



Molecular Taxonomy of Sporadic Amyotrophic Lateral Sclerosis Using Disease-Associated Genes

Giovanna Morello[†], Antonio Gianmaria Spampinato[†] and Sebastiano Cavallaro^{*}

Institute of Neurological Sciences, Italian National Research Council, Catania, Italy

OPEN ACCESS

Edited by:

Shannon K. McWeeney,
Oregon Health & Science
University, USA

Reviewed by:

Leonard Rabinow,
Université Paris-Sud, France
Don A. Baldwin,
Signal Biology Inc., USA

*Correspondence:

Sebastiano Cavallaro
sebastiano.cavallaro@cnr.it

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Neurogenomics,
a section of the journal
Frontiers in Neurology

Received: 12 December 2016

Accepted: 03 April 2017

Published: 19 April 2017

Citation:

Morello G, Spampinato AG and
Cavallaro S (2017) Molecular
Taxonomy of Sporadic Amyotrophic
Lateral Sclerosis Using Disease-
Associated Genes.
Front. Neurol. 8:152.
doi: 10.3389/fneur.2017.00152

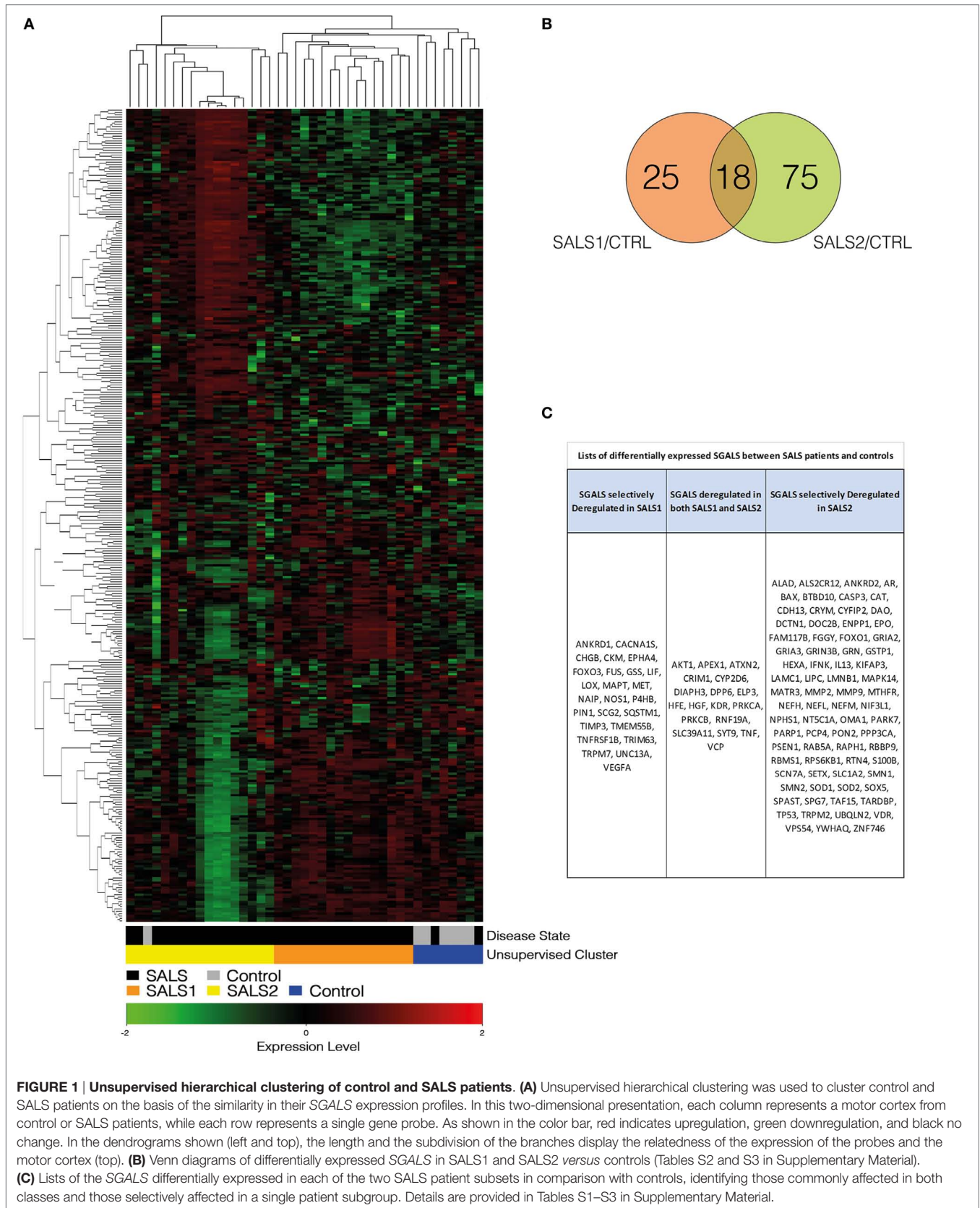
Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease characterized by selective loss of upper and lower motor neurons. Despite intensive research, the origin and progression of ALS remain largely unknown, suggesting that the traditional clinical diagnosis and treatment strategies might not be adequate to completely capture the molecular complexity underlying the disease. In our previous work, comprehensive genomic profiling of 41 motor cortex samples enabled to discriminate control from sporadic ALS patients and segregated these latter into two distinct subgroups, each associated with different deregulated genes and pathways. Interestingly, some deregulated genes in sporadic ALS were previously associated with familiar ALS, indicating shared pathogenic mechanisms between the two forms of disease. In this, we performed cluster analysis on the same whole-genome expression profiles using a restricted (203) subset of genes extensively implicated in monogenic forms of ALS. Surprisingly, this short and unbiased gene list was sufficiently representative to allow the accurate separation of SALS patients from controls and the stratification of SALS patients into two molecularly distinct subgroups. Overall, our findings support the existence of a molecular taxonomy for ALS and represent a further step toward the establishment of a molecular-based diagnosis and patient-tailored therapies.

Keywords: amyotrophic lateral sclerosis, genomics, expression profiling, molecular classification of disease subtypes, gene ontology analysis

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a genetically and clinically heterogeneous neurodegenerative disease characterized by progressive muscular paralysis reflecting degeneration of motor neurons in both the motor cortex and the spinal cord (1). Although ~10% of cases have a clear family history of ALS (FALS), the remaining 90% is considered sporadic (SALS) and probably associated with a polygenic and multifactorial etiology. Despite intensive research, the origin and progression of ALS remain largely unknown, suggesting that the traditional clinical diagnosis and treatment strategies might not be adequate to completely capture the molecular complexity underlying disease (2).

In the last few years, significant advances toward understanding the complex architecture of ALS have been achieved, thanks to the development of high-throughput profiling methods that have started to decipher genes and pathways involved in the disease pathogenesis (3–6). However, the analysis was often limited to small sample sizes, thus not allowing a transcriptionally driven classification of ALS subtypes.



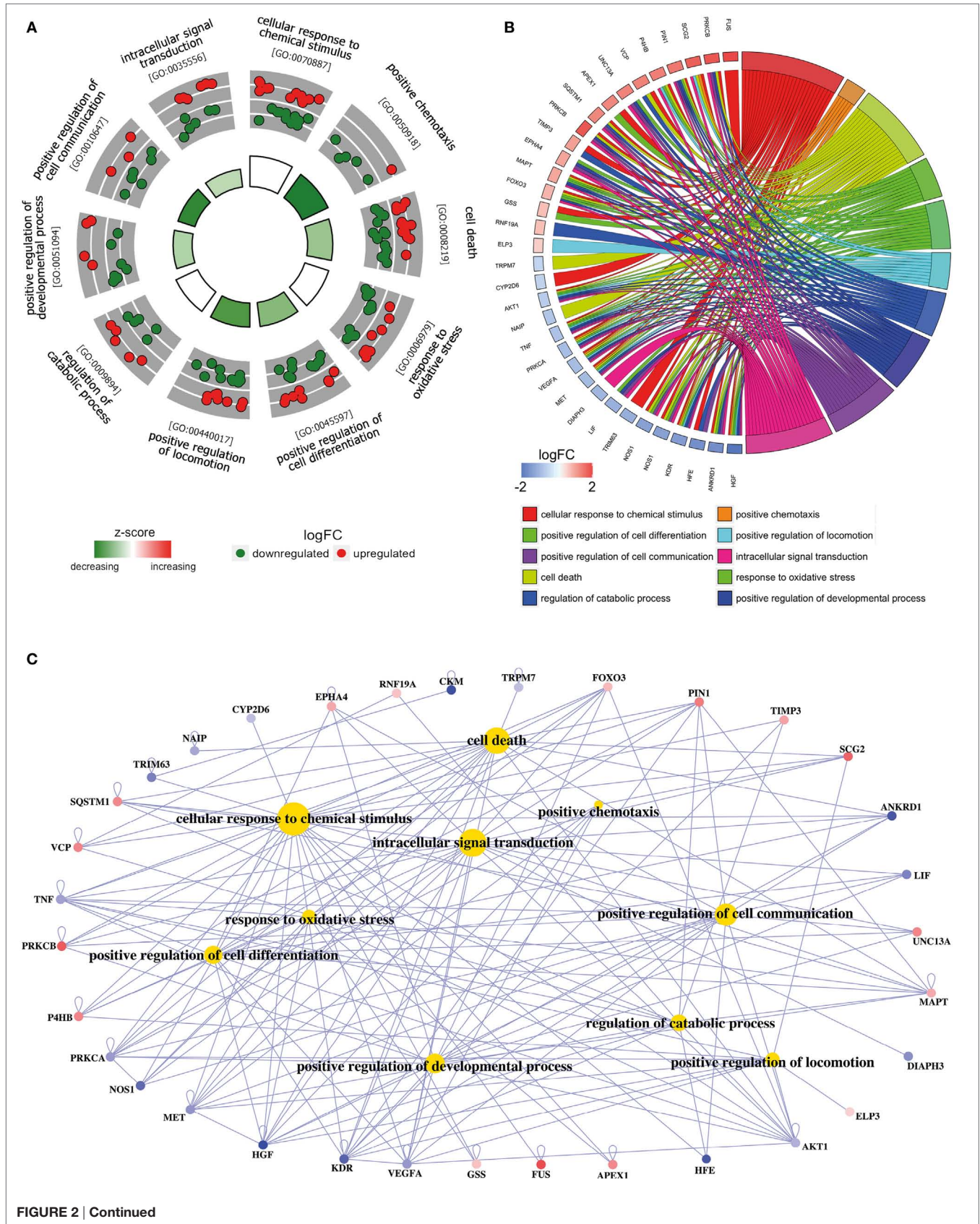


FIGURE 2 | Continued

Functional analysis of deregulated SGALS in SALS1. (A) The outer circle shows a scatter plot of the expression levels (logFC) for SALS1-related differentially expressed SGALS in each enriched gene ontology (GO) term: red circles indicate upregulation and blue ones downregulation. The inner ring is a bar plot where the height of the bar indicates the significance of GO terms (log₁₀-adjusted *P* value), and color corresponds to the z-score: green, decreased; red, increased; and white, unchanged. **(B)** The plot shows the relationship between statistically significant SGALS in SALS1 and their related GO terms; for each gene, the logFC value is shown by red/blue colored rectangles. **(C)** Gene-concept networks by GO analysis for differentially expressed SGALS in SALS1, using *GeneAnswers* Package. Yellow hubs correspond to the most enriched GO terms; red and blue nodes represent downregulation and upregulation of SGALS in SALS1, respectively. The size of each yellow hub is proportional to the statistical significance of the identified genes in corresponding GO categories. Details are provided in Table S4 in Supplementary Material.

In a previous work, for the first time, we demonstrated the remarkable power of genomic profiling to characterize and explore the hidden biological and molecular heterogeneity of SALS, enabling the identification of etiopathogenic mechanisms and potential therapeutic targets that were not put in evidence by considering SALS pathology as a single entity (7). In particular, we used an unsupervised hierarchical algorithm to cluster 41 motor cortex samples from control and SALS patients on the basis of their similarities measured over the most “hypervariable genes” (9,646 genes with a SD > 1.5). The transcriptome profiles enabled to differentiate controls from SALS patients and clearly distinguished two SALS subgroups, each associated with differentially expressed genes and pathways. Interestingly, some of deregulated genes in SALS patients were previously associated with FALS, indicating that pathological events associated with both forms of the disease may converge on common signaling pathways.

In this study, we used the full list of genes that have been implicated, up-to-date, in the causation and/or susceptibility of monogenic forms of ALS, investigating their expression patterns in SALS patients and evaluating their effectiveness in confirming the existence of a molecular taxonomy for ALS. Surprisingly, this short and unbiased list of genes was sufficiently representative to accurately separate SALS patients from controls and confirmed the presence of two molecularly distinct SALS subgroups.

MATERIALS AND METHODS

Data Sources and Gene Selection

For this study, we referred to our previously described transcriptome data set (7), available at ArrayExpress¹ with the accession number E-MTAB-2325. Briefly, this data set consists of mRNA expression profiles of 41,059 genes from 41 motor cortex samples of SALS and control subjects hybridized onto 4 × 44K Whole Human Genome Oligo expression microarrays (Agilent Technologies). Once obtained, raw intensity values were thresholded to 1, log₂-transformed, normalized, and baselined to the median of all samples by using GeneSpring GX (Agilent Technologies). In addition, the fold change (FC) of each gene was calculated between SALS patients and individual controls. A detailed description of the study design, subject characteristics, and experimental procedures is reported in the original publication (7).

In the present work, we reanalyzed our genomic data focusing on a list of 203 genes (here referred as SGALS) that were

reported to significantly influence ALS susceptibility (Table S1 in Supplementary Material). The SGALS gene set was obtained by integrating data from various source databases, including ALSod² and ALSGene.³ These are two freely available online bioinformatics repositories that provide systematic and in-depth qualitative and quantitative overviews of ALS genetic research, embracing known and putative risk factors, disease-causing genes, and other related genetic regions affected by different types of mutations (8).

Clustering Analysis and Statistical Investigation of SGALS Gene Expression Data

Once processed, normalized expression data were imported into R and filtered for the set of probes mapping to SGALS genes. The resulting expression profiles were clustered for visualization by using a Pearson centered correlation as a distance metric with average linkage rules in the tree building algorithm. In particular, unsupervised hierarchical cluster analysis was performed using the *hclust* function in R, while heatmaps were rendered using *heatmap.3* package from CRAN (9).

To assess the statistical significance of SGALS expression changes in SALS patients compared to healthy controls, statistical analysis was performed by using GeneSpring GX software package (version 14.5; Agilent Technologies). In particular, we used a one-way analysis of variance with a Benjamini–Hochberg’s false discovery rate (FDR)-controlling procedure followed by a *post hoc* Tukey’s test. Differentially expressed genes with an FDR-adjusted $P \leq 0.05$ and fold changes (FC) ≥ 1.5 were deemed as significant. Moreover, to achieve a precise estimation of the differential gene expression, multiple probes per gene were used. For hierarchical clustering (Figure 1A) and gene ontology (GO) enrichment analysis (Figures 2 and 3), relationships among genes are represented. When multiple probes corresponded to the same gene, the expression changes of probe sets were averaged.

Functional Characterization of Differentially Expressed SGALS Genes

To interpret the biological significance of gene expression changes, we performed a functional characterization of SGALS differentially expressed in both SALS subgroups. In particular, GO enrichment analysis of biological process (BP) was performed by using a hypergeometric test in the R Bioconductor

¹<http://www.ebi.ac.uk/arrayexpress/>.

²<http://alsod.iop.kcl.ac.uk/>.

³<http://www.alsgene.org>.

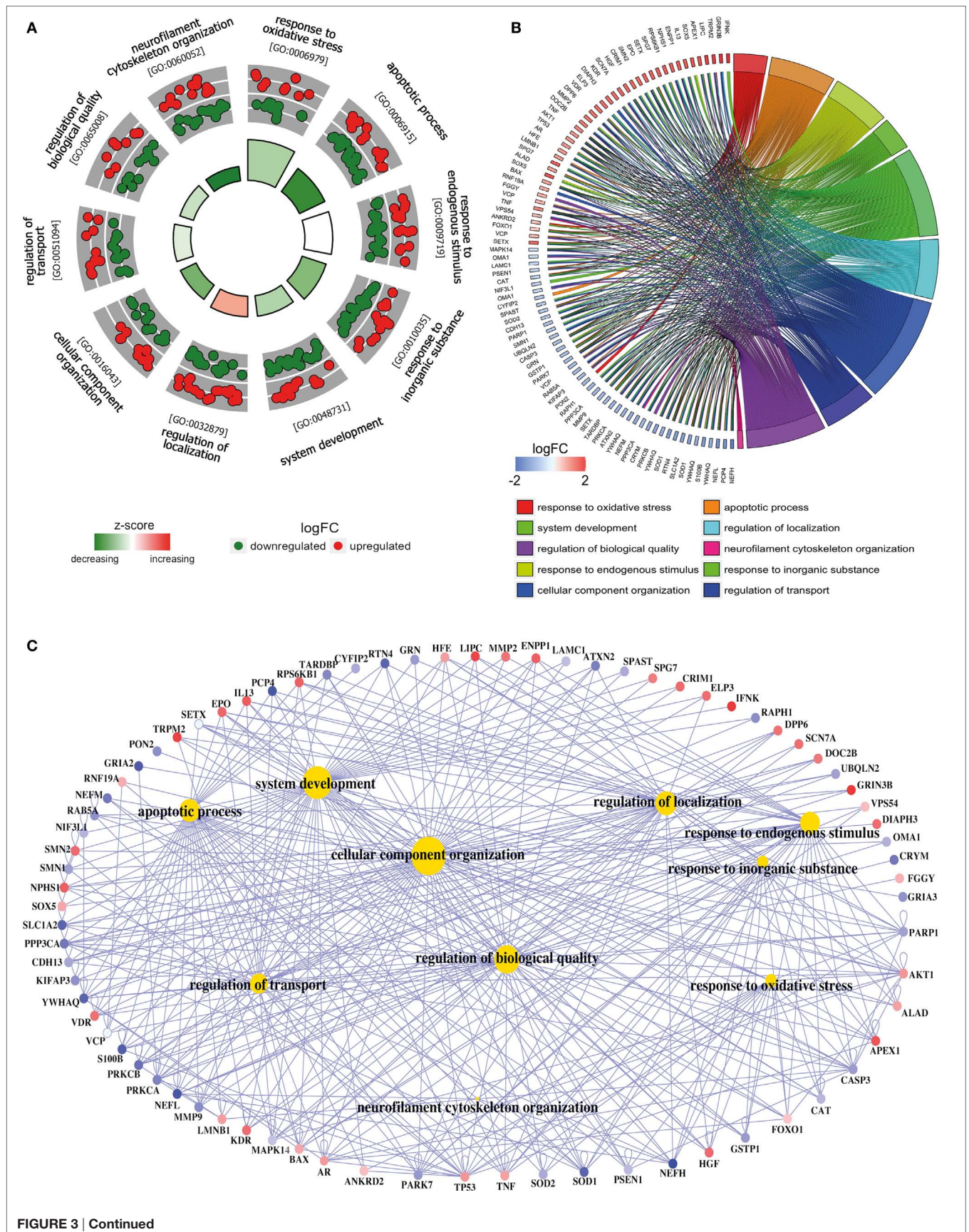


FIGURE 3 | Continued

Functional analysis of deregulated SGALS in SALS2. (A) The outer circle shows a scatter plot of the expression levels (logFC) for SALS2-related differentially expressed SGALS in each enriched gene ontology (GO) term: red circles indicate upregulation and blue ones downregulation. The inner ring is a bar plot where the height of the bar indicates the significance of GO terms (log₁₀-adjusted *P* value), and color corresponds to the z-score: green, decreased; red, increased; and white, unchanged. **(B)** The plot shows the relationship between statistically significant SGALS in SALS2 and their related GO terms; for each gene, the logFC value is shown by red/blue colored rectangles. **(C)** Gene-concept networks by GO analysis for differentially expressed SGALS in SALS2, using *GeneAnswers* Package. Yellow hubs correspond to the most enriched GO terms; red and blue nodes represent downregulation and upregulation of SGALS in SALS2, respectively. The size of each yellow hub is proportional to the statistical significance of the identified genes in corresponding GO categories. Details are provided in Table S5 in Supplementary Material.

package *GeneAnswers* (10). *P* values were then adjusted by using multiple testing adjustments with an FDR <0.05 as the significance threshold. In addition, to reduce potential errors due to the use of preselected gene sets (11), we also performed a “control” GO enrichment analysis on both the entire list of differentially expressed SGALS in the two SALS subgroups (Table S6 in Supplementary Material) and two randomly selected subsets of SGALS (*n* = 100) (Tables S7 and S8 in Supplementary Material).

To add quantitative molecular information to GO terms of interest, we used *GOCircle* and *GOChord* plot functions of *GOPLOT* R package (12), which permit to incorporate data derived from expression level measurements with those obtained from the functional annotation enrichment analysis. In addition, to provide a readable graphic representation of the complex relationship between SGALS and relative GO terms, the “concept-and-gene network” was constructed by Bioconductor package *GeneAnswers* (13).

RESULTS AND DISCUSSION

In this study, we reanalyzed our transcriptome data set (7), by focusing on a restricted subset of genes, previously identified as causative or susceptibility genes in ALS (here referred as SGALS) (Table S1 in Supplementary Material). Hierarchical clustering analysis for this minimal and unbiased gene panel resulted highly informative in supporting the existence of prominent molecular signatures capable of accurately separating controls from SALS patients and subdividing these latter into two greatly divergent subgroups (SALS1 and SALS2) (**Figure 1A**). These findings demonstrate that SGALS may not only cause the monogenic form of ALS but also act in the sporadic form, supporting the existence of shared pathogenic mechanisms between the two forms of the disease.

Next, a statistical significance test was performed to define genes that were differentially expressed between SALS patients and controls. Our analysis demonstrated that 118 of 203 SGALS were differentially deregulated in SALS patients and, interestingly, the majority of these genes were cluster specific (25 in SALS1 and 75 in SALS2), suggestive of a great divergence of SALS patient subgroups at the molecular level (**Figures 1B,C**; Tables S2 and S3 in Supplementary Material).

To provide a biologically meaningful interpretation of our results, we further performed a GO enrichment analysis for detecting significantly overrepresented GO categories (BP) among the set of differentially expressed SGALS, by using the TopGO package in R (12). The enrichment of GO annotation terms revealed

that, although some processes were commonly deregulated in both SALS subgroups (e.g., *regulation of apoptosis* and *cellular response to external/endogenous stimulus*), the majority of them were subgroup specific (**Figures 2 and 3**; Tables S4 and S5 in Supplementary Material). In particular, the most representative functional processes in SALS1 were annotated as involved in the *regulation of chemotaxis*, *regulation of cellular communication/differentiation*, and *intracellular signal transduction* (**Figures 2A,B**; Table S4 in Supplementary Material). Deregulated genes in SALS2 were associated with *response to oxidative stress*, *cytoskeleton organization*, *regulation of transport*, and *regulation of cellular localization* (**Figures 3A,B**; Table S5 in Supplementary Material). While these findings are consistent with previous evidence about the crucial role of these pathogenetic mechanisms in ALS (4, 7, 14–18), they suggest for the first time the differential involvement of these mechanisms in specific subsets of ALS patients, offering an useful starting point for the further development of personalized diagnostics and targeted therapies.

The existing interconnections between multiple differentially expressed SGALS and their related GO BPs were made more evident thanks to the generation of SALS subgroup-specific “Concept-and-Gene Networks” (**Figures 2C and 3C**). The functional interpretation of these GO-based genetic networks not only highlights the possibility that SGALS can exert their pathogenic effects *via* different multifactorial combinations but mostly provides important insights into candidate driver genes that may truly exert a fundamental influence on ALS (e.g., genes with a high number of connections/edges in the network). Among these, an interesting example is represented by *TNF*, a gene encoding a pro-inflammatory factor whose expression levels were found significantly elevated in serum of ALS patients (19). In agreement with these findings, we observed deregulated expression of *TNF* in both clusters of SALS patients, suggesting its potential role as a marker in SALS. *NEFH* is another attractive candidate gene encoding a neurofilament protein that plays an important role in axonal and dendritic intracellular transport and in the maintenance of neuronal caliber. Consistent with the observation that the decreased activity of NEFH was observed only in a small group of ALS patients (20, 21), we found low transcriptional levels of this gene in SALS2 patients, suggesting that this protein may represent a susceptibility factor for specific subgroups of SALS patients.

In summary, the SGALS panel was sufficiently representative to accurately differentiate SALS patients from controls, suggesting that sporadic and monogenic forms of ALS share common etiopathogenic mechanisms. In addition, the same unbiased gene list confirmed the stratification of SALS patients into two

molecularly distinct subgroups associated with a good number of cluster-specific candidate genes and unique biological features. Overall, our findings indicate the existence of a molecular taxonomy for ALS and represent a further step toward the establishment of a molecular-based diagnosis and patient-tailored therapies.

AUTHOR CONTRIBUTIONS

GM and AS wrote the manuscript. AS performed bioinformatics analyses. SC conceived, directed, and supervised the project. All authors have read and approved the final version of this manuscript and agreed to be accountable for all aspects of the work.

REFERENCES

1. Wijesekera LC, Leigh PN. Amyotrophic lateral sclerosis. *Orphanet J Rare Dis* (2009) 4:3. doi:10.1186/1750-1172-4-3
2. Agosta F, Al-Chalabi A, Filippi M, Hardiman O, Kaji R, Meininger V, et al. The El Escorial criteria: strengths and weaknesses. *Amyotroph Lateral Scler Frontotemporal Degener* (2015) 16(1–2):1–7. doi:10.3109/21678421.2014.964258
3. Pasinelli P, Brown RH. Molecular biology of amyotrophic lateral sclerosis: insights from genetics. *Nat Rev Neurosci* (2006) 7(9):710–23. doi:10.1038/nrn1971
4. Lederer CW, Torrisi A, Pantelidou M, Santama N, Cavallaro S. Pathways and genes differentially expressed in the motor cortex of patients with sporadic amyotrophic lateral sclerosis. *BMC Genomics* (2007) 8(1):26. doi:10.1186/1471-2164-8-26
5. Ravits J, Appel S, Baloh RH, Barohn R, Brooks BR, Elman L, et al. Deciphering amyotrophic lateral sclerosis: what phenotype, neuropathology and genetics are telling us about pathogenesis. *Amyotroph Lateral Scler Frontotemporal Degener* (2013) 14(Suppl 1):5–18. doi:10.3109/21678421.2013.778548
6. Renton AE, Chio A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. *Nat Neurosci* (2014) 17(1):17–23. doi:10.1038/nn.3584
7. Aronica E, Baas F, Iyer A, ten Asbroek AL, Morello G, Cavallaro S. Molecular classification of amyotrophic lateral sclerosis by unsupervised clustering of gene expression in motor cortex. *Neurobiol Dis* (2015) 74:359–76. doi:10.1016/j.nbd.2014.12.002
8. Lill CM, Abel O, Bertram L, Al-Chalabi A. Keeping up with genetic discoveries in amyotrophic lateral sclerosis: the ALSod and ALSGene databases. *Amyotroph Lateral Scler* (2011) 12(4):238–49. doi:10.3109/17482968.2011.584629
9. Zhao S, Guo Y, Sheng Q, Shyr Y. Advanced heat map and clustering analysis using heatmap3. *Biomed Res Int* (2014) 2014:986048. doi:10.1155/2014/986048
10. Feng G, Shaw P, Rosen ST, Lin SM, Kibbe WA. Using the bioconductor GeneAnswers package to interpret gene lists. *Methods Mol Biol* (2012) 802:101–12. doi:10.1007/978-1-61779-400-1_7
11. Bleazard T, Lamb JA, Griffiths-Jones S. Bias in microRNA functional enrichment analysis. *Bioinformatics* (2015) 31(10):1592–8. doi:10.1093/bioinformatics/btv023
12. Walter W, Sanchez-Cabo F, Ricote M. GOpot: an R package for visually combining expression data with functional analysis. *Bioinformatics* (2015) 31(17):2912–4. doi:10.1093/bioinformatics/btv300

ACKNOWLEDGMENTS

This work was supported by the Italian Ministry of Education, Universities and Research through grant CTN01_00177_817708. The authors gratefully acknowledge Cristina Cali, Alfia Corsino, Maria Patrizia D'Angelo, and Francesco Marino for their administrative and technical assistance. Research project: “DNA on Disk, Diagnostic platform and kits for human health in oncology, neurology, infectious diseases and poverty-related diseases.”

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fneur.2017.00152/full#supplementary-material>.

13. Feng G, Du P, Krett NL, Tessel M, Rosen S, Kibbe WA, et al. A collection of bioconductor methods to visualize gene-list annotations. *BMC Res Notes* (2010) 3:10. doi:10.1186/1756-0500-3-10
14. Bendotti C, Bao Cutrona M, Cheroni C, Grignaschi G, Lo Coco D, Peviani M, et al. Inter- and intracellular signaling in amyotrophic lateral sclerosis: role of p38 mitogen-activated protein kinase. *Neurodegener Dis* (2005) 2(3–4):128–34. doi:10.1159/000089617
15. Carter BJ, Anklesaria P, Choi S, Engelhardt JF. Redox modifier genes and pathways in amyotrophic lateral sclerosis. *Antioxid Redox Signal* (2009) 11(7):1569–86. doi:10.1089/ARS.2008.2414
16. Shao H, He Y, Li KC, Zhou X. A system mathematical model of a cell-cell communication network in amyotrophic lateral sclerosis. *Mol Biosyst* (2013) 9(3):398–406. doi:10.1039/c2mb25370d
17. Zhao W, Beers DR, Appel SH. Immune-mediated mechanisms in the pathoprogession of amyotrophic lateral sclerosis. *J Neuroimmune Pharmacol* (2013) 8(4):888–99. doi:10.1007/s11481-013-9489-x
18. Billaud M, Lohman AW, Johnstone SR, Biwer LA, Mutchler S, Isakson BE. Regulation of cellular communication by signaling microdomains in the blood vessel wall. *Pharmacol Rev* (2014) 66(2):513–69. doi:10.1124/pr.112.007351
19. Babu GN, Kumar A, Chandra R, Puri SK, Kalita J, Misra UK. Elevated inflammatory markers in a group of amyotrophic lateral sclerosis patients from northern India. *Neurochem Res* (2008) 33(6):1145–9. doi:10.1007/s11064-007-9564-x
20. Al-Chalabi A, Andersen PM, Nilsson P, Chioza B, Andersson JL, Russ C, et al. Deletions of the heavy neurofilament subunit tail in amyotrophic lateral sclerosis. *Hum Mol Genet* (1999) 8(2):157–64. doi:10.1093/hmg/8.2.157
21. Chen S, Sayana P, Zhang X, Le W. Genetics of amyotrophic lateral sclerosis: an update. *Mol Neurodegener* (2013) 8:28. doi:10.1186/1750-1326-8-28

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Morello, Spampinato and Cavallaro. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.