A MULTI-OMICS STUDY OF A MONOZYGOTIC TWIN PAIR DISCORDANT FOR AMYOTROPHIC LATERAL SCLEROSIS TO INVESTIGATE GENETIC AND EPIGENETIC FACTORS IMPLICATED IN THE DISEASE

Tosi, M.[1]; Zuccalà, M.[1]; Favero, F.[1]; Corrado, L.[1]; Croce, R.[1]; Basagni, C.[1]; Barizzone, N.[1]; Follia, L.[1]; De Marchi, F.[2]; Chinni, E.[3]; Cantello, R.[4]; Mazzini, L.[2]; Corà, D.[1]; Leone, M.[5]; D'Alfonso, S.[1]

[1] Department of Health Sciences, University of Eastern Piedmont UPO, Via Solaroli 17, 28100, Novara, Italy

[2] ALS Center AOU Maggiore della Carità, Corso Mazzini 18, 28100, Novara, Italy

[3] Thrombosis and Haemostasis Unit, Fondazione I.R.C.C.S. "Casa Sollievo della Sofferenza", Viale Cappuccini, S. Giovanni Rotondo, 71013, Foggia, Italy

[4] Department of Translational Medicine, University of Eastern Piedmont UPO, Via Solaroli 17, 28100, Novara, Italy

[5] SC Neurologia, Dipartimento di Scienze Mediche, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, 71013, Foggia, Italy

Introduction and aim of the study

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease, characterised by progressive death of upper and lower motor neurons. 90% of patients have no prior family history (sporadic ALS), while 10% of ALS patients have at least one other affected family member (familial ALS). This disease is phenotypically heterogeneous and its etiology is still poorly understood, as both genetic susceptibility and environmental exposure contribute to the pathogenesis.

To investigate genetic and epigenetic factors underlying ALS, we studied a monozygotic twin pair discordant for ALS. We applied a multi-omics approach, combining whole exome sequencing with genome-wide methylome- and transcriptome data from whole blood and PBMCs.

Materials and Methods

Transcriptome Analysis

- 8 samples: biological duplicate and, for the second blood sample, a technical triplicate
- 100 ng of RNA from PBMCs; library kit: Illumina TruSeq Stranded mRNA. mRNA sequencing was performed using NextSeq 500/550 High Output Kit v2.5 (150 Cycles - 2 X 75 read length, paired-end), obtaining a mean of 50 million reads per sample
- Quality controls were assured using FastQC. Data were analyzed with RSEM and STAR for the alignment of reads to the reference genome (GRCh38/hg38). We evaluated differentially expressed genes (DEGs) by DESeq2 with p.value adj < 0.1 and |log2FC|> 1. Lastly, pathway analysis has been conducted with different bioinformatic tools as G-Profiler, ToppGene, GSEA and IPA

Methylation Analysis

- 8 samples: a biological duplicate and, for each blood sample, a technical duplicate
- 500 ng of DNA from whole blood converted by using bisulphite conversion technique. We used the Infinium Methylation EPIC Array scanned on the NextSeq 550
- Quality controls were performed on Illumina GenomeStudio software. Results were analyzed using both GenomeStudio and the Chip Analysis Methylation Pipeline (ChAMP) Bioconductor package that allow the identification of differentially methylated probes (P.value adj ≤ 0.1; Δβ ≥ 0.25; Δβ ≤ -0.25)

Whole Exome Sequencing

- 2 samples: one blood sample per subject
- 50 ng of DNA according to Agilent Sure Select QXT Kit. WES has been processed using NextSeq 500/550 High Output Kit v2, producing 2x150 bp read lengths and 30X coverage across samples
- Fastq files were aligned on GRCh37 genome and BWA software produced the bam files. For each patient, a list of variations in a VCF format file was produced by GATK software and annotation of VCF files was performed by wANNOVAR software. CNV analysis has been performed by ExomeDepth tool. We tested ALS vs healthy twin and healthy vs ALS twin; then variants were classified as benign, pathogenic or of uncertain significance by ClassifyCNV Scores

Results

Transcriptome Analysis

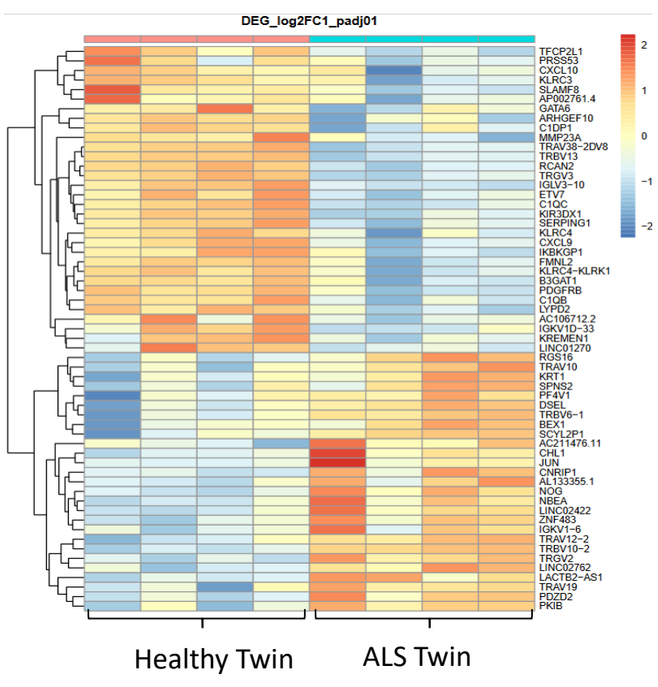


Figure 1: The heatmap shows the identified 59 DEGs for the healthy twin (left) and the ALS twin (right). Upregulated genes are represented in the red scale, while downregulated genes in the blue scale. On the right there are genes name, on the left genes are clustered by similar expression values

Methylation Analysis

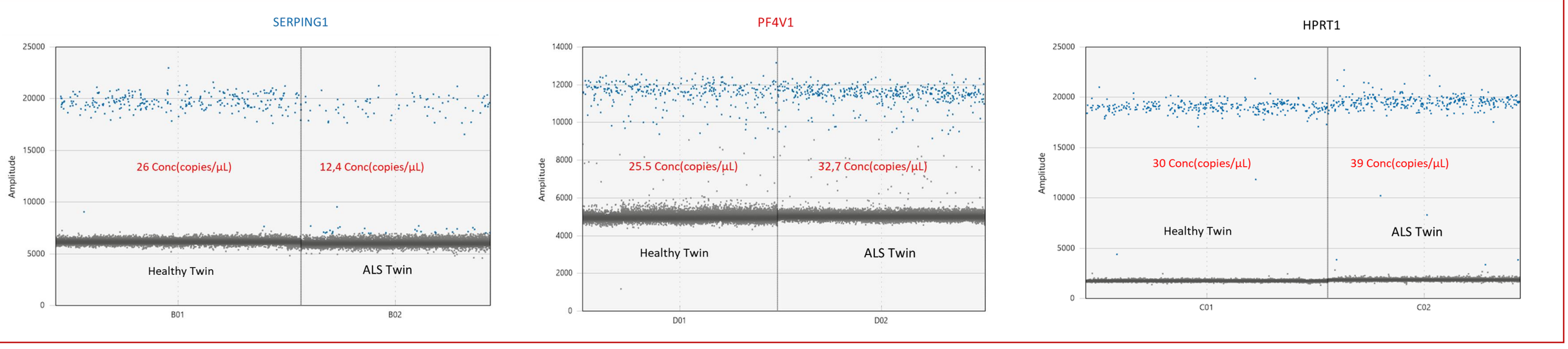
pvalue adj ≤ 0.1	P.Value Adj	CHR:POS	Gene	Cgi
cg18454685	0,004060966	17:48639239	CACNA1G	Body-island
cg27533288	0,054051711	10:118896776	VAX1	Body-island

Δβ ≤ - 0.25	Δβ	P.Value Adj	CHR:POS	Gene	Cgi
cg18686665	-0,2899808	0,692505608	2:629121		IGR-island
cg01032200	-0,3178169	0,711717766	1:155290641	RUSC1-AS1	Body-island

Δβ ≥ 0.25	Δβ	P.Value Adj	CHR:POS	Gene	Cgi
cg18565204	0,4725682	0,598885039	16:70298926	AARS	Body-opensea
cg18987683	0,2675134	0,915051596	3:160283058	KPNA4	1stExon-island

Table 1: Differentially Methylated Probes identified, based on p.value adj ≤ 0.1 or Δβ ≤ - 0.25 and Δβ ≥ 0.25

Figure 2: Example of 2 out of 6 validated DEGs by ddPCR. SERPING is confirmed to be downregulated in the ALS twin, while PF4V1 is upregulated. HPRT1 is the housekeeping gene



Whole Exome Sequencing

SNV	Healthy Twin vs ALS Twin	ALS Twin vs Healthy Twin
Different Variants	109	162
Pass Variants	10	25
Position	8 exonic 2 intronic	15 exonic 10 intronic
Exonic Function	3 synonymous 4 nonsynonymous 1 nonframeshift deletion	5 synonymous 8 nonsynonymous 1 nonframeshift deletion 1 nonframeshift insertion

Table 2: summary of SNVs identified in the ALS twin filtered for gnomAD_EXOME_ALL ≤ 0,00005 OR frequency 0

CNV	Healthy Twin vs ALS Twin	ALS Twin vs Healthy Twin
Deletions	2 of uncertain significance	3 of uncertain significance
Duplication	2 of uncertain significance	1 of uncertain significance

Table 3: summary of specific CNVs different for 60% and identified in the ALS twin

Conclusion

We studied a discordant twin pair for ALS considering three different omics, independently and in combination, to identify disease-relevant changes. Twins tested negative for mutations in main ALS-genes. From RNA-seq we identified 59 DEGs and validated 6 DEGs by ddPCR; functional analyses with distinct bioinformatic tools underlined a possible role of the immune system in the disease, as partially described in literature. We also identified 2 differentially methylated probes in *CACNA1G*, expressed mostly in brain, and *VAX1* genes and, filtering by Δβ values, we found 2 probes with Δβ ≤ - 0.25 in an intergenic region and in *RUSC1-AS1* gene and 2 probes with Δβ ≥ 0.25 in *AARS* and *KPNA4* genes. For exome analyses, 3 deletions and 1 duplication of uncertain significance were identified only in the ALS twin, while filtering for frequency and QC we were able to identify 25 variants (15 exonic, 10 intronic). Further understanding of these immunological results and the validation of methylation results by methylation-specific droplet digital PCR (ddMSP) combined with methylation-dependent restriction enzymes are ongoing to elucidate possible somatic genetic factors that could underlie susceptibility to sporadic ALS.