7th Sardinian International Summer School

Initiative financed and implemented with the support of Sardegna Ricerche and the Autonomous Region of Sardinia

REGIONE AUTÒNOMA DE SARDIGNA
REGIONE AUTONOMA DELLA SARDEGNA

SARDEGNA RICERCHE

"From Genome-wide association studies (GWAS) to function"

From July 09th to 13th, 2018

Sardinia Technology Park , Pula (CA), Italy

Directors:

Francesco Cucca (Italy)
Marcella Devoto (Italy, USA)
Giovanni Romeo (Italy)

Faculty:

Goncalo Abecasis (USA)
Myles Axton (USA)
Rachael Bashford-Rogers (UK)
Francesco Cucca (Italy)
Anna Di Rienzo (USA)
Daniel Gaffney (UK)
Arthur Gilly (UK)
Kylie James (UK)
Mauro Pala (Italy)
Clelia Peano (Italy)
Stephen Sawcer (UK)
David Schlessinger (USA)
Nicola Segata (Italy)
Carlo Sidore (Italy)
Nicole Soranzo (UK)
John Todd (UK)
Eleftheria Zeggini (UK)

SARDEGNA RICERCHE

IRGB — Istituto di Ricerca Genetica e Biomedica

EUROPEAN SCHOOL OF GENETIC MEDICINE

CRS4 — IDEAS BECOME LIFE

CENTRO RESIDENZIALE UNIVERSITARIO — Bertinoro

Web site: www.irgb.cnr.it/summerschool

Info: info_sc2018@irgb.cnr.it

FROM JULY 9TH TO JULY 13TH 2018, PULA (CA) - ITALY
7th Sardinian International Summer School
From GWAS to function

Directors: Francesco Cucca (Italy), Marcella Devoto (Italy, USA), Giovanni Romeo (Italy)

Faculty: Goncalo Abecasis, USA; Myles Axton, USA; Rachael Bashford-Rogers, UK; Francesco Cucca, Italy; Anna Di Rienzo, USA; Daniel Gaffney, UK; Arthur Gilly, UK; Kylie James, UK; Mauro Pala, Italy; Clelia Peano, Italy; Stephen Sawcer, UK; David Schlessinger, USA; Nicola Segata, Italy; Carlo Sidore, Italy; Nicole Soranzo, UK; John Todd, UK; Eleftheria Zeggini, UK.

MONDAY JULY 9ST
Morning Lectures
08:30-9:00 **Registration**
09:00-10:00 **Introduction to complex trait genetics**– *Eleftheria Zeggini (UK)*
10:00-11:00 **The genomic and functional architecture of human complex traits and diseases** – *Nicole Soranzo (UK)*
11:00-11:30 **Coffee break**
11:30-12:30 **From GWAS to function through natural selection**– *Anna Di Rienzo (USA)*
12:30-14:00 **Lunch Break**
Afternoon Workshops
14:00-15:30 **Computational tools for the identification of adaptive alleles**– *Anna Di Rienzo (USA)*
15:30-16:00 **Coffee break**
16:00-17:30 **Hands-on tutorial to Genome-wide Association Studies (GWAS)** – Arthur Gilly (UK)
TUESDAY JULY 10 ST
Morning Lectures
09:00-10:00 **Low-coverage genomewide sequencing approaches for population studies** – *Goncalo Abecasis (USA)*
10:00-11:00 **Understanding the function of human genetic variation using transcriptome analysis** – *Daniel Gaffney (UK)*
11:00-11:30 **Coffee break**
11:30-12:30 **Shotgun metagenomics for the study of the human microbiome** –*Nicola Segata (Italy)*
12:30-14:00 **Lunch Break**
Afternoon Workshops
14:00-15:30 **NGS variant calling** – *Carlo Sidore (Italy)*
15:30-16:00 **Coffee break**
16:00-17:30 **The eQTLs Catalog and LinDA browser** – *Mauro Pala (Italy)*
WEDNESDAY JULY 11 ST
Morning Lectures
09:00-10:00 **Genetic analysis of Neuroblastoma in African Americans** - *Marcella Devoto (USA)*
10:00-11:00 **MS genetic - pitfalls and prospects**- *Stephen Sawcer (UK)*
11:00-11:30 **Coffee break**
11:30-12:30 **Student presentations**
12:30-14:00 **Lunch Break**
Free Afternoon
THURSDAY JULY 12 ST
Morning Lectures
09:00-10:00 **High-throughput sequencing reveals insights into the relationships between B-cell antibody repertoire, phenotype and function in health, cancer and autoimmune disease** - *Rachael Bashford-Rogers (UK)*
10:00-11:00 **Exploring immunity at single-cell resolution**- *Kylie James (UK)*
11:00-11:30 **Coffee break**
11:30-12:30 **From genetics to clinic in autoimmune diabetes** - *John Todd (UK)*
12:30-14:00 **Lunch Break**
Afternoon Workshops
14:00-15:30 **Poster Session**
15:30-16:00 **Coffee break**
16:00-17:30 **Microbiome and metagenome Data Analysis** – *Clelia Peano (Italy)*
FRIDAY JULY 13 ST
Morning Lectures
09:00-10:00 **Genetics and Aging** – *David Schlessinger (USA)*
10:00-11:00 **From GWAS to function: The example of the Sardinia founder population** – *Francesco Cucca (Italy)*
11:00-11:30 **Coffee break**
11:30-12:30 **Publishing research to make sure society gains from your science** - *Myles Axton (USA)*
12:30-14:00 **Lunch Break**
Adjourn and Departure

Course Organizer: Dr. Andrea Angius   **Summer School Secretary:** Jessica Bazzoli
Please address all correspondence to*: info_sc2018@irgb.cnr.it*

For more information visit: https://www.irgb.cnr.it/summer-school/

7th Sardinian International Summer School

Initiative financed and implemented with the support of Sardegna Ricerche and the Autonomous Region of Sardinia

"From Genome-wide association studies (GWAS) to function"

From July 09th to 13th, 2018

Sardinia Technology Park ,
Pula (CA), Italy

Directors:

Francesco Cucca (Italy)
Marcella Devoto (Italy, USA)
Giovanni Romeo (Italy)

Faculty:

Goncalo Abecasis (USA)
Myles Axton (USA)
Rachael Bashford-Rogers (UK)
Francesco Cucca (Italy)
Anna Di Rienzo (USA)
Daniel Gaffney (UK)
Arthur Gilly (UK)
Kylie James (UK)
Mauro Pala (Italy)
Clelia Peano (Italy)
Stephen Sawcer (UK)
David Schlessinger (USA)
Nicola Segata (Italy)
Carlo Sidore (Italy)
Nicole Soranzo (UK)
John Todd (UK)
Eleftheria Zeggini (UK)

Web site: www.irgb.cnr.it/summerschool

Info: info_sc2018@irgb.cnr.it

Student Presentations

Making sense of the GWAS: colocalization and fine mapping of autoimmunity risk alleles with molecular QTLs
M Tardaguila, K Kundu, S Sawcer, N Soranzo

A comprehensive study reveals heterogeneous genetic landscape in primary and secondary microcephaly
P Boonsawat, Reza Asadollahi, P Joset, B Oneda, Laura Gogoll, S Azzarello-Burri, F Sheth, I Verma, M Zollino, S Passemard, R Bachmann-Gagescu, D Niedrist, M Papik, A Horn, R Masood, M Zweier, D Kraemer, A Verloes, H Sticht, B Plecko, B Latal, O Jenni, K Steindl, A Rauch.

Are Polygenic Risk Scores for Major Psychiatric Disorders associated with general or specific psychosis symptom dimensions?
D Quattrone, P Sham, E Vassos, C Gayer-Anderson, L Ferraro, G Tripoli, The EUGEI team, B Rutten, A Richards, M O'Donovan, J van Os, C Morgan, U Reininghaus, RM Murray, M Di Forti, Lewis C.

Evaluation of skin-related variants in African ancestry populations and their role in personal identification.
V Veltre, A Parisi, F De Angelis, G Biondi, O Rickards.

Poster Sessions

Modelling Mutation Rate of Hepatitis C virus: A Simulation Study
O Adesoji

NGS Data Analysis for the Identification of Rare and Common Variants Associated with Phenotypes of Interest
E Campana, M Cocca, G Girotto, P Gasparini

AIF-1 Gene polymorphisms do not confer susceptibility to Behcet's Disease: analysis of extended haplotypes in Sardinian Population.
MM Angioni, M Piga, F Paladini, S Lai, G Erre, A Floris, A Cauli, G Passiu, C Carcassi, R Sorrentino, A Mathieu

Discovery of URAT1 and GLUT9 Novel variant in hypouricemia subjects using Whole exome Sequencing analysis
S Cho, DH Cha, SK Cho

UHRF1: A Potential Independent Factor In CRC?
MG de Marino, L Muccillo, G Polcaro, M Mancini, V Colantuoni, IM Bonapace.

Network And Pathway-Based Analyses Of GWAS Data To Detect New Associations With Differentiated Thyroid Cancers
O Kulkarni, J Guibon, C Lonjou, P-E Sugier, J-F Deleuze, M-C Boutron-Ruault, C Rubino, A Kesminiene, P Guénel, F De Vathaire, T Truong, F Lesueur

Circulating levels of IL-1 family cytokines and receptors in autoimmune and neurodegenerative diseases
P Italiani, G Della Camera, D Melillo, B Swartzwelter, D Boraschi

A deep learning approach towards detecting positive selection
L Wyss, L Lorenzon, M Fumagalli

Structural Variation Discovery from Whole Genome Sequencing Data
A Kraft, D Plewczyński

Identification of Predictive Biomarkers of Lithium Response in Patients Affected By Bipolar Disorder
E Merkouri Papadima, C Melis, C Pisanu, D Congiu, R Ardau, G Severino, C Chillotti, N Orrù, S Orrù, C Carcassi, S Calza, M Del Zompo, A Squassina

17β-estradiol stimulates reactive species oxygen generation
MH Moghadasi, J Maleki, M Nourbakhsh, M Shabani, M Korani, S Manuchehr Nourazarian, MR Ostadali Dahaghi

Next generation technologies to reveal the molecular basis of complex diseases: the case study of acute lymphoblastic leukemia
K Pane, M Franzese

The eQTLs Catalog and LinDA browser: a platform for prioritising target genes of GWAS variants.
S Onano, F Cucca, M Pala

MicroRNA and mRNA Transcriptome Profiling in Pediatric Multiple Sclerosis
N Nuzziello, F Licciulli, A Consiglio, M Simone, RG Viterbo, G Grillo, S Liuni, M Trojano, M Liguori

Whole-transcriptome profiles of cancer and paired distant normal tissues from Colorectal Cancer Patients
G Pira, L Murgia, P Uva, A Scanu, F Sanges, R Cusano, MR Muroni, P Cossu Rocca, C Carru, A Angius, MR De Miglio.

A mixed-model methodology to correct technical artifacts and enable meta-analysis of sequence based association studies
C Murphy, V Plagnol, D Speed

Community analysis in interactomic regulation networks
D Liberati

Investigating Prostate Cancer Tumorigenesis In A RWPE-1 In Vitro Model Of Combined ERG Over Expression And PTEN Down Regulation
M Zocchi, M Mancini, M Mandruzzato, MG de Marino, C Cicalini, A Mascheroni, A Lunardi, IM Bonapace.

STUDENT PRESENTATIONS

MAKING SENSE OF THE GWAS: COLOCALIZATION AND FINE MAPPING OF AUTOIMMUNITY RISK ALLELES WITH MOLECULAR QTLS

Manuel Tardaguila1, Kousik Kundu1, Stephen Sawcer2, Nicole Soranzo1*,

1Department of Human Genetics, The Wellcome Trust Sanger Institute; 2Department of Clinical Neurosciences, Cambridge University

Resolving the genetic basis of human disease is one of the main challenges of present-day medicine. In the last decade, several hundreds of genomic loci have been robustly associated with susceptibility to autoimmune diseases, predominantly mapping to non-coding regulatory regions of the genome that are active in immune cells, but very few have yielded detailed insights into disease biology. Here we describe an analysis framework to identify and prioritise causal genetic variants and disease genes underpinning associations with fourteen autoimmune diseases e.g. multiple sclerosis (MS), celiac disease, IBD, T1D, and RA, extending published work1. We integrate genetic information with Quantitative Trait Locus (QTL) analyses of molecular phenotypes of gene expression, histone modifications, DNA methylation and transcription factor binding in neutrophils, monocytes and naïve CD4-T cells. Through colocalisation and fine-mapping, we identify putative molecular mechanisms for 346 unique disease loci, and resolve 110 to credible sets of 5 or less causal genetic variants to be assayed in targeted functional experiments. We describe the analytical rationale and results of this large scale effort leveraging high throughput sequencing at population scale, and show how this enhances the functional and mechanistic interpretation of genetic associations in the context of MS. As genetically informed linkage of disease and target gene almost doubles the success of phase II clinical trials2, we anticipate that population genomics-based integrative approaches will be central for target identification and prioritization in drug development pipelines of the omics era.

References
1.Chen, L. et al. Cell 167, 1398–1414.e24 (2016).
2.Cook, D. et al. Nature Reviews Drug Discovery 13, 419–431 (2014).

A COMPREHENSIVE STUDY REVEALS HETEROGENEOUS GENETIC LANDSCAPE IN PRIMARY AND SECONDARY MICROCEPHALY

Paranchai Boonsawat1, Reza Asadollahi1, Pascal Joset1, Beatrice Oneda1, Laura Gogoll1, Silvia Azzarello-Burri1, Frenny Sheth2, Ishwar Verma3, Marcella Zollino4, Sandrine Passemard5, Ruxandra Bachmann-Gagescu1, Dunja Niedrist1, Michael Papik1, Anselm Horn6, Rahim Masood1, Markus Zweier1, Dennis Kraemer1, Alain Verloes5, Heinrich Sticht6, Barbara Plecko7, Bea Latal8, Oskar Jenni8, Katharina Steindl1, and Anita Rauch1

1Institute of Medical Genetics, University of Zurich, Schlieren-Zurich, Switzerland; 2FRIGE's Institute of Human Genetics, Ahmedabad, India; 3Institute of Medical Genetics & Genomics, Sir Ganga Ram Hospital, New Delhi, India; 4Istituto di Genetica Medica, Università Cattolica del Sacro Cuore, Roma, Italy; 5PROTECT, INSERM, Université Paris Diderot, Paris, France; 6Institute of Biochemistry, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany; 7Division of Pediatric Neurology, University Children's Hospital, Zurich, Switzerland; 8Department of Developmental Pediatrics, University Children's Hospital, Zurich, Switzerland.

Microcephaly is an abnormally reduced head size which affects approximately 2–3% of the population worldwide. It is divided into primary (PM) if occurs prenatally or secondary microcephaly (SM) if develops postnatally, and can be caused by environmental or more commonly genetic factors. However, genetic causes are heterogeneous and most cases remain undiagnosed. To elucidate the genetic bases of human microcephaly, we performed a comprehensive study on 61 patients with microcephaly using high-resolution chromosomal microarray analysis and whole-exome sequencing. We determined pathogenic or likely pathogenic variants and therefore diagnosed a variety of genetic disorders in 45.9% of the cohort. Importantly, among these diagnosed patients we observed a recessive inheritance pattern in 70.6% of the patients with PM (n=17) and a de novo occurrence in 83.3% of the patients with SM (n=6), suggesting differential common inheritance between PM and SM. However, we found a comparable distribution of missense and truncating variants among these patients. In addition, we identified 8 high-level candidate genes of various pathways in 8 (13.1%) patients, for which we have provided additional evidence for pathogenicity including in silico predicted effect on protein structure, nonsense-mediated mRNA decay, or additional patients with similar phenotype. Taken together, we demonstrate here a comprehensive genetic landscape of human microcephaly and further propose a set of potentially novel microcephaly genes.

ARE POLYGENIC RISK SCORES FOR MAJOR PSYCHIATRIC DISORDERS ASSOCIATED WITH GENERAL OR SPECIFIC PSYCHOSIS SYMPTOM DIMENSIONS?

Quattrone D1; Sham P2; Vassos E1; Gayer-Anderson C1; Ferraro L3; Tripoli G1; The EUGEI team; Rutten B5; Richards A6; O'Donovan M6; van Os J5; Morgan C1; Reininghaus U5; Murray RM1; Di Forti M1; Lewis C1

1Institute of Psychiatry, Psychology & Neuroscience, King's College London; 2Centre for Genomic Sciences, Li KaShing Faculty of Medicine, The University of Hong Kong; 3Department of Experimental Biomedicine and Clinical Neuroscience, University of Palermo; 4MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff; 5School of Mental Health and Neuroscience, Maastricht University.

Background: Psychotic symptoms can be conceptualised as dimensions of psychopathology cutting across diagnostic boundaries. Thus, they might be considered enhanced quantitative phenotypes to relate to genetic variants as summarised by Polygenic Risk Scores (PRSs) for Major Mental Disorders (MMDs), including Schizophrenia (SZ), Bipolar Disorder (BP), and Major Depressive Disorder (MDD).

The objectives of this study were to: 1) identify the dimensional structure of symptoms at First Episode Psychosis (FEP), testing whether a bi-factor model statistically fits the conceptualization of psychosis as a single common construct (general psychosis factor) while also recognising multidimensionality (positive, negative, disorganized, manic, and depressive symptom factors); 2) examine the extent to which MMD PRSs indexed the phenotypic variance due to the general psychosis construct and to the specific symptom dimensions.

Methods The sample included 1182 FEP patients recruited as part of the EUGEI study. The MRC Sociodemographic Schedule and the OPerational CRITeria (OPCRIT) were used to collect sociodemographic information and assess psychopathology. DNA was extracted from blood or saliva samples collected from 940 participants. The following analysis steps were performed: 1) OPCRIT psychopathology items were analysed using multidimensional item response modelling in Mplus to estimate unidimensional, multidimensional, and bi-factor models of psychosis. Models' fit statistics were compared using Log-Likelihood, and Akaike and Bayesian Information Criteria. 2) SZ, BP, and MDD PRSs were built using the results from large mega-analyses from Working Groups of the Psychiatric Genomics Consortium. In PRSice, individuals' number of risk alleles in the target sample was weighted by the log odds ratio from the discovery samples, and summed into the three PRSs. 3) For the best data fitting psychosis model, linear regressions were estimated to predict symptom dimensions as a continuous outcome from the three PRSs, accounting for population stratification.

Results The best model fit statistics was observed for the bi-factor model including one general and five specific symptom factors compared with the other models. This indicated that there was a broad latent structure underlying the whole range of psychosis symptoms among five latent specific symptom dimensions. PRSs for SZ, BP, and MDD were calculated at the best model fitting P-value threshold. As expected, there was a substantial difference in discrimination of case-control status between SZ PRS and BP and MDD PRSs. A significant positive linear regression equation was observed for SZ PRS and mania dimension severity ($t(864)=2.74$, $p<0.01$), explaining 5% of the variance; whereas a significant negative linear regression equation was found for MDD PRS and the negative dimension severity ($t(864)=-1.75$, $p=0.05$), explaining 3% of the variance. No significant association was found for SZ, BP, or MDD PRSs and the general psychosis trait score.

Discussion These results suggest that a) psychosis at illness onset can be conceptualised as being composed of one general factor and five specific symptom dimensions, b) there is an association between mania dimension score and SZ PRS. Despite the need to both replicate these findings also using PGC new released GWAS to build better powered PRSs, psychosis symptom dimensions have clearly been shown to be a valid and a useful continuous quantitative phenotype across categorical disorders.

EVALUATION OF SKIN-RELATED VARIANTS IN AFRICAN ANCESTRY POPULATIONS AND THEIR ROLE IN PERSONAL IDENTIFICATION.

Virginia Veltre1, Arianna Parisi1, Flavio De Angelis1, Gianfranco Biondi2, Olga Rickards1

1University of Rome "Tor Vergata", Department of Biology, Rome, Italy, 2University of L'Aquila, Department of MESVA, L'Aquila, Italy

Pigment-related genetic variants point out their role in personal identification as they can be considered predictors for Forensic DNA Phenotyping (FDP) and mounting evidence suggest their bio-geographic inferential power to gain information about the individual geographical origin. The current research aims to explore the allelic status in several SNPs mapped in selected genes known to be involved in skin pigmentation: OCA2, HERC2, SLC45A2 and a novel intergenic region between BEND7/PRPF18. The genetic evaluation has been performed on 219 healthy people from African and African derived populations: Fon, Dendi, Bariba and Berba communities from Benin, Tuareg from Libya and Afroecuadorians. The genotypic results have been integrated with the available data from Phase 3-1KGP data release in order to obtain a selected populations panel and the HapMap project YRI, CHB, CEU, and MXL populations were used as an inferential model training set to test the likelihood of correct assignment to geographically differentiated human groups. Data reduction methods and two different classification algorithms based on Bayesian inference have been employed in order to compare the correct assignment likelihood. The proposed panel seems to properly interpret the geographic variation and some interesting evidence could be pointed out in African mixed populations, that seem to be differentially distributed if the total panel is considered. The results support the use of phenotypic inference by molecular information as an auxiliary tool in the personal identification through the use of bio-geographical ancestry information and outwardly visible characteristics such as dark skin tone.

POSTER SESSIONS

MODELLING MUTATION RATE OF HEPATITIS C VIRUS: A SIMULATION STUDY

Oluyomi Adesoji

Cologne Center for Genomics, University of Cologne, Germany

Hepatitis C Virus (HCV) is a frequently mutating virus that causes damages to the liver of humans. The mutation rate of the virus renders previous antibodies produced from exposure to treatments inactive. Therefore, to study the evolution of the virus in a subset of the infected population in Egypt, Shikoun et al. (2013) studied model based approaches for estimating mutation rate of hepatitis C Virus. Hence, the Idea behind this project was to simulate the statistical methods as well as genetic distances approaches employed by these authors. The aim of this project was to build an evolution model that could quantify the mutation rate at NS5B zone of the genotype 4 subtype of HCV between year 2007 and 2010. The HCV sequences were retrieved from "http://www.ncbi.nlm.nih.gov/protein" for the years under study. Multiple sequence alignment was applied as described in the paper and for each year the representative sequences were selected by hidden Markov model (HMM) using the Baulm-Welch Algorithm. Although the assumption of the algorithm is that the length of the model is known but the alternative algorithm implemented by the authors using profile hidden Markov model (pHMM) was not explicitly described. Further, the phylogenetic tree were constructed from the pairwise distance matrices obtained from the Juke-Cantor distance method and reconstructed using the Kimura distance method to account for purine to pyrimidine distances. The estimated difference in distance between the nucleotides over the years with reference to the year of origin (2007) showed that year 2008 has the highest difference. However, the mutation rate obtained by the author is quite different from that obtained in this simulation study. This was somewhat due to the slight differences in methods employed and software used.

Key Words: HMM, Alignment, Phylogenetic tree.

References

Shikoun, N., El Nahas, M., and Kassim, S. (2013). International Journal of Computer Science and Information Security, 11(3):30.

NGS DATA ANALYSIS FOR THE IDENTIFICATION OF RARE AND COMMON VARIANTS ASSOCIATED WITH PHENOTYPES OF INTEREST

E. Campana1, M. Cocca2, G. Girotto1,2, Paolo Gasparini1,2

1University of Trieste, Department of Medicine, Surgery and Health Sciences; 2I.R.C.C.S. Burlo-Garofolo.

Introduction: In recent years the technological advancements of Next Generation Sequencing (NGS) technologies and the drop of per sample sequencing cost led to the generation of big amounts of data. This leads to the requirement of a higher computational effort and the development of new methods for the data Quality Control (QC) and analysis. The aim of this project is to identify rare and common variants in NGS data belonging to different Italian Isolated populations and evaluate the association of detected variants with phenotypes of interest. Particular importance will be given to sequence alignment: for this purpose, we will compare results obtained with the traditional linear reference (BWA) and new nonlinear references based on graph theory [1,2]. Methods: Whole Genome Sequencing (WGS) and Whole Exome Sequencing (WES) data of ~2000 individuals from three isolated Italian populations (Friuli-Venezia-Giulia, Carlantino and Val Borbera) are available for the analyses. Quality control and variant calling will be performed using Samtools, Bcftools and GATK. Preliminary results: For a first set of 378 samples with low coverage WGS data using the linear-reference pipeline, we were able to define a reliable set of ~17M sites, most of them with a minor allele frequency less than 1%. We identified an average of 7.6K singletons per individuals. Among the singletons, the most represented functional categories are stop gained and frameshifts mutations. Conclusions: We expect that these new methods for alignment will increase the reliability of our data, allowing us to call variants with better sensitivity and specificity and improve the quality of the whole analysis.

References:

1: A. Novak et al. Genome Graphs. bioRxiv, 2018.
2: E. Garrison et al.. bioRxiv, 2017.

AIF-1 GENE POLYMORPHISMS DO NOT CONFER SUSCEPTIBILITY TO BEHCET'S DISEASE : ANALYSIS OF EXTENDED HAPLOTYPES IN SARDINIAN POPULATION.

Maria Maddalena Angioni1, Matteo Piga1, Fabiana Paladini2, Sara Lai1, Gianluca Erre3, Alberto Floris1, Alberto Cauli1, Giuseppe Passiu3, Carlo Carcassi4, Rosa Sorrentino2, Alessandro Mathieu1
1Rheumatology Unit, Department of Medical Sciences, University of Cagliari, Italy; 2Department of Biology and Biotechnology, University of Rome Sapienza, Italy; 3 Rheumatology Unit, Azienda Ospedaliera-University of Sassari, Italy; 4Medical Genetics, Department of Medical Sciences and Public Health, University of Cagliari, Cagliari, Italy.

Behçet's disease (BD) is a polygenic immune-mediated disorder characterized by a close association with the HLA-B*51 allele. The HLA region has a strong linkage disequilibrium (LD) and carries several genes variants (e.g. MICA, TNF-alpha) identified as associated to BD because of their LD with HLA-B*51. In fact, the HLA-B*51 is inherited as part of extended HLA haplotypes which are well preserved in patients with BD. Sardinian population is highly differentiated from other Mediterranean populations because of a distinctive genetic structure with very highly preserved HLA haplotypes. In order to identify other genes of susceptibility to BD within the HLA region we investigated the distribution of human Allograft inflammatory factor-1 (AIF-1) gene variants among BD patients and healthy controls from Sardinia. Six (rs2736182; rs2259571; rs2269475; rs2857597; rs13195276; rs4711274) AIF1 single nucleotide polymorphisms (SNPs) and related extended haplotypes have been investigated as well as their LD within the HLA region and with HLA-B51. Overall, 64 BD patients, 38 HLA-B*51 positive healthy controls (HC) and 70 random HC were enrolled in the study. HLA-B*51 was the only gene significantly more expressed (pc = 0.0021) in BD patients (40.6%) than in HC (9.8%). The rs2259571/T AIF-1 variant had a significantly reduced phenotypic, but not allelic, frequency in BD patients (72.1%; pc = 0.014) compared to healthy population (91.3%). That was likely due to the LD between HLA-B*51 and rs2259571/G (pc = 9E-5), even though the rs2259571/G distribution did not significantly differ between BD patients and HC. No significant difference in distribution of AIF1 SNPs haplotypes was observed between BD patients and HC and between HLA-B*51 positive BD patients and HLA-B*51 positive HC. Taken together, these results suggest that polymorphisms of AIF1 are not associated with the susceptibility to BD neither are protective of BD development in Sardinian population.

DISCOVERY OF URAT1 AND GLUT9 NOVEL VARIANT IN HYPOURICEMIA SUBJECTS USING WHOLE EXOME SEQUENCING ANALYSIS

Sunghwa Cho1,Do Hyun Cha2, Sung Kweon Cho2, MD, PhD
1College of pharmacy, Yonsei University, Incheon, Republic of Korea; 2Department of Health Sciences and Technology, SAIHST, Sungkyunkwan University, Seoul, Republic of Korea.

Objective: Renal hypouricemia is rare disorder associated with genetic mutant of renal transporters. Two types are currently reported (Type 1(OMIM: 220150) and Type 2 (OMIM: 612076)). Differentiating between inherited and non-inherited hypouricemia is challenging. Prevalence of hypouricemia is low in Asia. Japanese data reported the difference of it prevalence (0.579% ,West Japanese and 0.191% East Japanese). In this study, we attempted to investigated genetic inheritance of hypouricemia in Korean. Methods: 31 extreme hypouricemia Korean (<1.2mg/dl) were selected from Urban cohort of 179,381 subjects. Other selection criteria are 1) subjects who do not smoke or drink regularly 2) subjects who do not have any underlying conditions such as hypertension, diabetes and taking anti-hypertensive medication. Genetic analysis was performed using whole-exome sequencing. After identifying two SNPs (SLC22A12 c.774G>A (p.Trp258Stop) and c.269G>A (p.Arg90His) explaining 90%), we performed SNaPshot of 2 SNPs for 38 additional hypouricemia subjects. Further whole-exome sequencing was done in 3 unexplained subjects Results: 121 missense variants were determined after filtering out common variant (>1%) in Korean. After we filtered out known genes of renal hypouricemia (SLC22A12 and SLC2A9), 6 unsolved patients were remained. 5 novel variants of SLC22A12 (c.408C>A (p.Asn136Lys), c.674C>A (p.Thr225Lys), c.851G>A (p.Arg284Gln) and c.1253T>G (p.Leu418Arg) and c.1285G>A (p.Glu429Lys) and c.463A>G (p.Met155Val) of SLC2A9 were discovered. Homology modeling confirmed that all newly discovered variants are functionally related to uric acid transport. ASB 12, NEB and LRCH2-RBMXL3 are overlapping genes for 6 unsolved patients. Conclusion: This is the first study to introduce the genetic approach of hypouricemic patients. Screening test of 2 SNPs (p.Trp258Stop and p.Arg90His) is feasible for the future practice preventing acute kidney injury. Further study is needed for the 10% of unexplained area.

UHRF1: A POTENTIAL INDEPENDENT FACTOR IN CRC?

de Marino Maria Giovanna2, Muccillo Livio1, Polcaro Giovanna1, Mancini Monica2,Colantuoni Vittorio1, Bonapace Ian Marc2

1Department of Science and Technologies, University of Sannio, Benevento, Italy; 2Department of Biotechnology and Life Science, University of Insubria, Busto Arsizio, Italy

Ubiquitin-like with PHD finger domains 1 (URHF1) is a modular multi-domain protein required for DNA methylation maintenance and found to be overexpressed in several solid tumours. It is a cofactor that controls epigenetic silencing of many tumour suppressor genes, but its role in CRC remains unclear. We report an in silico analysis using The Cancer Genome Atlas dataset referred to colorectal adenocarcinoma (COADREAD). The TCGA dataset is made up by methylation data acquired by Illumina 450k, expression ones achieved from RNASeq, survival and stage indications.

We record that UHRF1 is a stage-independent factor and is also independent from staging system TNM (except for Lymph Nodes Metastasis index). Furthermore, UHRF1 expression is higher in tumours than in normal samples. Patients were separated into two categories according to UHRF1 overexpression, respectively above (UH) and below (UL) the median value. Surprisingly, survival analysis obtained as Kaplan Meier plots showed that UH had a better prognosis. A combined unsupervised analysis of the CpG islands' DNA methylation from the Illumina 450k and of the RNASeq of the TCGA data set, allowed us to correlate the expression of 83 genes, corresponding to 156 cytosines, with UHRF1 expression. By further classifying the results as a function of UHRF1 expression (UH and UL) and of DNA methylation levels (High - above the median value - and Low - below the median value - methylation), we subdivided patients into 4 groups: UH/LM, UH/HM, UL/LM, UL/HM. While the survival rates of the three UH/LM, UH/HM, UL/LM groups was similar, the prognosis of the UHRF1 Low overexpressed/Hypermethylated group was significantly (p-value =0.008) worst. This analysis enables to define the pivotal role of UHRF1 in methylation control in CRC and to characterize this gene as an independent risk factor. In addition, the study highlights a pool of selected genes that identifies patients with the worst prognosis.

NETWORK AND PATHWAY-BASED ANALYSES OF GWAS DATA TO DETECT NEW ASSOCIATIONS WITH DIFFERENTIATED THYROID CANCERS

O Kulkarni1, J Guibon2, C Lonjou1, P-E Sugier2, J-F Deleuze3, M-C Boutron-Ruault1, C Rubino2, A Kesminiene4, P Guénel2, F De Vathaire2, T Truong2, F Lesueur1

1Inserm U900, Institut Curie, Paris; 2Inserm U1018, CESP, Villejuif; 3CNRGH, Evry; 4Section of Environment and Radiation, IARC, Lyon, France.

Exposure to ionizing radiation and having a familial history of Differentiated Thyroid Cancer (DTC) are two major risk factors for DTC. Recent genome-wide association studies (GWAS) had limited sample size (<690 cases) and few loci were identified so far. Moreover, these studies examined genetic associations with DTC at the individual SNP or gene level, but higher level genetic association analyses using pathway and network-based analyses were not employed in the published datasets. Our goal was to apply such approaches to identify biological pathways and gene networks associated with DTC susceptibility in a multi-ethnic population with contrasted exposures to environmental and genetic factors. To achieved this a GWAS using the OncoArray chip (530,000 SNPs) augmented with over 14,000 custom SNPs known/suspected to be involved in DTC biology was performed in EPITHYR which involved 1,861 cases and 2,321 controls originated from metropolitan France, New Caledonia, French Polynesia, Cuba and the Chernobyl area. The SNP-based analysis with PLINK identified 296 SNPs with $P \leq 5 \times 10^{-8}$, including 256 custom SNPs. Several SNPs outside the well-characterised DTC susceptibility loci at 9q22, 14q13, 2q35 and 8p12 showed suggestive association ($P < 5 \times 10^{-5}$). We next conducted a gene-based analysis using VEGAS2, which confirmed previous findings for several DTCassociated SNPs in FOXE1 (9q22), PTCSC3 (14q13), DIRC3 (2q35) and NRG1 (8p12). Preliminary results from the pathway-based analysis using the Biosystems database indicate that the 'glial cell differentiation' and 'lateral line nerve and system development' GO pathways which have been previously associated with other endocrine-related cancers are enriched for DTC-associated loci. Suggestive association with the 'kinase activity', 'thyroid hormone metabolic process' and 'thyroid hormone generation' pathways was also evidenced. A network-based analysis using topological data from protein-protein interaction networks may also provide a global perspective to further characterize the EPITHYR population.

CIRCULATING LEVELS OF IL-1 FAMILY CYTOKINES AND RECEPTORS IN AUTOIMMUNE AND NEURODEGENERATIVE DISEASES

Paola Italiani, Giacomo Della Camera, Daniela Melillo, Benjamin Swartzwelter, Diana Boraschi
Institute of Protein Biochemistry (IBP), CNR, Naples, Italy

Background: Autoimmune, auto-inflammatory and neurodegenerative diseases are multifactorial, heterogeneous and genetically complex disorders. Although the mechanisms underlying these diseases are not yet understood, it is now recognized that inflammation could play a crucial role in their initiation and progression. Innate immune cells from the myeloid compartment are the main effectors of uncontrolled inflammation that is caused in great extent by the overproduction of inflammatory cytokines, especially those of IL-1 family. The IL-1 family encompasses inflammatory cytokines, anti-inflammatory cytokines, and soluble receptors able to avoid the binding of the inflammatory ligands with the membrane activating receptors. Thus, based on the anomalous activation of innate immune cells, the production of all these factors both locally and systemically may serve as diagnostic markers or therapeutic targets for these diseases. Methods: We have measured the levels of ten different factors of the IL-1/IL-1R family in serum of subjects with Systemic Lupus Erythematosus (SLE), IgG4-Related Diseases (IgG4-RD), and Alzheimer's disease (AD), compared to normal healthy subject. The inflammatory cytokines (IL-1 , IL-1β, IL-18, IL-33), the anti-inflammatory factors (IL-1Ra, IL-18BP), and the soluble receptors (sIL-1R1, sIL-1R2, sIL-1R3, sIL-1R4) were measured in sera by a custom-made multiarray ELISA assay (Quansys Biosciences, Logan, UT). Free IL-18 and active IL-1β were calculated as the amount of IL-18 not inhibited by IL-18BP, and the amount of IL-1β not inhibited by IL-1Ra and sIL-1R2.

Results: All IL-1 family members were measured in serum of inactive and active SLE patients. We confirmed the relevance of IL-18 and IL-18BP in discriminating SLE sera from normal controls. IL- 18 (both total and free) is higher in SLE patients with active disease. When comparing normal healthy subjects to SLE patients, a clear increase of the sIL-1R4 levels was observed. sIL-1R4 was also the main factors measured in serum of IgG4-RD patients. The level of this soluble receptor is higher in patients despite its ligand, the alarmin IL-33, is undetectable in the circulation. sIL-1R1, sIL-1R2, sIL-R3, and sIL-1R4 were measured in serum of subjects with Alzheimer's Disease (AD). All IL-1 family members were evaluated but only regulatory/inhibitory soluble receptors showed variations vs. controls. Three receptors (sIL-1R1, sIL-1R3 and sIL-1R4) were significantly elevated in AD patients. The elevation of sIL-1R4 in AD confirms the increase observed in SLE and IgG4-RD patients, and suggest that sIL-1R4 may be a marker of ongoing inflammation. Conclusions: The analysis of the circulating levels of IL-1 family showed significant increase of almost all soluble receptors (anti-inflammatory effectors) and no variation for the inflammatory cytokines of the IL-1 Family. IL-18 is the only inflammatory cytokines that show a variation in SLE and free IL-18 may have an important role in mediating inflammation in the active stages of disease. Soluble receptors of the IL-1 family may be involved in regulating inflammation not only in autoimmune diseases but also in neurodegenerative diseases, and levated circulating sIL-1R4 levels may represent the marker of an ongoing inflammation in different diseases.

A DEEP LEARNING APPROACH TOWARDS DETECTING POSITIVE SELECTION

Luis Wyss, Lucrezia Lorenzon, Matteo Fumagalli
Fumagalli Lab, Department of Life Sciences, Imperial College London

Although long neglected, recent evidence suggests that positive selection may be a major driver of evolution. However, current methods have difficulties pinpointing positive selection due to its elusive genetic signature. Using a deep learning algorithm, we can detect positive selection in simulated and in real population data. Here, we show a convolutional neural network, that can classify and quantify positive selection in populations with complex population histories. The convolutional neural network operates on genetic data in image format. We have developed a program for easy image creation. We also demonstrate superior robustness compared to support vector machines. This approach can be used to quantify the spread of lactase persistence in Europeans in particular, and in the future will also be applied to a further dataset of Southern American populations. Machine learning, and convolutional neural networks are shown to be a great tool for pattern detection in genetic and genomic data. We are currently applying them to population genetics problems, but they have great potential in any situation where vast amounts of genetic data have to be analysed. We believe that the tools developed by us will find broad applicability in the realm of genome-wide association studies (GWAS), not only due to their strength in detecting complex patterns and analysing large sets of data, but also thanks to their ease of use and simple transfer onto different situations. The approach to analyse genomic data in image format is likely to yield impressive results in the future, as we can currently see with the success of computer vision in non-biological fields.

STRUCTURAL VARIATION DISCOVERY FROM WHOLE GENOME SEQUENCING DATA

Agnieszka Kraft1,2, Dariusz Plewczyński2

1Department of Mathematics, Informatics, Mechanics, University of Warsaw, Poland; 2Centre of New Technologies, University of Warsaw, Poland

Structural variation discovery is a key element in understanding the impact of genome rearrangements on the 3D chromatin structure. Structural variants (SVs) are prone to arise in repetitive regions and can form complex internal structures, therefore variant discovery remain challenging and limited to sequencing data. SVs were first recognized in healthy human individuals, however there is evidence of their association with disease phenotypes - especially CNVs. Recently, disease genome sequencing studies has placed great effort to correlate structural diversity with the occurrence of complex diseases, such as cancer or autoimmune conditions, which aethiology is not well studied. While some variants apparently do not have phenotypic consequences by themselves, carrying a certain allele may predispose to other rearrangements in the same genomic region, in turn leading to a disease. Therefore family duos and trios studies focused on structural variation discovery may be advantageous when looking for the causes of diseases. In our research, we analyse a polish family of four in which one child has type 1 diabetes mellitus. Having provided short-read sequencing data, we use our discovery framework to predict variants carried by family members and analyse their inheritance patterns. We find heterozygous variants in parents whose accumulative effect may lead to the development of disease in T1D-child. Moreover, we determine whether discovered T1D-related variants have impact on the 3D chromatin structure. Based on Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET) data for GM12878 cell line, we analyse which of those candidate variants interrupt chromatin structural units, such as chromatin contact domains (CCDs), CCCTC-binding factor (CTCF) anchors, as well as various classes of genic and intergenic functional elements. Such analysis allows us to find structural differences in chromatin three-dimensional conformation between T1D-child and healthy sibling, enabling to link the phenotype with differences at the sequence level. Our approach can be the basis for future diagnostics and an attempt to explain the basis of many complex diseases.

IDENTIFICATION OF PREDICTIVE BIOMARKERS OF LITHIUM RESPONSE IN PATIENTS AFFECTED BY BIPOLAR DISORDER

Eleni Merkouri Papadima1, Carla Melis1, Claudia Pisanu1, Donatella Congiu1, Raffaella Ardau2, Giovanni Severino1, Caterina Chillotti2, Nicola Orrù2, Sandro Orrù,3, Carlo Carcassi,3, Stefano Calza4, Maria Del Zompo1,2, Alessio Squassina1

1Laboratory of Pharmacogenomics, Department of Biomedical Sciences, University of Cagliari, Cagliari; 2Unit of Clinical Pharmacology, 'San Giovanni di Dio' Hospital, Cagliari; 2S.C Medical Genetics, Binaghi Hospital, ASL Cagliari, Cagliari; 3Department of Medical Sciences, University of Cagliari, Cagliari; 4Biostatistics Unit, Department of Molecular and Translational Medicine, University of Brescia, Brescia.

Lymphoblastoid cells line (LCLs) derived from samples of bipolar disorder patients, Full Responder (FR) and Non-Responder (NR) to chronic lithium treatment were used for evaluating their miRNAs' expression profile through a Next Generation Sequencing (NGS) approach.

The sample consisted of LCLs derived from 24 bipolar patients, Sardinians for four generations, 12 NR and 12 FR to lithium treatment, according the "Retrospective Criteria of Long-Term Treatment Response in Research Subjects with Bipolar Disorder" and part of the Consortium on Lithium Genetics (ConLiGen). The LCLs from both groups were cultured either in presence or in the absence of 1mM LiCl, total RNA was extracted enriched in miRNAs and sequenced with MiSeq instrument (Illumina) and the reads mapped using miRBase database.

A comparative analysis between FR vs NR groups was performed and the effect of lithium treatment in vitro on the miRNA expression was estimated. Available transcriptomic data from the same cohort were used for a correlation analysis between the higher hits of the two datasets. Subsequently, target prediction with online miRNA target prediction software helped to narrow down and select miRNAs and targets for validation. The selected miRNAs and mRNAs were validated with qRT – PCR.

Two out of four selected miRNAs, miR-320a and miR-155-3p, and three of the corresponding mRNA targets (two targets of miR–320a and one of miR-155-3p) were validated as significantly differentially expressed. The three resulting miRNA – mRNA couples will undergo functional analysis using independent cell lines and miRNA mimics technology.

17B-ESTRADIOL STIMULATES REACTIVE SPECIES OXYGEN GENERATION

Mohamad Hosein Moghadasi1, Jafar Maleki2, Mitra Nourbakhsh2,3, Mohammad shabani*2, Mohsen Korani4, Seyed Manuchehr Nourazarian5, Mohammad reza Ostadali Dahaghi6

1Department of Hematology. Faculty of Medical Science,Tarbiat Modares University, Tehran, Iran; 2Department of Biochemistry, School of medicine, Iran University of medical sciences, Tehran, Iran; 3Metabolic Disorders Research Center, Endocrinology and Metabolism Research Institute, Tehran University of Medical Sciences, Tehran, Iran; 4Department of Biochemistry, School of medicine, Baqiyatallah University of medical sciences, Tehran, Iran; 5Department of laboratory sciences, Faculty of Paramedical sciences, Tabriz University of Medical sciences, Tabriz, Iran; 6Hematology- Oncology and Stem Cell Transplantation Research Center, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran.

Background: Experimental and epidemiological evidence supports a role for sex steroid hormones in the pathogenesis of ovarian cancer. Among steroid hormones, 17β –estradiol (E2) has the most potent effect on proliferation, apoptosis and metastasis.

Methods and materials: Ovarian adenocarcinoma cell line (OVCAR-3) was cultured and treated with various concentrations of E2, antioxidants (N-acetyle cysteine and Ebselen) and ICI182780 as an estrogen receptor antagonist. MTT test was performed to evaluate cell viability. NO and ROS levels were measured by Griess and DCFH-DA methods respectively.

Results: ROS levels as well as NO levels were increased in OVCAR-3 cells treated with E2. The increase in ROS production was in parallel with increased cell viability which indicates that estrogen-induced ROS can participate in cancer progression. ICI182780 abolished E2-induced ROS production. Progesterone was also effective in reducing ROS and NO generation.

Conclusions: NO and ROS are important molecules in signaling networks in cell. These molecules can be used as therapeutic targets for prevention and treatment of ovary cancer and other estrogen-induced malignancies.

NEXT GENERATION TECHNOLOGIES TO REVEAL THE MOLECULAR BASIS OF COMPLEX DISEASES: THE CASE STUDY OF ACUTE LYMPHOBLASTIC LEUKEMIA

Katia Pane, Monica Franzese

IRCCS SDN, Via E. Gianturco, 113, 80143 Naples, Italy

Acute lymphoblastic leukemia (ALL) is a malignant disorder originating from hematopoietic B- or T-cell precursors and is characterized by marked heterogeneity at the molecular and clinical levels. B-ALL and T-ALL comprise multiple subtypes defined by their primary chromosomal abnormality (mainly chromosomal translocations that give rise to chimeric fusion genes or broad aneuploidy) and defined by the cooperating secondary aberrations (deletions, amplifications, sequence mutations, and epigenetic lesions), which jointly contribute to leukemogenesis [1,2]. Recently, next-generation sequencing (NGS) technologies have identified many novel lesions in ALL and helped not only improve our understanding of its pathogenesis, but also helped to discover key biomarkers of diagnostic and prognostic importance [3]. The concept of NGS involves DNA, RNA, or miRNA sequencing through various approaches that coupled together create a multiassay approach that more likely can describe the complexity of events ongoing into the malignant cell [3].

The goal of this research project is to characterize the transcriptome landscape of patients with B-ALL using high throughput RNA-sequencing (RNA-seq) analysis compared to healthy blood donors and furtherly reveal the potentially related epigenomic alterations occurring in tumor cells.

We firstly carried out the analysis of transcriptional patterns between 3 B-ALL patients and 3 healthy blood donors. RNA-seq differentially expressed genes (DE genes),were identified and used for functional analysis by using Ingenuity pathway analysis and free web tools.

Interestingly, we found out a meaningful dysregulation of so far poorly characterized protein family. By using "wet-lab" approach we experimentally validate their occurrence, cellular localization and concentrations between B-ALL patients and healthy blood donors. Overall RNA-Seq functional analysis suggested their implications in signaling pathways related to induction of proliferation and cytoskeleton rearrangement. As future perspective we will evaluate the alterations occurring at chromatin level for the most aberrantly expressed genes by inferring their methylation profile output into our RNA-seq analysis.

References

[1] Iacobucci I, Mulligan CG.. J Clin Oncol. 2017 Mar 20; 35(9):975-983.
[2] Moorman AV Haematologica. 2016 Apr; 101(4):407-16.
[3] Montaño A et al., Cancers (Basel). 2018 Apr 7;10(4). pii: E110. doi: 10.3390/cancers10040110.

THE EQTLS CATALOG AND LINDA BROWSER: A PLATFORM FOR PRIORITISING TARGET GENES OF GWAS VARIANTS.

Stefano Onano1,2,3, Francesco Cucca1,2, Mauro Pala2

1Dipartimento di Scienze Biomediche, Università degli Studi di Sassari, Italy; 2Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Monserrato, Italy; 3The Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor (MI), USA.

The expression Quantitative Traits Loci (eQTLs) are genetic polymorphisms associated with changes in gene expression levels. They have been successfully used to pioritize the target genes of the variants associated with complex traits and diseases (GWAS variants). Up to date a few eQTLs databases exist and they collect only a small portion of the available datasets. We thus planned to build the largest publically available catalog of eQTLs, coupled with a browser, to optimize and simplify their interrogation. We collected and manually curated 51 eQTL public studies ranging from 2007 to date, corresponding to more than 95 sample types and 24 human populations for a total of 282719 cis-eQTLs and 33368 genes with at least one cis-eQTL (cis-eGenes). Most of the eQTLs studies were conducted in blood samples from healthy individuals of European ancestry.

We found that for 93% of the known protein-coding genes were eGenes, 20% of them intersecting ($r2{\geq}0.8$) with the NHGRI-EBI GWAS Catalog and 26% of whom considered as druggable. Futhermore, for those GWAS variants for which an eGene was known, we found that the NHGRI-EBI GWAS Catalog proposed an eQTL gene as candidate target only for the 70% of the times. Our eQTL-Catalog can be used as a reference to measure the degree of novelty for future eQTLs studies; it is provided within a platform with a web interface (LinDA) that we plan to implement with other types of quantitative traits (i.e. epigenetic, proteomic, metabolomics and microbiota) to better dissect the pleiotropy of the GWAS variants.

MICRORNA AND MRNA TRANSCRIPTOME PROFILING IN PEDIATRIC MULTIPLE SCLEROSIS

Nicoletta Nuzziello1, Flavio Licciulli1, Arianna Consiglio1, Marta Simone2, Rosa Gemma Viterbo2, Giorgio Grillo1, Sabino Liuni1, Maria Trojano2, Maria Liguori1

1National Research Council of Italy, Department of Biomedicine, Institute of Biomedical Technologies, Bari Section, 70125 Bari, Italy; 2Department of Basic Sciences, Neurosciences and Sense Organs, University of Bari, 70125 Bari, Italy

Multiple sclerosis (MS) is a chronic autoimmune disease of the central nervous system (CNS) that primarily affects young adults, although approximately 3-10% of all MS patients complain of the first symptom/s during childhood and adolescence (so called pediatric MS, PedMS). Susceptibility to develop MS, in pediatric as in adult ages, is determined by genetic, epigenetic and environmental factors.

By using a High-Throughput Next-generation Sequencing (HT-NGS) approach, miRNA and target mRNA profiles have been characterized in the peripheral blood of 19 PedMS patients and compared to 20 pediatric controls (PCs). An integrated bioinformatics and biostatistics analysis revealed 12 significantly upregulated miRNAs (miR-125a-5p, let-7b-5p, miR-942-5p, miR-221-3p, miR-652-3p, miR-182-5p, miR-185-5p, miR-181a-5p, let-7a-5p, miR-25-3p, miR-320a and miR-99b-5p), 1 significantly downregulated miRNA (miR-148b-3p) and a plethora of differential expressed (DE) target genes, in PedMS patients compared to PCs.

In addition, according to the transcriptional regulatory rule, a transcription factor (TF)-miRNA co-regulatory network was constructed, composed by regulatory relationships between TFs and DE miRNAs, TFs and DE target genes, and miRNAs and their targets. An enrichment analysis categorized in functional pathways associated with PedMS was also investigated, mostly related to immune system, oxidative stress and response to lipids.

In conclusion, this integrated analysis of miRNA and mRNA expression profiles enables to identify possible molecular signatures of PedMS, allowing to shed light in the pathogenesis of this multifactorial disease and adding further insights in the genetic background of PedMS.

WHOLE-TRANSCRIPTOME PROFILES OF CANCER AND PAIRED DISTANT NORMAL TISSUES FROM COLORECTAL CANCER PATIENTS

Giovanna Pira1, Luciano Murgia2, Paolo Uva3, Antonio Scanu2, Francesca Sanges2, Roberto Cusano3, Maria Rosaria Muroni2, Paolo Cossu Rocca2, Ciriaco Carru1, Andrea Angius4, Maria Rosaria De Miglio2. 1Department of Biomedical Sciences, University of Sassari, Sassari, Italy; 2Department of Clinical and Experimental Medicine, University of Sassari, Sassari, Italy; 3CRS4, Science and Technology Park Polaris, Piscina Manna, Pula, CA, Italy; 4Istituto di Ricerca Genetica e Biomedica (IRGB), CNR, Monserrato (CA), Italy.

Colorectal cancer (CRC) is one of the most frequent malignant tumor and the commonest cause of cancer death worldwide. The identification of specific biomarkers should be of great benefit for early diagnosis and develop of new targeted therapies to decrease the CRC mortality. Next-Generation Sequencing (NGS) techniques give us the complete genomic structure of neoplastic tissue and permits the identification of changes in the tumor pathogenesis. In this study, we performed high-throughput transcriptome sequencing on CRC and normal colon tissue (NCT). Total RNAs were extracted from 16 paired primary CRC and NCT. Whole-transcriptome analysis was performed using the Illumina TruSeq Stranded mRNA Library Prep Kit and the Illumina HiSeq3000. The RNAseq data were analyzed using the TopHat and Cufflinks protocols using GRCh37/hg19 as a reference. The transcriptome analysis revealed cancer-specific differentially expressed genes (DEGs) and differential alternative splicing. A total of 1378 DEGs were identified in CRC: 611 and 767 were significantly up and downregulated, respectively. Gene Ontology analysis revealed that CRC overexpressed DEGs were enriched in pathways involved in the cell cycle checkpoint, E2F transcription factor network, DNA damage response, WNT/beta-Catenin. While CRC downregulated DEGs affect Respiratory electron transport, Mitochondrial Fatty Acid beta-Oxidation, Phase II conjugation, Cytokine Signaling in Immune system. 12684 genes were found mutated. KRAS and NRAS mutations were identified in 56% of CRC. Interestingly, mutations were identified in BRAF, TP53, PTEN, SMAD4 and in FOX-O, ERB-B, AKT1, EGFR, CDKN1A genes, whose proteins are known members of pathways such as colorectal cancer, miRNAs in cancer and PI3K/AKT, respectively. RNA-sequencing technology revealed the variation landscape of CRC transcriptome. Our data raise the knowledge of the expression differences that underlying malignancy and revealing useful genes that may be used as diagnostic or prognostic markers.

A MIXED-MODEL METHODOLOGY TO CORRECT TECHNICAL ARTIFACTS AND ENABLE META-ANALYSIS OF SEQUENCE BASED ASSOCIATION STUDIES

Cian Murphy, Vincent Plagnol, Doug Speed
UCL Genetics Institute, University College London, UK.

High throughput DNA sequencing technologies, either whole-exome (WES) or whole-genome (WGS) sequencing are revolutionizing the diagnosis and novel gene discovery for rare disorders. As the field transitions from the early discovery for Mendelian to more complex diseases, there is substantial benefit in being able to combine data across studies, performing the type of meta-analysis for cases and controls that have proven to be so successful for genome-wide association studies (GWAS). However, WGS and WES are substantially more affected by sequencing errors and technical artifacts than genotyping arrays. As a consequence, meta-analysis of sequence based association studies are often dominated by spurious associations.

Here, we show that it is possible to take advantage of the type of mixed models developed initially to control for population structure in GWAS and apply these ideas to control for technical artifacts. Using a dataset of 5000 WES we demonstrate that substantial reduction in the association statistic-inflation can be achieved by applying these novel analytical techniques while preserving the sensitivity of the test. We focus on a subset of the phenotypes in this dataset to illustrate the ability of these novel methods to produce more interpretable results.

COMMUNITY ANALYSIS IN INTERACTOMIC REGULATION NETWORKS

Diego Liberati
National Research Council of Italy, Milano

The powerful though simplistic idea that every gene just codifies for its own protein, then free to interact within the cell, is nowadays over ever since. It would not account for most of the well- known epigenetic properties, modulating gene expression to the different context within, for instance, different organs, even in the same individual. Networks of interactomic actors, including genes and codified proteins themselves, are nowadays known, though not yet all completely understood, as responsible for the exhibited complex para-social interaction resulting in the beautiful diversity, within similarity, of key factors of life The same happened many years ago at a different scale within the central nervous system neuroscience, when the so-called "grand-mother" neuron - formerly believed to be responsible for memorizing the beloved - was then substituted by the task-recruited neural network including several actors, each of which also in turn still available to contribute to other tasks within (partially) different other networks. Still the same, at an even bigger scale, is everyday everybody's social multi-interactions experience of each homo "oeconomicus" of us. In general, such – ubiquitous, one would say - networks are not fully connected, yielding a case not directly tractable by many of the most diffused algorithms developed to analyze them irrespectively of the very chemo-physical nature of their nodes and arc interactions. In order to overcome such a first drawback, a so-called damping is usually introduced, making the network "artificially" fully connected, but with weights small enough associated to the artificially created arcs, in order to be able to consider the corrected network enjoying more or less the same properties of the original one. The investigated connected network is thus represented, as a first approximation, by a non oriented graph, where every arc does represent the possible reciprocal – thus symmetrical - influence of every protein, or in general actor – one of the two nodes connected by an arc could also be a gene – on every other. In such a framework, it is straightforward that a community is a kind of cluster of a subset of the whole interactomic network whose internal global interconnection is in some sense prevalent with respect to less important connections still existing among some nodes of such cluster and some other nodes not belonging to it. A precise specification of the above - quite fuzzy indeed - concept yields to the different definitions and techniques to identify communities proposed in literature, to some of which we shall be referring within the present contribution In order to investigate such a kind of networks, in fact, a simple but powerful idea, as proved by Google usefulness and consequential success, is to investigate the ranking of each actor relationships to each other, being such actor either an internet page, as in the original Page algorithm, or a gene - or codified protein - in our case. It is worth noticing that the recent randomized approach introduced by the prematurely late Roberto Tempo and coworkers [1] could be the technological key to drastically reduce, at the cheap price of a limited loss of precision, the over-helming computational complexity that would prevent to apply Page ranking to the analysis of every significant interactomic network, besides the almost-toy sub-networks already investigated mostly as a proof-of-concept, as for instance in Zaki and coworkers [2] In this contribution, we would like to investigate a possible complementary approach, recently proposed by Landi & Picardy [3], to our interactomic regulation network. As a public available benchmark, the data used in the reported work by Zaki [2] are also used, in order to investigate which features of our proposed approach are possibly useful as a complement to even improve the already powerful Page approach on one side, or eventually able to surrogate it in a less costly way, even taking into account the recalled randomized economy

References
[1]H.Ishii and R.Tempo, IEEE Trans. Auto. Control, 55(9), 1897, 2010.
[2]E.Hanna and N.Zaki, BMC Bioinformatics 15: 204, 2014
[3]P.Landi and C.Piccardi, Phys Rev E Stat Nonlin Soft Matter Phys. 89(1):012814, 2014

INVESTIGATING PROSTATE CANCER TUMORIGENESIS IN A RWPE-1 IN VITRO MODEL OF COMBINED ERG OVER EXPRESSION AND PTEN DOWN REGULATION

Zocchi Michele1, Mancini Monica1, Mandruzzato Martina1, de Marino Maria Giovanna1, Cicalini Claudia1, Mascheroni Arianna1, Lunardi Andrea2 and Bonapace Ian Marc1.
1DBSV, University of Insubria, Busto Arsizio, Italy; 2CIBIO, University of Trento, Trento, Italy.

Background and Rational: Prostate cancer (PCa) progression is largely dependent on epigenetic mechanisms, including concurrent global DNA hypomethylation and site-specific gene hypermethylation. In line with this, we have previously demonstrated that DNMT3A is essential for the promotion of PCa progression and EMT process activation. To better investigate tumour progression, together with Dr. Lunardi we have set up a PCa in vitro model based on the immortalized human epithelial prostate cell line RWPE-1, genetically engineered to simulate TRMPSS2/ERG over-expression and PTEN down-regulation, which represents a good cellular model to reproduce PCa onset and evolution.

Aim: The aim of the project is the identification of new methylation-dependent regulated genes involved in PCa progression and deregulated by the combination of ERG over-expression and PTEN down-regulation in RWPE-1 PCa cell model.

Results and Discussion: Dr. Lunardi has kindly provided us with the RWPE-1 cells engineered with a panel of doxycycline-based inducible vectors to mimic ERG over-expression alone or in combination with partial or total PTEN downregulation. Doxycycline treatment nicely induces ERG over-expression, while progressive PTEN downregulation inversely correlates with PI3K/AKT signalling. Given this premise, we will carry out the silencing of DNA methyltransferase enzymes, before and after ERG/PTEN deregulation and subsequently perform RNA-sequencing for coding and non-coding RNAs and DNA methylation analysis with the Illumina 850K MethylationEPIC BeadChip, to assess the differential regulation of gene expression and DNA methylation during the process. The combined analyses will allow identifying transcripts dependent or not on DNA methylation changes, induced or repressed by ERG and PTEN modulation. These will be candidate key drivers of tumour progression and will be verified by the analysis of TCGA-PRAD data sets.

With the support of

REGIONE AUTONOMA DELLA SARDEGNA

SARDEGNA
RICERCHE

EUROPEAN SCHOOL OF GENETIC MEDICINE