## SUPPLEMENTAL DIGITAL CONTENT

**Epigenetic Modifications in the Biology of Nonalcoholic Fatty Liver Disease: The Role of DNA-Hydroxymethylation and TET Proteins**

**Authors:** Carlos J Pirola PhD, Romina Scian PhD, Tomas Fernández Gianotti PhD, Hernán Dopazo PhD, Cristian Rohr B.Sc., Julio San Martino MD, Gustavo O Castaño MD, PhD, Silvia Sookoian MD, PhD

**Supplemental Figure F1**

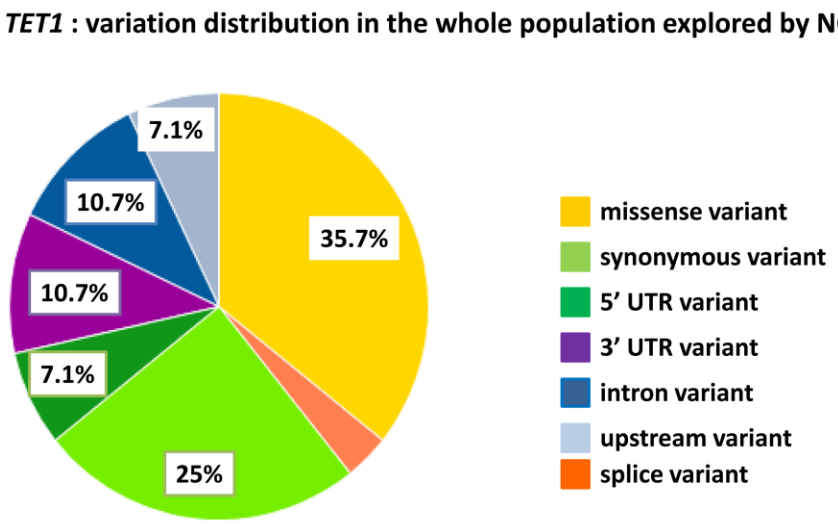


**Figure illustrates genetic variation of *TET1* in the whole population explored by next-generation sequencing: distribution of the predicted function of the variants.**
Information of variants annotation was extracted from dbSNP (http://www.ncbi.nlm.nih.gov/SNP/); information on predicted function was based on the bioinformatic tool variant Effect Predictor (VEP) (http://useast.ensembl.org/Homo_sapiens/Tools/VEP) and SnpEff platform (http://snpeff.sourceforge.net/) using University of California, Santa Cruz (UCSC) transcripts

**Epigenetic Modifications in the Biology of Nonalcoholic Fatty Liver Disease: The Role of DNA-Hydroxymethylation and TET Proteins**

**Authors:** Carlos J Pirola PhD, Romina Scian PhD, Tomas Fernández Gianotti PhD, Hernán Dopazo PhD, Cristian Rohr B.Sc., Julio San Martino MD, Gustavo O Castaño MD, PhD, Silvia Sookoian MD, PhD

**Supplemental Figure F2**

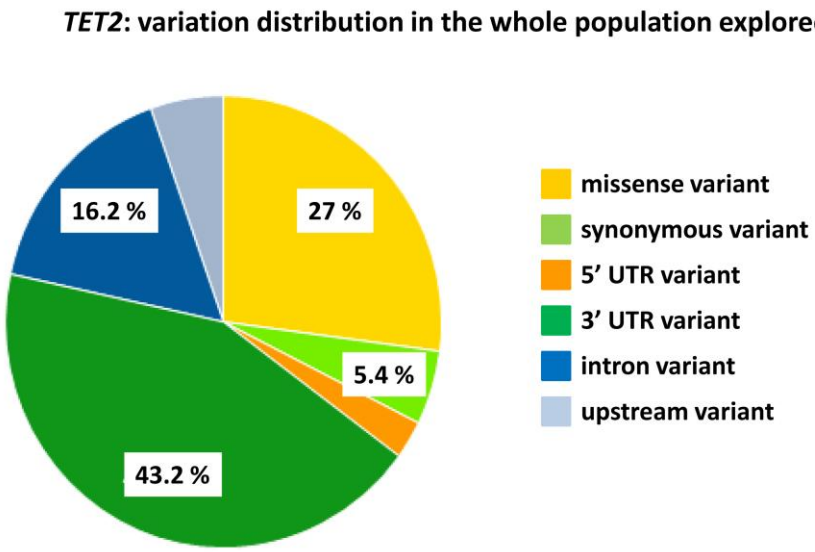


**Figure illustrates genetic variation of *TET2* in the whole population explored by next-generation sequencing: distribution of the predicted function of the variants.**
Information of variants annotation was extracted from dbSNP (http://www.ncbi.nlm.nih.gov/SNP/); information on predicted function was based on the bioinformatic tool variant Effect Predictor (VEP) (http://useast.ensembl.org/Homo_sapiens/Tools/VEP) and SnpEff platform (http://snpeff.sourceforge.net/) using University of California, Santa Cruz (UCSC) transcripts

**Epigenetic Modifications in the Biology of Nonalcoholic Fatty Liver Disease: The Role of DNA-Hydroxymethylation and TET Proteins**

**Authors:** Carlos J Pirola PhD, Romina Scian PhD, Tomas Fernández Gianotti PhD, Hernán Dopazo PhD, Cristian Rohr B.Sc., Julio San Martino MD, Gustavo O Castaño MD, PhD, Silvia Sookoian MD, PhD

**Supplemental Figure F3**

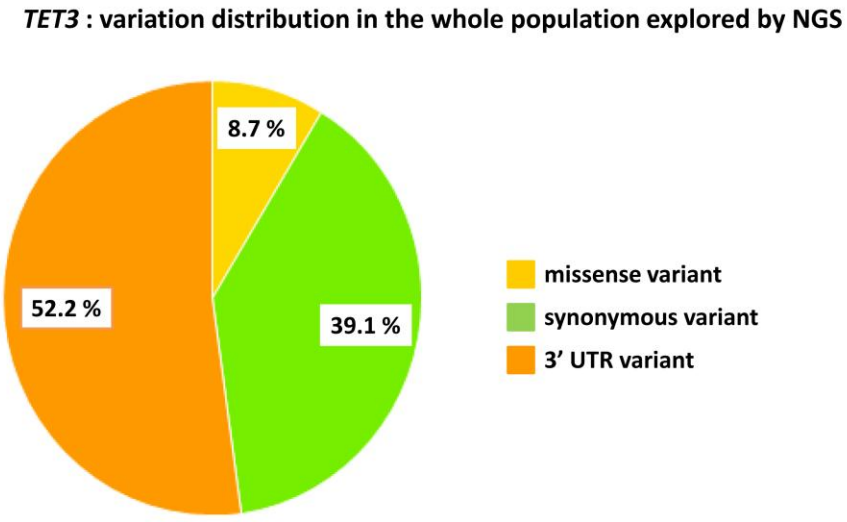


**Figure illustrates genetic variation of *TET3* in the whole population explored by next-generation sequencing: distribution of the predicted function of the variants.**
Information of variants annotation was extracted from dbSNP (http://www.ncbi.nlm.nih.gov/SNP/); information on predicted function was based on the bioinformatic tool variant Effect Predictor (VEP) (http://useast.ensembl.org/Homo_sapiens/Tools/VEP) and SnpEff platform (http://snpeff.sourceforge.net/) using University of California, Santa Cruz (UCSC) transcripts

**Epigenetic Modifications in the Biology of Nonalcoholic Fatty Liver Disease: The Role of DNA-Hydroxymethylation and TET Proteins**

**Authors:** Carlos J Pirola PhD, Romina Scian PhD, Tomas Fernández Gianotti PhD, Hernán Dopazo PhD, Cristian Rohr B.Sc., Julio San Martino MD, Gustavo O Castaño MD, PhD, Silvia Sookoian MD, PhD

**Supplemental Table S1**

| Gene | Forward primer 5'→3' | Reverse primer 5'→3' | Size, bp |
|---|---|---|---|
| **The primer sequences for mRNA gene expression** | | | |
| ***PPARGC1A*** | CCTGCATGAGTCTGTGCTCT | GCAAAGAGGCTGGTCTTCAC | 164 |
| ***PPIA*** | TTCCAGGGTTTATGTGTCA | CTCATCTTCAAATTTCTCC | 89 |
| **The primer sequences for mtDNA and nDNA amplification** | | | |
| **Mt F3212** | CACCCAAGAACAGGGTTTGT | TGGCCATGGGTATGTTGTTAA | 108 |
| **18S rRNA** | TAGAGGGACAAGTGGCGTTC3 | CGCTGAGCCAGTCA-GTGT | 105 |
| **The primers sequences for methylation-specific polymerase chain reaction** | | | |
| ***PPARGC1A*-M** | ATTTTTTATTGTTATGGGGGTAGTC | AAAAATATTTAAAAACGCAAACGAA | 143 |
| ***PPARGC1A*-U** | TTTTATTGTTATGGGGGTAGTTGA | AAAAAATATTTAAAAACACAAACAAA | 141 |

**Table that illustrates PCR primers used for exploration of liver *PPARGC1A* mRNA abundance and promoter mthylation status, and liver mtDNA copy number**

*PPARGC1A*=peroxisome proliferator-activated receptor gamma, coactivator 1 alpha; PPIA=cyclophilin. M: methylated, U: unmethylated.

**Epigenetic Modifications in the Biology of Nonalcoholic Fatty Liver Disease: The Role of DNA-Hydroxymethylation and TET Proteins**

**Authors:** Carlos J Pirola PhD, Romina Scian PhD, Tomas Fernández Gianotti PhD, Hernán Dopazo PhD, Cristian Rohr B.Sc., Julio San Martino MD, Gustavo O Castaño MD, PhD, Silvia Sookoian MD, PhD

**Supplemental Table S2**

| Gene symbol | Chromosome Location | Variant | db SNP | Alelles | MAF* | Predicted functional consequences | Ancestral allele |
|---|---|---|---|---|---|---|---|
| ***TET1*** | **10:68646098** (forward strand) | p.Ile1123Met Isoleucine (I) to Methionine (M) at position 1123 | rs3998860 | A/G | **0.27** (A) | Missense variant | A |
| ***TET2*** | **4:105275794** (forward strand) | p.Ile1762Val Isoleucine (I) to Valine (V) at position 1762 | rs2454206 | A/G | **0.24** (G) | Missense variant | A |

**Table that illustrates the main features of variants identified in *TET* locus by next generation sequencing and selected for following up in the replication study**

* MAF indicates minor allele frequency in the 1000 Genome Project.
TET1: Ten-Eleven Translocation-1 or tet methylcytosine dioxygenase 1, TET2: Ten-Eleven Translocation-2 or tet methylcytosine dioxygenase 2.

**Epigenetic Modifications in the Biology of Nonalcoholic Fatty Liver Disease: The Role of DNA-Hydroxymethylation and TET Proteins**

**Authors:** Carlos J Pirola PhD, Romina Scian PhD, Tomas Fernández Gianotti PhD, Hernán Dopazo PhD, Cristian Rohr B.Sc., Julio San Martino MD, Gustavo O Castaño MD, PhD, Silvia Sookoian MD, PhD

## SUPPLEMENTAL METHODS



### *1a. Physical, anthropometric, and biochemical evaluation*

The body mass index (BMI) was calculated as weight/squared height ($kg/m^2$) and was used as an index for relative weight. The waist and hip circumferences were also measured. Elevated blood pressure was defined as systolic arterial blood pressure (SABP) ≥ 130 mm Hg and/or diastolic (DABP) ≥ 85 mm Hg, or evidence of antihypertensive treatment.

Biochemical determinations: Blood was drawn from 12-hour fasting subjects that had been in a supine resting position for at least 30 min. Laboratory evaluation included serum ALT and AST, gamma glutamyl transferase (γGT), alkaline phosphatase (AP), glucose and insulin, total cholesterol and plasma triglycerides (TG). All biochemical determinations were measured using a Hitachi-912 Autoanalyzer (Roche, Diagnostic, Buenos Aires, Argentina) or Immulite 1000 (DPC, Buenos Aires, Argentina). Homeostasis Model Assessment (HOMA-IR) was used to evaluate an insulin resistance index and was calculated as follows: Fasting serum insulin (μU/ml) × Fasting plasma glucose (mmol/l) / 22.5.

### *1b. Liver biopsy and histopathological evaluation*

The degree of steatosis was assessed according to the system developed by Kleiner et al. [1] based on the percentage of hepatocytes containing macrovesicular fat droplets: grade 0 = <5%; grade 1 = 5-33%; grade 2 = 34% to 66% and grade 3 = >66%. NASH was defined as steatosis, accompanied by mixed inflammatory-cell infiltration, hepatocyte ballooning and necrosis,

glycogen nuclei, Mallory's hyaline, and any stage of fibrosis, including absent fibrosis [2]. Intra-acinar (lobular) inflammation was defined according to Brunt [2] as presence of cellular components of inflammation (polymorphonuclear leukocytes, lymphocytes and other mononuclear cells, eosinophils and microgranulomas) located in sinusoidal spaces, surrounding Mallory's hyaline or in hepatocellular necrosis. It was graded 0–3 and was defined as 0 (absent) = no foci; 1 = < 2 foci per 200 × field; 2 = 2 to 4 foci per 200 × field; and 3 = >4 foci per 200 × field. Ballooning was scored as: 0 = none; 1 = rare or few; and 2 = many. The severity of fibrosis was expressed on a 4-point scale, as follows: 0 = none; 1 =perivenular and/or perisinusoidal fibrosis in zone 3; 2 = combined pericellular portal fibrosis; 3 = septal/ bridging fibrosis; and 4 = cirrhosis.

### *1c-Liver Immunohistochemistry*

Four-micrometer sections were mounted onto silane-coated glass slides to ensure section adhesion through subsequent staining procedures. Briefly, sections were deparaffinized, rehydrated, and washed in phosphate buffer solution (PBS), before being treated with 3% $H_2O_2$ in PBS for 20 min at room temperature to block endogenous peroxidase. Following microwave heat-induced epitope retrieval in 0.1 M citrate buffer at pH 6.0 for 20 min, the slides were incubated with a dilution of 1:100 of TET1 (PAB21054, Abnova Taipei, Taiwan) and 1:500 of 5-hmc (Active Motif 39770, CA, USA). Immunostaining was performed using the VECTASTAIN Elite ABC Kit (Vector Lab. CA, USA) detection system. Subsequently, slides were immersed in a 0.05% 3,3'-diaminobenzidine solution in 0.1 M Tris buffer, pH 7.2, containing 0.01% $H_2O_2$. After brown color developed, slides were removed and the reaction was arrested by immersion in PBS. Hematoxylin was used as a counterstain for TET1 and eosin for 5-hmc. Immunostaining was semi-quantitatively evaluated in a blinded fashion regarding any of the histological and clinical characteristics of the patients by two independent observers. The extent of staining was scored according to its amount and intensity, using a 4-point scoring system, as follows: 0 = no staining; 1 = positive staining in less than 20% of cells; 2 = 21-50% of positive cells, and 3 = positive staining in more than 50% of cells. Scoring included nuclear and cytoplasm staining and the sections were observed by bright field microscopy, using an Axiostar plus (Carl Zeiss, Germany) microscope at a 400× magnification.

### *1d-RNA preparation and real-time RT-PCR for quantitative assessment of mRNA expression*

For RT-PCR, 1 to 3 μg of total RNA was reverse-transcribed using random hexamers and Moloney Murine Leukemia Virus (MMLV) reverse transcriptase (Promega, Wisconsin, USA). Real-time PCR was performed for quantitative assessment of mRNA expression in a StepOne Plus Real-Time PCR Systems (Applied Biosystems, CA 94404, USA). All the real-time PCR reactions were run in triplicate. The mRNA abundance of target genes was normalized to the amount of a housekeeping gene (cyclophilin A) to carry out comparisons between the groups. The selection of the housekeeping gene was based on the exploration of the most stable reference gene for testing liver mRNA expression among other housekeeping genes (β-actin, TATA box binding protein, cyclophilin A and glyceraldehyde-3-phosphate dehydrogenase) tested before starting this experiment. The geNorm program [3] was used to identify the appropriate reference control in our samples. The mRNA levels were expressed as the ratio of the estimated amount of the target gene relative to the cyclophilin A mRNA levels using fluorescence threshold cycle values (Ct) calculated for each sample, and the estimated efficiency of the PCR for each product was expressed as the average of all sample efficiency values obtained [4]. The specificity of

amplification and absence of primer dimers were confirmed using the melting curve analysis at the end of each run.

### *1e-Bisulfite treatment of DNA and methylation-specific polymerase chain reaction*

DNA bisulfite modification was based on bisulfite treatment of genomic DNA, thereby converting all the unmethylated cytosines to uracils, while conserving the methylated cytosines, by the EZ DNA Methylation Kit, according to the manufacturer's protocol **(**Zymo Research Corporation, Orange, CA, USA). The chemically modified DNA was subsequently used as a template for a methylation-specific polymerase chain reaction (MS-PCR) to determine the promoter methylation status of the selected CpG dinucleotides in the *PPARGC1A* promoter. The assay was based on a real-time quantitative PCR in an iCycler thermocycler (BioRad Hercules, CA, USA) using SYBR Green (Invitrogen, Buenos Aires, Argentina) as a fluorescent dye.

For the MS-PCR experiment, two pairs of primers were used: one pair was specific for bisulfite-modified methylated DNA (M primers) and the other pair was specific for bisulfite-modified unmethylated DNA (U primers). Thus, for each sample studied, two PCRs were performed simultaneously using the M primer and U primer pairs. Successful amplification from the M primers and U primers indicated methylation and unmethylation, respectively. For primer design, a sequence starting 2000 bp upstream from the transcriptional start site (TSS) of *PPARGC1A* was used in the MethPrimer program (http://www.urogene.org//methprimer/index1.html) to search for regions with potentially methylated CpG sites. For maximal discrimination between methylated and unmethylated alleles, M and U primers were designed to contain at least one CpG site at the 3'.

The level of methylated DNA is expressed as the ratio of the estimated amount for methylated DNA to the unmethylated DNA levels, calculated for each sample using the fluorescence threshold cycle (Ct) values for a previously estimated efficiency, according to a previous report [5]. We estimated the efficiency for each single sample tube using the slope of the exponential phase, as described by Ramakers et al. [6]. Furthermore, as expected, controls for unmethylated (a purified native amplicon) and fully methylated DNA (the same amplicon treated with the DNA methylase M.SssI (New England Biolabs, Ipswich, MA, USA) gave 0 and 100% DNA methylation patterns.

All the experiments were carried out in triplicate. The CV% was observed to be less than 5%. The specificity of amplification and the absence of primer dimers were confirmed by melting curve analysis at the end of each run and agarose electrophoresis.

To ensure the specificity of the method and to avoid variability in the results because of the presence of two CpG dinucleotides in the reverse primer, we designed a degenerated reverse primer that introduced a mismatch in the second CG site, and observed the absence of amplification, regardless of the target DNA (data not shown), indicating that primers recognize the status of both the CpG dinucleotides simultaneously.

### *1f-Quantification of mtDNA*

An assay based on real-time quantitative PCR was used for both nuclear DNA (nDNA) and mitochondrial DNA (mtDNA) quantification using SYBR Green as a fluorescent dye (Invitrogen) as previously described [7]. Results were presented as the mtDNA/nDNA ratio.

Real-time quantitative PCR was carried out in a BioRad iCycler (Bio-Rad). The calculation of DNA copy number involved extrapolation from the fluorescence readings in the mode of background subtracted from the BioRad iCycler according to Rutledge et al.[8]. Specificity of

amplification and the absence of primer dimers were confirmed by the melting curve analysis at the end of each run. The two target amplicon sequences (mtDNA and nDNA) were visualized in agarose 2% and purified by Qiagen Qiaex II, Gel extraction Kit (Tecnolab, Buenos Aires, Argentina), and dilutions of purified amplicons were used as the standard curve. The inter-assay variation coefficient was less than 20%. Specificity of the method was evaluated as was previously described [8;9].

### *1g-Next generation sequencing (NGS)*

DNA was isolated from whole blood, as previously described [10] and was quantified by a Qubit DNA high-sensitivity assay kit.

Library preparation for each sample was performed using the IT AmpliSeq 2.0 Beta kit following the manufacturer's instructions. Briefly, 10 ng of DNA was used as a template to generate the amplicon library for sequencing variation in the three genes. Genomic regions of interest were PCR amplified prior to sequencing, and the sequencing adaptors with short stretches of index sequences (96 barcodes) that enabled sample identification were ligated to the amplicons using the IT Xpress barcode adaptor kit. The prepared library was quantified using the Ion library TaqMan Quantitation Kit. Sequencing template preparation (emulsion PCR and beads-enrichment) from sequencing libraries was carried out using an Ion OneTouch Template Kit and Ion OneTouch system (Ion OneTouch Instrument and Ion OneTouch ES, Life Technologies, Carlsbad, CA, USA) according to the manufacturer's protocol. Prepared templates were sequenced using Ion Sequencing Kit v2. Ion Torrent Suite software 4.2.1 (Life Technologies, Carlsbad, CA, USA) was used for converting raw signals into base calls and extracting FASTQ files of sequencing reads. Number of nucleotide flows during sequencing was set to 500 cycles.

### *1h-Variant calling, estimation of quality control, data analysis and prediction of variant/ mutation effect*

The data obtained from the Ion Torrent PGM were processed using the Ion Torrent Suite Software v 4.2.1 (Life Technologies, Carlsbad, CA, USA). Variants were annotated with dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) IDs using SnpSift. *In silico* analysis, aimed at predicting gene and transcript functional consequences was performed by the bioinformatic tool variant Effect Predictor (VEP) (http://useast.ensembl.org/Homo_sapiens/Tools/VEP), employing ENSEMBL transcripts and SnpEff platform (http://snpeff.sourceforge.net/) using University of California, Santa Cruz **(**UCSC) transcripts. CellBase [11] available at http://docs.bioinfo.cipf.es/projects/cellbase was used to annotate variants with the phenotype information from HGMD, ClinVar, UNIPROT and COSMIC. Annotation in http://www.ncbi.nlm.nih.gov/SNP was also used to determine whether variants were novel or already associated with a phenotype.

Read alignment to the hg19 Human genome reference sequence was performed using torrent mapping alignment program **(**TMAP) included in PGM. The Ion Torrent variant caller plug-in was used to perform the variant calling with the germline low stringency settings. In order to filter erroneous base callings, control quality filtering steps were performed using proprietary Perl scripts. At the variant coverage depth > 20, we inspected the presence of strand bias by checking the number of bases covered in both strands for the variant, the reference =/ > 4 reads in both strands, and the quality for variant calling > 20. The retained variants were visually

examined using Integrative Genomics Viewer (IGV) software **(**http://www.broadinstitute.org/igv/**)** to check for any inconsistency in the base calls.

***1i-Replication study: Genotype and association analysis, power and sample size calculation, and population stratification***

To ensure genotyping quality, we included DNA samples as internal controls, hidden samples of known genotype, and negative controls (water). The overall genotype completion rate was 100%. To account for possible population stratification, we used a collection of 13 SNPs (rs6830727, rs12639788, rs1282807, rs1947745, rs7162312, rs12951674, rs7212346, rs1934869, rs9542666, rs11843545, rs9725124, rs2798659, and rs2199940) at different loci (located in chromosomes 4, 15, 17, 13, 1, and 3) and then analyzed the data with the Structure program Version 2 [12].
We found no evidence of stratification in our sample because Q values were similar for the cases and the controls. Moreover, the Structure program assigned a similar distance to clusters and no further improvement in the fitting model was achieved by adding up to four clusters (the ln of likelihood was maximum for K = 1).

## REFERENCES


(1) Kleiner DE, Brunt EM, Van NM et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 2005;41:1313-1321.

(2) Brunt EM, Kleiner DE, Wilson LA, Belt P, Neuschwander-Tetri BA. Nonalcoholic fatty liver disease (NAFLD) activity score and the histopathologic diagnosis in NAFLD: distinct clinicopathologic meanings. *Hepatology* 2011;53:810-820.

(3) Schlotter YM, Veenhof EZ, Brinkhof B et al. A GeNorm algorithm-based selection of reference genes for quantitative real-time PCR in skin biopsies of healthy dogs and dogs with atopic dermatitis. *Vet Immunol Immunopathol* 2009;129:115-118.

(4) Ruijter JM, Ramakers C, Hoogaars WM et al. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res* 2009;37:e45.

(5) Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001;25:402-408.

(6) Ramakers C, Ruijter JM, Deprez RH, Moorman AF. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett* 2003;339:62-66.

(7) Sookoian S, Rosselli MS, Gemma C et al. Epigenetic regulation of insulin resistance in nonalcoholic fatty liver disease: impact of liver methylation of the peroxisome proliferator-activated receptor gamma coactivator 1alpha promoter. *Hepatology* 2010;52:1992-2000.

(8) Rutledge RG. Sigmoidal curve-fitting redefines quantitative real-time PCR with the prospective of developing automated high-throughput applications. *Nucleic Acids Res* 2004;32:e178.

(9) Gianotti TF, Sookoian S, Dieuzeide G et al. A Decreased Mitochondrial DNA Content Is Related to Insulin Resistance in Adolescents. *Obesity* 2008;16:1591-1595.

(10) Sookoian S, Castano GO, Burgueno AL, Gianotti TF, Rosselli MS, Pirola CJ. A nonsynonymous gene variant in the adiponutrin gene is associated with nonalcoholic fatty liver disease severity. *J Lipid Res* 2009;50:2111-2116.

(11) Bleda M, Tarraga J, de MA et al. CellBase, a comprehensive collection of RESTful web services for retrieving relevant biological information from heterogeneous sources. *Nucleic Acids Res* 2012;40:W609-W614.

(12) Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000;155:945-959.