EMBO | EMBL Symposia

Multiomics to Mechanisms - Challenges in Data Integration. September 11th – 13th 2019
European Molecular Biology Laboratory. Heidelberg, Germany

UNIVERSITAT JAUME I
VNIVERSITAT DE VALÈNCIA
Facultat de Medicina i Odontologia
Departament de Medicina Preventiva i Salut Pública, Ciències de la Infermeria, Fisioteràpia i Medicina Legal

ciberobn
GENERALITAT VALENCIANA

Crosstalk between smoking and the genome in older subjects with metabolic syndrome through genomics, epigenomics and transcriptomics

Oscar Coltell[1,2], Eva M Asensio[2,3], Carolina Ortega-Azorín[2,3], Rocio Barragán-Arnal[2,3], Carmen Saiz[2,3], Olga Portoles[2,3], Jose V. Sorlí[2,3], Dolores Corella[2,3]

[1]: Department of Computer Languages and Systems, University Jaume I, 12071 Castellón, Spain. [2]: CIBEROBN. Instituto de Salud Carlos III, 28029 Madrid, Spain. [3]: Department of Preventive Medicine, University of Valencia, Valencia, 46010 Valencia, Spain.

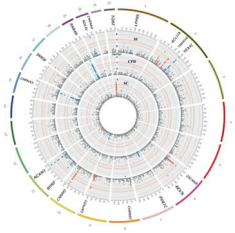
Background and aims

Tobacco smoking (Figure 1) is a major cause of cardiovascular diseases (CVD), and appears to have a multiplicative interaction with the other major CVD risk factors (lipids, hypertension, diabetes and others present in the metabolic syndrome (MetS). Several omics have analyzed the separate effects of tobacco smoking on the genome, epigenome, transcriptome, metabolome, etc. However an integrated omics approach can help to better understand the crosstalk between tobacco smoking and the genome.

Figure 1: Tobacco smoking)

Several GWAS have analyzed genetic polymorphisms associated with a higher risk of smoking, among them, the work carried out by Erzurumluoglu et al (Molecular Psychiatry, 2018), in more than 600,000 participants, identifying 40 new loci associated with smoking behavior (Figure 2), however, the Mediterranean population has been scarcely studied.

Figure 2: GWAS for tobacco smoking (Erzurumluoglu et al, Molecular Psychiatry, 2018)

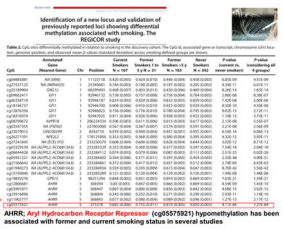


Likewise, several epigenome-wide association studies (EWAS), have consistently identified several differentially methylated loci associated with tobacco smoking (Figure 3).

Figure 3

Several individual studies and meta-analyses in different populations have reported a strong hypomethylation signal in the CpG cg05575921 locus depending on the smoking status . Among these studies, we show specific data in the REGICOR study, a population study carried out in the Mediterranean population (Figure 4), as well as the participants in the MESA study (Figure 5).

Figure 4: Methylation of the AHRR gene and other genes depending on the smoking status in the REGICOR study (Epigenetics, 2015).



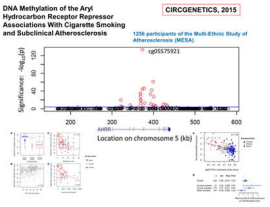


Figure 5: EWAS for methylation status depending on smoking in the MESA participants (Circulation Cardiovascular Genetics, 2015).

Another consistent association has been found with coagulation factor II (thrombin) receptor-like 3 (F2RL3, cg03636183). Hypomethylation of this gene has been associated with smoking in several studies and in different populations mimicking the AHRR effects (Figure 6 and Figure 7).

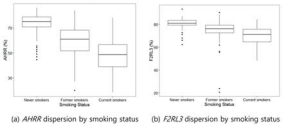
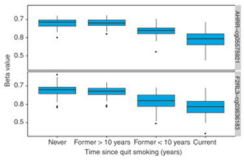


Figure 6: Hypomethylation of the AHRR and F2RL3 locus by smoking (Lee et al, PlosOne, 2017).

Figure 7: Hypomethylation of the AHRR and F2RL3 locus by smoking (Fasanelli et al, Nat Comunications, 2015).



In addition to these consistent loci, other methylated loci have been associated with tobacco smoking depending on the population analyzed. On the other hand, some transcriptomic-wide studies (TWAS) have identified several gene differentially expressed in blood from smokers vs non-smokers. Among these meta-analysis, Huan et al (Hum Mol Genet, 2016) analyzing 6 cohorts identified several differentially expressed genes including the LRRN3 (Leucine Rich Repeat Neuronal 3) (Figure 8).

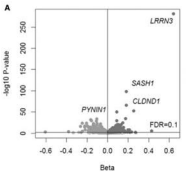


Figure 8: Differentially expressed genes in current smokers vs never smokers (Huan et al, 2016).

However, these studies have used isolated omics, and a multi-omics integration for tobacco smoking in needed. Therefore, our aim was to undertake a multi-omics analysis to better understand the genetic factors associated with current smoking in older subjects with MetS, as well as the effect of smoking on DNA methylation and on gene expression.

Methods

We carried out a study using genomics, epigenomics and transcriptomics (Figure 9) in subjects with MetS from the Mediterranean population.

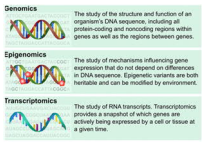


Figure 9: Multi-omics used

We have analyzed participants recruited from the PREDIMED Plus-Valencia study (located on the eastern Mediterranean coast, Spain). Eligible participants, recruited from several primary care health facilities in the Valencia field center, were community-dwelling adults (men, 55–75 years, women, 60–75 years) with a body-mass index (BMI) in the overweight or obesity range ((BMI) ≥ 27 and <40 kg/m2) and had at least three components of the metabolic syndrome. randomized participants included in the PREDIMED Plus trial. DNA and RNA were isolated from blood. The GWAS for tobacco smoking was carried out in all the participants with DNA available (n=450), EWAS and TWAS were carried out in a subsample of 88 participants having this analysis available. Anthropometric variables as well as tobacco smoking and other lifestyle variables were determined by trained staff and follow the PREDIMED Plus operations protocol detailed in the study Web site (http://www.predimedplus.com/). Smoking status was assessed according to the WHO classification (including current smokers, former smokers (including former smokers <1 year, from 1 to 5 years and more than 5 years) and never smokers. These subjects were aged 55 to 75 years, mean age 65+/-5 years, all with metabolic syndrome.

1 Genomic analyses and GWAS

DNA was isolated from blood (buffy-coats), and high-density genotyping was performed at the University of Valencia using the Infinium OmniExpress-24 v1.2 BeadChip genotyping array (Illumina Inc., San Diego, CA, USA), according to the manufacturer's protocol with appropriate quality standards. This array captures 713,599 markers. Allele detection and genotype calling were performed in the GenomeStudio genotyping module (Illumina, Inc., San Diego, CA, USA). Data cleaning was performed using standard analysis pipelines implemented in the Phyton programing language using the Numpy library modules combined with the PLINK. From the initial full set, those SNPs not mapped on autosomal chromosomes were filtered out. In addition, SNPs with a minor allele frequency (MAF) < 0.01 or those that deviated from expected Hardy-Weinberg equilibrium ($p < 1.0 \times 10-5$) were removed.

For GWAS, genetic association analyses were performed using PLINK v1.9 using and additive model. Subjects were re-classified into current smokers and non-smokers. Coefficients for the minor allele were estimated. Unadjusted and adjusted (for sex and age or for additional variables) general linear models were fitted. We used the conventional threshold of $p < 5 \times 10-8$ for genome-wide statistical significance. Since this threshold is very conservative for a small sample size, SNPs with p-values below $1 \times 10-5$ were also considered suggestive of genome-wide significance. SNPs were rank-ordered according to the minimum p-value in the genetic models.

2. RNA isolation and TWAs

RNA was isolated from fresh blood at baseline in 88 subjects (44 males and 44 females) with the Promega device and kit. RNA integrity was assayed by means of the 2100 Bioanalyzer with Eukaryote total RNA Nano Assay (Agilent Technologies, Santa Clara, CA, USA (Figure 10)). RNA integrity number (RIN) served as RNA integrity parameter (selection criteria RIN ≥ 9.0).Microarray experiments were performed at the Central Research Unit (University of Valencia). GeneChip Human Gene 2.0 ST Array containing over 41,000 transcripts and represent over 36,000 well-characterized human genes (Affymetrix, Santa Clara, CA, USA) was used for microarray analysis. GeneChip Operating Software supplied by Affymetrix was used to generate.CEL files. Partek Genomic suite was used for gene expression analysis. Multivariate models were used to adjust for potential confounders including batch effect, sex, age, leukocyte counts, etc.

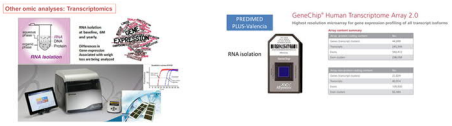


Figure 10: Transcriptomic analysis

3. Epigenome-wide methylation analysis

We performed genome-wide DNA methylation (EWAS) in the same 88 subjects using the Illumina Infinium 850K MethylationEPIC array (Figure 11). Differential methylation (M-values) was statistically analyzed with Partek Genomic Suite using ANCOVA models adjusted for potential confounders including batch effect, age, BMI, diabetes and leukocyte counts, among others.

Figure 11: Methylation analysis

Results

Table 1 shows the main characteristics of the Predimed Plus-Valencia study participants at baseline. Prevalence of current smokers was 11.4%.

Table 1: General characteristics of the PREDIMED Plus-Valencia participants at baseline

	Total (n=465)	Men (n=186)	Women (n=267)	P
Age (years)	65.1 ± 0.2	63.9 ± 0.4	66.1 ± 0.3	<0.001
Weight (kg)	84.5 ± 0.7	92.8 ± 1.0	78.0 ± 0.6	<0.001
BMI (kg/m2)	32.4 ± 0.2	32.3 ± 0.2	32.5 ± 0.2	0.029
Waist circumference (cm)	106.1 ± 0.5	111.2 ± 0.6	102.0 ± 0.6	<0.001
SBP (mm Hg)	141.6 ± 0.9	143.2 ± 1.3	136.9 ± 1.2	0.025
DBP (mm Hg)	81.9 ± 0.5	82.9 ± 0.7	79.7 ± 0.6	0.002
Total cholesterol (mg/dL)	199.4 ± 1.6	198.3 ± 2.8	202.6 ± 2.3	<0.001
LDL-C (mg/dL)	125.9 ± 1.5	121.6 ± 2.4	127.7 ± 1.9	0.044
HDL-C (mg/dL)	51.5 ± 0.5	47.5 ± 0.8	54.7 ± 0.7	<0.001
Triglycerides (mg/dL)	141.8 ± 2.9	138.2 ± 3.8	144.3 ± 4.2	0.295
Fasting glucose (mg/dL)	102.8 ± 1.3	102.8 ± 2.0	102.3 ± 1.7	0.802

A GWAS for current smoking was carried out on these subjects. Among the top-ranked SNPs (at P<1x10E-5 in a model adjusted for age and sex) (Table 2) we found the rs7942698-SLC1A2 (in a transporter essential for clearing glutamate from neuronal synapses) and the rs776404-NRG1 (Neuregulin signaling pathway), both previously linked to addictions and smoking.

P	SNP	ALLELES	MAF	CHR	GENE
4,727E-06	rs7942698	A	0,423323	11	SLC1A2
5,149E-06	rs2297072	G	0,164137	13	STARD13
5,331E-06	rs4756214	T	0,399361	11	SLC1A2
8,245E-06	rs776404	G	0,319888	8	NRG1
9,878E-06	rs2505470	C	0,058906	10	__
0,00001114	rs277273	G	0,282548	13	__
0,00001367	rs1103307	A	0,346845	11	SLC1A2

Table 2: Top-ranked SNPs for tobacco smoking (additive model adjusted for sex and age)

In the EWAS the top-ranked differentially methylated CpG sites were: cg05575921 in the AHRR (aryl hydrocarbon receptor repressor) gene ($P=2.73E-14$), and the cg03636183 in the F2RL3 (coagulation factor II (thrombin) receptor-like 3) gene ($P=6.90E-09$), replicating previous finding of the hypomethylation of the AHRR and F2RL3 sites in smokers.
In the TWAS, we identified the GPR15 gene as the most differentially expressed between smokers and non-smokers (6.43E-19). Likewise, the LRRN3 gene ($P=1.41E-08$) was another differentially expressed gene, in agreement with previous findings. For the LRRN3 gene we detected very consistent results both in the TWAS and in the EWAS analysis. There are preliminary results in a pilot study and more work has to be done, also increasing sample size.
Future research will focus on omics-integration analyses, as well as on analysis de modulation of the effects by other environmental factors such as diet, physical activity, etc.

Acknowledgements

This study was partially funded, by the Spanish Ministry of Health (Instituto de Salud Carlos III) and the Ministerio de Economía y Competitividad-Fondo Europeo de Desarrollo Regional (FEDER, "A way of making Europe") (grants CIBER 06/03, PRX17/00500, SAF2016–80532-R) and Premio Rei Jaume I (Valencia, Spain); the University Jaume I (grants P1–1B2013–54, COGRUP/2016/06 and UJI-B2018-69); the Fundació La Marató de TV3 (grant 538/U/2016); and the Generalitat Valenciana (grant PROMETEO2017/017).