almagro, 3 | 28029 Madrid | Spain
Tel.: +34912246900
www.cnio.es

Supported by the ESF Programme "Frontiers of Functional Genomics"

Canceromatics II: Multilevel Interpretation of Cancer Genome

CNIO FRONTIERS Meetings 2011

Spanish National Cancer Research Centre

CNIO FRONTIERS Meetings 2011

Canceromatics II: Multilevel Interpretation of Cancer Genome

28-30 MARCH 2011

Organisers

Søren Brunak, University of Copenhagen, Denmark
Stephen Chanock, National Institutes of Health, Bethesda, USA
Núria Malats, CNIO, Madrid, Spain
Chris Sander, Memorial Sloan-Kettering Cancer Center, New York, USA
Alfonso Valencia, CNIO, Madrid, Spain

CNIO FRONTIERS Meetings 2011

Spanish National Cancer Research Centre

Canceromatics II: Multilevel Interpretation of Cancer Genome

28-30 MARCH 2011

Canceromatics II: Multilevel Interpretation of Cancer Genome

CNIO FRONTIERS Meetings 2011

Summary

Canceromatics II: Multilevel Interpretation of Cancer Genome

CNIO FRONTIERS Meetings 2011

Detailed Progamme

MONDAY, MARCH 28th

09:00 **Welcome Address**
Alfonso Valencia, Núria Malats

Session I

Chairs: *Angela Brand, Søren Brunak*

09:15 **Stephen Chanock**, National Cancer Institute (NCI), National Institutes of Health, Bethesda, USA
Genome-wide association studies in cancer: discovery and characterization

09:55 **Francis Ouellette**, Ontario Institute for Cancer Research, Toronto, Canada
Finding the mutation in the alignment file!

10:35 ***Coffee break and poster session***

11:00 **Angela Brand**, Maastricht University, The Netherlands
Public Health Genomics goes personalized – lessons learned from cancer genomics

11:40 **Emmanuel Barillot**, Curie Institute, Paris, France
Modelling biological networks involved in tumor progression

12:20 **Yinyin Yuan**, Cancer Research UK, Cambridge Research Institute, UK
Short talk: *A sparse regulatory network of copy-number driven gene expression reveals putative breast cancer oncogenes*

13:00 ***Lunch and poster session***

14:15 **Open Discussion I. "Functional interpretation of cancer data"**

Session II

Chairs: *Francesca Ciccarelli, Lodewyk Wessels*

15:15 **Huanming Yang**, Beijing Genomics Institute, China
A natural combination of genomics and cancer research

15:55 ***Coffee break and poster session***

16:30 **Debora Marks**, Harvard Medical School, Boston, USA
Systems biology of microRNAs and cancer

17:10 **Henk Stunnenberg**, NCMLS, Radboud University Nijmegen Medical Centre, The Netherlands
The BLUEPRINT of hematopoiesis

17:50 **Zoltan Szallasi**, CBS, Technical University of Denmark, Lyngby, Denmark
Cancer genome data provide a comprehensive, "freeze-frame" view of dynamic genomic instability processes in human cancer samples

18:30 ***END of session***

TUESDAY, MARCH 29th

Session III

Chairs: *Anne-Lise Børresen-Dale, Stephen Chanock*

09:00 **Chris Sander**, Memorial Sloan-Kettering Cancer Center, New York, USA
Cancer genomics and network pharmacology

09:40 **Igor Jurisica**, Ontario Cancer Institute, Toronto, Canada
Network-based identification and systematic characterization of cancer signatures

10:20 ***Group picture. Coffee break and poster session***

11:00 **Rune Linding**, Technical University of Denmark (DTU), Lyngby, Denmark
Cancer & cellular information processing

11:40 **Lodewyk Wessels**, The Netherlands Cancer Institute, Amsterdam, The Netherlands
Oncogenic network identification from genomic data

12:20 **Nitin Kumar**, Institute of Molecular Life Sciences, Zurich, Switzerland
Short talk: *Non-neutral copy number alterations in cancer genome data*

13:00 ***Lunch and poster session***

14:15 **Open Discussion II. "Genome based stratified cancer treatment"**

Session IV

Chairs: *Nancy Cox, Igor Jurisica*

15:15 **Ivo Gut**, National Center for Genome Analysis (CNAG), Barcelona, Spain
The practicalities of the international cancer genome consortium analyses

15:55 ***Coffee break and poster session***

16:30 **Anne-Lise Børresen-Dale**, Oslo University Hospital Radiumhospitalet, Oslo, Norway
Integrated molecular profiles of primary breast tumors reveal differential interleukin signalling and improved prognostic power

17:10 **Alfonso Valencia**, CNIO, Madrid, Spain
Towards the protein network bases analysis of cancer genome data

17:50 **Francesca Ciccarelli**, IFOM-IEO Campus, Milan, Italy
Genomic instability and the evolution of cancer

18:30 ***END of session***

WEDNESDAY, MARCH 30th

Session V

Chairs: Debora Marks, Rune Linding

9:00 **Ken Fasman**, Dr. Miriam and Sheldon G. Adelson Medical Research Foundation, Needham, USA
Facilitating combination therapy development: and you thought it was difficult to invent a single drug

09:40 **Nancy Cox**, The University of Chicago, USA
Genome studies of cancer risk and response to chemotherapeutic agents

10:20 ***Coffee break and poster session***

10:50 **Jochen Supper**, Genomatix Software GmbH, Munich, Germany
Short talk: *Investigating and visualizing various levels of genetic variations for understanding the molecular causes of breast cancer*

11:10 **Núria Malats**, CNIO, Madrid, Spain
Bladder cancer Epidem&OMICS

11:50 **Søren Brunak**, CBS, Technical University of Denmark, Lyngby, Denmark
Integration of data from the molecular level and the clinic: disease gene finding and patient stratification

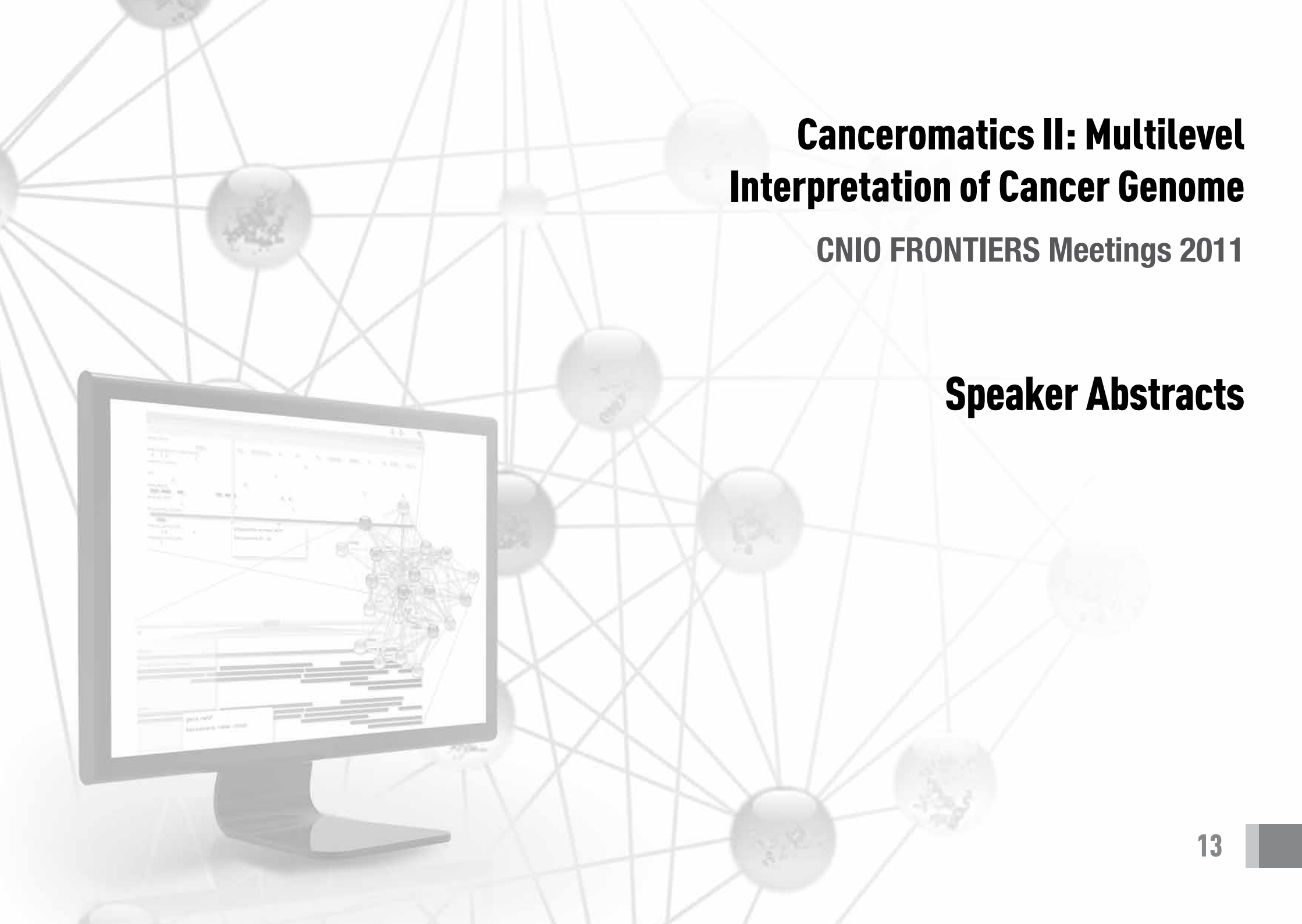
12:30 **Closing Remarks**
Chris Sander, Alfonso Valencia

13:00 ***END of session***

Note: Talks: 30 minutes / **Short talks:** 15 minutes
Discussion: 10 minutes after each **talk** / 5 minutes after each **short talk**

notes

CNIO FRONTIERS
Meetings 2011
Detailed programme



Canceromatics II: Multilevel Interpretation of Cancer Genome

CNIO FRONTIERS Meetings 2011

Speaker Abstracts

Canceromatics II: Multilevel Interpretation of Cancer Genome

CNIO FRONTIERS Meetings 2011

Session 1

Chairs: Angela Brand, Søren Brunak

Genome-wide association studies in cancer: discovery and characterization

Stephen Chanock

National Cancer Institute (NCI), National Institutes of Health, Bethesda, USA

Genome-wide association studies (GWAS) have emerged as an important tool for discovering regions of the genome that harbor genetic variants that confer risk for different types of cancers. The success of GWAS in the last five years has resulted from the convergence of new technologies that can genotype hundreds of thousands of single-nucleotide polymorphism markers together with comprehensive annotation of genetic variation applied to large scale studies. The agnostic approach has already yielded more than 125 distinct regions for nearly two dozen cancers and the susceptibility alleles discovered thus far are common with a frequency of >10% and each allele confers a small contribution to the overall risk for the disease. For nearly all regions conclusively identified by GWAS, the per allele effect sizes estimated are <1.3. With additional scanning and modeling, we can begin to appreciate the differences in genetic architectures for common variation only for specific diseases. We also can see the complex nature of germline genetic variation and its contribution to cancer susceptibility. For instance, in prostate cancer, there are probably more than 40 distinct regions harboring common susceptibility alleles already identified by GWAS, whereas in lung cancer, a disease strongly driven by exposure to tobacco products, so far, only three regions have been conclusively established. Interestingly, half a dozen regions have been associated with more than one distinct cancer type, suggesting these are nexus points in the genome, worthy of intense follow-up studies. GWAS are an important discovery tool that require extensive follow-up to map each region, investigate the biological mechanism underpinning the association and eventually test the optimal markers for assessing risk for a disease or its outcome, such as in pharmacogenomics, the study of the effect of genetic variation on pharmacological interventions. The success of GWAS has opened new horizons for exploration and highlighted the complex genomic architecture of disease susceptibility. Though initially prosecuted on a disease by disease basis, it is now possible to begin to look at common regions that influence heterogeneous traits relevant to cancer.

Finding the mutation in the alignment file!

Francis Ouellette

Ontario Institute for Cancer Research, Toronto, Canada

Computational biology approaches offer and present a multitude of tools and resources that would allow one to interpret and interrogate what is hidden in the many cancer genomes we have the opportunity to study. In this presentation I will summarize how the OICR executes this, and how we plan to make sense of our data, as well as that from the ICGC as a whole. I will also present some of our recent work on the identification of gene signatures to predict cancer outcomes based on the domain interaction network in human proteome. The results of the quantification in each patient were used to predict cancer outcome by a modified naïve Bayes classifier. In this study, we achieved a favorable accuracy, sensitivity and specificity on a set of well-documented gene expression profiles of breast cancer patients with different outcomes. We also compiled a list of cancer-associated gene signatures and domains, which provided testable hypotheses for further experimental investigation. Our approach proved successful on different independent breast cancer data sets as well as an ovarian cancer data set. This study constitutes a predictive method to classify cancer outcomes based on the relationship between the domain organization and protein network. I will also present how this kind of approach offers new possibilities and directions (and challenges) with data from the ICGC initiative.

Public Health Genomics goes personalized – lessons learned from cancer genomics

Angela Brand
Maastricht University, The Netherlands

With the dynamics of genomics and related fields like epigenomics and systems biology a more rapid translation of scientific evidence into products and a sharp decrease of sequencing costs is predicted. Ultimately the policy impact is higher than expected 5 years ago as the merger of diagnostics and pharmaceuticals (theranostics) and the individual risk modulation put a question mark behind market authorization and market regulation systems currently used in Europe. The personalization of risk prediction, drug response and the understanding of the etiology of diseases change the role of citizens and require a new degree of health literacy. A comprehensive model of future healthcare taking into account integrative genomics alongside with environmental, social and life style factors from various data sources as well as new policy tools will become essential to enable personalized strategies in prevention, early detection and treatment of disease.

The field of cancer is estimated to be the first, where personalized medicine reaches everyday healthcare practice. It has been acknowledged years ago that each tumor is unique with its own set of driver mutations which dysregulate key genomic pathways. Currently, cancer is classified based on its tissue of origin such as breast or colon, but due to full-genome sequencing and computational modelling much more powerful classification based on the tumor genotype is possible. The unique combination of activated signaling pathways will demand personalized diagnosis and treatment, depending on the level of metastasis and the availability of pharmaceuticals.

Thus, the real issues are rather to be found in the field of policy making and knowledge management, and, consequently, the real paradigm shift of health care systems depends on the willingness to restructure policies. Public Health Genomics aims to prepare policy makers and health care systems in time.

Modelling biological networks involved in tumor progression

Emmanuel Barillot
Curie Institute, Paris, France

Tumorigenesis and tumour progression are consequences of the deregulation of molecular networks controlling essential cellular processes such as cell cycle, DNA repair, cell death or survival among others. Improving therapeutic strategies rests therefore on the understanding and modelling of these biological networks.

By analysing literature we have constructed detailed maps of these networks using systems biology standard languages. Analysis of the structure of these networks allows their decomposition into functional modules and their organisation into interconnected layers.

We have then focused on the modelling of an important process which is involved in cancer: the mechanism of cell fate decision upon TNF and FAS death receptor engagement, by which a cell chooses to die by apoptosis or by necrosis, or to survive. We have conceived a mathematical model of cell fate decision, based on a logical formalisation of well-characterised molecular interactions. The dynamical modelling allowed us to analyse the temporal behaviour of a cell for different conditions, and to infer the phenotype reached (apoptosis, necrosis or survival). This was shown to recapitulate satisfactorily the experimental observation both for wild type and known mutants. The model also allows proposing prediction of mechanisms involved in the control of cell death.

In addition to the above mentioned several tools developed for network analysis and omics data analysis (in particular next identification of mutations by generation sequencing) will also be presented.

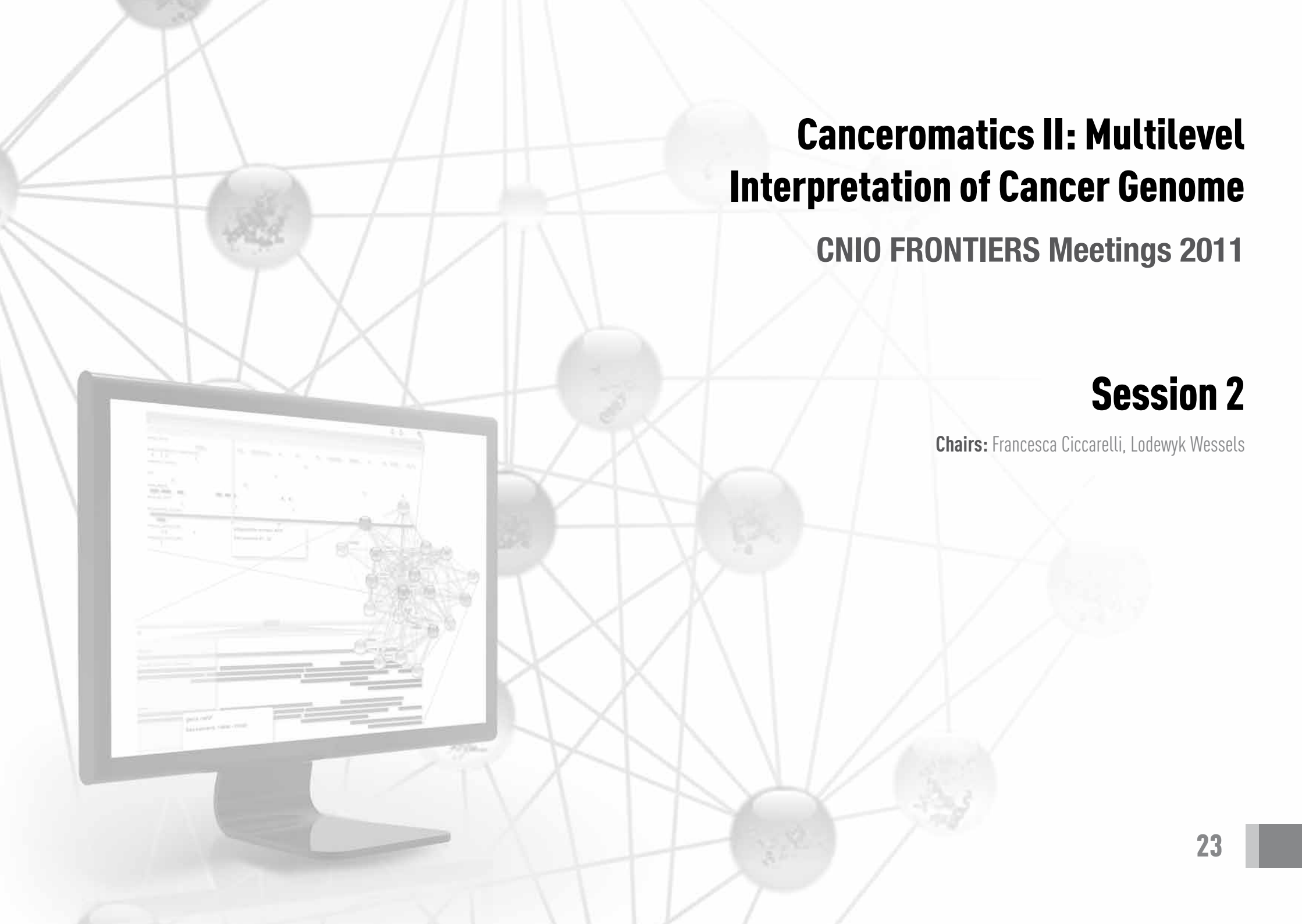
Short Talk

A sparse regulatory network of copy-number driven gene expression reveals putative breast cancer oncogenes

Yinyin Yuan, Christina Curtis, Carlos Caldas and Florian Markowetz
Cancer Research UK, Cambridge Research Institute, UK; University of Cambridge, UK

Copy number aberrations are recognized to be important in cancer as they may localize to regions harboring oncogenes or tumour suppressors. Such genomic alterations mediate phenotypic changes through their impact on expression. Both cis- and trans- acting alterations are impor tant since they may help to elucidate putative cancer genes. However, amidst numerous passenger genes, trans-effects are less well studied due to the computational difficulty in detecting weak and sparse signals in the data, and yet may influence multiple genes on a global scale. We propose an integrative approach to learn a sparse interaction network of DNA copy-number regions with their downstream transcriptional targets in breast cancer. With respect to goodness of fit on both simulated and real data, the performance of sparse network inference is no worse than other state-of the ar t models but with the advantage of simultaneous feature selection and efficiency. The DNA-RNA interaction network helps to distinguish copy-number driven expression alterations from those that are copy-number independent. Fur ther, our approach yields a quantitative copy-number dependency score, which distinguishes cis- versus trans-effects. When applied to a breast cancer dataset, numerous expression profiles were impacted by cis-acting copy-number alterations, including several known oncogenes such as GRB7, ERBB2 and LSM1. Numerous trans-acting alterations were also identified, impacting genes such as ADAM2 and BAGE, which warrant further investigation.

notes



Canceromatics II: Multilevel Interpretation of Cancer Genome

CNIO FRONTIERS Meetings 2011

Session 2

Chairs: Francesca Ciccarelli, Lodewyk Wessels

A natural combination of genomics and cancer research

Huanming Yang

Beijing Genomics Institute, China

Historically, cancer research was one of the major thinking-resources of the Human Genome Project (HGP) in the last century. In return, the completion of the HGP has provided cancer research with new strategies and novel technologies. Scientifically, genomics has supported the idea that cancer is a genomic disease not only involving one or a few genes. Technically, the advent of sequencing technology, especially its latest generation, has made feasible the large-scale, high-throughput and low-cost studies on cancer genome. Canceromatics is a natural and scientific combination of genomics and cancer research.

Cancer is the first cause of death in China at present. Approximately 1.6 million people die of cancer and more than 2.2 millions new cases are diagnosed The Chinese Cancer Genome Consortium (CCGC) was organized in 2008 as a publicly-funded network to coordinate a number of projects, sharing the common goals for unraveling the genomic alterations in cancers, performed by over 20 hospitals and research institutions in China. CCGC has announced the selection of gastric cancer and 12 other common cancers in China for its first phase.

CCGC is strongly supported technically by the biggest sequencing and bioinformatics platforms at BGI (formerly Beijing Genomics Institute), which has a capacity of completing 100 human genomes a day presently. Following its previous contribution to the HGP and HapMap projects, BGI and its collaborators have recently published the sequencing and analysis of the first Asian genome, "The 1000 human genomes", the human pan-genome, gut metagenome, transcriptome and methylome, as well as an ancient genome, in addition to the giant panda, ant, soybean, maize and other genomes. The preliminary results from studies on gastric cancer have proved the advantages of new-generation sequencing technologies and set up a successful model for other types of cancers.

Systems biology of microRNAs and cancer

Debora Marks

Harvard Medical School, Boston, USA

microRNAs are commonly dys-regulated in cancer tissues, and changes have been found associated with metastasis and survival. Small RNA eg siRNAs are now far long in development as therapies in several cancer types. But how do we predict the consequences of siRNAs and changes in microRNA levels have on all the molecular components in a cell? Recent work shows that individual regulation by microRNAs and siRNAs will depend on how they affect *many* global cellular components as well as characteristics of individual binding sites in target mRNA molecules. This means that small RNAs should be considered in the light of their global effect on cellular behaviour and not only individual targets. *Saturation* of protein machinery in the RNAi pathway may affect both siRNA and microRNA targeting. We have shown that targets of *endogenous* miRNAs are expressed at significantly higher levels after transfection, often showing upregulation of oncogenes. *Competition* between different mRNAs for the microRNAs or siRNAs could, according to basic chemical kinetics, affect the targeting quantitatively causing crosstalk between different mRNAs. We also show how mRNAs with a high turnover rate in the cell are more resistant to perturbation by small RNAs, with consequences for genome-wide siRNA studies and siRNA therapeutics.

I will talk about the consequences of this systems approach to microRNA and siRNA behaviour to understanding the genome wide changes in various cancer types such as liposarcoma, and ovarian cancer, using data from The Cancer Genome Atlas. Specifically, I will discuss the relationship of microRNA copy number changes and the consequences on the gene network, specific pathways and other clinical measurements in the cancer tissue data.

The BLUEPRINT of hematopoiesis

Henk Stunnenberg

Nijmegen Centre for Molecular Life Sciences (NCMLS), Radboud University Nijmegen Medical Centre, The Netherlands

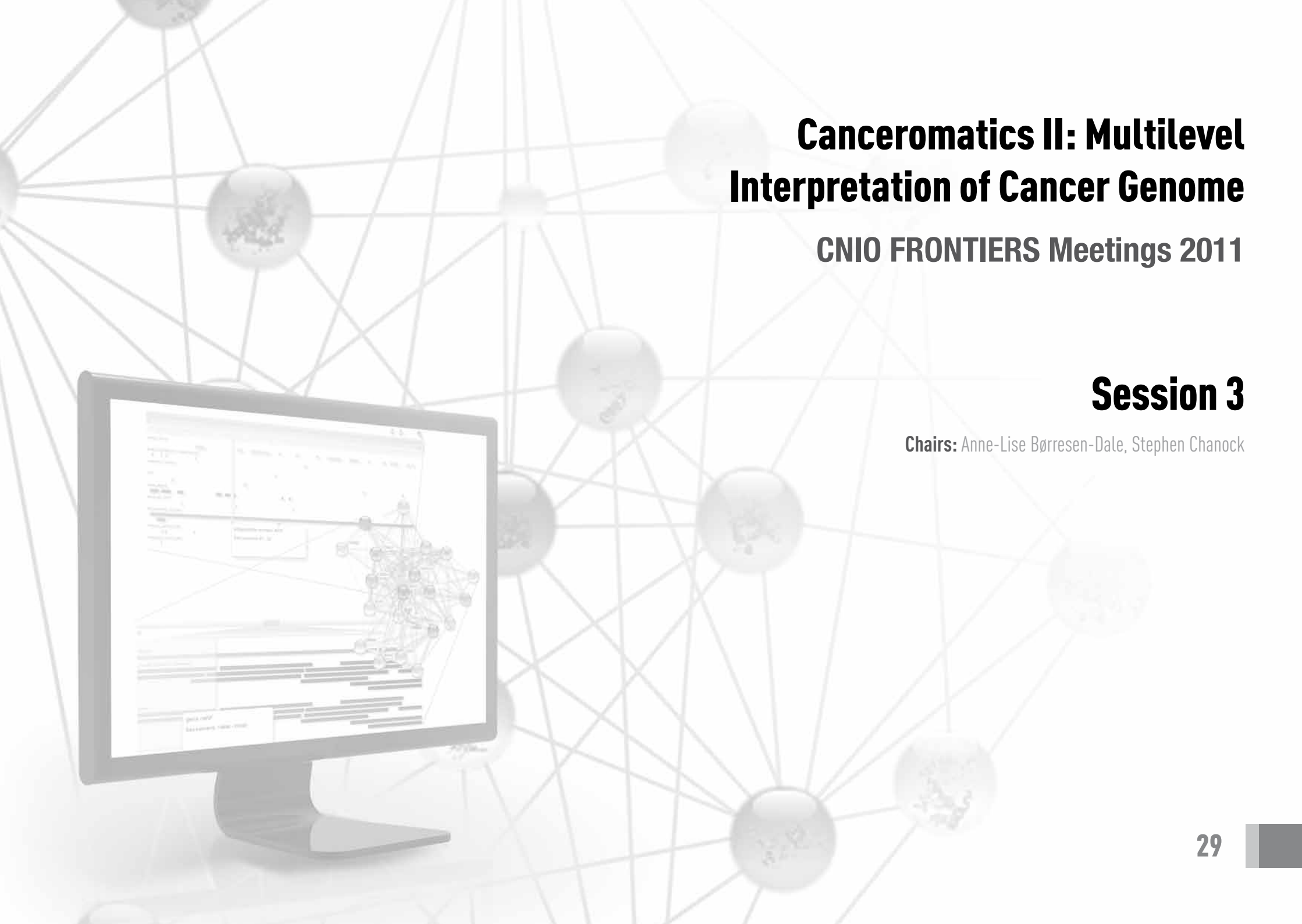
The regulation of gene expression is paramount in growth, development, differentiation, signaling, adaptation to the environment and many other processes. Gene expression is regulated at many levels, but primarily by binding of specific transcription factors to regulatory regions, resulting in the recruitment 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 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. The increasing sequence capacity enables for the first time the genome wide identification and integration 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. I will discuss our 'epigenetic/systems' biology approach to gain molecular insight into the action of oncofusion proteins in Acute Myeloid Leukaemia.

Cancer genome data provide a comprehensive, "freeze-frame" view of dynamic genomic instability processes in human cancer samples

Zoltan Szallasi

Centre for Biological Sequence Analysis (CBS), Technical University of Denmark, Lyngby, Denmark

Exploiting the "cancer-ome" data to understand the ***actual*** relevant processes in human cancer samples faces several, seemingly insurmountable challenges. In addition to the well-known issue of overfitting of high dimensional data sets, one is often a investigating a single human biopsy of high cellular heterogeneity. From these measurements one hopes to reconstruct complex dynamic processes, such as the subtype and level of genomic instability maintaining human cancer. Fortunately, these processes, such as chromosomal instability or DNA repair incompetence, seem to leave their footprints that are recognizable in "cancer-ome data" sets. In particular, based on biological knowledge, we postulated the existence of a specific pattern of allelic imbalance to characterize BRCA1 function incompetence. Interestingly, this pattern turned out to be a reliable predictor of monoagent platinum therapy in hormone receptor negative breast cancer and ovarian cancer in independent validation cohorts. We also applied our chromosomal instability measure to investigate whether the previously postulated "optimal zone" of genomic instability of human tumors exists and can be reliably identified. In several human cancer cohorts, untreated with chemotherapy, we could clearly demonstrate that over a certain threshold of genomic instability patients will have a significantly better outcome, supporting the idea that extreme levels of chromosomal instability are impeding the tumor's survival.



Canceromatics II: Multilevel Interpretation of Cancer Genome

CNIO FRONTIERS Meetings 2011

Session 3

Chairs: Anne-Lise Børresen-Dale, Stephen Chanock

Cancer genomics and network pharmacology

Chris Sander

Memorial Sloan-Kettering Cancer Center, New York, USA

The talk will bridge network analysis of complex cancer genotypes from The Cancer Genome Atlas and a new computational-experimental approach to network pharmacology that makes use of network inference algorithms inspired by statistical physics and aims to design optimal combination therapy.

(1) Network analysis of many individual cancer profiles led to the concept of *common specification* and *individual implementation* of *oncogenic programs* and to the notion of pathway-oriented therapeutic perturbation.

(2) To make perturbational cell biology more computable, we have developed a statistical method for deriving predictive network models from high-throughput molecular measurements of systematically perturbed cellular systems. The network models aim to predict quantitatively the molecular and physiological outcomes of unseen perturbations, such as drug treatments or multiple genetic alterations. We anticipate useful applications in network pharmacology and the design of combinatorial cancer therapy.

Network-based identification and systematic characterization of cancer signatures

Igor Jurisica

Ontario Cancer Institute, Toronto, Canada

Cancer development is a multi-step process that leads to uncontrolled tumor cell growth. Multiple pathways are involved, typically some signalling and regulatory pathways are activated, while others are suppressed. Existing prognostic and predictive biomarkers overlap only partially, and often do not validate on external cohorts or by other biological assays. Several technical reasons may explain this. Further evidence shows that several clinically identical gene signatures exist, and that the best signature may not comprise the most differential genes.

To address these challenges, we developed a system for an integrative analysis, prediction and characterization of signatures and relevant protein-protein interactions and microRNA: gene interactions.

These computational predictions will lead to improved human interactome coverage relevant to both basic and cancer biology, and importantly, may accelerate the development of novel therapies for many cancers.

Cancer & cellular information processing

Rune Linding

Technical University of Denmark (DTU), Lyngby, Denmark

Biological systems are composed of highly dynamic and interconnected molecular networks that drive biological decision processes. The goal of systems biology is to describe, quantify and predict the information flow and functional behaviour of living systems in a formal language and with an accuracy that parallels our characterisation of other physical systems such as Jumbo-jets. Decades of targeted molecular and biological studies have led to numerous pathway models of developmental and disease related processes. However, so far no global models have been derived from pathways, capable of predicting cellular trajectories in time, space or disease. The development of high-throughput methodologies has further enhanced our ability to obtain quantitative genomic, proteomic and phenotypic readouts for many genes/proteins simultaneously. Here, I will discuss how it is now possible to derive network models through computational integration of systematic, large-scale, high-dimensional quantitative data sets. I will review our latest advances in methods for exploring phosphorylation networks. In particular I will discuss how the combination of quantitative mass-spectrometry, systems-genetics and computational algorithms (NetworKIN [1] and NetPhorest [4]) made it possible for us to derive systems-level models of JNK and EphR signalling networks [2,3]. Finally, I shall discuss work we have done in comparative phospho-proteomics and network evolution[5,6].

References: http://www.lindinglab.org. [1]Linding et al., Cell 2007. [2]Bakal et al., Science 2008. [3]Jørgensen et al., Science 2009. [4]Miller et al., Science Signaling 2008. [5]Tan et al., Science Signaling 2009. [6]Tan et al., Science 2009

Oncogenic network identification from genomic data

Lodewyk Wessels

The Netherlands Cancer Institute, Amsterdam, The Netherlands

Biological complexity stems from the large number of possible interactions between molecular entities occurring at different scales. Gene-gene interactions include co-operating interactions (both genes required) redundant (either of the genes required) and mutually exclusive interactions (only one gene required). Typically, (combinations of) these interactions give rise to a specific phenotype, such as cancer. Pathways can be viewed as a compilation of such interactions, and dissecting pathways can be translated to the problem of mapping molecular interactions. High throughput insertional mutagenesis screens combined with same-sample gene expression data as well as copy number profiles from tumor (cell-line) cohorts offer the opportunity to systematically identify complex interactions and pathways. We will describe computational approaches to identify 1) candidate cancer genes; 2) complex interactions between these and 3) associate downstream target activities to regulatory gene interactions. These approaches reveal networks of molecular interaction that underlie cancer development furthering our understanding of oncogenesis and facilitating the development of targeted treatment.

Short Talk

Non-neutral copy number alterations in cancer genome data

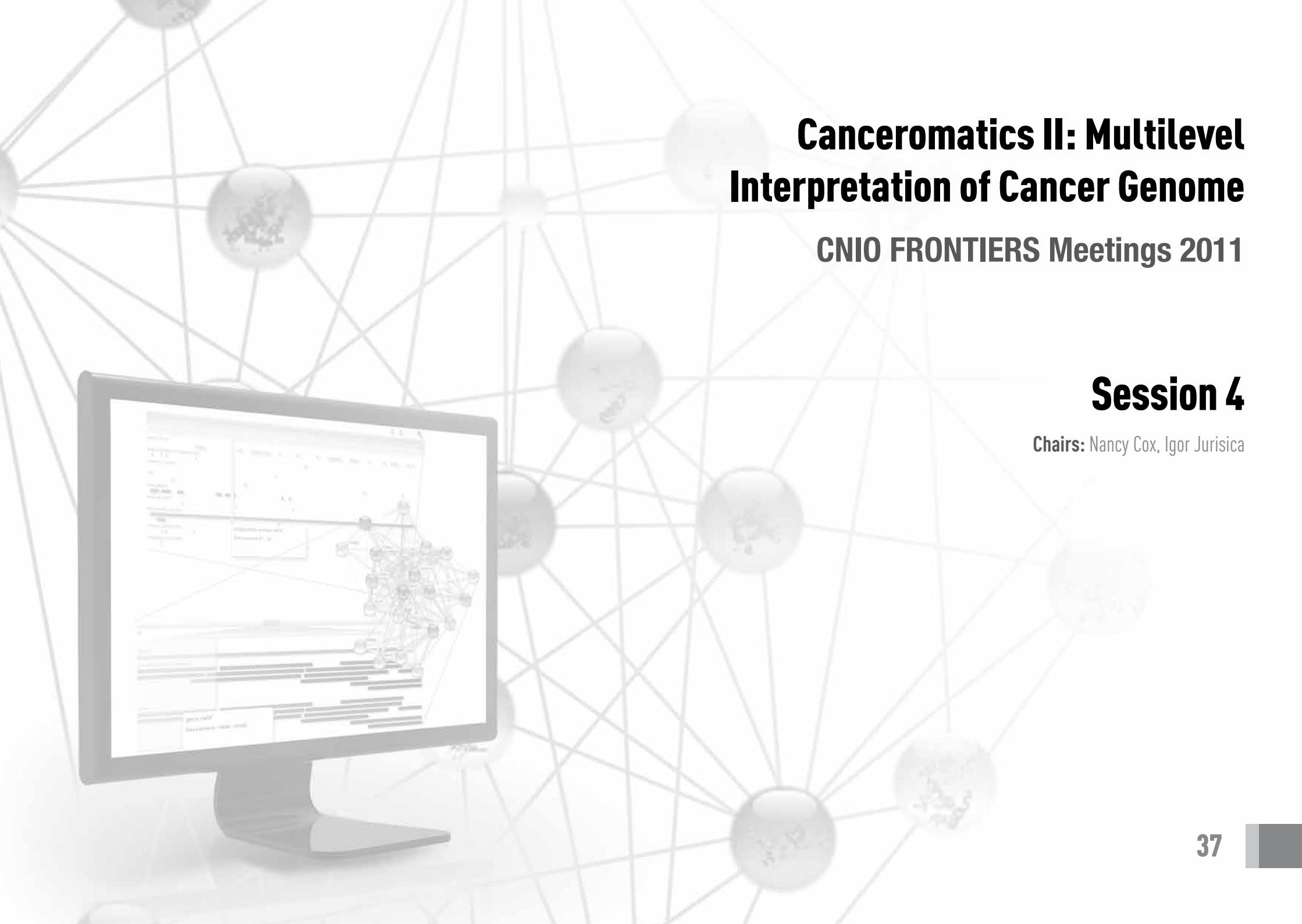
ST

Nitin Kumar, Christian Von Mering, Cai Haoyang and Michael Baudis
Institute of Molecular Life Sciences, University of Zurich, Switzerland

Cancer copy number alterations (CNA) from several studies have been focused on identifying cancer specific high frequent CNA. However, these studies have mostly been restricted to specific cancer types or single institution data sets. A cross comparison study across multiple cancer types should be able to identify genomic changes with either ubiquitous or entity specific impact. While some of the highly frequent changes can be found in many cancer types, some CNAs are specific for a limited number of entities. Statistically, such changes will have an impact on e.g. the clustering of cancer types. We call such cancer type distinguishing CNA ¨non-neutral CNA¨. Biologically, non-neutral CNA may provide useful information about genetic divergence among different cancer types. We address two questions considering CNA across virtually all cancer types: 1. Role of an individual CNA in distinguishing cancer types:To address this question we use a permutation based methodology to identify non-neutral CNA playing a role in clustering of cancers. 160 cancer types are clustered with hierarchical clustering and trees obtained on permutation are compared with original tree to find out CNA which result in a less structured tree. 2. Frequent CNA across cancer types: To find out frequent CNA across all cancer types binomial test is used, resulting in information of genomic regions which show CNA more often then by chance. Our CNA analysis methods were applied to 25579 samples from 160 cancer types as available through our group's Progenetix database (www.progenetix.org). According to our combined analysis, only a group of frequent CNA play a role in clustering cancer entities. However, a number of genomic imbalances with overall low frequency had a high impact on the detection of CNA profile similarities between different cancer entities. Those genomic regions with possible biological relevance may be overlooked when focussing on genomic ¨hot spot¨.

CNIO FRONTIERS
Meetings 2011
Session 3 | Tuesday, March 29th

notes



Canceromatics II: Multilevel Interpretation of Cancer Genome

CNIO FRONTIERS Meetings 2011

Session 4

Chairs: Nancy Cox, Igor Jurisica

The practicalities of the international cancer genome consortium analyses

Ivo Gut

National Center for Genome Analysis (CNAG), Barcelona, Spain

The objective of the International Cancer Genome Consortium is to fully characterize in the 50 most common forms of cancer 50 tumoural samples and then to validate observations in further 450 samples. Our main task is whole genome sequence analysis, but also whole exome analysis, RNA sequence analysis and epigenetic analysis is being tackled. Complete genome sequencing of many samples requires bringing together many different elements, starting from samples, preparation for sequencing, sequencing itself, data analysis, through to verification of results and translating a result into biological knowledge. The ICGC project is structured so that countries commit to cancer entities, such as Spain that is working on Chronic Lymphocytic Leukaemia. The ICGC project has been organised with working groups that are focussing on the key elements. The structure of the ICGC project will be outlined as well as some of the advances of the working groups. Some of the first observations of sequencing cancer genomes will be presented.

Integrated molecular profiles of primary breast tumors reveal differential interleukin signalling and improved prognostic power

V.N. Kristensen, C.J. Vaske, D. Haussler, B. Naume, O. Troyanskaya and Anne-Lise Børresen-Dale*

**Inst. Cancer Research, Oslo University Hospital Radiumhospitalet, Norway*

The accumulation of high throughput molecular profiles of tumors at various levels has been a long and costly process worldwide. Combined analysis of gene regulation at various levels may point to specific biological functions and molecular pathways that are deregulated in multiple epithelial cancers and reveal novel subgroups of patients for tailored therapy and monitoring. We have collected high throughput data at several molecular levels derived from fresh frozen samples from primary tumors and matched blood from approximately 110 breast cancer patients (the MicMa dataset) with information about presence of disseminated tumor cells (DTC) and long term follow-up. The MicMa set has been used in parallel pilot studies of whole-genome mRNA expression (1), arrayCGH (2), DNA-methylation (3), whole genome SNP-CGH array (4), whole-genome miRNA expression analyses (5), TP53 mutation status dependent pathways (6) and high throughput paired-end sequencing (7). We have explored the extent to which these new profiles, acquired by molecular analyses at various levels recapitulate the initially discovered expression subclasses and their prognostic potential. Further, we attempted to integrate these by modeling mRNA, CNAs, miRNAs, and methylation in a pathway context using Paradigm (Pathway Recognition Algorithm using Data integration on Genomic Models). We were able to show that combining mRNA expression and DNA copy-number leads to better discrimination of patients with respect to prognosis than any of the molecular levels studied separately. The pathways whose perturbations most strongly contributed to this classification were ERBB4, also previously found deregulated in breast and ovarian cancers, FoxM1 transcription, deregulated also in the ovarian and glioblastoma TCGA datasets, Angiopoientin receptor Tie2-mediated signaling and most notably the immune response (TCR) and interleukin signaling, IL4, IL6, IL12 and IL23 signaling. Massive interleukin signaling and similar sub-classification was observed also in DCIS (ductal carcinoma in situ) together with extracellular matrix and cell-cell adhesion regulating pathways.

[1]Naume B et al. (2007) Mol. Oncol. 1, 160-171; [2]Russnes HG and Vollan HK et al. (2010) Sci. Transl. Med. 2, 38ra47; [3]Ronneberg J A et al. (2011) Mol. Oncol. 5, 61-76; [4]Van LP et al. (2010) Proc. Natl. Acad. Sci. U. S. A. 107, 16910-16915; [5]Enerly E and Steinfeld I et al. (2011) PLoS. One. 6, e16915; [6]Joshi H et al. submitted (2011); [7]Stephens P et al. (2009) Nature. 462, 1005-1010.

Towards the protein network bases analysis of cancer genome data

Alfonso Valencia
CNIO, Madrid, Spain

I will introduce our general proposal (Baudot et al., Genome Biology 2009) for the use of protein interaction networks to analyze cancer associated variants, and the initial results of the analysis of pathways specific to certain cancer types (Baudot et al., EMBO Rep. 2010, Glaab et al., BMC Bioinfo 2010).

In this context the use of protein networks critically depends on the accuracy to which the specific interactions are described. Our efforts to increase the quality of the information on protein interactions are two fold.

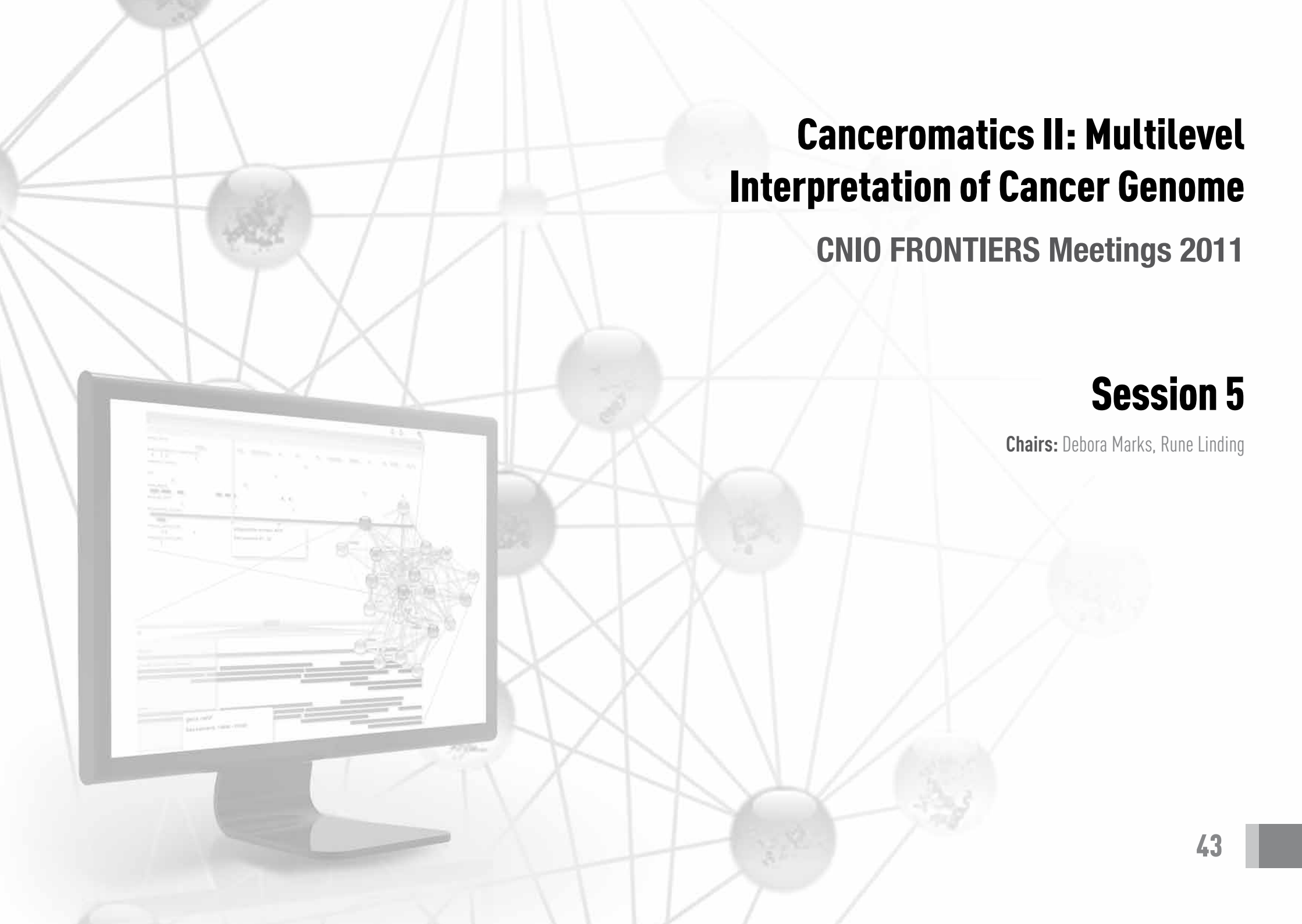
First, we are using evolutionary information to: a) discover networks of concerted interactions (Juan et al., PNAS 2008), b) detect residues potentially responsible of binding specificity (Rausell et al., PNAS 2009), and c) we are developing approaches to assess the specificity of interaction surfaces (Wass et al., Mol Sys Biol 2011).

Second, we are interested in complementing these bioinformatics prediction approaches with information derived directly from primary publications. An approach increasingly possible thanks to the development of the text mining methodology as reviewed in the BIOCREATVE II.5 challenge (Leitner et al., Nat Biotech 2010 and IEEE/ACM Trans Comput Biol Bioinfor 2010).

Genomic instability and the evolution of cancer

Francesca Ciccarelli
IFOM-IEO Campus, Milan, Italy

Cancer is often associated to genomic instability, defined as the dynamic acquisition and propagation of genomic modifications. In several cases these modifications are causally implicated in the initiation and development of the disease. The latest improvements in sequencing technologies have boosted the discovery of cancer-associated modifications, thus allowing the investigation of several aspects of cancer genomic instability. In this talk, our contribution to the understanding of the genetic and genomic determinants of cancer will be presented. Particular emphasis will be laid on the properties of genes whose mutations are causally implicated in cancer and on the detection of cancer-associated genomic instability.



Canceromatics II: Multilevel Interpretation of Cancer Genome

CNIO FRONTIERS Meetings 2011

Session 5

Chairs: Debora Marks, Rune Linding

Facilitating combination therapy development: and you thought it was difficult to invent a single drug

Kenneth Fasman
Dr. Miriam and Sheldon G. Adelson Medical Research Foundation, Needham, USA

More than 150 years ago, Paul Ehrlich led us off on a quest to create "magic bullets" to cure disease. And the molecular biology and genomics revolutions have only exacerbated the problem, seducing us to view disease through the lens of individual drug targets. But the small number of real successes in curing or arresting disease progression of the past few decades show us that instead of magic bullets we should be creating Swiss Army knives, automatic weapons, and MIRVs to attack these monsters.

This presentation will examine the challenges of identifying, developing and delivering combination therapies, with examples drawn from oncology and neural repair. Today, much of the emphasis of combination therapy development is on testing two or more already-approved compounds together in Phase IV (i.e., post-approval) clinical trials. We will instead focus on the earlier steps in the modern drug discovery and development process, reviewing the ramifications of combinatorial therapy development for systems biology, screening assays and strategies, structure-based drug design, ADME/tox, etc. We will also briefly consider strategic issues, such as dosing and phasing of components, and comparing the use of multiple therapeutic molecules versus single molecules with multiple pharmacophores or multiple sites of action.

Genome studies of cancer risk and response to chemotherapeutic agents

Nancy Cox
The University of Chicago, USA

We have conducted studies using second-generation analytic tools for GWAS using both data on response to chemotherapeutic agents in assays using a cell line model and on primary risk of cancer. I will describe here both some of these new methods of analysis and results of our studies with these methods looking across different chemotherapeutic agents and different types of cancer. One focus of the studies is to determine whether there are genetic risk factors that cut across drugs and drug classes, and also genetic risk factors that are common to both cancer risk and to response to chemotherapeutic agents.

Short Talk

Investigating and visualizing various levels of genetic variations for understanding the molecular causes of breast cancer

Jochen Supper, Philippe Youkharibache, Frederic Eyber and Martin Seifert
Genomatix Software GmbH, Münich, Germany

Next-generation sequencing (NGS) allows to characterize cells at the genomic, transcriptomic and epigenetic level. For a normal cell to become a cancer cell changes on all these levels accumulate over many years. Hence, to understand a certain type of cancer on the molecular level a comprehensive analysis based on all these levels is needed.

In this work, we investigate seven different breast cancer cell lines by analyzing and integrating various types of NGS datasets, these are: DNA-seq, RNA-seq, small RNAseq, DNA and methylation sequencing datasets. In the first level of analysis, all these datasets are analyzed individually to generate a compendium of genomic, transcriptomic and epigenetic alterations for each breast cancer cell line. This compendium includes gene fusions, SNPs, copy number variations (CNVs), alternative splicing, methylation sites in regulatory regions and small-RNA expression.

In the second level of analysis, the various alterations are combined to generate a comprehensive picture of cellular misregulation in cancer. Regulatory networks are generated that combine transcriptomic changes with CNVs, methylation alterations and expert-curated annotations. By linking different lines of evidence, the results of this work allow to unravel and underpin genetic alterations in cancer.

To present these results visually we developed a network and transcriptome viewer. The network viewer allows to depict gene expression, methylation and additional expert curated information in a network view for all breast cancer cell lines. The transcriptome viewer, in addition, allows to view changes in transcript expression and novel splice-junctions for all cell lines. Overall, this study pinpoints and visualizes many genomic, transcriptomic and epigenetic alterations that are potential causes for the different types of breast cancer under investigation.

Bladder cancer Epidem&OMICS

Núria Malats
CNIO, Madrid, Spain

At present, epidemiologists face several challenges in assessing and integrating the complexity of omics data with extensive information on environmental factors and behaviors at the individual level in large studies. Some of the challenges regard to the huge heterogeneity of the genetic alterations found through the new sequencying technologies emphasizesing the notion of individualized diseases; the complexity of highly correlated environmental exposures that are making us to move from a candidate to an agnostic/exploratory exposure analysis incorporating epigenomics and metabolomics markers to better dissect the still ¨missing exposurome¨; the notion that nothing is static during an individual's lifespan and the need to statistical modeling this dynamic process; the necessity to consider intermediate outcomes such as gene expression and/or protein function and, at the same time, the need to consider comprehensive phenomes to better account for the underlying genetic pleiotropic effect in human organisms living in selected environmental conditions. Following the current concept of epidemiological study design and statistical power requirements, very large sample sizes are needed to explore the underlying biological complexity in a meaningful manner. The integration of several types of data, from environmental exposures to epigenetics, metabolomics, as well as genetics, genomics, transcriptomics, and proteomics information, requires the development of innovative bioinformatics and data reduction techniques. I will exemplify some of this challenges with results from the Spanish Bladder Cancer / EPICURO Study.

Integration of data from the molecular level and the clinic: disease gene finding and patient stratification

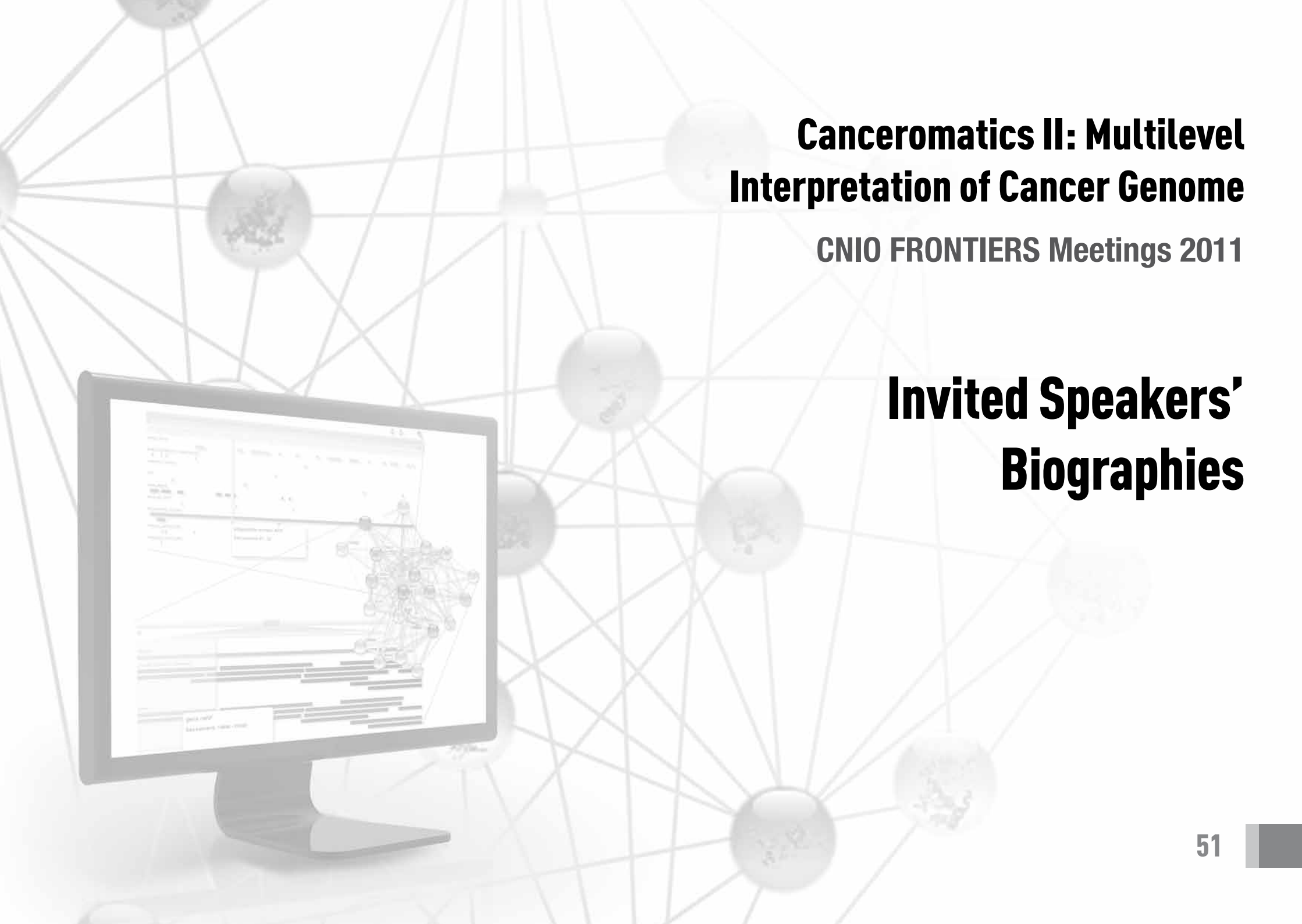
Søren Brunak

Centre for Biological Sequence Analysis (CBS), Technical University of Denmark, Lyngby, Denmark

Data integration within biology is a quite old idea which typically can produce more coherent views of biological systems when different structural levels are taken into account simultaneously. Combining molecular level data, such as DNA sequences, quantitative gene and protein expression data, small molecule abundances, is relatively common within bioinformatics, while it is less common to integrate the molecular level with the clinical level, for example text data from electronic patient records describing disease phenotypes of the individual. Electronic patient records remain a rather unexplored, but potentially rich data source for discovering correlations between diseases. We describe a general approach for gathering phenotypic descriptions of patients from medical records in a systematic and non-cohort dependent manner (e.g. in testis cancer). By extracting phenotype information from the free-text in such records we demonstrate that we can extend the information contained in the structured record data, and use it for producing fine-grained patient stratification and disease co-occurrence statistics. The approach uses a dictionary based on the International Classification of Disease ontology and is therefore in principle language independent. Combining molecular level data with clinical information and data on the chemical environment may add complementary types of knowledge which – together with genotype information from the individual – can reveal disease aetiology in novel ways.

CNIO FRONTIERS
Meetings 2011
Session 5 | Wednesday, March 30th

notes



Canceromatics II: Multilevel Interpretation of Cancer Genome

CNIO FRONTIERS Meetings 2011

Invited Speakers' Biographies

Stephen Chanock

National Cancer Institute (NCI), National Institutes of Health, Bethesda, USA

Dr. Chanock received his M.D. from Harvard Medical School in 1983 and completed clinical training in pediatrics, pediatric infectious diseases, and pediatric hematology/oncology and research training in molecular genetics at Boston Children's Hospital and the Dana-Farber Cancer Institute, Boston. Afterwards, he joined the NCI Pediatric Oncology Branch where he eventually became a Senior Investigator. Dr. Chanock's scientific research has been focused on the contribution of genetic variation to cancer risk and outcomes. In 2001, he became director of the NCI Core Genotyping Facility and in 2005 he became co-leader of the Cancer Genetic Markers of Susceptibility (CGEMS) project. In 2007, he became the Chief of the newly formed Laboratory of Translational Genomics. He is currently on sabbatical at the *Institut Curie* in Paris.

Francis Ouellette

Ontario Institute for Cancer Research, Toronto, Canada

Francis Ouellette is a bioinformatician who studies databases, protein networks, and data integration. He is the associate director of the Informatics and Biocomputing platform as well as a Principle Investigator at the Ontario Institute for Cancer Research in Toronto, Ontario. He also coordinates the Canadian Bioinformatics Workshop (CBW). Francis was trained at McGill, Calgary and Simon Fraser University. After working at the Yeast genome project at McGill University and at the NCBI as GenBank Coordinator, he went to UBC in 1998 where he was an Associate Professor in Medical Genetics and Director of the UBC Bioinformatics Centre. He moved to Toronto in August 2007 and is now applying his skills and interest to the study of cancer genomes.

Angela Brand

Maastricht University, The Netherlands

Prof. Angela Brand, MD PhD MPH (USA) is Director of the Institute for Public Health Genomics (IPHG) at the Faculty of Health, Medicine and Life Sciences at Maastricht University, the Netherlands, and Director of the European Centre for Public Health Genomics (ECPHG). She is Coordinator of the Public Health Genomics European Network (PHGEN), President of the Section Public Health Genomics within the European Public Health Association (EUPHA), Editor-in-Chief of the international journal Public Health Genomics, Excecutive Director of the Public Health Genomics international network GRaPH-Int, Associated Member of the international consortium Public Population Project in Genomics (P^3G) on biobanking, Advisory Board Member of Alacris Pharmaceuticals, Advisory Board Member of OncoTrack (IMI), Scientific Consultation Group Member of the European Centre for Disease Prevention and Control (ECDC), Co-Investigator of PIE (Pharmacogenomics & Innovation in Ethics) within the NIH Pharmacogenetics Research Network (PGRN), and Fellow of the Rockefeller Foundation; USA, and of the 21st Century Trust of the Wellcome Trust, UK.

Emmanuel Barillot

Curie Institute, Paris, France

Emmanuel Barillot earned an engineering degree from Ecole Centrale de Paris in 1988. After military obligations as an Air force officer he completed his PhD in Bioinformatics in 1992 at Généthon/CEPH in Paris where he headed the bioinformatics group up to the construction of the first physical map of the human genome in 1995. Then he created a research group in genomic data integration at Infobiogen, the French national bioinformatics center. In early 2000 he joined the National Agricultural Research Institute (INRA) to create a unit for plant bioinformatics, focusing on heterogeneous data integration. In late 2002 he joined institut Curie to create a bioinformatics core facility and then a research group in computational systems biology of cancer. Since 2008 he is heading the Department of Epidemiology, Biostatistics, Bioinformatics and Computational Systems Biology of Cancer (U900), a joint Institut Curie and INSERM research department, in partnership with Mines ParisTech.

His research focus on statistical analysis of large scale biological data, integration of heterogeneous biological data and biological network modeling in the context of cancer.

Huanming Yang

Beijing Genomics Institute, China

Dr. Yang received his Ph.D. from University of Copenhagen, Denmark, in 1988, and obtained his postdoctoral trainings in France and USA. As one of the co-founders of BGI, he and his collaborators have made a significant contribution to the HGP and HapMap projects, as well as to sequencing and analysing genomes of rice, cucumber, chicken, silkworm, giant panda, ant, maize, soybean, and many microorganisms since 1999. BGI, together with its collaborators, has published the first Asian's genome, human pan-genome, human ancient genome, and human gut metagenomes by means of new-generation sequencing technoloty and innovative bioinformatic tools recently in Science, Nature and other internationally prestigious journals.

Debora Marks

Harvard Medical School, Boston, USA

Debora started life in Medicine followed by a stint in the pharmaceutical industry, saw the light and went to do a degree in Mathematics, followed by a PhD in Mathematical Biology. She worked in Manchester University and Harvard in the Systems Biology department and has strong connections to Humboldt University in Berlin. She is fascinated by RNA genes, the evolvability of the human genome, how genetic variation with the environment produce the phenotypes we know and love; as well as those phenotypes which we love rather less, like diseases. Her work includes systems biology of microRNAs and cancer.

Henk Stunnenberg

Nijmegen Centre for Molecular Life Sciences (NCMLS), Radboud University Nijmegen Medical Centre, The Netherlands

In 1996, Hendrik G. Stunnenberg was appointed full professor and head of the Department of Molecular Biology, Science and Medical Faculties, Radboud University Nijmegen, The Netherlands. He was a group leader at the EMBL Heidelberg, Germany in the Gene Expression program from 1985 to 1996, where he studied the action of nuclear receptors and also contributed very significantly to deciphering basal transcription processes. He is a member of EMBO since 1992. He is the founder and main organizer of the biennial EMBL meeting on transcription.

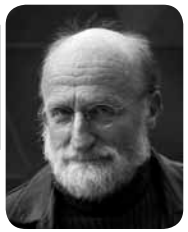
He is the coordinator of the EU-FP6 Integrated Project HEROIC that focuses on Highthroughput Epigenetic Regulatory Organisation In Chromatin of mouse Embryonic Stem cells and their derivatives. Within HEROIC, his group is performing ChIP-seq of histone modifications in wild type and KO lines. His group was amongst the first in Europe to establish ChIP in combination with next generation sequencing platform, and he acquired and is routinely running an Illumina GAII platform with paired-end module since 2007. The group participates in EU consortia focusing on ChIP-seq profiling of oncogenic translocation fusion proteins in AML and histone marks. His lab has established genome wide DNA methylation profiling at high resolution (MeDIP-seq and Methy lCap-seq) and is performing genome wide DNA methylation profiling in cancer. His lab is also involved in consortia aiming at establishing and implementing novel, highly sensitive protocols and novel cross-linking methods for ChIP and ChIP-seq on very small numbers of cells.

His research focuses on deciphering the genetic and epigenetic mechanisms of gene regulation during development, differentiation and in cancer.

Zoltan Szallasi

Centre for Biological Sequence Analysis (CBS), Technical University of Denmark, Lyngby, Denmark

Dr. Szallasi's group is developing computational methods to analyze the various genome scale molecular profiles of human cancer samples. They have produced several lines of evidence that genome scale molecular profiling of cancer is an effective method to identify the specific genomic instability subtype of a given tumor; and this information is often one of the key determinants whether a patient will respond to a given chemotherapeutic agent. The wider, epistemological context of their work is focused around how to exploit genome scale measurements to robustly characterize essential cancer biology while avoiding data overfitting inherent of high throughput measurements.



Chris Sander

Memorial Sloan-Kettering Cancer Center,
New York, USA

Chris Sander is acknowledged as one of the founders of the field of computational biology, an interdisciplinary field that aims to solve important problems in biology using techniques of mathematics, physics, engineering, and computer science. He is Head of the Computational Biology Center at Memorial Sloan Kettering Cancer Center and Tri-Institutional professor at Rockefeller and Cornell Universities.

Sander's current research interests are in computational and systems biology, including predictive simulations of biological processes, gene regulation by small RNAs, protein folding, cancer genomics and the development of combinatorial cancer therapy. He is active in the International Cancer Genomics Consortium, the NIH Cancer Genome Atlas Project, the NCI Integrative Cancer Biology Program and a leader in the bioPAX and PathwayCommons community efforts to create an open-source information resource for biological pathways.

Previously Sander served as Chief Information Science Officer with Millennium Pharmaceuticals, as Senior Scientist at the European Bioinformatics Institute in Cambridge, England, and as founding chair of the department of Biocomputing at the European Molecular Biology Laboratory in Heidelberg.

Sander supports open access publications and the transfer of military expenditures to scientific endeavors, especially those focused on the survival of humanity on the planet.

Igor Jurisica

Ontario Cancer Institute, Toronto, Canada

Igor Jurisica, Canada Research Chair in Integrative Computational Biology, is a Senior Scientist at the Ontario Cancer Institute, Associate Professor in the Departments of Computer Science and Medical Biophysics at University of Toronto and Visiting Scientist at IBM's Centre for Advanced Studies. He is also an Adjunct Professor at School of Computing, Queen's University and Graduate Program in Computer Science, York University.

His research focuses on integrative computational biology and the representation, analysis and visualization of high-dimensional data from high-throughput biology experiments in the cancer informatics context.

Interests include comparative analysis for mining different integrated data sets (e.g., protein-protein interactions, gene expression profiling, and high-throughput screens for protein crystallization).

Rune Linding

Technical University of Denmark (DTU),
Lyngby, Denmark

Dr Rune Linding is Professor and Research Group Leader for the Cellular Signal Integration Group (C-SIG) at the Technical University of Denmark (DTU), Center for Biological Sequence Analysis (CBS), Department of Systems Biology, Denmark. He performed his graduate work at the EMBL (Germany), where he pioneered computational signalling biology by developing popular tools like ELM, GlobPlot and DisEMBL for analysing post-translational modifications and intrinsic protein disorder. Dr Linding was Human Frontiers Science Program Postdoctoral Research Fellow jointly with Profs Tony Pawson and Mike Yaffe at Samuel Lunenfeld Research Institute (Canada) and Massachusetts Institute of Technology (USA), respectively. His postdoctoral work on the cellular phosphorylation networks and development of the NetworKIN algorithm pioneered Integrative Network Biology and led to the discovery of the importance of contextual kinase specificity. He started his own lab (the Cellular & Molecular Logic Team) at The Institute of Cancer Research (ICR) in London in 2007. At ICR his lab unravelled systems-level models of JNK and EphR kinase networks, demonstrated a link between specificity and oncogenecity of kinases and introduced the concept of Network Medicine. Dr Linding leads the NetPhorest community resource and have pioneered comparative phospho-proteomics and evolutionary studies of signalling networks. Dr Linding founded the Integrative Network Biology initiative (INBi) which aims to block cancer metastasis by integration of large-scale, high-dimensional quantitative genomic, proteomic and phenotypic data. His lab moved to DTU in 2011 and the long-term focus of his research group is on studying cellular signal processing and decision making.

CNIO FRONTIERS
Meetings 2011
Invited Speakers' Biographies

Lodewyk Wessels

The Netherlands Cancer Institute,
Amsterdam, The Netherlands

After obtaining a PhD in Electronic Engineering he joined the Delft University of Technology. In 2006 he became a faculty member and head of the Bioinformatics and Statistics group at the Netherlands Cancer Institute (NKI-AVL) in Amsterdam, The Netherlands. His research mainly focuses on developing novel computational approaches to exploit a wide variety of data sources to expand our understanding of the biology underlying oncogenesis and to improve cancer diagnostics and treatment. Dr Wessels is co-director of the Cancer Systems Biology Center at the NKI-AVL. The goal of this center is to build in silico models of signaling pathways in (breast) cancer to enable improved diagnosis and therapy selection.

Ivo Gut

National Center for Genome Analysis (CNAG), Barcelona, Spain

He qualified in Chemistry at the University of Basel in 1985 and obtained a PhD in Physical Chemistry from the same university in 1990. Then he joined the research group of Prof. Irene Kochevar at Harvard Medical School as a research fellow.

Between 1993 and 1996 he was research fellow with Dr. Stephan Beck at the Imperial Cancer Research Fund in London. Later, at the Max-Planck-Institute for Molecular Genetics he led a group in the Department for Vertebrate Genomics of Prof. Hans Lehrach. In the 11 years before joining the CNAG he worked at the Centre National de Génotypage of the Commissariat à l'Energie Atomique in Evry, France, first as Head of Technology Development and later as Associate Director under Prof. Mark Lathrop. In January 2010 he took on the direction of the Centro Nacional de Análisis Genómico.

His research interests are high-throughput nucleic acid analysis, sequencing, SNP genotyping, genomics, genetics, nucleic acid and protein analysis methods (molecular biological techniques and chemical modification), implementation of methods, automation and analysis. He is the author of over 100 research papers, inventor of 24 patents and patent applications, founder of 4 biotech companies (Genom Analytik GmbH, Biopsytec GmbH and Epigenomics AG, Integragen SA). He is the coordinator of the 12M€ EU FP7-funded Integrated Project READNA on DNA sequencing technology development.

Anne-Lise Børresen-Dale

Oslo University Hospital Radiumhospitalet, Oslo, Norway

Professor Børresen-Dale, Department of Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet is among the leading geneticists in research on molecular biology of breast cancer. She earned international reputation for pioneering research on gene expression profiling and molecular portraying of Breast Cancer into 5 subtypes, and has published more than 350 papers in peer reviewed International Journals and received a numerous of Awards. Her current research interests are "Exploring the systems Biology of Breast Cancer" using high dimensional data in integrated approaches.

Børresen-Dale was recently appointed to head the K.G. Jebsen Center for Breast Cancer Research.

Alfonso Valencia

CNIO, Madrid, Spain

PhD in Molecular Biology and Biochemistry, Madrid 1988. Postdoctoral EMBL-Heidelberg. Chris Sander's group, 1988-1994. PI Spanish Research Council (CSIC) 1994-2006. Full Professor of the Spanish Research Council. Director of the Spanish National Bioinformatics Institute (INB). Director of the Structural Biology and BioComputing Programme, CNIO. Member of EMBO since 2005. Member of the EMBL, Biozentrum Basel and Swiss Institute for

Bioinformatics Scientific Advisory Committees. Coordinator of the Spanish Bioinformatics network (1996 to 2004). EMBO fellowship (1989-1990). Award of the CSIC for the year accomplishments (2004, 2005). Founder, former VP and member of the board of the International Society for Computational Biology. Founder of the organization behind the European Conference of Computational Biology and co-organized the 2005 Conference. Member of the ESF programme on "Functional Genomics" (2000-2011). Founder and Organizer of the ISCB Text Mining Interest Group (Biolink). Coorganizer of the BioCreative community challenge. Assessor of CASP. Member of the "banquete" science-and-art society. Founder of the BioAlma company. Participant in international projects, including the NIH ENCODE consortium, the three main EU Networks: BioSapiens, Embrace and Enfin, and of the European infrastructure for Bioinformatics initiative. Executive Editor of Bionformatics (Oxford U. Press). Member of the Editorial Board of EMBO J. and EMBO Rep. Chair of the Systems Biology and/or Text Mining of the main Computational Biology Conferences. Research activity focussed on the study of protein structure, function and evolution at the genomic level in the context of cancer research.

Francesca Ciccarelli

IFOM-IEO Campus, Milan, Italy

Francesca Ciccarelli is Principal Investigator in evolutionary genomics of cancer at the European Institute of Oncology (Milan) since mid 2005. She graduated in Pharmaceutical Chemistry at the University of Bologna in 1998 and obtained the PhD in Natural Science at the University of Heidelberg in 2003. She has been research assistant in the group of Peer Bork at the EMBL for five years.

Dr. Ciccarelli has an extensive expertise in comparative and evolutionary genomics, which now applies to study the genetic determinant of cancer. Her lab focuses on two main lines of investigation:

- Cancer genome re-sequencing using next-gen technology
- Systems biology of cancer genes

Kenneth Fasman

Dr. Miriam and Sheldon G. Adelson Medical Research Foundation, Needham, USA

Dr. Ken Fasman is Chief Science Officer of the Adelson Medical Research Foundation, which promotes collaborative, goal-directed translational research in oncology and neurology.

From 1992-1996, Dr. Fasman was at Johns Hopkins University as Informatics Director of the Human Genome Database, and Assistant Professor of Biomedical Information Sciences, Neuroscience, and Biomedical Engineering. From 1996-1998, he was Sequencing Informatics Director at the MIT Center for Genome Research. From 1998-2007, Dr. Fasman was global director of Drug Development Strategy & Performance and VP of R&D Informatics at AstraZeneca PLC.

Dr. Fasman earned a B.S.E. in Electrical Engineering / Computer Science from Princeton, and a Ph.D. in Biomedical Engineering from Johns Hopkins.

Nancy Cox

The University of Chicago, USA

Nancy Cox is a quantitative human geneticist with a long-standing research program to identify and characterize genetic risk factors for complex human traits. The research in her lab focuses on the development of analytic methods to accomplish this goal, and she has funded research to develop methods for the analysis of GWAS, sequence data from the 1000 Genomes Project, and data from the new GTEx project, and in addition conducts studies in pharmacogenomics, mesothelioma, breast cancer, diabetes and diabetic complications, asthma and related phenotypes, Tourette Syndrome and obsessive compulsive disorder, autism, stuttering and specific language impairment. She completed a PhD in Human Genetics at Yale University in 1978, did post-doctoral research at Washington University, and worked briefly at the University of Pennsylvania, before joining the University of Chicago in 1987, where she has spent the remainder of her career, and is now Professor and Chief of Genetic Medicine at the University of Chicago.

Núria Malats

CNIO, Madrid, Spain

Dr. Malats is MD and PhD by the Universitat Autònoma de Barcelona. During her PhD, she carried out experimental and epidemiological work on pancreas cancer (PANKRAS2 Study) integrating multidisciplinary interests. From 1996-1998 she was visiting scientist at the International Agency for Research on Cancer (IARC-WHO), Lyon, in the group of Dr. Paolo Boffeta, where she worked in genetic epidemiology of lung cancer. Since 1997, she is a co-PI of the Spanish Bladder Cancer/EPICURO Study, a large case-control study on bladder cancer that integrates scientific interests from different disciplines in both cancer development and progression. In 1998 we got a position of Scientist by the Sistema Nacional de Salud-ISCIII and returned to Barcelona where she set up a group at Institut Municipal d'Investigació Mèdica (IMIM) and Centre de Recerca en Epidemiologia Ambiental (CREAL), leading and participating in both national and international competitively funded projects and coordinating the Spanish research network on bladder cancer (EPICUR-red). Dr. Malats joint CNIO in September 2007 focussing her work on the genetic and molecular epidemiology of pancreas, bladder, and breast cancer. She is the Coordinator and a co-PI of a large European case-control study on genetic epidemiology of pancreas cancer (PanGen-EU Study) and of an integrated European study on bladder cancer (ISBlaC Study). Dr. Malats is the Spanish delegate of the Public Health and Genomics European Network (PHGEN) and member of the Scientific Board of the ESUR and the European Association of Urology Foundation Research. She has over 100 publications, has coordinated and participated in teaching activities, and is external reviewer of national and international funding agencies and first rank scientific journals.

Søren Brunak

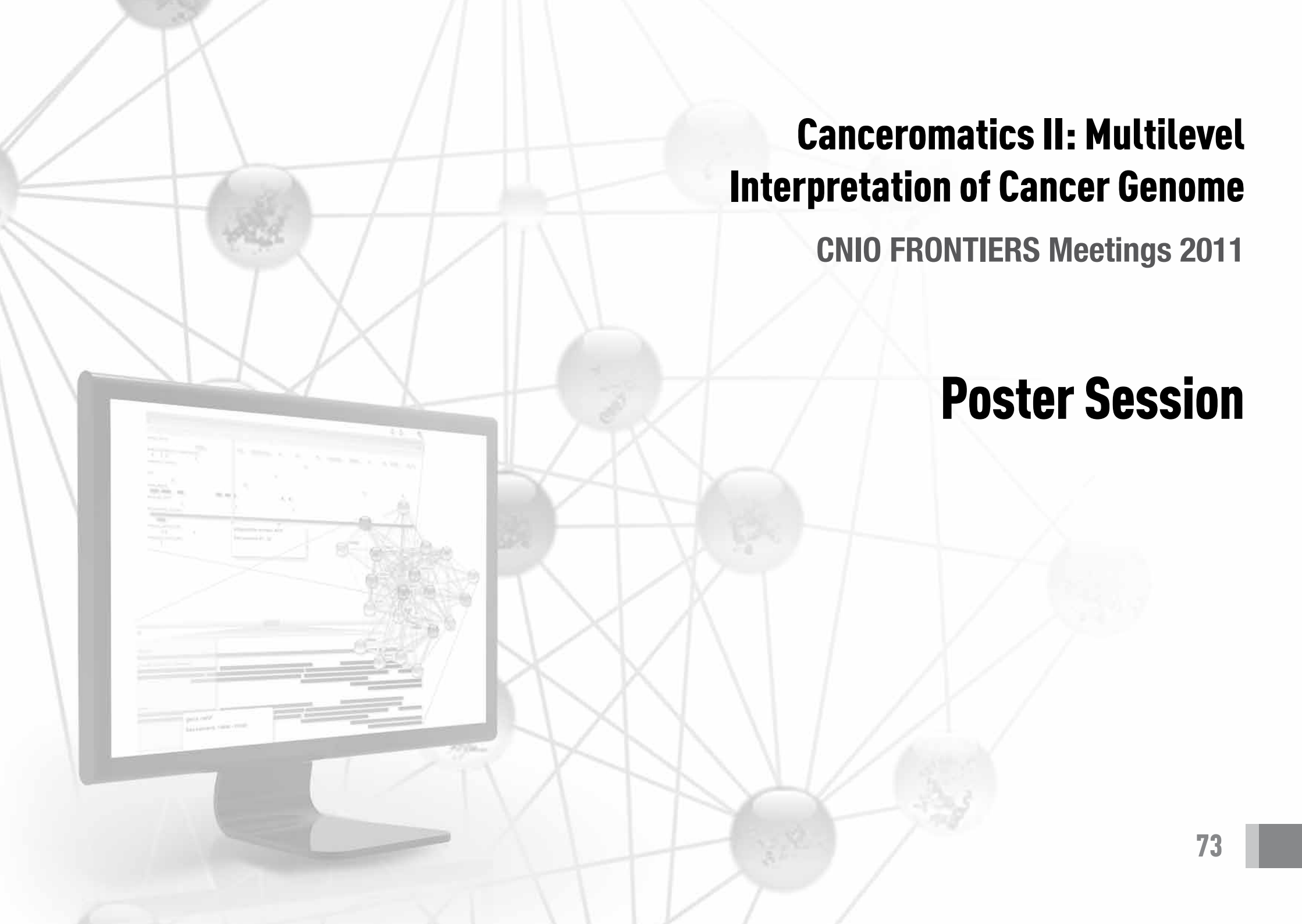
Centre for Biological Sequence Analysis (CBS), Technical University of Denmark, Lyngby, Denmark

Søren Brunak obtained his M.Sc. in Physics, in 1987 at the Niels Bohr Institute, University of Copenhagen, in 1991 his Ph.D. in Computational Biology at the Department of Structural Properties of Materials, Technical University of Denmark, and in 2002 a Dr.phil. (honoris causa) from the Natural Science Faculty of the Stockholm University.

He is member of the Danish Academy for the Natural Sciences since 1997, member of the Board of directors of the International Society for Computational Biology since 2001, of the Danish Academy of Technical Sciences since 2002, and of the Danish Royal Society of Science and Letters since 2004.

He is part of the scientific advisory committees of several scientific organizations, such as EMBL (Heidelberg), Ensembl at the European Bioinformatics Institute/Sanger Centre (chairman), the Bioinformatics Advisory Committee at the European Bioinformatics Institute (chairman), and the Max Planck Institute for Molecular Genetics (Berlin).

Dir. Ib Henriksens Price for Outstanding Science Achievement. Villum Kann Rasmussen Price for Research within the Natural and Technical Sciences (2006).



Canceromatics II: Multilevel Interpretation of Cancer Genome

CNIO FRONTIERS Meetings 2011

Poster Session

Copy number imbalances in human malignancies: a reference data set for systems biology

Haoyang Cai, Nitin Kumar and Michael Baudis
Institute of Molecular Life Sciences, University of Zurich, Switzerland

DNA copy number aberrations (CNAs) contribute to cancer pathogenesis and progression. Over the past several years, array comparative genomic hybridization (aCGH) techniques have proven their value for analyzing CNAs. By providing an additional layer of information, CNA data could prove valuable for systems biology approaches to modeling pathways and functional networks involved in oncogenesis.

However, no single resource does yet provide a global collection of interpreted

("called") oncogenomic array data. Such a resource is urgently needed for the cancer research community, which motivated us to undertake the effort to construct arrayMap - a database for annotated array CGH data.

Besides tracking publications for suitable aCGH data content, we explored NCBI's Gene Expression Omnibus (GEO) for oncogenomic data series. So far, we were able to identify 476 series with approximately 35,000 samples of human neoplasias. Important obstacles for generating a common reference data set are the necessities to a) determine the genomic positions for the tens to hundreds of thousands array probes with b) reference to a common genome Golden Path edition and c) to perform optimal copy number region segmentation for d) calling regions of copy number gain and loss while dealing with different data resolution and dynamics. So far, we have remapped 11395 arrays from 237 series for which we were able to obtain normalized probe values. For Affymetrix GenomeWide SNP arrays, we could circumvent these issues by performing a massive re-analysis of raw data sets. Here, we identified approx. 17,000 Affymetrix GenomeWide arrays in GEO, the majority of which was suitable for re-processing from raw data files.

As a public resource, the called copy number data as well as associated information will be presented through our regularly updated online database - arrayMap (www. arraymap.org) and through the Progenetix resource (www.progenetix.org).

Potential function of proteins encoded by chimeric transcripts

Milana Frenkel-Morgestern, Iakes Ezkurdia, David Pisano, Michael Tress and Alfonso Valencia
CNIO, Madrid, Spain

Chimeric RNAs are produced by means of two or more distinct transcripts, have been reported in different organisms including fruit fly, mouse, and human[1,2]. Thousands of chimeric transcripts were identified among the variety of ESTs for these organisms. In theory, those transcripts can greatly contribute to the complexity of the transcriptome and proteome of organisms[1,2], and nowadays are validated by the new generation sequencing technologies. However, a question of how the machinery producing chimeras once exists can be used to design new functional proteins still remains unclear.

We have studied the potential functional role of the chimeric proteins in human, mouse, fruit fly and yeast. For this purpose, we annotated both proteins participating in creation of the chimeric transcripts described by Li et al[2]. Using an integrated mass-spectrometry pipeline we found that chimeras combine signal peptides and trans-membrane domains of some membrane proteins in their formation. Moreover, some chimeras containing full functional domains, do not include the activation domain in the final chimeric protein, but only the DNA binding domain. In additional, for proteins functional as dimers, the function domain sometimes is missed, but dimerization domain is preserved. Finally, we report examples of novel chimeric proteins, which might produce in some cases the dominant negative effect in cells.

[1]Gingeras TR, Implications of chimeric non-co-linear transcripts, 2009, Nature, 461, pp 206-211; [2]Li X, Zhao L, Jiang H, Wang W, Short homologous sequences are strongly associated with the generation of chimeric RNAs in eukaryotes, 2009, J Mol Evol, 68:56-65.

Gene-gene interactions in a genome-wide prognostic study of bladder cancer: a novel analytical strategy

Jesús Herranz[1], **Stephen Chanock**[2], **Antoni Picornell**[1], **Nathaniel Rothman**[2], **Montserrat García-Closas**[2], **María L. Calle**[3], **Debra Silverman**[2], **Manolis Kogevinas**[4], **Francisco X. Real**[1,5] **and Núria Malats**[1]

[1]CNIO, Madrid, Spain; [2]National Cancer Institute, Bethesda, USA; [3]Universitat de Vic, Spain; [4]CREAL/IMIM-Hospital del Mar, Barcelona, Spain; [5]Universitat Pompeu Fabra, Barcelona, Spain

The study of gene-gene interactions is a fundamental step in the identification of susceptibility genetic variants involved in complex disease aetiology and progression. The exhaustive analysis of all gene-gene interactions in genomewide studies with survival data has not been yet addressed because several billions of interactions should be tested considering time-dependent tests, making it computationally unfeasible.

We propose a novel and feasible two-step strategy to analyse all pair-wise interactions in a genome-wide study of 1 million probes with survival data. The strategy is based in the adaptation of several tools designed for case-control studies and includes a screening step using logistic regression using Boost followed by Cox regression analysis.

We have studied prediction of tumour recurrence in 836 bladder cancer cases with non-invasive tumours from the Spanish Bladder Cancer /EPICURO Study. 585k SNP were selected for the analysis after having applied quality control filters and removing SNPs in LD. Out of the 171 billions of gene-gene interactions generated with the selected SNPs, we detected 23 interactions with P-values < 1E-10 and 235 interactions with P-values < 1E-9 worth to be further explored. The whole analysis lasted 10 days.

We propose an exhaustive and feasible strategy to analyse genomewide gene*gene interaction in a survival data setting. We have shown, in a real case, that this strategy is able to obtain a large number of significant gene-gene interactions in a reasonable computational time.

Predicting the pathogenicity of mutations within the human kinome

Jose M.G. Izarzugaza, **Angela del Pozo and Alfonso Valencia**

CNIO, Madrid, Spain

Human Protein Kinases are involved in a plethora of physiological functions. A broad number of mutations in the Protein Kinase superfamily have been reported to disrupt protein structure and/or function. Still, the majority of protein kinase mutations are tolerated without apparent significant effects. In a scenario where the mechanistic interpretation of the consequences of the mutations on the phenotypes is unfeasible, current research aims to identify correlations between mutations and human disease, particularly cancer, using indirect and circumstantial relations. Here we present the basis for the development of a computational method for the prediction of the impact of mutations in the function of protein kinases. We associate a number sequence-derived features to disease-associated kinase mutations, including: a) at the gene level, the membership to a Kinbase group and Gene Ontology terms. b) at the domain level, the occurrence of the mutation inside a PFAM domain, and c) at the residue level, properties including amino acid type, functional annotations from Swissprot and FireDB, specificity-determining positions, etc.

We examined the independent significance of these properties and their combination, using a Support Vector Machine (SVM). Interestingly, the kinase-specific features arise among the most discriminative information sources, which justifies the development of a predictor customized for this protein family. In summary, our study intends to provide a means for the fast and accurate priorization of kinase mutations. In addition, we discuss the benefits and pitfalls of using the information available for the development of a kinase-specific predictor with regard to other current prediction methods. Likely, in the future, the study of the distribution of propensities associated to specific mutation/kinase groups would be useful for the analysis of mechanisms by which mutations in the human kinome contribute to disease with a particular focus in cancer.

CNV detection and association with bladder cancer risk: an application to a 1M illumina SNP-array data

Gaëlle Marenne[1,2], Stephen J. Chanock[3], Luis Pérez-Jurado[4], Nathaniel Rothman[3], Benjamin Rodríguez[4], Manolis Kogevinas[5], Montse García-Closas[3], Debra T. Silverman[3], Francisco X. Real[1], Emmanuelle Génin[2] and Núria Malats[1]

[1]CNIO, Madrid, Spain; [2]Inserm UMR-S946, Univ. Paris Diderot, Institut Universitaire d'Hématologie, Paris, France; [3]National Cancer Institute, Bethesda, USA; [4]Universitat Pompeu Fabra, Barcelona, Spain; [5]Institut Municipal d'Investigació Mèdica, IMIM-Hospital del Mar, Barcelona, Spain

Copy number variants (CNV) influence phenotypic traits. Results from CNVassociation studies strongly depend on the accuracy of CNV assessment mainly if they are affected by differences between cases and controls. While they can be detected using SNP-arrays data by applying calling algorithms, previous studies highlighted the low sensitivity of these algorithms [Marenne, Hum Mutat 2011]. Here, we aim to 1) replicate the well-known association between the homologous deletion at the GSTM1 gene and bladder cancer risk, and 2) identify characteristics influencing sensitivity and false positive rate (FPR) of CNV detection algorithms. We used genetic data from the Spanish Bladder Cancer/EPICURO Study for 1150 cases and 1149 controls [Rothman, Nat Genet 2010] generated by the genome-wide Illumina 1M, jointly with 21 HapMap samples.

We observed over-measured SNP intensities in the GSTM1 region due to the high frequency of deletions in the Spanish population, leading to inaccurate CNV callings. Nevertheless, we replicated the association with bladder cancer using the SNP intensities from cases and controls as a raw CNV measurement. At the genome-wide level, we compared the CNV callings obtained with 4 algorithms applied to the HapMap sample data with those reported by Conrad [Nature 2010]. We observed that the sensitivity increased with CNV size and the FPR improved in regions where probe density was higher.

We reported a potential misclassification problem in calling CNVs from SNP-array in regions of common aberration, such as GSTM1. Nevertheless, our work supported the use of SNP-array data as a powerful genome screening tool to identify altered genomic areas associated with a particular disease. Furthermore, by accounting for CNV region's characteristics related to sensitivity and FPR, it may be possible to limit false positive association signals arising from differences in the accuracy of CNV detection in cases and controls.

Genomic analysis of pancreatic cancer

Raquel Martínez, Daniel Rico, Manuel Hidalgo and Alfonso Valencia

CNIO, Madrid, Spain

We have analyzed somatic copy number alterations (CNAs) of genomic regions in pancreas cancer and their association with the expression levels of genes therein. We profiled 30 different human pancreas adenocarcinomas. Gene expression levels were quantified using Affymetrix U133-plus, and to measure CNAs we have used two different platforms: Illumina Genotyping arrays 1M and Agilent aCGH 244K. After independent processing and normalization of each copy number dataset, we used the same methods for the segmentation and calling, finally obtainning the probability of the probes of being in a copy number state (loss/gained/normal). To compare the both platforms, we divided the autosomal genome in bins of 10 Kb. We observed that, when both platforms had probes in the same bin, the copy number was positively correlated in most of them. To integrate the information from both platforms, we calculated the median probability of all the probes mapping to each bin, so we finally had a unique probability of being gained or lost for each genomic bin. To associate CNAs and expression levels, we searched for the most frequently altered genes, calculating the average probability of all bins mapping to the same gene. We found 7077 frequently gained genes and 2743 genes lost at 0.5 frequency or higher. Then, for each of these genes, we grouped the samples in two classes: samples where that gene a gain/loss probability greater than 0.5, and the rest of samples (with lower probability of having the alteration.) We compared expression levels between the two classes of using Welch T-test and we found 354 genes that had their gene expression levels significantly different depending on their copy number (False Discovery Rate < 10%). 138 of them were genes over-expressed when gained and 216 genes were down-regulated when lost. Functional analysis of these genes showed association with cell cycle and DNA repair processes.

Deep sequencing of small non-coding RNAs in breast tumours

Helena Persson[1], Anders Kvist[1], Natalia Rego[3], Johan Staaf[1,2,4], Johan Vallon-Christersson[1,2,4], Göran Jönsson[1,2,4], Hugo Naya[3], Åke Borg[1,2,4] and Carlos Rovira[1,2,4]

[1]Lund University, Sweden; [2]SCIBLU Genomics, Centre for Integrative Biology at Lund University, Sweden; [3]Institute Pasteur Montevideo, Uruguay, [4]CREATE Health, Strategic Centre for Translational Cancer Research, Lund University, Sweden

MicroRNAs (miRNAs) are ~22 nt small non-coding RNAs that regulate gene expression mainly at the post-transcriptional level. They are frequently de-regulated in cancer and can act as tumour suppressors or oncogenes. We performed deep sequencing for breast tumours with paired adjacent and normal breast tissue for a total of 15 samples. With more than 6 million aligned sequences per sample the small RNA transcriptome can be described in great detail. The majority of miRNAs are common to all samples, but with large variation in expression levels, especially between tumours and normal breast tissue. In addition to differentially expressed miRNAs we identified hundreds of new miRNA loci and more than a hundred star sequences of known miRNAs. Mature miRNAs from 49% of our new miRNA precursors were also detected by sequencing of Argonaute 2-associated small RNAs immunoprecipitated from the MCF7 breast cancer cell line. Interestingly, 10% of our new miRNAs are located in regions with common high-level genomic amplifications in breast cancer. A small number of new miRNAs have been experimentally validated by qPCR in breast tumours and a panel of human tissues, and further characterisation is underway. In summary, our next-generation sequencing strategy reached depths that allowed identification of an astonishing number of new functional small RNAs.

Functional approach to a genome-wide prognostic study in bladder cancer

Antoni Picornell[1], Stephen J Chanock[2], Alfonso Valencia[1], Nathaniel Rothman[2], Montserrat García-Closas[2], Debra Silverman[2], Jesús Herranz[1], Kevin Jacobs[2], María L. Calle[4], Manolis Kogevinas[3], Francisco X. Real[1,5] and Núria Malats[1]

[1]CNIO, Madrid, Spain; [2]National Cancer Institute, Bethesda, USA, [3]CREAL/IMIM-Hospital del Mar, Barcelona, Spain; [4]Universitat de Vic, Spain; [5]Universitat Pompeu Fabra, Barcelona, Spain

Due to its chronic nature, patients diagnosed of bladder cancer require of periodical follow-ups affecting their quality of life. As a consequence bladder cancer represents the most expensive cancer in terms of medical costs per patient. No genetic marker has been proved of help in predicting bladder cancer evolution till present. The aim of the study is to identify groups of single nucleotide polymorphisms (SNPs) involved biological functions that predict bladder cancer evolution.

In this study we considered a cohort of 1,300 patients from the Spanish Bladder Cancer/EPICURO study. This was carried out between 1998 and 2001 in 18 hospitals from 5 Spanish regions. A 10-years follow-up provides detailed data on the outcomes considered (recurrence, progression and death due to cancer). Patients' DNA was genotyped by the Illumina Infinium platform of 1M of SNPs. A total of 1,071 patients with genetic, clinical and follow-up data are available.

The independent predictive value for each SNP was analyzed for each outcome by applying adjusted survival analysis based on Cox proportional hazard regression models. Analysis considered two bladder cancer subgroups, three outcomes of interest per each subgroup, and four modes-of-inheritance (MOI). Tables of SNPs for each outcome with the minimal p-value of the four MOI were obtained. Those tables were evaluated though a comparative test using three gene-set analysis methods: ALIGATOR, GSA-SNP and iGSEA4GWAS. These methods have an automatic mapping process to associate the SNP to a gene, when possible. Once assessed the association, a gene-based test statistic is performed and the calculation of pathway enrichment analysis is done. Results from this analysis will be presented.

The pathway-based approaches complement the common single-marker analyses, giving a wider perspective of the biological information underlying the predictive values obtained in the independent SNP analysis approach.

Gene losses in evolution and cancer

Daniel Rico, David de Juan and Alfonso Valencia
CNIO, Madrid, Spain

Genes during the evolution of species are duplicated and lost. We and others have reasoned that it should be easier to lose a gene, if the gene lost have a paralog. The hypothesis behind this idea is that functional redundancy may exist between paralogs (specially after recent duplications) and the function of a gene lost may be carried out by the other paralog.

We were interested in detecting gene/protein families that have suffered deletion events during mammalian evolution, and test the hypothesis that singleton genes (i.e., genes without detectable paralogs) are less frequently lost than genes with paralogs. To explore this, we inferred gene losses from Ensembl Compara phylogenetic trees. We took into account the presence of paralogs (or not) in the time that a gene loss occurred by estimating the number of paralogs that the reconstructed ancestral genome had at that moment (i.e., tree nodes at a given level). We found that singleton genes are lost at significantly lower frequency than genes that had paralogs. To test this hypothesis in a cellular system where we can actually detect gene losses (not just infer them from phylogenetic trees) we used cancer cell lines as an evolutionary model. We analyzed the genes that can be homozygously deleted (HD) in the cancer cell lines, and compared the frequencies of HD events between human singleton genes and genes with paralogs. Genes that have paralogs in human are lost most frequently than singletons in cancer cell lines. Our results support the hypothesis of functional redundancy between gene paralogs, a phenomenon that could have important implications in the carcinogenesis process.

Integrated genomic and transcriptomic analysis of basal-like breast cancer

Bruno Tesson[1,2,3,4], Guillem Rigaill[1,2,3,4,5], Anne Vincent-Salomon[1,6], David Gentien[1,7], Gordon Tucker[8], Stéphane Robin[5], Emmanuel Barillot[1,2,3] and Thierry Dubois[1,4]
[1]Institut Curie, Paris, France; [2]INSERM, U900, Paris, France; [3]Mines ParisTech, Fontainebleau, France; [4]Translational Research Department, Breast Cancer Biology, Paris, France; [5]AgroParisTech (département MMIP) / INRA (département MIA), Paris, France; [6]INSERM, U830, Paris, France; [7]Translational Research Department, Genomic Platform, Paris, France; [8]Institut de Recherche Servier, Croissy sur Seine, France

Basal-like breast carcinomas are a subtype of breast cancer associated with a poor prognosis. There is no targeted therapy for this disease; basal-like breast cancers are HER2 and hormone receptor negative and so cannot be treated with trastuzumab antibody (anti-HER2) or anti-hormonal therapies like for the other subtypes of breast cancer. Up to 15% of breast cancer patients have indicators of a basal-like pathology. Finding therapeutic targets for the treatment of basal-like tumors is thus a priority in cancer research.

In this project, biopsies for the different subclasses of breast cancer (46 basal-like, 33 HER2+/ER-, 35 luminal A and 40 luminal B) were obtained from the Biological Center of Curie Institute, and were characterized by immunohistochemistry by a pathologist. Molecular profiles using DNA (Affymetrix, SNP6.0) and RNA (Affymetrix, U133plus2) microarrays were obtained for these 154 breast cancer tissues, and for 11 healthy breast tissues. We identify thousands of genes that are differentially expressed in each cancer subclass compared to healthy tissues, and between different cancer subclasses. Segmentation of genomic profiles reveals recurrent copy number alterations for each breast cancer subclass and we observe different 'simplex', 'sawtooth' and 'firestorm' patterns in line with earlier reports. Finally, an integrated analysis of the genomic and transcriptomic data identifies genes that are recurrently altered and differentially expressed in basal like breast tumors versus healthy tissues. This integrative strategy will, in a next step, be extended to include data from miRNA microarrays and from Reverse Phase Protein Arrays and should enable the identification of well-known and putative cancer driving genes that are potential therapeutic targets.

Integrated genomics analysis of gene and miRNA expression profiles in clear cell renal carcinoma cell lines

Valentina Tinaglia[1*], Ingrid Cifola[2*], Fabio Frascati[2], Eleonora Mangano[2], Marta Biasolo[3], Maria Carla Proverbio[1], Stefania Bortoluzzi[3], Vitalba Di Stefano[4], Cristina Bianchi[4], Roberto Perego[4] and Cristina Battaglia[1]

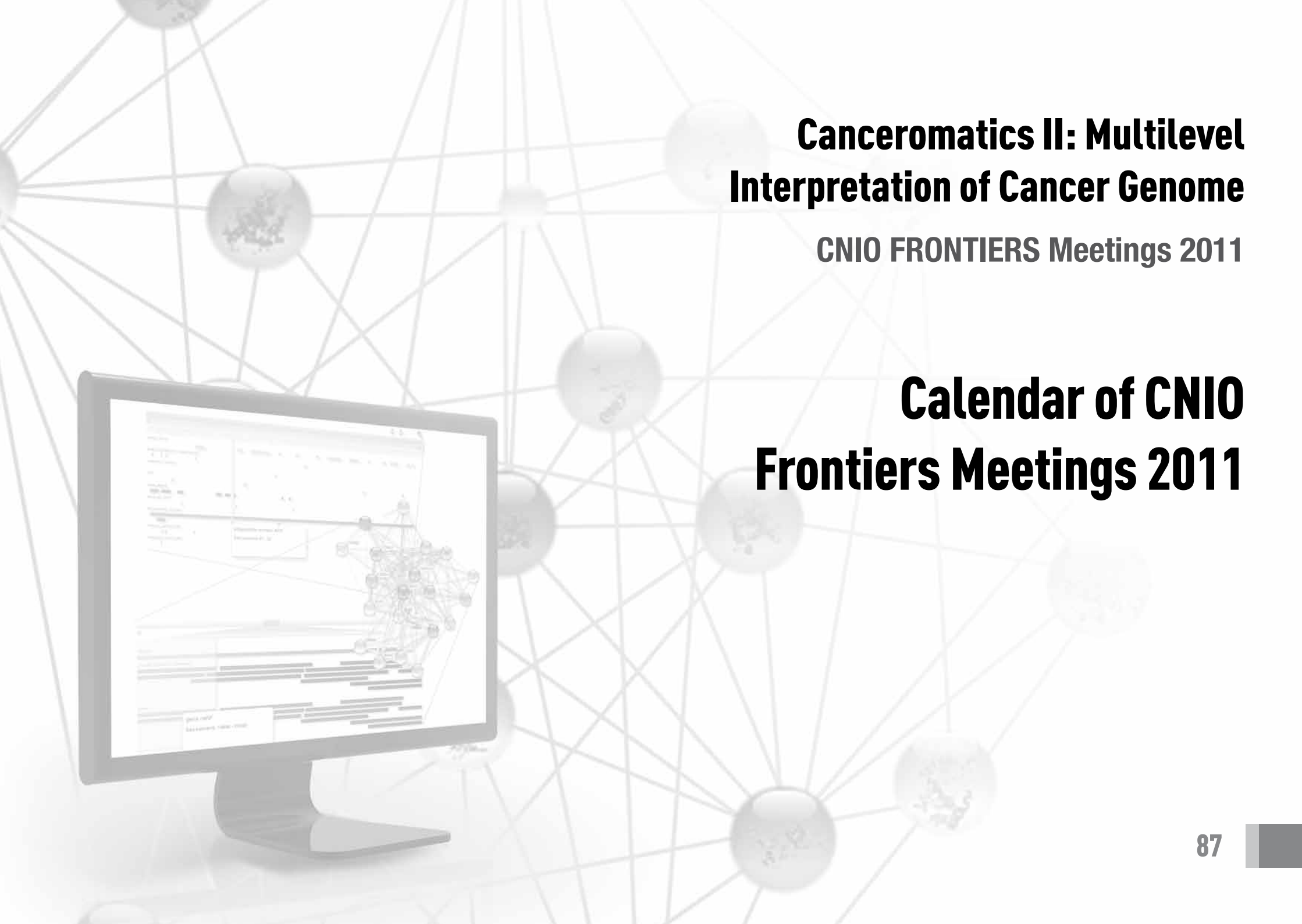
[1]Università degli Studi di Milano, Italy; [2]Institute for biomedical technologies, National Research Council, Milan, Italy; [3]University of Padua, Padova, Italy; [4]University of Milano-Bicocca, Milan, Italy; [*]equally contribute

Clear cell renal cell carcinoma (ccRCC) is the most common and malignant tumor in the adult kidney, representing 75-80% of renal primary malignancies. Inactivation of the von Hippel-Lindau (VHL) tumor suppressor gene, by mutation, deletion and promoter methylation, occurs in most sporadic ccRCCs and in all inherited cases. Recent evidences showed that microRNAs (miRNAs) are often dysregulated in many tumors, including ccRCC.

We used Caki-1, Caki-2 and A498 cell lines as in vitro model of ccRCC pathology, and HK-2 (normal proximal tubular epithelial cell line) as reference sample. We characterized the VHL status by direct sequencing and the HIF status by western blot. Affymetrix microarray technology was applied to assess miRNA (onto GeneChip® miRNA Array) and gene expression profiles (onto GeneChip® Human Gene 1.0 ST Array).

Analysis of common differentially expressed miRNAs (DEMs) in RCC cell lines outlined specific miRNAs to be involved in ccRCC and in cancer (e.g. miR-145, miR- 29a, miR-342-3p, miR-205, miR-183, miR-197, miR-132, miR-146a and miR-34a). Functional enrichment analysis of common differentially expressed genes (DEGs) highlighted some genes involved in leucocyte transendothelial migration, focal adhesion and p53 signalling pathways (e.g. ATM, FER, CDC27 and GRB10). Additionally, we conducted an integrated analysis to combine gene and miRNA expression profiles and to reconstruct miRNA-gene post-transcriptional regulatory networks involved in RCC pathology. We also compared our expression data with RCC datasets publicly available from NCBI GEO microarray repository. We selected potentially interesting miRNAs and target genes for further validation (by qPCR). This integrated analysis approach may help to unravel the molecular complexity characterizing ccRCC biology, and it will facilitate the elucidation of regulatory circuits important for tumorigenesis and the biological processes under relevant posttranscriptional regulation in ccRCC.

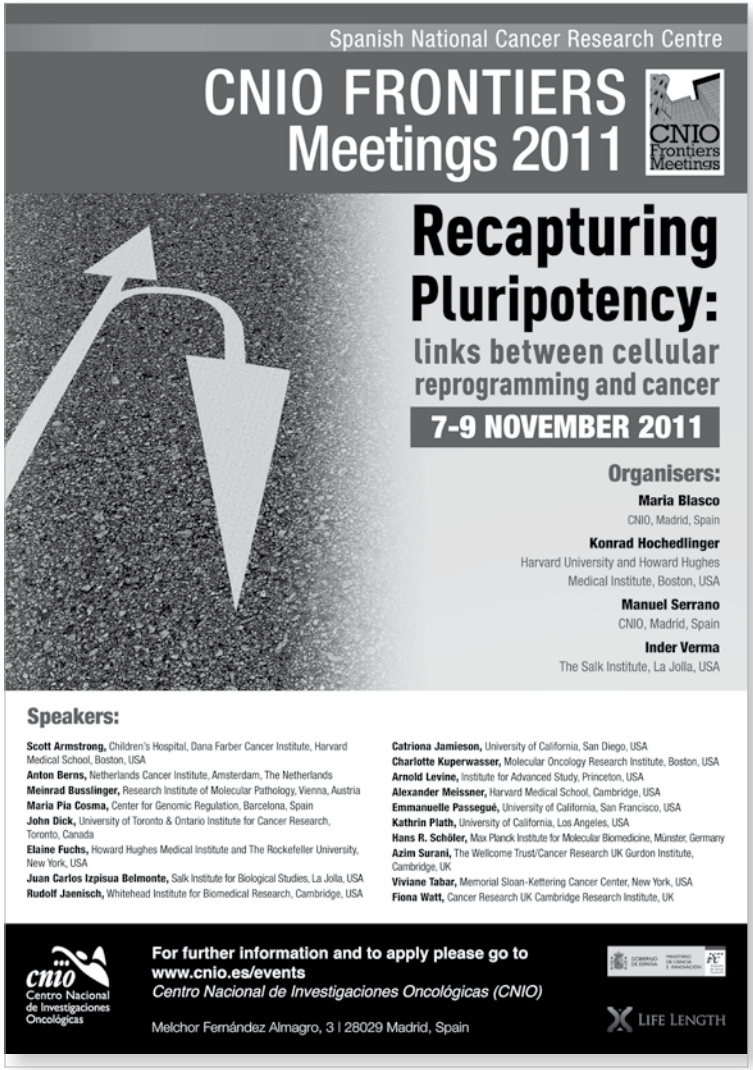
notes



Canceromatics II: Multilevel Interpretation of Cancer Genome

CNIO FRONTIERS Meetings 2011

Calendar of CNIO Frontiers Meetings 2011



CNIO FRONTIERS
Meetings 2011
CNIO Frontiers Meetings 2011

Canceromatics II: Multilevel Interpretation of Cancer Genome

CNIO FRONTIERS Meetings 2011

Previous CNIO Frontiers Meetings and CNIO Cancer Conferences

2011

Breast Cancer
07/02/2011 - 09/02/2011
Organisers: **Joaquín Arribas, José Baselga, Miguel Ángel Piris, Lajos Pusztai and Jorge Reis-Filho**

2010

Cancer pharmacogenetics: Personalizing medicine
22/11/2010 - 24/11/2010
Organisers: **Javier Benítez, William E. Evans, Miguel Martín and Magnus Ingelman-Sundberg**

Nature-CNIO Cancer Symposium on "Frontiers in Tumour Progression"
24/10/2010 - 27/10/2010
Organisers: **Douglas Hanahan, Zena Werb and Erwin Wagner**

4th ESO-CNIO Familial Cancer Conference
07/06/2010 - 08/06/2010
Organisers: **Javier Benítez, Rosalind Eeles and Hans Vasen**

Molecular cancer therapeutics
08/03/2010 - 10/03/2010
Organisers: **Gail Eckhardt, Roy S. Herbst and Manuel Hidalgo**

2009

The energy of cancer
02/11/2009 - 04/11/2009
Organisers: **Toren Finkel, David M. Sabatini, Manuel Serrano and David A. Sinclair**

Cancer-om-atics: Multilevelinterpretation of cancer genome data
06/07/2009 - 08/07/2009
Organisers: **Søren Brunak, Núria Malats, Chris Sander and Alfonso Valencia**

Stem cells and cancer
23/02/2009 - 25/02/2009
Organisers: **Elaine Fuchs, Maria A. Blasco, Eduard Batlle and Mirna Pérez-Moreno**

2008

Signalling upstream of mTOR
03/11/2008 - 05/11/2008
Organisers: **Dario R. Alessi, Tomi P. Mäkelä and Montserrat Sánchez-Céspedes**

Structure and mechanisms of essential complexes for cell survival
23/06/2008 - 25/06/2008
Organisers: **Niko Grigorieff, Eva Nogales and Jose María Valpuesta**

Development and cancer
04/02/2008 - 06/02/2008
Organisers: **Konrad Basler, Ginés Morata, Eduardo Moreno and Miguel Torres**

2007

Links between cancer, replication stress and genomic integrity
05/11/2007 - 07/11/2007
Organisers: **Oskar Fernández-Capetillo, Jiri Lukas, Juan Méndez and André Nussenzweig**

Myc and the transcriptional control of proliferation and oncogenesis
11/06/2007 - 13/06/2007
Organisers: **Robert N. Eisenman, Martin Eilers and Javier León**

Molecular mechanisms in lymphoid neoplasm
19/02/2007 - 21/02/2007
Organisers: **Elias Campo, Riccardo Dalla-Favera, Elaine S. Jaffe and Miguel Angel Piris**

2006

Telomeres and telomerase - CNIO / Joséf Steiner Cancer Conference
13/11/2006 - 15/11/2006
Organisers: **Maria A. Blasco and Jerry Shay**

Medicinal chemistry in oncology
02/10/2006 - 04/10/2006
Organisers: **Fernando Albericio, James R. Bischoff, Carlos García-Echeverria and Andrew Mortlock**

Inflammation and cancer
22/05/2006 - 24/05/2006
Organisers: **Curtis Harris, Raymond DuBois, Jorge Moscat and Manuel Serrano**

PTEN and the AKT route
08/05/2006 - 10/05/2006
Organisers: **Ana Carrera, Pier Paolo Pandolfi and Peter Vogt**

CNIO FRONTIERS
Meetings 2011
Previous CNIO Frontiers Meetings and CNIO Cancer Conferences

2005

Cancer and aging
07/11/2005 - 09/11/2005
Organisers: **Maria A. Blasco, Kathy Collins, Jan Hoeijmakers and Manuel Serrano**

MAP kinases and cancer
30/05/2005 - 01/06/2005
Organisers: **Philip Cohen, Roger Davis, Worcester, Chris Marshall and Ángel Nebreda**

Animal tumour models and functional genomics
07/03/2005 - 09/03/2005
Organisers: **Allan Balmain, Mariano Barbacid, Anton Berns and Tyler Jacks**

2004

Cadherins, catenins and cancer
29/11/2004 - 01/12/2004
Organisers: **Amparo Cano, Hans Clevers, José Palacios and Franz Van Roy**

Structural biology of cancer targets
27/09/2004 - 29/09/2004
Organisers: **Ernest Laue, Guillermo Montoya and Alfred Wittinghofer**

2003

Apoptosis and cancer
01/12/2003 - 03/12/2003
Organisers: **Gabriel Nuñez, Marisol Soengas and Scott Lowe**

Small GTPases in human carcinogenesis
16/06/2003 - 18/06/2003
Organisers: **Juan Carlos Lacal, Channing Der and Shuh Narumiya**

Targeted search for anticancer drugs
17/03/2003 - 19/03/2003
Organisers: **Amancio Carnero and David H. Beach**

2002

Mechanisms of invasion and metastasis
18/11/2002 - 20/11/2002
Organisers: **Joan Massagué and Richard Hynes**

The cell cycle and cancer
30/09/2002 - 02/10/2002
Organisers: **Marcos Malumbres, Charles Sherr and Jiri Bartek**

Cancer epigenetics: DNA methylation and chromatin
29/05/2002 - 31/05/2002
Organisers: **Manel Esteller and Stephen B. Baylin**

CNIO FRONTIERS
Meetings 2011
Previous CNIO Frontiers Meetings and CNIO Cancer Conferences

notes

As a non-profit organisation, we would like to thank all those who sponsored this conference. Such contribution help us to ensure that our conferences will continue to establish the CNIO as a point of reference for the international cancer research community

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN
Instituto de Salud Carlos III

Supported by the ESF Programme "Frontiers of Functional Genomics"

illumina®

EUROPEAN SCIENCE FOUNDATION

Centro Nacional de Investigaciones Oncológicas (CNIO)
Melchor Fernández Almagro, 3
28029 Madrid, Spain
www.cnio.es

Coordination and edition: Mar Pérez, Carolina Garcia Sabaté and Virginia de la Cruz, CNIO
Direction of art and production: Versus Gráficas S.A.
Photographic archive: CNIO

Printed in Spain