CRG
Centre for Genomic Regulation

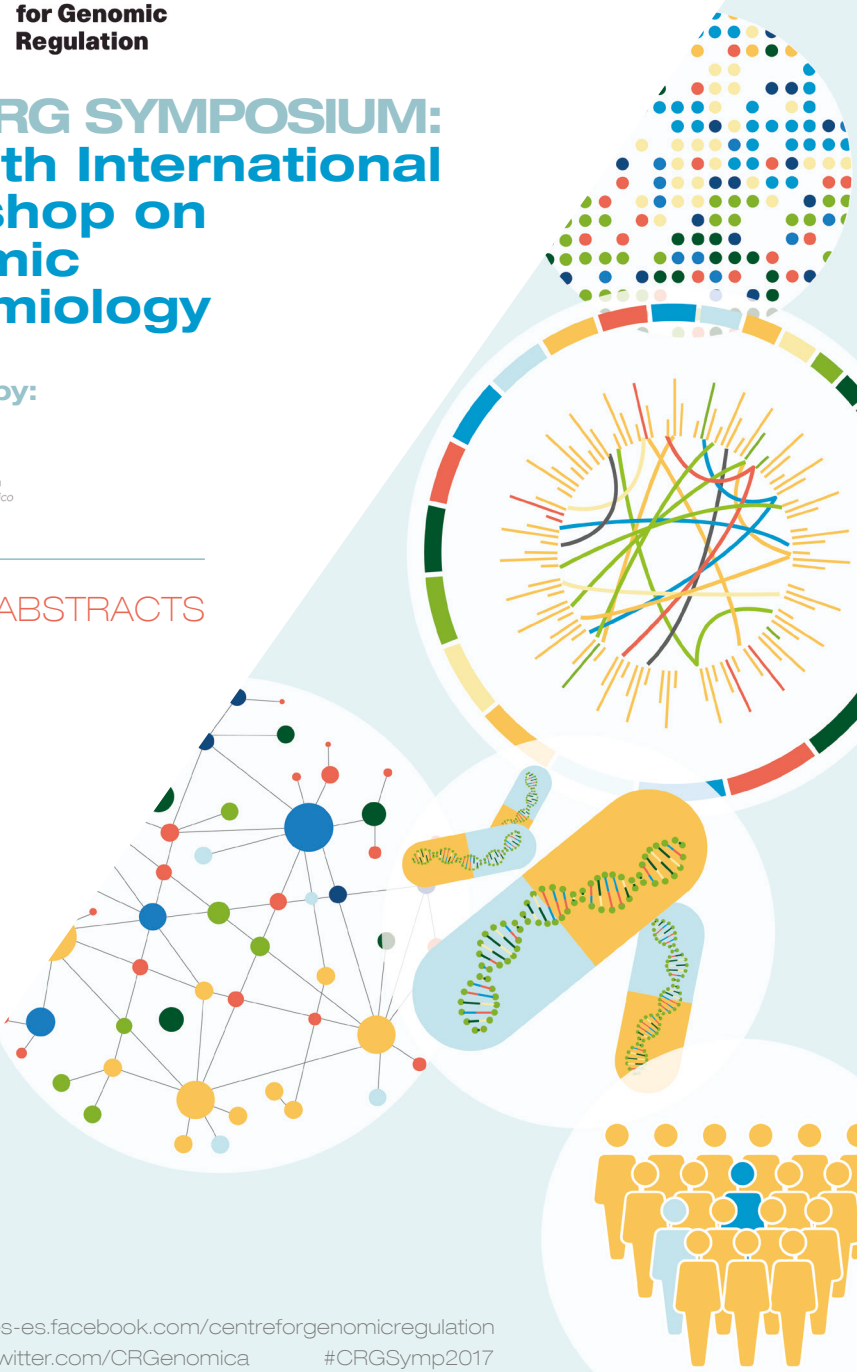
16th CRG SYMPOSIUM: Seventh International Workshop on Genomic Epidemiology

Organized by:

cnag

centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

BOOK OF ABSTRACTS



https://es-es.facebook.com/centreforgenomicregulation
https://twitter.com/CRGenomica #CRGSymp2017

Produced by: Centro Nacional de Análisis Genómico (CNAG)
and Centre for Genomic Regulation (CRG)

Texts and graphics: Invited speakers,
Speakers of short talks
CRG Department of Communication & Public Relations

Graphic design: Iris Joval (http://www.scienceillustration.net/)

16th CRG SYMPOSIUM: Seventh International Workshop on Genomic Epidemiology

PRBB Auditorium, Barcelona
20, 21 & 22 September 2017

Organized by:

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

Organizers:

William Cookson, Imperial College London, UK
Ivo Gut, CNAG-CRG, Barcelona, ES
Mark Lathrop, McGill University and Genome Quebec Innovation Centre, CA
Daniel E. **Weeks**, Department of Human Genetics, University of Pittsburgh, USA

CONTENTS

SYMPOSIUM PROGRAMME

Wednesday, Sept 20 th

14:00 - 14:15 Welcome

TECHNOLOGY

SESSION 1

14:15 - 14:45 **Ivo Gut**, CNAG-CRG, ES
Recent advances in nucleic acid sequencing technologies

14:45 - 15:15 **Holger Heyn**, CNAG-CRG, ES
Single-Cell RNA Sequencing of Complex Tissues

15:15 - 15:30 Selected Abstract talk, **Jessica Nordlund**
SPlinted Ligation Adapter Tagging (SPLAT), a novel library preparation method for whole genome bisulfite sequencing

15:30 - 15:45 Selected Abstract talk, **Nina Görner**
Metabolomics: The Missing Link in Precision Medicine

15:45 - 16:15 Coffee break

16:15 - 16:45 **Marc Marti-Renom**, CNAG-CRG, ES
Rational design of non-resistant targeted cancer therapies

16:45 - 17:00 Selected Abstract talk, **Masato Akiyama**
Genetic study of body mass index in Japanese population links cell-types to body weight regulation

17:00 - 17:30 **Manolis Kogevinas**, ISGlobal, ES
The exposome complements the genome: a new era in epidemiology

17:00 - 17:30 Welcome Reception + Poster Session (Terrasse)

Thursday, Sept 21 st

EPIGENETICS

SESSION 2

09:00 - 09:30 **Peter Jones**, Van Andel Institute, USA
Targeting Human Endogenous Retroviruses for Epigenetic Therapy

09:30 - 10:00 **Caroline Relton**, University of Bristol, UK
Causal inference in epigenetic epidemiology

10:00 - 10:30 **Iñaki Martin-Subero**, IDIBAPS, ES
Epigenetic insights into chronic lymphocytic leukemia biology and clinical behavior: from DNA methylomes to reference epigenomes

RESOURCES / BIG INITIATIVES

SESSION 3

10:30 - 11:00 **Hanns Lochmüller**, Newcastle University, UK
RD-Connect – linking, analysing and sharing multi-source omics data for Rare Disease Research

11:00 - 11:30 Coffee break

11:30 - 12:00 **Sergi Beltran**, CNAG-CRG, ES
The BBMRI-LPC call to sequence 900 Rare Disease exomes: a successful transnational collaborative initiative with EuroBioBank and RD-Connect

12:00 - 12:15 Selected Abstract talk, **Maria Alvarellos**
PharmGKB: The Pharmacogenomics Knowledgebase

12:15 - 12:45 **Roderic Guigó**, CRG, ES
The human transcriptome across tissues and individuals

12:45 - 14:00 Lunch

DISEASES - CANCER

SESSION 4

14:00 - 14:30 **Nuria Malats**, CNIO, ES
Towards the integration of Omics data in epidemiological studies: A Pancreatic Cancer Journey

14:30 - 14:45 Selected Abstract talk, **Evangelina López de Maturana**
Genome-scale data integration for deciphering pancreatic cancer aetiology

14:45 - 15:00 Selected Abstract talk, **Juan González**
Transcriptomic and epigenomic mechanism of mosaic loss of chromosome Y (LOY) in cancer

15:00 - 15:30 **Richard Houlston**, Institute Cancer Research, UK
Polygenic susceptibility to colorectal cancer: mechanisms and impact

15:30 - 15:45 Selected Abstract talk, **Izaskun Mallona**
DNA co-methylation networks for colon cancer patient stratification

15:45 - 16:00 **Sven Bocklandt**, Sr. Application Specialist, Bionano Genomics

Beyond NGS: Genome Mapping Reveals Structural Variation in Cancer and Genetic Disease

16:00 - 16:30 Coffee break

16:30 - 16:45 Selected Abstract talk, **Miranda Stobbe**
Recurrent somatic mutations reveal new insights into mutational processes in cancer

16:45 - 17:00 Selected Abstract talk, **Solip Park**
Systematic discovery of germline cancer predisposition genes through the identification of somatic second hits

17:00 - 17:30 **Ben Lehner**, CRG, ES
Mutations and their interactions in individuals

Friday, Sept 22 nd

09:00 - 09:30 **Nuria López-Bigas**, IRB, ES
Coding and non-coding cancer mutations

DISEASES - COMMON DISEASES

SESSION 4

09:30 - 10:00 **Florence Demenais**, Inserm-Université Paris Diderot, FR
New insights into the genetics of asthma

10:00 - 10:30 **Momoko Horikoshi**, RIKEN, JP
Genomic loci associated with birth weight identify genetic links between intrauterine growth and adult metabolic disease

10:30 - 11:00 **Patricia Munroe**, Queen Mary University of London, UK
New insights into blood pressure regulation from large-scale genetic studies

11:00 - 11:30 Coffee break

11:30 - 12:00 **George Thanassoulis**, McGill, CA
A Genomewide association study of aortic stenosis

12:00 - 12:15 Selected Abstract talk, **Jon Sánchez**
A Transcriptomic Investigation on Patient–Specific Comorbidities

12:15 - 12:30 Selected Abstract talk, **Oscar Lao**
Identification of polygenic adaptation in attention deficit hyperactivity disorder using GWAS data

12:30 - 13:00 **Stephanie Debette**, University Bordeaux, FR
Genomics of complex cerebrovascular disease

13:00 - 14:00 Lunch

ANALYTICAL METHODS AND RESOURCES

SESSION 5

14:00 - 14:30 **George Davey-Smith**, University of Bristol, UK
Mendelian randomization: what does the future hold?

14:30 - 15:00 **Dorret Boomsma**, Vrije Universiteit Amsterdam, The Netherlands
Contributions from twin studies to gene discovery: Multi-variate, Multi-rater, Multi-age GWAS of longitudinal aggression and attention problems

15:00 - 15:30 **Gonçalo Abecasis**, University of Michigan, USA
Sequencing and Analysis of 10,000s of Human Genomes: Challenges and opportunities

15:30 - 16:00 **Mark Caulfield**, Genomics England, UK
The 100,000 Genomes Project transforming healthcare

16:00 - 16:30 Closing Remarks

SYMPOSIUM FUNDING AGENCIES AND SPONSORS

The symposium organisers acknowledge the generous support from the following organisations:

Sponsored by:

illumina®

Atos

bionano GENOMICS

Roche

Agilent Technologies

Supported by:

EXCELENCIA SEVERO OCHOA

Member of the Board of Trustees:

Generalitat de Catalunya

upf. Universitat Pompeu Fabra Barcelona

GOBIERNO DE ESPAÑA MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

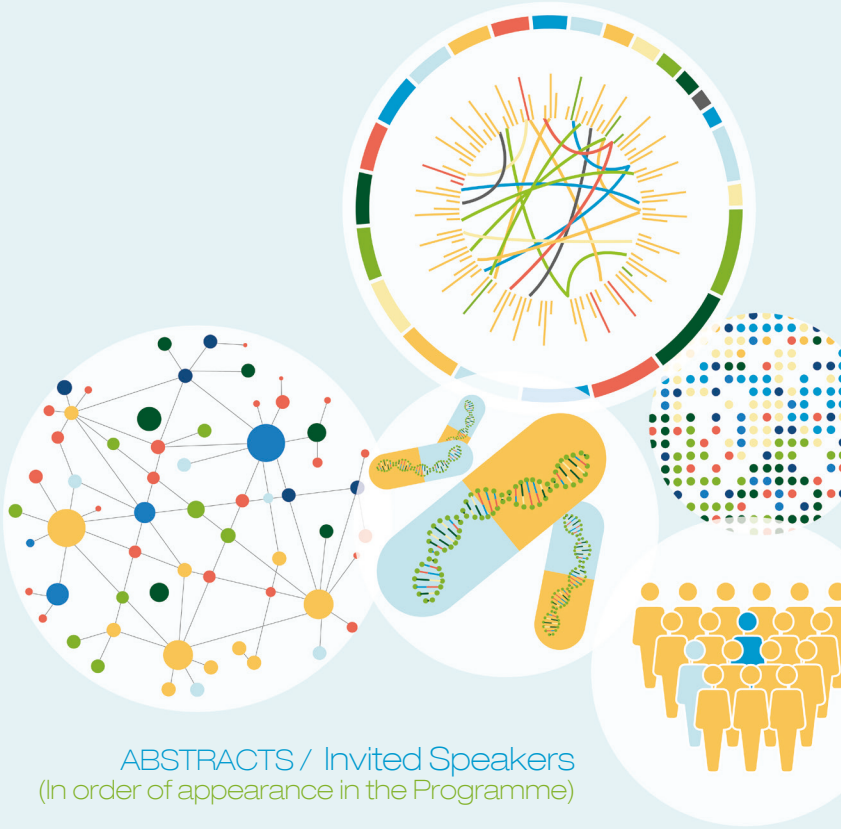
Obra Social "la Caixa"

Member of:

BIST Barcelona Institute of Science and Technology

Promotion Partner:

BIB BIOINFORMATICS BARCELONA



ABSTRACTS / Invited Speakers
(In order of appearance in the Programme)

*These abstracts should not be cited in bibliographies. Material contained herein should be treated as personal communication and should be cited as such only with consent of the author.

Advances in DNA sequencing technology

Ivo GUT

Centro Nacional de Analisis Genomico,

Center for Genomic Regulation, Barcelona, ES

The CNAG-CRG is a genome analysis center involved in many different types of genome research project, -fundamental and applied. In order to cope with the expectations of our collaborators we have established many analytical pipelines. A pipeline covers the planning stage of a project, understanding of input materials, sample preparation, sequencing, quality control procedures and data analysis. In order to be able to respond to a wide palette of needs we have established two sequencing systems, from Illumina and Oxford Nanopore Technologies in house and use additional techniques externally. With this we have the ability to do short-read sequencing with high base quality and long-read sequencing. The applications that we support are whole genome, whole exome, targeted sequencing, transcriptome analysis, epigenetic analysis and single cell sequencing. Projects touch human genetics, human diseases, such as cancer and rare diseases, epigenetics, model organisms and species of economic interest in the agrifood sector. CNAG has also developed computational solutions that facilitate data interpretation in a secure environment. Our collaborators can work on their data and results in our databases. These solutions are widely used in RD-Connect, an EU-funded project of the International Rare Disease Research Consortium, and in several Catalan and Spanish initiatives in Personalized Medicine.

NOTES

Single-Cell RNA Sequencing of Complex Tissues

Holger HEYN

Centro Nacional de Análisis Genómico,

Center for Genomic Regulation, Barcelona, ES

Single-cell RNA sequencing (scRNAseq) is at the forefront of techniques to chart molecular properties of individual cells. Recent methods are scalable to thousands of cells, enabling an unbiased sampling and in-depth characterization without prior knowledge. Consequently, studies aim to produce comprehensive cellular atlases of tissues, organs and organisms.

We implemented high-throughput scRNAseq processes for different protocols (MARS-seq, Smart-seq2, Seq-well) with >50,000 single cells sequenced and analyzed to date. We are joining computational, statistical and biological knowledge in order to determine best practices in single-cell research and to relate genome activity to cellular phenotypes. A systematic comparison of different scRNAseq protocols pointed to large differences in sensitivity of molecule capture, with a high degree of accuracy across the methods. We critically enlarged the scope of such methods by establishing cryopreservation as suitable method for sample transfer and clinical archiving.

We applied single-cell transcriptome analysis for cellular phenotyping to resolved heterogeneity during developmental processes, complex tissues or tumor evolution. To address the challenges of future large scRNAseq data sets, we developed an analytical framework for the sensitive detection of population markers and differentially expressed genes, being scalable to analyze millions of single cells. Analyzing 1.3 million cells from the mouse developing forebrain, we identified rare populations of neurons, for which we determined a previously not recognized heterogeneity associated to distinct differentiation stages, spatial organization and cellular function.

NOTES

Rational design of non-resistant targeted cancer therapies

Marc MARTI-RENOM

Centro Nacional de Analisis Genomico,

Center for Genomic Regulation, Barcelona, ES

Drug resistance is one of the major problems in targeted cancer therapy. A major cause of resistance is changes in the amino acids that form the drug-target binding site. Despite of the numerous eforts made to individually understand and overcome these mutations, there is a lack of comprehensive analysis of the mutational landscape that can prospectively estimate drug-resistance mutations. Here we describe and computationally validate a framework that combines the cancer-specifc likelihood with the resistance impact to enable the detection of single point mutations with the highest chance to be responsible of resistance to a particular targeted cancer therapy. Moreover, for these treatment- threatening mutations, the model proposes alternative therapies overcoming the resistance. We will exemplify the applicability of the model using EGFR-geftinib treatment for Lung Adenocarcinoma (LUAD) and Lung Squamous Cell Cancer (LSCC) and the ERK2-VTX11e treatment for melanoma and colorectal cancer. Our model correctly identiy the phenotype known resistance mutations, including the classic EGFR-T790M and the ERK2-P58L/S/T mutations. Moreover, the model predicts new previously undescribed mutations as potentially responsible of drug resistance. Finally, we will show a map of the predicted sensitivity of alternative ERK2 and EGFR inhibitors, with a particular highlight of two molecules with a low predicted resistance impact.

NOTES

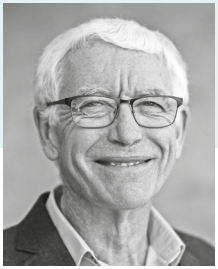
The exposome complements the genome: a new era in epidemiology

Manolis KOGEVINAS

ISGlobal, Barcelona Institute for Global Health, Barcelona, ES

Information on both environmental and genetic causes of disease is growing as a result of large-scale epidemiological research but exposure data including diet, lifestyle, environmental and occupational factors is often fragmentary, non-standardized, at crude resolution and often does not include estimates at the level of the individual. These limitations have recently been framed within the context of the "exposome", the environmental counterpart of the genome. This new approach comprehensively addresses the integration of external exposome (measured through advanced new methods in environmental exposure assessment) and the internal exposome (measured through biomarkers and omics technologies) at the individual level. This new methodology to population studies provides a holistic and consolidated approach to exposure science, disease causation and risk assessment at both individual and population level. I will present an overall view of the Exposome concept and present findings from recent large studies we conducted following this approach evaluating environmental exposures.

NOTES



Targeting Human Endogenous Retroviruses for Epigenetic Therapy

Peter JONES

Van Andel Research Institute, Grand Rapids, Michigan, USA

Vitamin C deficiency is found in patients with cancer and might complicate various therapy paradigms. Here we show how this deficiency may influence the use of DNA methyltransferase inhibitors (DNMTis) for treatment of hematological neoplasias. In vitro, when vitamin C is added at physiological levels to low doses of the DNMTi 5-aza-2'-deoxycytidine (5-aza-CdR), there is a synergistic inhibition of cancer-cell proliferation and increased apoptosis. These effects are associated with enhanced immune signals including increased expression of bidirectionally transcribed endogenous retrovirus (ERV) transcripts, increased cytosolic dsRNA, and activation of an IFN-inducing cellular response. This synergistic effect is likely the result of both passive DNA demethylation by DNMTi and active conversion of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) by ten-eleven translocation (TET) enzymes at LTR regions of ERVs, because vitamin C acts as a cofactor for TET proteins. In addition, TET2 knockout reduces the synergy between the two compounds. Furthermore, we show that many patients with hematological neoplasia are markedly vitamin C deficient. Thus, our data suggest that correction of vitamin C deficiency in patients with hematological and other cancers may improve responses to epigenetic therapy with DNMTis.

NOTES

Causal inference in epigenetic epidemiology

Caroline RELTON
University of Bristol, UK

Epigenetic variation is thought to play a role in mediating both the development of and adverse consequences of many diseases. Epigenetic mechanisms may be important determinants of health (where causality is important) or as useful biomarkers to help predict the occurrence and/or the consequences of disease (where causality is not important).

Epidemiological approaches can be applied to explore causal pathways and to identify and validate epigenetic biomarkers. The use of Mendelian randomization to strengthen causal inference will be presented. In addition, examples of the use of DNA methylation as a biomarker of previous, long term exposure and as a predictive biomarker of future health outcomes will be presented.

.

NOTES

Epigenetic insights into chronic lymphocytic leukemia biology and clinical behavior: from DNA methylomes to reference epigenomes

Iñaki MARTÍN-SUBERO

Biomedical Epigenomics Group, Institut d'investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, ES

Chronic lymphocytic leukemia (CLL) is the most frequent leukemia in the Western world. Over the last years, we have been studying the epigenome of CLL and linked our findings to the biological features and clinical behavior of the disease. From the DNA methylation perspective, CLLs can be classified into 3 categories with different clinico-biological characteristics based on epigenetic imprints of different normal B-cell maturation stages, which can be translated into the clinics using a small set of epigentic biomarkers. Beyond this imprint of the cellular origin, CLLs aquire extensive DNA methylation changes, which frequently affect enhnancer elements related to CLL pathogenesis as well as apparently non-functional elements such as heterochromatic and repressed regions. In spite of these novel insights into CLL pathogenesis, DNA methylation is only one layer of the epigenetic portfolio. In the context of the BLUEPRINT Project, we have been recently generating the reference epigenomes of CLL and normal B cell differentiation which include, in addition to the DNA methylome, ChIP-seq for 6 histone marks with non-overlapping functions, ATAC-seq for chromatin accessibility and RNA-seq. The integrative analysis of these epigenetic layers has revealed that CLL, as compared to normal B cells, show both de novo activation of regulatory elements of key CLL oncogenes and epigenomic patterns that are highly dynamic in different B cell subpopulations. During the presentation, both published and most recent unpublished data will be shown to provide a general overview of the epigenome of CLL.

NOTES

RD-Connect – linking, analysing and sharing multi-source omics data for Rare Disease Research

Hanns LOCHMÜLLER

John Walton Muscular Dystrophy Research Centre,
Institute of Genetic Medicine, Newcastle University, UK

6-8% of the European population – between 27 and 36 million people – are affected by one of the 5000-8000 distinct rare diseases. RD-Connect (www.rd-connect.eu) is a unique global infrastructure project that links up databases, registries, biobanks and clinical bioinformatics data used in rare disease research into a central resource for researchers worldwide. It has developed an integrated research platform in which complete clinical profiles are combined with -omics data and sample availability for rare disease research, in particular research funded under the International Rare Diseases Research Consortium (IRDiRC; www.irdirc.org).

RD-Connect's mechanism for sharing and analysis of rare disease genomic data begins with submission of the raw .bam or .fastq files, which is essential in order to allow data from multiple sequencing providers to be processed through a standard pipeline to ensure comparability. The raw data are stored for long-term access at the European Genome-phenome Archive (EGA), a secure, controlled-access repository, while the processed data are made accessible online for real-time analysis in the RD-Connect genomics analysis interface. Thanks to the initial collaboration with the two partner projects, NeurOmics and EURenOmics, the platform is a rich resource containing whole exomes and genomes of a large number of individuals with rare neuromuscular and rare renal disease, but it is also growing rapidly for other RD areas such as mitochondrial, neurogenetic and immunological disorders. Several thousand more datasets are planned for submission in the coming year, now that the platform is open for submission of data from all rare disease projects.

The collaboration with BBMRI-LPC is seen as a paradigm of good practice for European rare disease data sharing: researchers across Europe were provided with exome sequencing at no cost through a transnational access mechanism, but the project conditions mandated that biosamples must be made accessible

through EuroBioBank biobanks and phenotypic data must be submitted to the RD-Connect PhenoTips instance, while the resulting sequencing data will be automatically submitted to the RD-Connect platform. This workflow not only allows the researchers themselves to analyse their own cases but also ensures that the samples and data will be accessible to others in future, thus maximising the added value of the project for future research

NOTES

The BBMRI-LPC call to sequence 900 Rare Disease exomes: a successful transnational collaborative initiative with EuroBioBank and RD-Connect

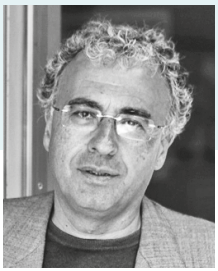
Sergi BELTRÁN

Centro Nacional de Análisis Genómico,

Center for Genomic Regulation, Barcelona, ES

The 2016 BBMRI-LPC WES Call offered a unique opportunity to genetically diagnose rare disease patients with biological samples from the EuroBioBank network. The program, launched in June 2016, offered free-of-charge Whole Exome Sequencing and analysis for a total of 900 samples from 10-30 coordinated projects, each with 2-3 principal investigators (PIs) from different countries. The PIs had to confirm they could provide phenotypic information using the Human Phenotype Ontology (HPO) and that the Informed Consent allowed data sharing in controlled access repositories and databases such as the EGA and RD-Connect. 17 projects were selected by the end of August and were asked to provide the samples by the end of September. Sequencing was conducted at the Centro Nacional de Análisis Genómico (CNAG-CRG, Spain) and at the Welcome Trust Sanger Institute (WTSI, UK) and was finished by early 2017. Data was processed through the RD-Connect validated analysis pipeline and was made available through its platform once all the commitments from the PIs were fulfilled. The RD-Connect platform allows the researchers to analyse and interpret their genotype:phenotype data privately for up to 6 months before it is shared with the rest of authorised users. It also enables anonymised data sharing through tools such as the GA4GH/IRDiRC MatchMaker Exhange and GA4GH Beacon. We will present the up-to-date diagnostic yield of the BBMRI-LPC WES while focusing on some success stories. We will also report on the challenges and lessons learned from conducting such a complex transnational collaborative multi-player initiative.

NOTES



The human transcriptome across tissues and individuals

Roderic GUIGÓ

Center for Genomic Regulations and Universitat Pompeu Fabra. Barcelona, ES

The pilot phase of the Genotype-Tissue Expression (GTEx) project has produced RNASeq from 1,641 samples originated from up to 43 tissues from 175 post-mortem donors, and constitutes a unique resource to investigate the human transcriptome across tissues and individuals. Clustering of samples based on gene expression recapitulates tissue types, separating solid from not solid tissues, while clustering based on splicing separates neural from non-neural tissues, suggesting that post-transcriptional regulation plays a comparatively important role in the definition of neural tissues .About 47 % of the variation in gene expression can be explained by variation of across tissues, while only 4% by variation across individuals. We find that the relative contribution of individual variation is similar for lncRNAs and for protein coding genes. However, we find that genes that vary with ethnicity are enriched in lncRNAs, whereas genes that vary with age are mostly protein coding. Among genes that vary with gender, we find novel candidates both to participate and to escape X-inactivation. In addition, by merging information on GWAS we are able to identify specific candidate genes that may explain differences in susceptibility to cardiovascular diseases between males and females and different ethnic groups. We find that genes that decrease with age are involved in neurodegenerative diseases such as Parkinson and Alzheimer and identify novel candidates that could be involved in these diseases. In contrast to gene expression, splicing varies similarly among tissues and individuals, and exhibits a larger proportion of residual unexplained variance. This may reflect that that stochastic, non-functional fluctuations of the relative abundances of splice isoforms may be more common than random fluctuations of gene expression. By comparing the variation of the abundance of individual isoforms across all GTEx samples, we find that a large fraction of this variation between tissues (84%) can be simply explained by variation in bulk gene expression, with splicing variation contributing comparatively little. This strongly suggests that regulation at the

primary transcription level is the main driver of tissue specificity. Although blood is the most transcriptionally distinct of the surveyed tissues, RNA levels monitored in blood may retain clinically relevant information that can be used to help assess medical or biological conditions.

NOTES

Towards the integration of Omics data in epidemiological studies: A Pancreatic Cancer Journey

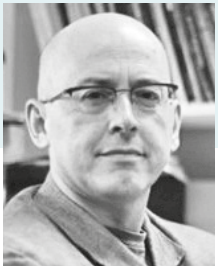
Nuria MALATS

Spanish National Cancer Research Centre, Madrid, ES

Disease prevention can highly benefit of a personalized medicine approach through the accurate discrimination of individuals at high risk of developing a specific disease from those at moderate and low risk. To this end precise risk prediction models need to be built. This endeavour requires a precise characterization of the individual exposome, genome, and phenome. Massive molecular omics data representing the different layers of the biological processes of the host and the non-host will enable to build more accurate risk prediction models. Epidemiologists aim to integrate omics data along with important information coming from other sources (questionnaires, candidate markers) that has been proved to be relevant in the risk assessment of complex diseases.

The vast proportion of pancreatic cancer is named sporadic because it does not aggregate within families and its aetiology is complex. Both genetic and non-genetic factors have been associated with sporadic pancreatic cancer though the magnitude of their risk is small/moderate. Therefore, cost-efficient primary and secondary prevention programs for sporadic pancreatic cancer should be based on multifactorial integrative scores to define high-risk populations. Steps towards the integration of genomics and non-genomics factors selected through an appropriate methodology are ongoing using the PanGenEU study resources. However, the integrative models in large-scale epidemiologic research still face numerous challenges, some of them at the analytical stage. I will comment on the efforts we do to better characterize pancreatic cancer risk factors and the strategies we plan to apply to build integrative predictive risk scores.

NOTES



Polygenic susceptibility to colorectal cancer: mechanisms and impact

Richard HOULSTON

The Institute of Cancer Research, London, UK

The assertion that much of the heritable risk to colorectal cancer (CRC) has a polygenic basis has been vindicated by recent studies. We shall consider the impact of this class of susceptibility to CRC risk and its biological basis.

NOTES

Mutations and their interactions in individuals

Ben LEHNER

Centre for Genomic Regulation, Barcelona, ES

We use data and diverse model systems to understand how genetic and non-genetic variation results in the phenotypic differences amongst individuals. Our recent work has focussed on three areas: (1) understanding the distribution of mutation rates and types across the genome in human cancer cells, (2) understanding how mutations combine to alter phenotypes at different molecular levels using deep mutation scanning experiments, and (3) understanding inter-generational influences on phenotypic variation.

NOTES

Coding and non-coding cancer mutations

Nuria LÓPEZ-BIGAS

Institut de Recerca Biomèdica, Barcelona, ES

Somatic mutations are the driving force of cancer genome evolution. The rate of somatic mutations appears to be greatly variable across the genome due to variations in chromatin organization, DNA accessibility and replication timing. However, other variables that may influence the mutation rate locally are unknown. I will discuss recent findings from our lab on how DNA-binding proteins and differences in exons and introns influence mutation rate. These finding have important implications for our understanding of mutational and DNA repair processes and in the identification of cancer driver mutations. Given the evolutionary principles of cancer, one effective way to identify genomic elements involved in cancer is by tracing the signals left by the positive selection of driver mutations across tumours. We analyze thousands of tumor genomes to identify driver mutations in coding and non-coding regions of the genome.

NOTES

New insights into the genetics of asthma

Florence DEMENAIS

Genetic Variation and Human Diseases, UMR-946, Inserm-Université Paris Diderot, Paris, FR

Asthma is a chronic disease affecting 300 million people worldwide and its prevalence varies between populations and ethnicity. Asthma results from multiple genetic and environmental factors. The contribution of genetic factors to asthma risk has been demonstrated in family studies, where heritability estimates range from 25%-80%. There have been considerable efforts to characterize the genetic determinants of asthma, including candidate gene studies, positional cloning studies and more recently genome-wide association studies (GWAS).

We recently conducted a meta-analysis of worldwide asthma GWAS (23,948 cases, 118,538 controls) from ethnically-diverse populations (which are part of the Trans-National Asthma Genetic Consortium). We identified five new asthma loci, uncovered two additional novel associations at two asthma loci previously reported in ancestry-specific populations, established asthma associations at two loci implicated previously in comorbidity of asthma plus hay fever, and confirmed nine known loci. All the asthma-associated loci identified in this study are enriched in enhancer marks and are likely to be involved in gene regulation. Interestingly, the best candidates at a number of loci are involved in immune response to viruses or bacteria, underlining the importance of infections in asthma risk. This study also provided evidence for overlap of asthma loci with loci underlying autoimmune diseases and other diseases that have an inflammatory component.

Future discoveries and progress in the understanding of the mechanisms underlying asthma might come by exploring more complex models (e.g., gene-gene and gene-environment interactions) of asthma phenotypes and through the joint analysis of asthma and other immune-mediated and inflammatory diseases. The integration of biological knowledge with GWAS outcomes through pathway or network analysis may serve this purpose at least in part. Genome-wide exploration of the epigenome while integrating information on genetic variation and environmental exposures histories may allow deciphering the gene regulatory mechanisms involved in asthma.

NOTES

Genomic loci associated with birth weight identify genetic links between intrauterine growth and adult metabolic disease

Momoko HORIKOSHI
RIKEN, JP

In observational epidemiology, birth weight is associated with higher risk of developing adult metabolic diseases such as type 2 diabetes (T2D). Both genetic and non-genetic factors underlie association between T2D and both ends of birth weight distribution. We focused on the genetic contribution by performing a large-scale trans-ethnic genome-wide association study (GWAS) of birth weight. GWAS has successfully identified 60 loci associated with birth weight, 9 of which overlap with T2D-associated loci. There is an overall inverse genetic correlation between birth weight and T2D, which supports the fetal insulin hypothesis. However, individual loci show heterogeneity in association with T2D and some T2D risk alleles are associated with higher birth weight. In most cases, this is a reflection of the maternal genotype, which raises maternal glucose and results in increased insulin-mediated fetal growth. We also observed enrichment of birth weight associations in imprinted genes. Future studies with large number of mother-child pairs (or trios) will enable to disentangle the genetic contribution of fetus and the parents.

NOTES

New insights into blood pressure regulation from large-scale genetic studies

Patricia MUNROE

William Harvey Research Institute,

Barts and The London School of Medicine and Dentistry,

Queen Mary University of London, UK

Over the past 10 years' substantial progress has been made mapping blood pressure loci using genome-wide association studies and deploying bespoke microarrays (most recently the Cardio-Metabochip and Exome chip arrays) in very large sample sizes. There are now over 400 blood pressure loci, providing hundreds of candidate genes, and new insights into pathways that are key in blood pressure regulation. My presentation will discuss the results of the most recent analyses, highlighting the key findings and on-going projects.

NOTES

A Genomewide association study of aortic stenosis

George THANASSOULIS
McGill University Health Center, CA

Aortic stenosis (AS) is the most prevalent valvular heart disease in developed countries. Currently, the LPA locus is the only genome-wide significant susceptibility locus for AS and no large scale genomewide association study has been performed. The identification of additional genetic loci could inform drug development for AS, a disease without effective medical therapy. We performed a genomewide association study in the Genetic Epidemiology on Adult Health and Aging Cohort with replication is several additional cohorts worldwide, as well as further validation with valve calcium, a subclinical phenotype. We replicate the LPA locus and provide evidence for additional loci associated with AS. Follow-up metabolomic studies suggest a key lipid pathway may be implicated in the genesis of valve calcification and AS with possible therapeutic implications.

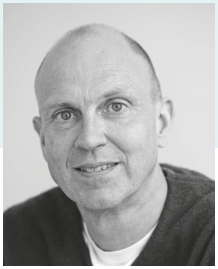
NOTES

Genomics of complex cerebrovascular disease

Stephanie DEBETTE
University of Bordeaux, FR

With increasing longevity, cerebrovascular disease has become the most common age-related source of disability and dependence, as well as a major determinant of dementia, and represents a huge societal burden. Compared to cardiovascular disease, unraveling the genomic underpinnings of cerebrovascular disease has proven more complex due to its much more heterogeneous nature. Recent multiethnic collaborative efforts have enabled important progress in revealing genetic determinants of stroke and MRI-defined covert vascular brain injury, primarily based on common variants but also early results on rare and low-frequency variants. Interestingly both a “lumping approach” combining all types of strokes with large numbers and strategies focusing on highly specific subtypes of cerebrovascular disease have proven efficient, with complementary findings. Analysis of shared genetic variation across different types of cerebrovascular disease and with other vascular phenotypes has also provided interesting mechanistic insight. Some of the findings have been followed up experimentally generating novel hypotheses on the biology of stroke.

NOTES



Mendelian randomization: what does the future hold?

George DAVEY-SMITH
University of Bristol, UK

Mendelian randomization is an increasingly widely used approach utilising genetic variants as causal anchors (or instrumental variables) in the analysis of observational epidemiological data. The approach circumvents many of the problems of conventional observational data analysis, including confounding, a variety of biases, and reverse causation. The method can be used to strengthen casual inference with regard to causes or treatments of disease, and drug target validation. This talk will speculate about ways in which Mendelian randomization may develop over the next few years to utilize multiple genetic variants, to further test the assumptions of the approach, to address potential mediation by epigenetic and other "omic" processes and ultimately to move towards hypothesis-free causality.

NOTES

Contributions from twin studies to gene discovery: Multi-variate, Multi-rater, Multi-age GWAS of longitudinal aggression and attention problems

Dorret BOOMSMA

Vrije Universiteit, Amsterdam, NL

Twin cohorts provide unique opportunities for investigations of the role of genetics in the etiology of complex human traits. Twins share environment throughout fetal periods and early years of life and twin designs harmonize this component of complex traits. The classical twin design compares mono- and dizygotic twin pair resemblance for uni- or multivariate traits and estimates the extent to which genes contribute to variation and covariation among traits. Twin cohorts tend to be population based, often have rich longitudinal follow up and increasingly collect DNA, RNA and biomarker data.

The EU-funded ACTION consortium (Aggression in Children: Unraveling gene-environment interplay to inform Treatment and InterventiON strategies) includes eight twin cohorts (from Scandinavia, The Netherlands, Australia and UK) which have phenotyped over 90.000 young twins for longitudinal aggressive behavior. The twin data revealed that genes substantially contribute to variation in aggression (AGG). The most important co-occurring behavioural problems with aggression in children and adults are attention (ATT) problems. Analyses of longitudinal observations of AGG and ATT revealed substantial stability with a significant portion of this stability attributable to genetic factors.

In children, behavior is assessed by parental or teacher ratings, leading to multivariate data, whose covariation is mainly due to genetic correlations. Genome wide association studies (GWAS) in children have considered AGG and ATT in univariate analyses, as have most other GWA meta-analyses projects either in children or in adults. Twin studies have been at forefront in leveraging the analyses of multivariate and longitudinal data, but GWA studies have so far made use of these approaches in a limited way. The analysis of multiple measures in GWA studies is non-trivial as repeatedly including the same subject in meta-analysis inflates type-1 error. The ACTION consortium is –in collaboration with other groups- carrying out a multivariate GWAS of the (developmental) genetic

etiology of AGG and ATT in 30,000 children. A multivariate meta-regression model is used (Nivard et al, doi: 10.1093/schbul/sbx031 2017), which relies on univariate GWASs of longitudinal phenotypes within cohorts. The multivariate analysis is performed at the meta-analysis stage, omitting the need for multivariate modeling in individual cohorts. The method enables the inclusion of dependent measures into a single analysis without inflating the false positive rate and allows for the estimation of age, rater or cohort specific effects for individual SNPs.

NOTES

Sequencing and Analysis of 10,000s of Human Genomes: Challenges and Opportunities

Gonçalo ABECASIS
University of Michigan, USA

Rapid advances in genome sequencing and genotyping technology are enabling increasingly detailed analysis of human genetic variation. In the next year, we expect to analyze >50,000 deeply sequenced human genomes, corresponding to ~5 million billion bases of raw sequence data. The generation, transfer and analysis of the data present many opportunities for scientific discovery - enabling better understanding of human history, biology and disease. It also presents varied computational and analytical challenges as well as opportunities to develop and implement new analytical strategies and modes of data sharing. I will illustrate these challenges and opportunities with examples from ongoing studies.

NOTES

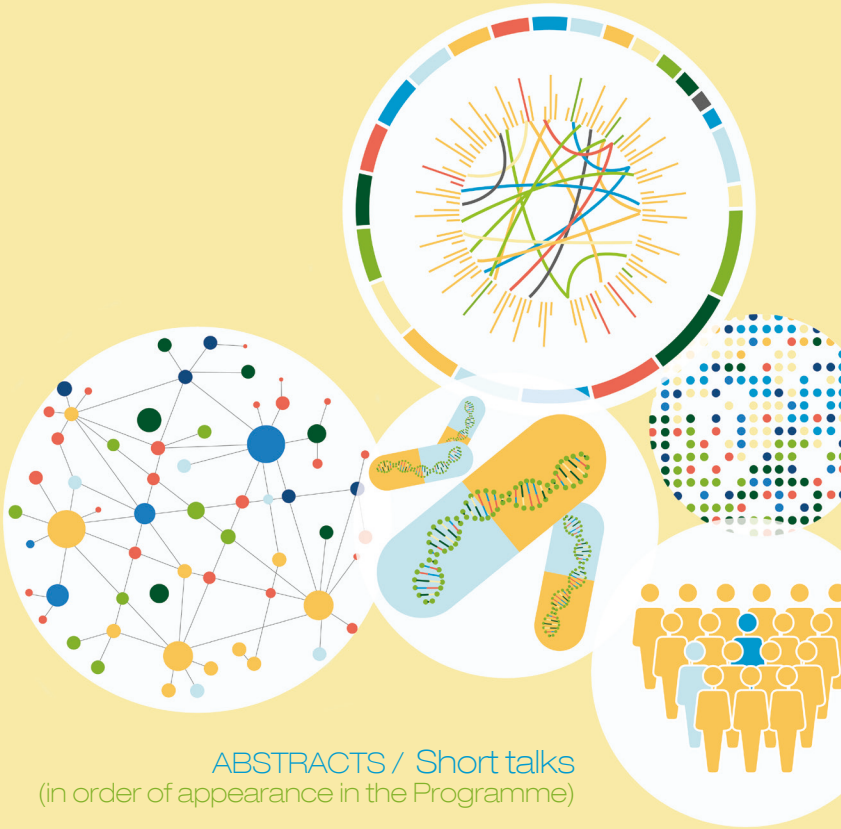
The 100,000 Genomes Project transforming healthcare

Mark CAULFIELD

Chief Scientist, Genomics England,
William Harvey Research Institute, Queen Mary University of London, UK

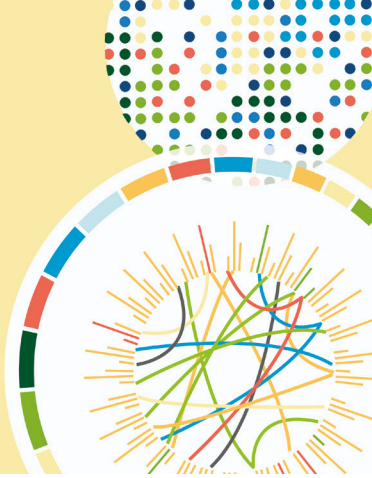
The 100,000 Genomes project is using whole genome sequencing to bring diagnoses to patients with rare inherited disorders, identify drivers to cancer and response to therapy and drivers to antimicrobial resistance in pathogens. This is transforming the capability and capacity of the National Health Service (NHS) to apply genomic medicine for patient benefit. To do this we have created 13 NHS Genomic Medicine Centres across England to enable generation of clinical data and sample flows from NHS patients with broad consent for whole genome sequencing. With our partner, Illumina we have one of the largest Next Generation Sequencing Centres in the World where we have sequenced 31,000 whole genomes. The value of this programme will be the alignment of the highest fidelity whole genome DNA sequence with detailed clinical data stored in pseudonymised format within a multi-petabyte data infrastructure. This will allow us to create a picture of life-course health and disease progression for participants. To drive up diagnoses for patients we have created the Genomics England Clinical Interpretation Partnership where 2600 clinicians and scientists will work on these data to enhance value for patients. Alongside this significant advance in NHS capability to utilise next generation sequencing for clinical care we have established over 700 person years of training to ensure we create the next generation of clinicians and scientists. This programme will ensure that our NHS has the capacity and capability to usher in a new era of Genomic Health and that the UK is amongst the most advanced in the world.

NOTES



ABSTRACTS / Short talks
(in order of appearance in the Programme)

*These abstracts should not be cited in bibliographies. Material contained herein should be treated as personal communication and should be cited as such only with consent of the author.



SPlinted Ligation Adapter Tagging (SPLAT), a novel library preparation method for whole genome bisulfite sequencing

Jessica NORDLUND[1]*, Amanda Raine[1], Tomas Axelsson[1], Pontus Larsson[1], Ulrika Liljedahl[1] and Ann-Christine Syvänen[1]

[1] LSNP&SEQ Technology Platform, National Genomics Infrastructure, Science for Life Laboratory, Uppsala University, Sweden

Whole genome bisulfite sequencing (WGBS) is a powerful tool for genome-wide profiling of DNA methylation at a single-base resolution. WGBS library preparation is challenging due to side effects of the bisulfite treatment, which leads to extensive DNA damage. WGBS constitutes a particular case of low diversity sequences where the base composition is reduced to three nucleotides (A,T,G) and a small fraction of Cs, which represent methylated cytosines. Sequencing of WGBS on Illumina platforms has historically been challenging, frequently leading to low data yields and inferior sequencing quality.

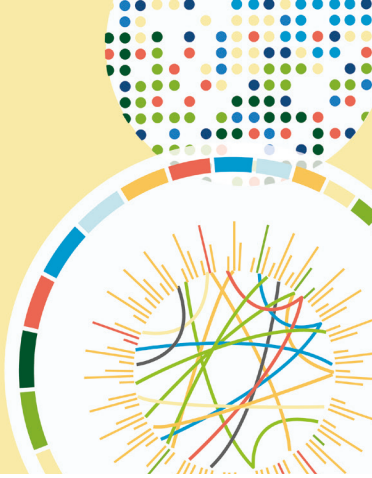
We recently developed a novel approach for WGBS library preparation that is based on splinted ligation adapter tagging (SPLAT) of single-stranded bisulfite-treated DNA [1]. This approach enables quick and efficient preparation of WGBS libraries from low-input DNA that compares favorably to commercial methods. Recent developments in patterned flowcell technology (Illumina) have significantly increased throughput and reduced the cost of WGBS. We performed WGBS on the HiSeq2500, HiSeqX, and NovaSeq platforms and demonstrate differences in quality scores across the different platforms and HSC/RTA versions. Despite the low Q-scores assigned to guanines by certain software versions, we observed only minor differences in global methylation levels across libraries prepared with the same method. Rather, we observe that global methylation rates vary depending on the choice of library preparation protocol.

This work was performed by the R&D team at the SNP&SEQ Technology Platform, which is part of the National Genomics Infrastructure within Science for Life Laboratory, a national center for molecular biosciences with focus on health and environmental research in Sweden. SNP&SEQ is an ISO/IEC 17025 accredited laboratory that offers next generation sequencing and genotyping services of high quality at the lowest possible costs to academic users. In addition to WGBS, we offer array-based DNA methylation analysis, SNP-genotyping, whole genome, linked-read, RNA, RRBS, and targeted sequencing services.

References

[1] *SPlinted Ligation Adapter Tagging (SPLAT), a novel library preparation method for whole genome bisulphite sequencing. 2017. Amanda Raine, Erika Manlig, et al. Nucleic Acids Research. 45 (6): e36*

NOTES



Metabolomics: The Missing Link in Precision Medicine

Nina GÖRNER[1], Lining Guo[1] and John Ryals[1]

[1] Metabolon Inc., Durham, North Carolina, USA

Metabolon has developed a platform technology capitalizing on advances in mass spectrometry, proprietary software and database analysis that provide unprecedented insight into biochemical pathways. Along with information delivered by whole genome or exome sequencing technologies, metabolite data provides ontology for determining genome wide associations. This knowledge is critical for determining penetrance of a gene determinant of interest and relating it to health status — whether part of a large population cohort or for an individual patient.

We present here work resulting from the analysis of plasma samples from 80 adults of normal health.

The comprehensive metabolic profiles provided a functional readout to assess the penetrance of gene mutations identified by whole-exome sequencing on these individuals. Conversely, metabolic abnormalities identifed by statistical analysis uncovered potential damaging mutations that were previously unappreciated. Additionally, we found metabolic signatures consistent with early signs of disease conditions and drug effects associated with efficacy and toxicity.

References

[1] *Plasma metabolomic profiles enhance precision medicine for volunteers of normal health. 2015. Lining Guo, Michael V. Milburn. PNAS. 2015 Sep 1; 112(35): E4901–E4910*

NOTES

Genetic study of body mass index in Japanese population links cell-types to body weight regulation

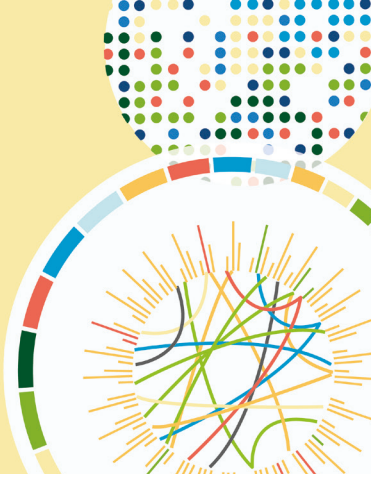
Masato AKIYAMA[1], Yukinori Okada[1, 2], Masahiro Kanai[1, 2], Michiaki Kubo[3] and Yoichiro Kamatani[1, 4]

[1] Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences.

[2] Department of Statistical Genetics, Osaka University Graduate School of Medicine.

[3] RIKEN Center for Integrative Medical Sciences.

[4] Center for Genomic Medicine, Kyoto University Graduate School of Medicine.



Although genome-wide association studies (GWAS) of body-mass index (BMI) have identified more than 100 susceptibility loci1, 2, investigations in Asian population have been limited. We carried out GWAS using 158,284 individuals who participated in the Biobank Japan project3. In the replication study, variants showed suggestive level of association ($P < 1.0 \times 10\text{-}5$) were evaluated in 15,146 individuals of Japanese population-based cohorts. A fixed effect meta-analysis of association studies in Japanese detected 85 loci satisfied genome-wide significant level ($P < 5.0 \times 10\text{-}8$). Of which, 51 were novel. We also performed a trans-ethnic meta-GWAS by MANTRA4 using publicly available result of Europeans which was provided by GIANT consortium1, and identified additional 61 new loci significantly associated (log10 Bayes' factor > 6). We compared the effect sizes of identified variants, and observed significant correlation ($r = 0.85$, $P = 8.69 \times 10\text{-}55$). Next, we conducted cell-type specificity analyses. First, we performed an enhancer enrichment analysis using the variants included in the 99% credible sets determined by the trans-ethnic fine-mapping with active enhancer information of the ChromHMM18-state model which was constructed by the Roadmap Epigenomics Consortium5. We found that associated variants showed significant enrichment for active enhancers of B-cell, adipose, and cells belonging to central nervous system (false discovery rate < 5%). Second, we evaluated overlaps between associated variants identified in Japanese population and H3K4me3 peaks of 34 cell-types using EpiGWAS software6. Significant overlaps were observed in B-cell, pancreas, and inferior temporal lobe of the brain ($P < 0.05$). Finally, we evaluated genetic correlations between BMI and blood cell counts by bivariate LD score regression7, and found a significant positive genetic correlation between BMI and lymphocyte count ($P = 6.46 \times 10\text{-}5$, $rg = 0.18$). These results provide pieces of genetic evidence that lymphocytes are relevant to body weight regulation in human.

References

[1] *Genetic studies of body mass index yield new insights for obesity biology. 2015. Adam E Locke, Bratati Kahali. Nature. 518, 197-206.*

[2] *Meta-analysis of genome-wide association studies in East Asian-ancestry populations identifies four new loci for body mass index. 2014. Wanqing, Wen, Wei Zheng. Hum. Mol. Genet. 23, 5492-504.*

[3] *Overview of the BioBank Japan Project: Study design and profile. 2017. Akiko Nagai, Makoto Hirata. J. Epidemiol. 27, 9-21.*

[4] *Transethnic meta-analysis of genomewide association studies. 2011. Andrew P. Morris. Genet. Epidemiol. 35, 809-822.*

[5] *Integrative analysis of 111 reference human epigenomes. 2015. Roadmap Epigenomics Consortium. Nature. 518, 317-30.*

[6] *Chromatin marks identify critical cell types for fine mapping complex trait variants. 2013. Gosia Trynka, Cynthia Sandor. Nat. Genet. 45, 124-30.*

[7] *An atlas of genetic correlations across human diseases and traits. 2015. Brendan Builk-Sullivan, Hilary K Finucane. Nat. Genet. 47, 1236-41*

NOTES

PharmGKB: The Pharmacogenomics Knowledgebase

Maria ALVARELLOS[1], Julia Barbarino[1], Li Gong[1], Katrin Sangkuhl[1], Caroline Thorn[1], Matt Devlin[1], Ryan Whaley[1], Mark Woon[1], Michelle Whirl-Carrillo[1], Russ B. Altman[1,2,3,4] and Teri E. Klein[1,2]

[1] Department of Biomedical Data Science.
[2] Department of Medicine - Biomedical Informatics Research.
[3] Department of Bioengineering.
[4] Department of Genetics.

Stanford University School of Medicine Shriram Center for Bioengineering and Chemical Engineering, 443 Via Ortega Rm 213 Stanford, CA 94305

The Pharmacogenomics Knowledgebase (PharmGKB https://www.pharmgkb.org)1 is a publicly available resource that collects, curates, and disseminates information about pharmacogenetics/genomics (PGx) by curating the primary literature, drug labels with PGx information, PGx-based dosing guidelines, and other relevant sources of PGx information. These sources support drug-gene variant clinical associations, drug-centered pathways and very important pharmacogene (VIP) summaries. Drug-gene associations with strong supporting evidence may also become the subjects of PGx-guided dosing guidelines developed by the Clinical Pharmacogenetics Implementation Consortium (CPIC https://cpicpgx.org/), a collaboration between PharmGKB and the Pharmacogenomics Research Network (PGRN http://www.pgrn.org/). PharmGKB also works with additional entities (https://www.pharmgkb.org/page/collaborators) to promote the understanding of PGx and its implementation into clinical care. These collaborations include but are not limited to, the development of a bioinformatics pipeline (PharmCAT) with the Pharmacogenomics Research Network (PGRN), working with Children's Mercy Hospital in Kansas City to build the PharmVar database to house PGx allele definitions and nomenclature, participating in the PGx working group of ClinGen and submission of curated variants to ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/). Finally, PharmGKB is also integral to the efforts of two data consortia: the International Clopirogrel Pharmacogenetics Consortium (ICPC), and the African American Cardiovascular Pharmacogenetics CONsorTium (ACCOUNT). PharmGKB is supported by the NIH/NIGMS R24 GM61374 and CPIC is supported by R24 GM115264.

References

[1] *Genetic studies of body mass index yield new insights for obesity biology. 2015. Adam E Locke, Bratati Kahali. Nature. 518, 197-206.*

[2] *Meta-analysis of genome-wide association studies in East Asian-ancestry populations identifies four new loci for body mass index. 2014. Wanqing, Wen, Wei Zheng. Hum. Mol. Genet. 23, 5492-504.*

[3] *Overview of the BioBank Japan Project: Study design and profile. 2017. Akiko Nagai, Makoto Hirata. J. Epidemiol. 27, 9-21.*

[4] *Transethnic meta-analysis of genomewide association studies. 2011. Andrew P. Morris. Genet. Epidemiol. 35, 809-822.*

[5] *Integrative analysis of 111 reference human epigenomes. 2015. Roadmap Epigenomics Consortium. Nature. 518, 317-30.*

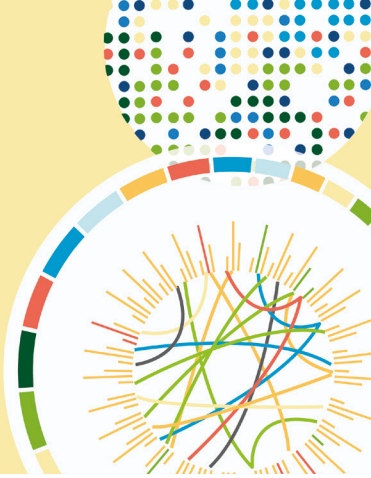
[6] *Chromatin marks identify critical cell types for fine mapping complex trait variants. 2013. Gosia Trynka, Cynthia Sandor. Nat. Genet. 45, 124-30.*

[7] *An atlas of genetic correlations across human diseases and traits. 2015. Brendan Builk-Sullivan, Hilary K Finucane. Nat. Genet. 47, 1236-41*

NOTES

Genome-scale data integration for deciphering pancreatic cancer aetiology

Evangelina LÓPEZ DE MATURANA[1,2], Lola Alonso 1,2, Isabel Martín-Antoniano 1,3, Francisco X. Real 4,5, Núria Malats 1,2, on behalf of the PanGenEU, ISBlaC and EPICURO Studies Investigators.



Pancreatic cancer (PC) is a complex disease whose frequency is increasing in Westernized countries. The knowledge of PC genetic and epigenetic basis is still incomplete, as limited number of genome-wide association studies (GWAS) have been conducted, and current research on the role of the methylome in PC is still in its infancy.

Here, we present a genome-scale data integration study to shed light on the PC aetiology. First, we performed a PC-GWAS (the first one conducted in European population) to identify novel hits associated to PC risk, using the resources of 1317 cases and 1616 controls from PanGenEU, ISBlaC and SBC/EPICURO studies. Then, we prioritized suggestive associations ($p<1\times10-4$) for further bioinformatics analysis, including evidence of functional impact, annotation of tagged genes and pathways, eQTLs and regulatory chromatin marks in pancreas, and mQTLs in leukocytes. Functional data were retrieved using DoriTool, a novel integrative pipeline developed in our lab. Moreover, we used data from assays conducted in pancreas/whole blood catalogued in the ENCODE project, and the DNA methylation profile in leukocytes of 265 controls from PanGenEU study.

We identified 139 novel variants, whose replication using the resources of PanScanI-III and PanC4 is undergoing. Most variants were in chromosomes 1, 6, 8 and 10. In particular, 8q24.21, a region previously associated with many cancers, harbored the largest number of the identified variants, followed by the 10q11.21 region. Variants in 7q34, 12q24.33 and 17q24.2 were potentially functional since either they were eQTLs in pancreas, they overlapped with regulatory elements, they tagged genes annotated in PC-related pathways or they were associated with mQTLs in leukocytes.

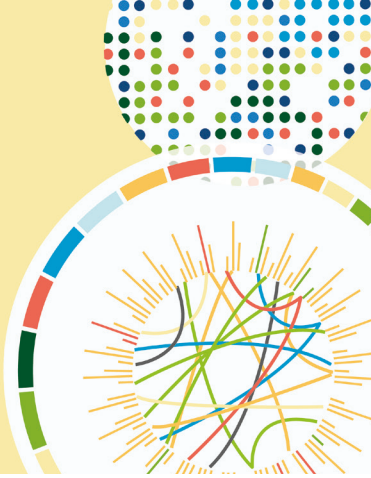
Our work adds important new knowledge to understand the genetics underlying PC through a post-GWAS functional in-silico analysis, by integrating additional –omics data that may reveal potential causal regions.

NOTES

Transcriptomic and epigenomic mechanism of mosaic loss of chromosome Y (LOY) in cancer

Juan GONZÁLEZ[1,2,*], Marcos López[1,3], Aina Jené[1] and Luis A Pérez-Jurado[3]

[1] Barcelona Institute for Global Health, Spain.
[2] CIBER in Epidemiology and Publich Health, Spain.
[3] Genetics Unit, Universitat Pompeu Fabra, Barcelona, Spain.



The greatest contribution of genetic epidemiological studies to date has been provided by GWAS. Deeper analyses of existing data, mainly based on exploiting genotype-array intensity data, would help in addressing the missing heritability problem. In addition to CNVs, intensity data enables a measure of loss of chromosome Y (LOY). It has been demonstrated that LOY is associated with aging and age-related disease such as cancer1,2 of Alzheimer's disease3. However, the clinical relevance and pathogenic mechanisms of LOY have been poorly evaluated to date due to the difficulties to analyze at population scale. In order to fill this gap, a total of 3,487 tumor samples belonging to The Cancer Genome Atlas project were analyzed. LOY status was inferred from SNP-array data using our MADloy tool4. Association analyses revealed that the pooled odds ratio (OR) of association between LOY and tumor samples was 11.0 (CI95% 6.6 – 18.3). Differences in the magnitude of the OR were observed across cancer types, with no association with prostate cancer and high association with kidney tumor. Transcriptomic data analyses showed that only genes in chromosome Y were down-regulated having strong effect in kidney, bladder and lung cancer. Epigenomic studies provided a large list of CpGs associated with LOY in Y chromosome as well as in autosomes. Gene enrichment analysis revealed a range of pathways involved in LOY that vary across tumors. These include cell cycle, metabolic processes and proteasome. Some of deregulated genes shared sites of the transcription factors such as E2F, ETS and CEBPB. In summary, this study shows how the reanalysis of existing GWAS along with functional data in epidemiological studies may help in discovering new genomic variants that contribute a larger proportion of trait variability than the one obtained by simply analyzing SNP.

References

[1] *Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer.*
2014. Fosberg, et al.. Nature Genetics. 46, 624-628.

[2] *Mosaic loss of chromosome Y is associated with common variation near TCL1A.*
Zhou et al.. Nature Genetics.2016. 48,563-8.

[3] *Mosaic Loss of Chromosome Y in Blood Is Associated with Alzheimer Disease.*
2016. Dumansky et al.. Am J Hum Genet. 98, 1208–1219.

[4] *MADloy: mosaic alteration detector of LOY.*
Accesible at https://github.com/isglobal-brge/madloy_v1.

NOTES

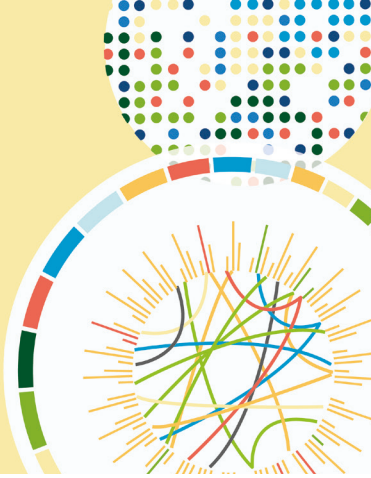
DNA co-methylation networks for colon cancer patient stratification

Izaskun MALLONA[1], Anna Díez-Villanueva[2], Francisco Chen[1], Víctor Moreno[2] and Miguel A. Peinado[1*]

[1] Health Research Institute Germans Trias i Pujol (IGTP), Predictive and Personalized Medicine of Cancer Program, Can Ruti Campus, Barcelona

[2] Colorectal Cancer Group, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona

* mpeinado@igtp.cat



Epigenetic alterations are a widespread signature of human cancer, including generalized changes in DNA methylation status and variability. Whilst frequently affecting small to large regions of co-localized CpGs (i.e. CpG island shores), non adjacent loci also co-methylate.

To integrate localized and trans DNA methylation changes we propose a novel network-based analysis leveraging DNA methylation changes in sympathy across colorectal cancer patients. Aimed to detect both contiguous and non-contiguous effects, it extracts highly connected modules of co-methylated CpGs, whose DNA methylation profiles are next classified into discrete classes, allowing samples classification according to the DNA methylation status of their members.

Built upon Infinium 450k array measures from four datasets independently (two cohorts including tumor and matched normal data), the tumor co-methylation network is robust and partly overlaps the normal colon's. Structurally, CpGs at inactive promoters co-methylate frequently, while active promoters are depleted. Module-aware DNA methylation clustering of tumoral samples conflate normal-like and hypermethylated profiles; such classification is associated to differential gene expression and somatic mutation rate.

In summary, we have developed a method to extract shared patterns of co-methylation regardless of loci contiguity, thus overriding the traditional view of local effects of DNA methylation changes on the regulation of neighboring genes. Featuring DNA methylation at the CpG level, with no aggregation nor binning, our method reports potential heterogeneity at the gene sub levels (exons, UTRs...) and including loci outside genes (i.e. enhancers). The co-methylation aware samples classification shows association to molecular features of clinical relevance.

NOTES

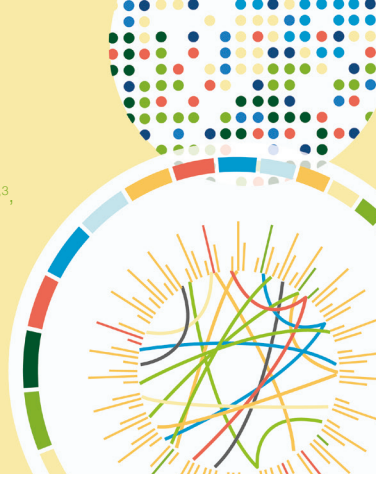
Recurrent somatic mutations reveal new insights into mutational processes in cancer

Miranda STOBBE[1,2,*], G.A. Thun[1,2], J.P. Whalley[1,2], M. Oliva[1,2,3], E. Raineri[1,2], I.G. Gut[1,2] and the PCAWG consortium of the ICGC

[1] CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain.

[2] Universitat Pompeu Fabra (UPF), Barcelona, Spain.

[3] University of Chicago Department of Medicine & Institute for Genomics and Systems Biology, Chicago, IL, United States



Although mutation rates in tumours are higher than spontaneous mutation rates in unaffected tissues, with over 3 billion base pairs the sheer size of the human genome makes it improbable that by chance an identical mutation is found at exactly the same position in the cancer genomes from different donors. Despite these odds, 1,057,935 (2.4%) of the ~43.4 million Somatic Single-base Mutations (SSMs) are recurrent in the 2,583 cancer genomes of the Pan-Cancer Analysis of Whole Genomes project of the International Cancer Genome Consortium. Furthermore, of the ~2.1 million Somatic Insertion/deletion Mutations (SIMs) 186,576 (8.7%) are recurrent. This suggests that there may be non-random mutational processes acting upon specific parts of the genome. To prove this hypothesis, the characteristics of the somatic mutations were captured with a total of 42 features, e.g. percentage of C>A SSMs, number of recurrent SIMs. Next, these features served as input for a principal component analysis followed by hierarchical clustering on the resulting principal components. The eight main clusters capture different mutational processes with varying levels of recurrency reflected in them. One of the clusters captures microsatellite instable tumours and is characterized by a high percentage of recurrent SIMs and of 1 bp C/G deletions in the context of a 5-10 bp C/G homopolymer. The cluster of mostly lung cancer samples is characterized by a high level of C>A SSMs, a high percentage of 1 bp C/G deletions and a low level of recurrency. A cluster dominated by oesophagus adenocarcinoma samples showed high levels of recurrent T>G and T>C SSMs, which potentially is linked to DNA damage caused by gastric reflux. By analysing the 2,583 samples together we were able to borrow information across the 39 tumour types and identify non-random mutational processes along with potential mechanisms that could explain them.

NOTES

Systematic discovery of germline cancer predisposition genes through the identification of somatic second hits

Solip PARK[1,2], Fran Supek1-3 and Ben Lehner1,2,4*

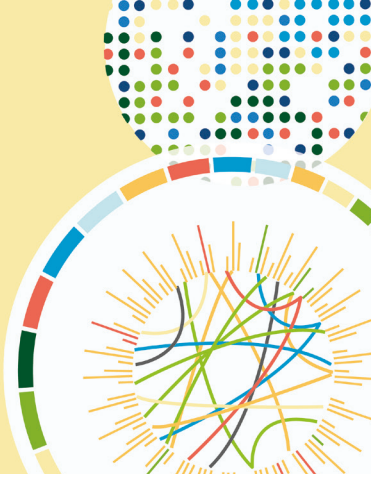
[1] CEMBL-CRG Systems Biology Unit, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology, Barcelona, Spain.

[2] Universitat Pompeu Fabra (UPF), Barcelona, Spain.

[3] Division of Electronics, Rudjer Boskovic Institute, Zagreb, Croatia.

[4] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

* Correspondence should be addressed to B.L. (ben.lehner@crg.eu).



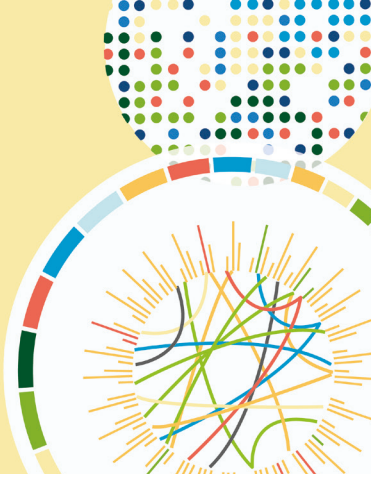
Cancer is caused by both somatically-acquired and inherited genetic variants. Large-scale tumor sequencing has revolutionized the identification of somatic driver alterations but has had limited impact on the identification of inherited cancer predisposition genes (CPDGs). Here we present a statistical method, ALFRED, that uses Knudson's two-hit hypothesis to systematically identify CPDGs from cancer genome data. Applied to ~10,000 tumor exomes the approach identifies known and new CPDGs that contribute to cancer through a combination of rare inherited variants and somatic loss-of-heterozygosity (LOH). Rare inherited variants in these genes contribute substantially to cancer risk, including to a median of ~6.8% of tumors across 17 cancer types, ~18% of ovarian carcinomas, ~9.7% of breast tumors, and ~7.2% of lung adenocarcinomas.

NOTES

A Transcriptomic Investigation on Patient–Specific Comorbidities

Jon SÁNCHEZ-VALLE[1,2], Héctor Tejero[2], David Juan[3], Vera Pancaldi[1,2], Fatima Al-Shahrour[2], Alfonso Valencia[1,4]

[1] Barcelona Supercomputing Center, BSC.
[2] Spanish National Cancer Research Centre, CNIO.
[3] Universitat Pompeu Fabra, UPF.
[4] ICREA.



Comorbidity is defined as a tendency to a lower/higher risk of developing a disease when already suffering from a previous one. Previous transcriptomic meta-analyses point to gene expression deregulation as potential driver of both inverse and direct comorbidity relations between central nervous system disorders (CNSd) and cancer 1-2. Despite the existence of know comorbidities at the population level, there is great variability in the protection for each specific patient.

The purpose of this study was to analyse comorbidity relations at a patient level using transcriptomic data, in order to detect patient clusters and specific molecular patterns driving both inverse and direct comorbidity relations. 108 diseases were analysed.

Differential gene expression analyses were conducted comparing each patient with all the control samples with the same gender. Fisher's exact tests were used to compare lists of the top 500 differentially up- and down-regulated genes. Pairs of patients with significant overlaps between genes deregulated in the same direction are considered to be directly comorbid, while those presenting significant overlaps between genes up-regulated in one disease and down-regulated in the other one, and vice versa, are considered to be inversely comorbid. For each disease we extracted patient clusters that optimizes the intra- and inter-disease coherence, based on patient-patient interactions. Using this procedure we obtained 548 clusters of which 458 present a significant inter-disease clustering while keeping the coherence in their own disease clusters. The results are represented in a graphical system that allows the direct evaluation of the clusters, the individual cases and the gene expression patterns associated to each cluster and relation.

The results obtained point to the need to analyse comorbidity relations breaking the general classification of diseases in smaller groups with specific genetic characteristics. In the long run, this type of analysis will open the door to personalized drug repurposing applications.

References

[1] *Molecular Evidence for the Inverse Comorbidity between Central Nervous System Disorders and Cancers Detected by Transcriptomic Meta-analyses. 2014. Kristina Ibañez, Cesar Bollousa. Plos Genetics. https://doi.org/10.1371/journal.pgen.1004173*

[2] *A molecular hypothesis to explain direct and inverse co-morbidities between Alzheimer's disease, Glioblastoma and Lung cancer. 2017. Jon Sánchez-Valle, Hector Tejero. Scientific Reports in the press.*

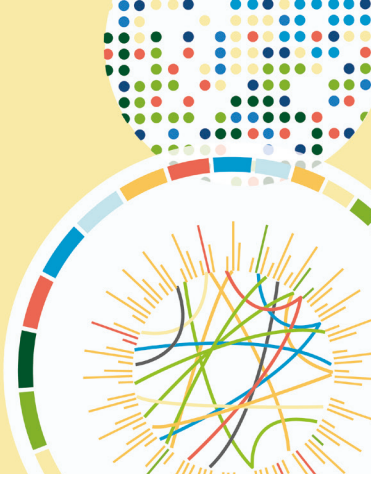
NOTES

Identification of polygenic adaptation in attention deficit hyperactivity disorder using GWAS data

Oscar LAO[1*], Paula Esteller-Cucala[1], Iago Maceda[1], The Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH), Psychiatric Genomics Consortium (PGC), Bru Cormand[2*]

[1] CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona. Institute of Science and Technology (BIST) Barcelona. Universitat Pompeu Fabra (UPF), 08003 Barcelona.

[2] Department of Genetics, Microbiology and Statistics, Faculty of Biology, Uiversity of Barcelona, Barcelona; Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona; Centro de Investigación en red de Enfermedades Raras (CIBERER), Spain; Institut de Recerca Sant Joan de Déu (IR-SJD), Esplugues de Llobregat, Spain.



According to the Out of Africa hypothesis, anatomically modern humans (AMH) appeared in Africa ~200 kilo years ago (kya). During the human Diaspora out of Africa, AMH adapted to the new environments they encountered (i.e. (1)). There is increasing evidence that human adaptation also occurred at a psychological traits (2). On top, adaptation process could have been fuelled by interbreeding with homo archaic species -such as Neanderthal and Denisovan- present at that time in Eurasia (3). Most of these adaptations are due to complex phenotypes involving many genes. However, identifying signatures of polygenic adaptation remains challenging and prior knowledge about the functional background in the surveyed loci is essential (4). This knowledge comes mainly from the output from genome wide association studies (GWAS). In the present study we have analysed the signatures of recent polygenic selection and the effect of archaic inbreeding in a complex behaviour trait such as ADHD taking advantage of the effect size estimates of millions of SNPs reported at a GWAS on ADHD conducted by the Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH) and the Psychiatric Genomics Consortium (PGC) on approximately 20K patients and 35K controls. Our results suggest that current North European populations show signatures of recent (<2,000 years ago) polygenic adaptation in the alleles that are protective for ADHD at a genomic level. Furthermore, analysis of ancient AMH European samples (ranging from ~2,000 years ago (ya) to ~45 kya) shows that the mean number of ADHD risk alleles per SNP linearly correlates with the sampling time more than expected under the hypothesis of neutrality. Finally, the analysis of the genetic variation of Altai Neanderthal suggests that this sample carries more ADHD risk alleles than current and ancient AMH samples. Furthermore, we observed that introgressed Neanderthal alleles influence ADHD susceptibility in current populations.

References

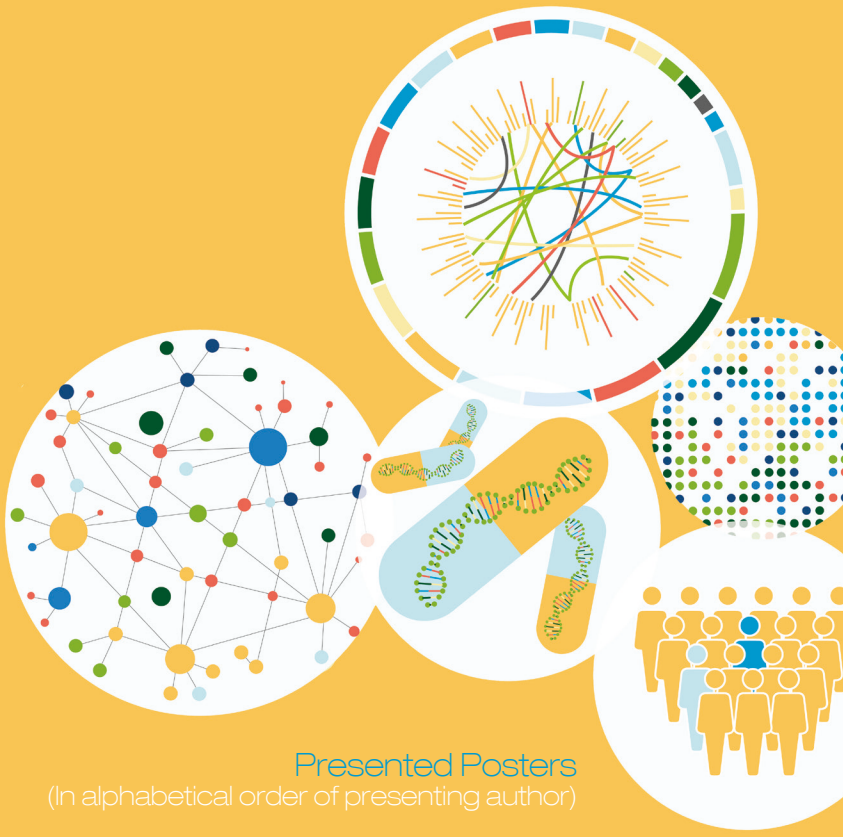
[1] *Evidence of Polygenic Adaptation in the Systems Genetics of Anthropometric Traits. 2016. R. Polimanti, B. Z. Yang, H. Zhao, J. Gelernter, PloS One. 11, e0160654.*

[2] *Widespread signatures of positive selection in common risk alleles associated to autism spectrum disorder. 2017. R. Polimanti, J. Gelernter,. PLoS Genet. 13, e1006618.*

[3] *Tracing the peopling of the world through genomics. 2017. R. Nielsen et al., Nature. 541, 302–310.*

[4] *A population genetic signal of polygenic adaptation. 2014. J. J. Berg, G. Coop, PLoS Genet. 10, e1004412.*

NOTES



Presented Posters

(In alphabetical order of presenting author)

P01 M. Henar Alonso[1], Rebeca Sanz-Pamplona[1], Victor Moreno[1,2] on behalf of the CORECT consortium.

Somatic mosaicism events in colorectal cancer patients

[1] Biomarkers and Susceptibility Unit, Institut Català d'Oncologia, IDIBELL and CIBERESP, L'Hospitalet de Llobregat, Barcelona
[2] University of Barcelona, L'Hospitalet de Llobregat, Barcelona

---

P02 Gemma Armengol[1] and Maria Cabezas[1], Mireia Camós[2,3,4], Susana Rives[2,4,5], África García-Orad[6,7], Josep Lluís Dapena[8], María Rosa Caballín[1,*].

Impact of polymorphisms in apoptosis-related genes on the outcome of childhood acute lymphoblastic leukemia

[1] Unit of Biological Anthropology, Department of Animal Biology, Plant Biology and Ecology, Faculty of Biosciences, UniversitatAutònoma de Barcelona, Bellaterra, Catalonia, Spain
[2] Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat, Catalonia, Spain.
[3] Hematology Laboratory, Hospital Sant Joan de Déu Barcelona, Esplugues de Llobregat, Catalonia, Spain.
[4] National Biomedical Research Institute on Rare Diseases (CIBER ER), Instituto de Salud Carlos III, Madrid, Spain.
[5] Pediatric Hematology and Oncology Department, Hospital Sant Joan de Déu Barcelona, Esplugues de Llobregat, Catalonia, Spain
[6] Department of Genetics, Physic Anthropology and Animal Physiology, University of the Basque Country, UPV/EHU, Leioa, Bizkaia, Spain
[7] BioCruces Health Research Institute, Barakaldo, Bizkaia, Spain.
[8] Service of Pediatric Hematology and Oncology, Hospital Universitari Vall d'Hebron, Barcelona, Catalonia, Spain.

P03 Robert Castelo[2] and Daniel Costa[1,2].

Gene expression profiling in archived neonatal dried blood spots reveals RNA changes associated with the fetal inflammatory response in extremely preterm newborns

[1] Dept. of Pediatrics, Hospital de Figueres
[2] Dept. Experimental & Health Sciences, Universitat Pompeu Fabra

---

P04 Anna Díez-Villanueva*[1] Berta Martín[2], Isabel Padrol[1], Mireia Obón-Santacana[1,2], Marco A. Fernandez[2], Llorenç Coll[2], Rafael de Cid[2], Miquel Àngel Peinado[2] and Víctor Moreno[1].

Identification of inherited DNA methylation loci in trios by BS-seq analysis

[1] Unit of Biomarkers and Susceptibility, Cancer Prevention and Control Programme, Catalan Institute of Oncology-IDIBELL, L'Hospitalet de Llobregat, Catalonia, Spain.

[2] Institute for Health Science Research Germans Trias i Pujol (IGTP), Program for Predictive and Personalized Medicine of Cancer (PMPPC), Can Ruti Biomedical Campus, Badalona, Catalonia, Spain.

P05 María del Mar González[1,*], Amanda Ramos[1,2,3], Maria Pilar Aluja[1] and Cristina Santos[1].

Sensitivity of mitochondrial DNA heteroplasmy detection using Next Generation Sequencing

[1] Unitat Antropologia Biològica, Dep. Biologia Animal, Biologia Vegetal i Ecologia, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Barcelona. Spain.
[2] Faculdade de Ciências e Tecnologia, Universidade dos Açores, Ponta Delgada, Portugal.
[3] Instituto de Investigação e Inovação em Saúde - Instituto de Biologia Molecular e Celular (IBMC), Universidade do Porto, Portugal.

P06 S. Laurie[1,2], D. Piscia[1,2], J. Protasio[1,2], A. Papakonstantinou[1,2], I. Martínez[1,2], A. Cañada[3,4], J.M. Fernández[3,4], R. Kaliyaperumal[5], S. Lair[6], P. Sernadela[7], T. Katsila[8], G.P. Patrinos[8], O.J. Buske[9], M. Girdea[9], M. Brudno[9], A. Blavier[6], J. Dawson[10], R. Thompson[10], H. Lochmüller[10], M. Bellgard[11], D. Spalding[12], M. Roos[5], P.A.C. 't Hoen[5], A. Valencia[3,4], D. Salgado[13,14], C. Béroud[13,14,15], I. Gut[1,2], S. Beltran[1,2] and the RD-Connect Consortium.

Using the RD-Connect analysis platform to solve and share rare disease cases

[1] Centro Nacional de Análisis Genómico (CNAG-CRG), Center for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), Barcelona, Spain
[2] Universitat Pompeu Fabra (UPF), Barcelona, Spain
[3] Centro Nacional de Investigaciones Oncológicas (CNIO), Madrid, Spain
[4] Instituto Nacional de Bioinformática (INB), Spain
[5] Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands
[6] Interactive Biosoftware, Rouen, France
[7] DETI/IEETA, University of Aveiro, Portugal
[8] Department of Pharmacy, School of Health Sciences, University of Patras, University Campus,Patras,Greece
[9] Centre for Computational Medicine, Hospital for Sick Children and University of Toronto, ON, Canada
[10] Institute of Genetic Medicine, MRC Centre for Neuromuscular Diseases, Newcastle University, UK
[11] Centre for Comparative Genomics, Murdoch University, Perth, Western Australia
[12] European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Cambridge, United Kingdom
[13] Aix-Marseille Université, Marseille, France
[14] Inserm, UMR_S 910, Marseille, France
[15] APHM, Hôpital TIMONE Enfants, Laboratoire de Génétique Moléculaire, Marseille, France

P07 Nadezda Lipunova[1,2], Richard T. Bryan[1], Jean-Baptiste Cazier[1,3], Anke Wesselius[2], Kar K. Cheng[4], Frederik-Jan van Schooten[5], Maurice P. Zeegers[1,2].

GWAS for tumour size, grade, stage, and age of onset in NMIBC patients in West Midlands Bladder Cancer Prognosis Programme

[1] Institute of Cancer and Genomic Sciences, University of Birmingham, United Kingdom;
[2] Department of Complex Genetics, Maastricht University, The Netherlands;
[3] Centre for Computational Biology, University of Birmingham, United Kingdom,
[4] Institute for Applied Health Research, University of Birmingham, United Kingdom;
[5] Department of Genetic and Molecular Toxicology, Maastricht University, The Netherlands

---

P08 Adriana Lopez-Doriga[1], Rebeca Sanz-Pamplona[1], Pere Gelabert[2], M. Henar Alonso[1], Víctor Moreno[1,3,4].

Estimation of mitochondria abundance from Exome Sequencing and association with clinical and molecular factors in colon cancer

[1] Institut Català Oncologia (ICO)
[2] Institut de Biologia Evolutiva (IBE) CSIC, UPF
[3] Institut d'Investigació Biomèdica de Bellvitge (IDIBELL)
[4] Department of Clinical Sciences, Faculty of Medicine, Universitat de Barcelona

---

P09 Isabel Martín-Antoniano[1,3*], Lola Alonso[1,2*], Miguel Madrid[4], Evangelina López de Maturana[1,2], Núria Malats[1,2].

DoriTool: A bioinformatics integrative tool for post-association functional annotation

[1] Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Centre, Madrid
[2] Centro de Investigación Biomética en red Cáncer (CIBERONC), Spain.
[3] Instituto de Medicina Molecular Aplicada, Facultad de Medicina, Universidad San Pablo CEU. Madrid, Spain.
[4] Structural Computational Biology Group, Spanish National Cancer Research Centre, Madrid, Spain

---

P10 Leslie Matalonga[1,2], R. Tonda, D. Piscia[1,2], S. Laurie[1,2], J. P. Whalley[1,2], R. Thompson[3], H. Lochmüller[3], I. Gut[1,2], S. Beltran[1,2].

Semi-automated generation of custom clinical genomic reports for rare diseases

[1] Centro Nacional de Análisis Genómico (CNAG-CRG), Center for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), Barcelona, Spain
[2] Universitat Pompeu Fabra (UPF), Barcelona, Spain
[3] Institute of Genetic Medicine, MRC Centre for Neuromuscular Diseases, Newcastle University, UK

P11 Cristina Rodríguez-Antona[1,5,†,*], María Santos[1], Mikko Niemi[2], Masahiro Hiratsuka[3], Masaki Kumondai[3], Magnus Ingelman-Sundberg[4] and Volker M. Lauschke[4,†].

Novel copy number variants in pharmacogenes contribute to interindividual differences in drug pharmacokinetics

[1] Hereditary Endocrine Cancer Group, Human Cancer Genetics Programme, Spanish National Cancer Research Centre (CNIO), Spain.
[2] Department of Clinical Pharmacology, University of Helsinki and Helsinki University Hospital, Finland.
[3] Laboratory of Pharmacotherapy of Life-Style Related Diseases, Graduate School of Pharmaceutical Sciences, Tohoku University, Sendai, Japan.
[4] Section of Pharmacogenetics, Department of Physiology and Pharmacology, Karolinska Institutet, Sweden.
[5] ISCIII Center for Biomedical Research on Rare Diseases (CIBERER), Spain.
[†] Co-senior authorship.

P12 Justin P. Whalley[*1,2] and Ivo G. Gut[1,2] on behalf of RD-Connect and PCAWG.

Framework for quality assessment in whole exome and whole genome sequencing

[1] CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain.
[2] Universitat Pompeu Fabra (UPF), Barcelona, Spain.

P13 Leszek P. Pryszcz[1*], Katarzyna Misztal[1,2], Matthias Bochtler[1,2], Cecilia L. Winata[1,3].

Deciphering the role of RNA editing in vertebrate development

[1] IIMCB, Warsaw.
[2] IBB, Warsaw.
[3] Max Planck.

CRG
Centre for Genomic Regulation

PRBB Building
Dr. Aiguader, 88
08003 Barcelona, Spain
Tel.: +34 93 316 01 00 Fax: +34 93 316 00 99
comunicacio@crg.eu http://www.crg.eu

Sponsored by:

illumina® Atos bionano GENOMICS Roche Agilent Technologies

Supported by:

EXCELENCIA SEVERO OCHOA

Member of:

BIST Barcelona Institute of Science and Technology

Promotion Partner:

BIB BIOINFORMATICS BARCELONA

Members of the Board of Trustees:

Generalitat de Catalunya

GOBIERNO DE ESPAÑA MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

upf. Universitat Pompeu Fabra Barcelona

Obra Social "la Caixa"