Revealing Clusters of Connected Pathways Through Multisource Data Integration in Huntington's Disease and Spastic Ataxia

Andrea C. Kakouri, Christiana C. Christodoulou, Margarita Zachariou, Anastasis Oulas, George Minadakis, Christiana A. Demetriou, Christina Votsi, Eleni Zamba-Papanicolaou, Kyproula Christodoulou, and George M. Spyrou, *Senior Member, IEEE*

***Abstract*—The advancement of scientific and medical research over the past years has generated a wealth of experimental data from multiple technologies, including genomics, transcriptomics, proteomics, and other forms of –omics data, which are available for a number of diseases. The integration of such multisource data is a key component toward the success of precision medicine. In this paper, we are investigating a multisource data integration method developed by our group, regarding its ability to drive to clusters of connected pathways under two different approaches: first, a disease-centric approach, where we integrate data around a disease, and second, a gene-centric approach, where we integrate data around a gene. We have used as a paradigm for the first approach Huntington's disease (HD), a disease with a plethora of available data, whereas for the second approach the *GBA2,* a gene that is related to spastic ataxia (SA), a phenotype with sparse availability of data. Our paper shows that valuable information at the level of disease-related pathway clusters can be obtained for both HD and SA. New pathways that classical pathway analysis methods were unable to reveal, emerged as necessary "connectors" to build connected pathway stories formed as pathway clusters. The capability to integrate multisource molecular data, concluding to something more than the sum of the existing information, empowers precision and personalized medicine approaches.**

***Index Terms*—Network integration, systems bioinformatics, precision medicine, Huntington's disease (HD), spastic ataxia (SA), GBA-2 related diseases.**

Manuscript received January 19, 2018; revised May 17, 2018 and July 2, 2018; accepted August 5, 2018. Date of publication August 29, 2018; date of current version January 2, 2019. The work of A. C. Kakouri, C. C. Christodoulou, M. Zachariou, A. Oulas, G. Minadakis, and G. M. Spyrou are supported by the European Commission Research Executive Agency (REA) under Grant BIORISE 669026, under the Spreading Excellence, Widening Participation, Science with and for Society Framework. This work was supported in part by H2020-WIDESPREAD-04-2017-Teaming Phase 1, under Grant 763781, Integrated Precision Medicine Technologies. (Andrea C. Kakouri and Christiana C. Christodoulou are co-first authors.) *(Corresponding author: George M. Spyrou.)*

A. C. Kakouri is with the Neurogenetics Department and Bioinformatics Group, the Cyprus Institute of Neurology and Genetics, Nicosia 1683, Cyprus (e-mail: andreak@cing.ac.cy).

C. C. Christodoulou is with the Neurology Clinic D and Bioinformatics Group, the Cyprus Institute of Neurology and Genetics, Nicosia 1683, Cyprus (e-mail: christianachr@cing.ac.cy).

M. Zachariou, A. Oulas, G. Minadakis, and G. M. Spyrou are with the Bioinformatics Group, the Cyprus Institute of Neurology and Genetics, Nicosia 1683, Cyprus (e-mail: margaritaz@cing.ac.cy; anastasioso@cing.ac.cy; georgem@cing.ac.cy; georges@cing.ac.cy).

C. A. Demetriou and E. Papanicolaou-Zamba are with the Neurology Clinic D, the Cyprus Institute of Neurology and Genetics, Nicosia 1683, Cyprus (e-mail: christianad@cing.ac.cy; ezamba@cing.ac.cy).

C. Votsi and C. Kyproula are with the Neurogenetics Department, the Cyprus Institute of Neurology and Genetics, Nicosia 1683, Cyprus (e-mail: votsi@cing.ac.cy; roula@cing.ac.cy).

All the authors are also with the Cyprus School of Molecular Medicine, Nicosia, Cyprus.

Digital Object Identifier 10.1109/JBHI.2018.2865569

I. Introduction

THE majority of human diseases are relatively complex and involve a combination of multiple genetic interactions, epigenetic, environmental and lifestyle factors. However, there are rare cases of human diseases such as HD and SA that are caused by a single gene. These are known as monogenic diseases [1], [2].

HD is a rare progressive neurodegenerative disease with autosomal dominant inheritance. HD is caused by a tri-nucleotide expansion of CAG repeats in the *huntingtin (HTT)* gene. Even though HD is caused by a single gene, the *HTT* protein interacts with various other proteins which is a challenge in terms of understanding the pathophysiology and molecular mechanisms involved in disease development as well as in the phenotypic variability observed in terms of age of onset and progression rate [3]. The disease age of onset is usually between 35–50 years old. However, there is substantial variation in the onset and progression of the disease observed among HD patients [4], which still remains to be explained.

A comprehensive range of several different molecular mechanisms have been suggested to describe how the CAG triplet expansion and HD mutation leads to cellular dysfunction and death. There is a likely connection between reactive oxygen species, neuro-inflammation, protein-dysfunction, abnormal protein-protein interactions and other mechanisms [3]. However, the precise molecular mechanisms responsible for HD pathogenesis that are induced by *HTT*, still remain unknown. Nevertheless, a lot of research focuses on the study of HD; therefore, there is a vast availability of multisource data that can be used for further elucidation of the pathogenetic mechanisms. Although a single causative gene is involved in the development of HD, multiple genetic interactions in combination with the

factors mentioned above, contribute to the complexity of this disease.

Another group of monogenic diseases are the autosomal recessive cerebellar ataxias (ARCA). ARCAs cover a clinically and genetically heterogeneous group of rare neurodegenerative diseases with an age of onset usually before the age of 20. This group of diseases usually affects the cerebellum, the spinocerebellar tract and/or the sensory tracts of the spinal cord [2], [5], [6]. The prominent clinical characteristics of ARCAs are progressive gait and limb ataxia, usually accompanied by other neurological or additional features [5], [7]. Hereditary spastic paraplegias (HSP) encompass a distinct group of neurodegenerative diseases, characterised by progressive spasticity and weakness of the lower limbs, which can be accompanied by additional neurological and non-neurological symptoms [7], [8]. The corticospinal tract and posterior columns are usually involved in the development of this group of diseases. Often, a partial clinical overlap between ARCA and HSP can be observed, which results in more complex phenotypes termed as SA [9]. The most commonly mutated genes in SA include the *SACS, FXN*, and *SPG7* [7]. Recently, the involvement of *GBA2* in SA has been reported [2], [10], [70]. The classification of ARCAs still remains controversial due to clinical overlap and high genetic heterogeneity reported within this group of diseases [11]. The *GBA2* gene codes for the non-lysosomal glucosylceramidase (GBA2), which is responsible for catalysing the conversion of the sphingolipid glucosylceramide to glucose and ceramide. Although the involvement of *GBA2* in the sphingolipid metabolism has been characterised, the molecular mechanisms that lead to the pathogenesis of SA are still unknown [12]. Thus, we can hypothesize that GBA2 is involved in the pathophysiology of SA by its role in the sphingolipid metabolism, but also by a different function that is still unknown, which contributes to a Sphingolipid metabolism-independent mechanism that drives the development and/or progression of SA.

The most complex and monogenic diseases consist of overlap at the pathway level, genetic interactions and profiles and at the molecular mechanisms that are involved in their pathophysiology [13]. The advancement of scientific and medical research over the past years, has generated a wealth of experimental data from multiple sources such as genomics, transcriptomics, proteomics, and other forms of –omics data, which are available for a number of diseases [14]. Such data are utilized by scientists, clinicians and bioinformaticians to investigate the dysregulation occurring at the different –omics levels and provide clarification in the underlying molecular mechanisms, genes and aetiology of the disease [14], [15]. The human genome is not only complex but also regulated at multiple levels. Events taking place in the cell are not only interdependent, but also interactive and therefore, this provides a challenge in the integration of heterogeneous data to discover relevant biological signatures and predict phenotypical characteristics [16]. To address the challenge of data integration, a new sub-discipline of bioinformatics known as systems bioinformatics, has emerged. Systems bioinformatics is a powerful approach, which involves computational and mathematical modelling of biological systems. It is a multidisciplinary field approach, which integrates knowledge and information from several fields, including biology, chemistry, computer science and mathematics, in order to provide a complete understanding on the biological mechanisms involved in a disease [17]. More specifically, systems bioinformatics works by integrating data derived from different levels (genes, differentially expressed genes (DEGs), protein-protein interactions (PPIs), single nucleotide polymorphisms (SNPs) and other forms of genetic variation, pathways, drugs) and sources in the form of a network. This enables the extraction of information from complex biological networks that were previously unobtainable with the normal bioinformatics analyses of –omics data (for a comprehensive review of Systems Bioinformatics approaches the reader is referred to [17]). Systems Bioinformatics approaches can contribute towards the goals of precision and personalised medicine by facilitating (1) the development of a more customized approach based on patient-specific clinical, genetic, environmental and lifestyle profiles, (2) disease-risk estimation and prediction (3) delay of disease onset and progression, (4) effective treatments, (5) identification of putative drug targets and (6) computational drug discovery and repurposing.

Networks have become a valuable tool in bioinformatics, allowing the researcher to visualize in a simple manner the relationship between genes, metabolites, proteins, drugs and signalling pathways obtained from experimental data and available literature, and to identify meaning on the molecular mechanisms by which a gene can affect the development and/or progression of a disease [18]. A network is made up of entities known as "nodes" and the links between them, known as "edges". The nodes can represent genes, proteins, miRNAs, metabolites and small molecules, while the edges (links) can represent metabolic, regulatory, protein-protein interactions, co-expression, co-localization and signalling pathways. The network offers a simple snapshot of cellular states [18]. Furthermore, the application of a network approach in human diseases, plays a vital role in representing the different levels of data and knowledge, and provides further characterization of complex biological patterns, therefore, enabling the generation and visualization of numerous network types, such as, disease-pathway and disease-drug [16], [19].

In this work we employed a multi-source data integration method developed by our group [20] in order to explore clusters of connected pathways for two very different cases in terms of data availability, (a) HD, a disease with a plethora of available data, and (b) SA, a disease with sparse data availability. For HD we implemented the integration as it was previously done for Alzheimer's' Disease (AD) using a *disease-centric* approach, where data were integrated around a disease [20]. For SA we adapted our method to a *gene-centric* approach, where we integrated data with respect to a specific gene related to SA (GBA2). Previous studies have demonstrated different multi-omics integration approaches. For example, Brandi *et al*. (2017), Dimitrakopoulos *et al.* (2018) and Koscielny *et al.* (2017) have applied their methods to study HD, cancer and drug targeting respectively. The novelty of our method [20]. lies in the fact that, the gene-specific scores are calculated not only based on the sum of available information for each gene, but also takes into account the topology of the super-network constructed by integrating all the types of data (see Methods for more details).

Our work is structured in the following manner. Firstly we retrieved data using Malacards as our source for HD and GBA2-related diseases, including genes, pathways, variants, drugs and differentially expressed genes (DEGs). Secondly we constructed a super network aggregating the different information sources. From that a gene-specific score was obtained through the combination of both the topology of the super network and the gene-specific information gene-set enrichment analysis was performed using the top scored genes for each disease and top resulting pathways were mapped on the KEGG pathway-pathway reference network. Finally, additional pathways were added in order to create a sub- network of minimally connected pathways for each disease and clustering was performed to group the final set of pathways.

II. Methods

A. Data Resources

Data for HD and SA were downloaded in November 2017 from the human disease database MalaCards [21]. In MalaCards the search term Huntington's disease was used as the parent term (MalaCards ID HNT016). In the case of HD, the availability of multi-source information, allowed for the extraction of specific genetic, molecular and therapeutic information related to HD.

In the case of SA, due to the fact that the disease name has recently been introduced and detailed characterisation and classification is still pending, we could not retrieve enough information in order to proceed with our methodology. Therefore, we shifted our approach from disease-centric to gene-centric and selected diseases that are correlated to the *GBA2* gene. *GBA2* (GeneCard ID GC09M035726) was searched in GeneCards [22] and eight related diseases were identified: i) Spastic Paraplegia 46, Autosomal Recessive (MalaCards ID SPS109), ii) Spastic Paraplegia 56, Autosomal Recessive (MalaCards ID SPS101), iii) Autosomal Recessive Cerebellar Ataxia (MalaCards ID ATS307) and iv) Gaucher's Disease (MalaCards ID GCH001), v) Spastic Ataxia (MalaCards ID SPS008) vi) Autosomal recessive cerebellar ataxia with late onset spasticity (MalaCards ID ATS112), vii) spastic paraplegia 46 (MalaCards ID SPS174) and viii) splenic sequestration (MalaCards ID SPL009). Splenic sequestration was excluded as it is unrelated to neurodegenerative diseases. (MalaCards ID ATS112) and spastic paraplegia 46 (MalaCards ID SPS174) were also excluded, as no additional information was retrieved from these two searches. This is due to the fact that there is a large clinical and genetic overlap in ARCAs, and this may lead to the categorisation of the same disease as two different diseases or vice versa. We named the selected group of diseases '*GBA2*-related diseases' in order to proceed with our approach. Information on the *GBA2*-related diseases was retrieved from MalaCards using the parent search terms i) - v), as detailed above. The list of genes associated with each *GBA2*-related disease (according to MalaCards) can be found in Table I.

TABLE I
List of Genes Associated With *GBA2*-Related Diseases, According to Malacards

Spastic ataxia	Autosomal recessive cerebellar ataxia	Gaucher's disease	Spastic paraplegia 46, autosomal recessive	Spastic paraplegia 56, autosomal recessive
SACS	*SYNE1*	*GBA*	***GBA2***	*CYP2U1*
KIF1C	*SETX*	*PSAP*	*CYP2U1*	*DDHD2*
MTPAP	*ANO10*	*CHIT1*	*SPG21*	***GBA2***
AFG3L2	*AFP*	*CBA3*	*COX6A1*	*AP4S1*
VAMP1	*SNX14*	*SCARB2*	*TRIM2*	*RTN2*
MARS2	*COQ8A*	*PKLR*	*DNAJB2*	*COX6A1*
SPG7	*PMPCA*	*MTX1*		*AP4B1*
GBA2	*TDP1*	*ACP5*		
	SPG7	*ARSA*		
	CWF19L1	***GBA2***		
	VLDLR	*BGLAP*		
	GBA2	*UGCG*		
	UBA5	*CCL18*		
	FXN	*MRC1*		
	ATM	*SNCA*		
	STUB1			
	APTX			
	GRID2			
	ZNF592			

For both HD and *GBA2*-related diseases, multiple sources of information were obtained from MalaCards, based on the: (i) Genes involved, (ii) Pathways related, (iii) Variants-SNPs and Copy Number Variations (CNVs), (iv) Drugs related and (v) Differentially expressed genes.

MalaCards provided a list of drugs for HD which was parsed with a local copy of DrugBank [23] in order to extract the information of known gene targets for each drug. The same approach was applied to *GBA2*-related diseases, and known gene-drug targets were only obtained for spastic ataxia, autosomal recessive cerebellar ataxia and Gaucher's disease.

For HD, non-coding RNA and more specifically microRNA (miRNA) regulatory information was retrieved from three miRNA-gene targets databases: 1) MirTarBase [24], 2) MicroCosm [25] and 3) TargetScan [26]. miRNAs that had been previously associated with HD were retained from the mir2Disease database [27] and parsed with the miRNA-gene target lists downloaded from the three databases mentioned above.

In the case of *GBA2*-related diseases, the list of genes that were associated with this group of diseases was analysed in Cytoscape [28] along with the Cytoscape plugin CyTargetLinker [29] in order to identify the target genes. The miRNA-gene target list for *GBA2*-related diseases was further validated using the three miRNA-gene target databases mentioned above (mirtarbase 6.1, microcosm-2012-12-05, and targetscan-2012-12-05). Only the miRNA-gene targets which were found in at least two out of three databases were retained for further analysis. This was also applied in the case of HD miRNA-gene target list.

For both HD and *GBA2*-related diseases, SNPs-gene targets lists as well as pathway information were also obtained from

MalaCards. It is important to note that, only the super-pathways involved along with the pathway score and genes involved were obtained. In the case of *GBA2*-related diseases, each disease was searched in MalaCards separately and the results were combined to get a single list for each type of data.

GeneMANIA [30] web interface was used to extract protein-protein interactions and genetic interactions of the genes (MalaCards) associated with HD and *GBA2*-related diseases with other known genes. GeneMANIA [30] uses the BioGRID [31] and PathwayCommons [32] databases to retrieve the above mentioned data.

B. Network Construction

Following the methodology proposed by [20], we constructed a different network per data source available for HD and *GBA2*-related diseases using Cytoscape. Six different data sources were obtained for both diseases and used for the construction of six corresponding networks. The biological networks produced were of the following types: 1) gene-drug network, 2) gene-pathway network, 3) gene-variant (SNPs and CNVs) network, 4) gene-microRNA network, 5) gene-gene network (based on PPIs) and 6) gene-gene network (based on genetic interactions).

In order to retrieve gene-specific information, we transformed all networks to a gene-gene network. Thus, three new networks (in addition to the two existing gene-gene networks i.e., PPI and genetic interaction network) were constructed in which the genes were associated based on the common drugs, pathways and miRNAs they share. These networks were then added into the two remaining networks from the PPI and genetic interaction network. Two tables were constructed including all pairwise combinations of genes involved in both HD and *GBA2*-related diseases, as well as the weighted edge from the 5 types of biological networks mentioned above (for pairs of genes with no available information the value 0 was entered in the table). In addition to the gene-gene networks, gene-specific information for example for drugs (how many target each gene), miRNAs (how many target each gene), pathways (how many involve each gene), variations (how many found on each gene) and differential expression was extracted where available for both cases.

C. The Multi-Source Information Super-Network

We constructed a Multi-source Information (MI) super network having weighted edges (E_{MI}) obtained by the weighted sum of the edge vectors for each pair of genes:

$$E_{MI} = w^e_{Drug} * E_{Drug} + w^e_{Mir} * E_{Mir} + w^e_{Path} * E_{Path} + w^e_{PPI} * E_{PPI} + w^e_{GI} * E_{GI} \quad (1)$$

Here, for a given pair of genes, E_{Drug} is the total number of disease-related drugs which target both genes, E_{Mir} is the total number of disease-related miRNAs which target both genes, E_{Path} is the total number of disease-related pathways associated with both genes, E_{PPI} is the total number of PPIs that both disease related genes are involved and E_{GI} is in binary form with value one if the two disease related genes interact genetically.

The edge vectors E_X $(x \in \{Drug, Mir, Path, PPI, GI\})$ are normalised per vector to be canonical (from 0 to 1). As in our previous work, all respective sources were considered to be equivalent and their weights w^e_x were set to 1 divided by the number of sources so that their sum satisfies the condition

$$\sum_x w^e_x = 1,\ x \in \{Drug, Mir, Path, PPI, GI\}. \quad (2)$$

The Multi-source Information Gain (MIG) is a characteristic score per gene and is comprised by two parts,

$$MIG = w * MIG_n + (1 - w) * MIG_e, \quad (3)$$

where the first term MIG_n represents the normalised integrated gene-specific information (i.e., node characteristics) and MIG_e represents the normalised integrated gene-gene information (based on the topology of the MI super network) and corresponds to the weighted degree of the MI super network. In this study, we have considered equal contribution to the score of the gene-specific information and of the topology of the integrated gene-gene super network.

The gene-specific information is given by

$$MI\,G_n = w^n_{Drug} * N_{Drug} + w^n_{Mir} * N_{Mir} + w^n_{Path} * N_{Path} + w^n_{Var} * N_{Var} + w^n_{Deg} * FC \quad (4)$$

where N_{Drug} is the total number of disease-related drugs which target the gene, N_{Mir} in the total number of disease-related miRNAs which target the gene, N_{Path} is the total number of disease-related pathways associated with a gene, N_{Var} the total number of genetic variations associated with disease-related (both CNVs and SNPs) and FC is the absolute fold change of the DEGs associated with disease-related (both in brain and in blood). Again, all respective sources were considered to be equivalent and their weights w^n_x were set to 1/(Number of sources) so that their sum satisfies the condition:

$$\sum_x w^n_x = 1,\ x \in \{Drug, Mir, Path, Var, Deg\}. \quad (5)$$

The igraph package in R [33] was used to generate and analyze the MI super network and to compile the MIG score.

D. Pathway Analysis

Gene-set enrichment analysis was performed using the EnrichR [34] package in R. The enrichment was obtained using "KEGG2016" as the database option. The enrichment analysis was performed on the genes of the sorted gene list with a score equal or larger to the median of the MIG score. We calculated a pathway-specific (PS) score [20] for each of the pathways obtained through enrichment analysis based on the sum of the MIG scores of each gene participating in the respective pathway,

$$PS_j = \sum_i MIG^i, \quad \mathrm{i} = 1 : \mathrm{N_j}, \quad (6)$$

where $\mathrm{N_j}$ is the number of genes within in our data set which are part of the j pathway.

E. KEGG Pathway-Pathway Reference Network

We have used an overall human pathway-pathway network (as developed in [20] using R language) in order to use it as a reference network map for retrieving information about the

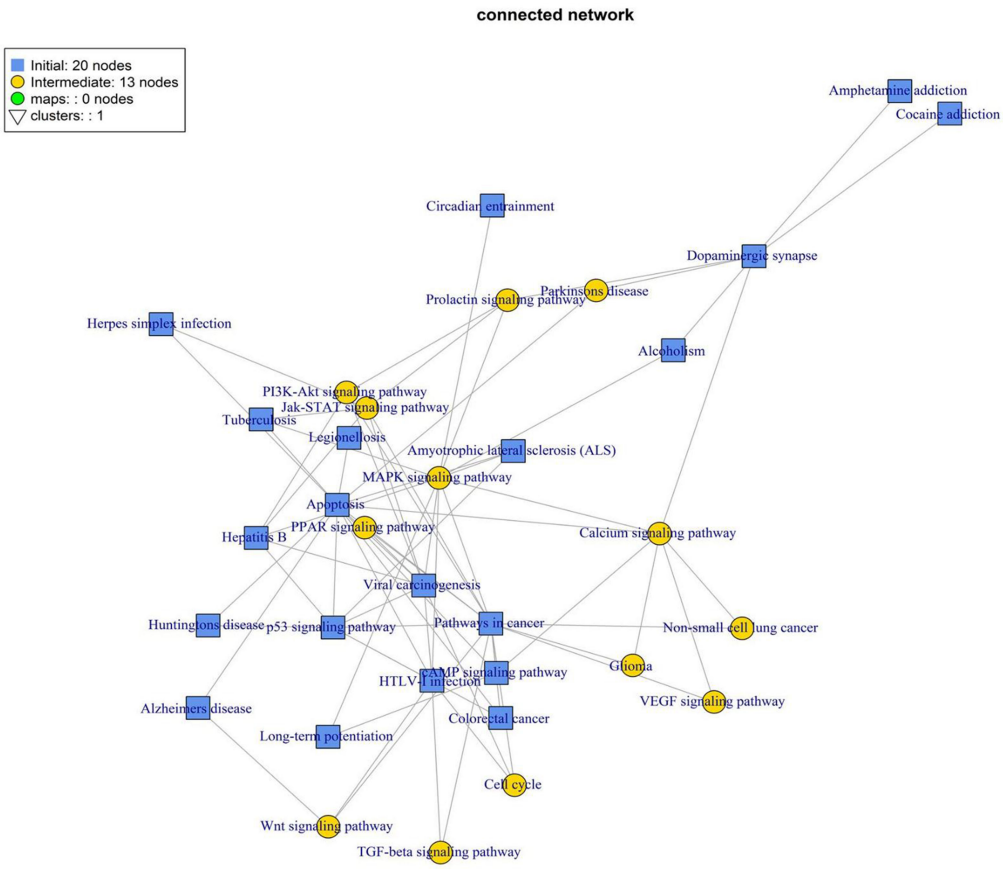

Fig. 1. Network of HD enriched pathways mapped on KEGG network. The top 20 pathways obtained from the enrichment analysis are indicated by the blue squares. The complementary pathways obtained by the shortest path algorithm are indicated by the yellow nodes.

functional interconnections between pathways of interest. The reference pathway-network is based on the functional connection among pathways found in KEGG [35], a well-established reference database for pathways and contains 314 human pathways as well as 91 maps. Links to non-human pathways have been excluded.

F. Identifying and Adding Missing Pathways

Following the pathway enrichment analysis, the top resulting pathways with a certain cutoff (20 top pathways based on the EnrichR combined score for HD and 10 top pathways for GBA2-related diseases) were mapped on the KEGG pathway-pathway reference network. This mapping resulted in a subnetwork of important for the disease pathways which were linked to each other based on the functional connections from the KEGG reference network of pathways. As the resulting subnetworks were not fully connected we implemented an approach to ensure minimum connectivity, i.e., that each node/pathway is connected to at least one other node. By identifying all the possible connections that may exist between the pathways of interest we ensured the conservation of connections that may be important for the specific disease under study. This process led to the inclusion of additional pathways, leading to an overall network where all the pathways are fully connected to each other. In order to achieve this goal for a given set of pathways (nodes), a specific algorithm finds and calculates all the shortest paths within the reference network in order to interconnect all the nodes with at least one link to another node. After this stage, the algorithm chooses the extra pathway nodes that belong to the calculated shortest paths to be included in the final subnetwork. In the case that there is more than one shortest path with the same length then all possible shortest paths are included into the final subnetwork. In order to divide the fully connected subnetwork to sub-groups of pathways we employed a clustering method based on the edge-betweenness property of networks. For more details on the shortest-path method and the clustering method that we used, see [20].

III. Results & Discussion

A. HD and GBA2-Related Diseases Enriched Pathways Mapped on KEGG Networks

Enrichment analysis using EnrichR as mapped on the KEGG database reference network revealed pathways suggested to be related to HD and *GBA2*-related diseases, such as Spastic Ataxia.

1) HD: Following the integration of the HD-related multi-source data, the top-scored genes with a score equal or larger to the median of the MIG score were used as a gene- set for enrichment analysis. The top 20 HD-related enriched pathways (sorted by EnrichR combined score) obtained from enrichment were then mapped on the KEGG reference pathway-pathway network. The result of this mapping is illustrated in Fig. 1 where

TABLE II
PATHWAYS FOR HD

Pathway	KEGG reference	Combined score	Path sum score	Participation in cluster
INITIAL PATHWAYS				
Huntington's Disease	hsa05016	90.122	5.177	2
Amyotrophic Lateral Sclerosis	hsa05014	42.054	1.616	2
Hepatitis B	hsa05161	31.466	2.584	3
Alzheimer's Disease	hsa05010	28.013	1.996	4
Viral Carcinogenesis	hsa05203	27.236	2.564	3
Herpes infection	hsa05168	25.710	2.092	2
Apoptosis	hsa04210	23.747	1.700	2
Cocaine addiction	hsa05030	22.836	1.532	6
Long term potentiation	hsa04720	20.825	1.068	1
cAMP signalling pathway	hsa04024	19.276	1.959	1
Tuberculosis	hsa05152	18.967	2.422	3
p53 signalling pathways	hsa04115	18.625	1.518	2
Circadain entrainment	hsa04713	18.314	1.278	7
HTLV-1	hsa05166	16.765	1.575	6
Dopaminergic synapse	hsa04728	15.619	1.278	6
Pathways in cancer	hsa05200	12.906	1.946	1
Alcholism	hsa05034	12.836	1.532	6
Legionellosis	hsa05134	12.357	1.356	5
Colorectal cancer	hsa05210	12.327	1.035	2
Amphetamine addiction	hsa05031	12.174	1.096	6
COMPLEMENTARY PATHWAYS INCLUDED IN INITIAL ENRICHMENT ANALYSIS				
MAPK signalling pathway	hsa04010	5.557	1.305	3
TGF-beta signalling pathway	hsa04350	3.785	0.795	1
Wnt signalling pathway	hsa04310	2.350	0.589	4
Cell cycle	hsa04110	2.331	0.589	3
Jak-STAT signalling pathway	hsa04630	2.229	0.619	3
Parkinson's Disease	hsa05012	2.109	0.873	6
Calcium signalling pathway	hsa04020	1.955	0.299	1
PI3K-Akt signalling pathway	hsa04151	0.425	0.799	3
Glioma	hsa05214	-1.894	0.161	1
Non-small cell lung cancer	hsa05233	-2.279	0.161	1
COMPLEMENTARY PATHWAYS NOT INCLUDED IN INITIAL ENRICHMENT ANALYSIS				
PPAR signalling pathway	hsa03320			1
VEGF signalling pathway	hsa04370			1
Prolactin signalling pathway	hsa04917			3

the above-mentioned 20 pathways are shown in blue squares. Complementary missing pathways were identified and added to the network based on the shortest path algorithm (represented by yellow circles in Fig. 1). Out of the 13 complementary pathways, only 10 were included in the enrichment analysis results, yet below the 20 pathway cut-off. The shortest path algorithm introduced 10 additional complementary pathways which are the following: MAPK signalling pathway, TGF-beta signalling pathway, Wnt signalling pathway, cell cycle, Jak-STAT signalling pathway, Parkinson's disease, Calcium signalling pathway, PI3K-Akt signalling pathway, Glioma and non-small cell lung cancer, the complementary pathways all seem to have a vast connection with HD. The three remaining pathways (PPAR signalling pathway, VEGF, signalling pathway and prolactin signalling pathway) were not included in the enrichment analysis results. The PPAR signalling pathway consists of three sub-types (PPARα, PPARγ, and PPARβ/δ). The PPARα subtype plays a role in energy metabolism. PPARγ subtype plays a role in cellular absorption of lipids through anabolic pathways. PPARγ co-activator 1α (PGC1α) it is a transcriptional regulator of genes that are involved in energy metabolism, the mHTT interacts with PGC1α, this leads to impairment of its function in HD. The dysfunctional energy metabolism, in HD patients may likely be contributed to the dysregulation of the PPAR signalling pathway [36]. For VEGF, signalling pathway, genetic studies have shown that reduced VEGF levels can cause neurodegeneration due the impairing of neural tissue perfusion, which then causes hypoxia which in turn leads to neuronal cell death. The perfusion deficits often leads to the onset of clinical symptoms [37]. For prolactin signalling pathway, published

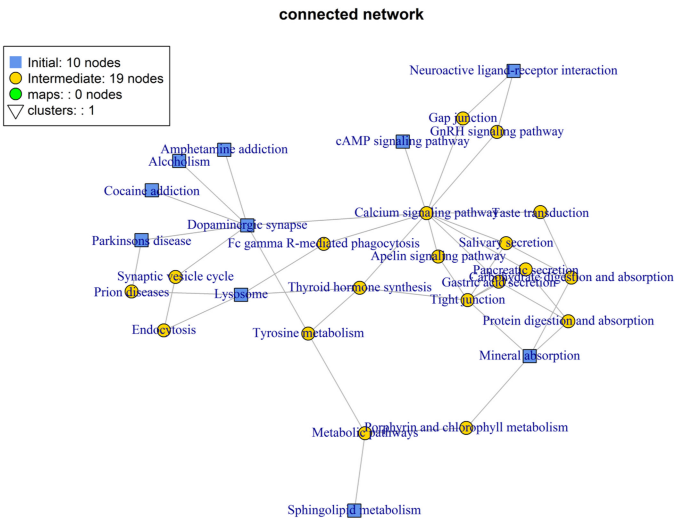

Fig. 2. Network of SA enriched pathways mapped on KEGG network. The top 10 pathways obtained from the enrichment analysis are indicated by the blue squares. The complementary pathways obtained by the shortest path algorithm are indicated by the yellow nodes.

literature suggests that pre-manifesting HD patients had decreased levels of prolactin, which is expected since HD affects both metabolic function and hormone secretion as well as energy regulation. Overall, these results could possibly shed light on the pathogenesis of the disease and offer prospects for biomarker development [38]. As anticipated the Huntington's disease pathway (KEGG entry hsa05016) was identified to be on the top of the list of enriched pathways with a p-value of 2.573846e-21 and combined score of 90.1225008. Other well characterized Huntington's related pathways that were also included in the top scoring EnrichR pathways are shown in Table II along with the p-value, combined score, path sum score.

An interesting finding was that other neurodegenerative disease pathways were also identified in the top scoring pathways, such as Amyotrophic lateral sclerosis (ALS) (hsa05014, p-value, 9.213013e-11 and combined score 42.0546949) and Alzheimer's disease (hsa05010, p-value 1.300089e-07 and combined score 28.0135632).

The above mentioned HD associated pathways additionally received high PS scores according to the methodology proposed by [20].

2) GBA2-Related Diseases: We selected the top 10 pathways for *GBA2*-related diseases according to EnrichR combined score (indicated as blue nodes in Fig. 2). Interestingly, the Sphingolipid Metabolism (KEGG entry hsa00600) was not on top of the list of enriched pathways, but it was the fourth, with combined score and mean path score 21.731 and 2.402 respectively (Table III). We would have expected the Sphingolipid metabolism to be on the top of the list, as the function of GBA2 enzyme is primarily correlated with sphingolipids. As mentioned previously, the *GBA2* gene codes for the non-lysosomal β-glucosylceramidase, an enzyme that catalyses the hydrolysis of glucosylceramide (GlcCer) into glucose and ceramide. Although the function of this enzyme has been characterised, the involvement of Sphingolipid Metabolism in the development of Spastic ataxia and related neurodegenerative diseases of the cerebellum is still unclear. The Dopaminergic synapse (KEGG entry hsa04728) was listed as the top scoring pathway (Combined score: 31.498 and sum path score: 1.250), while second on the list of EnrichR pathways was the Lysosome pathway (KEGG entry hsa04142) (Combined score: 30.509 and sum path score: 4.149). Furthermore, the pathway of Parkinson's disease appeared in the top scored pathways (Combined score: 22.565 and sum path score: 1.338).

There were 19 additional pathways generated by the shortest path algorithm (represented as yellow nodes in Fig. 2), five out of which were identified through the enrichment analysis.

From the five complementary pathways identified through the enrichment analysis, three show high correlation with the function of neurons: *1) Gap junction pathway* – Gap junctions are intercellular membrane channels that allow direct communication between adjacent cells in the developing and mature central nervous system (CNS). Hemi-channels and connexons are formed by connexin proteins that join together into hexamers. There is a variety of connexins expressed in CNS and peripheral nervous system (PNS), such as Cx32 in oligodentrocytes and microglia and Cx43 in microglia and endothelial cells. These are characterised as the main neuronal connexins and play a critical role in neurodevelopment. The function of Cx43 in gap junction communication between astrocytes is crucial, as it affects the proliferation and differentiation of neural stem cells [39], while Cx32 has been described to be have a role in the homeostasis of myelinated axons. Deficiencies in Cx32 have been implicated in cases of peripheral neuropathy. More specifically, mutations in Cx32 cause X-linked Charcot-Marie-Tooth, a group of inherited disorders that affect the peripheral nervous system [40]. Therefore mutations in genes coding for connexins or defects in the expression, synthesis, trafficking and turnover of connexins, can have an effect in cell death of neurons and subsequently in neurodevelopment and disease [39]. *2) Calcium signalling pathway* - Calcium (Ca^{2+}) is a ubiquitous second messenger, especially important in excitable cells, since it stimulates the release of neurotransmitter into the synaptic cleft [41]. Moreover, neuronal differentiation depends on calcium transients, which control neurotransmitter phenotype, dendritic morphology, and axonal growth and guidance [42]. Deregulation of Ca^{2+} signalling is a crucial aspect in the pathogenesis of neurodegenerative diseases, such as Alzheimer's disease (AD) and Parkinson's disease (PD) [43]. *3) Synaptic vesicle cycle pathway* – The release of neurotransmitter in the synaptic cleft is facilitated by exocytosis of synaptic vesicles, in order to mediate electrical impulses in the postsynaptic neuron. Synaptic vesicles undergo a trafficking cycle, in order to promote repeated rounds of neurotransmitter release [44]. The synaptic vesicle cycle pathway is affected in a number of neurodegenerative diseases, such as AD and PD, suggesting that deficiencies in this process might be a key process in disease pathogenesis.

The shortest path algorithm provided 14 pathways, which were not included in the EnrichR results (Table III). One of those pathways is the Apelin pathway. Apelin is an endogenous ligand for the G protein-coupled receptor APJ, both highly expressed in various parts of the CNS, including the cerebellum

TABLE III
PATHWAYS FOR *GBA2*-RELATED DISEASES

Pathway	KEGG reference	Combined score	Path sum score	Participation in cluster
INITIAL PATHWAYS				
Dopaminergic synapse	hsa04728	31.498	1.2500000	1
Lysosome	hsa04142	30.509	4.1493429	1
Parkinson's disease	hsa05012	22.565	1.338	1
Sphingolipid metabolism	hsa00600	21.731	2.402	3
Neuroactive ligand-receptor interaction	hsa04080	21.063	1.100	4
Cocaine addiction	hsa05030	19.391	0.750	1
Alcoholism	hsa05034	15.973	0.916	1
Amphetamine addiction	hsa05031	12.669	0.583	1
cAMP signaling pathway	hsa04024	6.994	0.500	2
Mineral absorption	hsa04978	6.741	0.333	2
COMPLEMENTARY PATHWAYS INCLUDED IN INITIAL ENRICHMENT ANALYSIS				
Gap junction	hsa04540	5.575	0.333	4
Metabolic pathways	hsa01100	5.273	2.608	3
Calcium signalling pathway	hsa04020	3.628	0.333	2
Porphyrin and chlorophyll metabolism	hsa00860	1.351	0.199	3
Synaptic vesicle cycle	hsa04721	1.050	0.166	1
COMPLEMENTARY PATHWAYS NOT INCLUDED IN INITIAL ENRICHMENT ANALYSIS				
Tyrosine metabolism	hsa00350			3
Thyroid hormone synthesis	hsa04918			2
Taste transduction	hsa04742			2
Carbohydrate digestion and absorption	hsa04973			2
Gastric acid secretion	hsa04971			2
Protein digestion and absorption	hsa04974			2
Pancreatic secretion	hsa04972			2
Salivary secretion	hsa04970			2
Tight junction	hsa04530			2
Prion diseases	hsa05020			1
Endocytosis	hsa04144			1
Fc gamma R-mediated phagocytosis	hsa04666			1
GnRH signalling pathway	hsa04912			4
Apelin signalling pathway	hsa04371			2

and spinal cord [45], [46]. The formation of the Apelin/APJ complex was reported to have a possible role in the protection against neuronal loss in the spinal cord. This was reported in relation to the pathogenesis of ALS [45]. In SA the cerebellum, spinocerebellar tract and/or the sensory tracts of the spinal cord are affected, suggesting the examination of any possible involvement of the Apelin pathway in the pathogenesis or progression of the disease. Furthermore, abnormalities in the Apelin/APJ system affect motor neurons and therefore the movement of patients. Similarly, SA is characterised by defects in movement as mentioned previously.

Another interesting pathway obtained, is the Prion disease pathway. Prion diseases comprise a group of fatal neurodegenerative diseases, characterized by abnormal folding of prior proteins into large amyloid plaques and fibrous structures, which accumulate in the nervous system [47]. Creutzfeldt-Jacob disease (CJD) is a human prion disease. The clinical features of CJD include progressive dementia, myoclonus, which is described as the involuntary bouncing of a muscle or a group of muscles and ataxia [48]. The molecular mechanisms involved in the development and progression of prion disease might benefit the understanding of SA pathophysiology, as the two diseases present an overlap in their clinical features.

B. Clusters of Pathways on HD and GBA2-Related Diseases

Following the enrichment analysis, the mapping on the KEGG database reference network and the addition of complementary pathways in order to create a sub-network of minimally connected pathways for both HD and SA diseases, clustering was performed to group the final set of pathways in clusters.

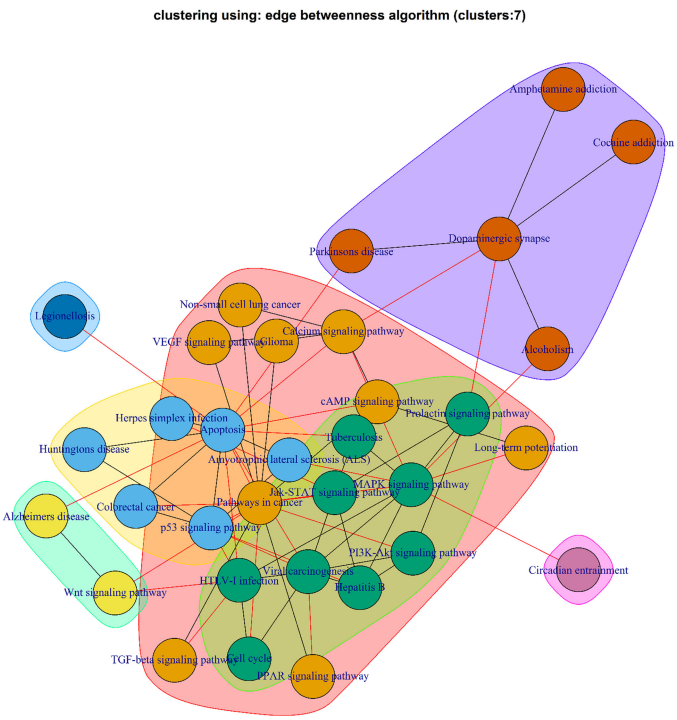

Fig. 3. Cluster pathways on HD. There is a total of seven clusters, each one shaded by a different colour.

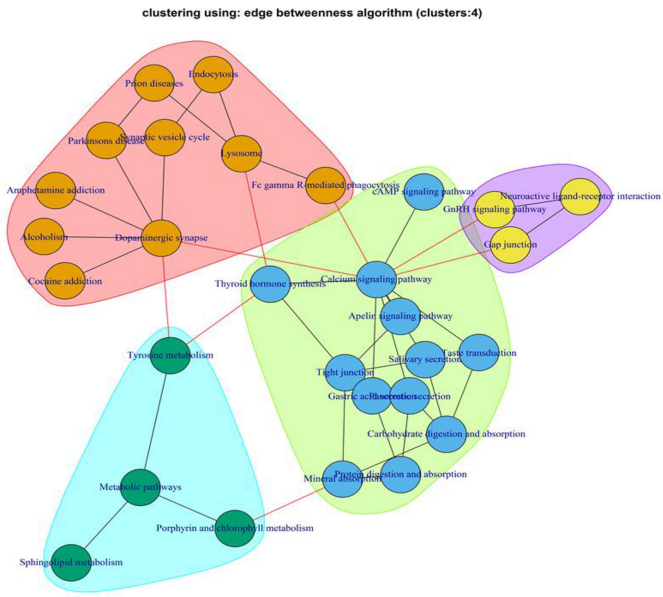

Fig. 4. Cluster pathways on *GBA2*-related diseases. There is a total of four clusters, each one shaded by a different colour.

Seven clusters were identified for HD and four clusters for *GBA2*-related diseases based on the network property of edge-betweenness. The clusters are shown in different coloured circles and backgrounds as seen in Fig. 3 for HD and Fig. 4 for *GBA2*-related diseases. *1)* HD

Cluster 1, was one of the largest clusters with nine pathways in total. Cluster 1 represents signalling pathways such as cAMP, TGF-beta, Ca^{2+}, PPAR and VEGF which all play a possible role in contributing to the disease pathogenesis and clinical characteristics of HD. Additional pathways in the cluster that are involved include long term potentiation, pathways involved in cancer, non-small cell lung cancer and glioma.

In cluster 1 cAMP signalling pathway and long term potentiation were identified from the initial pathway analysis. The Adenosine 3′,5′-cyclic monophosphate (cAMP) signalling pathway has a key role in cell survival and neuro-secretion. When there is a decrease in the intracellular concentration of cAMP this leads to a decrease in synaptic neurotransmission, thus, affecting neuronal transmission. Previous studies have demonstrated that there is a functional interaction between intracellular signalling pathways mediated by Ca^{2+} and cAMP (Ca^{2+}/cAMP signalling interaction) as both participate in various cellular responses, including neurotransmitter/hormone exocytosis, neuronal survival and protection [41], [49].

The dysregulation of calcium mediated signalling which is in cluster 1 has been associated with other neurodegenerative diseases such as AD (in cluster 4), PD (in cluster 6) and ALS (in cluster 2). Previous studies have demonstrated that repetitive application of glutamate causes elevations in cytosolic Ca^{2+} levels in medium spiny neurons (MSN), the enhanced cytosolic Ca^{2+} signalling in striatal MSN has been linked to HD neuropathology. The mechanism of Ca^{2+} toxicity in HD most likely involves the activation of calpains and excessive Ca^{2+} accumulation in mitochondria [49].

Long-term potentiation (LTP) (in cluster 1) is the persistent strengthening of synapses; there is a pattern of synaptic activity that produces a long lasting increase in the neuronal transmission between two neurons. In HD there is impairment of the LTP which is likely correlated to the size of the trinucleotide CAG repeat. The disruption of the LTP causes cognitive impairment, which is one of the earliest symptoms that is experienced in HD patients [50]. Furthermore, it is likely that the dopaminergic neurons are also affected which also adds to the cognitive impairment [50], and addictive behaviour such as Alcoholism, Amphetamine and cocaine addiction that is observed in cluster 6 along with dopaminergic synapses.

The transforming growth factor beta (TGF-Beta) signalling pathway (in cluster 1), is involved in a vast majority of cellular processes such as cell growth, differentiation, apoptosis, cellular homeostasis, neurotrophins and neuro-protection in developing embryos and adults. Alternations in this signalling pathway are not only linked to a number of human diseases, but also likely connect to HD [51].

A study found that asymptomatic HD patients showed a decrease in TGF-beta1 levels in the peripheral blood, this reduction is likely to be linked to the length of the CAG trinucleotide expansions [51]. TGF-beta 1 signalling is observed in other neurodegenerative diseases such as Alzheimer's disease (AD), were a reduction in TGF-beta 1 signalling increases the deposition of amyloid plaques [51].

The neurodegenerative disease, ALS (in cluster 2) appears in the same cluster as HD (in cluster 2). ALS and HD might appear different in terms of neuronal degeneration at first. For instance, in ALS there is degeneration in the upper and lower

motor-neurons and in HD there is degeneration in the medium spiny neurons of the basal ganglia which is located in the central nervous system (CNS). However, despite the differences of neuronal vulnerability, ALS and HD share some commonality in terms of their pathophysiological mechanisms such as inflammation, oxidative stress, protein aggregation (HTT for HD and TDP-43 in ALS) and mitochondrial dysfunction [52] and risk factors. Non-neuronal factors such as weight loss that is caused by a dysregulation in energy metabolism, in spite of increased food intake is a characteristic that is observed in HD and ALS patients, (this can be seen in cluster 2 of Fig. 2). Furthermore, apoptosis (in cluster 2) along with ALS and HD, is a common pathway involved in the already mentioned neurodegenerative diseases. The up-regulation of pro-apoptotic proteins (Bid and Bax) may possibly lead to the cleavage of HTT The resulting toxic fragment may promote toxic effects in the cell by disrupting transcription, therefore eliciting development of cell dysfunction and eventually cell death. In addition, increased levels of pro-apoptotic proteins may gradually contribute to neuronal cell death in HD as seen in Fig. 2 [53].

In cluster 4, AD is a neurodegenerative disease that leads to neuronal cell death, initially observed in the hippocampus. The hippocampus is located in the CNS, and plays a vital role in memory formation and storage. The main symptom of AD is dementia; HD patients also suffer from dementia in the late stages of the disease [54]. Although in HD the neurons of the basal ganglia are the first to undergo neuronal degeneration, as the disease progresses and worsens over time other areas of the brain also become affected. Notably, both AD and HD share some common key underlying molecular mechanisms, such as inflammation, oxidative stress, protein aggregation (HTT for HD and Aβ oligomers in AD) and mitochondrial dysfunction. Wnt signalling identified also in cluster 4, plays a role in neuroprotection, synaptic maintenance and neuronal functions under normal psychological conditions. Wnt signalling dysfunction has been linked to AD [55]. In HD, the mHTT is described as altering both the stability and levels of β-Catenin, an important molecule in the Wnt signalling with a role in cell adhesion and signal transduction. However, it still remains largely unknown whether dysfuction of Wnt signalling can cause HD pathology [56].

Cluster 3 contains the mitogen activated protein kinase (MAPK) signalling pathway. Kinase signalling pathways are assumed to contribute to HD pathophysiology as they are well-known to counter toxic metabolic changes induced by mHTT and assist in maintaining neuronal survival [57]. Alternations in the kinase signalling pathways possibly have implication in the maintenance of chronic neuro-inflammation observed in HD patients. Mutant HTT (mHTT) affects the signalling at the upstream portions of the ERK and JNK pathways which become activated. The modulation of the ERK pathway indicates that the pathway is linked to cell survival while the inhibition of JNK was demonstrated to effectively suppress pathogenesis. Previous studies have suggested that pharmacological intervention of the MAPK pathways and more specifically ERK activation may be used as a therapeutic intervention for HD [57].

2) GBA2-Related Diseases: The top three out of 10 pathways obtained by EnrichR analysis (Dopaminergic synapse pathway, Lysosome pathway and Parkinson's disease pathway) are present in the second largest cluster (cluster 1) of 10 pathways in total, including the previously mentioned synaptic vesicle cycle pathway.

PD is characterized by a loss of dopaminergic (DA) neurons in the substantia nigra (SN). A number of mechanisms may participate in the DA neurons damage and cell death, including cellular disturbances produced by misfolding and aggregation of the synaptic protein α-synuclein, as well as the disruption of macro-autophagy, which is responsible for the destruction and degradation of abnormal proteins [58]. In SN DA neurons from PD patients, autophagosomes and lysosomal marker proteins are increased and decreased, respectively, suggesting failure in the function of this system in these patients [59], [60], as well as in patients of other neurodegenerative disorders, such as AD, HD and ALS [61]. The system of autophagy-lysosome pathway provides explanation for the association of Parkinson's disease pathway with the Endocytosis and Lysosome pathways, which are found in the same cluster, as well as with the Prion disease which results after accumulation of abnormal proteins in the nervous system, as mentioned above. Moreover, studies have described the role of α-synuclein proteins in the regulation of the synaptic vesicle cycling [62], [63], thus explaining the association of the Parkinson's disease pathway and Synaptic vesicle pathway in the same cluster. PD has also been clinically linked with Gaucher disease, a neurodegenerative disease caused by mutations in the *GBA* gene. *GBA* codes for the lysosomal glucosylceramidase (GBA), which plays an important role in lysosomal protein degradation. Abnormalities in GBA can cause accumulation of α-synuclein proteins, leading to neurotoxicity and neuronal cell death [64]. Moreover, GlcCer has been reported to play a key role in the misfolding and aggregation of the synaptic protein α-synuclein, by stabilizing soluble oligomeric intermediates [64]. Thus, if impairment of the function of the GBA2 enzyme results in accumulation of GlcCer in neurons, then the above pathways can be studied to strengthen our understanding on the molecular mechanisms that lead to SA. The dopaminergic Synapse pathway is also present in cluster 1 and is associated with neurodegenerative disorders, including PD, HD and multiple sclerosis (MS). The neurotransmitter dopamine plays an important role in motor coordination, emotions, cognition, memory and other functions. Deficiencies in the dopaminergic system and its receptors in the basal ganglia structures is an important aspect in the pathogenesis of PD, as the loss of dopamine can result to imbalance in the motor networks and subsequently to abnormal stimulation or inhibition of movement [65]. This again may provide information on the causes of movement defects present in SA patients. Furthermore, the clinical features of PD include spasticity [66]. This overlap in the clinical phenotype between PD and SA provides further knowledge on the pathogenetic mechanisms of SA. The Sphingolipid pathway and Tyrosine metabolism are found in cluster 3. Studies have pointed the relevance of the enzyme tyrosine hydroxylase (TyrOH) in various affective disorders as well as neurological disorders, such as PD. TyrOH is responsible for the catalysis of the conversion of tyrosine to L-DOPA, which is part of the biosynthesis of the neurotransmitters dopamine, adrenaline,

and norandrenaline. Abnormalities in the biosynthesis of these molecules have been implicated as potential mediating factors in PD [67], [68]. Alterations in sphingolipid homeostasis have been implicated in several neurodegenerative diseases. Its role in the development of SA phenotype is still unknown. However, previous studies have reported that defects in ceramide *de novo* biosynthesis due to CERS1 deficiency are associated with: a) progressive neuronal death, particularly Purkinje cell loss in mice, and b) abnormal accumulation of lipofuscin with ubiquitylated proteins in many brain regions, thus suggesting that apart from neurodegeneration, alteration of can also accelerate some aspects of aging [69]. Ceramide is the main structural component of complex sphingolipids such as sphingomyelin and glycosphingolipids. Its homeostasis is primarily regulated by a *de novo* biosynthetic pathway catalysed mainly by the enzyme ceramide synthase 1 (CERS1) in the brain. However, the role of GBA2 is directly correlated with ceramide, as ceramide can result after degradation of glucosylceramide by GBA2. Moreover, according to the STRING online database, GBA2 and CERS1 are likely to have an interaction, which has not been experimentally confirmed yet. Therefore, we can hypothesize that *GBA2* mutations might exert their pathogenic effects by affecting this interaction, thus causing ceramide biosynthesis deficiency. Accumulation of GlcCer in the brain tissue can possibly affect the GSL homeostasis and survival or function of neuronal cells [10], [12]. Accumulation of glucosylceramides in the ER of neuronal cells would also change calcium homeostasis and lead to neurological symptoms [10]. Whether the presence of Tyrosine metabolism and Sphingolipid metabolism in the same cluster results because of the fact that they are both found under the umbrella of the super pathway Metabolism, or due to a possible role in the development and/or progression of SA remains unclear.

IV. Conclusion

In this work we have extended a recently developed integration approach [20] for two neurodegenerative disorders, namely HD and SA. Our approach was successfully applied to both diseases illustrating the potential and flexibility of this method. Our first contribution lies in the extension of the disease-centric approach method beyond diseases with a plethora of available data, such as AD and HD, to a gene-centric approach for diseases with limited data such as SA. Specifically for HD we used a disease-centric approach, where we integrated data around a disease whereas for SA we used a gene-centric approach focusing on GBA2-related diseases.

In the latter case, by modifying our method to be gene-centric and integrating information around the SA associated gene *GBA2,* we were able to amount sufficient data despite the sparse availability shed light to possible mechanisms underlying SA.

Our second contribution lies in enriching our knowledge through multisource data integration on the possible molecular mechanisms involved with regards to two rare diseases. Nevertheless, we acknowledge the limitation of lack of statistical evidence of our findings since there is lack of solid ground truth regarding the genes and the mechanisms that are implicated in these diseases. Correlation does not infer causality and thus wet lab experiments are required to validate the findings. Overall, our work can be used as a stepping stone for further developing methodologies which have the potential to facilitate research in the field of precision and personalised medicine.

References

[1] R. K. R. Kalathur, J. Pedro Pinto, B. Sahoo, G. Chaurasia, and M. E. Futschik, "HDNetDB: A molecular interaction database for network-oriented investigations into Huntington's disease," *Sci. Rep.*, vol. 7, no. 1, Dec. 2017, Art. no. 5216.

[2] C. Votsi, E. Zamba-Papanicolaou, L. T. Middleton, M. Pantzaris, and K. Christodoulou, "A novel GBA2 gene missense mutation in spastic ataxia," *Ann. Human Genetics*, vol. 78, no. 1, pp. 13–22, Jan. 2014.

[3] C. M. Everett and N. W. Wood, "Trinucleotide repeats and neurodegenerative disease," *Brain*, vol. 127, no. 11, pp. 2385–2405, Aug. 2004.

[4] D. C. Rubinsztein, "How does the Huntington's disease mutation damage cells?," *Sci. Aging Knowl. Environ.*, vol. 2003, no. 37, Sep. 2003, Art. no. PE26.

[5] B. L. Fogel and S. Perlman, "Clinical features and molecular genetics of autosomal recessive cerebellar ataxias," *Lancet Neurol.*, vol. 6, no. 3, pp. 245–257, Mar. 2007.

[6] S. Vermeer *et al.*, "Targeted next-generation sequencing of a 12.5 Mb homozygous region reveals ANO10 mutations in patients with autosomal-recessive cerebellar ataxia," *Amer. J. Human Genetics*, vol. 87, no. 6, pp. 813–819, Dec. 2010.

[7] S. T. de Bot, M. A. A. P. Willemsen, S. Vermeer, H. P. H. Kremer, and B. P. C. van de Warrenburg, "Reviewing the genetic causes of spastic-ataxias," *Neurology*, vol. 79, no. 14, pp. 1507–1514, Oct. 2012.

[8] S. Salinas, C. Proukakis, A. Crosby, and T. T. Warner, "Hereditary spastic paraplegia: Clinical features and pathogenetic mechanisms," *Lancet Neurol.*, vol. 7, no. 12, pp. 1127–1138, Dec. 2008.

[9] L. Ruano, C. Melo, M. C. Silva, and P. Coutinho, "The global epidemiology of hereditary ataxia and spastic paraplegia: A systematic review of prevalence studies," *Neuroepidemiology*, vol. 42, no. 3, pp. 174–183, 2014.

[10] M. Hammer *et al.*, "Mutations in GBA2 cause autosomal-recessive cerebellar ataxia with spasticity," *Amer. J. Human Genetics*, vol. 92, no. 2, pp. 245–251, Feb. 2013.

[11] C. Votsi and K. Christodoulou, "Molecular diagnosis of autosomal recessive cerebellar ataxia in the whole exome/genome sequencing era," *World J. Neurol.*, vol. 3, no. 4, pp. 115–128, 2013.

[12] M. Aureli *et al.*, "Current and novel aspects on the nonlysosomal beta-Glucosylceramidase GBA2," *Neurochem. Res.*, vol. 41, pp. 210–220, 2016.

[13] B. Shen, H.-B. Shen, T. Tian, Q. Lü, and G. Hu, "Translational bioinformatics and computational systems medicine," *Comput. Math. Methods Med.*, vol. 2013, May 2013, Art. no. 375641.

[14] D. Ayers and P. J. Day, "Systems medicine: The application of systems biology approaches for modern medical research and drug development," *Mol. Biol. Int.*, vol. 2015, pp. 1–8, Aug. 2015.

[15] A. Mardinoglu and J. Nielsen, "Systems medicine and metabolic modelling," *J. Internal Med.*, vol. 271, no. 2, pp. 142–154, Feb. 2012.

[16] S. Huang, K. Chaudhary, and L. X. Garmire, "More is better: Recent progress in multiomics data integration methods," *Frontiers Genetics*, vol. 8, Jun. 2017, Art. no. 84.

[17] A. Oulas *et al.*, "Systems bioinformatics: Increasing precision of computational diagnostics and therapeutics through network-based approaches," *Brief Bioinform.*, pp. 1–19, 2017.

[18] M. Suderman and M. Hallett, "Tools for visually exploring biological networks," *Bioinformatics*, vol. 23, no. 20, pp. 2651–2659, Oct. 2007.

[19] X. Ma, Z. Liu, Z. Zhang, X. Huang, and W. Tang, "Multiple network algorithm for epigenetic modules via the integration of genome-wide DNA methylation and gene expression data," *BMC Bioinf.*, vol. 18, no. 1, Dec. 2017, Art. no. 72.

[20] M. Zachariou, G. Minadakis, A. Oulas, S. Afxenti, and G. M. Spyrou, "Integrating multisource information on a single network to detect disease-related clusters of molecular mechanisms," *J. Proteomics*, vol. 188, pp. 15–29, Mar. 2018.

[21] N. Rappaport *et al.*, "MalaCards: An amalgamated human disease compendium with diverse clinical and genetic annotation and structured search," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D877–D887, Jan. 2017.

[22] G. Stelzer *et al.*, "The GeneCards suite: From gene data mining to disease genome sequence analyses," in *Current Protocols in Bioinformatics*, vol. 54. Hoboken, NJ, USA: Wiley, 2016, no. 1, pp. 1.30.1–1.30.33.

[23] D. S. Wishart *et al.*, "DrugBank 5.0: A major update to the DrugBank database for 2018," *Nucleic Acids Res.*, vol. 46, no. D1, pp. D1074–D1082, Jan. 2018.

[24] C.-H. Chou *et al.*, "miRTarBase update 2018: A resource for experimentally validated microRNA-target interactions," *Nucleic Acids Res.*, vol. 46, no. D1, pp. D296–D302, Jan. 2018.

[25] S. Griffiths-Jones, H. K. Saini, S. van Dongen, and A. J. Enright, "miRBase: Tools for microRNA genomics," *Nucleic Acids Res.*, vol. 36, no. Database, pp. D154–D158, Dec. 2007.

[26] B. P. Lewis, C. B. Burge, and D. P. Bartel, "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets," *Cell*, vol. 120, no. 1, pp. 15–20, Jan. 2005.

[27] Q. Jiang *et al.*, "miR2Disease: A manually curated database for microRNA deregulation in human disease," *Nucleic Acids Res.*, vol. 37, no. Database, pp. D98–D104, Jan. 2009.

[28] P. Shannon *et al.*, "Cytoscape: A software environment for integrated models of biomolecular interaction networks," *Genome Res.*, vol. 13, no. 11, pp. 2498–2504, Nov. 2003.

[29] M. Kutmon, F. Ehrhart, E. L. Willighagen, C. T. Evelo, and S. L. Coort, "CyTargetLinker app update: A flexible solution for network extension in Cytoscape," *F1000Research*, vol. 7, p. 743, Jun. 2018.

[30] D. Warde-Farley *et al.*, "The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function," *Nucleic Acids Res.*, vol. 38, no. suppl_2, pp. W214–W220, Jul. 2010.

[31] C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers, "BioGRID: A general repository for interaction datasets," *Nucleic Acids Res.*, vol. 34, pp. D535–D539, Jan. 2006, Art. no. 90001.

[32] E. G. Cerami *et al.*, "Pathway Commons, a web resource for biological pathway data," *Nucleic Acids Res.*, vol. 39, no. Database, pp. D685–D690, Jan. 2011.

[33] A. Gabor Csardi and M. Gabor Csardi, "The igraph package: Routines for simple graphs, network analysis," 2008.

[34] M. V. Kuleshov *et al.*, "Enrichr: A comprehensive gene set enrichment analysis web server 2016 update," *Nucleic Acids Res.*, vol. 44, no. W1, pp. W90–W97, Jul. 2016.

[35] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, "KEGG: New perspectives on genomes, pathways, diseases and drugs," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D353–D361, Jan. 2017.

[36] M. Kiaei, "Peroxisome proliferator-activated receptor-gamma in amyotrophic lateral sclerosis and Huntington's disease," *PPAR Res.*, vol. 2008, 2008, Art. no. 418765.

[37] E. Storkebaum and P. Carmeliet, "VEGF: A critical player in neurodegeneration," *J. Clin. Investigation*, vol. 113, no. 1, pp. 14–18, Jan. 2004.

[38] R. Wang *et al.*, "Metabolic and hormonal signatures in pre-manifest and manifest Huntington's disease patients," *Frontiers Physiol.*, vol. 5, Jun. 2014, Art. no. 231.

[39] A. B. Belousov, J. D. Fontes, M. Freitas-Andrade, and C. C. Naus, "Gap junctions and hemichannels: Communicating cell death in neurodevelopment and disease," *BMC Cell Biol.*, vol. 18, no. S1, Jan. 2017, Art. no. 4.

[40] K. A. Kleopa, "The role of gap junctions in charcot-marie-tooth disease," *J. Neurosci.*, vol. 49, pp. 17753–17760, 2011.

[41] E. Pchitskaya, E. Popugaeva, and I. Bezprozvanny, "Calcium signaling and molecular mechanisms underlying neurodegenerative diseases," *Cell Calcium*, vol. 17, pp. 87–94, Jun. 2017.

[42] S. S. Rosenberg and N. C. Spitzer, "Calcium signaling in neuronal development," *Cold Spring Harbor Perspective Biol.*, vol. 3, no. 10, Oct. 2011, Art. no. a004259.

[43] T.-S. Tang, X. Chen, J. Liu, and I. Bezprozvanny, "Dopaminergic signaling and striatal neurodegeneration in Huntington's disease," *J. Neurosci.*, vol. 27, no. 30, pp. 7899–7910, Jul. 2007.

[44] T. C. Südhof, "The synaptic vesicle cycle," *Annu. Rev. Neurosci.*, vol. 27, no. 1, pp. 509–547, Jul. 2004.

[45] A. Kasai *et al.*, "Apelin deficiency accelerates the progression of amyotrophic lateral sclerosis," *PLoS One*, vol. 6, no. 8, Aug. 2011, Art. no. e23968.

[46] B. Cheng, J. Chen, B. Bai, and Q. Xin, "Neuroprotection of apelin and its signaling pathway," *Peptides*, vol. 37, no. 1, pp. 171–173, Sep. 2012.

[47] A. Aguzzi, C. Sigurdson, and M. Heikenwaelder, "Molecular mechanisms of prion pathogenesis," *Annu. Rev. Pathol.*, vol. 3, no. 1, pp. 11–40, Feb. 2008.

[48] M. Weller and A. Aguzzi, "Movement disorders reveal Creutzfeldt–Jakob disease," *Nature Rev. Neurol.*, vol. 5, no. 4, pp. 185–186, Apr. 2009.

[49] A. Caricati-Neto, A. G. García, and L. B. Bergantin, "Pharmacological implications of the Ca^{2+}/cAMP signaling interaction: From risk for antihypertensive therapy to potential beneficial for neurological and psychiatric disorders," *Pharmacol. Res. Perspective*, vol. 3, no. 5, Oct. 2015, Art. no. e00181.

[50] G. M. Dallérac *et al.*, "Impaired long-term potentiation in the prefrontal cortex of Huntington's disease mouse models: Rescue by D_1 dopamine receptor activation," *Neurodegenerative Disease*, vol. 8, no. 4, pp. 230–239, 2011.

[51] G. Battaglia *et al.*, "Early defect of transforming growth factor β1 formation in Huntington's disease," *J. Cellular Mol. Med.*, vol. 15, no. 3, pp. 555–571, Mar. 2011.

[52] E. Buck *et al.*, "Comparison of Sirtuin 3 levels in ALS and Huntington's disease—differential effects in human tissue samples vs. transgenic mouse models," *Frontiers Mol. Neurosci.*, vol. 10, May 2017, Art. no. 156.

[53] M. A. Hickey and M.-F. Chesselet, "Apoptosis in Huntington's disease," *Prog. Neuro-Psychopharmacol. Biol. Psychiatry*, vol. 27, no. 2, pp. 255–265, Apr. 2003.

[54] M. Y. Davis, C. D. Keene, S. Jayadev, and T. Bird, "The co-occurrence of Alzheimer's disease and Huntington's Disease: A neuropathological study of 15 elderly Huntington's disease subjects," *J. Huntingtons Disease*, vol. 3, no. 2, pp. 209–217, Jan. 2014.

[55] M. S. Arrázola, C. Silva-Alvarez, and N. C. Inestrosa, "How the Wnt signaling pathway protects from neurodegeneration: The mitochondrial scenario," *Frontiers Cellular Neurosci.*, vol. 9, 2015, Art. no. 166.

[56] P. Dupont, M.-T. Besson, J. Devaux, and J.-C. Liévens, "Reducing canonical Wingless/Wnt signaling pathway confers protection against mutant Huntingtin toxicity in Drosophila," *Neurobiol. Disease*, vol. 47, no. 2, pp. 237–247, Aug. 2012.

[57] K. R. Bowles and L. Jones, "Kinase Signalling in Huntington's disease," *J. Huntingtons Disease*, vol. 3, no. 2, pp. 89–123, Jan. 2014.

[58] P. P. Michel, E. C. Hirsch, and S. Phane Hunot, "Understanding dopaminergic cell death pathways in Parkinson Disease," *Neuron*, vol. 90, no. 4, pp. 675–691, 2016.

[59] P. Anglade *et al.*, "Apoptosis and autophagy in nigral neurons of patients with Parkinson's disease," *Histol. Histopathol.*, vol. 12, no. 1, pp. 25–31, Jan. 1997.

[60] Y. Chu, H. Dodiya, P. Aebischer, C. W. Olanow, and J. H. Kordower, "Alterations in lysosomal and proteasomal markers in Parkinson's disease: Relationship to alpha-synuclein inclusions," *Neurobiol. Disease*, vol. 35, no. 3, pp. 385–398, Sep. 2009.

[61] T. Pan, S. Kondo, W. Le, and J. Jankovic, "The role of autophagy-lysosome pathway in neurodegeneration associated with Parkinson's disease," *Brain*, vol. 131, no. 8, pp. 1969–1978, Jan. 2008.

[62] S. Chandra, G. Gallardo, R. Fernández-Chacón, O. M. Schlüter, and T. C. Südhof, "α-Synuclein cooperates with CSPα in preventing neurodegeneration," *Cell*, vol. 123, no. 3, pp. 383–396, Nov. 2005.

[63] K. J. Vargas, S. Makani, T. Davis, C. H. Westphal, P. E. Castillo, and S. S. Chandra, "Synucleins regulate the kinetics of synaptic vesicle endocytosis," *J. Neurosci.*, vol. 34, no. 28, pp. 9364–9376, Jul. 2014.

[64] J. R. Mazzulli *et al.*, "Gaucher disease glucocerebrosidase and α-synuclein form a bidirectional pathogenic loop in synucleinopathies," *Cell*, vol. 146, no. 1, pp. 37–52, Jul. 2011.

[65] C. Rangel-Barajas, I. Coronel, and B. Florán, "Dopamine receptors and neurodegeneration," *Aging Disease*, vol. 6, no. 5, pp. 349–368, Sep. 2015.

[66] J. Massano and K. P. Bhatia, "Clinical approach to Parkinson's disease: Features, diagnosis, and principles of management," *Cold Spring Harbor Perspectives Med.*, vol. 2, no. 6, Jun. 2012, Art. no. a008870.

[67] K. E. Goodwill, C. Sabatier, C. Marks, R. Raag, P. F. Fitzpatrick, and R. C. Stevens, "Crystal structure of tyrosine hydroxylase at 2.3 Å and its implications for inherited neurodegenerative diseases," *Nature Struct. Biol.*, vol. 4, no. 7, pp. 578–585, Jul. 1997.

[68] Z. Q. Xu *et al.*, "Immunohistochemical studies on phosphorylation of tyrosine hydroxylase in central catecholamine neurons using site- and phosphorylation state-specific antibodies," *Neuroscience*, vol. 82, no. 3, pp. 727–738, Feb. 1998.

[69] L. Zhao *et al.*, "A deficiency of ceramide biosynthesis causes cerebellar purkinje cell neurodegeneration and lipofuscin accumulation," *PLoS Genetics*, vol. 7, no. 5, pp. 1–12, May 2011.

[70] E. Martin *et al.*, "Loss of function of Glucocerebrosidase GBA2 Is responsible for motor neuron defects in hereditary spastic paraplegia," *Amer. J. Human Genetics*, vol. 92, no. 2, pp. 238–244, 2013.