PONTIFICAL XAVIERIAN UNIVERSITY

Basic sciences faculty

Biology

GRADE PROJECT

REM sleep Behaviour Disorder module: A network medicine approach

Tain Mauricio Velasco Luquez

Directed by

Andrés Mauricio Pinzón PhD

Bogotá D.C.

2017

REM sleep Behaviour Disorder module: A network medicine approach

By

Tain Mauricio Velasco Luquez

A Thesis Submitted to the Pontifical Xavierian University

in Partial Fulfillment of the

requirements for the degree of

Biologist

Approved by

Andrés Mauricio Pinzón PhD, Director	Janet Gonzáles Santos PhD, co-Director
	Emiliano Barreto Hernández PhD, Jury

Pontifical Xavierian University

Bogotá D.C.

2017

Nota de advertencia

Artículo 23 de la Resolución N° 13 de Julio de 1946

"La Universidad no se hace responsable por los conceptos emitidos por sus alumnos en sus trabajos de tesis. Solo velará porque no se publique nada contrario al dogma y a la moral católica y por que las tesis no contengan ataques personales contra persona alguna, antes bien se vea en ellas el anhelo de buscar la verdad y la justicia"

Preface

Before you lies the ultimate product of a very enthralling and intellectually demanding five months of my life. Here, I use for the first time a network approach to unravel the molecular mechanism underneath the Rapid Eye Movement Sleep Behaviour Disorder aetiology, a very promising parasomnia regarding its biomarker potential for synucleinopathies, such as Parkinson's Disease.

Despite I undertook this project with little practical information and armed with no more than a puerile scientific enthusiasm and curiosity for the unknown, it has taught me not only the practical techniques of network science and all its paraphernalia, but also has helped me to forge a solid professional foundation in research.

It is required to reflect upon my mother, my grandmother and my father for their unconditional support and the willing ear. Mother, you are my life example and I deeply admire your intellectual capabilities and tender demeanour, inherited from my grandmother's restless and passionate discipline. Father, thank you for your creative ideas and life lessons, indispensable to inspire me and my work. I want to thank Andrés Pinzón, for his wise and opportune guide, and to Janet Gonzáles Santos for her benevolent and genuine advice. In addition, I would like to thank my laboratory colleagues, particularly Juan David Henao and Daniel Osorio, for deliberate over exciting and creative research ideas and methodological issues. I would particularly like to single out Dennis Castillo Figueroa, who, in spite of the scientific abyss that I opened in our friendship, he is always ready to skilfully draw an implacable critique, connecting, again and again, our scientific paths.

There being no further purposes in this writing, I hope you, dear reader, find my words entertaining.

Tain Velasco Luquez

Bogotá D.C., June 4, 2017

Abstract

The Rapid eye movement sleep Behaviour Disorder (RBD) is a promising biomarker for early diagnosis of synucleinopathies, such as Parkinson's Disease (PD), however, little is known about its molecular pathogenesis. The meagre attempts to investigate it use a fragmentary single-gene centred approach, disregarding key molecular entities, interactions and biological processes in the RBD aetiology. Under the systemic paradigm of network medicine, RBD arises from the disruption of a disease module in the network of a cell' physical interactions, or Human Interactome (HI). This research aimed to characterise, for the first time, the RBD-module in an up-to-date high-quality brain-specific HI, employing the DIAMOnD algorithm for disease module prediction. The HI and the RBD-module were validated topologically, utilising exponential random graph model fitting and comparisons against random expectation, and functionally, using a novel network enrichment analysis test, for the former, and the evolutionary framework of the duplication-acquisition model, for the latter. DIAMOnD expanded the molecular pathogenesis of RBD from 26 isolated molecular entities to over 400 interacting molecular entities. The RBD-module not only recovers the well-known role of immune and signalling processes in the sleep regulation, represented by overenrichment with biological process and cellular compartments related to cytokine and serotonin metabolism, but also supports an alternative RBD aetiological hypothesis including limbic structures and suggesting a wider pathogenic view than previously considered for RBD. This study shows the utility of the holistic approach of network medicine to improve our disease understanding and paves the way for further studies guided to disclose the molecular relationship between RBD and synucleinopathies, especially PD, strengthening its potential as their biomarker.

Keywords:

Network medicine, disease module, REM sleep behaviour disorder, Parkinson's Disease, diagnosis biomarker

Contents

List of Figures

List of Tables

Acronyms

CAGE Cap Analysis of Gene Expression. 15

DLB Dementia with Lewi Bodies. 10

GRN Gene Regulatory Network. 15

HPA Human Protein Atlas. 15

MSA Multiple System Atrophy. 10

NEAT Network Enrichment Analysis Test. 17, 18

PD Parkinson's disease. 9

PPI Protein-Protein physical Interactions. 14

RBD Rapid eye movement sleep Behaviour Disorder. 9, 14, *Glossary:* Rapid eye movement sleep Behaviour Disorder

REM Rapid Eye Movement. 10

TF Transcription Factor. 15

Glossary

Average Path Length ($< d >$) It is a measure of connectivity, calculated by averaging the minimum number of edges between two given nodes, *a.k.a.* geodesic distance, as follows:

$$< d >= 1/N(N-1) * \sum_{i,j=1,N;i\neq j} d_{i,j}$$

Where $d_{i,j}$ is the distance between the nodes i and j and N is the number of nodes in the network (Kolaczyk & Csárdi, 2014). 16

Biomarker "A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention." (Biomarkers Definitions Working Group., 2001). They are valuable tools to identify at-risk population of a certain disease, track the disease prognosis, aid in the process of disease's staging and to predict clinical responses. 10

Degree Given a network G with vertices $n \in N$ and edges $m \in M$, then the degree k_i denote the number of edges of the i^{th} node, which in turn enable the definition of the total number of edges in the network (M)

$$M = 1/2 \sum_{i=1}^{N} k_i$$

The average degree of the network $< k >$ is defined as

$$< k >= 1/E \sum_{i=1}^{N} k_i = 2M/E$$

(Kolaczyk & Csárdi, 2014; 'Network Science by Albert-László Barabási', n.d.). 16, 19

Diameter In the network context, the diameter is a topological feature measured as the longest geodesic distance between the farthest nodes in the network (Kolaczyk & Csárdi, 2014). 19

Diseasome A comprehensive network where the disease phenome, representing all genetic disorders, and the disease genome, representing all disease genes associated to these phenotypes, are linked into two main networks, as Goh et al. (2007) proposed it: 1. The Human Disease Network, whose nodes are diseases and edges are shared genes among them, and 2. The Disease Gene Network, whose nodes are genes and edges are drawn if two genes are implicated in the same disease. 12

Emergent property A system's feature only perceivable when the system is seen as an indissoluble complex whole (Moriello, 2013). Contrarily, while some properties emerge in the system, others submerge into it and are no longer discernible (Moriello, 2013). *Exampli gratia* Only when all cellular elements and its interactions at different organisational levels are perceived holistically, the emergent property of life arises, whereas individual properties of these elements or their interactions, such as quaternary structures of proteins or allosteric regulations between enzymes. 12

Exponential Random Graph Models (p^*) Statistichal model which employs a logistic regression for parameter estimation of the form:

$$P(Y_{ij}|Y_{i'j'},\theta) = logistic \sum_{h=1}^{k} \theta_h \delta_h^{ij}(Y)$$

Where Y_{ij} is a binary random variable indicating if there is an edge between a pair of vertices i, j, $Y_{i'j'}$ is a binary random variable for the other vertices and θ is the coefficient of a network statistic δ (Kolaczyk & Csárdi, 2014).. 7, 16, 21

Human Interactome "complete repertoire of genetic interactions potentially encoded by an organism's genome" (Sanchez et al., 1999). Therefore, it includes not only the proteome, protein-protein interaction network, but also the transcriptome, regulome, metabolome, transient and long-lasting interactions among all cellular entities, both intra and extracellular, such as proteins, DNA, RNA, lipids and carbohydrates. 9, 14

Hypergeometric distribution Distribution that models the number of successful cases in a random sample without replacement. In the network context such distribution is very useful when performing set enrichment analysis, where the successful cases are those links between the query set (*e.g.* the genes of interest) and the target set (*e.g.* set of functional terms). It is defined as:

$$N_{AB} \sim hypergeom(n = d_A, K = d_B, N = d_V)$$

Where d_A, d_B and d_V represents the degree of the query set, the degree of the functional terms set and the total degree between them, respectively (Signorelli, Vinciotti, & Wit, 2016). 18

Network Theoretical set of concepts for representing systems in the form of nodes and links among them (Moriello, 2013). Graph theory is inextricably intertwined with networks, as a graph is the underlying mathematical object of every network, enabling the application of formal quantitative analyses on them. 9

Pathophenotype Hallmark phenotype of a disease. 10

Prodrome Period preceding the main symptomatology of a given disease. *E.g.* PD, as a progressive pathology, exhibit non-motor symptoms, such as constipation, RBD, olfactory loss, inter alia, which precede the cardinal motor symptoms emergence for decades (Postuma & Berg, 2016). 10, 11

Rapid eye movement sleep Behaviour Disorder Parasomnia where the patient have vivid dreams during the REM sleep phase associated to a loss of atonia (Boeve, 2010). 9, 14

Synucleinopathy Neurodegenerative disease which exhibit an atypical aggregation of the α-Synuclein protein. Grouping Parkinson's disease, Dementia with Lewi Bodies and Multiple System Atrophy, inter alia. 9, 14

Transitivity *a.k.a.* Clustering Coefficient (cl) convey information regarding the frequency with which the triplets in the network form triangles, thus informing about the closeness among nodes (Kolaczyk & Csárdi, 2014).

$$cl(G) = 3\tau_{\triangle}(G)/\tau_3(G)$$

Where $\tau_{\triangle}(G)$ is the number of triangles in the graph G and τ_3 is the number of connected triplets of the form $\wedge$. 16

1 Introduction

With the advent of high-throughput technologies in molecular biology, a bulk of genomic information is being generated at an unprecedented rate (Stephens et al., 2015), empowering the devising of a completely contemporary, despite its antiqueness, approach to understanding biological systems, as such, systems (Kitano, 2002). Contrary to a set of fragmentary elements that interact in a simple and, to some extent, naive way, as is inherently encoded in the analysis as the *sine qua non* enquiry method, characteristic of the reductionism paradigm. A paradigm shift is, therefore, taking place (Chan & Loscalzo, 2012). Medicine is being penetrated by such shift, going from a reactive medicine focused on the disease, to a predictive, preventive, personalised and participatory medicine, focused on the health (Wang, Maron, & Loscalzo, 2015; Flores, Glusman, Brogaard, Price, & Hood, 2013; Silverman & Loscalzo, 2012).

As Thomas Rau stated, "to treat different, you have to think different" (Biological Medicine Network, 2015). In such brand new way of thinking, network as the pinnacle representation of systems, play a critical role, offering the mathematical formalism and a practical tool to understand phenotypes of complex systems, such as disease and health in human cells (Barabási, Gulbahce, & Loscalzo, 2011). The birth of network medicine was imminent, a brand new marriage between the systemic paradigm and medicine, towards a revolutionary health system. In network medicine, a disease arises from the disruption of a sub-network or module in the Human Interactome (HI), and where the pathogenic entities responsible for such disease are located (Goh et al., 2007; Feldman, Rzhetsky, & Vitkup, 2008).

Rapid eye movement sleep Behaviour Disorder (RBD) is a parasomnia where patients have, often violent, oneiric behaviours while in REM sleep, causing serious harm to themselves and their spouses (Boeve, 2010; Arnulf, 2012). Due to its high conversion rate to synucleinopathies (80 %) and its large lead period to neurodegeneration (3 - 34 years), RBD stands out as one of the most promising prodromal biomarkers of Parkinson's disease (PD), an incurable neurodegenerative synucleinopathy with a high social and economical burden worldwide (Kowal, Dall, Chakrabarti, Storm, & Jain, 2013; Hirsch, Jette, Frolkis, Steeves, & Pringsheim, 2016). Therefore, RBD not only presents itself as an early diagnostic tool but also offers an invaluable window to treat and study the progression of PD beforehand (Postuma, 2014). Notwithstanding the anatomical knowledge regarding the aetiology of RBD, pinpointing the pontomedullar region in the brainstem as the pathogenic source, little is known about the molecular mechanisms implicated in the RBD pathogenesis, with the meagre attempts employing a reductionistic single-gene centred approach (Gan-Or, Mirelman, et al., 2015; Gan-Or, Girard, et al., 2015; Gan-Or et al., 2016; Gan-Or et al., 2017; Fernández-Santiago et al., 2016), thus, hampering our RBD understanding and its applications as a biomarker.

This research aimed to obtain insight concerning the RBD aetiology, employing the power of the disease module hypothesis under the network medicine paradigm. The most up-to-date high-quality brain-specific HI was constructed, exhibiting evolutionary network hallmarks, upon which the identification of RBD-causing and non-direct RBD-causing molecular entities (*i.e.* genes, proteins, biological processes) were identified, supporting an alternative hypothesis for RBD aetiology with a strong limbic involvement, thereby, widening the perception of RBD pathogenesis. Identification of RBD module and its molecular mechanisms is the first step towards a holistic understanding of RBD and its relationships with other synucleinopathies, paving the way for further studies transgressing the merely reductionistic approaches and imposing network medicine.

2 Research problem statement

RBD is a multifactorial parasomnia characterised by an, usually violent, enacting of dreams during the Rapid Eye Movement (REM) sleep phase (Schenck, Bundlie, Ettinger, & Mahowald, 1986; Arnulf, 2012), generating injuries and sleep disruptions to the patients and their spouses (Olson, Boeve, & Silber, 2000). RBD is one of the most promising prodromal biomarkers of several synucleinopathies, including PD and Dementia with Lewi Bodies (DLB), as it has the highest specificity (Postuma, Lang, Gagnon, Pelletier, & Montplaisir, 2012), conversion rate (Iranzo et al., 2014; Postuma, Gagnon, Bertrand, Génier Marchand, & Montplaisir, 2015; Schenck, Boeve, & Mahowald, 2013) and diagnostic strength (Postuma & Berg, 2016) of the proposed biomarkers up to date. Additionally, RBD exhibit a median lead period to neurodegeneration of around 11 years (Postuma et al., 2009), ranging from 3 (Olson et al., 2000) up to 34 years (Claassen et al., 2010), period in which its progression has been linked with the progression of cognitive impairment in PD patients (Sixel-Döring, Zimmermann, Wegener, Mollenhauer, & Trenkwalder, 2016). Therefore, RBD not only provides an exceptional tool to identify at-risk population, which in turn enables early diagnose, stratification, testing on, and treatment of PD, DLB and Multiple System Atrophy (MSA) patients, but also as a potential clinical progression marker of these pathologies, offering and unprecedented window to study the evolution of such diseases (Postuma, 2014). Despite its relevance as prodromal biomarker, to the author's knowledge, little is known about the molecular aetiology governing RBD in humans, with the meagre attempts employing a reductionistic single-gene centred approach, (Gan-Or, Mirelman, et al., 2015; Gan-Or, Girard, et al., 2015; Gan-Or et al., 2016; Gan-Or et al., 2017; Fernández-Santiago et al., 2016; W. J. Zhang, Shang, Peng, Zhou, & Sun, 2017) thus, restricting our disease understanding and its applications.

Network is a pragmatical operable representation of complex systems (*e.g.* the cell), in the form of nodes connected by edges, enabling its study and better understanding at different organisational levels (*i.e.* molecular entity, metabolic pathway, network module, cellular compartment, inter alia) from a holistic standpoint (Moriello, 2013; Nurse & Hayles, 2011; Kitano, 2002). For instance, the HI is the network representing the cellular interactions in a given condition and cell type (Sanchez et al., 1999), whose nodes are molecular entities (*e.g.* DNA, RNA, protein, inter alia) and its edges represent physical or functional interactions among them (Vidal, Cusick, & Barabási, 2011). Under the network medicine paradigm, a disease arises from the perturbation of several HI components (Goh et al., 2007; Feldman et al., 2008) and thus can be mapped into a disease module, a tissue-specific and highly interconnected set of functionally related nodes whose perturbation generates the pathophenotype, in the HI (Barabási et al., 2011; Barabási, 2007; Kitsak et al., 2016). Disease module identification is the first step towards a holistic understanding of the pathology in question, as it moves our disease knowledge beyond the merely reductionistic disease-causing nodes, allowing the identification at different organisational levels of novel disease-causing and non-direct disease-causing entities and key interactions among them (Sharma et al., 2015; Ghiassian et al., 2016), thereby, enhancing our pathophenotype's understanding.

Inasmuch as RBD is a multifactorial disease (Arnulf, 2012), a network approach, well-fitted to cope with such complex interactions through the disease module concept, is suitable to unravel the RBD aetiology and to discover, from a holistic standpoint, novel molecular entities at different organisational levels related to its pathogenesis (Silverman & Loscalzo, 2012). Providing a holistic outlook of the structural and functional organisation governing the RBD disease module not only would shed

light on novel RBD-related molecular mechanisms, but also would pave the way for further studies guided to disclose its relevance as a prodromal biomarker for PD, DLB and MSA at the molecular level. See figure S1 for problem tree.

2.1 Research question

What are the molecular entities at different organisational levels composing the Rapid eye movement sleep Behaviour Disorder module?

3 Theoretical framework

3.1 Rapid eye movement sleep Behaviour Disorder

RBD is a multifactorial (Arnulf, 2012) parasomnia where patients enact their dreams due to a loss of atonia whilst in REM sleep phase (Boeve, 2010). Such dreams are significantly loaded with aggression and vigorous motor behaviours related to threatening (Fantini, Corona, Clerici, & Ferini-Strambi, 2005), consequently, serious harm can be inflicted to the patients and their spouses, even verging on lethality (Schenck, Lee, Bornemann, & Mahowald, 2009). Worldwide RBD population prevalence is unknown, but Ohayon and Schenck (2010) estimated it, through telephone questionnaires, to be 0.5 % in a representative sample from the United Kingdom, however, it is likely a sub-estimation loaded with false positives and false negatives (Arnulf, 2012). Regarding the local scope, so far, the only article referring to RBD in Colombia is a clinical report of a patient treated with trazodone (Chica-Urzola, 2015), making it evident that, besides the molecular aetiological aspects that this research attempts to enrich, more research is needed concerning the demographic and epidemiologic dimensions of RBD both at the global and regional scope.

Recently, forasmuch as RBD has a conversion rate to synucleinopathies of up to 80 % (Schenck et al., 2013), it has been under scrutiny for its potential as a prodromal biomarker of these pathologies, from which PD stands out as one of the most prominent due its elevated socio-economic burden (Kowal et al., 2013), its high prevalence (Pringsheim, Jette, Frolkis, & Steeves, 2014) and high incidence (Hirsch et al., 2016) worldwide. For instance, roughly half of PD patients suffer RBD (weighted prevalence of 42.3 %) (X. Zhang, Sun, Wang, Tang, & Xie, 2017) though prospective studies have reported a prevalence of up to 70 % (Neikrug et al., 2014), with a conversion rate from RBD to PD of up to 50 % (Schenck et al., 2013). Similarly, 73 % (Muntean, Sixel-Döring, & Trenkwalder, 2013) - 100 % (Vetrugno et al., 2004) of DLB patients suffer RBD and Boeve et al. reported that 92 % of RBD patients with symptoms of degenerative dementia suffer DLB (1998). These findings encourage the RBD research as a crucial step towards its use as a prodromal biomarker for the most important synucleinopathies, which in turn yields, as Postuma (2014) pointed out, a valuable opportunity for 1. Early therapeutic intervention, 2. Testing of potential biomarkers, 3. Anticipatory study of synucleinopathies' epidemiology and 4. Investigate the progression of synucleinopathies.

Since its first diagnosis by Schenck et al. (1986), little has been assessed about the RBD anatomical pathogenesis in humans, instead, the traditional cat and rat models, although contradictory regarding

some findings obeying species-specific mechanisms, have revealed that disruption of the sleep-wake regulatory circuit in the ponto-medullary axis, specifically, degeneration of the sublaterodorsal tegmental nucleus, might induce RBD (Luppi et al., 2011). Other nuclei in the brain stem have been implicated in the non-human animal models, such as the ventral gigantocellular reticular nucleus (Luppi et al., 2011) and the ventral mesencephalic reticular formation (Lai, Hsieh, Nguyen, Peever, & Siegel, 2008), nevertheless, certainty regarding the appositeness of these structures in humans is still to come (Boeve, 2010). Neuroimaging approaches employing single photon emission computed tomography, magnetic resonance imaging and positron emission tomography, not only have confirmed metabolic and neurostructural changes in areas previously implicated with RBD such as the hippocampus and pontomesencephalic tegmentum, but also have steered toward cortical and subcortical areas as novel regions potentially implicated in the RBD aetiology at the tissue organisational level (Boucetta et al., 2016; Holtbernd et al., 2014; Wu et al., 2014). Notwithstanding the efforts to decipher the complex anatomical circuit responsible for RBD in humans employing non-human animal models, the potential aetiological mechanisms befalling at a slender organisational level have been widely overlooked, therefore, and taking into account that translation from models to humans do not holds always (Burns, Li, Mehta, Awad, & Morgan, 2015), it is required a novel approach to dig deeper into the molecular basis of RBD in humans.

3.2 Network science and the systemic paradigm

Reductionism employs the analysis as the enquiry method, which fragment the studied system into its minimal components in order to, *caeteris paribus*, look for linear interactions among them, thus, decontextualising these components. Although its relevance during the XIX century, reductionism have proved to be insufficient to explain a plethora of dimensions of the cell as a biological system, and the different states of life as its emergent property, such as disease networks (Goh et al., 2007), neural networks (Watts & Strogatz, 1998), protein-protein interaction (PPI) networks (Vázquez, Flammini, Maritan, & Vespignani, 2003) and signalling networks (Ma'ayan et al., 2005), among others. Network constitutes the pinnacle operable representation of a system as nodes connected through edges, which, in conjunction with the advent of high-throughput technologies to generate biological data, reinforced systems biology as the *de facto* paradigm to approach biomedical research (Wang et al., 2015). Hence, aside from being insufficient to explain biological systems, the reductionism paradigm is inappropriate insofar as a system is more than the sum of its parts, demanding, thus, a synthetic enquiry to fully understand a complex system, such as the cell, and its different states, such as disease and healthy (Barabási & Oltvai, 2004; Kitano, 2002).

Network medicine is an attempt to explain diseases under the systemic paradigm, on the empirical premise that disease is a state of the biological system, arising from the disruption of complex interactions among several cellular components (Barabási et al., 2011). It is the case even for previously believed monogenic diseases, such as phenylketonuria and cystic fibrosis, which have been proved to be oligogenic instead (Badano & Katsanis, 2002). Network medicine had its iconic start with the seminal paper by Goh et al. (2007), in which all known genotype-phenotype associations to date were studied systematically in its inherent complexity, rather than the traditional fragmentary single gene-single disease approach, revealing that most human diseases shared a genetic origin illustrated in the human diseasome. The utility of such approach goes beyond the sole depiction of the common pathogenic basis and interconnectedness among several human diseases, it constitutes

an outstanding instrument to mathematically frame and explain complex diseases at different organisational levels (*i.e.* molecular entity, metabolic pathway, network module, cellular compartment, inter alia) (Silverman & Loscalzo, 2012), enhancing our disease understanding markedly (Loscalzo & Barabási, 2011).

The HI is the "complete repertoire of genetic interactions potentially encoded by an organism's genome" (Sanchez et al., 1999) of a given human cell type in a specific time, and it constitutes the baseline foundation behind network medicine analyses (Zanzoni, Soler-López, & Aloy, 2009). Derived from network medicine, the disease module hypothesis stands that disease-related (*i.e.* disease-causing and non-direct disease-causing) nodes are functionally organised in a densely connected sub-network in the HI, rather than randomly scattered throughout it, from whose perturbation arises the disease (Goh et al., 2007; Feldman et al., 2008) and which is tissue-specific (Kitsak et al., 2016). Disease module identification lies at the core of network medicine analyses and constitutes the first step towards a systemic pathophenotype's understanding, as it lays the foundation for holistic aetiological examination, through prediction of novel disease-related nodes and pathways (Oti, Snel, Huynen, & Brunner, 2006; Köhler, Bauer, Horn, & Robinson, 2008; Navlakha & Kingsford, 2010; Sharma et al., 2015), effective drug development, through enhanced pharmacological target selection (Pawson & Linding, 2008; Stern, Schurdak, Bahar, Berg, & Taylor, 2016; Hart & Xie, 2016), better inter-disease linkage, through comorbidity studies (Ko, Cho, Lee, & Kim, 2016), and improved diagnosis accuracy, through systemic nosology (Loscalzo, Kohane, & Barabási, 2007) and biomarkers identification (Potashkin, Santiago, Ravina, Watts, & Leontovich, 2012).

Invariably, the completeness of the HI will determine not only the integrity of the disease modules, as missing links will exclude disease-related nodes, but also their relationship at higher organisational levels (*e.g.* at pathways or module level) (Menche et al., 2015; Zanzoni et al., 2009). Unfortunately, due to historical technological constraints, both the HI and the majority of disease modules remain highly incomplete and biased towards the most studied nodes, and the inherited bias of the most employed techniques (*i.e.* yeast-2-hybrids assays, GWAS and co-expression) (Menche et al., 2015; Rolland et al., 2014). For instance, as Menche et al. (2015) showed, many disease modules include less than 20 % of the potential disease-related nodes. Despite its inherited issues, a carefully curated HI can be constructed and the corresponding disease module identified if the disease in question possesses enough disease-causing genes recognised (Menche et al., 2015).

3.3 Towards the future medicine in the present

Systemic approaches, and particularly network medicine, are fundamental players in the transition from a reactive medicine, focused on fighting diseases, to the so called P4 medicine, for predictive, preventive, personalised and participatory, whose focus is health preserving (Hood & Auffray, 2013). In such framework, network medicine lay the groundwork for expanding the personalised understanding of a given pathophenotype, as 1. it allows the identification of biomarkers, enabling early diagnose, stratification, intervention and prevention of disease; 2. the characterisation of unascertained disease-modifying mechanisms, which in turn shed light on novel aetiological and pathogenesis insights required to develop predictive models for disease presentation; 3. the better designing of therapeutic interventions by means of systemic exploration of potential drug-target interactions; and 4. for participatory and multidisciplinary patient-healthcare system interactions, going beyond

the doctor's office and penetrating other spheres such as the family and friend relationships of the patient (Flores et al., 2013; Hood & Auffray, 2013; Galas & Hood, 2009). These insights not only would have repercussions in the healthcare systems, which *per se* would have a big improvement, but also would reduce the temporal, social and economic burden of current medical organisations. Inducing, thus, profound changes in the perception of medicine (Flores et al., 2013). The 100k project is the most palpable example that we are living in a P4 medicine era (Hood & Price, 2014).

4 Objectives

Characterise the Rapid eye movement sleep Behaviour Disorder (RBD) disease module at the node, edge, biological process and cellular compartment organisational levels, in order to strengthen the potential use of RBD as a prodromal biomarker of synucleinopathies.

1. Construct the Human Interactome (HI)
2. Identify the putative RBD module

5 Methods

5.1 Human interactome assembly

5.1.1 Interactions data gathering

Considering the highly tissue and cell specific nature of the disease module (Kitsak et al., 2016) only tissue-specific and cell-line specific data relative to the brain and its cell types were incorporated into the HI. In order to avoid multi-mapping when converting between gene or protein IDs, only those nodes mapped to an Entrez Gene ID were used. As long as they were free, different sources of experimentally validated physical interactions were employed, ensuring the completeness and quality of the HI (Barabási et al., 2011). The contemplated sources are:

1. Binary Protein-Protein physical Interactions (PPI) represent pairwise physical interactions among proteins. High-quality PPI (*i.e.* More than two events of curation, reported in at least two separate experiments or publications and with 3D structures from PDB where more than two distinct proteins have been identified in a complex) were retrieved from the databases Intact, MINT, HPRD, DIP, BioPlex and bioGRID, employing the web server APID on 02/02/2017. Moreover, APID parse the data into the 2.5.4 version of the PSI-MI XML format adopted by the HUPO Proteomics Standard Initiative for the expedite data comparison, exchange and verification (Kerrien et al., 2007), thus easing its use. The proteome (*a.k.a.* map of PPI) generated has $\sim$ 42 % of coverage of known human PPIs (Alonso-López et al., 2016) and, by excluding binary singleton relationships (interactions supported by a single piece of experimental evidence), it is less likely to include false PPI, a product of curation errors with a recovery rate similar to the random expectation (Rolland et al., 2014). The detailed workflow of APID is available here.

A given protein was considered expressed in the brain if its healthy tissue reliability score from The Human Protein Atlas was "Supportive" or "Approved" and its expression level was detected (Uhlén et al., 2015) (a detailed description of the reliability and expression score is available here), thus ensuring the interactions' quality. The Human Protein Atlas (HPA) is a recent attempt to unravel the tissue-specific human proteome from a gene-centric approach, in which RNAseq and immunofluorescence labelling were employed to determine the tissue-specificity of a given transcript and in-house and commercial antibody labelling coupled with immunohistochemistry imaging were employed to determine its subcellular localisation at a single-cell level (Uhlén et al., 2015).

2. Gene Regulatory Network (GRN), where regulatory elements (*e.g.* promoters, enhancers, insulators, Transcription Factor (TF), RNA, inter alia) are linked through physical contacts, are key components in understanding diseases, as they are enriched with disease causing genetic variants (Lee & Young, 2013; Jimenez-Sanchez, Childs, & Valle, 2001) and play a critical role in cell homoeostasis (Ma'ayan et al., 2005). Recently, Marbach et al. (2016) created a tissue-specific and cell-line specific compendium of high-quality inferred regulatory networks including TF-enhancer, TF-promoter and enhancer-promoter relationships employing FANTOM5 project data. FANTOM5 use Cap Analysis of Gene Expression (CAGE), a high-throughput quantitative technology to map transcription starting sites and assess gene expression profiles simultaneously in a cell, tissue or condition specific way. Only adult data grouped into "Neurons & fetal brain", "Nervous system & adult hindbrain" and "Adult forebrain" clusters from (Marbach et al., 2016) was used to generate the regulatory network.

3. Metabolic and signalling interactions, where two nodes (*i.e.* metabolites, metabolic-genes, signalling-genes) are connected if they share the same pathway, was downloaded from KEGG on 03/05/2017. In order to consider the inextricably intertwine between metabolic and signalling networks, they were merged into a single brain-specific MetaboSignal network, following the guides in PPI section and using the R package by Rodriguez-Martinez et al. (2016). From now on this type of interactions will be referred as metabosignal, for brevity.

Table 1 lists a set of common topological attributes (please refer to Glossary section for definitions) assessed on the interactome and compared with previous human interactomes, proteomes and regulomes, to place in context the network here constructed. Refer to supplementary information S2 for the complete research workflow.

5.1.2 Validation

Poisson distribution is a signature of random processes governing a given network attribute (Solomonoff & Rapoport, 1951; Erdos & Rényi, 1959). Consequently, deviation from such distribution can be interpreted as a non-random process taking place to generate the network property (Kolaczyk & Csárdi, 2014). By setting such premise as the foundation, one can compare observed interactome' attributes to random ones, drawn from similar networks as the interactome, expecting a poor resemblance between them; supporting the idea that a relevant biological process is giving place to the observed feature instead of pure randomness. However, it is computationally expensive to simulate the massiveness of the HI, therefore, for the next two attributes, a stratified random sample with replacement of 6945 (roughly half of HI) vertices was taken from the HI, ensuring thus, the repres-

entation of all interaction sources and reducing the sampling error. For comparisons, Z-scores were computed for raw scores of 95 %. The compared attributes are:

1. The transitivity or Clustering Coefficient (cl) convey information regarding the frequency of triplets in the network forming triangles, thus inform about the closeness of nodes. Observed transitivity was compared against the average transitivity calculated from 1000 random graphs rewired with degree preserving randomization (further detail in the Module validation section).

2. The average path length ($< d >$)is a measure of connectivity, calculated by averaging the minimum number of edges between two given nodes, *a.k.a.* geodesic distance. Observed APL was compared against the average APL calculated from 1000 random graphs rewired with degree preserving (further detail in the Module validation section).

For further validate the HI, instead of comparing its properties against similar networks randomly drawn, a random model was fitted to the HI, to see whether such model is able to explain the observed degree distribution $P(k)$ and the probability of any tie occurring in the HI. Exponential Random Graph Models (ERGM) are statistical models intended to describe the probability of any tie in the network as a function of linear predictors, very much like a generalised linear model of regression (Kolaczyk & Csárdi, 2014). The random Bernoulli model (a variation of the Erdős-Rényi model) stands that, for each pair of vertices i, j, the probability to draw a link between them (Y_{ij}) is independent of the other nodes' probability ($Y_{i',j'}$), for any $i', j' \neq i, j$ (Kolaczyk & Csárdi, 2014). Therefore, the probability of any tie is calculated as $p = \frac{\exp(\theta)}{1+\exp(\theta)}$, for any θ value of an estimated network parameter (*e.g.* degree distribution) thought to be ruling the apparition of edges. Commonly, model fitting is used to found the parameters governing the network topology at a local level (*i.e.* edge level) (Goodreau, Kitts, & Morris, 2009), thought not common, model fitting can be used to validate the HI by looking for a poor fit between an observed property in the network at the global level, such as the $P(k)$, and the modelled one. For this, a goodness-of-fit test was carried out comparing the observed HI $P(k)$ against 10 000 simulations and the $log - odd$ to draw an edge in the HI was calculated.

Several real-world networks (Albert, Jeong, & Barabási, 1999, 2000; Goh et al., 2007; MacArthur, Sánchez-Garćia, & Ma'ayan, 2010; Faloutsos, Faloutsos, & Faloutsos, 1999), including previous proteomes (Rolland et al., 2014), exhibit the scale-freeness property, in which the $P(k)$ follows a power-law, thus inheriting key features, such as the Achilles-heel property (Albert et al., 2000) and the presence of hubs (highly connected nodes) (Albert et al., 1999)), which in turn offer valuable information regarding the interactome architecture and dynamics. Contrarily to tradition, the scale-freeness of the interactome was assessed following the quantitative guidelines for accurate heavy-tailed distributions fitting by Clauset, Shalizi, and Newman (2009). The maximum likelihood, a method well fitted for large datasets like the HI, estimators for the distribution parameters k_{min}, the minimum degree for which the power-law fits the $P(k)_{HI}$ (HI's degree distribution), and α, the scaling parameter of the distribution $P(k) \propto k^{-\alpha}$, were obtained by minimising the Kolmogorov-Smirnov distance (K-S) between the observed probability distribution and the best power-law model (Clauset et al., 2009). A goodness-of-fit test with a bootstrapping of 10 000 simulations was used to assess the uncertainty of the estimators (Clauset et al., 2009). Finally, the observed $P(k)_{HI}$ was compared against the discrete Poisson distribution, which represents random networks, and the exponential distribution, presumed to fit the $P(k)_{HI}$, employing the log-likelihood ratio ($\mathcal{R}$) and the Vuong's test for $p_{value}(\mathcal{R})$, both, two-sided and one-sided (Clauset et al., 2009).

5.2 Identification and validation of the RBD module in the HI

5.2.1 Seed nodes gathering

RBD nodes, to be used as seed for putative RBD-related nodes prediction, were obtained from the OMIM, Comparative Toxicogenomics Database (CTD), DisGeNET and MalaCards databases, employing the keywords from the Disease Ontology, a correlational database for accurate disease terms, "REM sleep behavior disorder", "rapid eye movement sleep behavior disorder", "REM sleep parasomnia" and "REM sleep behavior". These databases have at least one step of manual curation events, ensuring the interaction's quality. Furthermore, a literature search was performed to complement the database searches (Gan-Or, Mirelman, et al., 2015; Gan-Or et al., 2016; Gan-Or et al., 2017; Gan-Or, Girard, et al., 2015; Fernández-Santiago et al., 2016; W. J. Zhang et al., 2017).

5.2.2 Module identification

Broadly, three approaches to predict novel putative disease-modifying nodes can be distinguished: 1. Neighbourhood-based, 2. Graph partitioning and 3. Diffusion-based (Barabási et al., 2011; Navlakha & Kingsford, 2010). The neighbourhood approach departs from the observation that gene products related to the same disease are more likely to interact with each other (Goh et al., 2007), and therefore, new gene products can be associated with a disease if they lie in the same loci of the disease and also interacts with known disease-related gene products (Oti et al., 2006). Graph partitioning methods rely on the disease module hypothesis, thus linking gene products if they belong to the same disease or functional module (Feldman et al., 2008). Lastly, diffusion-based procedures assign a score to each node regarding its distance to the seed nodes, based on a random-walk, and the configuration of these nodes' edges (Köhler et al., 2008).

Irrespective of the method, most of community (a synonym for network module) detection algorithms employ the density of edges of each node to link putative nodes, however, inasmuch as disease nodes usually are not very densely connected (*i.e.* they are in the periphery of the network) (Feldman et al., 2008; Goh et al., 2007), such algorithms are not able to distinguish between the functional module (*i.e.* neighbourhood exerting a common function) and the disease module, which can overlap (Ghiassian, Menche, & Barabási, 2015; Sharma et al., 2015). Therefore, the DIAMOnD algorithm, a graph partitioning method, (Ghiassian et al., 2015) was used to predict the RBD module, as it is based on the connectivity significance of the seed nodes' interactions instead of pure density, hence, it prevents the inclusion of hub nodes merely by its large number of interactions, ensuring the pathobiological relevance of the nodes included. However, inasmuch as DIAMOnD ranks according to the connectivity significance all nodes in the network, it requires a biological criterion to decide how many predicted nodes should be included in the RBD module (the sum of the proto-module and the predicted nodes). For this, a Network Enrichment Analysis Test (NEAT) for Gene Ontology (GO) biological processes terms was carried out employing the Signorelli et al. approach (see next section for details).

5.2.3 Module validation

NEAT is a method to disclose the functional relevance of a set of nodes taking advantage of the topological position of each node in the network (Signorelli et al., 2016). Such approach allows the reliable identification of the biological processes in which the RBD seed nodes are involved. Unlike previous methods (Alexeyenko et al., 2012), NEAT creates the null distribution, from which the enrichment significance will be calculated, by modelling the observed number of links between the RBD nodes and a set of GO biological processes terms as an hypergeometric distribution. Improving thus, not only the computation time but also the statistical accuracy, as the null distribution is derived from the observed data set instead of forcing it to be normal (Signorelli et al., 2016), which is clearly not the case for the networks here consolidated (figure S6). To ensure the quality of the enrichment, only high quality experimental and computational GO evidence codes were included (*i.e.* EXP, IDA, IMP, IGI, IEP, ISS, ISA and ISO). In addition to functionally validate the RBD module for overenriched biological processes, the network enrichment analysis was also calculated every DIAMOnD iteration (*i.e.* every addition of a new ranked node), as a biological criterion to stop the addition of nodes to the RBD module (Ghiassian et al., 2015). In order to control the type I error due to multiple testing, the p_{value} of the enrichment was adjusted using the Benjamini and Hochberg (B-H) method, by reducing the False Discovery rate (Bouaziz, Jeanmougin, & Guedj, 2012).

Furthermore, following the same premise as in the HI validation section, the transitivity and the average path length of the putative RBD module were compared against the random expectation from 1000 degree-preserving randomisations of networks with the same number of nodes and degree per node as the RBD-module. These random graphs, as defined by Erdos and Rényi (1959), were computed as $G(n, p)$, where n is the number of nodes and p is the edge probability calculated as $p = \frac{d=1}{n}$. Comparisons were made employing Z-score for a raw score of 95 %. For further validate the biological relevance of the RBD-module, an ERGM was fitted to it, comparing the geodesic distance distribution and the degree distribution against 100 bootstrapping simulated networks (Kolaczyk & Csárdi, 2014). The probability to randomly draw an edge in the RBD-module and the GO enrichment for Cellular Compartment, following the parameters aforementioned, were carried out.

5.3 Software and hardware

The statistical programming language R 3.4.0 ('R: A Language and Environment for Statistical Computing', n.d.) was employed in most analyses, except for the implementation of DIAMOnD in Python 2. Following the guidelines for reproducible computational research (Stodden et al., 2016), scripts, workflows, raw and processed data is publicly available on GitHub, as well as the citation of all R packages used. Analyses were performed in the server of the Bioinformatics and Systems Biology Group (GIBBS) located at the Instituto de Genética of the Universidad Nacional de Colombia.

6 Results & discussion

6.1 Human interactome assembly

6.1.1 Interactions data gathering

The resulting brain-specific HI is the most comprehensive and up-to-date simple (*i.e.* no multi edges nor loops) undirected network, with 3 057 105 edges among 17 329 vertices (Figure S3). Table 1 place in context some topological properties of the HI (Interactome 1) compared against previous interactomes, proteomes and regulomes. The elevated mean degree ($< k >$ = 355.2; k_{max} = 14672) of the HI compared with the other networks is remarkable, and obeys to the colossal number of unprecedented interactions (20.8 times larger than the previous biggest one) here consolidated. Regulatory represent the vast majority of these interactions (97.4 %), letting metabosignal and PPI behind with 2.2 % and 0.4 % of interactions, respectively (figure S4 A). Such overrepresentation of regulatory interactions may obey to the combination of several causes: 1. The presence of very promiscuous TFs, promoters and enhancers that, through duplication events along the evolutionary history, have expanded the regulatory program exponentially (Pougach et al., 2014; Ferreira et al., 2013), thus reflecting its inherent complexity; 2. There are many more tissues used for regulatory interactions (33) than those used for either metabosignal or PPI (5), generating a bias toward regulatory interactions; 3. The level of curation of regulatory interactions, due to their inferred nature, is lower than metabosignal and PPI, which are manually curated and highly supported, respectively.

Table 1: Comparison of topological properties among recent HI, proteomes and a regulome. The large number of edges in the Interactome 1 and Interactome 2, a subset of the Interactome 1, here consolidated, which in turn lead to a high $< k >$ of the brain-specific HI and a bimodal $P(k)$, obeys to the massive contribution of promiscuous TF, enhancers and promoters to regulatory interactions. Topological network properties: M = Number of edges, N = Number of nodes, $< k >$ = Mean degree, $< cl >$ = Global transitivity (*a.k.a.* Clustering coefficient), $P(k)$ = Degree distribution, $< d >$ = Average path length and d_{max} = Diameter, NA = Not available or not apply

Citation	Network type	M	N	$< k >$	$< cl >$	$P(k)$	$< d >$	d_{max}
(Sharma et al., 2015)	HI	101032	11643	17.3	0.19	Scale-free	3.7	NA
(Menche et al., 2015)	HI	141296	13460	21	0.17	Scale-free	3.6	12
(Rolland et al., 2014)	Proteome	13944	4303	6.3	0.05	Scale-free	4.1	NA
(Mohammadi & Grama, 2016)	Proteome	147444	14658	NA	NA	NA	NA	NA
(Jolma et al., 2013)	Regulatory	3563	830	NA	NA	NA	NA	NA
(Kitsak et al., 2016)	HI	141296	13460	21	0.17	Scale-free	3.6	12
Interactome 1	HI brain-specific	3077707	17329	355.2	0.06	Bimodal	2.1	8
Interactome 2	HI brain-specific	1234593	16628	148.4957	0.04	Bimodal	2.3	14

If the first assertion were true, then, the regulatory program must exhibit the hallmarks of these duplication process. Indeed it does, through the $P(k)$. Unlike many biological and human-made networks, the HI is not scale-free, as even when the power-law is significantly valid (p_{value} = 0.32, 10 000 bootstrapping; figure S5) it fits only the 0.8 % (7664$< k <$7787) of the k range and have a particularly high scaling parameter α = 4.6 ($2 < \alpha < 3$ for typical scale-free networks) indicating more evenness in the degree than expected for a scale-free network, but also revealing a bimodal

distribution (figure S6 A). The HI inherits such distribution from regulatory interactions, as PPI and metabosignal are both unimodal (figure S6 C,D), hence supporting the first assertion.

Whereas the very pervasive scale-free networks evolve through a preferential-attachment, in which new nodes bind preferentially to hubs, (de Solla Price, 1965; Barabási & Albert, 1999) following the duplication-divergence model (Vázquez et al., 2003; Pastor-Satorras, Smith, & Solé, 2003), where new nodes arise from duplication and a subsequent loss of interactions lead to divergence of these paralogs, there is a type of non-scale free networks with bimodal degree distribution that evolve through the preferential attachment following the duplication-acquisition model (Ferreira et al., 2013), where new nodes arise from duplication events preferentially attaching hubs with high clustering coefficient instead of hubs with low clustering coefficient, thus, generating two type of hubs: intramodular and intermodular (Fraser, 2005; Li, Huang, Xia, & Sun, 2006). The former, also called "party hubs", tend to interact with most of their partners in a single spatio-temporal frame and belong to a functional module (Han et al., 2004; Li et al., 2006), evolve slower than intermodular hubs and are physically and functionally constrained (Fraser, 2005). Contrarily, intermodular hubs, *a.k.a.* "date hubs", interact with many partners in a variable spatio-temporal frame and tend to connect modules (Han et al., 2004), are more prone to duplication than party hubs, do exhibit pleiotropy (Li et al., 2006) and evolve faster (Fraser, 2005).

Clustering coefficient is successfully able to distinguish these two type of nodes in regulatory interactions (figure S7), indicating that the preferential duplication-acquisition model is plausible for the regulatory program. It is in strongly agreement with the empirical facts that gene-products with substrate promiscuity, just as the TFs in regulatory interactions which headed the top-10 hubs in the network (table S1), are more prone to duplication events (Conant & Wolfe, 2008) and roughly 90 % of Eukaryotic genes are believed to evolved by duplication events (Teichmann & Babu, 2004; Ferreira et al., 2013). Consequently, it is exciting and reasonable to imagine an evolutionary scenario where the first assertion takes place.

In order to reduce the influence of the second assertion and also to diminish the dimensionality of the HI, only whole-brain regulatory interactions were included in the HI (table 1, as Interactome 2) at the expense of the 32 area and cell-specific brain files (Marbach et al., 2016). The third assertion is somewhat inauspicious as, even though it is not manually curated as PPI or metabosignal, the validation of Marbach et al. has a high standard, only drawing interactions supported by the best predictive practices (Banf & Rhee, 2017). It is worth noting that, even after filtering with expression data, false positive and false negative interactions might be present in the three sources. Further investigations must take care of them, either by adding an additional filter or by including another type of regulatory data.

6.1.2 Validation

Inasmuch as it is significantly unlikely that a random process originates the observed $P(k)_{HI}$ when compared against a power-law distribution ($\mathcal{R}$ = 5.165587, $p_{value}(\mathcal{R})$ = 2.396 854 $\times$ 10^{-7}; figure S5) and an exponential distribution ($\mathcal{R}$ = -14.08696, $p_{value}(\mathcal{R})$ = 4.568 209 $\times$ 10^{-45}; figure S5), it is valid to think that a biological process is taking place to generate the observed degree distribution. Moreover, the observed transitivity is significantly higher than the random expectation (Z-score = 116.793, 95% confidence; figure 1 A) and the average path length is significantly smaller than the random expect-

ation (Z-score = -2.224, 95% confidence; 1 B). *i.e.* , A random network, with the same number of nodes and edges as the HI, would not have neither the clustering level nor the short paths of the HI. For further validation of the HI, the Bernoulli Exponential Random Graph Models (ERGM) was fitted to it, founding a very low probability ($log - odd$ = 0.0089) to randomly draw any link of the HI and an overall poor fitting between the observed geodesic distance and the predicted by the Bernoulli model (figure S8 C).

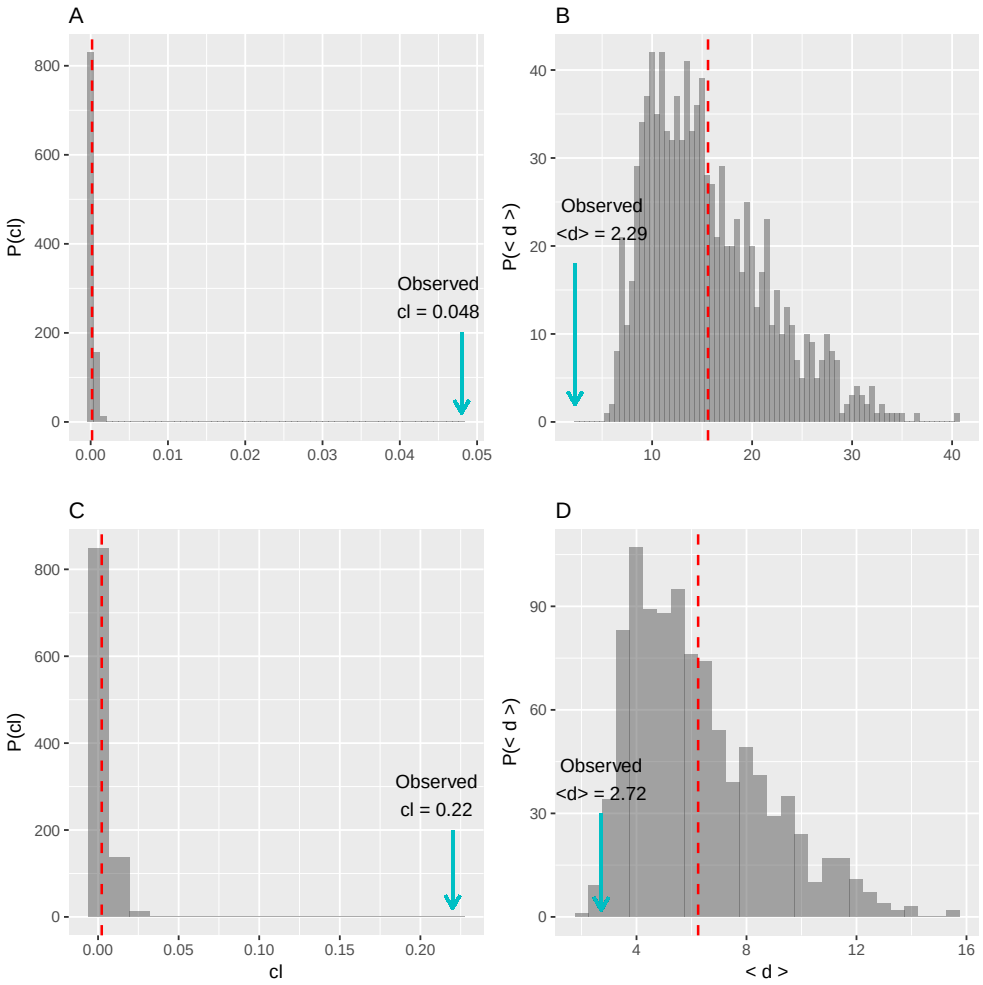

Figure 1: Transitivity (A and C) and average path length (B and D) comparison between the observed and the random expectation from 1 000 degree preserving randomisation for the RBD-module (bottom) and a sample from the HI (upper). In $G(n, d)$,d= 0.0024 for RBD-module and d = 0.0001 for the HI (please refer to the RBD-module validation for further detail). The Z-score for A, B, C and D are 116.793, -2.224, 14.78 and -2.50 respectively. In brief, the observed HI and RBD clustering coefficient and average path length are significantly greater and smaller, respectively, than one would expect for a random network with similar features, thus confirming the biological validity of the HI and RBD-module. Red dotted line is the mean. Note the close resemblance between the HI cl and <d> values and the ones from the stratified random sample, endorsing its representativeness.

6.2 Identification and validation of the RBD-module in the HI

After the exhaustive database and literature search for the RBD-related genes to be used as seed nodes, a list of 26 genes were consolidated (Table S2).

According to the disease module hypothesis, the seed nodes should form a distinguishable module in the HI (a proto-module) (Feldman et al., 2008; Barabási et al., 2011) however, RBD-seed nodes do not, as there are no edges in the HI linking them (figure 2 A). This might obey to 1. The small number of seed nodes present in the HI (24/26) reflecting the little molecular knowledge regarding RBD, in which case, either there are truly no biological interactions among them, or the science simply have not linked them yet, or 2. The inherent incompleteness of the HI, in which case is required a more comprehensive and carefully constructed novel HI. Notwithstanding the strongly influence of the HI's completeness in the detection of the RBD module, it is more plausible that the number of seed nodes used were not enough to distinguish the RBD proto-module, as it was demonstrated that, even in an incomplete interactome, a thoroughly consolidated set of seed nodes is able to make the proto-module pop up (Menche et al., 2015). This is further supported when including the 1st order neighbours of the RBD-seed nodes, where the RBD proto-module not only is observable but connect all seed nodes present in the HI (figure 2 B). Furthermore, considering that most diseases exhibit less than 20 % of their nodes in the observable proto-module (Menche et al., 2015) it is not surprising that the RBD proto-module is only observable when including the 1st neighbours.

After running the iterative network enrichment (figure S9), the first 400 most significantly connected DIAMOnD nodes were included in the high-quality RBD-module, ever assembled (figure 2 C). As other disease modules (Ghiassian et al., 2015), it is significantly more clustered than expected by chance (cl = 0.22, Z-score = 38.47; figure 1 C), even more than the HI (table 1), and its average paths length is significantly smaller than the random expectation ($< d >$ = 2.72, Z-score = -2.50; figure 1 D). Moreover, as the geodesic distance distribution (figure S8 A) and the degree distribution (figure S8 B) shows, the RDB-module represents a biologically plausible representation of the molecular landscape governing the disease, further supported by the low probability ($log - odd$ = 0.069) to randomly draw any edge of the module.

The majority of the ten most overenriched biological process in the RBD-module are immune-related (table S3). This is surprising, as it is in strong agreement with an alternative and barely explored hypothesis for RBD pathogenesis (Iranzo et al., 2005), in which is the limbic system, instead of the traditional pontomedullar region, the area in which the aetiological causation of RBD lies. It derives from the, previously believed coincidental, observations that autoimmune limbic encephalitis, without pontomedullar lesions, concomitantly exhibit RBD (Vale, Fernandes do Prado, do Prado, Povoas Barsottini, & Pedroso, 2016; Lin, Liu, & Hsu, 2009; Compta, Iranzo, Santamaŕia, Casamitjana, & Graus, 2007; Limousin et al., 2009; Iranzo et al., 2005; Adams, McKeon, Silber, & Kumar, 2011). This is further supported by several immune response-related overenriched cellular components in the RBD-module (table S3) and the widely accepted somnogenic effect of proinflammatory cytokines (Venancio & Suchecki, 2015). Although the extensive anatomical crosstalk between the limbic system and the pontomedullary system was already known, it had not been contemplated the possibility for the RBD aetiology to lie in a different anatomical region than the traditional one. Thus, these results reinforce the exiguously supported hypothesis for a strong limbic influence in the RBD pathogenesis, illustrating the capacity of network approaches to depicting a different disease landscape; further supporting the usefulness of a holistic approach for aetiological analyses.

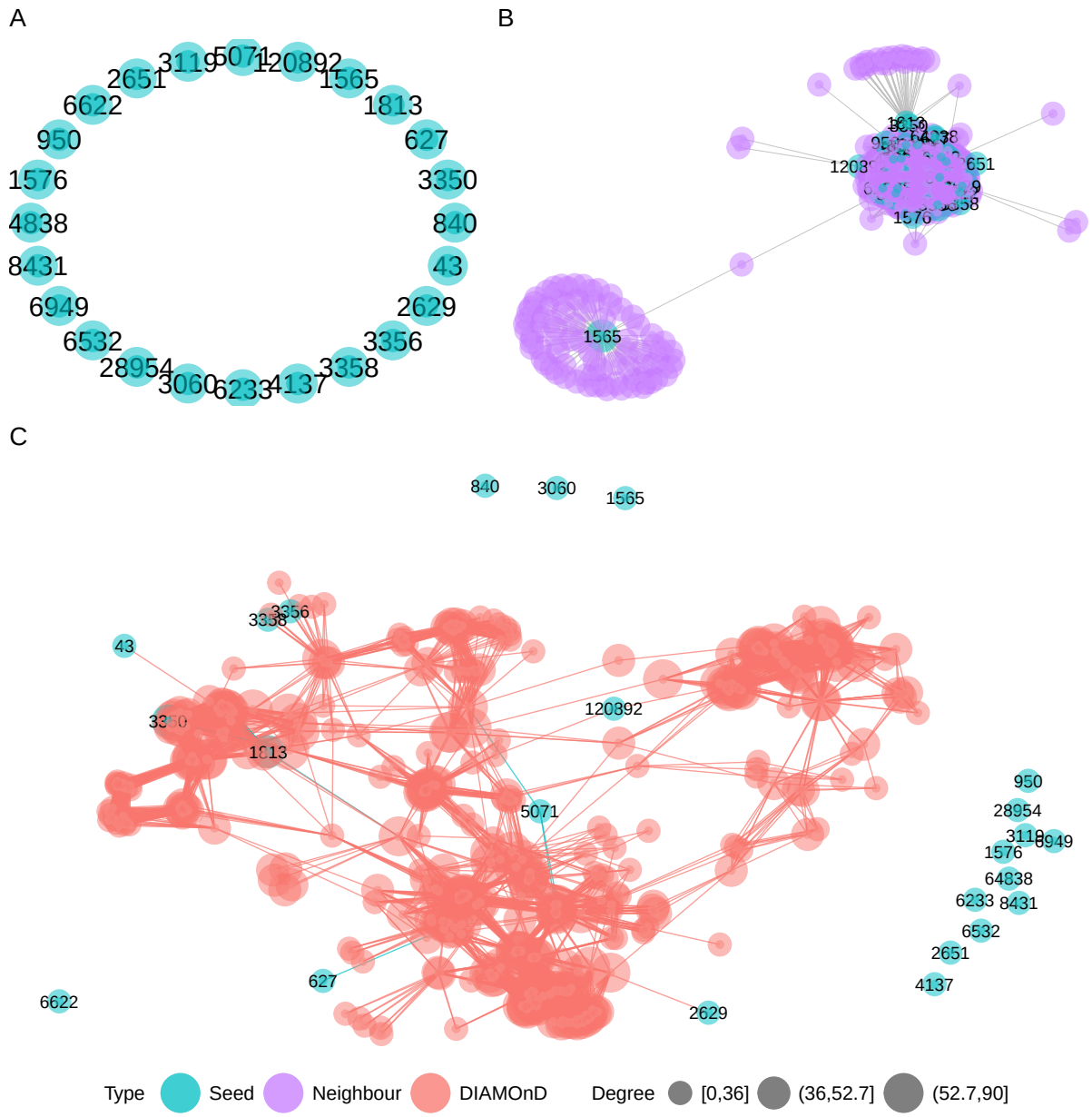

Figure 2: Fractionated proto-module containing the RBD-seed nodes (A). Proto-module plus 1st order neighbours (B) in which all seed nodes get connected, indicating that even though the exhaustive list of seed nodes here consolidated is not enough to make the module pop up, only adding the partners of these seed nodes, which I believe soon will be covered by science, is enough to show the proto-module. The first ever constructed RBD-module (C), showing biological entities and interactions previously unlinked to the aetiology of RBD (red nodes). As shown in figure S8 B, there are some seed nodes that remain disconnected from the module. For the sake of aesthetics, the first 58 hubs were removed from C. Labelled nodes are Entrez gene ID.

The RBD-module recovers well-established characteristics of RBD, such as the strong influence of serotonin signalling, illustrated in the fact that the 5th and 7th best-ranked DIAMOnD nodes correspond to its receptors HTR1D and HTR1F, respectively. HTR1, common in the REM-switch and hippocampus regions, are G_i-coupled receptors which, by repressing the adenylate cyclase, inhibit the cAMP meta-

bolism, which in turn down-regulate the cAMP-response element binding protein (CREB) (Berumen, Rodŕiguez, Miledi, & García-Alcocer, 2012), a well-known TF responsible for sustained arousal in areas involved in the REM-switch, such as the locus coeruleus (Graves, 2003). This may explain the overenriched biological processes, cellular compartments and best-ranked DIAMOnD nodes (Neuropeptide Y, HCAR-1, Somatostatin-28 and -14, Oxytocin, GHSR and endothelin-2) involved in cAMP signalling and G-coupled signal transduction (table S3). HTR1 also modulate the conductance of K^+ channels in the dorsal raphe nucleus, a member of the REM-switch, and the entorhinal cortex in the limbic system (Deng, Poudel, Rojanathammanee, Porter, & Lei, 2007), which is in concordance with the concomitant presence of RBD and K^+ channel antibody–associated limbic encephalitis (Iranzo et al., 2005).

7 Conclusion

Currently, the disease module of RBD remains unascertained and little is known about the molecular entities, at different organisational levels, related to its aetiology. This research has successfully assembled the most up-to-date high-quality brain-specific HI and characterise, for the first time in science, the RBD-module, shedding light on the molecular mechanisms implicated in a barely supported alternative hypothesis for RBD pathogenesis. Both the HI and the RBD-module recover interesting biological features previously reported in other biological networks, thus validating the innovative methodology employed, but also, the former exhibits the hallmarks of a recently proposed evolutionary scenario and the latter suggests a broader and more comprehensive molecular picture of RBD than previously considered. Moreover, this research not only has settled a solid foundation for further studies guided to evaluate the potential of RBD as biomarker of synucleinopathies, specially PD, but also has supported a systemic approach for aetiological analyses under the network medicine paradigm.

The present investigation has limitations though, mainly regarding to the completeness of the HI, which in turn determines further conclusions in the disease module, as discussed in the RBD-module section. To alleviate these limitations, is strongly advised to employ an unified framework to achieve the tissue-specificity, one that takes advantage of non-hard cut-offs and the network topology, such as the one proposed by Mohammadi and Grama (2016) including the frame discussed in Venkatesan et al. (2009) and Rolland et al. (2014), thus, avoiding biases towards certain type of interaction. The research would also benefit from a comparison of the DIAMOnD results against other prediction methodologies, such as the diffusion-based method previously discussed, in order to further validate the RBD-module. A further validation of the results shown here in the wet laboratory is required to fully ensure its validity.Lastly, It is worth noting that this research is fully reproducible following the freely available code mentioned early.

8 Funding

No funding source supported the present research.

References

Adams, C., McKeon, A., Silber, M. H., & Kumar, R. (2011). Narcolepsy, REM sleep behavior disorder, and supranuclear gaze palsy associated with Ma1 and Ma2 antibodies and tonsillar carcinoma. *Archives of neurology*.

Albert, R., Jeong, H., & Barabási, A.-L. (1999). The diameter of the world wide web. *Nature*.

Albert, R., Jeong, H., & Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*.

Alexeyenko, A., Lee, W., Pernemalm, M., Guegan, J., Dessen, P., Lazar, V., … Pawitan, Y. (2012). Network enrichment analysis: extension of gene-set enrichment analysis to gene networks. *BMC bioinformatics*.

Alonso-López, D., Gutiérrez, M. A., Lopes, K. P., Prieto, C., Santamaŕia, R., & De Las Rivas, J. (2016). APID interactomes: providing proteome-based interactomes with controlled quality for multiple species and derived networks. *Nucleic Acids Research*.

Arnulf, I. (2012). REM sleep behavior disorder: Motor manifestations and pathophysiology. *Movement disorders : official journal of the Movement Disorder Society*.

Badano, J. L. & Katsanis, N. (2002). Beyond Mendel: an evolving view of human genetic disease transmission. *Nature Reviews Genetics*.

Banf, M. & Rhee, S. Y. (2017). Computational inference of gene regulatory networks: Approaches, limitations and opportunities. *BBA - Gene Regulatory Mechanisms*.

Barabási, A.-L. (2007). Network Medicine — From Obesity to the "Diseasome". *New England Journal of Medicine*.

Barabási, A.-L. & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*.

Barabási, A.-L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*.

Barabási, A.-L. & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society Series B (…*.

Berumen, L. C., Rodŕiguez, A., Miledi, R., & Garćia-Alcocer, G. (2012). Serotonin Receptors in Hippocampus. *The Scientific World Journal*.

Biological Medicine Network. (2015). Rethinking Medicine: The Biological Medicine Network.

Biomarkers Definitions Working Group. (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.*

Boeve, B. F. (2010). REM sleep behavior disorder. *Annals of the New York Academy of Sciences*.

Boeve, B. F., Silber, M. H., Ferman, T. J., Kokmen, E., Smith, G. E., Ivnik, R. J., … Petersen, R. C. (1998). REM sleep behavior disorder and degenerative dementia: an association likely reflecting Lewy body disease. *Neurology*.

Bouaziz, M., Jeanmougin, M., & Guedj, M. (2012). Multiple Testing in Large-Scale Genetic Studies. In *Data production and analysis in population genomics*.

Boucetta, S., Salimi, A., Dadar, M., Jones, B. E., Collins, D. L., & Dang-Vu, T. T. (2016). Structural Brain Alterations Associated with Rapid Eye Movement Sleep Behavior Disorder in Parkinson's Disease. *Scientific Reports*.

Burns, T. C., Li, M. D., Mehta, S., Awad, A. J., & Morgan, A. A. (2015). Mouse models rarely mimic the transcriptome of human neurodegenerative diseases: A systematic bioinformatics-based critique of preclinical models. *European journal of …*.

Chan, S. Y. & Loscalzo, J. (2012). The emerging paradigm of network medicine in the study of human disease. *Circulation Research*.

Chica-Urzola, H. L. (2015). Trazodona en el trastorno del comportamiento del sueño MOR. *Revista Colombiana de Psiquiatría*.

Claassen, D. O., Josephs, K. A., Ahlskog, J. E., Silber, M. H., Tippmann-Peikert, M., & Boeve, B. F. (2010). REM sleep behavior disorder preceding other aspects of synucleinopathies by up to half a century. *Neurology*.

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-Law Distributions in Empirical Data. *SIAM Review*.

Compta, Y., Iranzo, A., Santamaría, J., Casamitjana, R., & Graus, F. (2007). REM sleep behavior disorder and narcoleptic features in anti-Ma2-associated encephalitis. *SLEEP*.

Conant, G. C. & Wolfe, K. H. (2008). Turning a hobby into a job: How duplicated genes find new functions. *Nature Reviews Genetics*.

de Solla Price, D. J. (1965). Networks of Scientific Papers. *Science*.

Deng, P. Y., Poudel, S. K. S., Rojanathammanee, L., Porter, J. E., & Lei, S. (2007). Serotonin Inhibits Neuronal Excitability by Activating Two-Pore Domain K+ Channels in the Entorhinal Cortex. *Molecular Pharmacology*.

Erdos, P. & Rényi, A. (1959). On random graphs I. *Publ. Math. Debrecen*.

Faloutsos, M., Faloutsos, P., & Faloutsos, C. (1999). On power-law relationships of the internet topology. *ACM SIGCOMM computer communication review*.

Fantini, M. L., Corona, A., Clerici, S., & Ferini-Strambi, L. (2005). Aggressive dream content without daytime aggressiveness in REM sleep behavior disorder. *Neurology*.

Feldman, I., Rzhetsky, A., & Vitkup, D. (2008). Network properties of genes harboring inherited disease mutations. *Proceedings of the National Academy of Sciences of the United States of America*.

Fernández-Santiago, R., Iranzo, A., Gaig, C., Serradell, M., Fernández, M., Tolosa, E., … Ezquerra, M. (2016). Absence of LRRK2 mutations in a cohort of patients with idiopathic REM sleep behavior disorder. *Neurology*.

Ferreira, R. M., Rybarczyk-Filho, J. L., Dalmolin, R. J. S., Castro, M. A. A., Moreira, J. C. F., Brunnet, L. G., & de Almeida, R. M. C. (2013). Preferential Duplication of Intermodular Hub Genes: An Evolutionary Signature in Eukaryotes Genome Networks. *PLoS ONE*.

Flores, M., Glusman, G., Brogaard, K., Price, N. D., & Hood, L. (2013). P4 medicine: how systems medicine will transform the healthcare sector and society. *Per Med*.

Fraser, H. B. (2005). Modularity and evolutionary constraint on proteins. *Nature Genetics*.

Galas, D. J. & Hood, L. (2009). Systems Biology and Emerging Technologies Will Catalyze the Transition from Reactive Medicine to Predictive, Personalized, Preventive and Participatory (P4) Medicine. *Interdisciplinary Bio Central*.

Gan-Or, Z., Girard, S. L., Noreau, A., Leblond, C. S., Gagnon, J.-F., Arnulf, I., … Rouleau, G. A. (2015). Parkinson's Disease Genetic Loci in Rapid Eye Movement Sleep Behavior Disorder. *Journal of Molecular Neuroscience*.

Gan-Or, Z., Mirelman, A., Postuma, R. B., Arnulf, I., Bar-Shira, A., Dauvilliers, Y., … Rouleau, G. A. (2015). GBA mutations are associated with Rapid Eye Movement Sleep Behavior Disorder. *Annals of clinical and translational neurology*.

Gan-Or, Z., Mohsin, N., Girard, S. L., Montplaisir, J. Y., Ambalavanan, A., Strong, S., … Rouleau, G. A. (2016). The role of the melanoma gene MC1R in Parkinson disease and REM sleep behavior disorder. *Neurobiol. Aging*.

Gan-Or, Z., Montplaisir, J. Y., Ross, J. P., Poirier, J., Warby, S. C., Arnulf, I., … Rouleau, G. A. (2017). The dementia-associated APOE ε4 allele is not associated with rapid eye movement sleep behavior disorder. *Neurobiol. Aging*.

Ghiassian, S. D., Menche, J., & Barabási, A.-L. (2015). A DIseAse MOdule Detection (DIAMOnD) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome. *PLoS Comput Biol*.

Ghiassian, S. D., Menche, J., Chasman, D. I., Giulianini, F., Wang, R., Ricchiuto, P., … Loscalzo, J. (2016). Endophenotype Network Models: Common Core of Complex Diseases. *Scientific Reports*.

Goh, K.-I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., & Barabási, A.-L. (2007). The human disease network. *Proceedings of the National Academy of Sciences*.

Goodreau, S. M., Kitts, J. A., & Morris, M. (2009). Birds of a Feather, Or Friend of a Friend?: Using Exponential Random Graph Models to Investigate Adolescent Social Networks. *Demography*.

Graves, L. A. (2003). Genetic Evidence for a Role of CREB in Sustained Cortical Arousal. *Journal of Neurophysiology*.

Han, J.-D. J., Bertin, N., Hao, T., Goldberg, D. S., Berriz, G. F., Zhang, L. V., … Vidal, M. (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*.

Hart, T. & Xie, L. (2016). Providing data science support for systems pharmacology and its implications to drug discovery. *Expert Opin Drug Discov*.

Hirsch, L., Jette, N., Frolkis, A., Steeves, T., & Pringsheim, T. (2016). The Incidence of Parkinson's Disease: A Systematic Review and Meta-Analysis. *Neuroepidemiology*.

Holtbernd, F., Gagnon, J.-F., Postuma, R. B., Ma, Y., Tang, C. C., Feigin, A., … Montplaisir, J. (2014). Abnormal metabolic network activity in REM sleep behavior disorder. *Neurology*.

Hood, L. & Auffray, C. (2013). Participatory medicine: a driving force for revolutionizing healthcare. *Genome Medicine*.

Hood, L. & Price, N. D. (2014). Promoting Wellness & Demystifying Disease:The 100K Project. *Clinical OMICs*.

Iranzo, A., Fernández-Arcos, A., Tolosa, E., Serradell, M., Molinuevo, J. L., Valldeoriola, F., … Santamaŕıa, J. (2014). Neurodegenerative disorder risk in idiopathic REM sleep behavior disorder: study in 174 patients. *PLoS …*.

Iranzo, A., Graus, F., Clover, L., Morera, J., Bruna, J., Vilar, C., … Santamaŕıa, J. (2005). Rapid eye movement sleep behavior disorder and potassium channel antibody-associated limbic encephalitis. *Annals of Neurology*.

Jimenez-Sanchez, G., Childs, B., & Valle, D. (2001). Human disease genes. *Nature*.

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K. R., Rastas, P., … Taipale, J. (2013). DNA-binding specificities of human transcription factors. *Cell*.

Kerrien, S., Orchard, S., Montecchi-Palazzi, L., Aranda, B., Quinn, A. F., Vinod, N., … Hermjakob, H. (2007). Broadening the horizon – level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biology*.

Kitano, H. (2002). Systems Biology: A Brief Overview. *Science*.

Kitsak, M., Sharma, A., Menche, J., Guney, E., Ghiassian, S. D., Loscalzo, J., & Barabási, A.-L. (2016). Tissue Specificity of Human Disease Module. *Scientific Reports*.

Ko, Y., Cho, M., Lee, J.-S., & Kim, J. (2016). Identification of disease comorbidity through hidden molecular mechanisms. *Scientific Reports*.

Köhler, S., Bauer, S., Horn, D., & Robinson, P. N. (2008). Walking the Interactome for Prioritization of Candidate Disease Genes. *The American Journal of Human …*.

Kolaczyk, E. D. & Csárdi, G. (2014). *Statistical Analysis of Network Data with R*. Springer New York.

Kowal, S. L., Dall, T. M., Chakrabarti, R., Storm, M. V., & Jain, A. (2013). The current and projected economic burden of Parkinson's disease in the United States. *Movement disorders : official journal of the Movement Disorder Society*.

Lai, Y. Y., Hsieh, K.-C., Nguyen, D., Peever, J., & Siegel, J. M. (2008). Neurotoxic lesions at the ventral mesopontine junction change sleep time and muscle activity during sleep: an animal model of motor disorders in sleep. *Neuroscience*.

Lee, T. I. & Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. *Cell*.

Li, L., Huang, Y., Xia, X., & Sun, Z. (2006). Preferential Duplication in the Sparse Part of Yeast Protein Interaction Network. *Molecular Biology and Evolution*.

Limousin, N., Dehais, C., Gout, O., Héran, F., Oudiette, D., & Arnulf, I. (2009). A brainstem inflammatory lesion causing REM sleep behavior disorder and sleepwalking (parasomnia overlap disorder). *Sleep medicine*.

Lin, F.-C., Liu, C.-K., & Hsu, C.-Y. (2009). Rapid-eye-movement sleep behavior disorder secondary to acute aseptic limbic encephalitis. *J. Neurol.*

Loscalzo, J. & Barabási, A.-L. (2011). Systems biology and the future of medicine. *Wiley interdisciplinary reviews. Systems biology and medicine*.

Loscalzo, J., Kohane, I., & Barabási, A.-L. (2007). Human disease classification in the postgenomic era: a complex systems approach to human pathobiology. *Molecular systems biology*.

Luppi, P.-H., Clément, O., Sapin, E., Gervasoni, D., Peyron, C., Léger, L., … Fort, P. (2011). The neuronal network responsible for paradoxical sleep and its dysfunctions causing narcolepsy and rapid eye movement (REM) behavior disorder. *Sleep Medicine Reviews*.

Ma'ayan, A., Jenkins, S. L., Neves, S., Hasseldine, A., Grace, E., Dubin-Thaler, B., … Iyengar, R. (2005). Formation of regulatory patterns during signal propagation in a Mammalian cellular network. *Science*.

MacArthur, B. D., Sánchez-Garćıa, R. J., & Ma'ayan, A. (2010). Microdynamics and Criticality of Adaptive Regulatory Networks. *Physical review letters*.

Marbach, D., Lamparter, D., Quon, G., Kellis, M., Kutalik, Z. a. n., & Bergmann, S. (2016). Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nature Methods*.

Menche, J., Sharma, A., Kitsak, M., Ghiassian, S. D., Vidal, M., Loscalzo, J., & Barabási, A.-L. (2015). Uncovering disease-disease relationships through the incomplete interactome. *Science*.

Mohammadi, S. & Grama, A. (2016). A convex optimization approach for identification of human tissue-specific interactomes. *Bioinformatics (Oxford, England)*.

Moriello, S. A. (2013). *Ciencias de la Complejidad*. Editorial Nueva Librerı́a.

Muntean, M.-L., Sixel-Döring, F., & Trenkwalder, C. (2013). No Difference in Sleep and RBD between Different Types of Patients with Multiple System Atrophy: A Pilot Video-Polysomnographical Study. *Sleep Disorders*.

Navlakha, S. & Kingsford, C. (2010). The power of protein interaction networks for associating genes with diseases. *Bioinformatics (Oxford, England)*.

Neikrug, A. B., Avanzino, J. A., Liu, L., Maglione, J. E., Natarajan, L., Corey-Bloom, J., … Ancoli-Israel, S. (2014). Parkinson's disease and REM sleep behavior disorder result in increased non-motor symptoms. *Sleep medicine*.

Nurse, P. & Hayles, J. (2011). The Cell in an Era of Systems Biology. *Cell*.

Ohayon, M. M. & Schenck, C. H. (2010). Violent behavior during sleep: prevalence, comorbidity and consequences. *Sleep medicine*.

Olson, E. J., Boeve, B. F., & Silber, M. H. (2000). Rapid eye movement sleep behaviour disorder: demographic, clinical and laboratory findings in 93 cases. *Brain*.

Network Science by Albert-László Barabási. (n.d.).

R: A Language and Environment for Statistical Computing. (n.d.).

Oti, M., Snel, B., Huynen, M. A., & Brunner, H. G. (2006). Predicting disease genes using protein-protein interactions. *Journal of medical genetics*.

Pastor-Satorras, R., Smith, E., & Solé, R. V. (2003). Evolving protein interaction networks through gene duplication. *Journal of Theoretical Biology*.

Pawson, T. & Linding, R. (2008). Network medicine. *FEBS letters*.

Postuma, R. B. (2014). Prodromal Parkinson's disease–using REM sleep behavior disorder as a window. *Parkinsonism & related disorders*.

Postuma, R. B. & Berg, D. (2016). Advances in markers of prodromal Parkinson's disease. *Nature Reviews Neurology*.

Postuma, R. B., Gagnon, J.-F., Bertrand, J.-A., Génier Marchand, D., & Montplaisir, J. Y. (2015). Parkinson risk in idiopathic REM sleep behavior disorder: preparing for neuroprotective trials. *Neurology*.

Postuma, R. B., Gagnon, J.-F., Vendette, M., Fantini, M. L., Massicotte-Marquez, J., & Montplaisir, J. (2009). Quantifying the risk of neurodegenerative disease in idiopathic REM sleep behavior disorder. *Neurology*.

Postuma, R. B., Lang, A. E., Gagnon, J.-F., Pelletier, A., & Montplaisir, J. Y. (2012). How does parkinsonism start? Prodromal parkinsonism motor changes in idiopathic REM sleep behaviour disorder. *Brain*.

Potashkin, J. A., Santiago, J. A., Ravina, B. M., Watts, A., & Leontovich, A. A. (2012). Biosignatures for Parkinson's disease and atypical parkinsonian disorders patients. *PLoS …*.

Pougach, K., Voet, A., Kondrashov, F. A., Voordeckers, K., Christiaens, J. F., Baying, B., … Verstrepen, K. J. (2014). Duplication of a promiscuous transcription factor drives the emergence of a new regulatory network. *Nature Communications*.

Pringsheim, T., Jette, N., Frolkis, A., & Steeves, T. D. L. (2014). The prevalence of Parkinson's disease: a systematic review and meta-analysis. *Movement disorders : official journal of the Movement Disorder Society*.

Rodriguez-Martinez, A., Ayala, R., Posma, J. M., Neves, A. L., Gauguier, D., Nicholson, J. K., & Dumas, M.-E. (2016). MetaboSignal: a network-based approach for topological analysis of metabotype regulation viametabolic and signaling pathways. *Bioinformatics*.

Rolland, T., Taşan, M., Charloteaux, B., Pevzner, S. J., Zhong, Q., Sahni, N., … Vidal, M. (2014). A proteome-scale map of the human interactome network. *Cell*.

Sanchez, C., Lachaize, C., Janody, F., Bellon, B., Röder, L., Euzenat, J., … Jacq, B. (1999). Grasping at molecular interactions and genetic networks in Drosophila melanogaster using FlyNets, an Internet database. *Nucleic Acids Research*.

Schenck, C. H., Boeve, B. F., & Mahowald, M. W. (2013). Delayed emergence of a parkinsonian disorder or dementia in 81 of older men initially diagnosed with idiopathic rapid eye movement sleep behavior disorder: a 16-year update on a previously reported series. *Sleep medicine*.

Schenck, C. H., Bundlie, S. R., Ettinger, M. G., & Mahowald, M. W. (1986). Chronic behavioral disorders of human REM sleep: a new category of parasomnia. *SLEEP*.

Schenck, C. H., Lee, S. A., Bornemann, M. A. C., & Mahowald, M. W. (2009). Potentially Lethal Behaviors Associated With Rapid Eye Movement Sleep Behavior Disorder: Review of the Literature and Forensic Implications. *Journal of Forensic Sciences*.

Sharma, A., Menche, J., Huang, C. C., Ort, T., Zhou, X., Kitsak, M., … Barabási, A.-L. (2015). A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes in asthma. *Human molecular genetics*.

Signorelli, M., Vinciotti, V., & Wit, E. C. (2016). NEAT: an efficient network enrichment analysis test. *BMC bioinformatics*.

Silverman, E. K. & Loscalzo, J. (2012). Network medicine approaches to the genetics of complex diseases. *Discov Med*.

Sixel-Döring, F., Zimmermann, J., Wegener, A., Mollenhauer, B., & Trenkwalder, C. (2016). The Evolution of REM Sleep Behavior Disorder in Early Parkinson Disease. *SLEEP*.

Solomonoff, R. & Rapoport, A. (1951). Connectivity of random nets. *The Bulletin of Mathematical Biophysics*.

Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., … Robinson, G. E. (2015). Big Data: Astronomical or Genomical? *PLoS Biol*.

Stern, A. M., Schurdak, M. E., Bahar, I., Berg, J. M., & Taylor, D. L. (2016). A Perspective on Implementing a Quantitative Systems Pharmacology Platform for Drug Discovery and the Advancement of Personalized Medicine. *J Biomol Screen*.

Stodden, V., McNutt, M., Bailey, D. H., Deelman, E., Gil, Y., Hanson, B., … Taufer, M. (2016). Enhancing reproducibility for computational methods. *Science*.

Teichmann, S. A. & Babu, M. M. (2004). Gene regulatory network growth by duplication. *Nature Genetics*.

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., … Pontén, F. (2015). Tissue-based map of the human proteome. *Science*.

Vale, T. C., Fernandes do Prado, L. B., do Prado, G. F., Povoas Barsottini, O. G., & Pedroso, J. L. (2016). Rapid Eye Movement Sleep Behavior Disorder in Paraneoplastic Cerebellar Degeneration: Improvement with Immunotherapy. *SLEEP*.

Vázquez, A., Flammini, A., Maritan, A., & Vespignani, A. (2003). Modeling of Protein Interaction Networks. *Complexus*.

Venancio, D. P. & Suchecki, D. (2015). Prolonged REM sleep restriction induces metabolic syndrome-related changes: Mediation by pro-inflammatory cytokines. *Brain, Behavior, and Immunity*.

Venkatesan, K., Rual, J.-F., Vázquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., … Vidal, M. (2009). An empirical framework for binary interactome mapping. *Nature Methods*.

Vetrugno, R., Provini, F., Cortelli, P., Plazzi, G., Lotti, E. M., Pierangeli, G., … Montagna, P. (2004). Sleep disorders in multiple system atrophy: a correlative video-polysomnographic study. *Sleep medicine*.

Vidal, M., Cusick, M. E., & Barabási, A.-L. (2011). Interactome networks and human disease. *Cell*.

Wang, R.-S., Maron, B. A., & Loscalzo, J. (2015). Systems medicine: evolution of systems biology from bench to bedside. *Wiley interdisciplinary reviews. Systems biology and medicine*.

Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of ”small-world” networks. *Nature*.

Wu, P., Yu, H., Peng, S., Dauvilliers, Y., Wang, J., Ge, J., … Zuo, C. (2014). Consistent abnormalities in metabolic network activity in idiopathic rapid eye movement sleep behaviour disorder. *Brain*.

Zanzoni, A., Soler-López, M., & Aloy, P. (2009). A network medicine approach to human disease. *FEBS letters*.

Zhang, W. J., Shang, X. L., Peng, J., Zhou, M. H., & Sun, W. J. (2017). Expression of prion protein in the cerebrospinal fluid of patients with Parkinson’s disease complicated with rapid eye movement sleep behavior disorder. *Genet. Mol. Res.*

Zhang, X., Sun, X., Wang, J., Tang, L., & Xie, A. (2017). Prevalence of rapid eye movement sleep behavior disorder (RBD) in Parkinson's disease: a meta and meta-regression analysis. *Neurol. Sci.*

9 Supplementary information

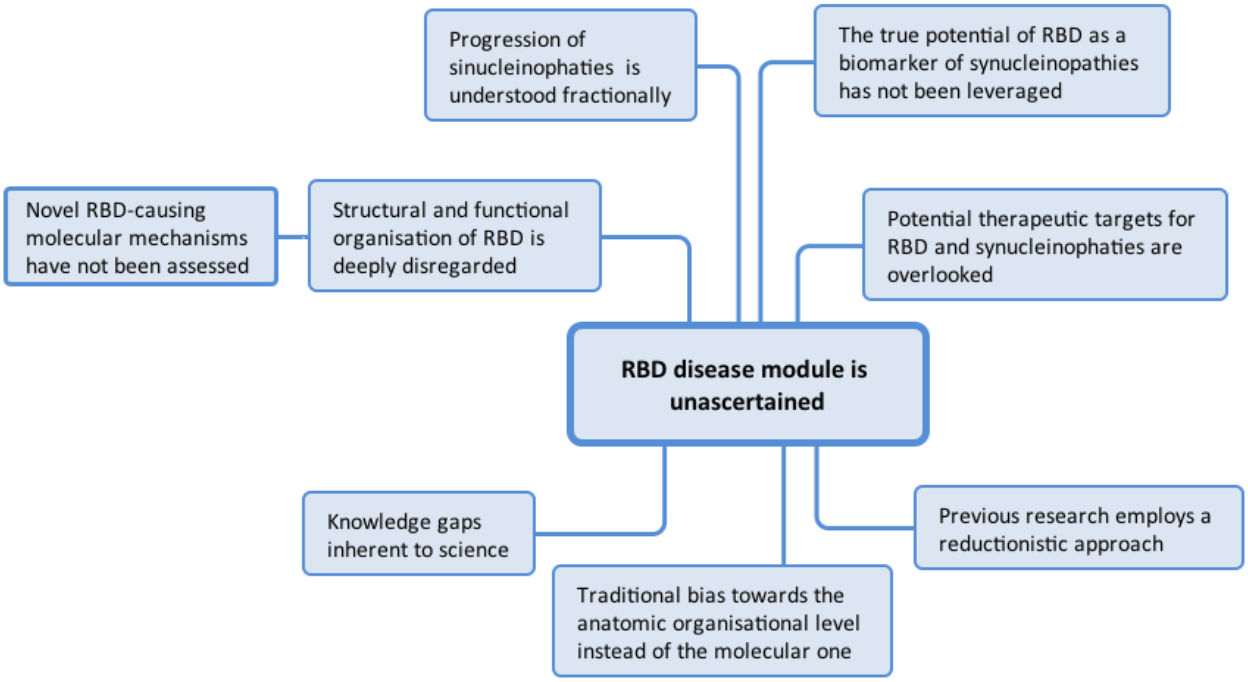

Figure S1: Problem tree showing the causes (bottom squares), generating the research problem (central square), which in turn produces some consequences (upper squares).

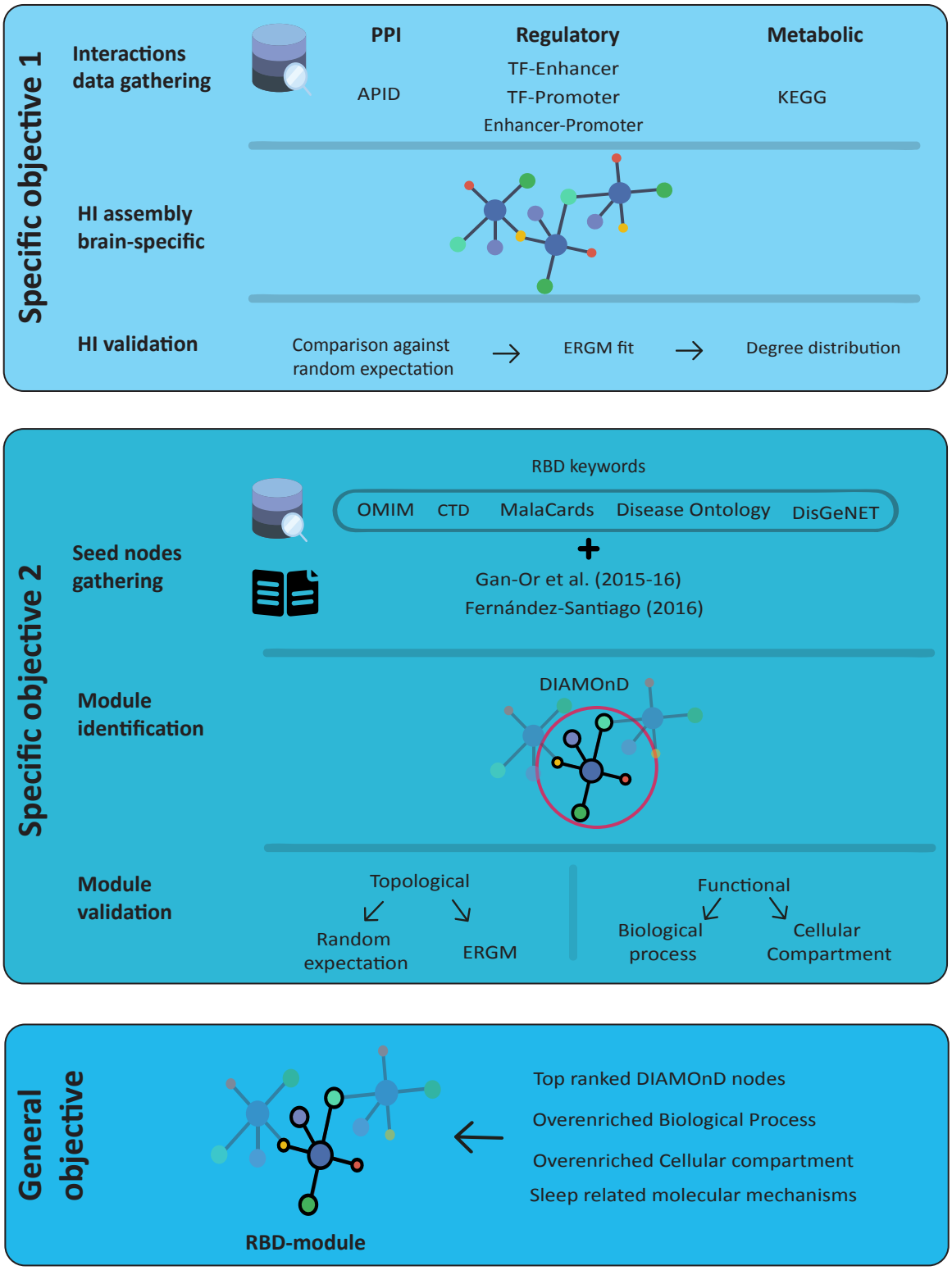

Figure S2: Schematic representation of the methodological workflow followed in the present research.

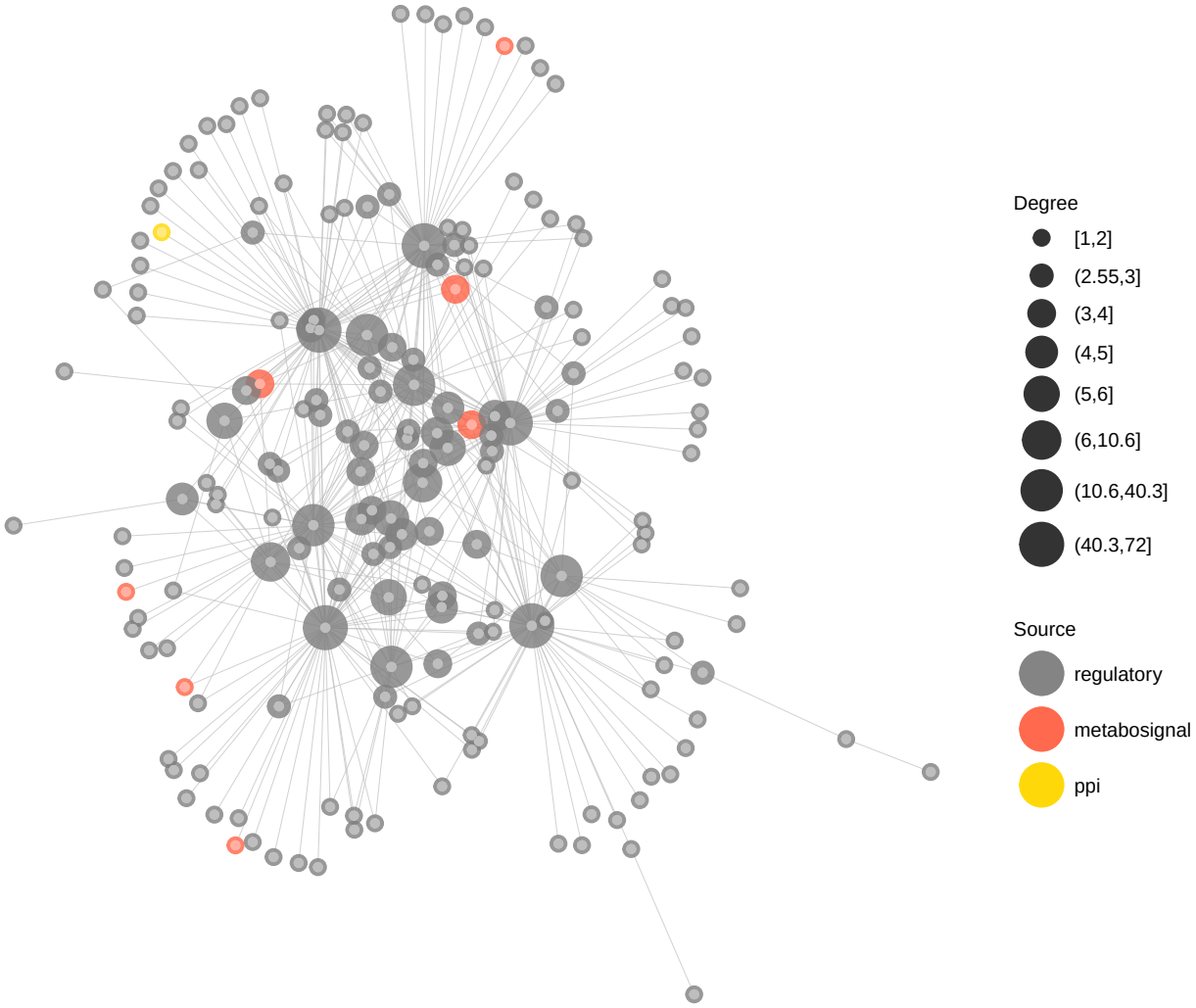

Figure S3: Random sample of 300 nodes and their respective edges from the HI version 1, drew only for the sake of aesthetics. Regulatory interactions dominate the nodes in the interactome and also represent most of the hubs.

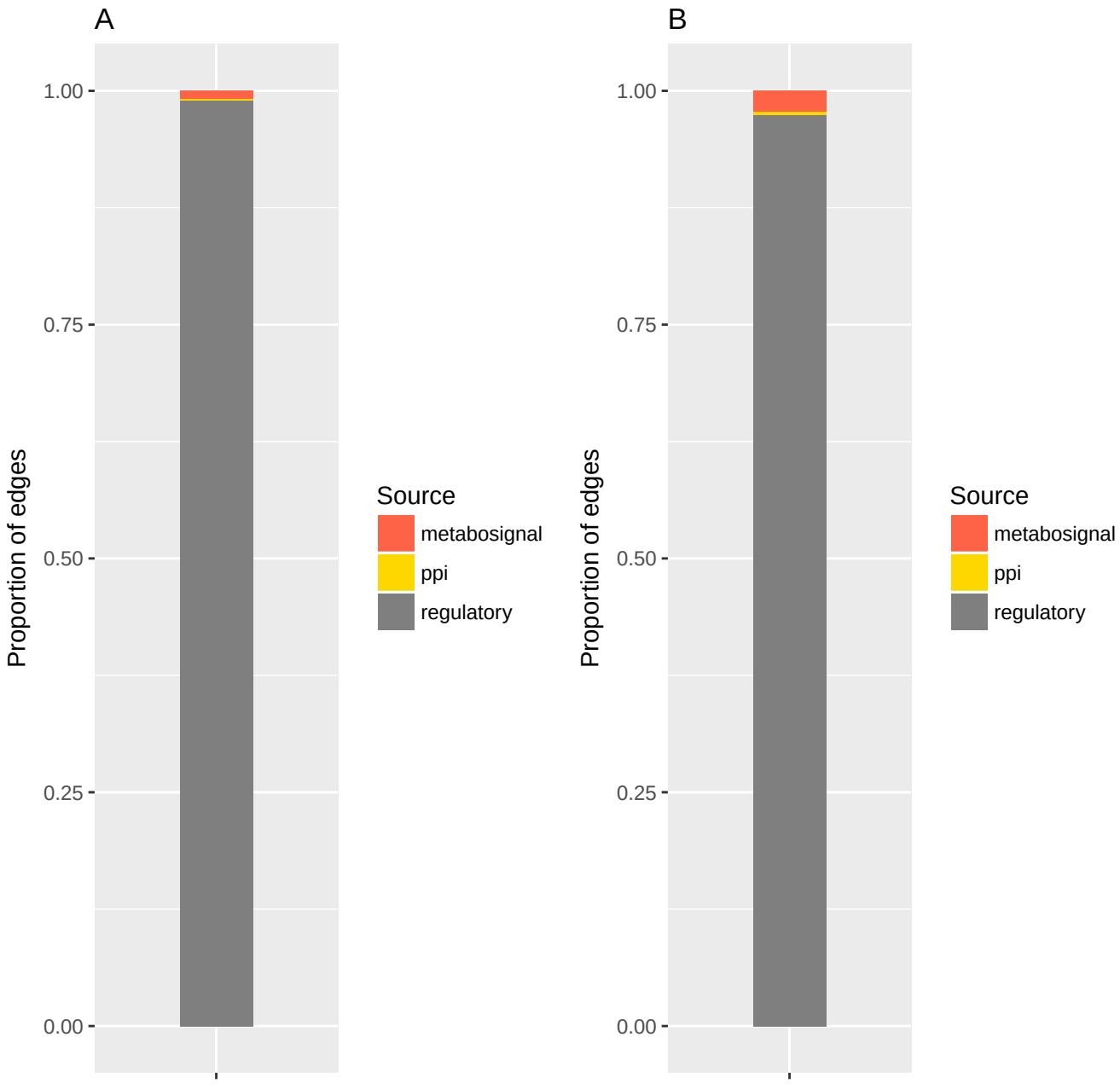

Figure S4: Proportion of edges per source in the HI 1, including all regulatory interactions, (A) and HI 2, including only whole-brain regulatory interactions, (B). Despite the regulatory overrepresentation persist when employing only whole-brain regulatory interactions, the dimensionality of the interactome reduces by a factor of 2.5, compared against HI 1 (table 1).

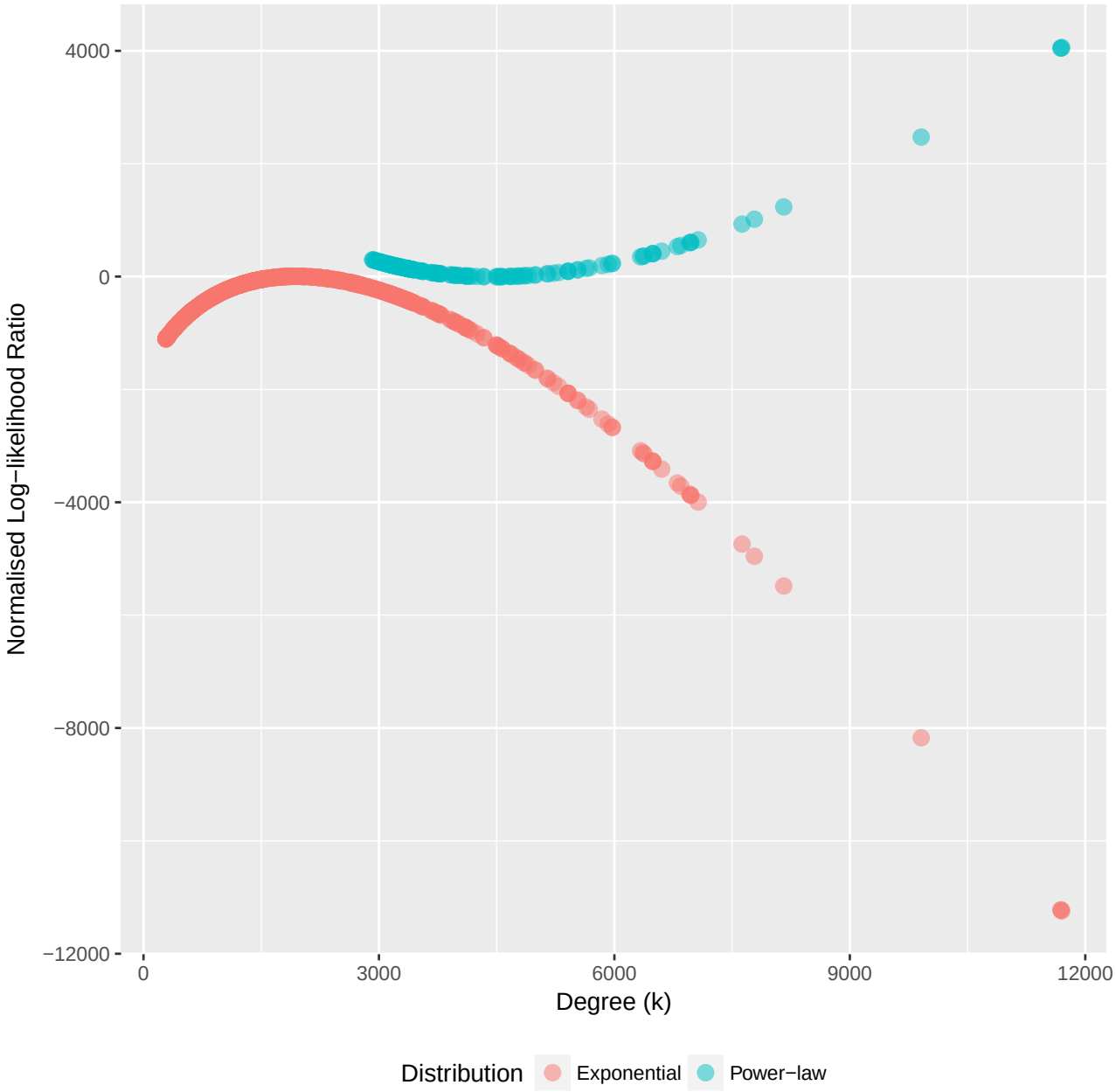

Figure S5: Comparison of distributions to explain the $P(k)_{HI}$. The power-law is significantly closer to $P(k)_{HI}$ than the Poisson distribution (one-sided p_{value} = 1.198427e-07, two-sided p_{value} = 2.396854e-07), as seen in the log-likelihood ratio($\mathcal{R}$), which goes towards positive values (blue line). Similarly, when comparing exponential distribution against the Poisson, the former is significantly more plausible than the latter (two-sided p_{value} = 4.568209e-45), represented as a trend of $\mathcal{R}$ towards negative values (red line) (Clauset, Shalizi, & Newman, 2009).

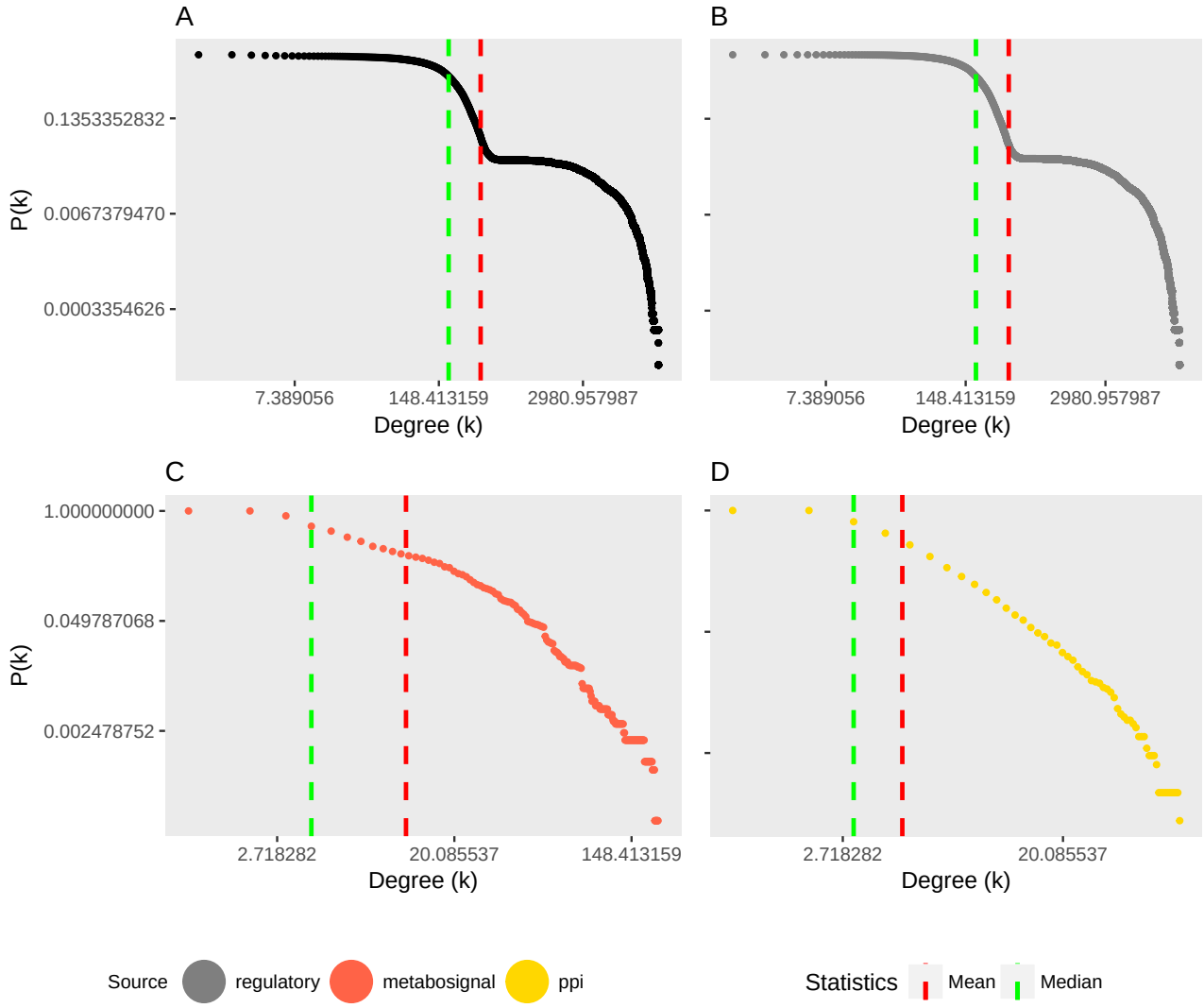

Figure S6: Log-log plots for the cumulative degree distribution ($P(k)$) for the HI 2 (A), regulatory interactions (B), metabosignal interactions (C) and PPI (D). The bimodal distribution of the HI is inherited from regulatory interactions, as both, PPI and metabosignal follows an unimodal distribution.

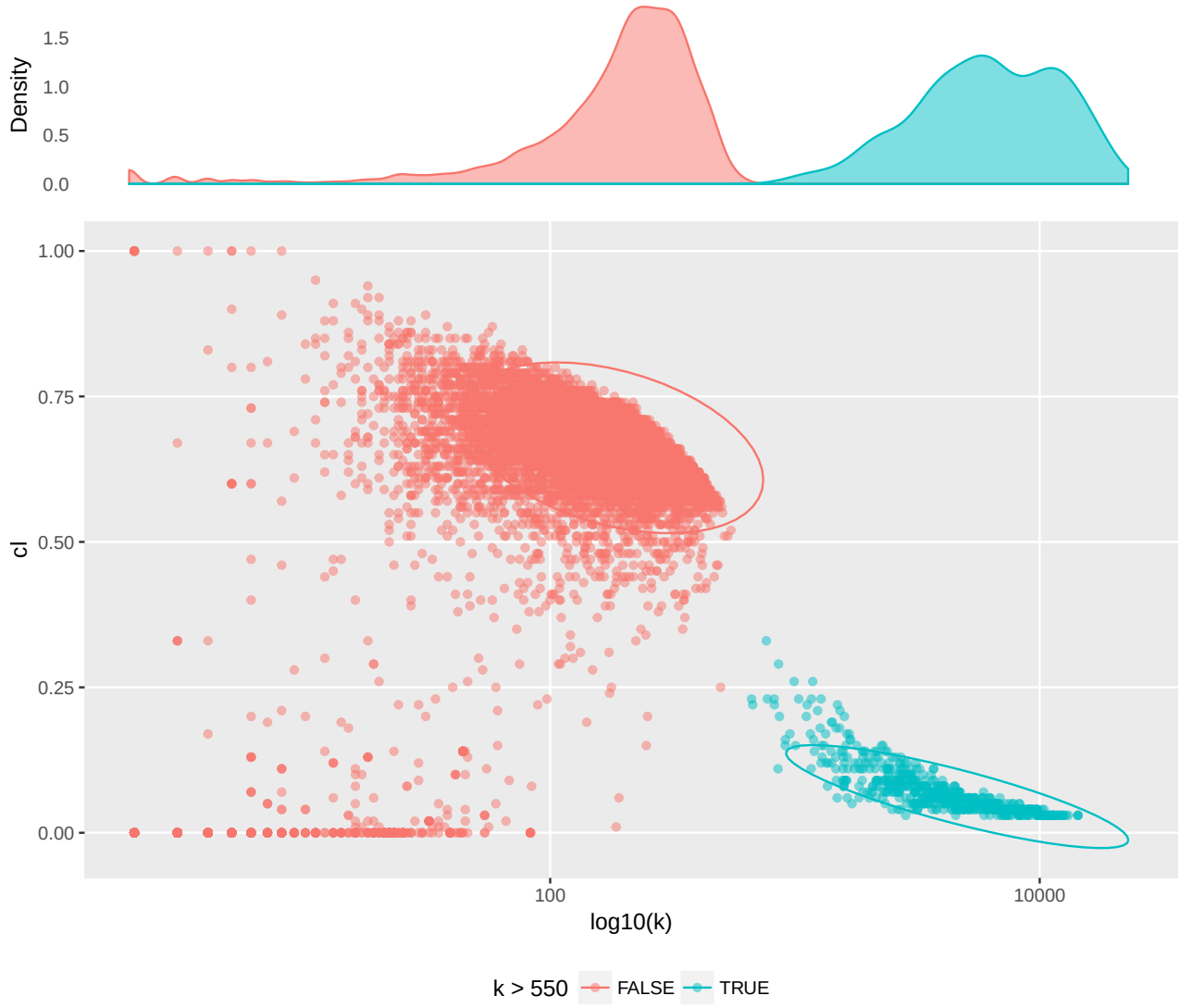

Figure S7: Clustering coefficient (cl) is successfully able to separate the two type of hubs in regulatory interactions: intramodular or party hubs (red), with a high clustering coefficient keeping integrity of modules, and intermodular or date hubs (blue), with a lower clustering coefficient and a tendency to connect modules. The analogy by (Fraser, 2005) with exon shuffling exemplify very well the evolutionary scenario where intermodular hubs in regulatory interactions evolve through the duplication-acquisition model (Ferreira et al., 2013) here supported, *"In an abstract sense,... The idea of evolutionary innovation occurring through exon shuffling, a process in which the swapping between different genes of DNA segments encoding conserved protein domains allows new proteins to evolve, by creating new domain combinations... Functional modules may be similar to protein domains, which, once established, are usually best left intact; new combinations of and connections between modules (or domains) may lead to advantageous changes more often than would changes that disrupt modules (or domains)... In some sense, it is similar to the idea of evolution occurring through the co-option of existing modules for new purposes"*.

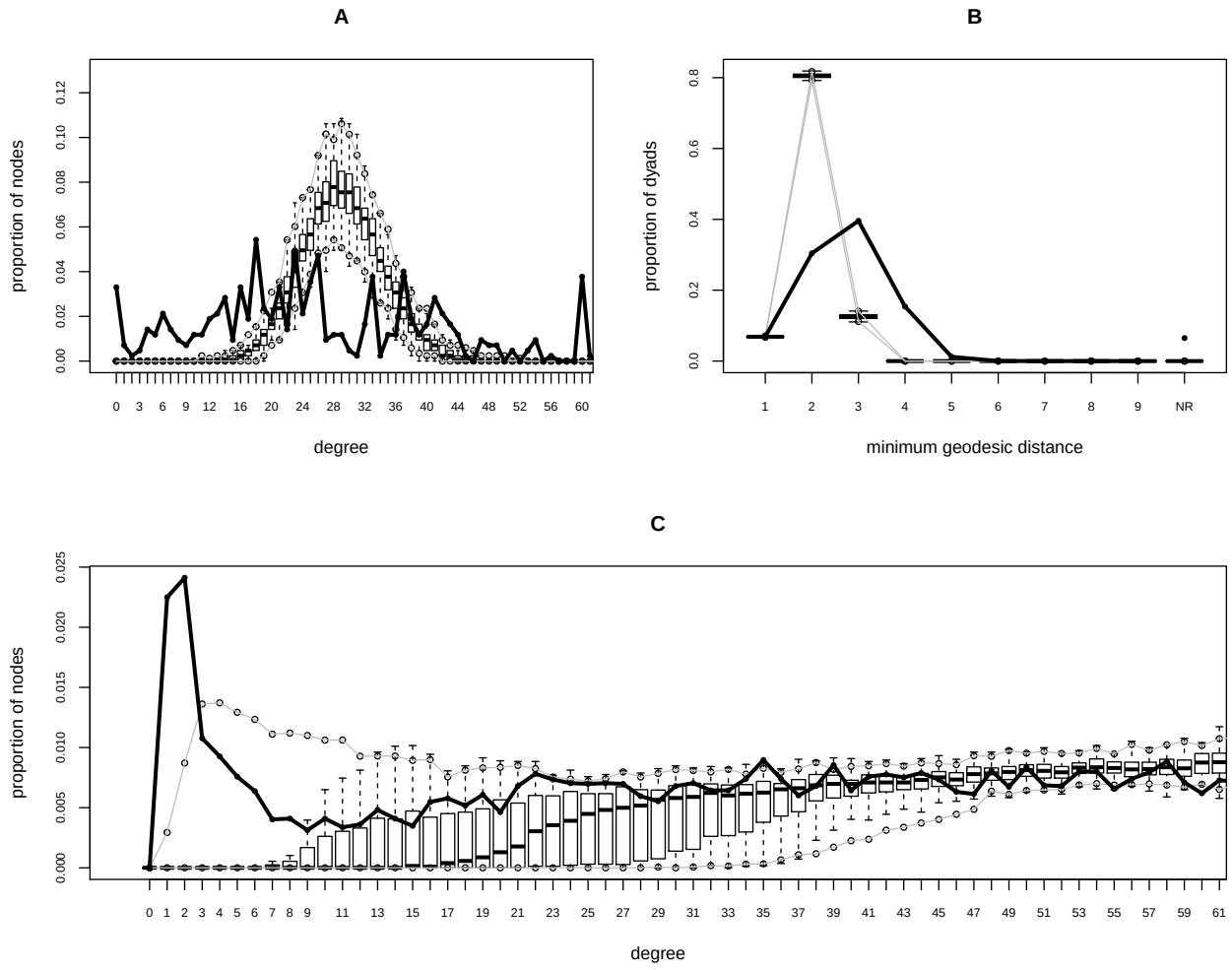

Figure S8: Goodness-of-fit test for the RBD-module by fitting an ERGM and comparing the degree distribution (A) and the geodesic distance (B). Additionally, B shows that most nodes in the RBD-module are reachable (Non Reachable ~ 0). Goodness-of-fit test for the HI comparing the degree distribution is shown in C. The bold black line is the observed value and the boxplots represent the simulated values showing minimums and maximums, as well as the 10^{th} and 90^{th} quantiles. In all cases, the overall poor fitting between the simulated and the observed value is quite remarkable, supporting the biological relevance of the RBD-module and the HI. For the sake of aesthetics, only the first 60 values of d are shown.

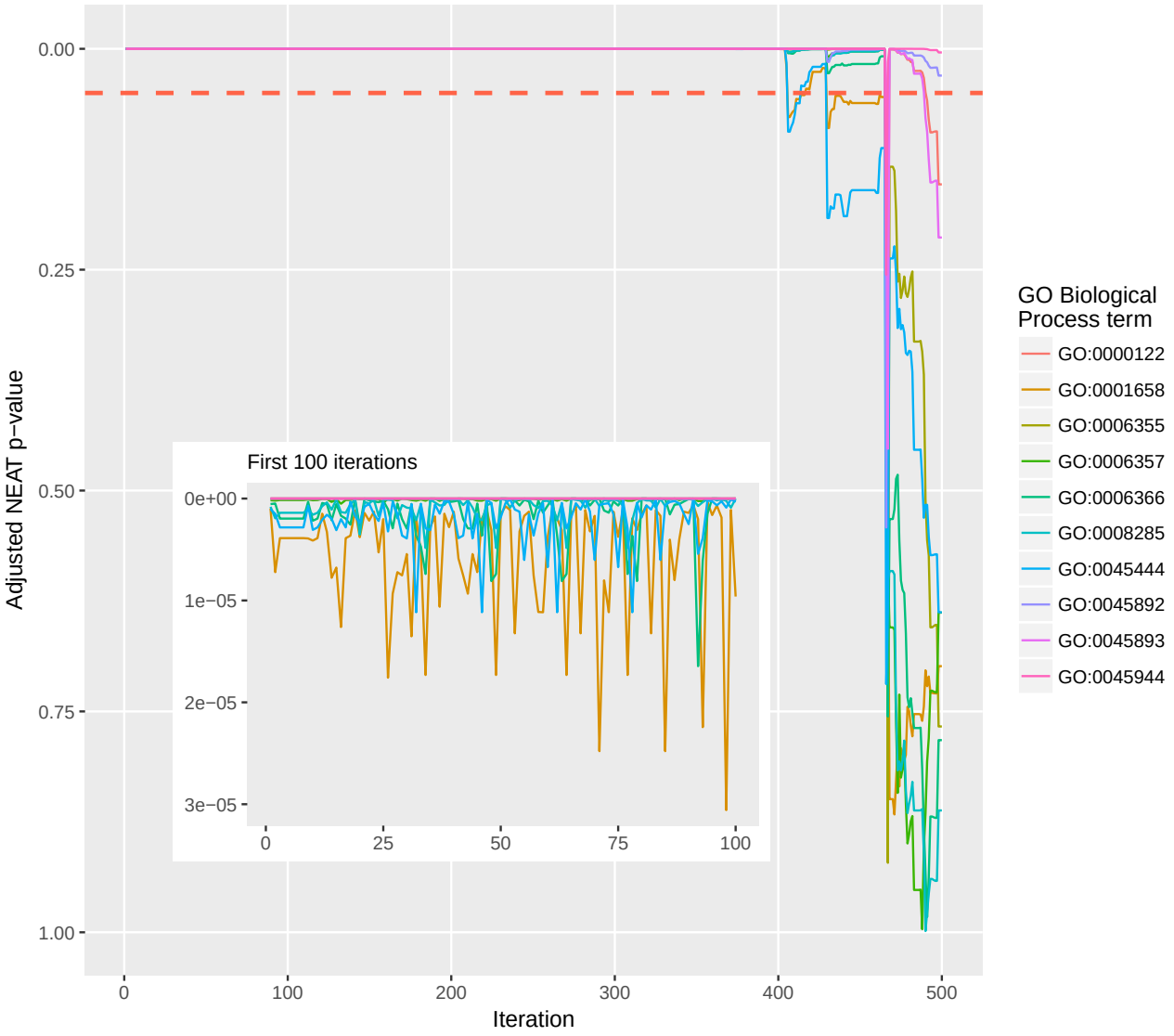

Figure S9: GO Biological process enrichment of DIAMOnD nodes. Behaviour of the enrichment significance for the ten most enriched Gene Ontology Biological processes in the RBD seed nodes as DIAMOnD nodes are added to the RBD module. The red dotted line represents the p_{value} = 0.05, thus, by extrapolating to the abscissa, the first 400 predicted DIAMOnD nodes were included to the RBD-module as novel entities related to the RBD aetyology, matching the drop in the enrichment significance. The first $\sim$ 400 iterations are anchored to 0 because they are very significant (*e.g.* 1×10^{-170}), as seen in the inset of the first 100 nodes with a rescaled ordinate. The ordinates represent the B-H adjusted p_{value}. A fully interactive version of the plot, for further inspection, is available here.

Table S1: The main function of the first ten intermodular and intramodular hubs in regulatory interactions. Most of these nodes are TF, further supporting the evolutionary scenario of duplication-acquisition (Ferreira et al., 2013) in the regulatory interactions of the HI.

	Entrez gene ID	Gene name	Function
Intermodular	3725	Jun proto-oncogene, AP-1 transcription factor subunit	TF
	6667	Sp1 transcription factor	TF
	2313	Fli-1 proto-oncogene, ETS transcription factor	TF
	7020	Transcription factor AP-2 alpha	TF
	5078	Paired box 4	TF
	5451	POU class 2 homeobox 1	TF
	5669	Pregnancy specific beta-1-glycoprotein 1	Antibody
	199699	DAN domain BMP antagonist family member 5	Morphogen
	3172	Hepatocyte nuclear factor 4 alpha	TF
	6688	Spi-1 proto-oncogene	TF
Intramodular	8266	Ubiquitin like 4A	Ubiquitin like
	26973	Cysteine and histidine rich domain containing 1	TF
	1272	Contactin 1	Antibody
	7010	TEK receptor tyrosine kinase	Receptor
	2776	G protein subunit alpha q	Signal transduction
	64412	GDNF inducible zinc finger protein 1	TF
	3673	Integrin subunit alpha 2	MEC support
	51119	SBDS, ribosome assembly guanine nucleotide exchange factor	TF
	7675	Zinc finger protein 121	TF
	23657	Solute carrier family 7 member 11	Transport

Table S2: RBD seed genes collected from databases and literature. Inferred C = Inferred association via a curated chemical interaction from the CTD database. When separated by a "/" then, several sources or evidence are supporting the gene.

Gene	Evidence	Source
ACHE	Inferred	MalaCards humandisease
BDNF	Inferred C	CTD
CASP7	Inferred C	CTD
SCARB2	linkage	(Gan-Or, Girard, et al., 2015)
CYP2D6	Inferred C	CTD
CYP3A4	Inferred C	CTD
DRD2	Inferred C	CTD
GBA	linkage	(Gan-Or, Mirelman, et al., 2015)
GCNT2	literature	DisGeNET
HCRT	literature/Inferred	DisGeNET/MalaCards humandisease
HLA-DQB1	TextMining	HuGE
HTR1A	Inferred C	CTD
HTR2A	Inferred C	CTD
HTR2C	Inferred C	CTD
MAPT	linkage	(Gan-Or, Girard, et al., 2015)
PARK2	literature	DisGeNET
PSG5	literature	DisGeNET
RPS27A	Inferred	MalaCards humandisease
SLC6A3	literature/Inferred	MalaCards humandisease
SLC6A4	Inferred C	CTD
SNCA	literature/Inferred	MalaCards humandisease
TCOF1	Inferred	MalaCards humandisease
NR0B2	Inferred C	CTD
REM1	literature	DisGeNET
FNDC4	Inferred C	CTD
LRRK2	Inferred	MalaCards human disease

Table S3: Ten most enriched GO terms in the RBD-module organised by ontology. Most Biological processes (BP) and Cellular Compartments (CC) are immune and signalling related.

GOID	Term	Ontology
GO:0002232	leukocyte chemotaxis involved in inflammatory response	BP
GO:0002673	regulation of acute inflammatory response	BP
GO:0002720	positive regulation of cytokine production involved in immune response	BP
GO:0002879	positive regulation of acute inflammatory response to non-antigenic stimulus	BP
GO:0090276	regulation of peptide hormone secretion	BP
GO:1904322	cellular response to forskolin	BP
GO:0018105	peptidyl-serine phosphorylation	BP
GO:0007189	adenylate cyclase-activating G-protein coupled receptor signaling pathway	BP
GO:0006171	cAMP biosynthetic process	BP
GO:0048009	insulin-like growth factor receptor signaling pathway	BP
GO:0005891	voltage-gated calcium channel complex	CC
GO:0005834	heterotrimeric G-protein complex	CC
GO:0005886	plasma membrane	CC
GO:1990454	L-type voltage-gated calcium channel complex	CC
GO:0005942	phosphatidylinositol 3-kinase complex	CC
GO:0005943	phosphatidylinositol 3-kinase complex, class IA	CC
GO:0000159	protein phosphatase type 2A complex	CC
GO:0032281	AMPA glutamate receptor complex	CC
GO:1902494	catalytic complex	CC
GO:0034704	calcium channel complex	CC