



## Research article

# Inferring gene regulatory networks of ALS from blood transcriptome profiles

Xena G. Pappalardo <sup>a,1</sup>, Giorgio Jansen <sup>a,b,1</sup>, Matteo Amaradio <sup>a</sup>, Jole Costanza <sup>c</sup>, Renato Umeton <sup>d</sup>, Francesca Guarino <sup>a,f</sup>, Vito De Pinto <sup>a,f</sup>, Stephen G. Oliver <sup>b,\*\*</sup>, Angela Messina <sup>e,f,\*</sup>, Giuseppe Nicosia <sup>a,b,\*\*\*</sup>

<sup>a</sup> Department of Biomedical and Biotechnological Sciences, University of Catania, Catania, Italy

<sup>b</sup> Department of Biochemistry, University of Cambridge, Cambridge, UK

<sup>c</sup> The National Institute of Molecular Genetics "Romeo and Enrica Invernizzi", Milano, Italy

<sup>d</sup> Department of Informatics and Analytics, Dana-Farber Cancer Institute, Boston, MA, USA

<sup>e</sup> Department of Biological, Geological and Environmental Sciences, University of Catania, Catania, Italy

<sup>f</sup> National Institute of Biostructures and Biosystems, Section of Catania, Catania, Italy

## ARTICLE INFO

## Keywords:

Amyotrophic lateral sclerosis (ALS)  
Algorithm for the reconstruction of accurate cellular networks - adaptive partitioning version (ARACNe-AP)  
Gene expression profile (GEP)  
Gene regulatory network (GRN)  
Reverse engineering methods

## ABSTRACT

One of the most robust approaches to the prediction of causal driver genes of complex diseases is to apply reverse engineering methods to infer a gene regulatory network (GRN) from gene expression profiles (GEPs). In this work, we analysed 794 GEPs of 1117 human whole-blood samples from Amyotrophic Lateral Sclerosis (ALS) patients and healthy subjects reported in the GSE112681 dataset. GRNs for ALS and healthy individuals were reconstructed by ARACNe-AP (Algorithm for the Reconstruction of Accurate Cellular Networks - Adaptive Partitioning). In order to examine phenotypic differences in the ALS population surveyed, several datasets were built by arranging GEPs according to sex, spinal or bulbar onset, and survival time. The designed reverse engineering methodology identified a significant number of potential ALS-promoting mechanisms and putative transcriptional biomarkers that were previously unknown. In particular, the characterization of ALS phenotypic networks by pathway enrichment analysis has identified a gender-specific disease signature, namely network activation related to the radiation damage response, reported in the networks of bulbar and female ALS patients. Also, focusing on a smaller interaction network, we selected some hub genes to investigate their inferred pathological and healthy subnetworks. The inferred GRNs revealed the interconnection of the four selected hub genes (*TP53*, *SOD1*, *ALS2*, *VDAC3*) with p53-mediated pathways, suggesting the potential neurovascular response to ALS neuroinflammation. In addition to being well consistent with literature data, our results provide a novel integrated view of ALS transcriptional regulators, expanding information on the possible mechanisms underlying ALS and also offering important insights for diagnostic purposes and for developing possible therapies for a disease yet incurable.

\* Corresponding author. Department of Biological, Geological and Environmental Sciences, University of Catania, Catania, Italy.

\*\* Corresponding author.

\*\*\* Corresponding author. Department of Biomedical and Biotechnological Sciences, University of Catania, Catania, Italy.

E-mail addresses: [sgo24@cam.ac.uk](mailto:sgo24@cam.ac.uk) (S.G. Oliver), [mess@unict.it](mailto:mess@unict.it) (A. Messina), [giuseppe.nicosia@unict.it](mailto:giuseppe.nicosia@unict.it) (G. Nicosia).

<sup>1</sup> First authors.

<https://doi.org/10.1016/j.heliyon.2024.e40696>

Received 20 July 2024; Received in revised form 21 November 2024; Accepted 24 November 2024

Available online 26 November 2024

2405-8440/© 2024 Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

A major challenge in transcriptomic research involves inferring the relationships among transcription factors (TFs) and their targets from gene expression profiles (GEPs) by making use of reverse engineering methods [1–3]. Reverse engineering is the process of deconstructing and analysing a pre-existing system (natural or artificial) to elucidate its underlying design, functionality, and components, with the aim of comprehending its design principles, functionality, and/or structure. It is often also for the purpose of replicating, enhancing, or redesigning the system. In the field of genomics, reverse engineering refers to the analytical process of deducing the regulatory networks and functional relationships among genes, proteins, and other molecular components within a given biological system. This computational approach involves the integration of experimental data, computational pipelines, and statistical techniques to infer the complex interactions that govern gene expression, protein-protein interactions, and metabolic/developmental pathways. Genomics-oriented reverse engineering aims to uncover the underlying biological mechanisms and signalling cascades, often with the goal of understanding pathological states and processes, identifying therapeutic targets, or elucidating the dynamics of cellular systems. This process is fundamental to unravelling the complexities of biological systems and advancing precision medicine.

Through this approach, the connectivity among naturally occurring genes is reconstructed in a large-scale gene regulatory network (GRN). Inferring a GRN from GEPs is useful not only for approximating patterns of gene regulation, but also for characterizing the impact that changes in gene expression may have on the structural and functional control of the GRN in response to certain cellular and environmental stimuli [4–6]. GRN analysis involves the reconstruction of the GRNs that represent interactions (the edges) between TFs and target genes (the nodes), enabling researchers to infer regulatory relationships and identify key drivers of disease processes. To date, GRN inference has yielded important insights into the molecular mechanisms of some complex and multifactorial diseases, such as cancers and neurodegenerative diseases [7–12]. In particular, the use of GRNs offers particular advantages for the elucidation of the mechanisms of some rare diseases such as amyotrophic lateral sclerosis (ALS). These present severe problems for clinicians, such as misleading or late diagnosis, high mortality rates, lack of therapy, or poor data management and patient selection. We contend that the use of GRNs to study complex and multifactorial diseases such as ALS offers great opportunities, allowing deregulated pathways, biomarkers, and potential therapeutic targets to be identified [13]. ALS is a heterogeneous neurodegenerative disorder caused by the degeneration of motoneurons [14]. The majority of cases (90 %), have no familial correlation (sALS) while the remaining 10 % of ALS cases are hereditary (fALS) [14]. One of the hallmarks of ALS is the presence of several inclusions formed by apparently disparate misfolded proteins, which are unrelated in sequence, structure, function, and localization [15]. ALS predominately affects rather than women (3:1 ratio) with survival rates depending on clinical manifestations and genetic defects. In fact, survival ranges from 2 to 5 years from onset in most patients to 10 years or more in about 10 % of cases [16]. Currently, more than 30 candidate genes have been identified as being involved in ALS pathogenesis (<http://alsod.iop.kcl.ac.uk/>). Four major genes are associated with the onset of ALS: *SOD1*, *TDP-43*, *FUS* and *C9orf72* [17]. These are joined by more than 20 other minor genes (e.g. *ALS2*, *SETX*, *VAPB*, *FIG4*, *ERBB4*, *MATR3*, *ANG*, *OPTN*, *VCP*, *UBQLN2*, *CHMP2B*, *PFN1*, *hNRP1 A2/B1*, *TUBA4A*, *NEK1*, *UNC12A*, *ANXA11*) [18–20]. Mutations in these genes are responsible for about 70 % of fALS and about 12 % of sALS [18]. ALS can have a spinal onset (SALS), which is more frequent and has better survival prospects, or a bulbar onset (BALS), which is less frequent, and results in shorter survival times [21]. However, the two forms are difficult to distinguish in the early stages of the disease. In addition to these issues, and genetic variability between the sporadic and familial forms of ALS, there is a lack of definite information on factors influencing risk, rate of disease progression, and survival. These problems are compounded by gender differences and site of disease onset (SALS and BALS) [21]. It is likely that not all genes directly and/or indirectly involved in the disease have been identified to date. Also to be considered the limitations of preclinical animal models in advancing our understanding of ALS, a disease that in 90 % of cases has sporadic onset. Among the criticisms raised of animal models those of not taking into account the multifactorial nature of the diseases and thus not recapitulating the complexity of human neurodegenerative diseases [22]. However, most of our mechanistic knowledge about ALS comes from studies in animal models with a specific and well-defined genetic cause. The proteins encoded by ALS genes have apparently different functions, which prevents drawing a single picture representing the pathological mechanism of ALS. Therefore, targeting predictive methods to find links between the known mechanisms of ALS may represent a great opportunity for the expansion of mechanistic knowledge about the disease.

By reconstructing networks from diverse patient samples (and control samples), GRN analysis can identify regulatory modules that are consistently altered in ALS, regardless of the specific genetic mutations involved [23]. To appreciate and assess the value of GRN analysis in ALS study, it is important to compare it with other commonly used methods, such as differential expression analysis, weighted gene co-expression network analysis (WGCNA), and genome-wide association studies (GWAS). While differential expression analysis identifies individual genes that are up- or down-regulated, it lacks insight into gene interactions and is often noisy due to ALS heterogeneity. WGCNA analysis reveals gene clusters but does not capture causal relationships, leading to potential misinterpretations. GWAS identifies genetic variants linked to ALS but struggles to explain their functional role in disease progression. In contrast, GRN analysis uncovers regulatory interactions between genes, providing a systems-level view of gene regulation. This enables the identification of key regulatory nodes and understanding how genetic variants influence gene expression and disease pathways, complementing GWAS findings. Hence, compared with the traditional methods like differential expression analysis, co-expression network analysis and GWAS, our approach, based on GRN analysis by ARACNe-AP, provides a more mechanistic and comprehensive understanding of gene regulation in ALS [8,23,24].

The aim of this research work was to elucidate some still poorly understood aspects of the ALS by focusing on as yet unknown gene interconnections. Using a reverse engineering approach, we performed a large-scale *in silico* analysis [25,26] that allowed us to map the network of interactions between TFs that control ALS gene expression and identify novel gene interactions in the ALS transcriptional program.

## 2. Results

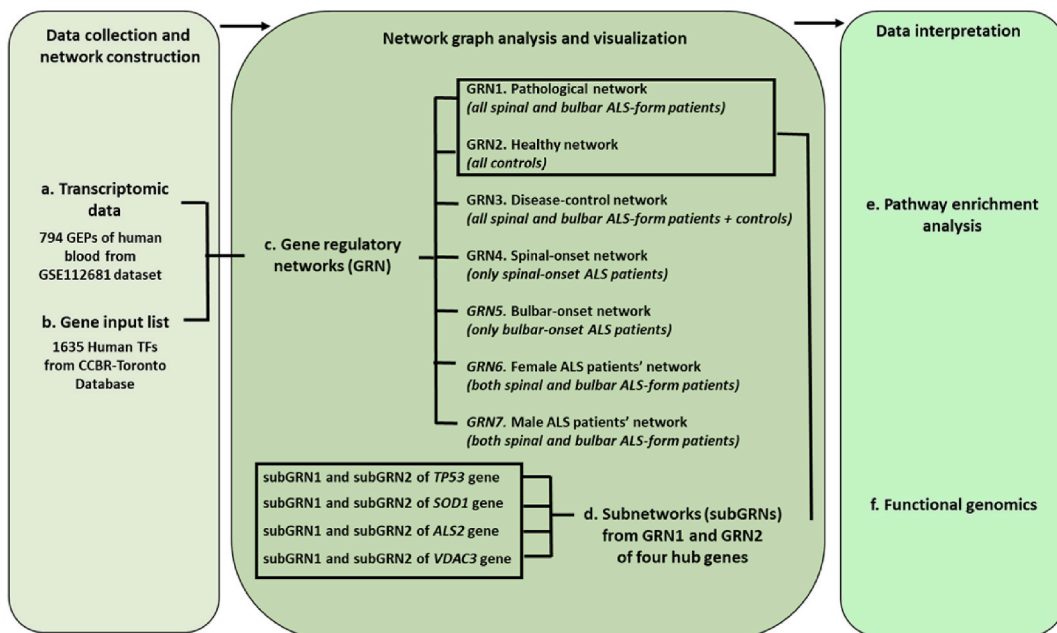
### 2.1. Characterization of ALS phenotypic networks

The workflow to build the GRNs and subnetworks (subGRNs) from 794 ALS blood GEPs reported in GSE112681 dataset [25,26] is schematically represented in Fig. 1 a–f.

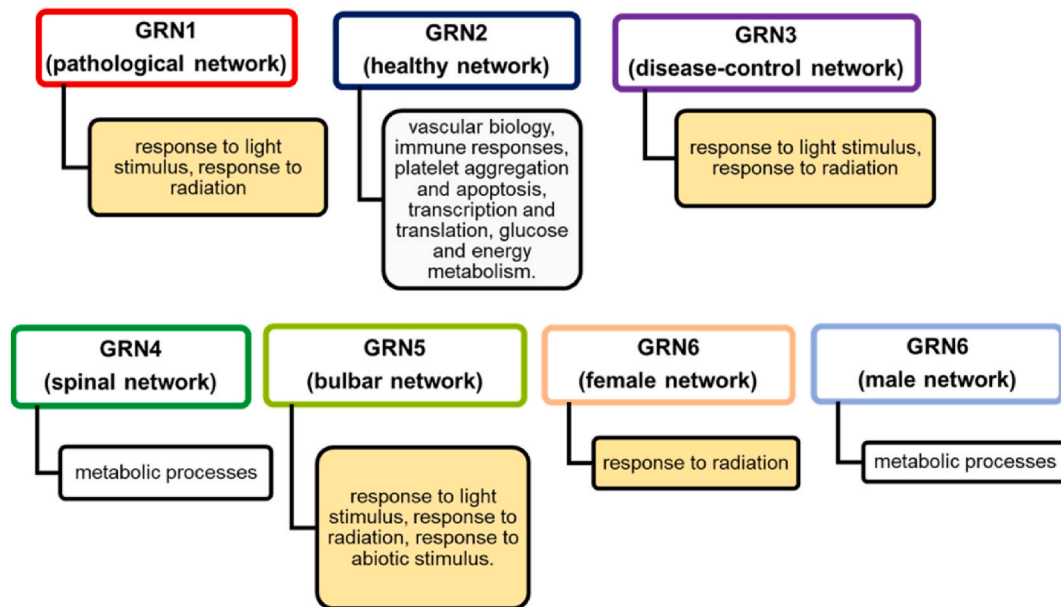
All gene regulatory networks and subnetworks identified from ARACNe-AP were produced by processing data on 1635 human TFs (Table S1) [27]. The GO database was then used to compare the activity of the pathways associated with the different ALS phenotypic networks, obtained by ARACNe-AP and filtered with 95th percentile (perc) of mutual information (MI) (see 5. *Materials and Methods*). We thus identified seven lists of hub genes (Table S2) and the following ALS-associated networks: the ‘pathological’ (GRN1) versus the ‘healthy’ (GRN2), the ‘disease-control’ (GRN3), the ‘spinal onset – SALS’ (GRN4), the ‘bulbar onset - BALS’ (GRN5), the ‘female’ (GRN6) and the ‘male’ (GRN7) network. More details can be found in the hub gene lists (Table S2) and GRN1-7 plots (Figs. S1–S2).

The GO pairwise comparison between the pathological and healthy network revealed protein metabolism and radiation damage response among the top pathways in the pathological network (GRN1) (Fig. 2; Table 1 a–b). In contrast, the biological processes in the healthy network (GRN2) (Fig. 2; Table 1 a) were related to vascular biology involving several immune responses, platelet aggregation and apoptosis, transcription and translation, glucose and energy metabolism. These findings are consistent with the GO terms of another study focused on characterizing the normal heterogeneity of gene expression in blood cells [28]. Furthermore, as the characteristics of the data used to train the ARACNe-AP model have a huge influence on the information that can be obtained from the gene relationships [23], we separated ALS samples of both sexes in two groups, spinal-SALS and bulbar-BALS, to prepare GRN4-5 (Table 1 b), and then we analysed the opposite clustering of GEPs, assembling female and male patients of both at disease-onset, respectively in GRN6-7 (Fig. 2; Table 1 c).

In these four networks (GRN4-7), GEPs of ALS patients were further arranged for survival time (<1 year, from 1 to 3 years, from 3 to 5 years, from 5 to 10 years), as schematically represented in Figs. S1a–b (see 5. *Materials and Methods*). The stratification of ALS patients was aimed at identifying the most different pathways mediating the spinal and bulbar ALS phenotypes, as well as the otherwise undetectable gender response to disease onset. Another issue to consider when interrogating eventual gender modulation or gender prevalence in spinal or bulbar ALS trajectories is the possible bias of qualitative analysis of biochemical pathways related to sex hormones regulation [29]. For this reason, the hub gene lists of GRN6-7 (Table S2) were cleared of genes implicated in gonadal differentiation [29] (see 5. *Materials and Methods*). Interestingly, the pathway analysis showed no significant differences between female and male patients (Table 1 c), with the exception of the radiation response in the female network (Fig. 2). Furthermore, the hub genes (Table S2) were compared with the pattern of ALS survival-associated upregulated genes identified by Swindell et al. [25], to identify a functional link between survival risk markers and our transcriptional regulators. We thus found in the spinal, bulbar and



**Fig. 1a–f.** Research pipeline of ALS consensus network building. The algorithm ARACNe-AP used as input files: (a) 794 GEPs from GSE112681 dataset, and (b) a list of regulators comprising human TFs (see Table S1) to build: (c) GRN1-7, pruned by 95th percentile of mutual information; and, (d) four subnetworks of some hub genes selected among top 5% (*TP53*) and among 30% (*SOD1*, *ALS2*, *VDAC3*) of the GRN1 and GRN2. Data analysis and interpretation was carried out by genomic approaches of (e) biological process enrichments for GRN1-7 and (f) functional gene clustering for hub genes to investigate underlying pathophysiological mechanisms and their normal counterparts.



**Fig. 2. Representation of key findings of pathway analysis of GRN hub genes.** In shaded boxes, it is highlighted the most significant difference obtained from the GO pairwise comparison (see Table 1a–c) associated with the ‘response to light/radiation’ emerged in certain phenotypic cluster assembled in GRNs.

female networks (GRN4–6) two genes (*SLC40A1*, *CLEC7A*) out of the 30 upregulated genes associated with worse prognosis and no targets among the 31 upregulated genes associated with good prognosis. The male network (GRN7) was represented only by the *SLC40A1* gene. Double crossover was detected in GRN1 and GRN3, while in the ‘healthy network’ GRN2 the *VRK3* gene was evaluated as associated with good prognosis.

A first analysis performed by comparing the filtered hub gene lists (Table S2) shows numerical variation between the different groups: GRN1 (n = 1313), GRN2 (n = 1305), GRN3 (n = 1381), GRN4 (n = 1339); GRN5 (n = 1239); GRN6 (n = 1415), GRN7 (n = 1259). As expected, most genes are shared between GRNs, with some genes specific to GRN1 (n = 7), GRN2 (n = 251), GRN3 (n = 23), GRN4 (n = 18); GRN5 (n = 46); GRN6 (n = 91), GRN7 (n = 21). These differences could underlie the observed variations in GRN morphology (Figs. S1–2). Specifically, the shape of the pathological gene regulatory pattern (GRN1) looks like an inverted ‘L’ with an oblong central stem and a small upward folded leg. Several small lateral branches of nodes are also evident, including a more descending branch of genes. Some large nodes (red circles) are scattered in the central stem. The shape of the normal pattern (GRN2) results from a mainly horizontal distribution, with a rounded central body, nodes on the inside and a flexed lateral tail. The disease control network (GRN3) retains the pathological shape and healthy branches appear to have been lost or heavily pruned (Fig. S1).

The spinal (GRN4) and bulbar (GRN5) networks appear similar, both having a symmetrical division with a cluster of hubs on each side, but the bottleneck is more densely connected in the spinal network than in the bulbar network (Fig. S2). Comparing the graphical view of the female (GRN6) and male (GRN7) networks, a number of differences become apparent: (i) a dividing line and a narrower constraint point in the male network (GRN7); (ii) in the female network (GRN6) the branches are mirrored, with a more regular distribution than in the male ~~SLA~~ ALS network developed on one side only (Fig. S2). The analysis of geometric differences in the spatial organization of GRNs may suggest concepts describing the pairwise relationship between the shape of network and their gene connections associated with ontological enrichments.

## 2.2. Analysis of *TP53*, *SOD1*, *ALS2* and *VDAC3* sub-networks

Surprisingly, in the top 5 % of hub genes in networks generated from ALS patient data, we found no known ALS-associated genes except for the tumour suppressor gene *TP53* (OMIM#191170). Although the *TP53* function has been extensively associated to neoplastic processes, this gene is also known to play a role in neurodegeneration, resulting in neural apoptosis in response to several insults [30,31]. In addition, recent studies showed that *TP53* is a central driver in ALS associated with mutant *SOD1* or *C9orf72* [32,33] and, more broadly, that it is included in the ALS transcriptional signature [34]. Next, searching among hub genes for known ALS-associated genes, we found, in the top 30 % of the GRNs, the superoxide dismutase 1 gene (*SOD1*; OMIM#147450) and the Alsin Rho guanine nucleotide exchange factor *ALS2* gene (*ALS2*; OMIM#606352) [35]. The plot of *SOD1* subnetwork revealed no common targets between pathological and healthy conditions. In particular, subGRN1 showed 14 new putative disease targets, all strongly implicated in morphological development and cell viability, while several known *SOD1*-interacting partners involved in mitochondrial metabolism were found in the healthy subGRN2. In contrast, most of the *ALS2*-interacting genes found in both sub-networks have never been associated with ALS. To sum up, we identified 8 new genes out of 11 in subGRN1 and another 4 new genes, encoding

**Table 1a-c**

Pathway analysis of: a. Pathological (GRN1), healthy (GRN2) and disease-control network (GRN3); b. spinal-onset (GRN4) and bulbar-onset ALS network (GRN5); c. female (GRN6) and male ALS patients' network (GRN7). GO analysis of hub gene list (Table S2) calculated with 95th perc of MI of ARACNe-AP algorithm was performed using Fisher's exact test and Bonferroni correction for  $P < 0.05$ ; Panther Overrepresentation test (release 20210224; Panther v.16.0 released 2020-12-01). In bold, the first ranked GO pathways of interest.

	GO Biological Process	Ref. list	Our list	Expected	Fold Enrichment	+/-	P value	
<b>A</b>								
<b>Pathological network (GRN1)</b>	<b>response to light stimulus</b>	317	<b>46</b>	19,81	2,32	+	1,20E-02	
	<b>response to radiation</b>	453	<b>59</b>	28,31	2,08	+	7,56E-03	
	negative regulation of metabolic process	2952	249	184,47	1,35	+	1,70E-02	
	negative regulation of biological process	5187	405	324,14	1,25	+	8,08E-03	
	regulation of metabolic process	6586	496	411,57	1,21	+	1,46E-02	
	cellular protein metabolic process	3219	268	201,16	1,33	+	1,75E-02	
	cellular macromolecule metabolic process	4442	353	277,58	1,27	+	1,43E-02	
	cellular process	15074	1060	941,99	1,13	+	5,55E-10	
	protein metabolic process	3807	308	237,9	1,29	+	2,20E-02	
	primary metabolic process	6869	511	429,25	1,19	+	3,43E-02	
	<b>Healthy network (GRN2)</b>	regulation of transport	1732	158	107,98	1,46	+	4,34E-02
		regulation of biological process	11528	806	718,72	1,12	+	1,52E-02
		biological regulation	12249	848	763,67	1,11	+	2,72E-02
		multicellular organismal process	6584	494	410,48	1,2	+	1,67E-02
regulation of cellular process		11076	777	690,54	1,13	+	2,25E-02	
cellular process		15074	1054	939,79	1,12	+	3,08E-09	
<b>Disease-control network (GRN3)</b>	<b>response to light stimulus</b>	317	<b>47</b>	20.84	2.26	+	1.38E-02	
	<b>response to radiation</b>	453	<b>60</b>	29.78	2.01	+	1.54E-02	
	cellular protein metabolic process	3219	281	211.63	1.33	+	1.20E-02	
	cellular macromolecule metabolic process	4442	371	292.04	1.27	+	9.11E-03	
	cellular process	15074	1122	991.03	1.13	+	3.63E-12	
	macromolecule metabolic process	5743	458	377.57	1.21	+	3.53E-02	
	primary metabolic process	6869	538	451.60	1.19	+	1.83E-02	
	negative regulation of biological process	5187	420	341.01	1.23	+	2.73E-02	
	regulation of biological process	11528	849	757.90	1.12	+	1.16E-02	
	biological regulation	12249	899	805.30	1.12	+	3.58E-03	
	regulation of metabolic process	6586	521	432.99	1.20	+	1.05E-02	
	regulation of cellular process	11076	820	728.18	1.13	+	1.13E-02	
	<b>B</b>							
<b>Spinal network (GRN4)</b>	cellular protein metabolic process	3219	274	205,85	1,33	+	1,29E-02	
	cellular macromolecule metabolic process	4442	367	284,06	1,29	+	1,85E-03	
	cellular process	15074	1087	963,95	1,13	+	8,27E-11	
	macromolecule metabolic process	5743	451	367,25	1,23	+	1,02E-02	
	protein metabolic process	3807	317	243,45	1,3	+	9,56E-03	
	primary metabolic process	6869	530	439,26	1,21	+	3,90E-03	
	negative regulation of metabolic process	2952	251	188,77	1,33	+	4,77E-02	
	regulation of metabolic process	6586	507	421,16	1,2	+	1,21E-02	
	<b>Bulbar network (GRN5)</b>	<b>response to light stimulus</b>	317	<b>44</b>	18,69	2,35	+	1,03E-02
		<b>response to radiation</b>	453	<b>55</b>	26,7	2,06	+	2,63E-02
<b>response to abiotic stimulus</b>		1125	<b>108</b>	66,31	1,63	+	2,39E-02	
response to organic substance		2688	219	158,45	1,38	+	1,52E-02	
negative regulation of metabolic process		2952	235	174,01	1,35	+	3,01E-02	
negative regulation of biological process		5187	389	305,75	1,27	+	1,91E-03	
regulation of metabolic process		6586	472	388,22	1,22	+	8,69E-03	
negative regulation of cellular process		4785	353	282,06	1,25	+	4,54E-02	
regulation of primary metabolic process		5707	412	336,41	1,22	+	3,69E-02	
cellular process		15074	996	888,56	1,12	+	1,81E-08	
<b>C</b>								
<b>Female network (GRN6)</b>		<b>response to radiation</b>	453	<b>60</b>	30,42	1,97	+	3,02E-02
	negative regulation of biological process	5187	426	348,32	1,22	+	4,67E-02	
	multicellular organismal process	6584	525	442,13	1,19	+	4,93E-02	
	cellular process	15074	1143	1012,25	1,13	+	9,27E-12	
	negative regulation of biological process	5187	426	348,32	1,22	+	4,67E-02	
	multicellular organismal process	6584	525	442,13	1,19	+	4,93E-02	
	cellular process	15074	1143	1012,25	1,13	+	9,27E-12	
<b>Male network (GRN7)</b>	cellular protein metabolic process	3219	263	193,5	1,36	+	4,04E-03	
	cellular macromolecule metabolic process	4442	342	267,02	1,28	+	1,09E-02	
	cellular process	15074	1010	906,12	1,11	+	1,89E-07	
	protein metabolic process	3807	302	228,85	1,32	+	5,47E-03	
	primary metabolic process	6869	495	412,91	1,2	+	2,24E-02	
	cellular protein metabolic process	3219	263	193,5	1,36	+	4,04E-03	
	cellular macromolecule metabolic process	4442	342	267,02	1,28	+	1,09E-02	

**Panther classification system:** The 2nd column contains the number of genes in the reference list that map to this annotation data category. The 3rd column contains the number of genes in your uploaded list that map to this annotation data category. The 4th column contains the expected value, which is the number of genes you would expect in your list for this category, based on the reference list. The 5th column shows the Fold Enrichment of the genes observed in the uploaded list over the expected (number in your list divided by the expected number). If it is greater than 1, it indicates that the category is overrepresented in your experiment. Conversely, the category is underrepresented if it is less than 1. The 6th column has either a + or -. A plus sign indicates over-representation of this category in your experiment: you observed more genes than expected based on the reference list (for this category, the number of genes in your list is greater than the expected value). Conversely, a negative sign indicates under-representation. The 7th column is the raw p-value as determined by Fisher's exact test.

zinc-finger proteins in subGRN2. Within the same cut-off point as the *SOD1* and *ALS2* genes, we also identified *VDAC3* (OMIM#610029), a gene encoding an important mitochondrial protein also thought to be related to neurodegeneration due to its role as a sensor of redox state and protection of mitochondria from oxidation [36,37]. Moreover, a putative impact of *VDAC3* in ALS has recently been highlighted by our group [37]. Briefly, the regulatory scenario related to *VDAC3* showed 12 genes in subGRN1, an additional 12 genes in subGRN2 and two common nodes. Considering the still limited knowledge about the regulation and transcriptional interactions of *VDAC3*, the inferred GRNs provide important information to define the role of this gene in the pathological context. Data produced by these analyses summarize the plot information of molecular interactions related to the first-neighborhood of four hub genes, distinguishing two types of results, red for the pathological sub-network (subGRN1) and blue for the healthy sub-network (subGRN2), further grouped as 'known' and 'novel' target genes, based on the available literature.

### 2.2.1. TP53 sub-networks

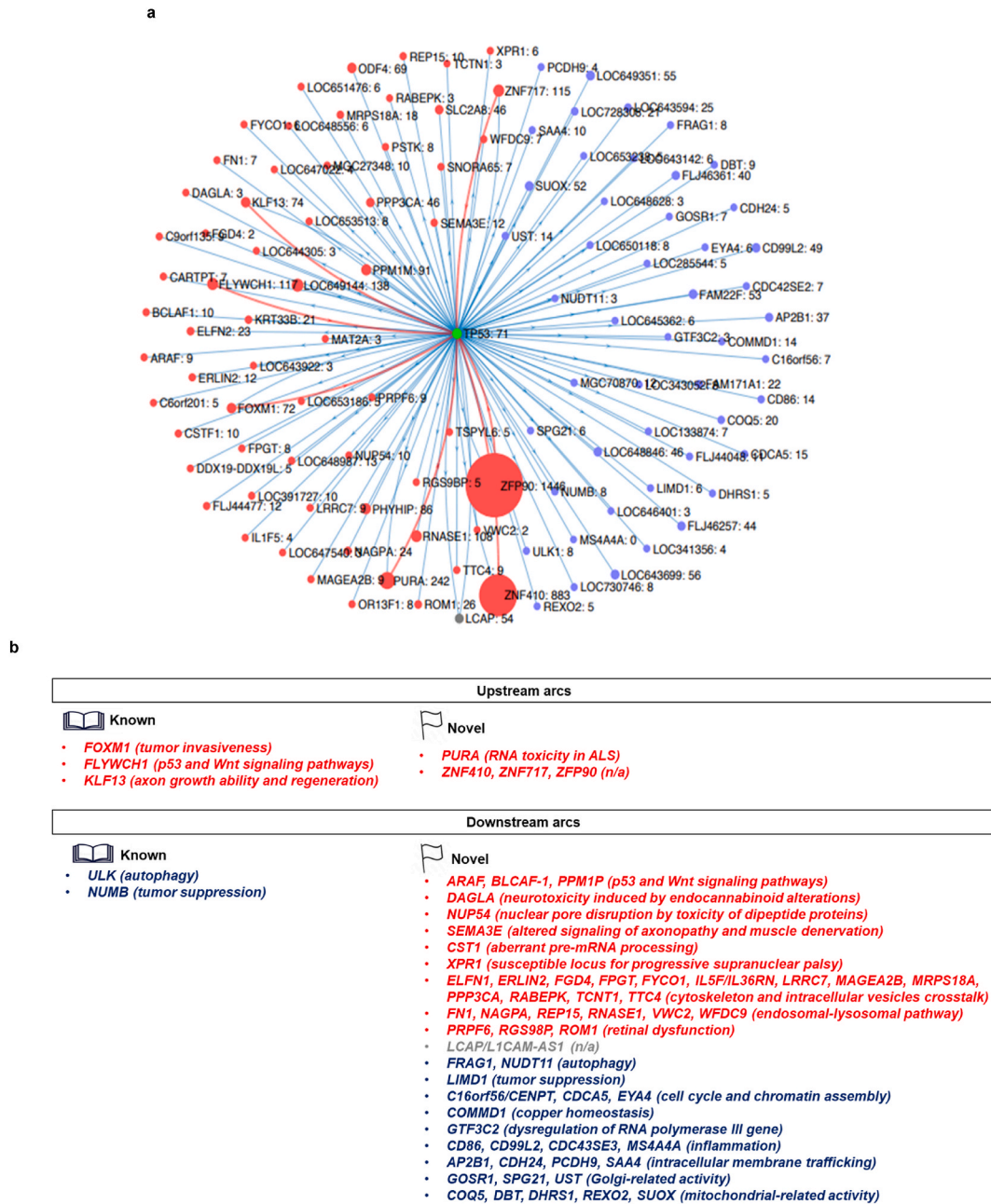
Consistent with the *TP53* gene's functions as both a transcriptional regulator (Table S1) and effector, two types of connections were deduced from this hub as *incoming edges*, i.e. a set of putative upstream genes acting as TFs, and *outgoing edges* representing p53 downstream targets (Fig. 3a–b). The analysis of *TP53*-related subGRN1 and subGRN2 revealed the articulated transcriptional dynamics of the disease. Moreover, most of the p53 targets in both sub-networks are unique and describe poorly investigated activities of *TP53* with respect to known mechanisms. Regarding the upstream genes identified by our analysis, 4/7 TFs have never been associated with ALS (*PURA*, *ZFP90*, *ZNF410*, *ZNF717*) (Fig. 3b), nor have their 43 downstream interactors. In the subGRN2 sub-network, 24/26 new downstream targets were identified and only one gene was a common target between the two sub-networks; this had also not been detected previously (Fig. 3b).

Seven TFs (*FOXM*, *FLYWCH1*, *KLF13*, *PURA*, *ZFP90*, *ZNF410*, *ZNF717*) whose transcripts have been identified in studies of ALS patients are found in our *TP53* sub-network (Fig. 3a–b). In addition to some zinc-finger factors, we detected the *PURA* (Purine Rich Element Binding Protein A) gene, encoding a highly conserved protein with regulatory roles in replication and transcription, RNA transport and mRNA translation. *PURA* is also involved in RNA foci formation, as recently found in ALS/FTD [38]. All downstream pathological targets observed were novel and most of their functions are associated with some neurodegeneration mechanisms, including ALS-related ones. For instance, we detected the genes encoding a-Raf, Bclaf1 and Ppm1m, known mediators of p53-related oncogenic signalling [39]. Other predicted interactions, although not directly related to *TP53*, are known to play a role in ALS. One of these involves the *DAGLA* gene, encoding diacylglycerol lipase alpha, the main enzyme involved in the biosynthesis of endocannabinoids, important for their antioxidant properties and action in multiple neurotoxic pathways and also related to ALS [40]. Another interesting finding concerns *NUP54*, the gene encoding nucleoporin 54 - a member of the nuclear pore complex, which mediates the shuttling of RNA cargoes, proteins, and TFs from the cytoplasm to the nucleus. Indications for the pathological cytoplasmic accumulation of RNA-binding proteins have been addressed due to defects in the nuclear import and Nup abnormalities in ALS and related neurodegenerative diseases [41].

In addition, the *SEMA3E* gene, belonging to the semaphorin gene family, encodes extracellular signalling proteins involved in the cytoskeleton organization and development of many tissues, especially in the nervous system for axonal growth. In particular, semaphorins have been extensively investigated in ALS pathobiology as candidates of axonopathy and muscle denervation [42]. Another member of the *TP53* downstream group, the *CSTF1* gene, encodes one of three subunits of the cleavage stimulation factor implicated in the polyadenylation and 3'-end cleavage of pre-mRNAs. In this regard, upregulation of *CSTF1* linked to altered RNA metabolism was revealed in the *SOD1*<sup>G86R</sup> mouse model [43]. Recently, it has been reported that *CSTF1* expression is significantly elevated in the peripheral blood of Alzheimer's disease (AD) patients compared with controls, proving its potential diagnostic value [44]. Also of interest is that *XPR1* was identified as a novel susceptibility locus for progressive supranuclear palsy, whose clinical manifestations overlap with ALS and Parkinson's disease (PD) [45]. In addition, in a study in an ALS *Drosophila melanogaster* model with defects in *FUS* protein, it was shown that knockdown of *Xrp1* in motor neurons rescues the phenotypes of ALS-mutant *FUS* flies [46].

For a clear presentation of the data, targets sharing some functional properties and cooperate in the same pathway were clustered in two main groups: the cytoskeleton and crosstalk between intracellular vesicles (*ELFN2*, *ERLIN2*, *FGD4*, *FPGT*, *FYCO1*, *IL5F/IL36* R N, *LRR7*, *MAGEA2B*, *MRPS18A*, *PPP3CA*, *RABEPK*, *TCTN1*, *TTC4*) and the endosomal-lysosomal pathway (*FN1*, *NAGPA*, *REP15*, *RNASE1*, *VWC2*, *WFDC9*). Both pathways, already associated with ALS, have recently come to attention, showing a synergistic connection between mechanisms and organelles (vesicular transport, Golgi, endolysosomal and mitochondrial systems) involved in the regulation of neuronal trafficking [47].

Overall, we can speculate that *TP53* interaction maps are congruent with altered blood-specific transcriptional regulation for two main reasons: (i) the impairment of the endolysosomal system implies a dysfunction of autophagy, considered one of the key p53-mediated signalling events, as already reported for several neurodegenerative disorders [48]; (ii) the accumulation of extracellular vesicles as vehicles of misfolded proteins detectable in ALS patients' biological fluids, such as in blood and cerebrospinal fluid;



**Fig. 3a–b.** The *TP53* subnetwork. (a) The *TP53* subnetwork was obtained by including all the genes that have  $P < 1.0 \times 10^{-8}$  based on their pairwise MI with *TP53*. For representation purposes, only the first neighbors are shown. Red nodes indicate the target genes interacting with *TP53* in the pathological network. Blue nodes represent the predicted interactions in the healthy network. Grey node is referred to common target in both networks. Red lines indicate the upstream inferred arcs in the pathological network (i.e. putative *TP53*-related TFs). The size of each circle is proportional to the number of the gene interactions. For hubs with more than 100 interactions, the exact number of first neighbors is shown beside the gene symbol.

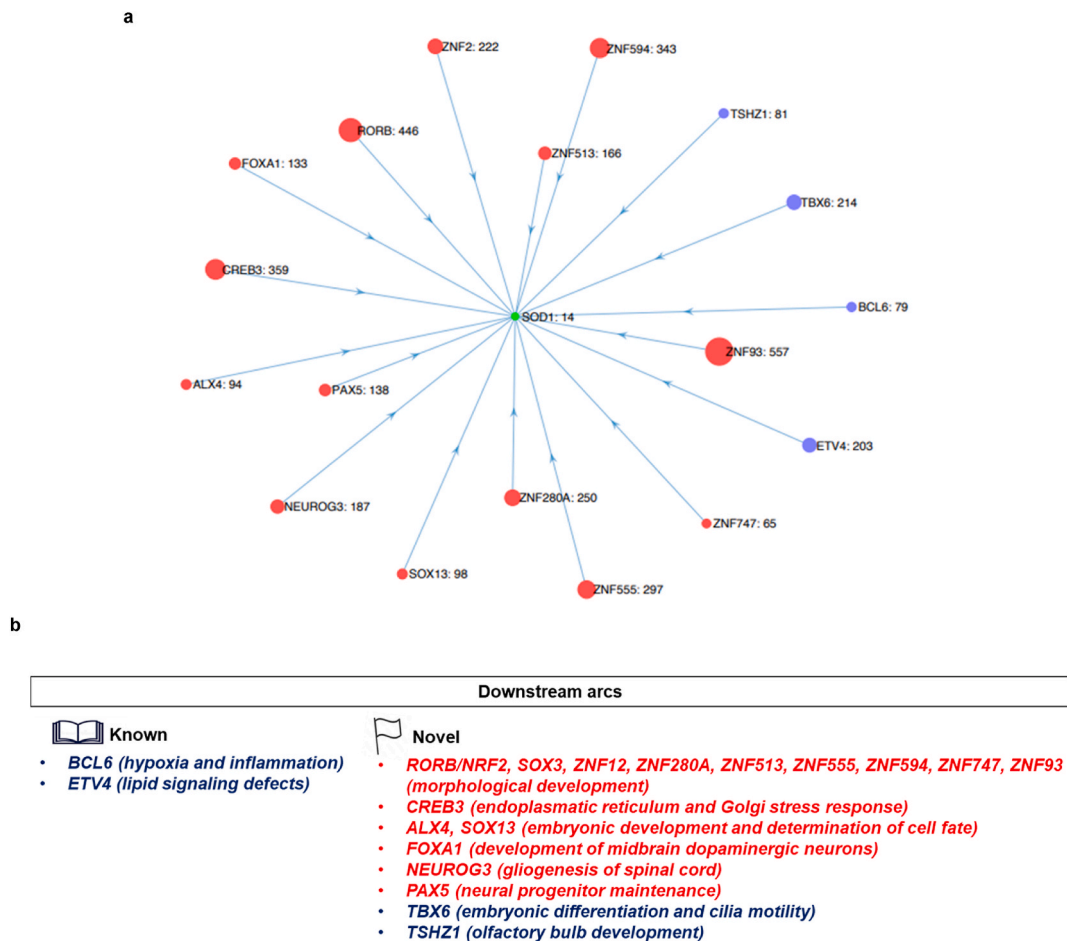
In subGRN1, *TP53* shows to be modulated by seven putative *TP53*-related TFs (*FLYWCH1*, *FOXM1*, *KLF13*, *PURA*, *ZFP90*, *ZNF410*, *ZNF717*), and in turn, to modulate 43 genes (*ARAF*, *BCLAF1*, *C6orf201*, *C9orf135*, *CARTPT*, *CSTF1*, *DAGLA*, *DDX19-DDX19L*, *ELFN2*, *ERLIN2*, *FGD4*, *FN1*, *FPGT*, *FYCO1*, *IL1F5*, *KRT33B*, *LRRRC7*, *MAGEA2B*, *MAT2A*, *MRPS18A*, *NAGPA*, *NUP54*, *ODF4*, *OR13F1*, *PHYHIP*, *PPM1M*, *PPP3CA*, *PRPF6*, *PSTK*, *RABEPK*, *REP15*, *RGS98P*, *RNASE1*, *ROM1*, *SEMA3E*, *SLC2A8*, *SNORA65*, *TCNT1*, *TSPYL6*, *TTC4*, *VWC2*, *WFDC9*, *XPR1*).

The subGRN2 contains 26 genes (*AP2B1*, *C16orf56*, *CD86*, *CD99L2*, *CDC42SE2*, *CDC45*, *CDH24*, *COMMD1*, *COQ5*, *DBT*, *DHRS1*, *EYA4*, *FRAG1*, *GOSR1*, *GTF3C2*, *LIMD1*, *MS4A4A*, *NUDT11*, *NUMB*, *PCDH9*, *REXO2*, *SAA4*, *SPG21*, *SUOX*, *ULK1* and *UST*). The only common target (grey node) between subGRN1 and subGRN2 is a non-coding RNA gene, *LCAP* (*L1CAM-AS1*).

(b) Analysis of known (book symbol) and novel (flag symbol) interacting partners in the pathological and healthy subnetworks were schematically summarized. n/a = not available information.

suggesting their use as clinical biomarkers [49].

As for the healthy subnetwork, although most of the *TP53*-interacting partners were unknown to date, the identified signalling pathways correlate well with many responses evoked by p53. Among these are indeed found autophagy, tumour suppression, cell cycle and chromatin assembly, dysregulation of RNA polymerase III gene and inflammation [39]. We also found the expected enrichment of interplaying organellar-secretory pathways, demonstrating the relevance of p53 in controlling survival mechanisms involving autophagy and endocytosis. An interesting target was the *COMMD1* gene, whose functions range from copper homeostasis and sodium transport to regulation of the NF- $\kappa$ B pathway and expression of hypoxia-inducible factor-1 $\alpha$  (*HIF-1 $\alpha$* )-mediated genes [50]. Although there is no evidence for a direct interaction with p53, several *COMMD1*-influenced factors may induce activation of the p53 pathway, such as neurotoxicity due to copper accumulation, copper-mediated oxidative stress through protein aggregation between *COMMD1* and mutated *SOD1*, and altered modulation with nuclear lamin A, leading to premature senescence [51]. Therefore, the failure to identify the *TP53*-*COMMD1* interaction in the pathological subnetwork allows us to postulate that altered regulation of copper homeostasis may be one of the main contributors in ALS pathogenesis, as well as one of the most studied mechanisms in *SOD1* mutant mice to reverse the signs of ALS disease by copper therapy [52]. The common downstream target of both subnetworks was *LCAP/LICAM-AS1* (lung carcinoma-associated 10), a novel long intergenic non-coding RNA (lncRNA). As few studies have investigated lncRNA-disease associations, we found only one transcriptional analysis (*E*-GEOD-59612) in the Open Targets Platform (<https://platform.opentargets.org/>), showing significant downregulation ( $\text{Log}_2$  -1.9; p-value  $2.28 \times 10^{-3}$ ) in controls compared with glioma patients [53].



**Fig. 4a–b.** The *SOD1* subnetwork. (a) The *SOD1* subnetwork was obtained by including all the genes that have  $P < 1.0 \times 10^{-8}$  based on their pairwise MI with *SOD1*. For representation purposes, only the first neighbors are shown. Red nodes indicate the target genes interacting with *SOD1* in the pathological network (subGRN1). Blue nodes represent the predicted interactions in the healthy network (subGRN2). The size of each circle is proportional to the number of the gene interactions. For hubs with more than 100 interactions, the exact number of first neighbors is shown beside the gene symbol. In subGRN1, *SOD1* shows to interact with 14 genes (*ALX4*, *CREB3*, *FOXA1*, *NEUROG3*, *PAX5*, *RORB*, *SOX13*, *ZNF2*, *ZNF280A*, *ZNF513*, *ZNF555*, *ZNF594*, *ZNF747*, *ZNF93*), and with 4 genes (*BCL6*, *ETV4*, *TBX6*, *TSHZ1*) in the subGRN2. (b) Analysis of known (book symbol) and novel (flag symbol) interacting partners in the pathological and healthy subnetworks were schematically summarized.



### 2.2.2. SOD1 sub-networks

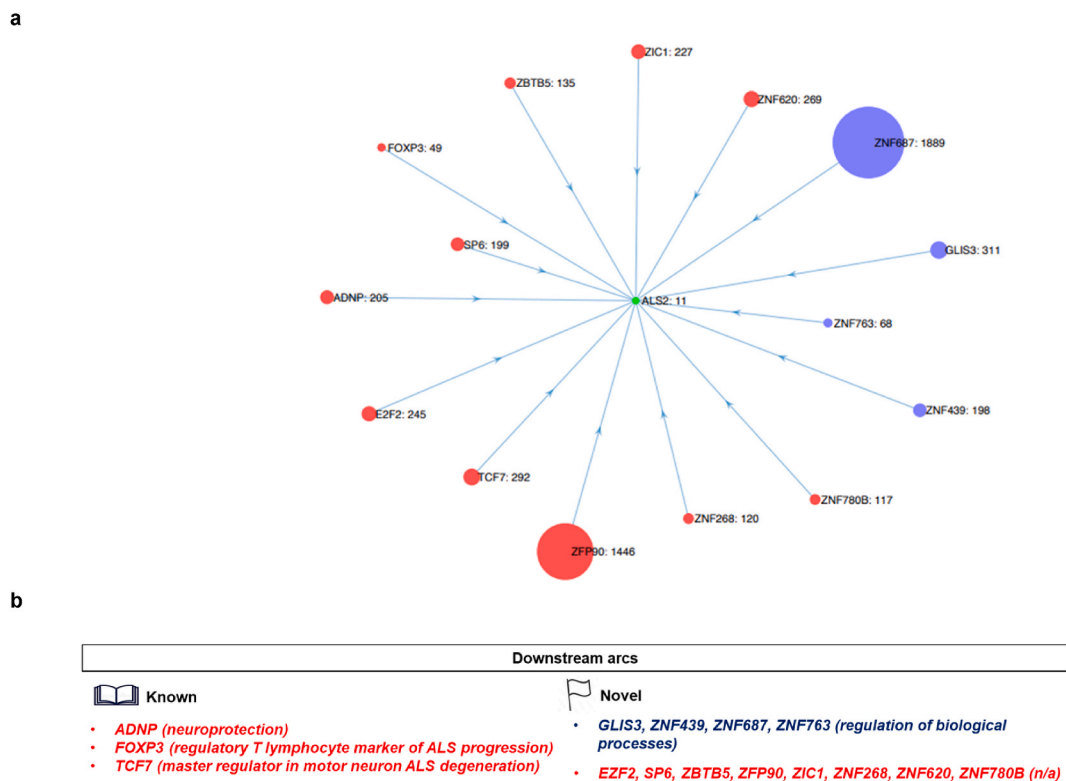
In the *SOD1* pathological subnetwork, we found a remarkable enrichment of genes associated with pathways related to neuronal development, namely: *ALX4*, *FOXA1*, *NEUROG3*, *PAX5*, *RORB/NR1F2*, *SOX13*, *ZNF2*, *ZNF280A*, *ZNF513*, *ZNF555*, *ZNF594*, *ZNF747* and *ZNF93* (Fig. 4a–b). In contrast, genes involved in the role of counteracting oxidative stress, which relate to the main physiological function performed by the *SOD1* gene, are mostly found in the healthy subnetwork (Fig. 4a–b). This evidence is compatible with the notion that supersaturated proteins, like *SOD1*, may be ~~are~~ involved in ALS due to their propensity to aggregate, rather than because of their biological function [15]. In this regard, it is known that wild-type *SOD1* produces misfolded aggregates in spinal cord of ALS patients [54].

Considering *FOXA1*, this gene is known to regulate midbrain dopaminergic neuron development [55], and the *NEUROG3* gene controls the spinal cord gliogenesis [56].

Evidence of their interaction has been provided and, in particular, co-expression of *Neurog3* and *Foxa2* has been observed in the mouse ventral spinal cord, probably aimed at controlling the correct specification of ventral progenitors of serotonergic neurons.

[57]. The *PAX5* gene, known for its involvement in the maintenance of neural progenitors in the human brain, has been identified as a marker of selectively vulnerable excitatory neurons in the entorhinal cortex, also expressed in neurofibrillary inclusions of AD patients [58]. Regarding the *CREB3* gene, it is strongly associated with the expression of several genes involved in endoplasmic reticulum (ER) and Golgi stress response [59]. In contrast, *ALX4* and *SOX13* are genes implicated in embryonic development and cell fate determination [60]. Taken together, these findings underscore a lesser known activity of *SOD1* as a putative neuromodulator of synaptic transmission and activation of the transduction pathway involving muscarinic acetylcholine receptors, the most abundant receptors type in the CNS [61]. Moreover, it is noteworthy that the calculated interactive pattern was not based on the mutated form of *SOD1*, but the gene relationships detected involved the *SOD1* wild type, as a hub gene, and its binding partners that are presumably affected in the pathological condition. This could help to understand the possible steps that *SOD1* is able to move in the molecular chessboard, especially in light of the pathogenic implication of wild-type *SOD1* in ALS cases not linked to *SOD1* mutations [62].

In contrast, genes involved in the role of counteracting oxidative stress, which is consistent with the main physiological function performed by the *SOD1* gene, are mostly found in the healthy subnetwork (Fig. 4a–b). Among them, we found the *BCL6* and *ETV4*

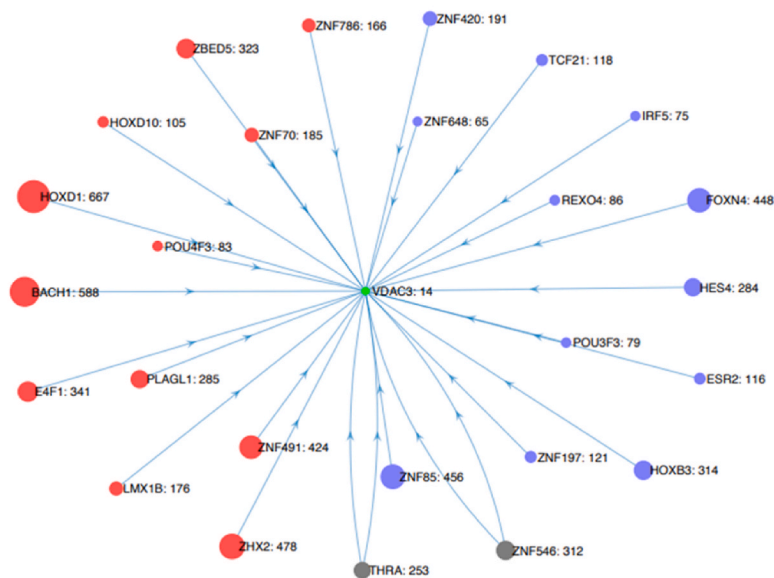


**Fig. 5a–b.** The *ALS2* subnetwork. (a) The *ALS2* subnetwork was obtained by including all the genes that have  $P < 1.0 \times 10^{-8}$  based on their pairwise MI with *ALS2*. For representation purposes, only the first neighbors are shown. Red nodes indicate the target genes interacting with *ALS2* in the pathological network (subGRN1). Blue nodes represent the predicted interactions in the healthy network (subGRN2). The size of each circle is proportional to the number of the gene interactions. For hubs with more than 100 interactions, the exact number of first neighbors is shown beside the gene symbol. *ALS2* shows to interact with 11 genes (*ADNP*, *E2F2*, *FOXP3*, *SP6*, *TCF7*, *ZBTB5*, *ZFP90*, *ZIC1*, *ZNF268*, *ZNF620*, *ZNF780B*) in subGRN1, and with 4 genes (*GLIS3*, *ZNF439*, *ZNF687*, *ZNF763*) in subGRN2. (b) Analysis of known (book symbol) and novel (flag symbol) interacting partners in the pathological and healthy subnetworks were schematically summarized. n/a = not available information.

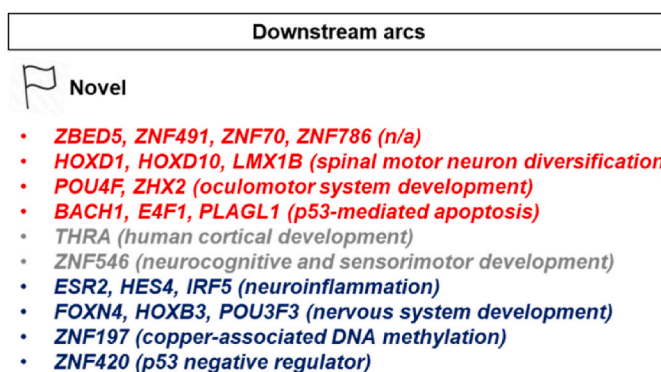
genes, which have not yet been identified as *SOD1* partners. The authors of the GSE112681 dataset [25,26] used in this study, identified the *BCL6* gene among the upregulated genes associated with hypoxia and inflammation. *BCL6*, in the context of inflammatory and cell death pathways, can determine both activation and inhibition of *TP53* transcription and/or p53 translation, depending on genomic instability [63]. The potential *BCL6-SOD1* complex could therefore represent a novel intermediate of p53-mediated mitochondrial apoptosis [64]. The role of *ETV4* has been more documented in the pathological than in the normal context. In particular, apoptosis seems to be triggered by upregulation of *ETV4* in response to lipid signalling defects, as observed in oligodendrocytes of *SOD1*<sup>G37R</sup> mice [65].

Therefore, we cannot exclude that the unusual interplay among these genes is an information inferred by ARACNe-AP simulations. In fact, not all impaired gene interactions lead to up or down-regulation, but sometimes they are missed in the pathological condition.

a



b



**Fig. 6a–b.** The *VDAC3* subnetwork. (a) The *VDAC3* subnetwork was obtained by including all the genes that have  $P < 1.0 \times 10^{-8}$  based on their pairwise MI with *VDAC3*. For representation purposes, only the first neighbors are shown. Red nodes indicate the target genes interacting with *VDAC3* in the pathological network (subGRN1). Blue nodes represent the predicted interactions in the healthy network (subGRN2). Grey nodes are the two common targets shared by both networks. The size of each circle is proportional to the number of the gene interactions. For hubs with more than 100 interactions, the exact number of first neighbors is shown beside the gene symbol. *VDAC3* shows to interact with 12 genes (*BACH1*, *E4F1*, *HOXD1*, *HOXD10*, *LMX1B*, *PLAGL1*, *POU4F3*, *ZBED5*, *ZHX2*, *ZNF491*, *ZNF70*, *ZNF786*) in subGRN1, and with 12 genes (*ESR2*, *FOXN4*, *HES4*, *HOXB3*, *IRF5*, *POU3F3*, *REXO4*, *TCF21*, *ZNF197*, *ZNF420*, *ZNF648*, *ZNF85*) in subGRN2. Only two common targets (grey nodes) between GRN1 and GRN2 are found (*THRA*, *ZNF546*). (b) Analysis of known (book symbol) and novel (flag symbol) interacting partners in the pathological and healthy subnetworks were schematically summarized. n/a = not available information.

### 2.2.3. ALS2 sub-networks

The regulatory framework obtained for the *ALS2* gene (Fig. 5a–b) shows that both subnetworks are enriched mainly in zinc-finger transcription factors (ZF-TFs) which control the expression of multiple genes involved in cell proliferation and cell fate.

Although the identified ZF-TFs are not all well characterized, we found some of them associated with a highly CNS-specific expression pattern. In particular, the *ZIC1* gene (expressed in the cerebellum [66]) and the *ZFP90* gene (expressed in the caudate nucleus and substantia nigra [67]) are found in subGRN1; while the *ZNF687* gene (expressed in the amygdala, corpus callosum, cerebellum, caudate nucleus, and spinal cord [67]) is located in subGRN2. Furthermore, the *ALS2* network analysis revealed at least three genes known to be associated with pathological events (*ADNP*, *FOXP3*, *TCF7*), supporting the validity of the predictions. In fact, there is ample evidence of the neurotrophic function of the *ADNP* (activity-dependent neuroprotective protein) gene, which is implicated in many neurodegenerative disorders (AD, PD and ALS) but also in neuropsychiatric disorders, such as schizophrenia and autism [68,69]. With regard to *FOXP3*, its upregulation in the T-helper cells of ALS patients has been defined as a potential marker of the early phase of the disease [70]. In our case, however, the original datasets did not distinguish ALS samples according to clinical progression, and so we were unable to prepare more time-specific matrices (see 5. *Material and Methods*). Nevertheless, this observation could be interpreted as a useful confirmation of blood-specific regulatory signals. Finally, the *TCF7* (Transcription Factor 7) gene has been described as one of the “23 potential master regulators in ALS motor neuron degeneration” [13].

### 2.2.4. VDAC3 sub-networks

The networks obtained for the *VDAC3* gene (Fig. 6a–b), represent the first attempt to characterize and interpret the regulatory pathways related to this gene.

In the pathological subGRN1 subnetwork, we found some genes specifically implicated in defining the characteristics of the spinal motor neurons (*HOXD1*, *HOXD10*, *LMX1B*) and in the development of the oculomotor system (*BRN3A/POU4F*, *ZHX2*) [71]. Indeed, *HOXD1* and *HOXD10* genes are known to coordinate the dorsal and lumbar specification of motor neurons [72] (Fig. 6a–b). *LMX1B* is an evolutionarily conserved TF that, in association with *LMX1A*, orchestrates the complex temporal and spatial gene expression of spinal cord dorsal horn neurons and midbrain dopaminergic neurons [73]. The *POU4F1* gene (also known *BRN3A*) regulates the morphology and functional diversity of retinal ganglion cells [74], while the *ZHX2* gene functions as a transcriptional repressor of the developing retina [75]. In addition, a small group of genes in the *VDAC3* network (*BACH1*, *E4F1*, *PLAGL1*) has been recognized to be part of the p53 pathway. In particular, *BACH1* is a transcriptional suppressor that heterodimerizes with small Maf proteins and binds to Maf recognition elements in the promoters of target genes related to the oxidative stress response [76]. *E4F1* is another transcriptional repressor, which can also function as a ubiquitin ligase mediating the ubiquitination of chromatin-associated p53 and directly controlling genes involved in mitochondrial functions and cell-cycle checkpoints [77], its deficiency results in oxidative stress-mediated cell death. The *PLAGL1* gene, also known as *Zac1*, seems to be implicated in neural plasticity mechanisms in the mouse model [78,79] as well as in the mitochondrial apoptosis, being a direct target of p53 [80]. In the healthy subGRN2 subnetwork, we found three genes (*ESR2*, *HES4*, *IRF5*) functionally related to neuroinflammatory mechanisms. *ESR2* gene encodes a member of the estrogen receptor family that influences mitochondrial membrane permeabilization in response to toxic events, such as ischemic brain injury and neurodegeneration [81]. Epigenetic alteration of the *HES4* gene promoter is associated with striatal degeneration in post-mortem brains of Huntington disease patients [82]. The *IRF5* gene is a known key mediator of innate and acquired immunity able to regulate numerous immune cells and pro-inflammatory cytokines [83]. Interestingly, its involvement in mediating neuronal injury was identified in a recent work, showing that knockdown of *IRF5* in NSC-34 cells inhibits apoptosis and mitochondrial redox homeostasis [84].

The subGRN2 also showed a group of genes associated with CNS development, such as *FOXP4*, implicated in retinogenesis and spinal neurogenesis [85]. In the same network the following genes have been also found: *HOXB3*, mainly investigated in PD mice models for its role in the development of dopaminergic neurons [86]; *POU3F3*, also named *Brain-1*, a well-known inhibitor of apoptosis and involved in neural development; it appears to play a central role in developmental delays and/or intellectual disability [87]. Moreover, the *ZNF197* gene regulates the DNA methylation pattern in the placenta, which is influenced by copper metabolism [88], and *ZNF420* (also named *Apak*) is a KRAB-type zinc-finger protein that negatively regulates p53-mediated apoptosis [89]. Also, the involvement of the *THRA* gene in human cortical development has recently been discovered, with neurocognitive and sensory-motor deficits in cases of its downregulation [90]. These results support the known role of mitochondrial ROS sensor attributed to *VDAC3* [36] together its involvement in the organelle functions. These *VDAC3* functions are well correlated with the oxidative stress linked to mitochondrial dysfunction and neurodegeneration [91]. Specifically, the presence of overoxidized cysteines in *VDAC3* from ALS samples is reported [37] while in physiological contexts the single cysteines of *VDAC3* are devoid of such modifications, thus able to form S-S bonds with protein interactors [36,37]. This would suggest that oxidative damage typical of ALS produces structural and functional defects at *VDAC3* capable of both modifying the pool of its interactors and limiting its role as a sensor of mitochondrial oxidative state.

Overall, the analysis of the *VDAC3* subnetworks, pathological and healthy, suggests that *VDAC3* may have a role in neuro-inflammatory mechanisms but also in the development of spinal motor neurons and the oculomotor system. These findings complement the expression data from the Expression Atlas repository ([www.ebi.ac.uk/gxa/home](http://www.ebi.ac.uk/gxa/home)) which, for humans, report higher levels of *VDAC3* gene in the cerebellum, diencephalon and spinal cord [79].

## 3. Discussion

Over the past 15–20 years, different computational methods have been developed to infer gene regulatory networks for the analysis

of transcriptomics data from different sources [92–94]. Of particular interest is the ability of differential co-expression analysis methods to identify regulatory genes underlying various phenotypes. These methods are mostly useful for identifying disease regulatory genes in pathological phenotypes [95], but also offer interesting insights for developing effective therapeutic approaches. Unfortunately, to date, studies aimed at inferring gene regulatory networks in ALS are lacking and no synthetic biology approach has been used to address this topic.

In this study, adopting a reverse engineering strategy based on the use of the ARACNe-AP algorithm, we analysed transcriptomic data from the blood of a large sample of ALS patients and their controls. By this approach we produced novel data on ALS and already known information that strengthen our study. Exactly, we inferred 7 different ALS regulatory genes networks (GRNs) for many phenotypic conditions of the disease. Specifically, GRN-1 and GRN3-7 were relative to survival time, site of onset (spinal or bulbar), and probability of developing the disease according to the sex. In each of these transcriptional regulatory networks, we detected the most relevant ALS biological pathways and potential ALS-promoting mechanisms, some of which completely novel. Among them, we identified signs of increased vulnerability in male patients and the activation of the radiation response-related network as a disease-specific signature for female BALS patients.

In this regard, one study demonstrated reduced metabolism in BALS patients [96], while another showed in SALS upregulation of the miR-23a-3p, a microRNA associated with motor neurons loss in BALS [97]. Taken together, these findings highlight the relevance of the distinction between SALS and BALS for predicting disease outcome and assessing treatment.

Furthermore, we divided the large-scale data into smaller and more informative network segments, obtained thus subnetworks of some hub genes (*TP53*, *SOD1*, *ALS2*, *VDAC3*) that could function as potential pathological drivers of ALS. By investigating the pathological and healthy subnetworks of the selected hub genes, we discovered different disease-related pathways and, interestingly, their interconnection with p53-mediated pathways. This result suggests a potential neurovascular response to ALS neuroinflammation.

In addition, data from ALS blood [98,99] confirm that patient blood cells respond to systemic and local signals with changes in gene expression similar to those found in the spinal cord or brain of ALS patients. This underlines the value of blood for diagnostic purposes, to assess the onset and progression of ALS. Moreover, the research strategy used in this study responds well to the growing need to study the disease with alternative approaches, also aimed at using investigation methods that do not involve animal models. In fact, the limitations of preclinical mouse models in recapitulating the complexity of the human disease are well known.

Overall, the synthetic biology approach applied in this study has allowed to successfully address some of the new challenges of ALS research: the analysis of big data released in patient databases; developing a broad and detailed overview of the molecular mechanisms of the disease; reducing animal studies; paving the way for the search in the blood of ALS “gold standard” biomarkers; laying the foundations for the design of new potential therapies for ALS.

### 3.1. Identification of biological pathways and topological key motifs in ALS GRNs

The prevalence of certain biological processes in a network over others may denote the activation of a regulatory axis with a more or less significant impact on the phenotype. Our analysis revealed few some differences and altered functions in pathway enrichment, with some punctual changes resulting in a noticeable regulatory shift of specific groups of hub genes. In particular, pairwise comparison of pathological (GRN1), healthy (GRN2) and disease control (GRN3) networks revealed clear differences in the pathway representation of affected and healthy individuals (Table 1a). Indeed, in GRN2 we mainly observed an enrichment related to common circulatory pathways (e.g. immune response and platelet aggregation), confirming already known data [28]. Interestingly, in GRN1 and GRN3 the top pathway found was ‘radiation response’ (GO:0009314). Under this macro-category are the response pathways to light stimuli and ionizing radiation, comprising almost 60 genes encoding for the antioxidant, anti-inflammatory and DNA repair machinery. This finding underlines the role of radiation-induced oxidative stress that correlates well with certain inflammatory mechanisms in ALS, such as altered lipid and protein metabolism and energy imbalance [100]. The identification of this peculiar oxidative stress condition in blood samples from ALS patients supports the so-called “neurovascular scenario”, according to which alterations in gene expression may be associated with some altered ALS signalling pathways due to blood-brain barrier damage [101]. This suggests that the adoption of diagnostically valid and non-invasive blood biomarkers related to genes of the ALS regulatory network, could be advantageous not only to initiate preventive actions, but also as an aid in monitoring the course [75]. Furthermore, it is known that low-dose radiation ~~can~~ induces increased oxidative metabolism and redox imbalance of the central nervous system (CNS) microenvironment, mitochondrial dysfunction and protein degradation, leading to neural apoptosis in many neurodegenerative and inflammatory diseases [102]. Among the 60 genes we identified by GO analysis (Table 1; Table S2) there are 3 (*CCND1*, *MMP1* and *MMP3*) suspected to play a role in ALS and related to radiation-induced oxidative stress. In particular, the cyclin D1 (*CCND1*), not only is a radiation-responsive gene but also interacts with the *FUS* gene by blocking its translation. This action is known to be toxic for motor neurons [103]. Regarding the matrix metalloproteinase genes, *MMP1* and *MMP3*, they represent potential biomarkers of motor neuron degeneration since their deregulation is associated with ALS [78]. The identification of reliable biomarkers is one of the main goals of ALS research [25]. ALS is still an incurable disease, with few treatment options useful only to slightly slow down its course. Having ‘gold standard’ biomarkers would certainly help the development of new drugs for this disease.

In addition, we performed a topological analysis of the inferred GRNs, since it is known that the structure of a network can be as informative as its complexity. Of particular interest is the analysis of disease-associate networks because pathogenesis usually induces topological variations useful to distinguish the pathology [104]. Our results shows that the genomic data from GRN1-GRN3 reflect well the ‘topological’ characteristics of their networks (Fig. S2). In particular, a common L-shaped spatial conformation of gene interactions in all three networks (GRN1-GRN3) is easily recognizable. This could denote a hierarchical pattern of molecular signals expressed in blood as a gene signature of the target tissue. Moreover, the density and orientation of genes in GRN1 overlap with those in GRN3,

sharing a similar arrangement of the central body and side branches (Fig. S2). These data reveal thus features in common between the profiles of ALS patients (i.e. the GO:0009314 “radiation response” pathway) and those of controls. Also, some transcripts of the healthy regulatory profiles in GRN2 are missing or severely impaired in the pathological GRN1 network and the GRN3 fusion network (Fig. S2). These results are in line with what is known, in general, about the topology of networks. Namely, that the interacting components of a network with a constant topology have evolved to perform specific cellular functions [105]. Although parameters that do not change the topology of a network may equally influence its response dynamics [106].

Concerning the GO signal (GO:0009314) of “radiation response” was not found in SALS networks (GRN4) or “male ALS patients” (GRN7), enriched instead in pathways related to protein metabolism. The signal “response to radiation” was instead found in both BALS - patients with reduced survival (GRN5) (Table 1b) and “female ALS patients” (GRN6) networks (Table 1c) normally less affected than men by the disease and more prone to spinal onset (Figs. S1a–b). This leads to the hypothesis that the gene enrichment characteristic of BALS women may derive from the impairment of the protective effects of antioxidant genes, essential to attenuate/counteract the deleterious effects of the disease [79]. Therefore, radiation-induced oxidative damage would seem to have an important role in the pathogenesis of ALS. There are several reports that associate ALS with defective DNA repair. For example, FUS pathology has been shown to disrupt DNA damage response signalling (DDR) [107,108]. While an important role in the NHEJ mechanism of DNA double-strand break repair and mismatch repair was discovered for *TDP43* [109,110]. On the other hand, loss of key DDR components leads not only to accumulation of DNA damage but also to marked proteotoxic stress and genome instability [111]. The latter condition can be induced by radiation exposure, and interestingly, it has been reported that low-dose radiation can increase the risk of neurodegeneration [112]. Regarding the radiation-induced oxidative damage, is it not surprising that edaravone, prescribed to ALS patients for its anti-oxidant and anti-inflammatory effects, was studied to assess its ability to protect cells from the oxidising action of radiation [113].

As for GRNs 1–3, also for GRNs 4–7, the ontological correspondence coordinates well with the physical model of the networks, a: GRN4 resembles GRN6 and GRN5 resembles GRN7 (Fig. S2). This shows that the networks reconstructed by ARACNe are plausibly matched and are able to predict two models of the same disease but distinct for phenotypic profile, a less resilient one composed of male BALS patients and a more resilient one composed of female SALS patients. These data correlate well with findings from other studies focusing on the influence of sex and site of onset in ALS progression. Moreover, they hint at the importance of sex-specific factors in understanding the pathophysiology of ALS, as also suggested by others [114]. Interestingly, a recent study on iPSC motor neurons showed that few genes are differentially expressed between ALS patients and healthy controls, although the number of differentially expressed genes in ALS is much higher in males than in females. In particular, in ALS males, an enrichment of stress response pathways related to TNF, NF- $\kappa$ B, and radiation exposure was found. Instead, in ALS females, pathways related to ribosomal function, oxidative phosphorylation, as well as mitochondria were more involved [115]. Also, it has emerged the idea that sexual dimorphism in ALS results from interactions between the central nervous system and peripheral organs, implicating the vascular, endocrine, metabolic, musculoskeletal, and immune systems, which are physiologically different between males and females [116].

In order to identify putative prognostic markers of ALS specifically associated with patients’ survival time, we implemented disease stratification by enhancing the pathway-level analysis of gender responses to the type of ALS site of onset. The obtained results suggest the existence of a lowest common denominator among the different phenotypes associated with ALS patients [118]. By analyzing the 61 gene signatures for predicting survival of ALS patients, we showed that the same two markers of poor prognosis (*SLC40A1*, *CLEC7A*) are found in the spinal, bulbar and female networks. This confirms the accurate inference of the algorithm in predicting significant regulators of gene perturbations. Contextually, we also questioned the significance of the absence of the *CLEC7A* marker in the male network. This gene is a member of the superfamily of proteins containing the C-type lectin/C-type lectin-like domain (CTL/CTLD), typically expressed by cells of the innate immune response. Recent evidence in ALS patients [119] and in mouse models of ALS [120] evaluated it as a key gene of microglial activity, associated with short survival and severe condition in a male-dependent manner. The results obtained surprised us because, considering the ALS case history, we expected more worse prognosis targets in the male network than in the female network. The Cox proportional-hazards model used by Swindell et al. [25] identifies gene predictors of ALS survival with a different approach from ours. In fact, ARACNe-AP analysis can obtain information on the transcriptional relationships of hub genes, not on changes in gene expression levels [24]. In particular, the construction of the matrix of gender-distinct disease profiles allowed us to identify some subtle regulatory differences more stringently. In this regard, loss of *CLEC7A* could be an aggravating factor that exacerbates the vulnerability of male patients. However, this is a hypothesis that requires further study to be confirmed. In any case, our results are in line with other evidence on the role of this gene and lead to narrow the assessment of a differential neuro-inflammatory response in male and female due to the absence of a neuroprotective mechanism caused by *CLEC7A*. The inferred GRN thus reveals some hidden pathways of initiation, suggesting that attention should be turned to as yet under-investigated aspects [5,69].

### 3.2. ALS transcriptional networks of *TP53*, *SOD1*, *ALS2* and *VDAC3* genes

*In silico* analysis of the subGRNs sub-networks of the four selected hub genes produced interesting and useful data to better understand ALS. First, within the individual subnetworks deduced for each hub gene, we found several novel target genes, not yet associated with ALS and putatively useful as disease biomarkers. Also, for each analysed hub gene, the relative involvement in specific pathways was highlighted, many of which related to mechanisms of neurodegeneration, some directly related to ALS, others more generic but no less interesting. For example, the neuronal development or spinal motor neuron pathways found associated, respectively, with the pathological subnetworks of *SOD1* and *VDAC3*. Interestingly, *VDAC3* also seems to be involved in the development of the oculomotor system, the only apparatus whose muscles are not affected by ALS. *TP53* appears to promote some preferential pathways involving autophagy and neuronal apoptosis in the ALS contest, which are also correlated to endolysosomal and

mitochondrial metabolism genes [30]. Another interesting result emerged from the elaboration of the *ALS2* gene subnetwork. Indeed, its transcriptional map was the one with the highest number of ALS-related hits. A data of great interest for drug design on new target molecules.

Certainly, a strong finding from this analysis is that we have revealed the interconnection of the four selected hub genes (*TP53*, *SOD1*, *ALS2*, *VDAC3*) with p53-mediated pathways, which suggests a potential neurovascular response to ALS neuroinflammation. This is confirmed by the results of other studies that identify p53 as a key regulator of neurodegeneration caused by ALS-associated genes [33,34]. Our data therefore reinforce the previous results and, in addition, provide evidence that the aforementioned mechanisms may have a primary role in the onset of the ALS, both sporadic and genetic. Furthermore, neuroinflammation also correlates well with other pathways associated with the 4 hub genes identified by us, namely the response to radiation damage or to the oxidative stress associated with it. In this regard, *VDAC3*, functioning as an oxidative stress sensor, could well guide this response by controlling various pathways involved, including mitochondrial metabolism itself [91]. It is therefore possible that DNA damage, as well as the related genomic instability, may have a significant role in the onset of the disease.

Overall, the information on the molecular mechanisms associated with ALS resulting from this analysis, in addition to being overall consistent with the data reported in the literature, have the value of providing a new integrated vision of the transcriptional regulators of ALS. A vision that not only helps to broaden knowledge about the disease but is also useful to identify certain molecular features on which to focus future studies also with the intent of discovering effective ALS biomarkers.

### 3.3. Study limitations and strengths

We are aware that our study is burdened by some weaknesses. Probably the main ones are related to the quality and quantity of input data (e.g. GEP) used to generate the GRNs. In this regard, the limited number and accessibility of repository datasets [121] made it difficult to collect a sufficiently large sample to prepare input matrices for ARACNe-AP with case and control GEPs from different tissue samples. In particular, the limited availability of GEPs based on human/murine spinal cord and neurons prevented the expected cross-validation between tissues and species, forcing us to scale down the initial research design by looking for information from microarray data from blood of ALS patients. However, there are several research lines [122–127] that demonstrate that blood is a promising alternative to the use of other biological samples. Also, some studies have evaluated the link between pathological processes in the brain and transcriptional levels in blood cells of ALS subjects [99,127–129].

In fact, although not directly affected by the disease, blood is still as “informative” as other tissues considered more suitable (e.g. cerebrospinal fluid, CSF) thanks to the continuous exchanges at the blood-CSF barrier. In addition, blood is easily available and sampled by patients and therefore highly suitable for meta-analysis studies with statistically valid results, unthinkable to obtain from CSF or spinal cord. Therefore, the transcriptional signature in peripheral blood cells is a useful surrogate for evaluating CNS gene expression. This makes blood an excellent biological sample in which to search for ALS biomarkers also useful to speed up the diagnosis, especially in the early stages of ALS, and to follow patients in the follow-up of treatment.

Another gap of this study stems from the limited information retrieved on the exact number of sporadic and familial patients, genetic variations, sample ID associated with sporadic and familial profiles of the GSE112681 dataset we used. Probably, this constraint hindered the opportunity to perform more accurate inferences and interpretations on the stratification of the ALS population and thus elaborate more deeply the regulatory mechanisms of the genotype-phenotype correlation. However, we believe that most of weaknesses may be compensated by the size of the final sample analysed and the potential of the algorithm used.

Among the strength points of our GRN analysis is the ability of ARACNe-AP to model the complex nonlinear relationships between genes. Indeed, by capturing the interactions between multiple TFs and their target genes, GRN analysis can reveal how perturbations in one part of the network propagate through the overall system, leading to disease. This is particularly important in ALS for two reasons. First, because the effects of genetic mutations may not be immediately evident at the level of individual genes, becoming more evident when considering the broader regulatory context. Second, it is particularly useful if the input data, as in the present study, lack some anamnestic details without affecting the robustness of the reverse engineered models.

## 4. Conclusions

In this work, blood networks of ALS patients and control subjects were reverse engineered by ARACNe-AP algorithm to determine the specific gene interactions of each network, as well as the pathways linking the normal state to the ALS condition. The potential of the designed reverse engineering framework, based on the processing of transcriptome data useful for predicting phenotypic differences between pathological and healthy subjects, allowed us to identify a considerable number of potential ALS-promoting mechanisms and putative transcriptional biomarkers previously unknown. Noteworthy is the presence of *SOD1* among the hub genes in the blood of non-*SOD1* ALS patients. Among the most interesting results emerging from our predictive study is the characterization of the radiation-induced oxidative pathway (GO:0009314), which associated the “female gender with ALS” group (less vulnerable to the disease than the “male gender” group) with bulbar onset (ALS patients with worse prognosis) with the weakest pathway signal among the most significant pathological mechanisms usually investigated [117]. Another significant finding is the presumed transcriptional cooperation of hub genes playing a role in inter-organelle communication. In particular, communication between the ER, mitochondria, lysosomes, and Golgi apparatus correlates well with the known activity of p53 in counteracting neuroinflammatory response [25,130].

Overall, the originality and importance of the study consist in having predicted/identified, for the first time:

1. Connections between single molecular mechanisms of ALS related to mutations on specific genes;
2. Molecular pathways associated with specific ALS phenotypes, related to the site of onset and/or to the sex of the individual;
3. *VDAC3* as a hub gene of GRNs in ALS.
4. Putative pathological targets of *TP53* in ALS;
5. Possible disease biomarkers in the blood of patients, from which to design putative therapies against ALS.

Taken together, our results demonstrate the utility of an integrated systems biology approach to obtain a detailed view of the gene interaction networks putatively involved in the regulation of the molecular mechanisms related to ALS. The proposed approach is not limited to this biological context but could also find application in other disease contexts.

## 5. Materials and methods

### 5.1. The algorithm for the reconstruction of accurate cellular networks - adaptive partitioning version (ARACNe-AP) pipeline

In order to map the transcriptional interplay that may contribute to ALS pathogenesis, we used the reverse engineering algorithm for the reconstruction of accurate cellular networks - adaptive partitioning version (ARACNe-AP) [24] and GEPs of the GSE112681 dataset [25,26]. These datasets were generated in the Netherlands from peripheral blood samples taken from a panel of ALS patients and from a control group of neurologically healthy individuals. The ARACNe-AP algorithm, developed by Andrea Califano [8,23,24] is an advanced computational methodology in the field of computational systems biology. ARACNe-AP is designed to infer regulatory relationships within biological systems by statistically analyzing gene expression data. It utilizes MI and statistical significance tests to deduce direct interactions between genes, thereby revealing potential gene regulatory networks. This algorithm aims to provide a rigorous and data-driven approach for the reconstruction of accurate cellular networks, contributing to a deeper understanding of biological processes at the genomic level.

The inference of gene regulatory networks using the ARACNe-AP algorithm can be described in a series of 9 steps:

1. *Data Acquisition and Pre-processing*:
  - Begin with gene expression data, typically in the form of microarray or RNA-Seq datasets [131].
  - Pre-process the data by normalizing and filtering to remove noise and systematic biases.
2. *MI Calculation*:
  - Calculate the MI between pairs of genes. MI measures the statistical dependence between two variables - in this case, the expression levels of genes.
3. *Statistical Significance Assessment*:
  - Estimate the statistical significance of the MI scores to identify meaningful interactions while controlling for false positives.
  - Typically, permutation tests or other statistical methods are used to determine significance thresholds.
4. *Network Construction*:
  - Create an initial network where nodes represent genes, and edges represent MI scores between gene pairs.

These edges could transduce a signal of:

**Activation:** A directed arc from a gene (node) A to a gene (node) B indicates an activation effect. In other words, gene A promotes or stimulates the expression of gene B. This can occur when gene A produces a protein or TF that acts as an activator for gene B.

**Inhibition:** A directed arc from gene A to gene B can also represent an inhibition effect. This means that gene A suppresses or reduces the expression of gene B. This can occur when gene A encodes for a protein or TF that acts as a repressor for gene B.

- Apply a significance threshold to retain only significant interactions, thereby reducing the network to a subset of high-confidence edges.
5. *Data Processing Inequality (DPI) Pruning*:
    - Apply the DPI principle to remove indirect interactions and false positives from the network.
    - DPI pruning ensures that only direct interactions are retained in the network.
  6. *Network Refinement*:
    - Further refine the network by removing any remaining redundant or spurious interactions, using techniques such as the Minimum Spanning Tree (MST) algorithm.
    - This step helps to simplify the network structure.
  7. *Network Visualization and Analysis*:
    - Visualize the resulting gene regulatory network to gain insights into the relationships between genes.
    - Analyse network properties, such as hubs, clusters, and central nodes, to identify key regulators and functional modules.
  8. *Biological Interpretation*:
    - Functionally annotate the genes in the network to understand their roles in biological processes.
    - Investigate the potential regulatory pathways and feedback loops within the network.
  9. *Iterative Process*:
    - Gene regulatory network reconstruction is often an iterative process, with refinement based on additional data and experimental findings.

## 5.2. Gene expression dataset analysis

The transcriptomic data that we used to infer the GRNs derive from 1117 whole blood RNA samples included in the SuperSeries GSE112681 (composed by following SubSeries GSE112676 and GSE112680), available from the Gene Expression Omnibus (GEO) database ([www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo)); (see Refs. [25,26]). In details, among the 1117 microarray expression profiles reported in GSE112681, 397 are patients with ALS, 75 are patients defined as ‘ALS-mimics’ who affected by diverse ALS-like conditions (the most common diagnoses were benign fasciculations, spinal muscular atrophy, and myelopathy), and 645 are neurological healthy volunteers matched for gender and age. Samples were hybridized to two different platforms: Illumina’s HumanHT-12 version 3 and version 4 BeadChips according to manufacturer’s protocol (Illumina, Inc., San Diego, CA, U.S.A.). Some phenotypic data of GSE112681 are reported in Refs. [25,26], and these allowed us to prepare different networks based on gender (239M/158F), survival status (from one year up to 10 years); and site of onset (251 spinal-onset vs. 146 bulbar-onset samples) (Figs. S1a–b). The gender composition of the original cohort study includes 60.2 % males (M) and 39.8 % females (F) of ALS patients; 55.3 % M and 44.7 % F in the control group. Regarding the genetic mutation screening of familial and sporadic ALS, the C9orf72 expansion is found in 5.2 % of GSE112676 and 12.8 % of GSE112680; however, this feature has not been annotated in the corresponding GEO file of the patient. A very particular feature reported in both datasets is the diagnostic form of ALS patients, the spinal-onset (SALS) and bulbar-onset ALS (BALS). In the GSE112676 were diagnosed with the 61.4 % of SALS patients and the 38.6 % of BALS patients, while the 66.9 % of SALS and the 33.1 % BALS in the GSE112680. Moreover, the average age of ALS patients is 63.6 (55–72 years), and 62.3 for controls (55–69 years). The survival time of ALS patients ranges from less than one year to more than 10 years (in just a few cases).

According to the properties of the study sample reported in GSE112681, inclusion and exclusion criteria were defined to answer to our research question. Hence, the large-screening dataset (GSE112681) is eligible (*inclusion criteria*) as ARACNe-AP algorithm requires  $\geq 100$  GEPs to build GRNs [24]. With the aim to ensure the representativeness of ALS population and the statistical power of the sample-specific networks analysis [132–134], we assembled ARACNe-AP matrices to generate GRNs using a reduced set from the original entire sample of 1117 profiles, selecting 794 expression profiles exclusively associated to patients with confirmed ALS diagnosis (*inclusion criteria*). Therefore, the previously mentioned 75 ALS-mimics (*exclusion criteria*) were excluded from the current study sample.

## 5.3. Gene network inference

For optimal analysis, ARACNe-AP meets two certain requirements, large dataset of GEPs ( $\geq 100$ ), as first input, and the list of gene regulators (e.g. TFs) of the study organism [24], as second input.

The consensus network inferred with ARACNe-AP was built using a subset of 794/1117 GEPs (GSE112681) [25,26], and a list of 1635 human TFs available from database of Terrence Donnelly Centre for Cellular and Biomolecular Research (CCBR), University of Toronto (1.01 version) ([humantf.ccb.utoronto.ca/download.php](http://humantf.ccb.utoronto.ca/download.php)) (Table S1) [27] to determine whether the inferred activities of each regulator were significantly correlated using a conservative MI threshold ( $p \leq 1.0 \times 10^{-8}$ , i.e.,  $p \leq 0.05$  Bonferroni corrected for all candidate interactions).

With regards of data normalization, each gene matrix prepared for ARACNe-AP contains as input data quantile normalized values of GEPs processed as reported in GSE112681 [25,26]. In details, the used raw microarray data files contain the raw intensity measurements (raw counts) directly from the microarray scanner where the proper gene annotations are available to map probe sets to gene IDs. Background correction methods are used to correct for this noise. Microarray data have been normalized to make the distributions of probe intensities similar across all arrays. It has been used the standard procedure of the quantile normalization. Microarray expression have been log<sub>2</sub>-transformed to stabilize the variance and make the data more normally distributed. This step is very important because ARACNe-AP uses mutual information, which assumes that the data are continuous and appropriately scaled. In practice the data were normalized using quantile normalization with IlluminaGUI in R. After quantile normalization, gender was checked by expression of gender-specific Y chromosomal probes (JARID1D and RPS4Y1). Samples with gender mismatches were excluded. Subsequently expression values were log<sub>2</sub> transformed and quantile normalized. Arrays were projected along the first and second principal component and outliers were dismissed. Moreover, the genes with very low variance across samples have been filtered out since they are unlikely to contribute meaningful information to network inference. Finally, using k-nearest neighbors, missing values were identified in the data set, samples or genes with missing values were removed. It is crucial to ensure that the final input file has no missing values, as ARACNe-AP requires complete data for accurate MI calculation.

As previously specified, the clinical group of ALS-mimics of the original dataset was not included to avoid potential confounding variables in the regulatory analysis. The network inferred with ARACNe-AP is a directed graph that consists of a series of arches representing the connections between regulators (predefined list of input genes, e.g. TFs) and their targets. Each connection is computed by the correlation measure of MI that indicates the pairwise connection strength. The huge complexity of the network output was filtered using the statistical parameter of 95th perc of MI to select gene connections with higher-degree MI corresponding with the strongest gene regulators [135] (Table S2). To identify hub genes and their related ALS and healthy control subnetworks, subGRN1 and subGRN2 respectively, we used the top 5 % and 30 % ranked arcs of the complete network between the regulators and their targets [24].

We prepared different matrices with 794 GEPs from GSE112681 to build the following seven GRNs: ALS patients (both genders, SALS and BALS condition;  $n = 397$ ) named “pathological network” (GRN1); only controls ( $n = 397$ ), defined as “healthy network” (GRN2); all samples of GRN1 and GRN2 ( $n = 694$ ), defined as “disease-control network” (GRN3); only SALS patients ( $n = 251$ ), called “SALS network” (GRN4); only BALS patients ( $n = 146$ ), “BALS network” (GRN5); only female patients (SALS and BALS including;  $n =$



158), “female ALS network” (GRN6); and, only male patients (SALS and BALS including;  $n = 239$ ), “male ALS network” (GRN7).

Before running all GRNs simulations, the hub gene list (Table S2) was pruned of probes associated with uncharacterized or unknown functions (e.g. LOC, KIAA, FAM, FLJ, DKFZp, C#orf) to reduce random noise [136]. In addition, before generating the two gender-specific GRNs (GRN6-7), the hub gene lists of GRN6-7 (Table S2) were cleared of genes implicated in gonadal differentiation [29]. To simulate a temporal signature of disease state in relation to the different ALS onset site and gender, all GEPs were preprocessed stratifying patients based on the survival duration, as illustrated in Figs. S1a–b. The clustering of survival data was accomplished to compute GRN4-5 and GRN6-7, in order to analyse differences in the gene regulatory pattern among these GRNs and uncover possible critical mediators or interdependencies of genes/biological pathways involved in the ALS progression mechanisms and prognostic outcomes.

#### 5.4. Pathway analysis

We analysed the list of ARACNe-AP inferred genes at 95th perc of MI (Table S2) by using Gene Ontology (GO) (geneontology.org/). We also used the hierarchically organized underlying GO-terms calculated by Fisher’s exact test and using the Bonferroni correction for  $P < 0.05$ , and the Panther Overrepresentation test (release 20210224; Panther v.16.0 released 2020-12-01) in order to find the biological pathways of each reconstructed network [137].

#### CRedit authorship contribution statement

**Xena G. Pappalardo:** Writing – review & editing, Writing – original draft, Formal analysis. **Giorgio Jansen:** Writing – original draft, Visualization, Software, Formal analysis, Data curation. **Matteo Amaradio:** Visualization, Software. **Jole Costanza:** Software, Data curation. **Renato Umeton:** Software, Data curation. **Francesca Guarino:** Writing – review & editing, Supervision. **Vito De Pinto:** Writing – review & editing, Supervision, Funding acquisition. **Stephen G. Oliver:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Angela Messina:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Conceptualization. **Giuseppe Nicosia:** Supervision, Methodology, Conceptualization.

#### Data availability statement

All data generated or analysed during this study are included in this published article (and its Supplementary Information files).

#### Ethical approval

Not applicable.

#### Funding

This research was funded by the European Union-Next Generation EU - PRIN 2022 PNRR call, Project No. P20225J5NB, to Angela Messina.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

G.N. Author thanks Valerie Wood (Cambridge) for helpful discussions. Work in the Cambridge laboratory was supported by a Wellcome Trust - MRC UK award to SGO (grant code: 089703/Z/09/Z). GN thanks the University of Cambridge for the extended research periods and the opportunity to have been Visiting Professor during the Sabbatical year. Moreover, GN thanks Wolfson College, the Department of Biochemistry and the Cambridge Systems Biology Centre for the hospitality and warm welcome that allowed for peaceful studies, novel interesting collaborations, and fruitful research activities. SGO would like to dedicate this paper to the memory of Justin Yerbury, who was not only a superb scientist but whose courage and dedicatio were inspiration.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e40696>.

## Abbreviations

ALS	Amyotrophic Lateral Sclerosis
AD	Alzheimer's disease
ADNP	Activity-dependent neuroprotective protein
ALS/FTD	ALS/fronto-temporal dementia
ALS2	Alsin Rho guanine nucleotide exchange factor
ALX4	ALX Homeobox 4
ARACNe-AP	Algorithm for the Reconstruction of Accurate Cellular Networks - adaptive partitioning version
BACH1	BTB domain and cnc homolog 1
BCL6	BCL6 Transcription Repressor
COMMD1	Copper metabolism MURR1 domain-containing 1
CREB3	CAMP Responsive Element Binding Protein 3
CSTF1	Cleavage stimulation factor subunit 1
E4F1	E4F Transcription Factor 1
ER	Endoplasmatic reticulum
ESR2	Estrogen receptor (ER)- $\beta$
ETV4	ETS Variant Transcription Factor 4
FOXA1	Forkhead Box A1
FOXN4	Forkhead Box N4
FOXP3	Forkhead box P3
FUS	FUS RNA Binding Protein
GEP	Gene expression profile
GO	Gene Ontology
GRN	Gene regulatory network
HES4	Hairy and enhancer of split 4
HOXB3	Homeobox B3
HOXD1	Homeobox D1 protein
HOXD10	Homeobox D10 protein
IRF5	Interferon regulatory factor 5
LCAP/LICAM-AS1	Lung carcinoma-associated 10
LMX1A	LIM Homeobox Transcription Factor 1 Alpha
LMX1B	LIM Homeobox Transcription Factor 1 Beta
lncRNA	Long intergenic non-coding RNA
NEUROG3	Neurogenin 3
PACAP	Pituitary adenylate cyclase-activating peptide
PAX5	Paired Box 5
PD	Parkinson's disease
PLAG1	Pleiomorphic adenoma gene like 1
POU3F3	POU Class 3 Homeobox 3
POU4F1	POU Class 4 Homeobox 1
SEMA3E	Semaphorin 3E
SOD1	Superoxide dismutase 1
SOX13	SRY-Box Transcription Factor 13
SubGRN	Subnetwork (sub gene regulatory network)
TBX6	T-Box Transcription Factor 6
TCF7	Transcription Factor 7
TF	Transcription Factor
THRA	Thyroid hormone receptor alpha
TP53	Tumour protein 53
TSHZ1	Teashirt Zinc Finger Homeobox 1
VDAC3	Voltage-dependent anion-selective channel protein 3
VIP	Vasoactive intestinal peptide
ZHX2	Zinc Fingers and Homeoboxes 2

## References

- [1] B. He, K. Tan, Understanding transcriptional regulatory networks using computational models, *Curr. Opin. Genet. Dev.* 37 (2016) 101–108, <https://doi.org/10.1016/j.gde.2016.02.002>.
- [2] P. Unger Avila, T. Padvitski, A.C. Leote, H. Chen, J. Saez-Rodriguez, M. Kann, et al., Gene regulatory networks in disease and ageing, *Nat. Rev. Nephrol.* 20 (2024) 616–633, <https://doi.org/10.1038/s41581-024-00849-7>.
- [3] P. Badia-I-Mompel, L. Wessels, S. Müller-Dott, R. Trimbou, R.O. Ramirez Flores, R. Argelaguet, et al., Gene regulatory network inference in the era of single-cell multi-omics, *Nat. Rev. Genet.* 24 (2023) 739–754, <https://doi.org/10.1038/s41576-023-00618-5>.
- [4] C. Carré, A. Mas, G. Krouk, Reverse engineering highlights potential principles of large gene regulatory network design and learning, *NPJ Syst Biol Appl* 3 (2017) 17, <https://doi.org/10.1038/s41540-017-0019-y>.
- [5] A. Vashishat, P. Patel, G. Das Gupta, B. Das Kurmi, Alternatives of animal models for biomedical research: a comprehensive review of modern approaches, *Stem Cell Rev Rep* 20 (2024) 881–899, <https://doi.org/10.1007/s12015-024-10701-x>.
- [6] B. Mishra, N. Kumar, M.S. Mukhtar, Systems biology and machine learning in plant-pathogen interactions, *Mol. Plant Microbe Interact.* 32 (2019) 45–55, <https://doi.org/10.1094/MPMI-08-18-0221-FI>.

- [7] C. Manzoni, P.A. Lewis, R. Ferrari, Network analysis for complex neurodegenerative diseases, *Current Genetic Medicine Reports* 8 (2020) 17–25, <https://doi.org/10.1007/s40142-020-00181-z>.
- [8] K. Basso, A.A. Margolin, G. Stolovitzky, U. Klein, R. Dalla-Favera, A. Califano, Reverse engineering of regulatory networks in human B cells, *Nat. Genet.* 37 (2005) 382–390, <https://doi.org/10.1038/ng1532>.
- [9] A. Bakulin, N.B. Teyssier, M. Kampmann, M. Khoroshkin, H. Goodarzi, pyPAGE: a framework for Addressing biases in gene-set enrichment analysis-A case study on Alzheimer's disease, *PLoS Comput. Biol.* 20 (2024) e112346, <https://doi.org/10.1371/journal.pcbi.1012346>.
- [10] N. Khayer, S.-A. Marashi, M. Mirzaie, F. Goshadrou, Three-way interaction model to trace the mechanisms involved in Alzheimer's disease transgenic mice, *PLoS One* 12 (2017) e0184697, <https://doi.org/10.1371/journal.pone.0184697>.
- [11] J.U. Loers, V. Vermeirssen, A single-cell multimodal view on gene regulatory network inference from transcriptomics and chromatin accessibility data, *Briefings Bioinform.* 25 (2024) bbae382, <https://doi.org/10.1093/bib/bbae382>.
- [12] F. Tomasi, M. Pozzi, M. Lauria, Investigating the mechanisms underlying resistance to chemotherapy and to CRISPR-Cas9 in cancer cell lines, *Sci. Rep.* 14 (2024) 5402, <https://doi.org/10.1038/s41598-024-55138-x>.
- [13] B. Ikiz, M.J. Alvarez, D.B. Ré, V. Le Verche, K. Politi, F. Lotti, et al., The regulatory machinery of neurodegeneration in in vitro models of amyotrophic lateral sclerosis, *Cell Rep.* 12 (2015) 335–345, <https://doi.org/10.1016/j.celrep.2015.06.019>.
- [14] P. Masrori, P. Van Damme, Amyotrophic lateral sclerosis: a clinical review, *Eur. J. Neurol.* 27 (2020) 1918–1929, <https://doi.org/10.1111/ene.14393>.
- [15] P. Ciryam, I.A. Lambert-Smith, D.M. Bean, R. Freer, F. Cid, G.G. Tartaglia, et al., Spinal motor neuron protein supersaturation patterns are associated with inclusion body formation in ALS, *Proc. Natl. Acad. Sci. U. S. A.* 114 (2017) E3935–E3943, <https://doi.org/10.1073/pnas.1613854114>.
- [16] C.A. Brown, C. Lally, V. Kupelian, W.D. Flanders, Estimated prevalence and incidence of amyotrophic lateral sclerosis and SOD1 and C9orf72 genetic variants, *Neuroepidemiology* 55 (2021) 342–353, <https://doi.org/10.1159/000516752>.
- [17] K. Boylan, Familial amyotrophic lateral sclerosis, *Neurol. Clin.* 33 (2015) 807–830, <https://doi.org/10.1016/j.ncl.2015.07.001>.
- [18] M. Ghasemi, R.H. Brown, Genetics of amyotrophic lateral sclerosis, *Cold Spring Harb Perspect Med* 8 (2018) a024125, <https://doi.org/10.1101/cshperspect.a024125>.
- [19] A. Al-Chalabi, J. Andrews, S. Farhan, Recent advances in the genetics of familial and sporadic ALS, *Int. Rev. Neurobiol.* 176 (2024) 49–74, <https://doi.org/10.1016/bs.irm.2024.04.007>.
- [20] E.L. Feldman, S.A. Goutman, S. Petri, L. Mazzini, M.G. Savelieff, P.J. Shaw, et al., Amyotrophic lateral sclerosis, *Lancet* 400 (2022) 1363–1380, [https://doi.org/10.1016/S0140-6736\(22\)01272-7](https://doi.org/10.1016/S0140-6736(22)01272-7).
- [21] S. Shellikeri, V. Karthikeyan, R. Martino, S.E. Black, L. Zinman, J. Keith, et al., The neuropathological signature of bulbar-onset ALS: a systematic review, *Neurosci. Biobehav. Rev.* 75 (2017) 378–392, <https://doi.org/10.1016/j.neubiorev.2017.01.045>.
- [22] A. Granzotto, B. Vissel, S.L. Sensi, Lost in translation: inconvenient truths on the utility of mouse models in Alzheimer's disease research, *Elife* 13 (2024) e90633, <https://doi.org/10.7554/eLife.90633>.
- [23] A. Lachmann, F.M. Giorgi, G. Lopez, A. Califano, ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information, *Bioinformatics* 32 (2016) 2233–2235, <https://doi.org/10.1093/bioinformatics/btw216>.
- [24] A.A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R.D. Favera, et al., ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context, *BMC Bioinform.* 7 (2006) S7, <https://doi.org/10.1186/1471-2105-7-S1-S7>.
- [25] W.R. Swindell, C.P.S. Kruse, E.O. List, D.E. Berryman, J.J. Kopchick, ALS blood expression profiling identifies new biomarkers, patient subgroups, and evidence for neutrophilia and hypoxia, *J. Transl. Med.* 17 (2019) 170, <https://doi.org/10.1186/s12967-019-1909-0>.
- [26] W. van Rheenen, F.P. Diekstra, O. Harschnitz, H.-J. Westeneng, K.R. van Eijk, C.G.J. Saris, et al., Whole blood transcriptome analysis in amyotrophic lateral sclerosis: a biomarker study, *PLoS One* 13 (2018) e0198874, <https://doi.org/10.1371/journal.pone.0198874>.
- [27] S.A. Lambert, A. Jolma, L.F. Campitelli, P.K. Das, Y. Yin, M. Albu, et al., The human transcription factors, *Cell* 172 (2018) 650–665, <https://doi.org/10.1016/j.cell.2018.01.029>.
- [28] K. Gruden, M. Hren, A. Herman, A. Blejec, T. Albrecht, J. Selbig, et al., A “crossomics” study analysing variability of different components in peripheral blood of healthy caucasoid individuals, *PLoS One* 7 (2012) e28761, <https://doi.org/10.1371/journal.pone.0028761>.
- [29] M.G. Butler, S.K. Rafi, A. McGuire, A.M. Manzardo, Currently recognized clinically relevant and known genes for human reproduction and related infertility with representation on high-resolution chromosome ideograms, *Gene* 575 (2016) 149–159, <https://doi.org/10.1016/j.gene.2015.08.057>.
- [30] A. Szybińska, W. Leśniak, P53 dysfunction in neurodegenerative diseases - the cause or effect of pathological changes? *Aging Dis* 8 (2017) 506–518, <https://doi.org/10.14336/AD.2016.1120>.
- [31] R. Kumar, S. Haider, Protein network analysis to prioritize key genes in amyotrophic lateral sclerosis, *IBRO Neurosci Rep* 12 (2022) 25–44, <https://doi.org/10.1016/j.ibneur.2021.12.002>.
- [32] M. Maor-Nof, Z. Shipony, R. Lopez-Gonzalez, L. Nakayama, Y.-J. Zhang, J. Couthouis, et al., p53 is a central regulator driving neurodegeneration caused by C9orf72 poly(PR), *Cell* 184 (2021) 689–708.e20, <https://doi.org/10.1016/j.cell.2020.12.025>.
- [33] S.K. Suthar, S.-Y. Lee, The role of superoxide dismutase 1 in amyotrophic lateral sclerosis: identification of signaling pathways, regulators, molecular interaction networks, and biological functions through bioinformatics, *Brain Sci.* 13 (2023) 151, <https://doi.org/10.3390/brainsci13010151>.
- [34] O.J. Ziff, J. Neeves, J. Mitchell, G. Tyzack, C. Martinez-Ruiz, R. Luisier, et al., Integrated transcriptome landscape of ALS identifies genome instability linked to TDP-43 pathology, *Nat. Commun.* 14 (2023) 2176, <https://doi.org/10.1038/s41467-023-37630-6>.
- [35] S. Mathis, C. Goizet, A. Soulages, J.-M. Vallat, G.L. Masson, Genetics of amyotrophic lateral sclerosis: a review, *J. Neurol. Sci.* 399 (2019) 217–226, <https://doi.org/10.1016/j.jns.2019.02.030>.
- [36] S. Reina, S.C. Nibali, M.F. Tomasello, A. Magri, A. Messina, V. De Pinto, Voltage Dependent Anion Channel 3 (VDAC3) protects mitochondria from oxidative stress, *Redox Biol.* 51 (2022) 102264, <https://doi.org/10.1016/j.redox.2022.102264>.
- [37] M.G.G. Pittalà, R. Saletti, S. Reina, V. Cunsolo, V. De Pinto, S. Foti, A high resolution mass spectrometry study reveals the potential of disulfide formation in human mitochondrial voltage-dependent anion selective channel isoforms (hVDACs), *Int. J. Mol. Sci.* 21 (2020) E1468, <https://doi.org/10.3390/ijms21041468>.
- [38] M. Proske, R. Janowski, S. Bacher, H.-S. Kang, T. Monecke, T. Köhler, et al., PURA Syndrome-causing mutations impair PUR-domain integrity and affect P-body association, *Elife* 13 (2024), <https://doi.org/10.7554/eLife.93561.2>.
- [39] K.H. Vousden, D.P. Lane, p53 in health and disease, *Nat. Rev. Mol. Cell Biol.* 8 (2007) 275–283, <https://doi.org/10.1038/nrm2147>.
- [40] D.R. Smith, C.M. Stanley, T. Foss, R.G. Boles, K. McKernan, Rare genetic variants in the endocannabinoid system genes CNR1 and DAGLA are associated with neurological phenotypes in humans, *PLoS One* 12 (2017) e0187926, <https://doi.org/10.1371/journal.pone.0187926>.
- [41] B. Khalil, M. Linsenmeier, C.L. Smith, J. Shorter, W. Rossoll, Nuclear-import receptors as gatekeepers of pathological phase transitions in ALS/FTD, *Mol. Neurodegener.* 19 (2024) 8, <https://doi.org/10.1186/s13024-023-00698-1>.
- [42] A. Birger, M. Ottolenghi, L. Perez, B. Reubinoff, O. Behar, ALS-related human cortical and motor neurons survival is differentially affected by Sema3A, *Cell Death Dis.* 9 (2018) 256, <https://doi.org/10.1038/s41419-018-0294-6>.
- [43] C. Marques, M. Fischer, C. Keime, T. Burg, A. Brunet, J. Scekic-Zahirovic, et al., Early alterations of RNA metabolism and splicing from adult corticospinal neurons in an ALS mouse model, *bioRxiv* (2019) 667733, <https://doi.org/10.1101/667733>.
- [44] Markid T. Yeganeh, M.A. Hosseinpour Feizi, M. Talebi, M. Rezaadeh, M. Khalaj-Kondori, Gene expression investigation of four key regulators of polyadenylation and alternative adenylation in the periphery of late-onset Alzheimer's disease patients, *Gene* 895 (2024) 148013, <https://doi.org/10.1016/j.gene.2023.148013>.
- [45] J.A. Chen, Z. Chen, H. Won, A.Y. Huang, J.K. Lowe, K. Wojta, et al., Joint genome-wide association study of progressive supranuclear palsy identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases, *Mol. Neurodegener.* 13 (2018) 41, <https://doi.org/10.1186/s13024-018-0270-8>.
- [46] M. Mallik, M. Catinozzi, C.B. Hug, L. Zhang, M. Wagner, J. Bussmann, et al., Xrp1 genetically interacts with the ALS-associated FUS orthologue caz and mediates its toxicity, *J. Cell Biol.* 217 (2018) 3947, <https://doi.org/10.1083/jcb.201802151>.

- [47] J. Root, P. Merino, A. Nuckols, M. Johnson, T. Kukar, Lysosome dysfunction as a cause of neurodegenerative diseases: lessons from frontotemporal dementia and amyotrophic lateral sclerosis, *Neurobiol. Dis.* 154 (2021) 105360, <https://doi.org/10.1016/j.nbd.2021.105360>.
- [48] B.M. Edens, N. Miller, Y.-C. Ma, Impaired autophagy and defective mitochondrial function: converging paths on the road to motor neuron degeneration, *Front. Cell. Neurosci.* 10 (2016) 44, <https://doi.org/10.3389/fncel.2016.00044>.
- [49] D. Gagliardi, N. Bresolin, G.P. Comi, S. Corti, Extracellular vesicles and amyotrophic lateral sclerosis: from misfolded protein vehicles to promising clinical biomarkers, *Cell. Mol. Life Sci.* 78 (2021) 561–572, <https://doi.org/10.1007/s00018-020-03619-3>.
- [50] Z. Jiang, W. Chen, J. Zhou, Q. Peng, H. Zheng, Y. Yuan, et al., Identification of COMMD1 as a novel lamin A binding partner, *Mol. Med. Rep.* 20 (2019) 1790–1796, <https://doi.org/10.3892/mmr.2019.10419>.
- [51] R. Weiskirchen, L.C. Penning, COMMD1, a multi-potent intracellular protein involved in copper homeostasis, protein trafficking, inflammation, and cancer, *J. Trace Elem. Med. Biol.* 65 (2021) 126712, <https://doi.org/10.1016/j.jtemb.2021.126712>.
- [52] J.S. Lum, M.L. Brown, N.E. Farrowell, L. McAlary, D. Ly, C.G. Chisholm, et al., CuATSM improves motor function and extends survival but is not tolerated at a high dose in SOD1G93A mice with a C57BL/6 background, *Sci. Rep.* 11 (2021) 19392, <https://doi.org/10.1038/s41598-021-98317-w>.
- [53] B.J. Gill, D.J. Pisapia, H.R. Malone, H. Goldstein, L. Lei, A. Sonabend, et al., MRI-localized biopsies reveal subtype-specific differences in molecular and cellular composition at the margins of glioblastoma, *Proc. Natl. Acad. Sci. U. S. A.* 111 (2014) 12550–12555, <https://doi.org/10.1073/pnas.1405839111>.
- [54] L.I. Grad, J.J. Yerbury, B.J. Turner, W.C. Guest, E. Pokrishevsky, M.A. O'Neill, et al., Intercellular propagated misfolding of wild-type Cu/Zn superoxide dismutase occurs via exosome-dependent and -independent mechanisms, *Proc. Natl. Acad. Sci. U. S. A.* 111 (2014) 3620–3625, <https://doi.org/10.1073/pnas.1312245111>.
- [55] E. Gasser, H.C. Johannssen, T. Rüllicke, H.U. Zeilhofer, M. Stoffel, Foxa1 is essential for development and functional integrity of the subthalamic nucleus, *Sci. Rep.* 6 (2016) 38611, <https://doi.org/10.1038/srep38611>.
- [56] J. Lee, Y. Wu, Y. Qi, H. Xue, Y. Liu, D. Scheel, et al., Neurogenin3 participates in gliogenesis in the developing vertebrate spinal cord, *Dev. Biol.* 253 (2003) 84–98, <https://doi.org/10.1006/dbio.2002.0868>.
- [57] A.L. Carcagno, D.J. Di Bella, M. Goulding, F. Guillemot, G.M. Lanuza, Neurogenin3 restricts serotonergic neuron differentiation to the hindbrain, *J. Neurosci.* 34 (2014) 15223–15233, <https://doi.org/10.1523/JNEUROSCI.3403-14.2014>.
- [58] J.A. Blake, M.R. Ziman, Pax genes: regulators of lineage specification and progenitor cell maintenance, *Development* 141 (2014) 737–751, <https://doi.org/10.1242/dev.091785>.
- [59] L. Sampieri, P. Di Giusto, C. Alvarez, CREB3 transcription factors: ER-Golgi stress transducers as hubs for cellular homeostasis, *Front. Cell Dev. Biol.* 7 (2019) 123, <https://doi.org/10.3389/fcell.2019.00123>.
- [60] H. Kayserili, E. Uz, C. Niessen, I. Vargel, Y. Alanay, G. Tuncbilek, et al., ALX4 dysfunction disrupts craniofacial and epidermal development, *Hum. Mol. Genet.* 18 (2009) 4357–4366, <https://doi.org/10.1093/hmg/ddp391>.
- [61] P. Mondola, S. Damiano, A. Sasso, M. Santillo, The Cu, Zn superoxide dismutase: not only a dismutase enzyme, *Front. Physiol.* 7 (2016) 594, <https://doi.org/10.3389/fphys.2016.00594>.
- [62] Y. Furukawa, E. Tokuda, Does wild-type Cu/Zn-superoxide dismutase have pathogenic roles in amyotrophic lateral sclerosis? *Transl. Neurodegener.* 9 (2020) 33, <https://doi.org/10.1186/s40035-020-00209-y>.
- [63] S.M. Ranuncolo, L. Wang, J.M. Polo, T. Dell'Oso, J. Dierov, T.J. Gaymes, et al., BCL6-mediated attenuation of DNA damage sensing triggers growth arrest and senescence through a p53-dependent pathway in a cell context-dependent manner, *J. Biol. Chem.* 283 (2008) 22565–22572, <https://doi.org/10.1074/jbc.M803490200>.
- [64] S. Pedrini, D. Sau, S. Guareschi, M. Bogush, R.H. Brown, N. Nanche, et al., ALS-linked mutant SOD1 damages mitochondria by promoting conformational changes in Bcl-2, *Hum. Mol. Genet.* 19 (2010) 2974–2986, <https://doi.org/10.1093/hmg/ddq202>.
- [65] S. Sun, Y. Sun, S.-C. Ling, L. Ferraiuolo, M. McAlonis-Downes, Y. Zou, et al., Translational profiling identifies a cascade of damage initiated in motor neurons and spreading to glia in mutant SOD1-mediated ALS, *Proc. Natl. Acad. Sci. U. S. A.* 112 (2015) E6993–E7002, <https://doi.org/10.1073/pnas.1520639112>.
- [66] J. Aruga, N. Yokota, M. Hashimoto, T. Furuichi, M. Fukuda, K. Mikoshiba, A novel zinc finger protein, zic, is involved in neurogenesis, especially in the cell lineage of cerebellar granule cells, *J. Neurochem.* 63 (1994) 1880–1890, <https://doi.org/10.1046/j.1471-4159.1994.63051880.x>.
- [67] T. Nagase, R. Kikuno, K.I. Ishikawa, M. Hirose, O. Ohara, Prediction of the coding sequences of unidentified human genes. XVI. The complete sequences of 150 new cDNA clones from brain which code for large proteins in vitro, *DNA Res.* 7 (2000) 65–73, <https://doi.org/10.1093/dnares/7.1.65>.
- [68] S. Sragovich, N. Amram, A. Yeheskel, I. Gozes, VIP/PACAP-Based drug development: the ADNP/NAP-Derived mirror peptides SKIP and D-SKIP exhibit distinctive in vivo and in silico effects, *Front. Cell. Neurosci.* 13 (2019) 589, <https://doi.org/10.3389/fncel.2019.00589>.
- [69] I. Gozes, The ADNP syndrome and CP201 (NAP) potential and hope, *Front. Neurol.* 11 (2020) 608444, <https://doi.org/10.3389/fneur.2020.608444>.
- [70] Chehreh Bargh S. Rashid, A. Tafakhori, F. Masoumi, F. Rahmani, M. Ahmadi, A. Namdar, et al., Evaluation of regulatory T lymphocytes and IL2Ra and FOXP3 gene expression in peripheral mononuclear cells from patients with amyotrophic lateral sclerosis, *Ir. J. Med. Sci.* 187 (2018) 1065–1071, <https://doi.org/10.1007/s11845-018-1793-2>.
- [71] N. Prakash, E. Puellas, K. Freude, D. Trümbach, D. Omodei, M. Di Salvo, et al., Nkx6-1 controls the identity and fate of red nucleus and oculomotor neurons in the mouse midbrain, *Development* 136 (2009) 2545–2555, <https://doi.org/10.1242/dev.031781>.
- [72] M. Misra, E. Sours, C. Lance-Jones, Hox transcription factors influence motoneuron identity through the integrated actions of both homeodomain and non-homeodomain regions, *Dev. Dynam.* 241 (2012) 718–731, <https://doi.org/10.1002/dvdy.23763>.
- [73] H. Doucet-Beaupré, C. Gilbert, M.S. Profes, A. Chabrat, C. Pacelli, N. Giguère, et al., Lmx1a and Lmx1b regulate mitochondrial functions and survival of adult midbrain dopaminergic neurons, *Proc. Natl. Acad. Sci. U. S. A.* 113 (2016) E4387–E4396, <https://doi.org/10.1073/pnas.1520387113>.
- [74] T.C. Badea, H. Cahill, J. Ecker, S. Hattar, J. Nathans, Distinct roles of transcription factors brn3a and brn3b in controlling the development, morphology, and function of retinal ganglion cells, *Neuron* 61 (2009) 852–864, <https://doi.org/10.1016/j.neuron.2009.01.020>.
- [75] Y. Kawamura, K. Yamanaka, B. Poh, H. Kuribayashi, H. Koso, S. Watanabe, The role of Zfx2 transcription factor in bipolar cell differentiation during mouse retinal development, *Biochem. Biophys. Res. Commun.* 503 (2018) 3023–3030, <https://doi.org/10.1016/j.bbrc.2018.08.088>.
- [76] S. Piras, A.L. Furfaro, L. Brondolo, M. Passalacqua, U.M. Marinari, M.A. Pronzato, et al., Differentiation impairs Bach1 dependent HO-1 activation and increases sensitivity to oxidative stress in SH-SY5Y neuroblastoma cells, *Sci. Rep.* 7 (2017) 7568, <https://doi.org/10.1038/s41598-017-08095-7>.
- [77] M. Lacroix, G. Rodier, O. Kirsh, T. Houles, H. Delpech, B. Seyran, et al., E4F1 controls a transcriptional program essential for pyruvate dehydrogenase activity, *Proc. Natl. Acad. Sci. U. S. A.* 113 (2016) 10998–11003, <https://doi.org/10.1073/pnas.1602754113>.
- [78] H. Sakai, Y. Fujii, N. Kuwayama, K. Kawaji, Y. Gotoh, Y. Kishi, Plag1 regulates neuronal gene expression and neuronal differentiation of neocortical neural progenitor cells, *Gene Cell.* 24 (2019) 650–666, <https://doi.org/10.1111/gtc.12718>.
- [79] F. Zinghirino, X.G. Pappalardo, A. Messina, G. Nicosia, V. De Pinto, F. Guarino, VDAC genes expression and regulation in mammals, *Front. Physiol.* 12 (2021) 708695, <https://doi.org/10.3389/fphys.2021.708695>.
- [80] G. Rozenfeld-Granot, J. Krishnamurthy, K. Kannan, A. Toren, N. Amariglio, D. Givol, et al., A positive feedback mechanism in the transcriptional activation of Apaf-1 by p53 and the coactivator Zac-1, *Oncogene* 21 (2002) 1469–1476, <https://doi.org/10.1038/sj.onc.1205218>.
- [81] S.R. Burstein, H.J. Kim, J.A. Fels, L. Qian, S. Zhang, P. Zhou, et al., Estrogen receptor beta modulates permeability transition in brain mitochondria, *Biochim. Biophys. Acta Bioenerg.* 1859 (2018) 423–433, <https://doi.org/10.1016/j.bbabi.2018.03.006>.
- [82] G. Bai, I. Cheung, H.P. Shulha, J.E. Coelho, P. Li, X. Dong, et al., Epigenetic dysregulation of hairy and enhancer of split 4 (HES4) is associated with striatal degeneration in postmortem Huntington brains, *Hum. Mol. Genet.* 24 (2015) 1441–1456, <https://doi.org/10.1093/hmg/ddu561>.
- [83] C.D. Thompson, B. Matta, B.J. Barnes, Therapeutic targeting of IRFs: pathway-dependence or structure-based? *Front. Immunol.* 9 (2018) 2622, <https://doi.org/10.3389/fimmu.2018.02622>.
- [84] Y. Li, L. Yu, C. Liu, W. Duan, S. Zhang, Z. Li, et al., IRF5 knockdown reverses TDP-related phenotypes partially by increasing TBK1 expression, *Brain Res.* 1798 (2023) 148155, <https://doi.org/10.1016/j.brainres.2022.148155>.

- [85] X. Chen, Y. Yao, J. Guan, X. Chen, F. Zhang, Up-regulation of FoxN4 expression in adult spinal cord after injury, *J. Mol. Neurosci.* 52 (2014) 403–409, <https://doi.org/10.1007/s12031-013-0166-6>.
- [86] L. Cheng, C.Y.J. Quek, L.W. Hung, R.A. Sharples, N.A. Sherratt, K.J. Barnham, et al., Gene dysregulation is restored in the Parkinson's disease MPTP neurotoxic mice model upon treatment of the therapeutic drug Cull(atsm), *Sci. Rep.* 6 (2016) 22398, <https://doi.org/10.1038/srep22398>.
- [87] L. Snijders Blok, T. Kleefstra, H. Venselaar, S. Maas, H.Y. Kroes, A.M.A. Lachmeijer, et al., De novo variants disturbing the transactivation capacity of POU3F3 cause a characteristic neurodevelopmental disorder, *Am. J. Hum. Genet.* 105 (2019) 403–412, <https://doi.org/10.1016/j.ajhg.2019.06.007>.
- [88] E. Kennedy, T.M. Everson, T. Punshon, B.P. Jackson, K. Hao, L. Lambertini, et al., Copper associates with differential methylation in placenta from two US birth cohorts, *Epigenetics* 15 (2019) 215–230, <https://doi.org/10.1080/15592294.2019.1661211>.
- [89] L. Yuan, C. Tian, H. Wang, S. Song, D. Li, G. Xing, et al., Apak competes with p53 for direct binding to intron 1 of p53AIP1 to regulate apoptosis, *EMBO Rep.* 13 (2012) 363–370, <https://doi.org/10.1038/embor.2012.10>.
- [90] J. Lin, P. Huang, W. Chen, C. Ye, H. Su, X. Yao, Key molecules and pathways underlying sporadic amyotrophic lateral sclerosis: integrated analysis on gene expression profiles of motor neurons, *Front. Genet.* 11 (2020) 578143, <https://doi.org/10.3389/fgene.2020.578143>.
- [91] T. Hosaka, H. Tsuji, A. Tamaoka, Biomolecular modifications linked to oxidative stress in amyotrophic lateral sclerosis: determining promising biomarkers related to oxidative stress, *Processes* 9 (2021) 1667, <https://doi.org/10.3390/pr9091667>.
- [92] M. Marku, V. Pancaldi, From time-series transcriptomics to gene regulatory networks: a review on inference methods, *PLoS Comput. Biol.* 19 (2023) e1011254, <https://doi.org/10.1371/journal.pcbi.1011254>.
- [93] W. Liu, X. Sun, L. Peng, L. Zhou, H. Lin, Y. Jiang, RWRNET: a gene regulatory network inference algorithm using random walk with restart, *Front. Genet.* 11 (2020) 591461, <https://doi.org/10.3389/fgene.2020.591461>.
- [94] W. Liu, Y. Jiang, L. Peng, X. Sun, W. Gan, Q. Zhao, et al., Inferring gene regulatory networks using the improved markov blanket discovery algorithm, *Interdiscip. Sci.* 14 (2022) 168–181, <https://doi.org/10.1007/s12539-021-00478-9>.
- [95] S. van Dam, U. Vösa, A. van der Graaf, L. Franke, J.P. de Magalhães, Gene co-expression analysis for functional classification and gene-disease predictions, *Briefings Bioinf.* 19 (2018) 575–592, <https://doi.org/10.1093/bib/bbw139>.
- [96] A. Canosa, A. Martino, A. Giuliani, C. Moglia, R. Vasta, M. Grassano, et al., Brain metabolic differences between pure bulbar and pure spinal ALS: a 2-[18F] FDG-PET study, *J. Neurol.* 270 (2023) 953–959, <https://doi.org/10.1007/s00415-022-11445-9>.
- [97] Y. Liu, M. Ding, S. Pan, R. Zhou, J. Yao, R. Fu, et al., MicroRNA-23a-3p is upregulated in plasma exosomes of bulbar-onset ALS patients and targets ERBB4, *Neuroscience* 524 (2023) 65–78, <https://doi.org/10.1016/j.neuroscience.2023.05.030>.
- [98] J.-L.C. Mougeot, Z. Li, A.E. Price, F.A. Wright, B.R. Brooks, Microarray analysis of peripheral blood lymphocytes from ALS patients and the SAFE detection of the KEGG ALS pathway, *BMC Med. Genom.* 4 (2011) 74, <https://doi.org/10.1186/1755-8794-4-74>.
- [99] MdR. Rahman, T. Islam, F. Huq, J.M.W. Quinn, M.A. Moni, Identification of molecular signatures and pathways common to blood cells and brain tissue of amyotrophic lateral sclerosis patients, *Inform. Med. Unlocked* 16 (2019) 100193, <https://doi.org/10.1016/j.imu.2019.100193>.
- [100] W.S. Tae, J.H. Sung, S.H. Baek, C.N. Lee, B.J. Kim, Shape analysis of the subcortical nuclei in amyotrophic lateral sclerosis without cognitive impairment, *J. Clin. Neurol.* 16 (2020) 592–598, <https://doi.org/10.3988/jcn.2020.16.4.592>.
- [101] R.H. Chipika, F. Christidi, E. Finegan, Hi Shing S. Li, M.C. McKenna, K.M. Chang, et al., Amygdala pathology in amyotrophic lateral sclerosis and primary lateral sclerosis, *J. Neurol. Sci.* 417 (2020) 117039, <https://doi.org/10.1016/j.jns.2020.117039>.
- [102] D.B. Castellanos, C.A. Martín-Jiménez, F. Rojas-Rodríguez, G.E. Barreto, J. González, Brain lipidomics as a rising field in neurodegenerative contexts: perspectives with Machine Learning approaches, *Front. Neuroendocrinol.* 61 (2021) 100899, <https://doi.org/10.1016/j.yfrne.2021.100899>.
- [103] T.J. Tracey, F.J. Steyn, E.J. Wolvetang, S.T. Ngo, Neuronal lipid metabolism: multiple pathways driving functional outcomes in health and disease, *Front. Mol. Neurosci.* 11 (2018) 10, <https://doi.org/10.3389/fnmol.2018.00010>.
- [104] J. Wu, X. Fang, H. Huang, W. Huang, L. Wang, X. Xia, Construction and topological analysis of an endometriosis-related exosomal circRNA-miRNA-mRNA regulatory network, *Aging (Albany NY)* 13 (2021) 12607–12630, <https://doi.org/10.18632/aging.202937>.
- [105] X. Chen, T. Wang, Y. Guan, Q. Ouyang, C. Lou, L. Qian, The topological characteristics of biological ratio-sensing networks, *Life* 13 (2023) 351, <https://doi.org/10.3390/life13020351>.
- [106] N.A. Shah, C.A. Sarkar, Variable cellular decision-making behavior in a constant synthetic network topology, *BMC Bioinf.* 20 (2019) 237, <https://doi.org/10.1186/s12859-019-2866-6>.
- [107] M. Naumann, A. Pal, A. Goswami, X. Lojewski, J. Japtok, A. Vehlou, et al., Impaired DNA damage response signaling by FUS-NLS mutations leads to neurodegeneration and FUS aggregate formation, *Nat. Commun.* 9 (2018) 335, <https://doi.org/10.1038/s41467-017-02299-1>.
- [108] H. Wang, S. Rangaswamy, M. Kodavati, J. Mitra, W. Guo, E.N. Guerrero, et al., RT2 PCR array screening reveals distinct perturbations in DNA damage response signaling in FUS-associated motor neuron disease, *Mol. Brain* 12 (2019) 103, <https://doi.org/10.1186/s13041-019-0526-4>.
- [109] J. Mitra, M.L. Hegde, A commentary on TDP-43 and DNA damage response in amyotrophic lateral sclerosis, *J. Exp. Neurosci.* 13 (2019) 1179069519880166, <https://doi.org/10.1177/1179069519880166>.
- [110] V.E. Provasek, A. Bacolla, S. Rangaswamy, J. Mitra, M. Kodavati, I.O. Yusuf, et al., RNA/DNA binding protein TDP43 regulates DNA mismatch repair genes with implications for genome stability, *bioRxiv* 5 (16) (2024) 594552, <https://doi.org/10.1101/2024.05.16.594552>, 2024.
- [111] A. Ainslie, W. Huiting, L. Barazzuol, S. Bergink, Genome instability and loss of protein homeostasis: converging paths to neurodegeneration? *Open Biol* 11 (2021) 200296, <https://doi.org/10.1098/rsob.200296>.
- [112] C.C. Rodgers, Low-dose X-ray imaging may increase the risk of neurodegenerative diseases, *Med. Hypotheses* 142 (2020) 109726, <https://doi.org/10.1016/j.mehy.2020.109726>.
- [113] C. Bailly, Potential use of edaravone to reduce specific side effects of chemo-, radio- and immuno-therapy of cancers, *Int. Immunopharm.* 77 (2019) 105967, <https://doi.org/10.1016/j.intimp.2019.105967>.
- [114] M. Grassano, C. Moglia, F. Palumbo, E. Koumantakis, P. Cugnascio, S. Callegaro, et al., Sex differences in amyotrophic lateral sclerosis survival and progression: a multidimensional analysis, *Ann. Neurol.* 96 (2024) 159–169, <https://doi.org/10.1002/ana.26933>.
- [115] M.J. Workman, R.G. Lim, J. Wu, A. Frank, L. Ornelas, L. Panther, et al., Large-scale differentiation of iPSC-derived motor neurons from ALS and control subjects, *Neuron* 111 (2023) 1191–1204.e5, <https://doi.org/10.1016/j.neuron.2023.01.010>.
- [116] S.M. Jacob, S. Lee, S.H. Kim, K.A. Sharkey, G. Pfeffer, M.D. Nguyen, Brain-body mechanisms contribute to sexual dimorphism in amyotrophic lateral sclerosis, *Nat. Rev. Neurol.* 20 (2024) 475–494, <https://doi.org/10.1038/s41582-024-00991-7>.
- [117] E. Pollari, G. Goldsteins, G. Bart, J. Koistinaho, R. Giniatullin, The role of oxidative stress in degeneration of the neuromuscular junction in amyotrophic lateral sclerosis, *Front. Cell. Neurosci.* 8 (2014) 131, <https://doi.org/10.3389/fncel.2014.00131>.
- [118] N. Jankovska, R. Matej, Molecular pathology of ALS: what we currently know and what important information is still missing, *Diagnostics* 11 (2021) 1365, <https://doi.org/10.3390/diagnostics11081365>.
- [119] J. Eshima, S.A. O'Connor, E. Marschall, NYGC ALS Consortium, R. Bowser, C.L. Plaisier, et al., Molecular subtypes of ALS are associated with differences in patient prognosis, *Nat. Commun.* 14 (2023) 95, <https://doi.org/10.1038/s41467-022-35494-w>.
- [120] M. MacLean, J. Juraneck, S. Cuddapah, R. López-Diez, H.H. Ruiz, J. Hu, et al., Microglia RAGE exacerbates the progression of neurodegeneration within the SOD1G93A murine model of amyotrophic lateral sclerosis in a sex-dependent manner, *J. Neuroinflammation* 18 (2021) 139, <https://doi.org/10.1186/s12974-021-02191-2>.
- [121] R.I. Amann, S. Baichoo, B.J. Blencowe, P. Bork, M. Borodovsky, C. Brooksbank, et al., Toward unrestricted use of public genomic data, *Science* 363 (2019) 350–352, <https://doi.org/10.1126/science.aaw1280>.
- [122] J. Sun, J.J. Carrero, U. Zagai, M. Evans, C. Ingre, Y. Pawitan, et al., Blood biomarkers and prognosis of amyotrophic lateral sclerosis, *Eur. J. Neurol.* 27 (2020) 2125–2133, <https://doi.org/10.1111/ene.14409>.
- [123] J.K. Kühlwein, W.P. Ruf, K. Kandler, S. Witzel, C. Lang, M.A. Mulaw, et al., ALS is imprinted in the chromatin accessibility of blood cells, *Cell. Mol. Life Sci.* 80 (2023) 131, <https://doi.org/10.1007/s00018-023-04769-w>.

- [124] K.A. Staats, D.R. Borchelt, M.G. Tansey, J. Wymer, Blood-based biomarkers of inflammation in amyotrophic lateral sclerosis, *Mol. Neurodegener.* 17 (2022) 11, <https://doi.org/10.1186/s13024-022-00515-1>.
- [125] S. Zucca, S. Gagliardi, C. Pandini, L. Diamanti, M. Bordoni, D. Sproviero, et al., RNA-Seq profiling in peripheral blood mononuclear cells of amyotrophic lateral sclerosis patients and controls, *Sci. Data* 6 (2019) 190006, <https://doi.org/10.1038/sdata.2019.6>.
- [126] N. Grima, S. Liu, D. Southwood, L. Henden, A. Smith, A. Lee, et al., RNA sequencing of peripheral blood in amyotrophic lateral sclerosis reveals distinct molecular subtypes: considerations for biomarker discovery, *Neuropathol. Appl. Neurobiol.* 49 (2023) e12943, <https://doi.org/10.1111/nan.12943>.
- [127] S. Köks, K. Rallmann, M. Muldmaa, J. Price, A.L. Pfaff, P. Taba, Whole blood transcriptome profile identifies motor neurone disease (MND) RNA biomarker signatures, *bioRxiv* 9 (27) (2024) 615372, <https://doi.org/10.1101/2024.09.27.615372>, 2024.
- [128] P.F. Sullivan, C. Fan, C.M. Perou, Evaluating the comparability of gene expression in blood and brain, *Am J Med Genet B Neuropsychiatr Genet* 141B (2006) 261–268, <https://doi.org/10.1002/ajmg.b.30272>.
- [129] U.G. Vijayakumar, V. Milla, M.Y. Cynthia Stafford, A.J. Bjourson, W. Duddy, S.M.-R. Duguez, A systematic review of suggested molecular strata, biomarkers and their tissue sources in ALS, *Front. Neurol.* 10 (2019) 400, <https://doi.org/10.3389/fneur.2019.00400>.
- [130] W. van Rheenen, A. Shatunov, A.M. Dekker, R.L. McLaughlin, F.P. Diekstra, S.L. Pulit, et al., Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis, *Nat. Genet.* 48 (2016) 1043–1048, <https://doi.org/10.1038/ng.3622>.
- [131] D. Kim, A. Tran, H.J. Kim, Y. Lin, J.Y.H. Yang, P. Yang, Gene regulatory network reconstruction: harnessing the power of single-cell multi-omic data, *NPJ Systems Biology and Applications* 9 (2023) 51, <https://doi.org/10.1038/s41540-023-00312-6>.
- [132] R.P.A. van Eijk, H.-J. Westeneng, S. Nikolakopoulos, I.E. Verhagen, M.A. van Es, M.J.C. Eijkemans, et al., Refining eligibility criteria for amyotrophic lateral sclerosis clinical trials, *Neurology* 92 (2019) e451–e460, <https://doi.org/10.1212/WNL.0000000000006855>.
- [133] N.A. Goyal, J.D. Berry, A. Windebank, N.P. Staff, N.J. Maragakis, L.H. van den Berg, et al., Addressing heterogeneity in amyotrophic lateral sclerosis CLINICAL TRIALS, *Muscle Nerve* 62 (2020) 156–166, <https://doi.org/10.1002/mus.26801>.
- [134] T. Lanciano, A. Savino, F. Porcu, D. Cittaro, F. Bonchi, P. Provero, Contrast subgraphs allow comparing homogeneous and heterogeneous networks derived from omics data, *GigaScience* 12 (2022) giad010, <https://doi.org/10.1093/gigascience/giad010>.
- [135] S. Uygun, C. Peng, M.D. Lehti-Shiu, R.L. Last, S.-H. Shiu, Utility and limitations of using gene expression data to identify functional associations, *PLoS Comput. Biol.* 12 (2016) e1005244, <https://doi.org/10.1371/journal.pcbi.1005244>.
- [136] L.D. Miller, P.M. Long, L. Wong, S. Mukherjee, L.M. McShane, E.T. Liu, Optimal gene expression analysis by microarrays, *Cancer Cell* 2 (2002) 353–361, [https://doi.org/10.1016/s1535-6108\(02\)00181-2](https://doi.org/10.1016/s1535-6108(02)00181-2).
- [137] K. Manjang, S. Tripathi, O. Yli-Harja, M. Dehmer, G. Glazko, F. Emmert-Streib, Prognostic gene expression signatures of breast cancer are lacking a sensible biological meaning, *Sci. Rep.* 11 (2021) 156, <https://doi.org/10.1038/s41598-020-79375-y>.